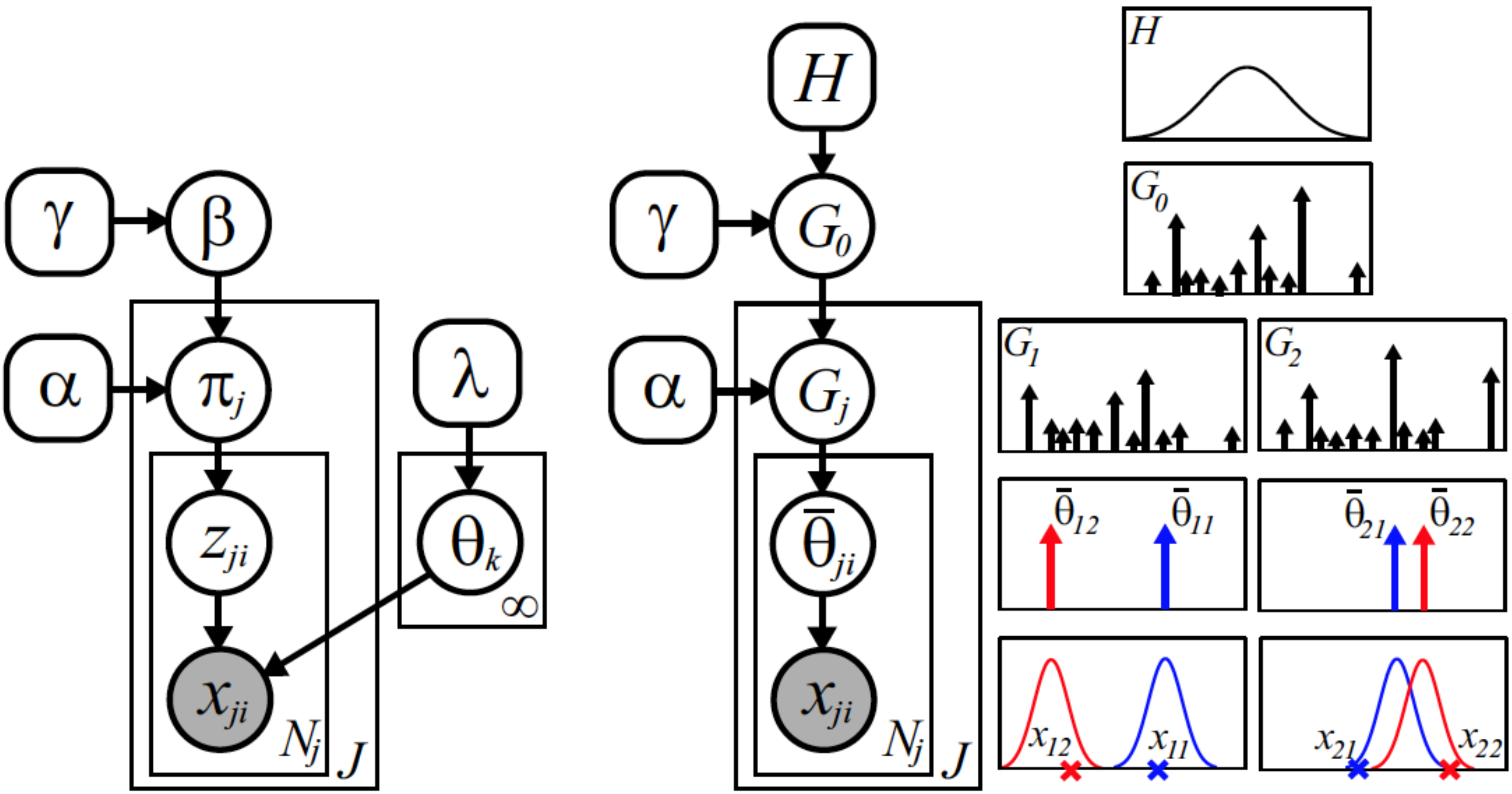
Applied Bayesian Nonparametrics

Special Topics in Machine Learning
Brown University CSCI 2950-P, Fall 2011

October 4: Hierarchical Dirichlet Processes in Computer Vision

Hierarchical Dirichlet Process



Hierarchical Dirichlet Process

$$G_0(\theta) = \sum_{k=1}^{\infty} \beta_k \delta(\theta, \theta_k)$$

$$\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma)$$

$$\theta_k \sim H(\lambda) \qquad k = 1, 2, \ldots$$

$$G_j(\theta) = \sum_{t=1}^{\infty} \widetilde{\pi}_{jt} \delta(\theta, \widetilde{\theta}_{jt})$$

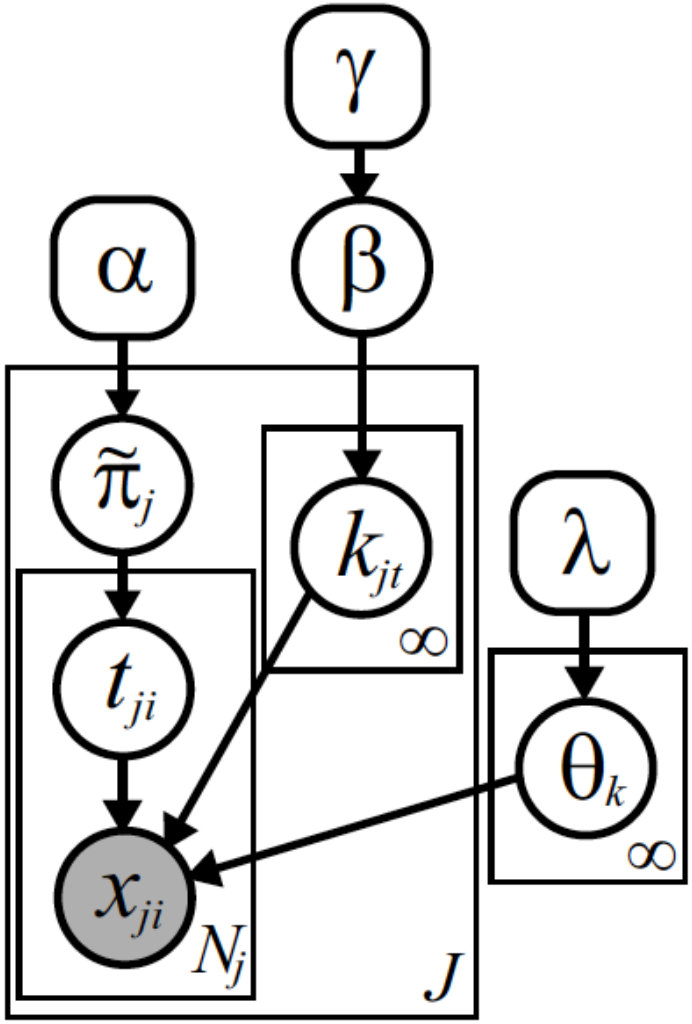
$$\widetilde{\boldsymbol{\pi}}_j \sim \mathrm{GEM}(\alpha)$$

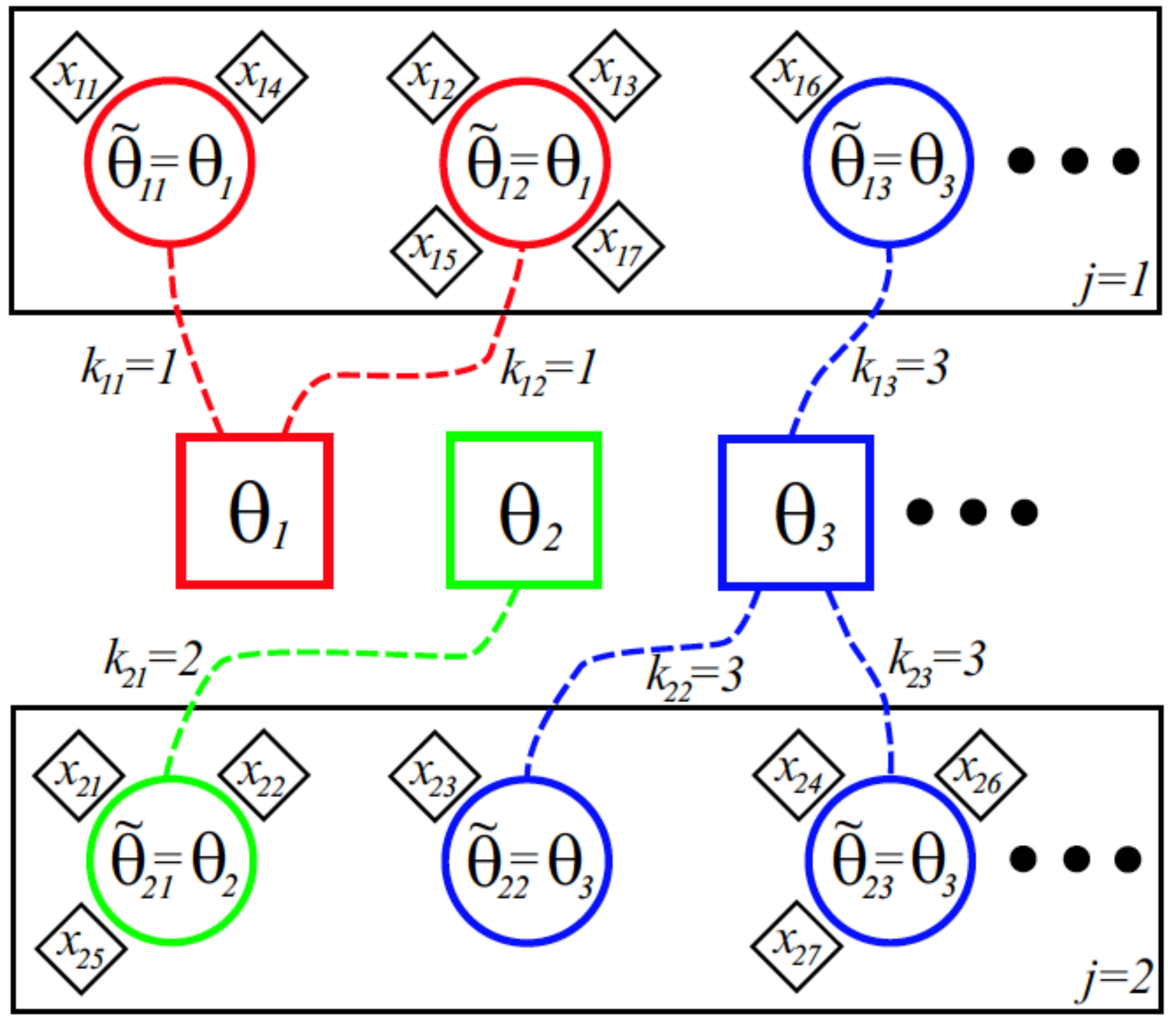
$$\widetilde{\theta}_{jt} \sim G_0 \qquad t = 1, 2, \ldots$$

$$G_j(\theta) = \sum_{k=1}^{\infty} \pi_{jk} \delta(\theta, \theta_k)$$

$$\pi_{jk} = \sum_{t | k_{jt} = k} \widetilde{\pi}_{jt}$$

Chinese Restaurant Franchise



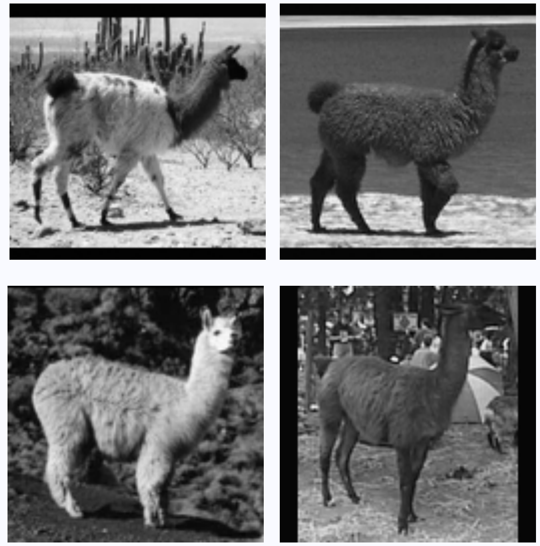


$$p(t_{ji} \mid t_{j1}, \ldots, t_{ji-1}, \alpha) \propto \sum_t N_{jt} \delta(t_{ji}, t) + \alpha \delta(t_{ji}, \bar{t})$$

$$p(k_{jt} \mid \mathbf{k}_1, \ldots, \mathbf{k}_{j-1}, k_{j1}, \ldots, k_{jt-1}, \gamma) \propto \sum_k M_k \delta(k_{jt}, k) + \gamma \delta(k_{jt}, \bar{k})$$

