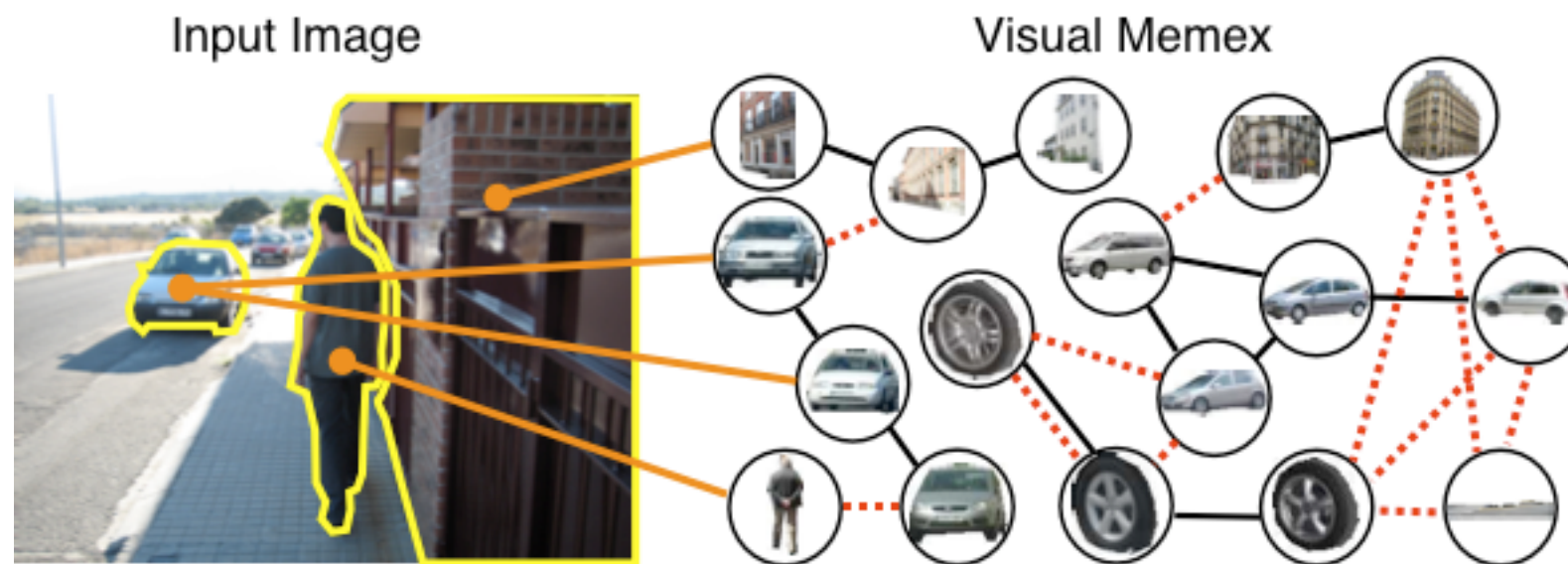


Beyond Categories: The Visual Memex Model for Reasoning About Object Relationships



Tomasz Malisiewicz and Alexei (Alyosha) Efros
NIPS 2009

Presented at CMU's misc-read on 11/4/2009

Overview

- Motivation
- Visual Memex Model
- Evaluation Task: Context Challenge
- Results and Comparisons to Baselines
- Examples

Understanding an Image



slide by Fei Fei, Fergus & Torralba

Object naming

sky

building

flag

face

banner

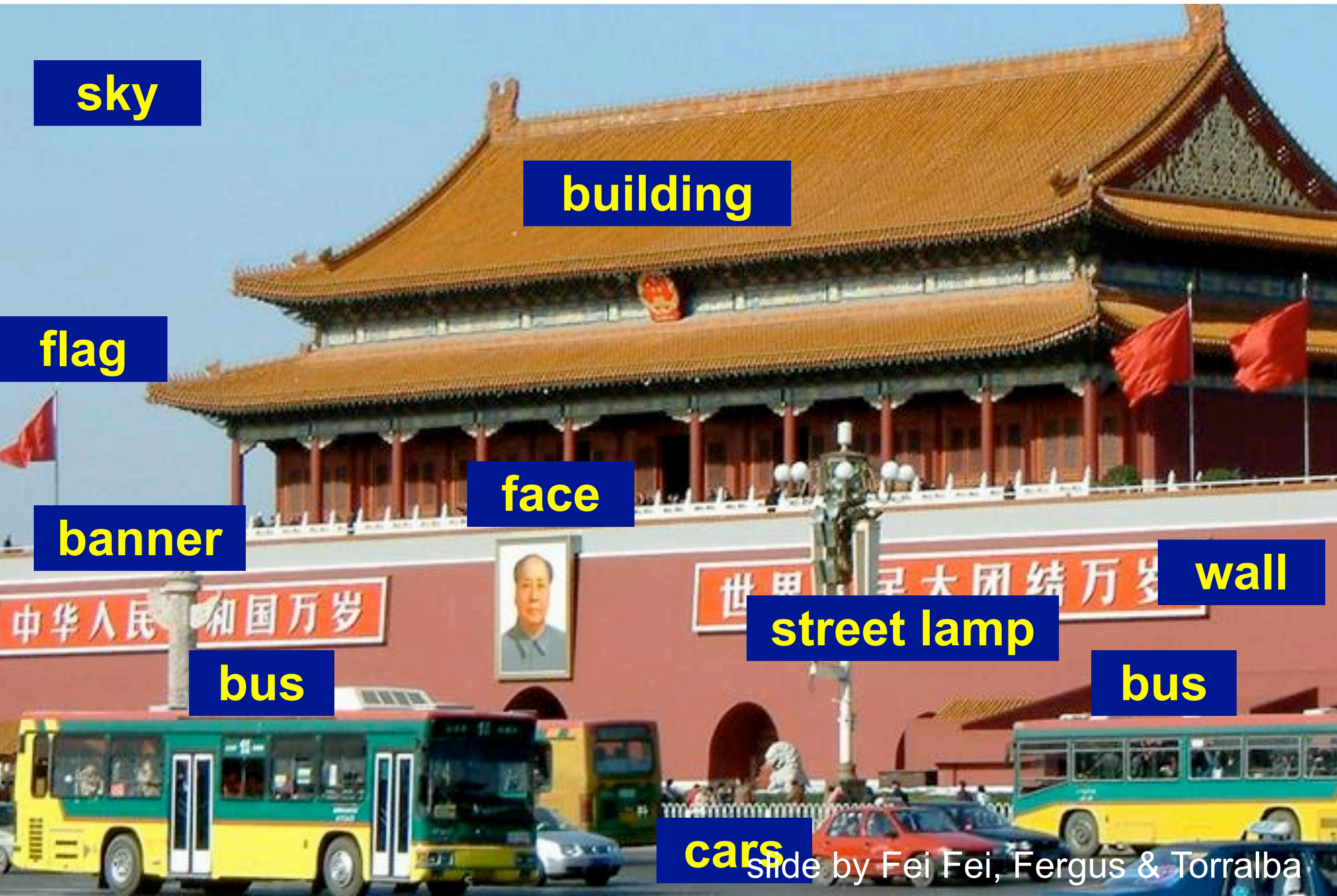
wall

street lamp

bus

bus

cars



Object naming / Object categorization



sky

building

flag

face

banner

wall

street lamp

bus

bus

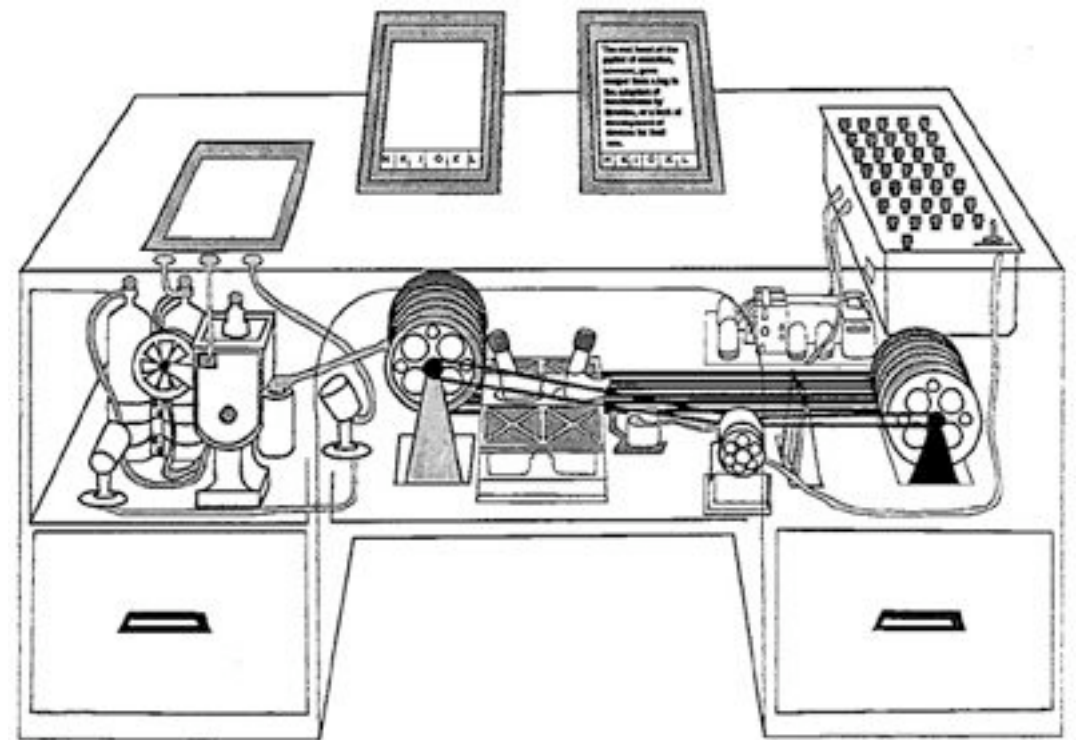
cars

slide by Fei Fei, Fergus & Torralba

Vannevar Bush' Memex



Memex: A device to organize concepts and ideas via a web of associations (and not indexing)