Visual Object Categorization

Bicycles



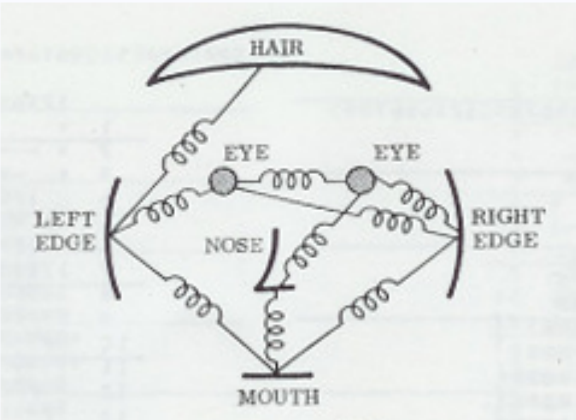
Llamas

Cannons

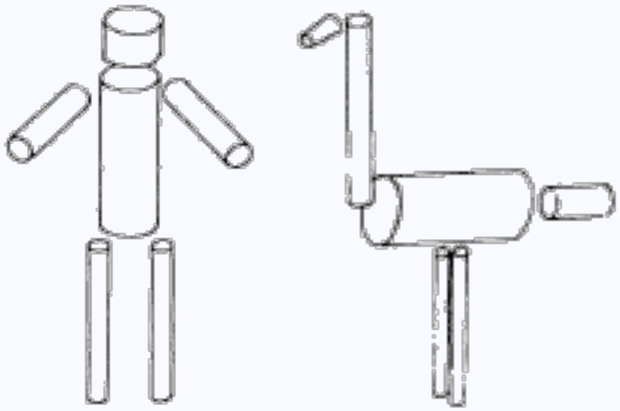
GOALS:

- Visually *recognize* and *localize* object categories
- Robustly *learn* appearance models from few examples
  - Use hierarchical models to *transfer* knowledge among categories
  - Nonparametric, *Dirichlet process* prior gives flexibility

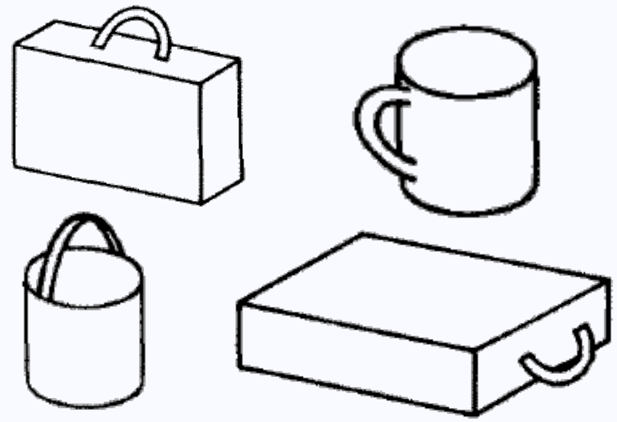
Part-Based Models for Objects



Pictorial Structures
Fischler & Elschlager, 1973

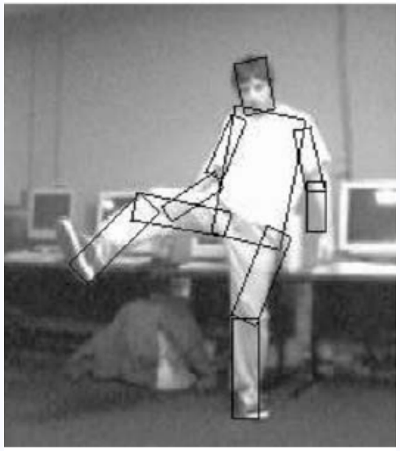


Generalized Cylinders
Marr & Nishihara, 1978

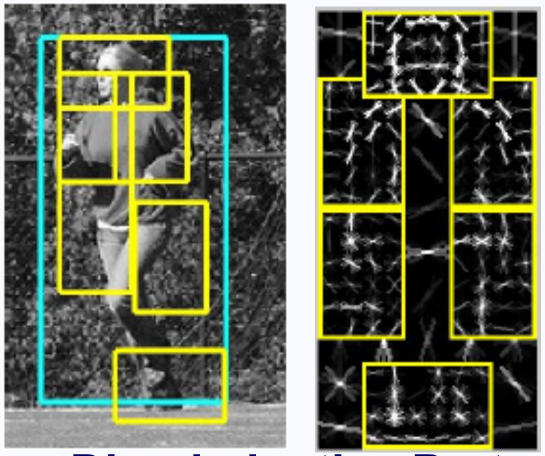


Recognition by Components
Biederman, 1987

Constellation Model
Perona, Weber, Welling, Fergus, Fei-Fei, 2000 to …

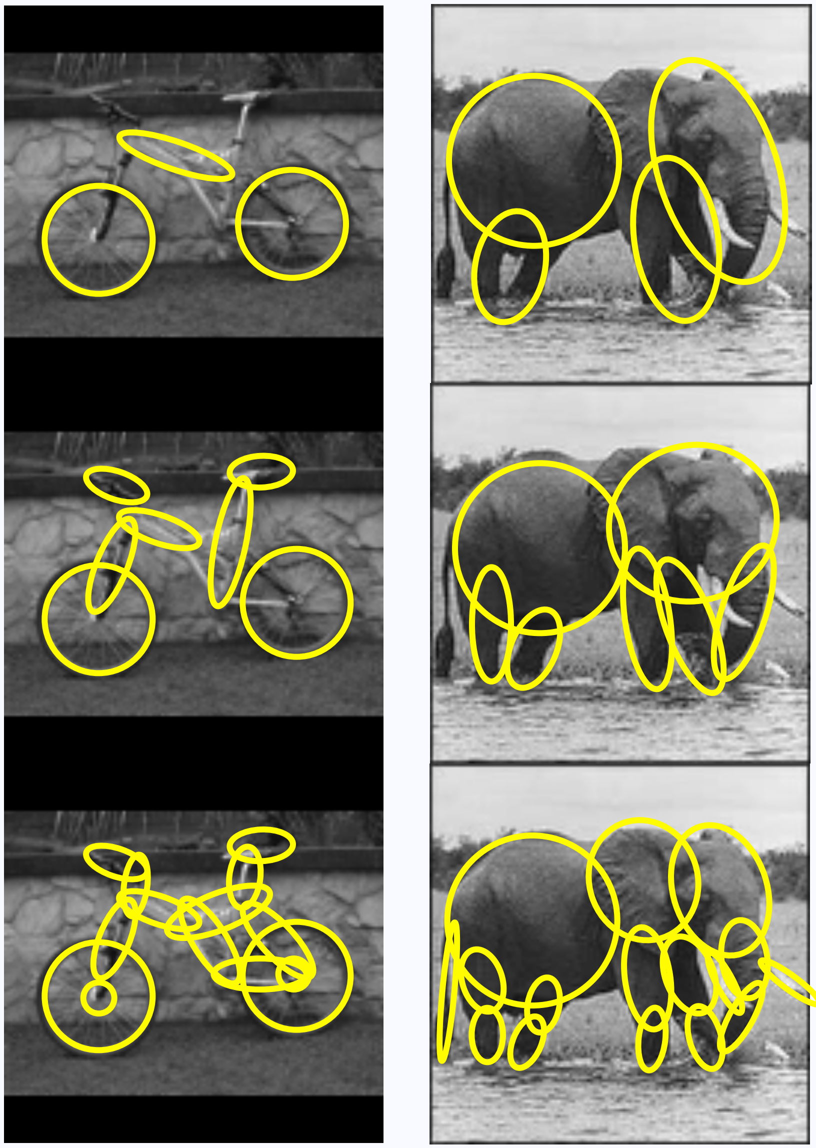


Efficient Matching
Felzenszwalb & Huttenlocher, 2005



Discriminative Parts
Felzenszwalb, McAllester, Ramanan, 2008 to …

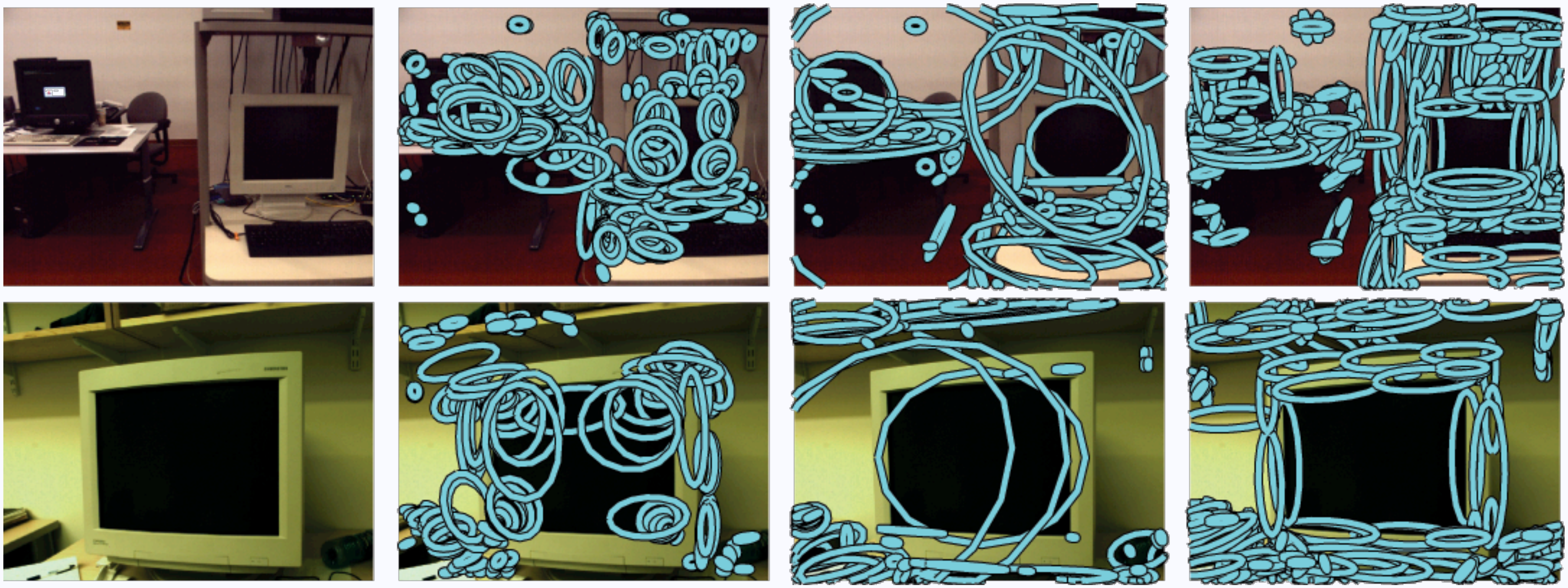
Counting Objects & Parts



How many parts?

How many objects?