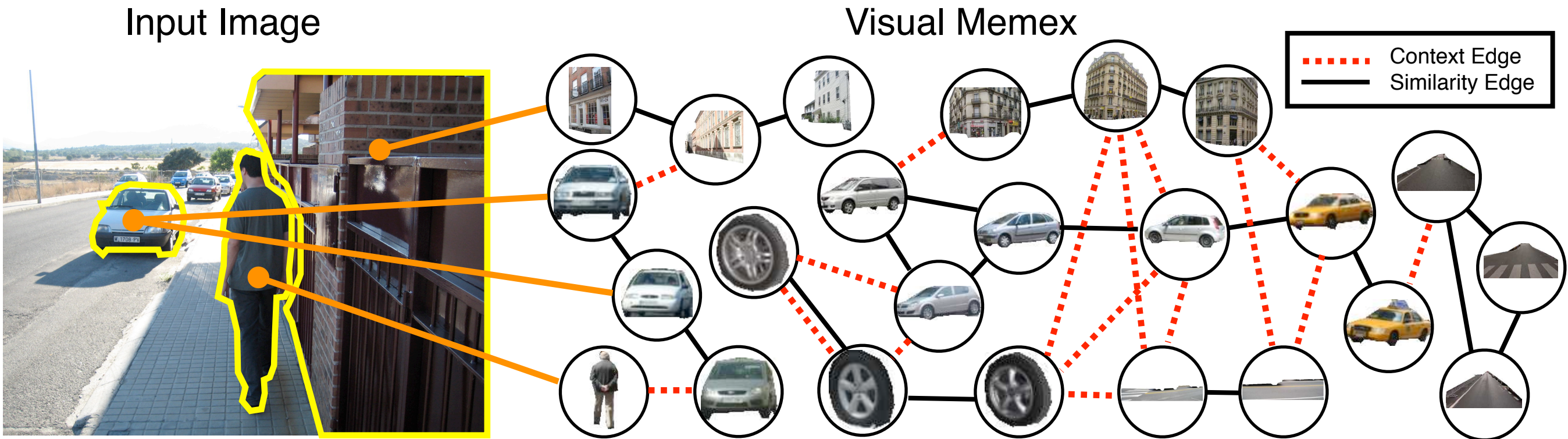


V. B. "As We May Think," The Atlantic, 1945

Visual Memex Model

Input Image

Visual Memex



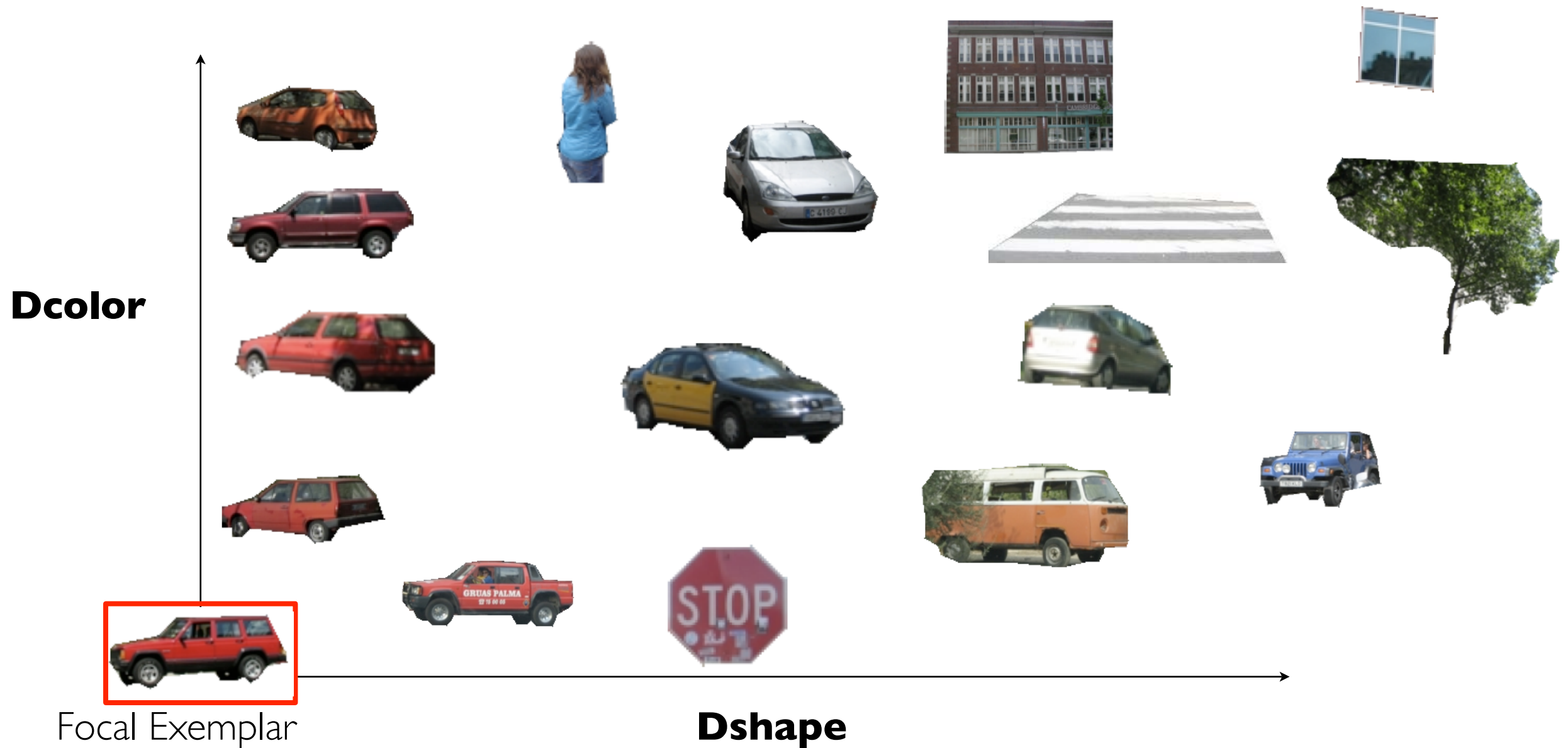
Learning Similarity Edges

- Per-exemplar distance function learning algorithm from CVPR 2008



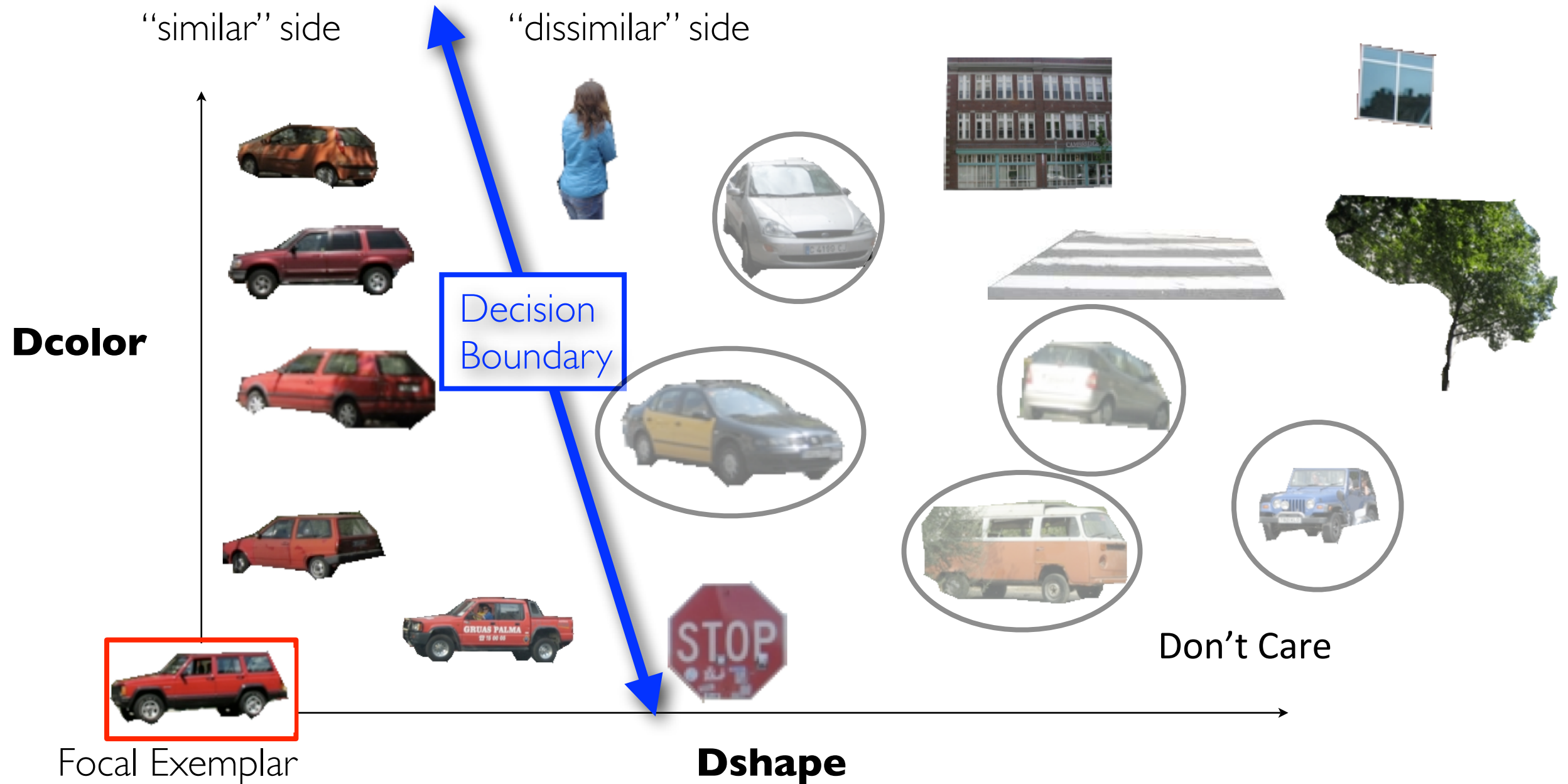
Type	Name	Dimension
Shape	Centered Mask	32x32=1024
	BB Extent	2
	Pixel Area	1
Texture	Right Boundary Tex-Hist	100
	Top Boundary Tex-Hist	100
	Left Boundary Tex-Hist	100
	Bottom Boundary Tex-Hist	100
	Interior Tex-Hist	100
Color	Mean Color	3
	Color std	3
	Color Histogram	33
Location	Absolute Mask	8x8=64
	Top Height	1
	Bot Height	1

Learning Similarity Edges



$$f(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{\lambda}{2} \|\mathbf{w}\|^2 + \sum_{i \in C} \alpha_i L(-(\mathbf{w}^T \mathbf{d}_i + b)) + \sum_{i \notin C} L(\mathbf{w}^T \mathbf{d}_i + b) - \sigma \|\boldsymbol{\alpha}\|^2$$

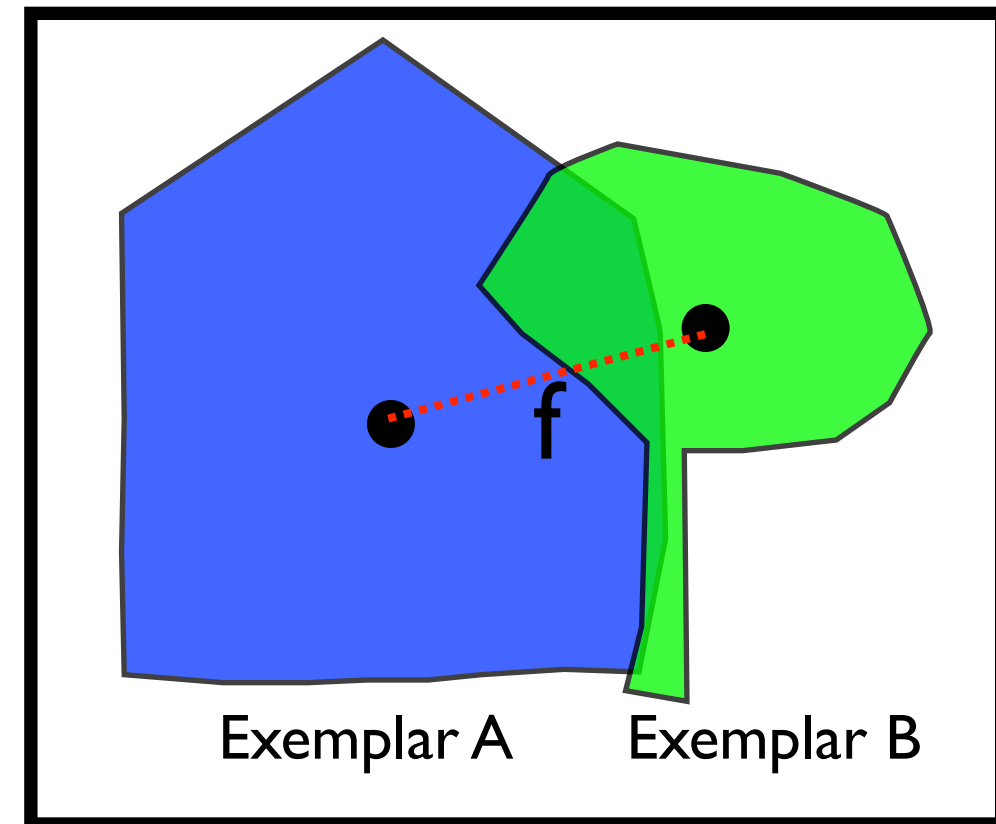
Learning Similarity Edges



$$f(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{\lambda}{2} \|\mathbf{w}\|^2 + \sum_{i \in C} \alpha_i L(-(\mathbf{w}^T \mathbf{d}_i + b)) + \sum_{i \notin C} L(\mathbf{w}^T \mathbf{d}_i + b) - \sigma \|\boldsymbol{\alpha}\|^2$$

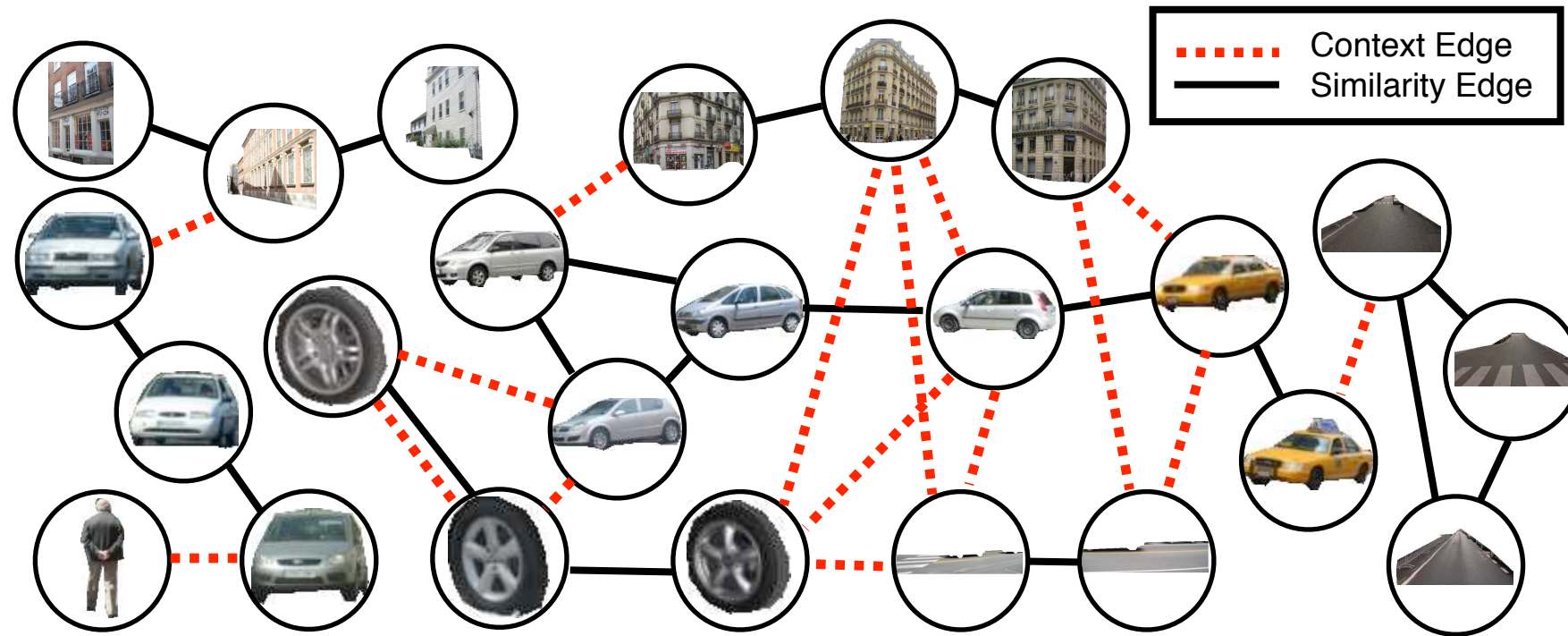
Context Edges

- Compute pairwise feature for each pair of exemplars in a single image
- Overlap, Scale, Displacement, Bottom-pixel Height $\mathbf{f} \in \mathbb{R}^{10}$
- No appearance!