From Images to Features



Affinely Adapted Harris Corners

Maximally Stable Extremal Regions

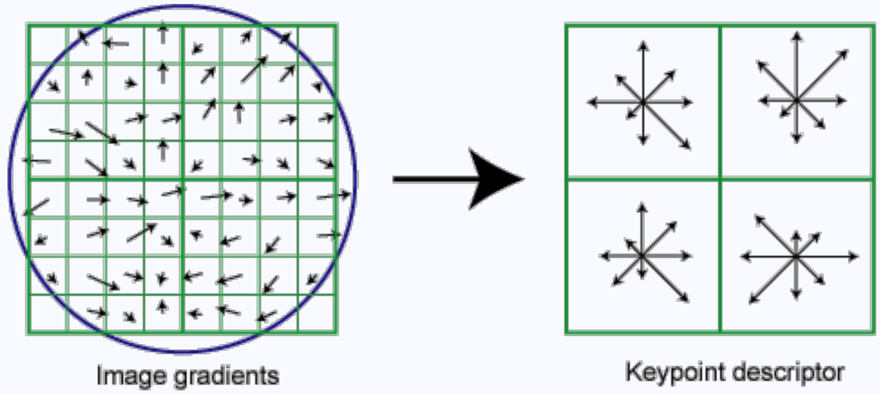
Linked Sequences of Canny Edges

- Some invariance to lighting & pose variations
- Dense, multiscale, over-segmentation of image

A Discrete Feature Vocabulary

SIFT Descriptors

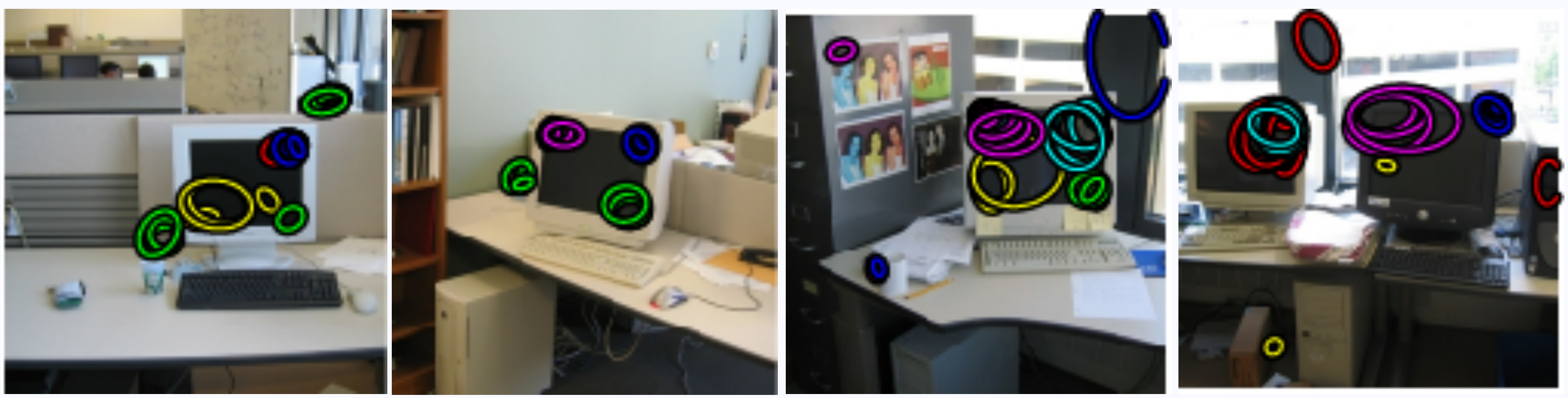
- Normalized histograms of orientation energy
- Compute ~1,000 word dictionary via K-means
- Map each feature to nearest *visual word*



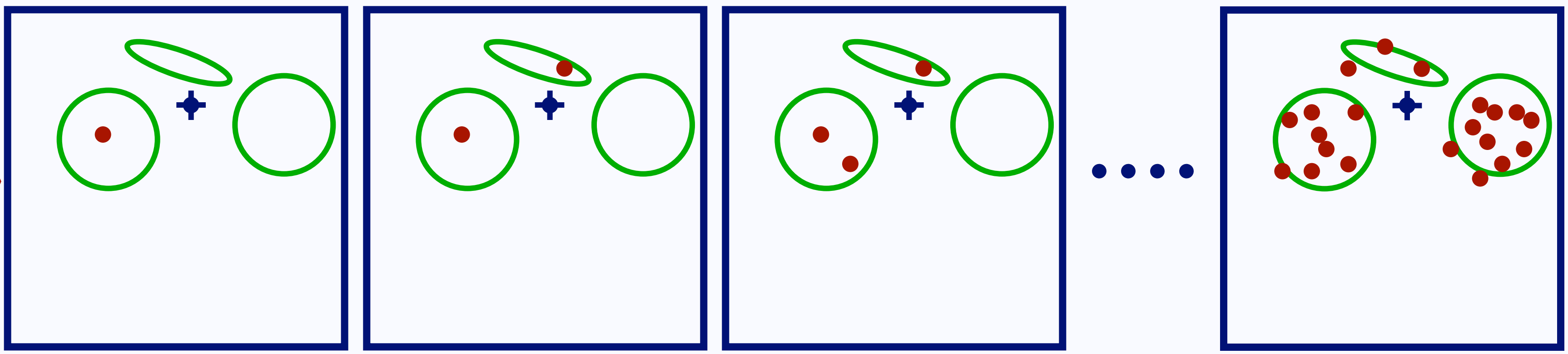
Lowe, IJCV 2004

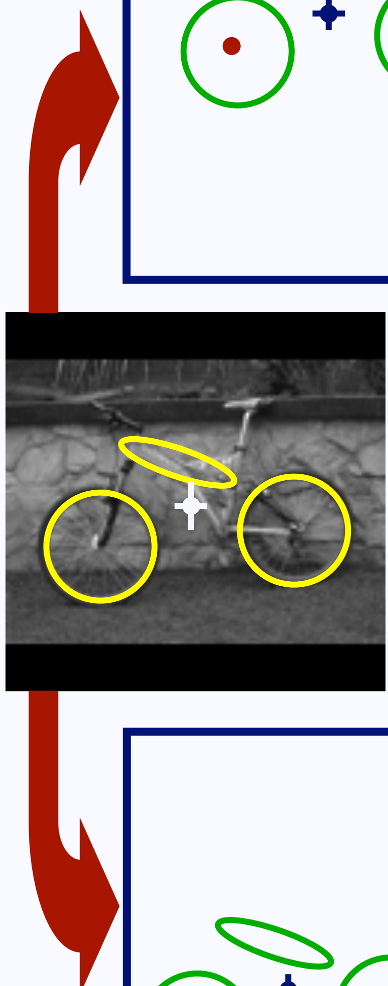
$w_{ji} \longrightarrow$ appearance of feature i in image j

$v_{ji} \longrightarrow$ 2D position of feature i in image j



Generative Model for Objects

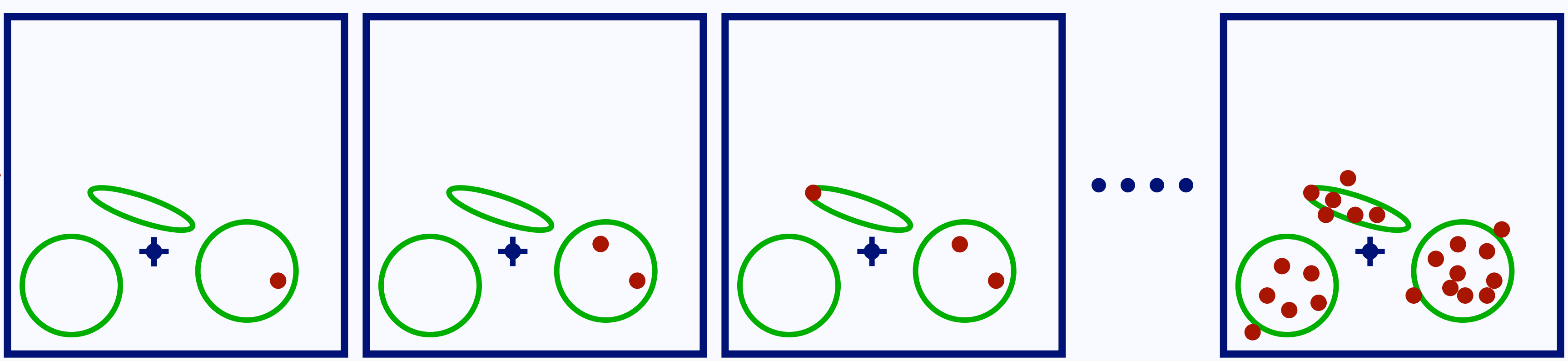




For each image: Sample a reference position

For each feature:

- Randomly choose one part
- Sample from that part's feature distribution



Objects as Mixture Models

- For a fixed reference position, our generative model is equivalent to a finite mixture model:

$$p(w_{ji}, v_{ji} | \rho_j) = \sum_{k=1}^{K} \pi_k \underbrace{\eta_k(w_{ji})}_{\text{Pr(appearance | part)}} \underbrace{\mathcal{N}(v_{ji}; \mu_k + \rho_j, \Lambda_k)}_{\text{Pr(position | part)}}$$

Feature appearance

Feature position

Pr(part)

- How many parts should we choose?
  - ➢ Too few reduces model accuracy
  - ➢ Too many causes overfitting & poor generalization

Objects as Distributions

$$p(w_{ji}, v_{ji} | \rho_j) = \sum_{k=1}^{\infty} \pi_k \underbrace{\eta_k(w_{ji})}_{\text{Pr(appearance | part)}} \underbrace{\mathcal{N}(v_{ji}; \mu_k + \rho_j, \Lambda_k)}_{\text{Pr(position | part)}}$$

Feature appearance

Feature position

DP prior

- Parts are defined by *parameters*, which encode distributions on visual features:

$$\theta_k = \{\eta_k, \mu_k, \Lambda_k\}$$

- Objects are defined by *distributions* on the infinitely many potential part parameters:

$$G(\theta) = \sum_{k=1}^{\infty} \pi_k \delta(\theta, \theta_k) \qquad \boldsymbol{\pi} \sim \text{Stick}(\alpha)$$

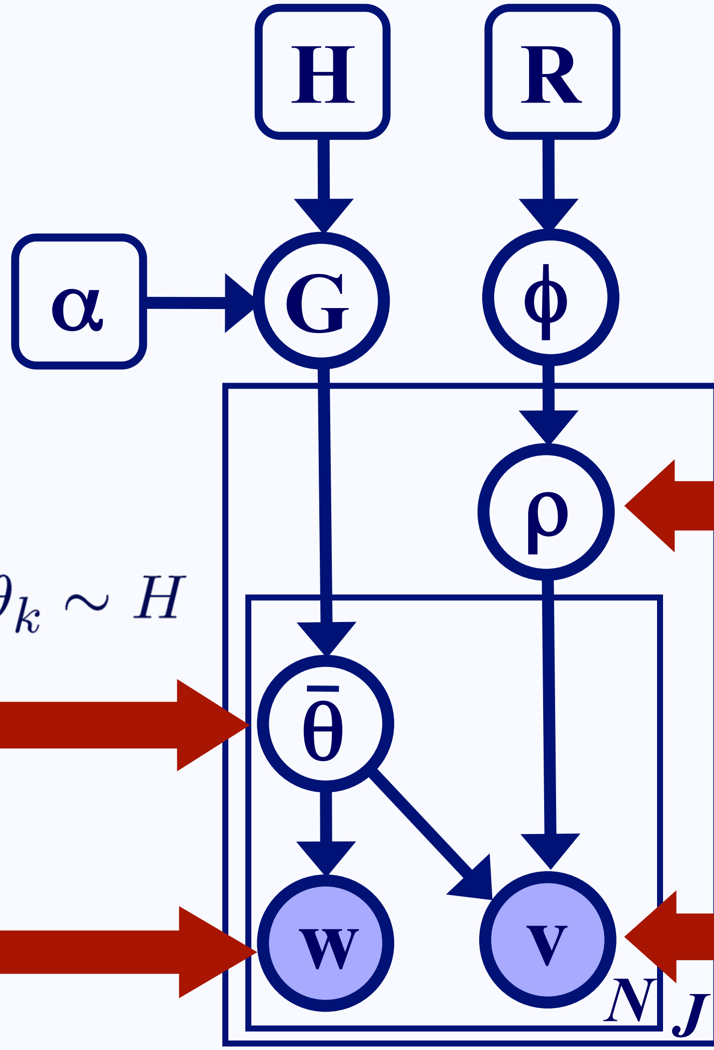
Dirichlet Process Object Model

Part-based object model sampled from DP prior:

$$G \sim \mathrm{DP}(\alpha, H)$$

$$G(\theta) = \sum_{k=1}^{\infty} \pi_k \delta(\theta, \theta_k)$$

$$\boldsymbol{\pi} \sim \mathrm{Stick}(\alpha) \qquad \theta_k \sim H$$



For each of N features, sample *part parameters:*

$$\bar{\theta}_{ji} \sim G(\theta)$$

Sample multinomial *feature appearance:*

$$w_{ji} \sim \bar{\eta}_{ji}(w)$$

$$\bar{\theta}_{ji} = \{\bar{\eta}_{ji}, \bar{\mu}_{ji}, \bar{\Lambda}_{ji}\}$$

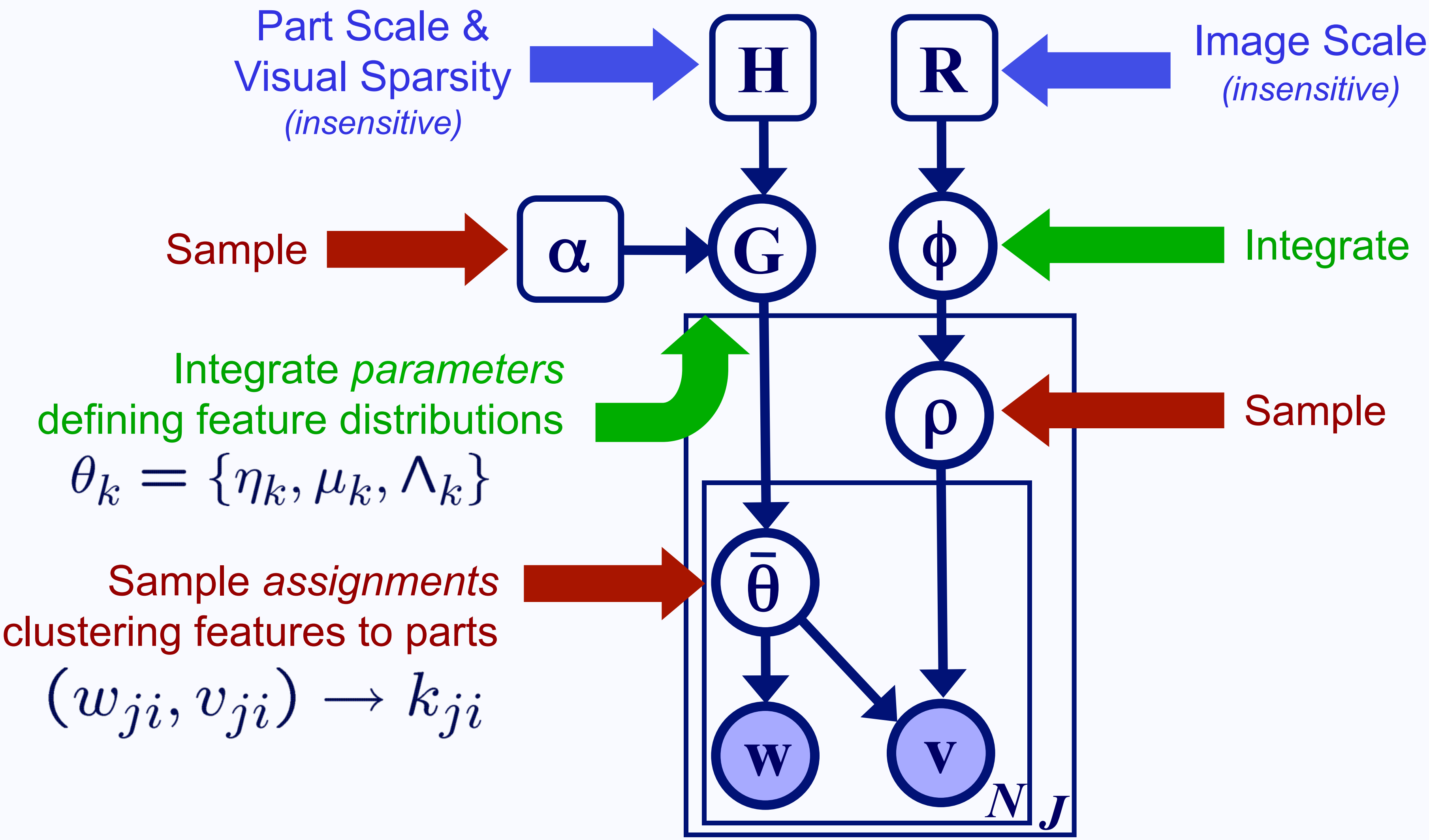
For each of J images, sample a *reference position:*

$$\rho_j \sim \mathcal{N}(\rho; \phi)$$

Sample Gaussian *feature position:*

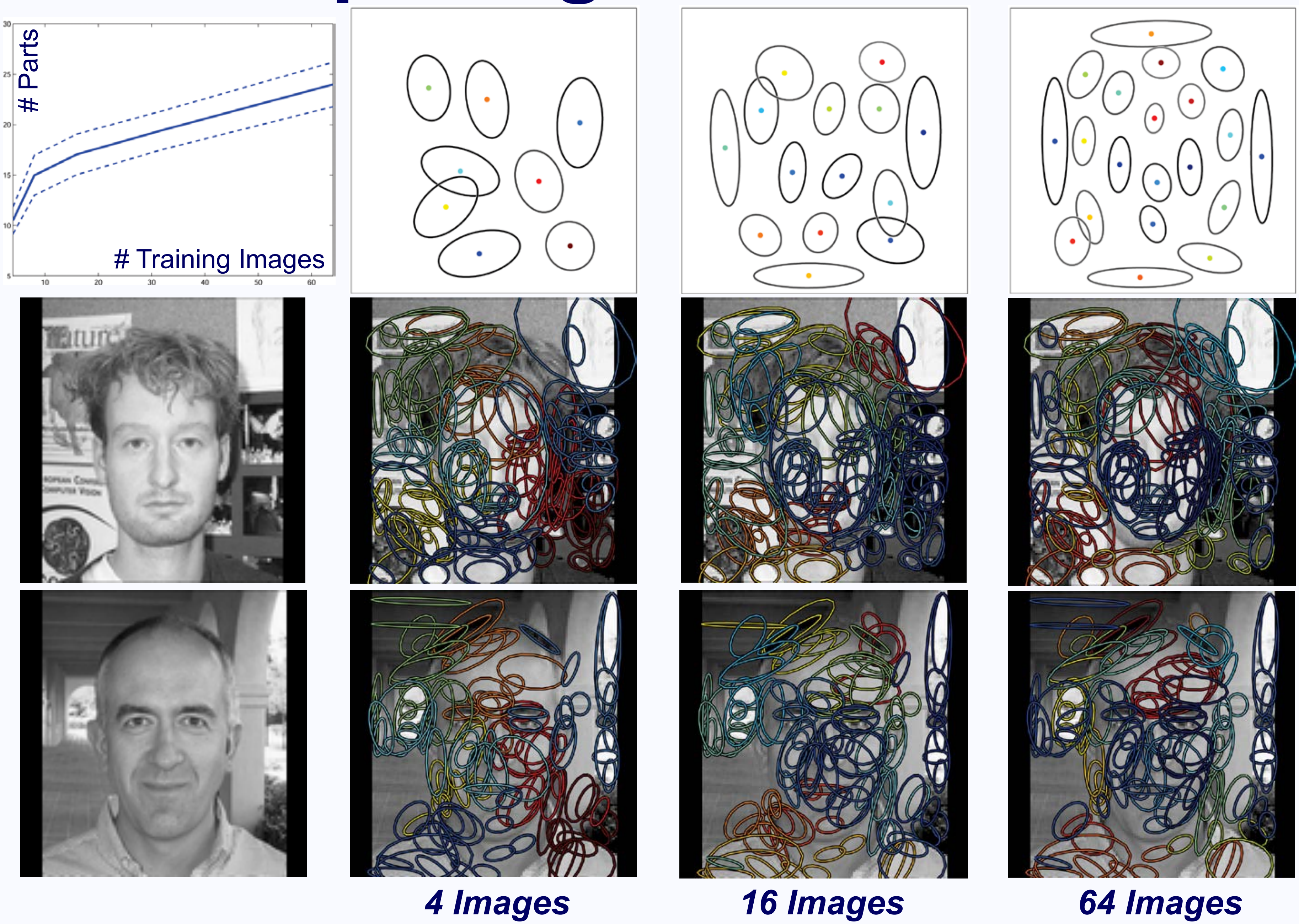
$$v_{ji} \sim \mathcal{N}(v; \bar{\mu}_{ji} + \rho_j, \bar{\Lambda}_{ji})$$

Learning DPs: Gibbs Sampling

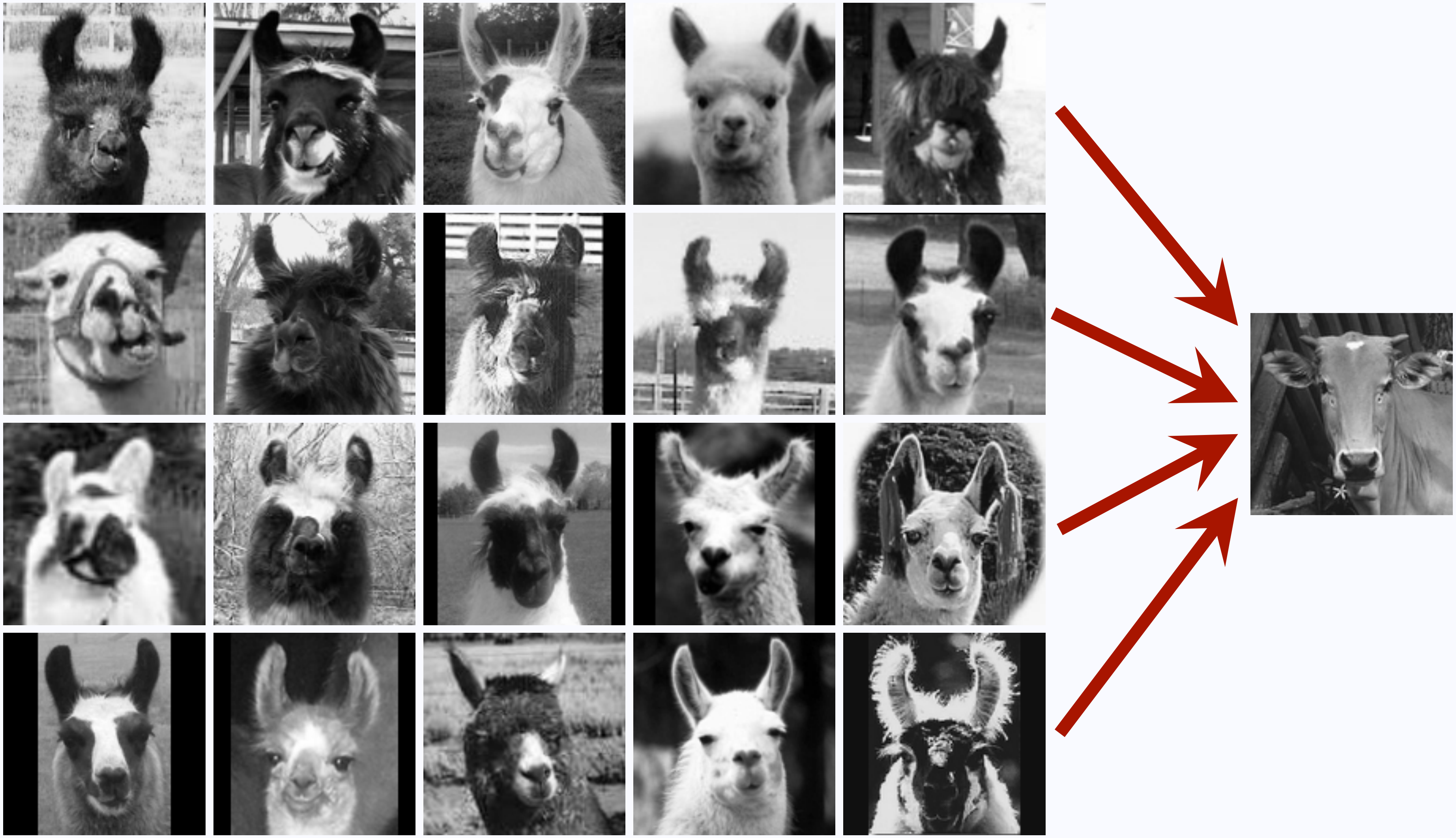


Dirichlet processes have many desirable analytic properties, which lead to efficient *Rao-Blackwellized* learning algorithms