$$K(\mathbf{f}, \mathbf{f}') = e^{-\alpha_1 || \mathbf{f} - \mathbf{f}' ||^2}$$

Visual Memex Model



$$G = (V, E_S, E_C, \{D\}, \{f\})$$

$$N = 87,802$$

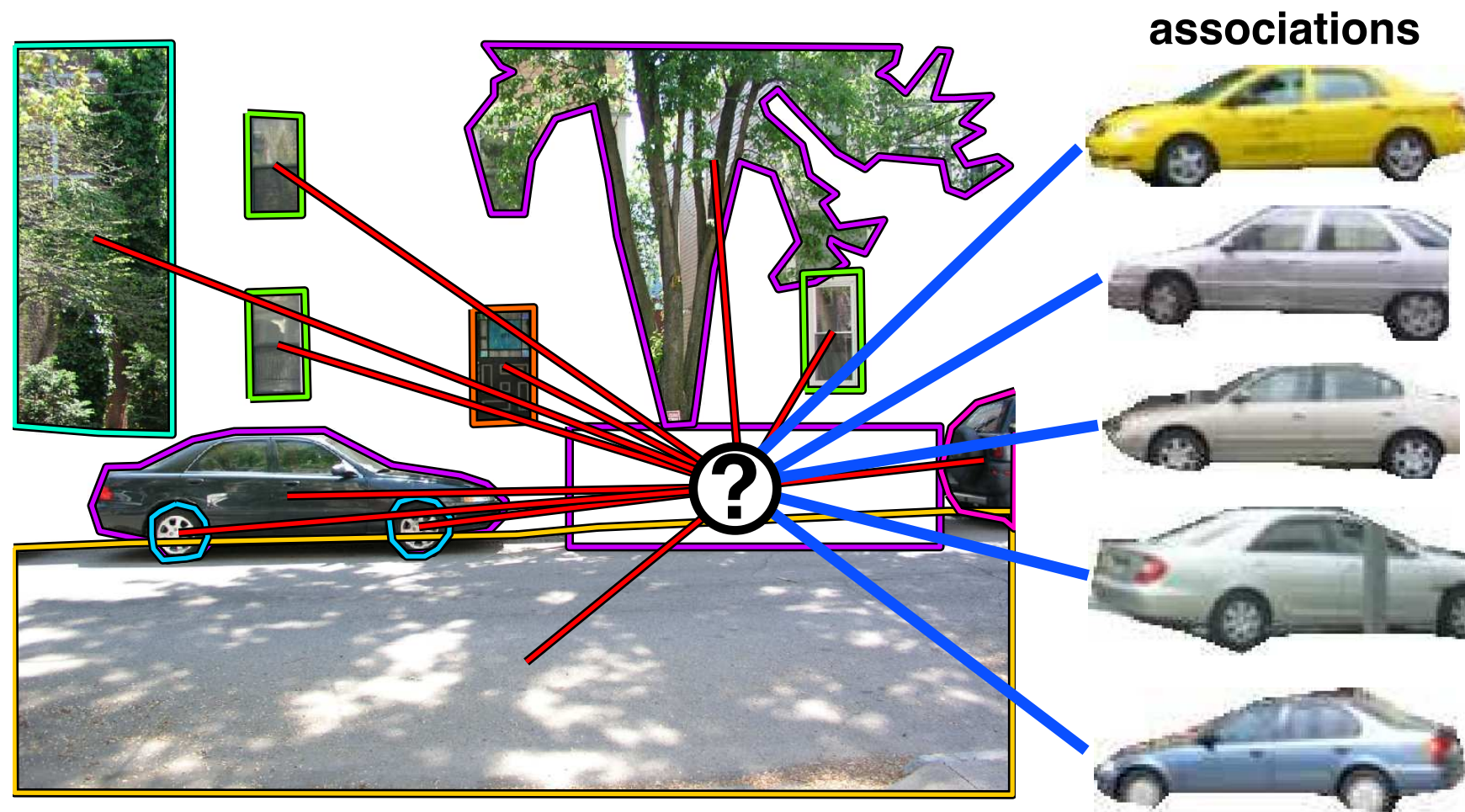
$$|E_S| = 276,782$$

$$|E_C| = 989,106$$

Evaluation Task: Context Challenge

- Compare against Category-Based models
- Perform context-only recognition “without local appearance”
- Motivated by Torralba: “How far can you go without running an object detector?”
- Our context is object-centered as opposed to Torralba’s scene-centered GIST context

Context Challenge



Visual Memex Inference

- **Goal:** Create exemplar associations for hidden region given K supporting object regions
- **Approach:** Score the assignment of every exemplar to the hidden region

Visual Memex Inference

Appearance Features of supporting objects $\{S_1, \dots, S_K\}$

Soft Associations between supporting objects and Memex exemplars $A_j^a = e^{-D_a(S_j)}$

Exemplar-Exemplar Pairwise Potential $\Psi(e_i, e_j, \mathbf{f}_{ij})$

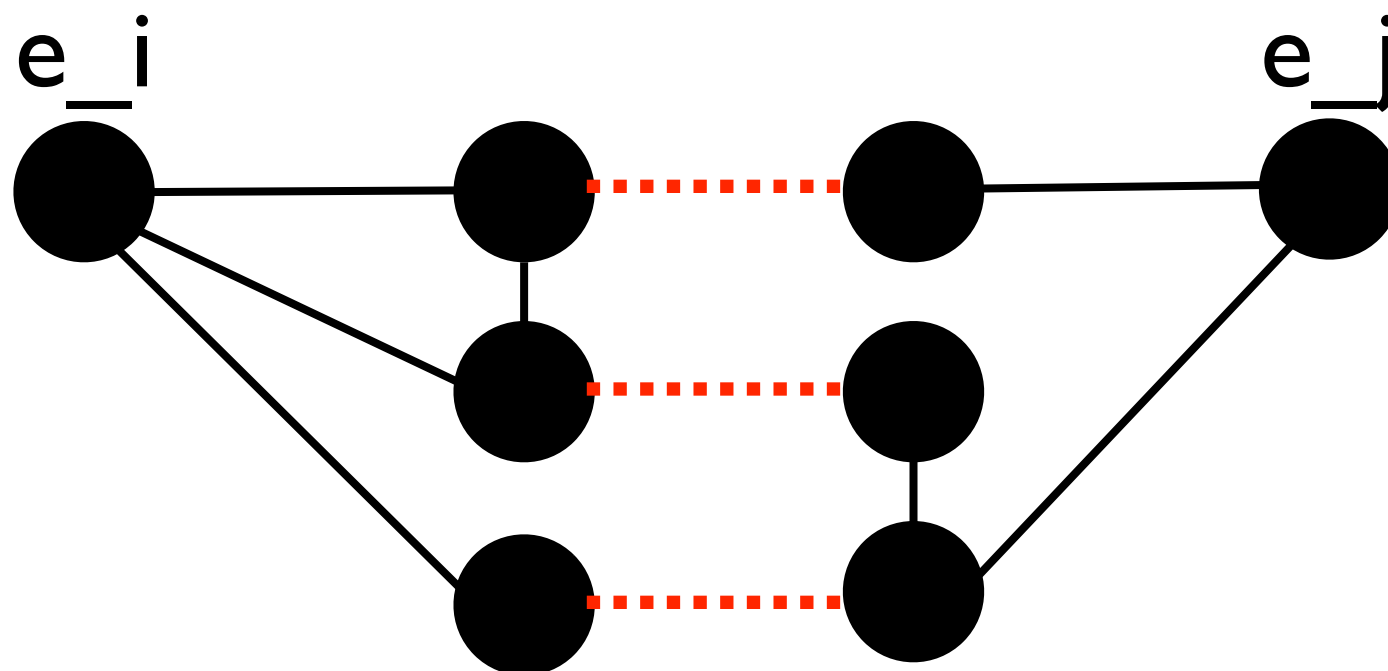
$$p(e_i | A_1, \dots, A_K, \mathbf{f}_{i1}, \dots, \mathbf{f}_{iK}) \propto \prod_{j=1}^K \sum_{a=1}^N A_j^a \Psi(e_i, e_a, \mathbf{f}_{ij})$$

“Densification” Potential

Visual Memex Adjacency Matrix

$$W_{uv} = [(u, v) \in E_S]$$

$$\log \Psi(e_i, e_j, \mathbf{f}_{ij}) = \frac{\sum_{(u,v) \in E_C} W_{iu} W_{jv} K(\mathbf{f}_{ij}, \mathbf{f}_{uv})}{\sum_{(u,v) \in E_C} W_{iu} W_{jv}}$$