Decomposing Faces into Parts

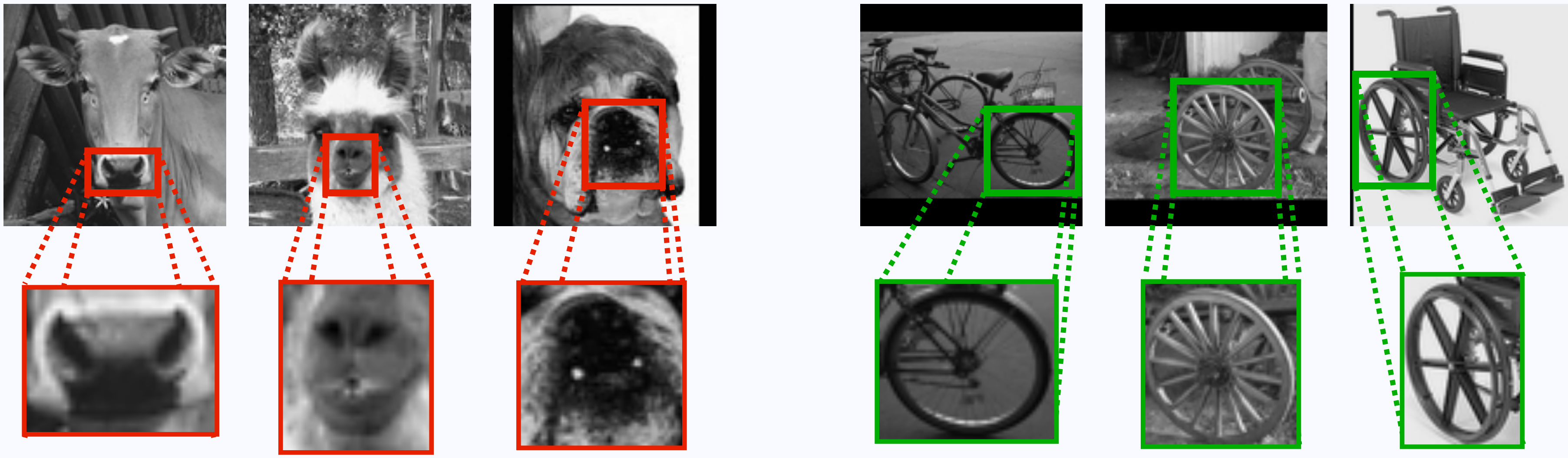


Generalizing Across Categories



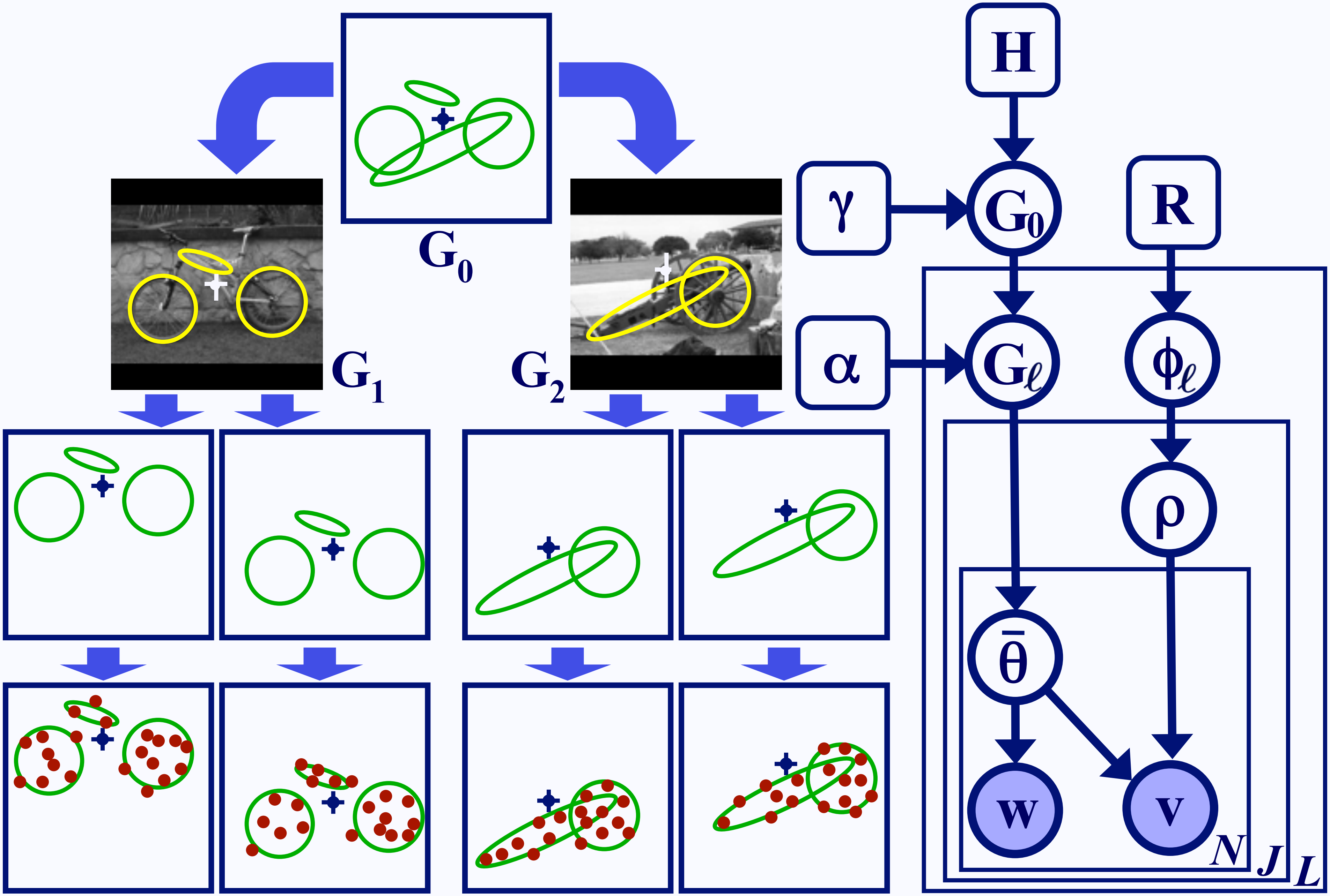
Can we transfer knowledge from one object category to another?

Learning Shared Parts

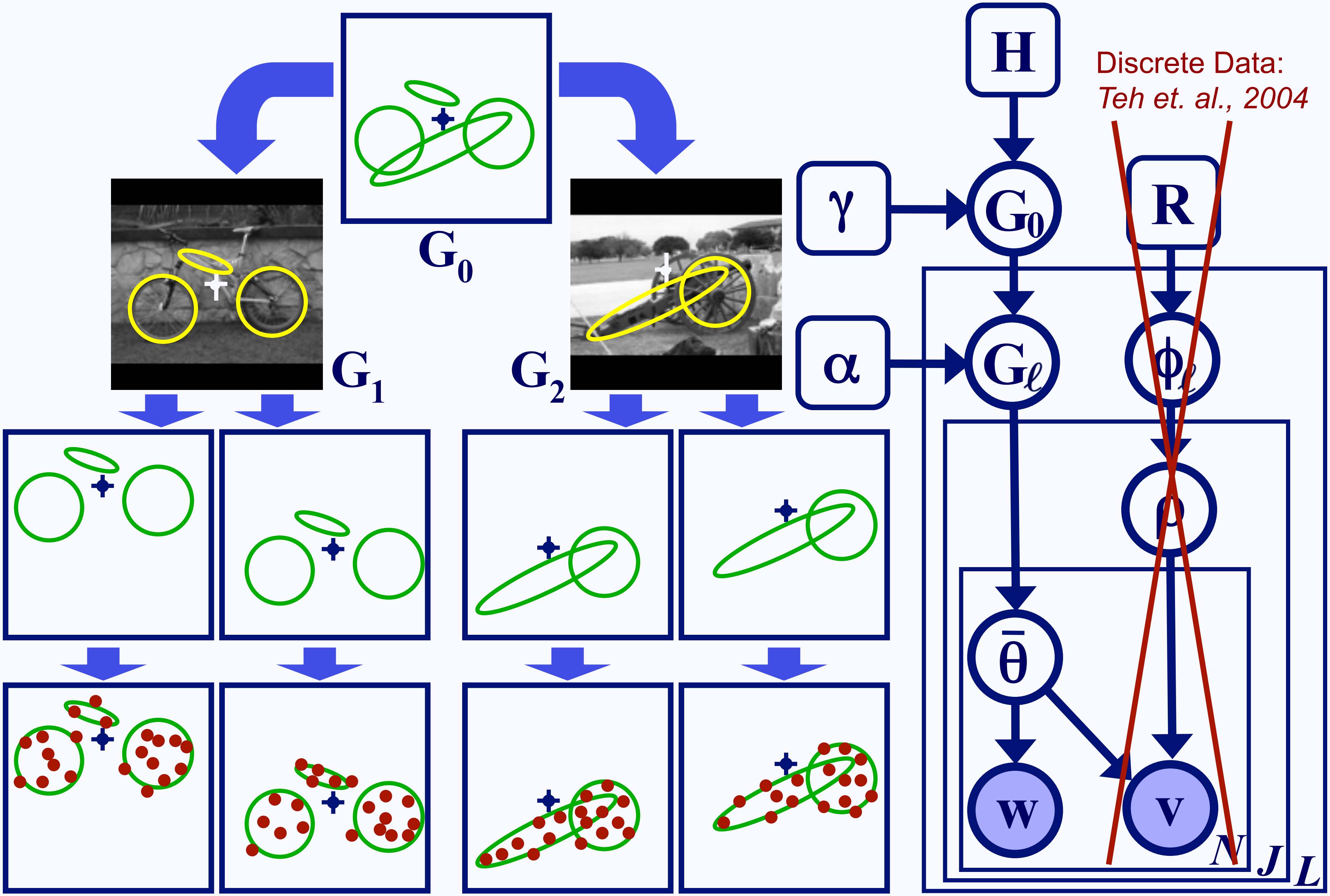


- Objects are often locally similar in appearance
- Discover *parts* shared across categories
  - ➢ How many total parts should we share?
  - ➢ How many parts should each category use?

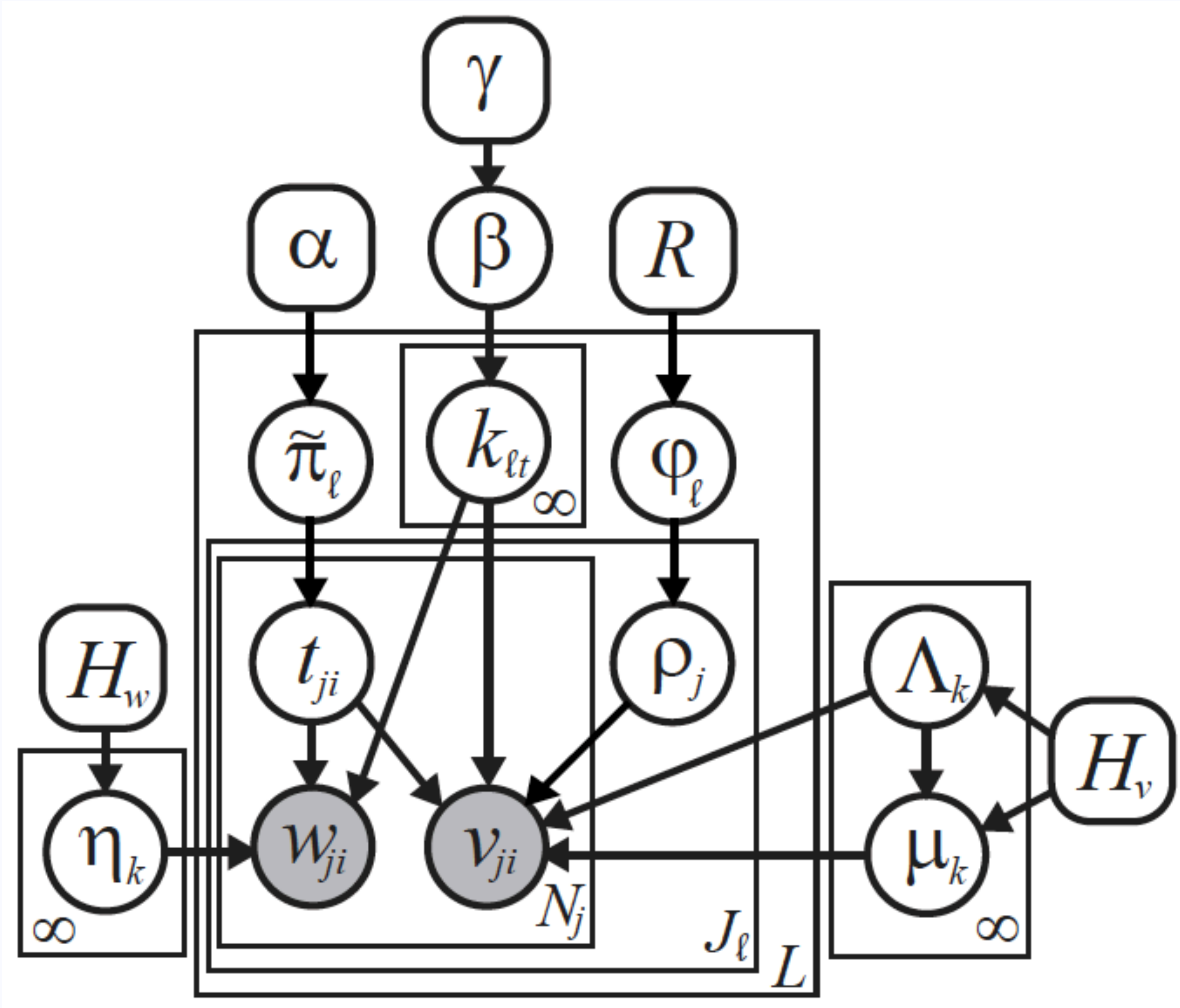
Hierarchical DP Object Model



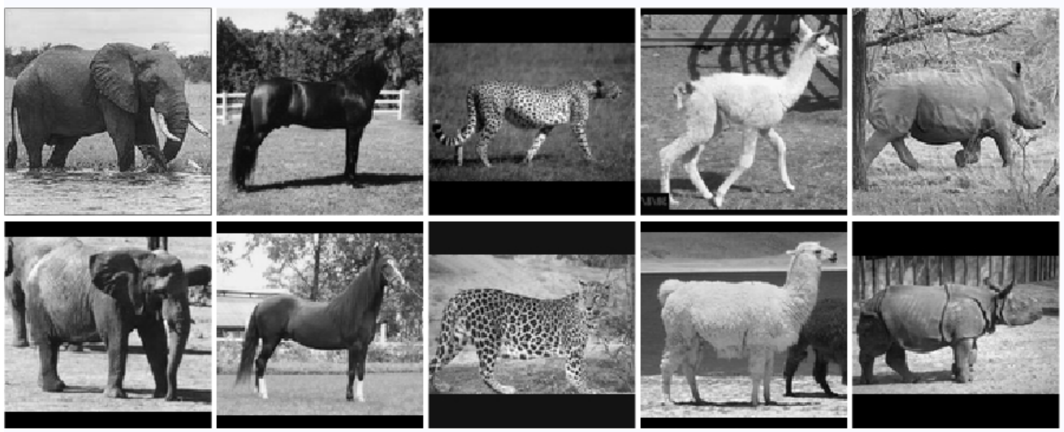
Hierarchical DP Object Model



Chinese Restaurant Franchise

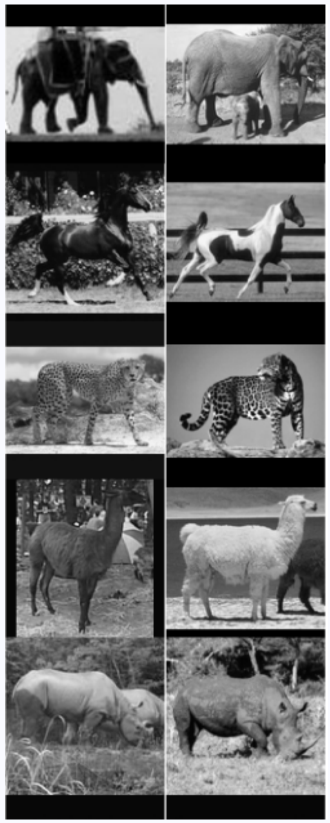


Sharing Parts: 16 Categories

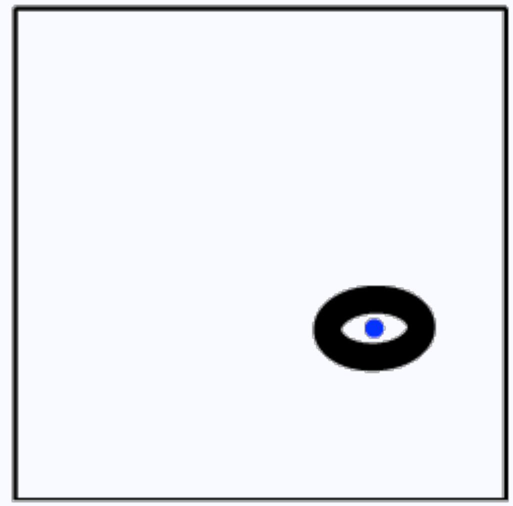


- Caltech 101 Dataset (Li & Perona)
- Horses (Borenstein & Ullman)
- Cat & dog faces (Vidal-Naquet & Ullman)
- Bikes from Graz-02 (Opelt & Pinz)
- Google…

Visualization of Shared Parts

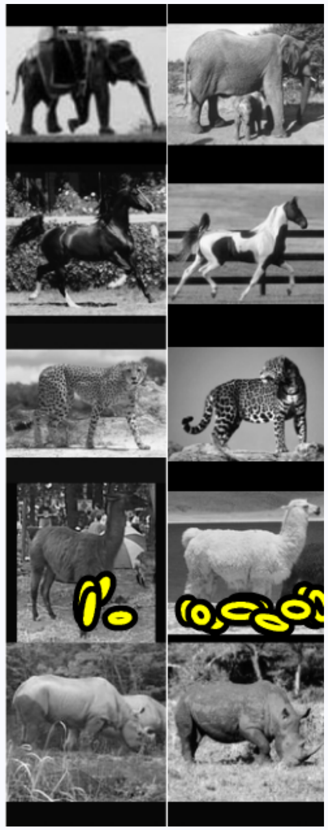


Pr(appearance | part)

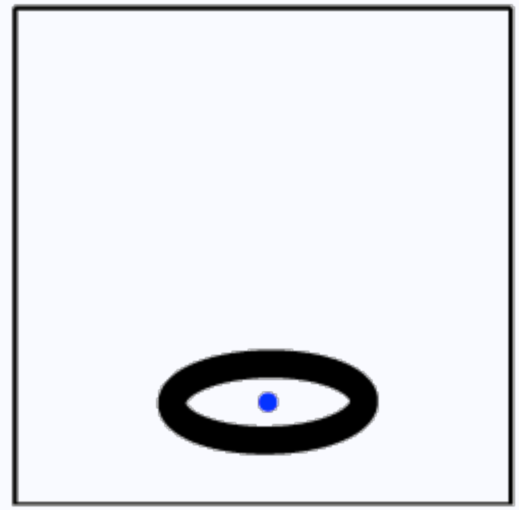


Pr(position | part)

Visualization of Shared Parts

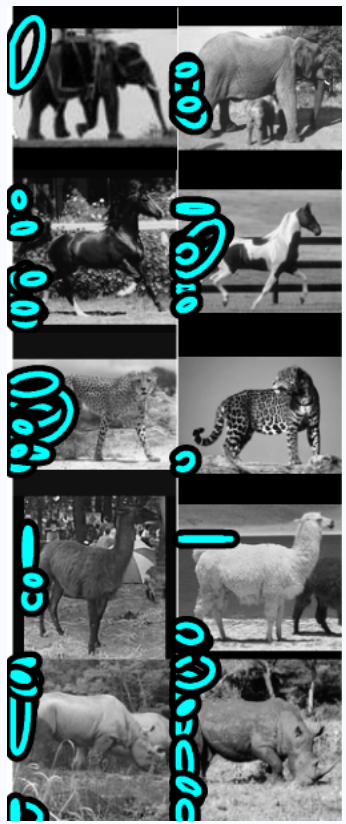


Pr(appearance | part)

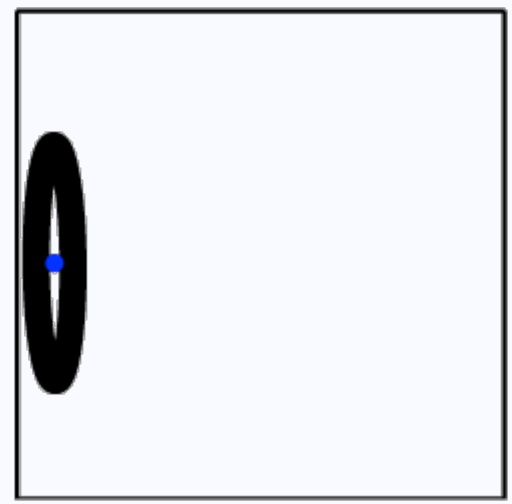


Pr(position | part)

Visualization of Shared Parts

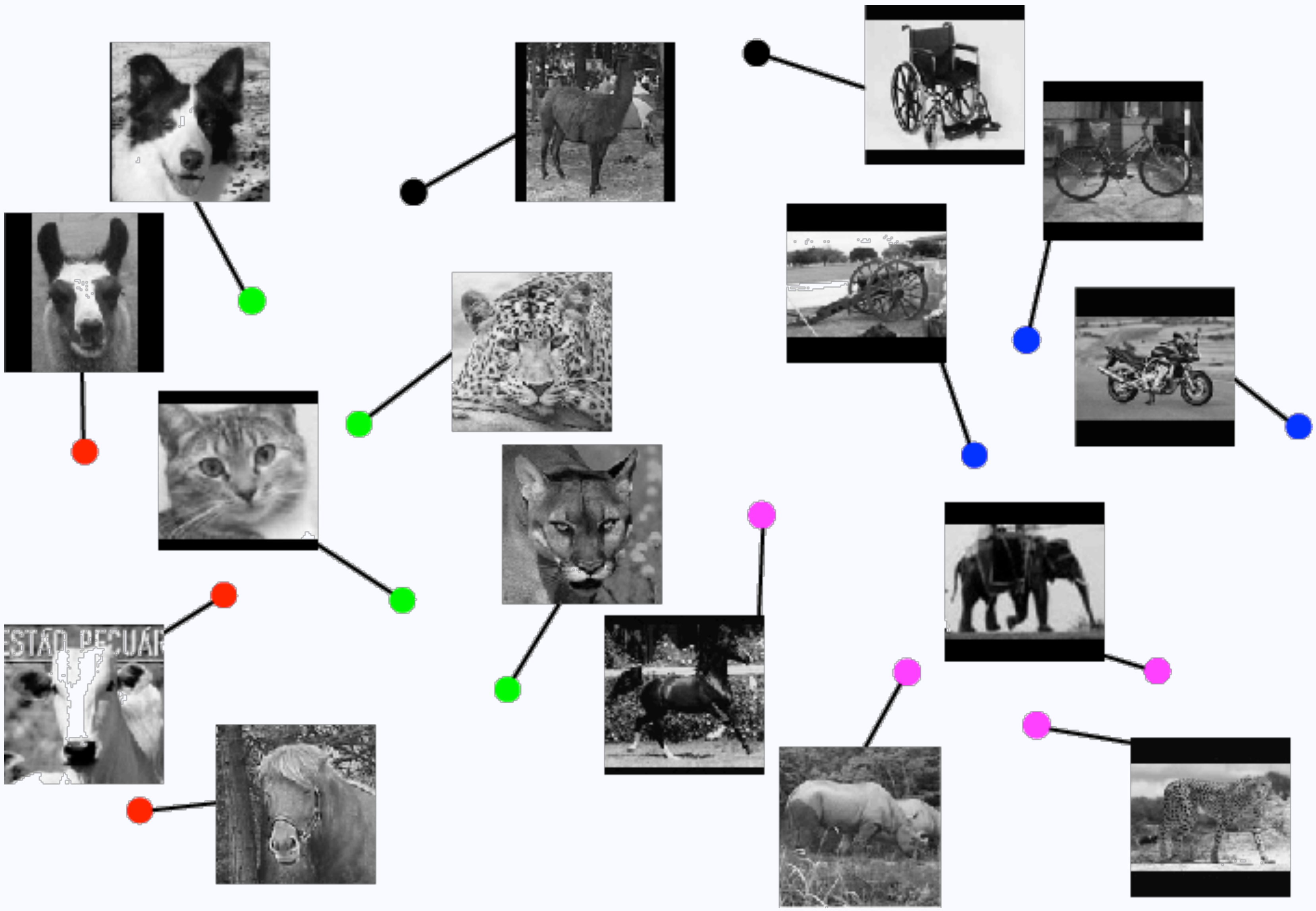


Pr(appearance | part)