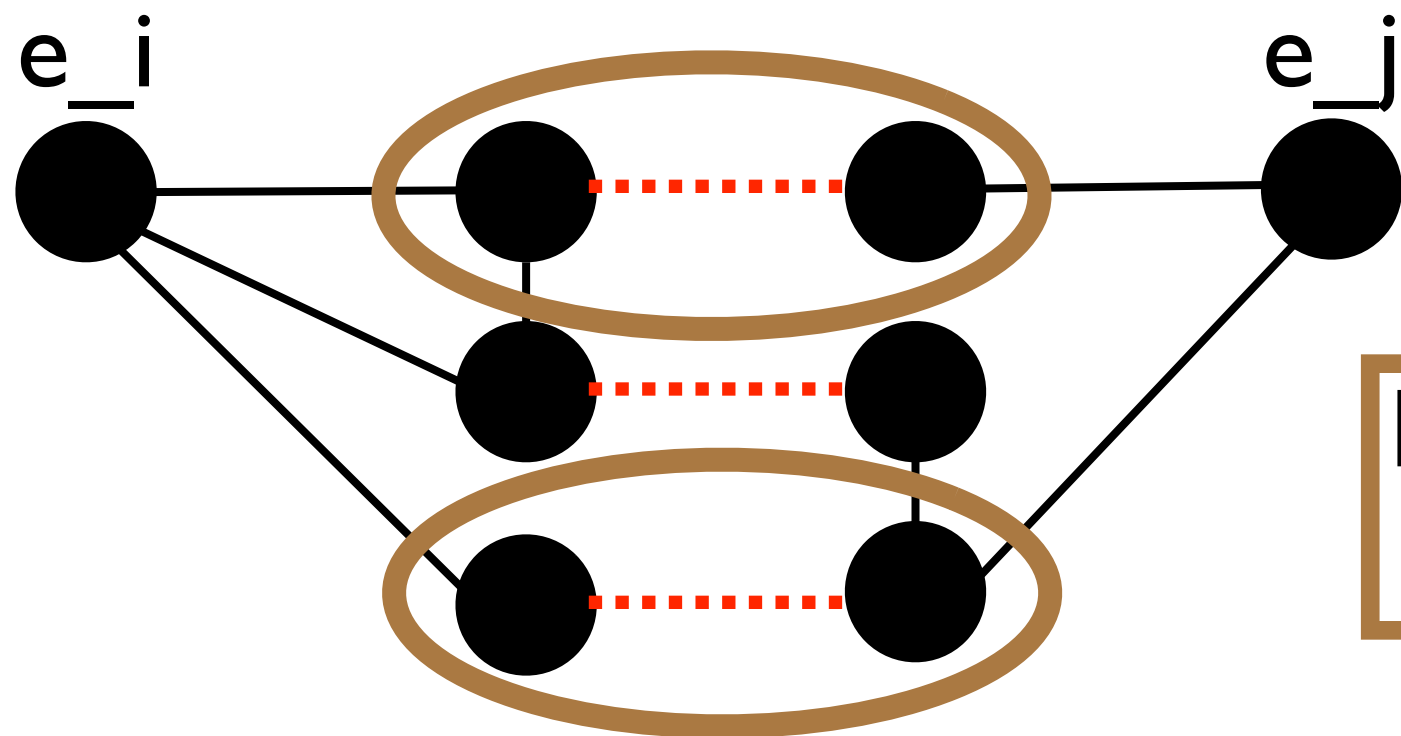


“Densification” Potential

Visual Memex Adjacency Matrix

$$W_{uv} = [(u, v) \in E_S]$$

$$\log \Psi(e_i, e_j, \mathbf{f}_{ij}) = \frac{\sum_{(u,v) \in E_C} W_{iu} W_{jv} K(\mathbf{f}_{ij}, \mathbf{f}_{uv})}{\sum_{(u,v) \in E_C} W_{iu} W_{jv}}$$



KDE with these
features

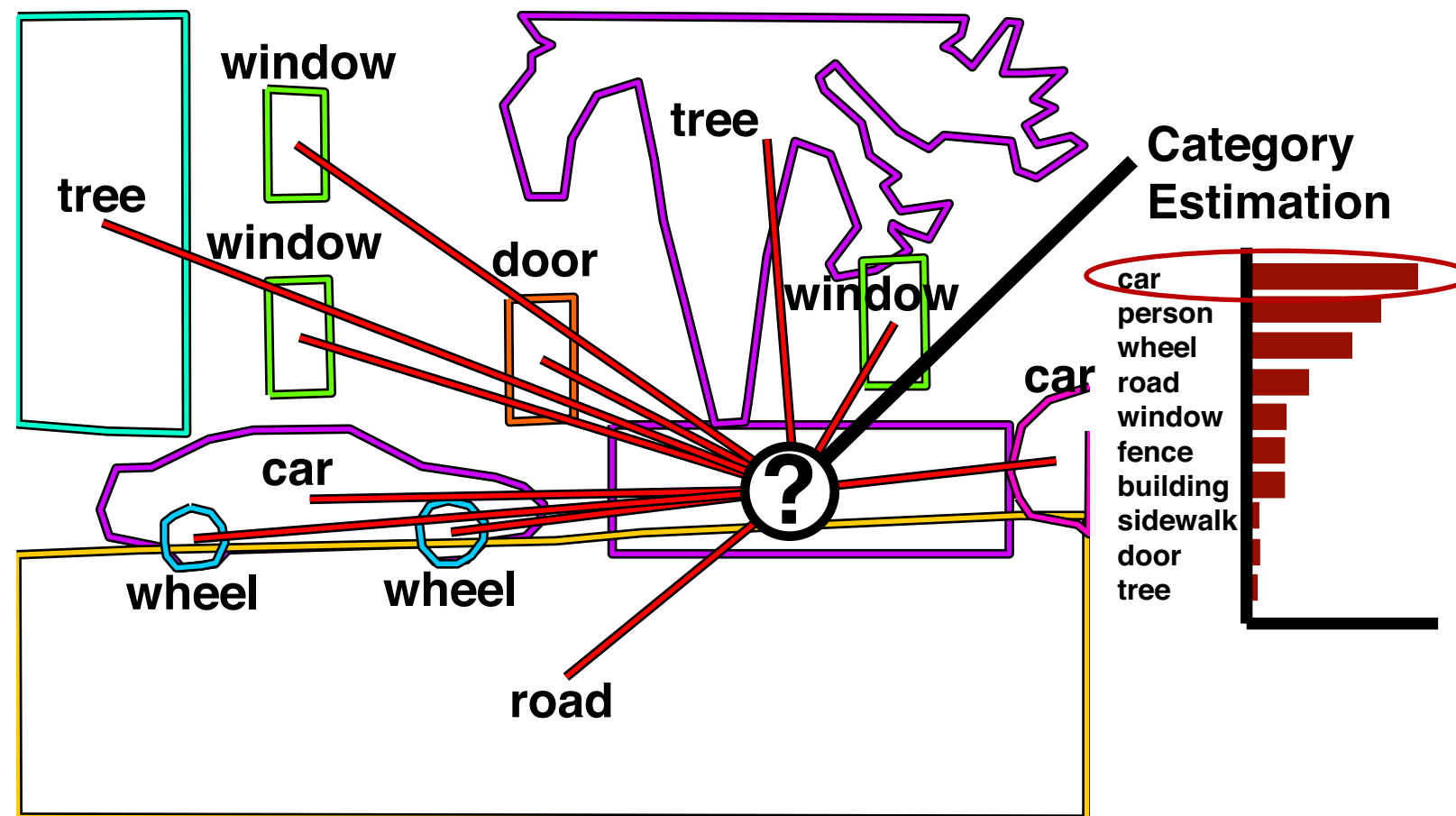
Baselines

- Category-based parametric CoLA model
(Serge Belongie's Vision Group)
- Category-based Reduced Memex
(categories for objects + nonparametric
context model)

CoLA Baseline

$$p(y_i | y_1, \dots, y_K, \mathbf{f}_{i1}, \dots, \mathbf{f}_{iK}, \boldsymbol{\theta}) = \frac{1}{Z} \prod_{j=1}^K \Psi(y_i, y_j, \mathbf{f}_{ij}, \boldsymbol{\theta})$$

$$\log \Psi(y_i, y_j, \mathbf{f}_{ij}, \boldsymbol{\theta}) = [h(\mathbf{f}_{ij})^T] \boldsymbol{\theta}(y_i, y_j)$$



Reduced KDE Memex

- No distance function learning
- Just connect all exemplars of the same category

$$\log \Psi(y_i, y_j, \mathbf{f}_{ij}) = \frac{\sum_{(u,v) \in E_C} \delta_{y_i y_u} \delta_{y_j y_v} K(\mathbf{f}_{ij}, \mathbf{f}_{uv})}{\sum_{(u,v) \in E_C} \delta_{y_i y_u} \delta_{y_j y_v}}$$

Results

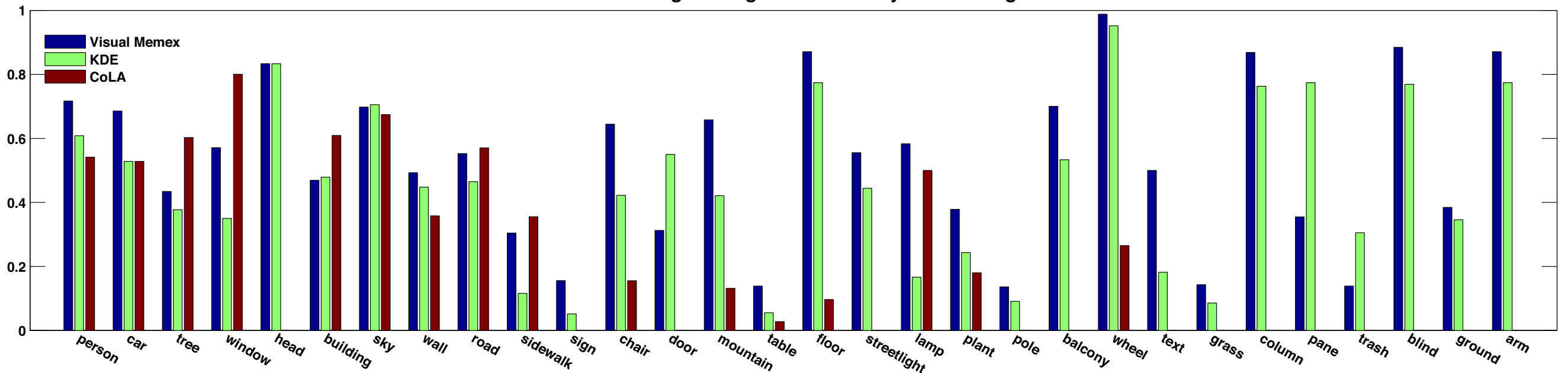
- 200 Densely labeled LabelMe images
- Solve separate Context Challenge task for a single object while fixing supporting objects

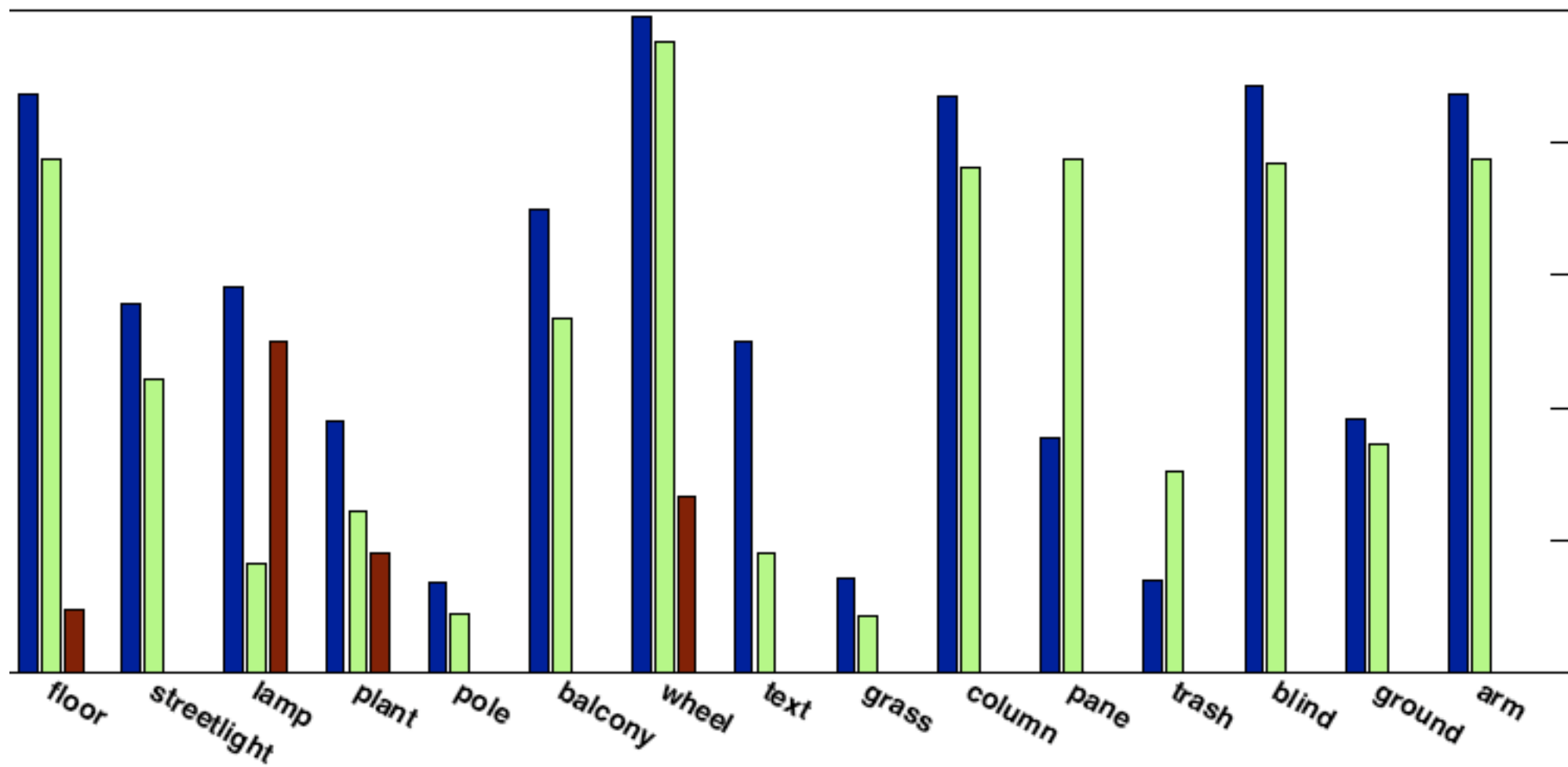
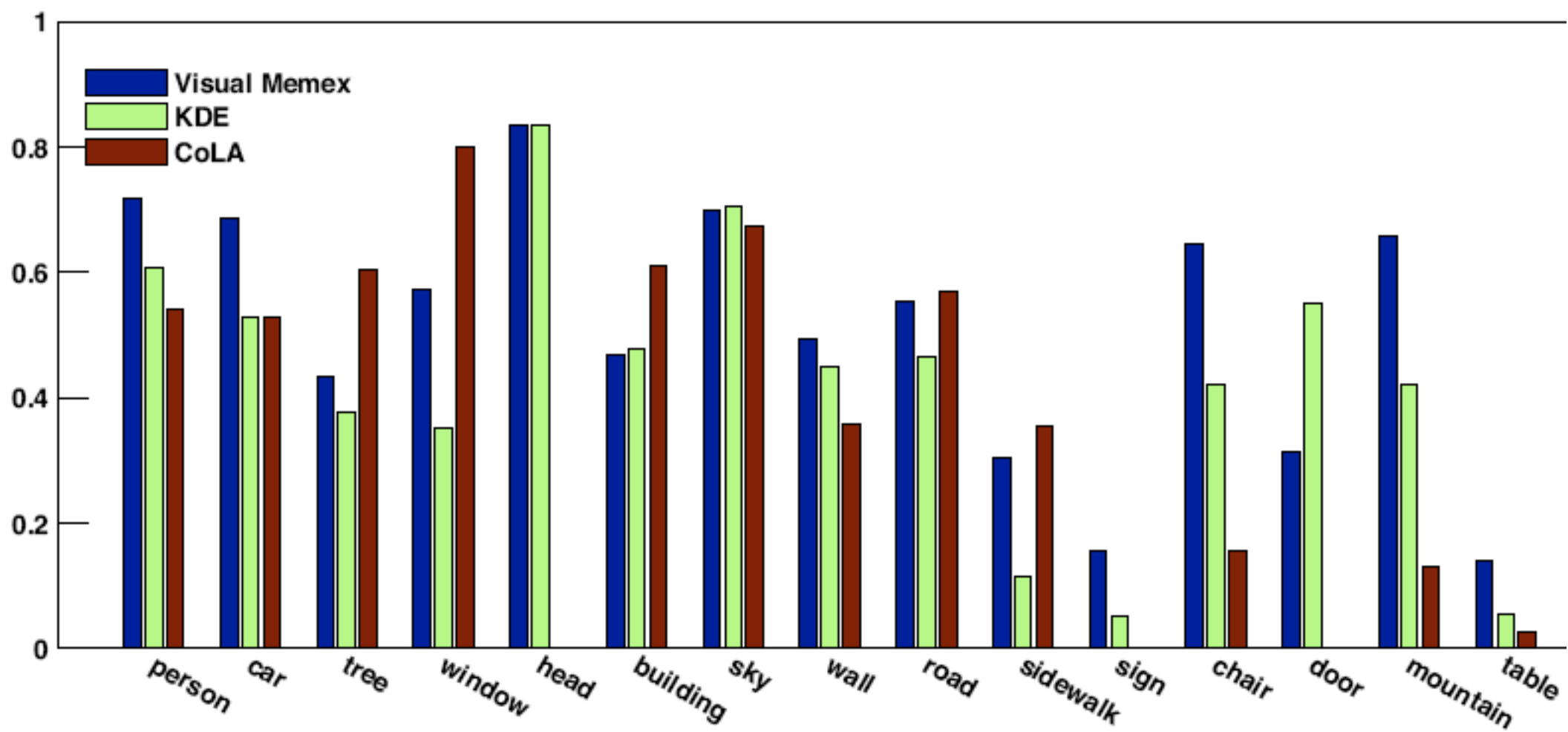
Results