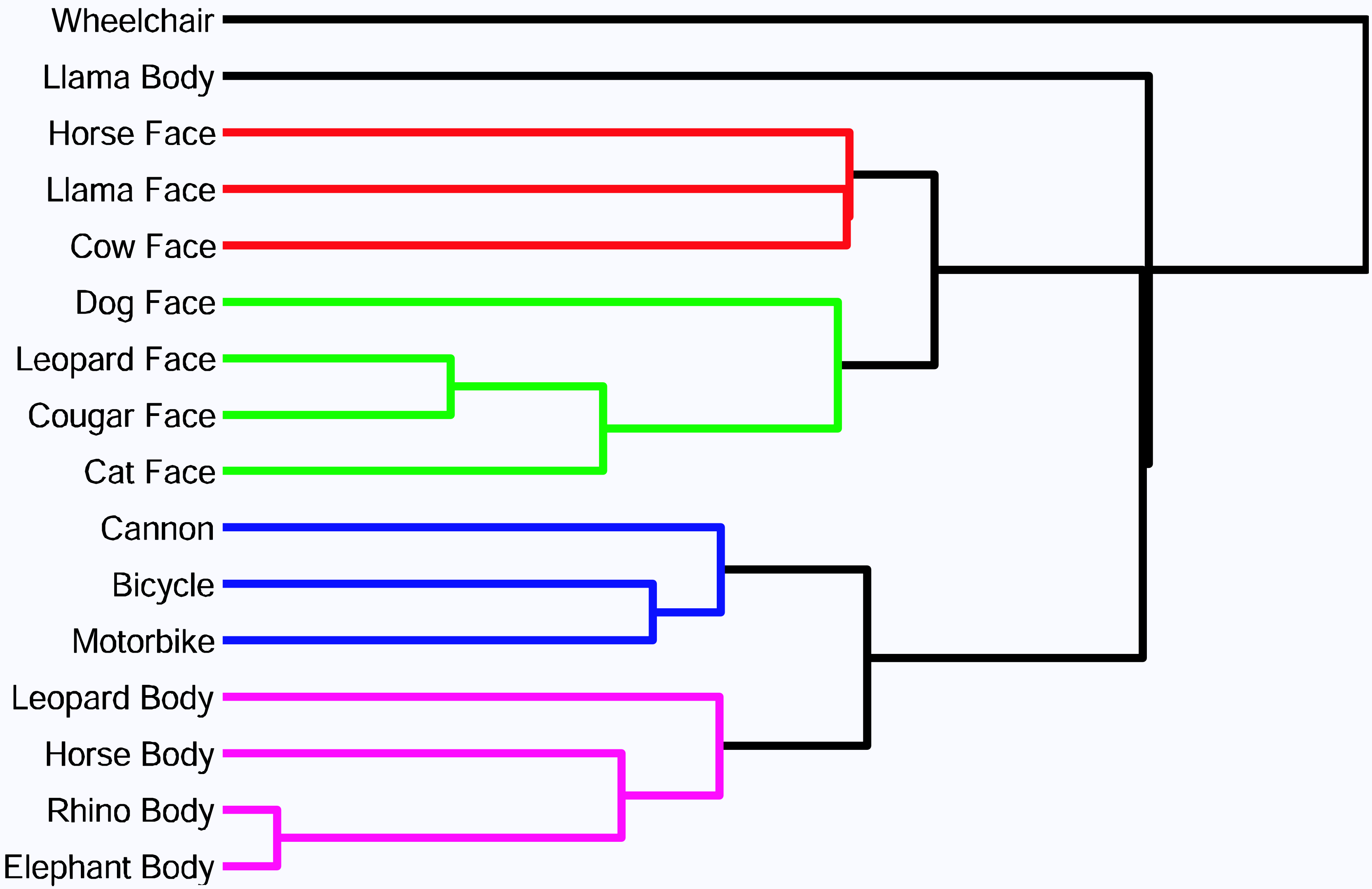
Pr(position | part)

Visualization of Part Densities



MDS Embedding of Pr(part | object)

Visualization of Part Densities

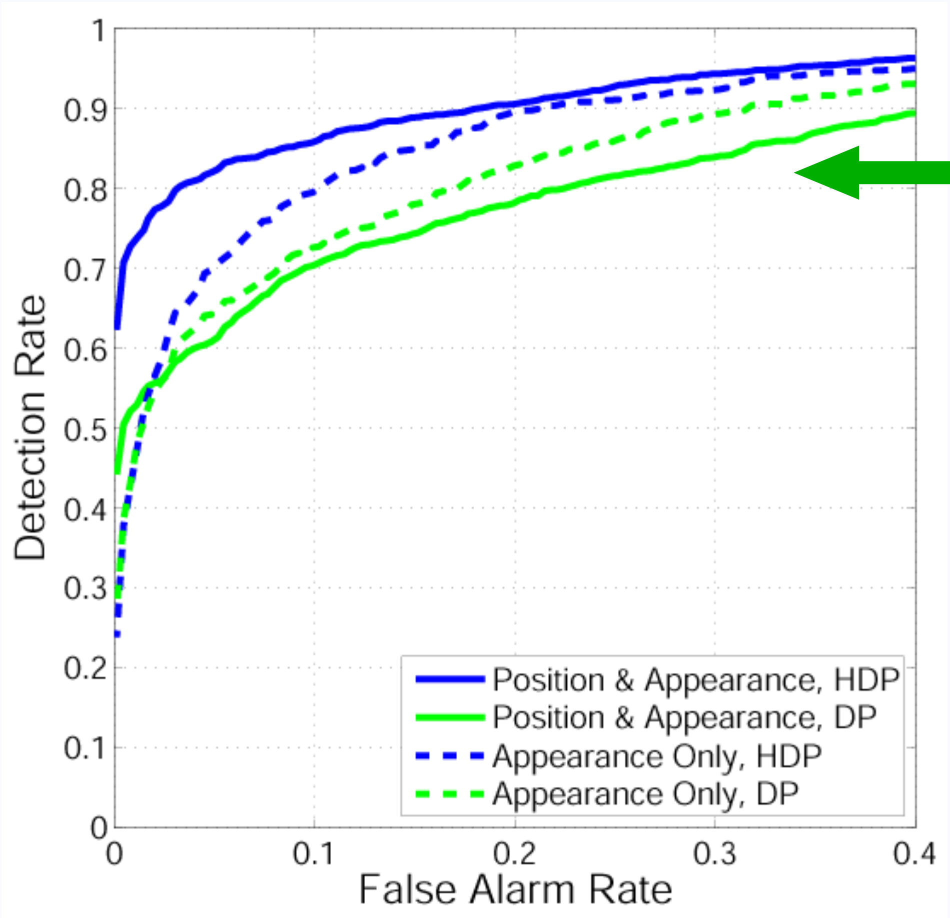


Hierarchical Clustering of Pr(part | object)

Detection Task

versus

Detection Results

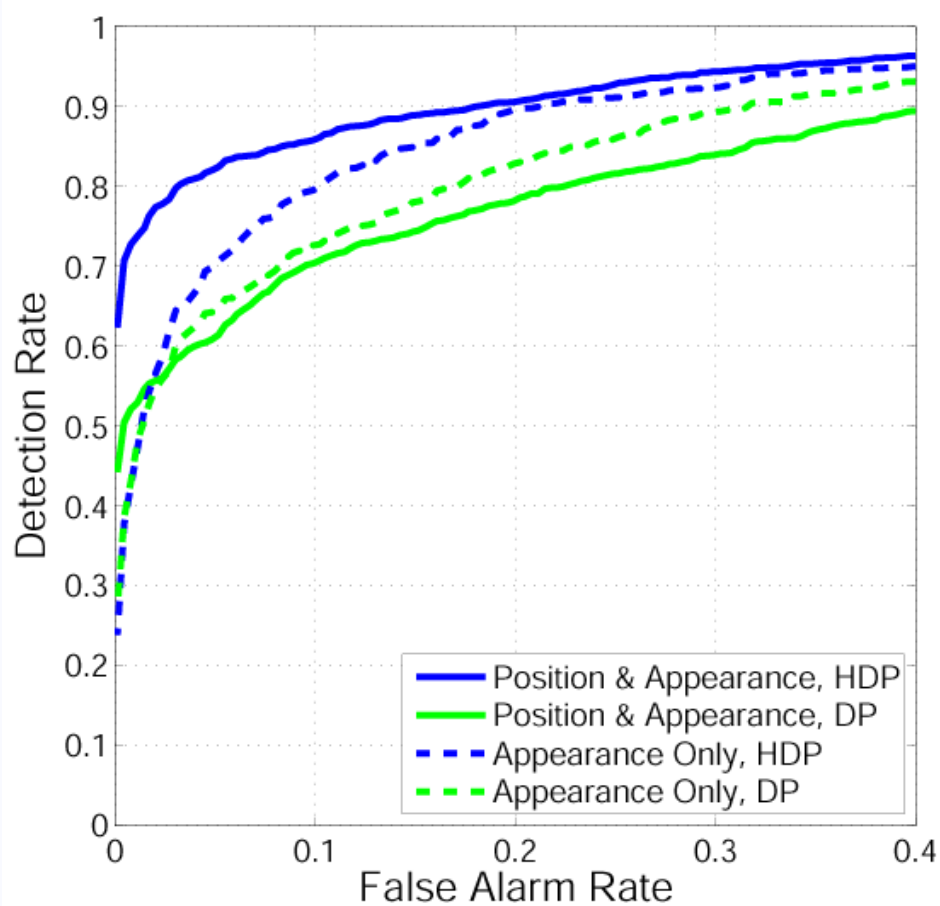


6 Training Images per Category
(ROC Curves)

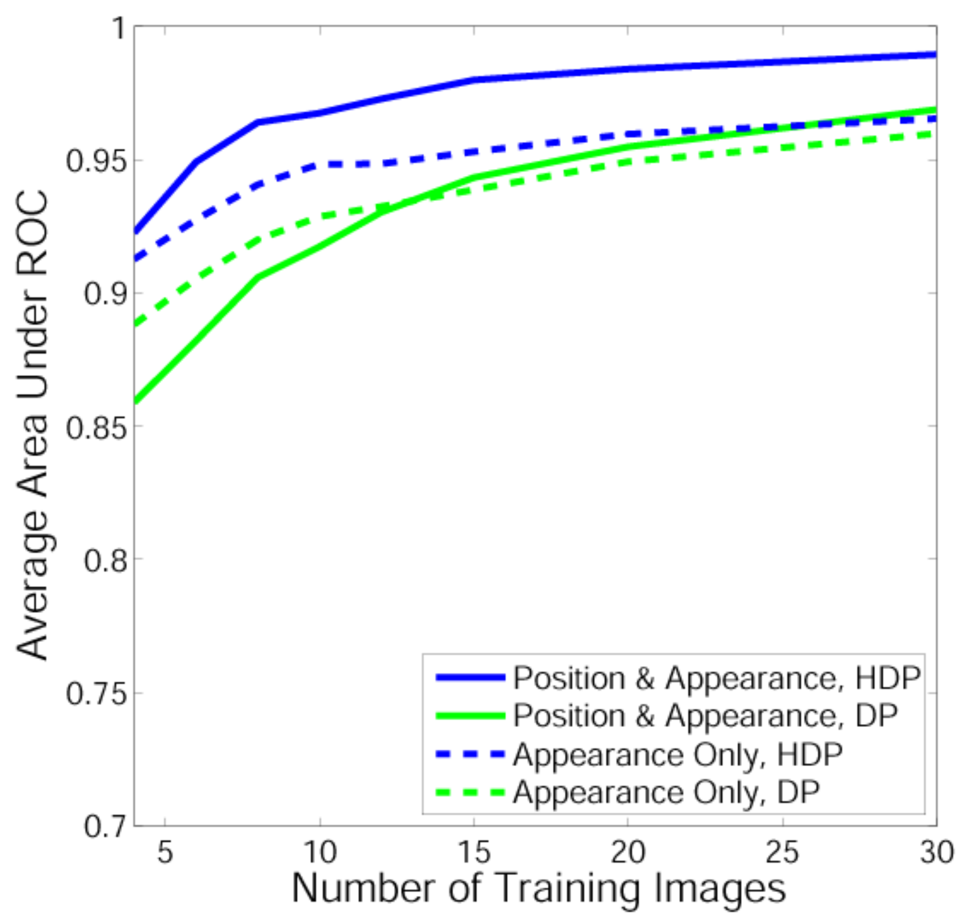
Shared Parts
more accurate than
Unshared Parts

Modeling feature positions ***improves shared*** detection, but ***hurts unshared*** detection

Detection Results

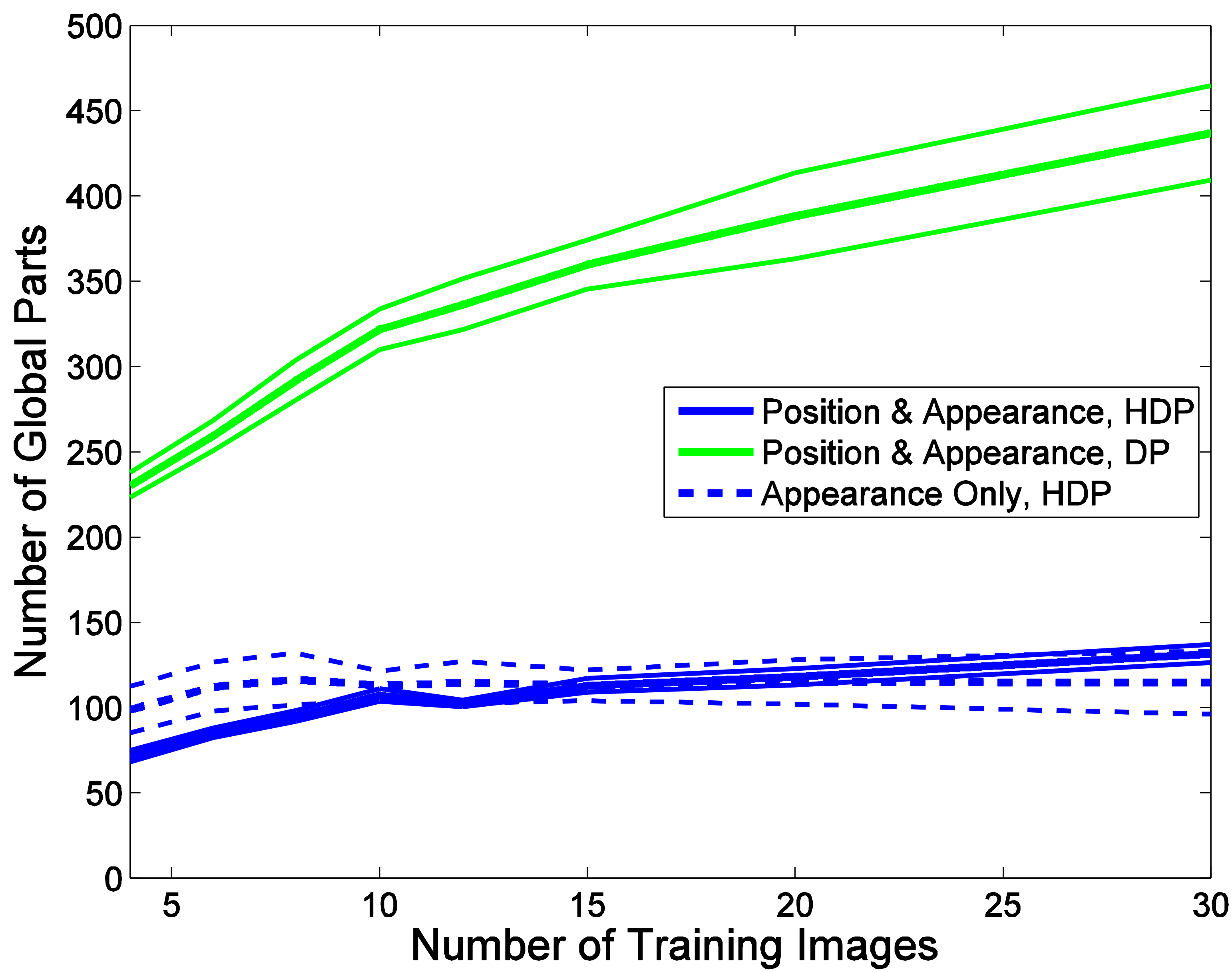


6 Training Images per Category
(ROC Curves)



Detection vs. Training Set Size
(Area Under ROC)

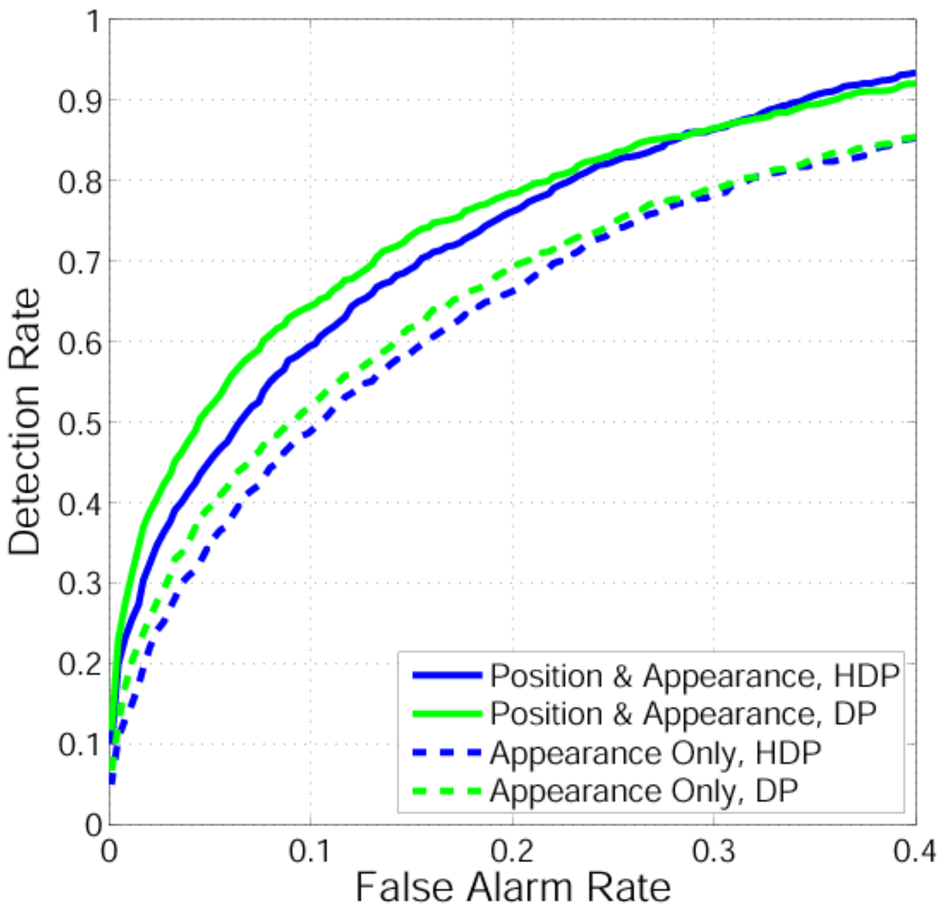
Sharing Simplifies Models



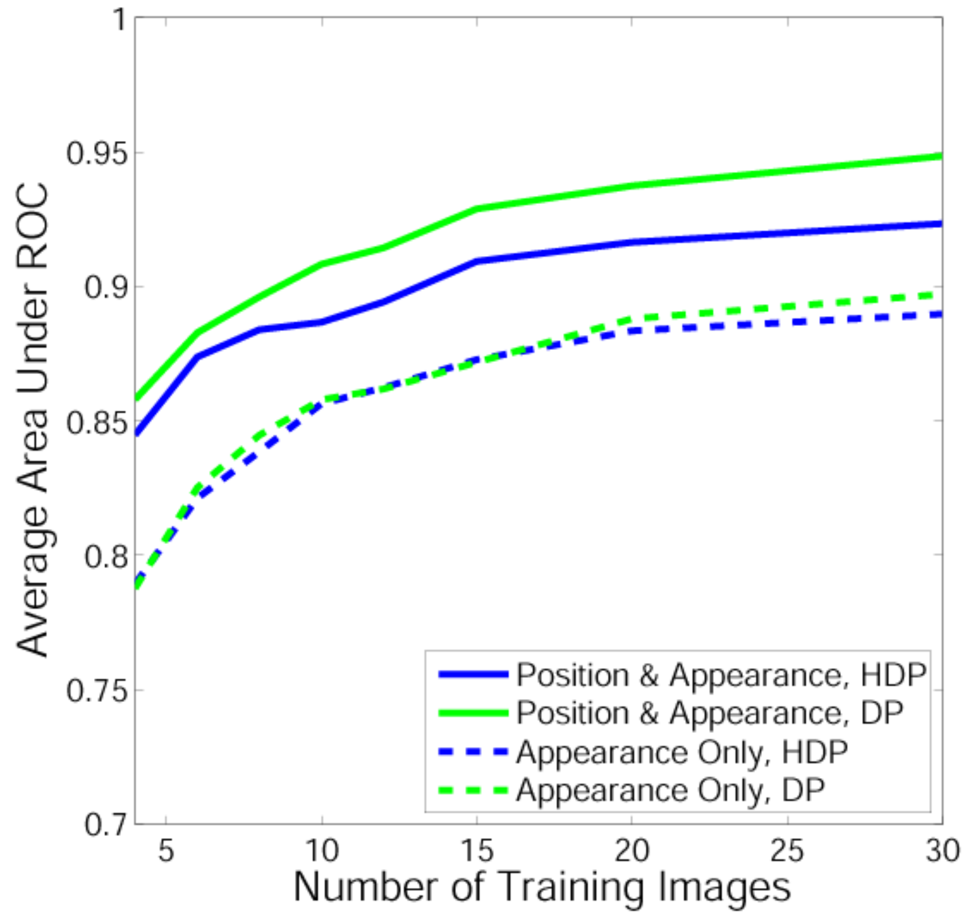
Recognition Task

versus

Recognition Results



6 Training Images per Category
(ROC Curves)



Detection vs. Training Set Size
(Area Under ROC)