	Overall	Per-Category
Visual Memex	0.527	0.534
KDE Memex	0.430	0.454
CoLA	0.457	0.213

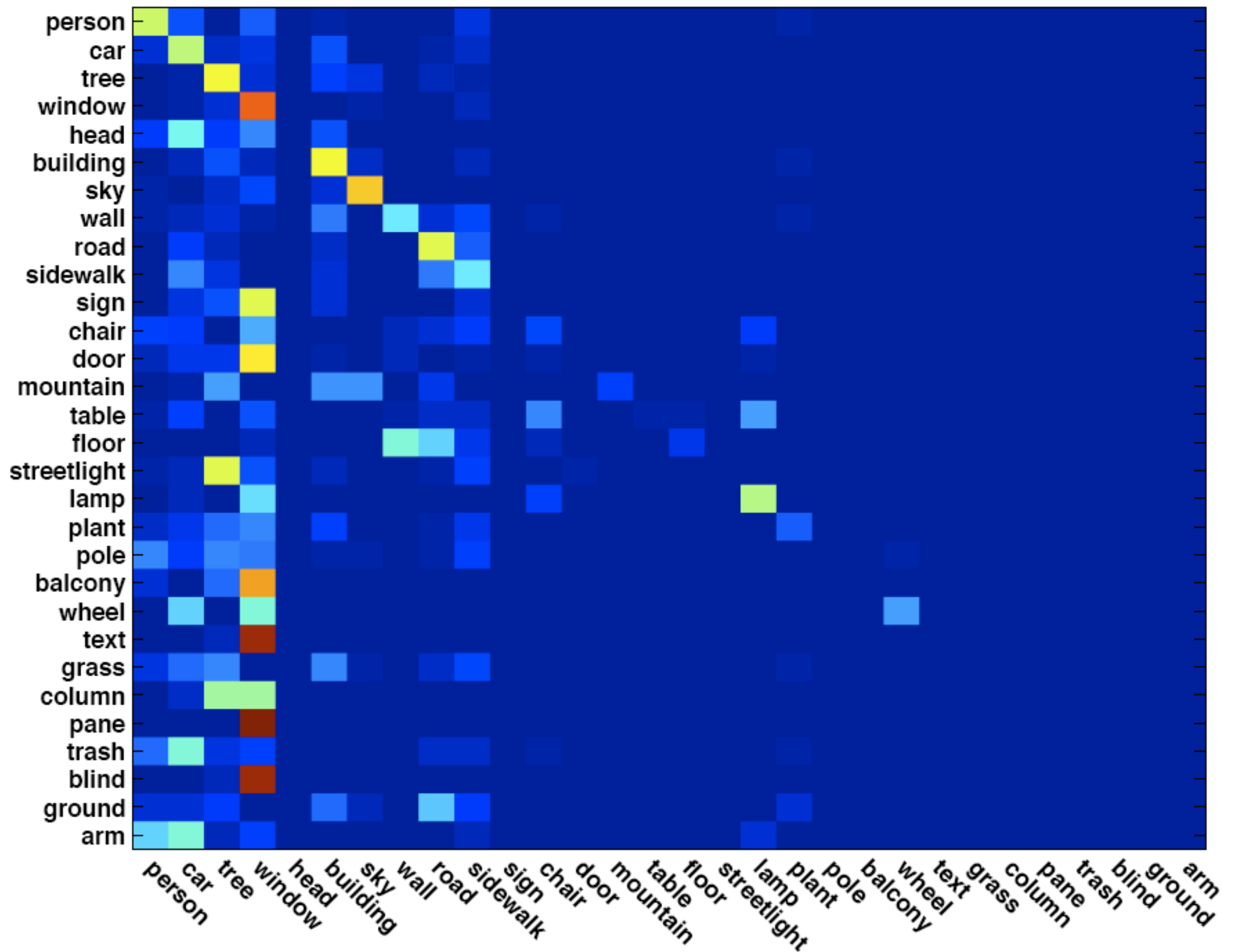
Per-Category Results

Context Challenge Recognition Accuracy for 30 Categories

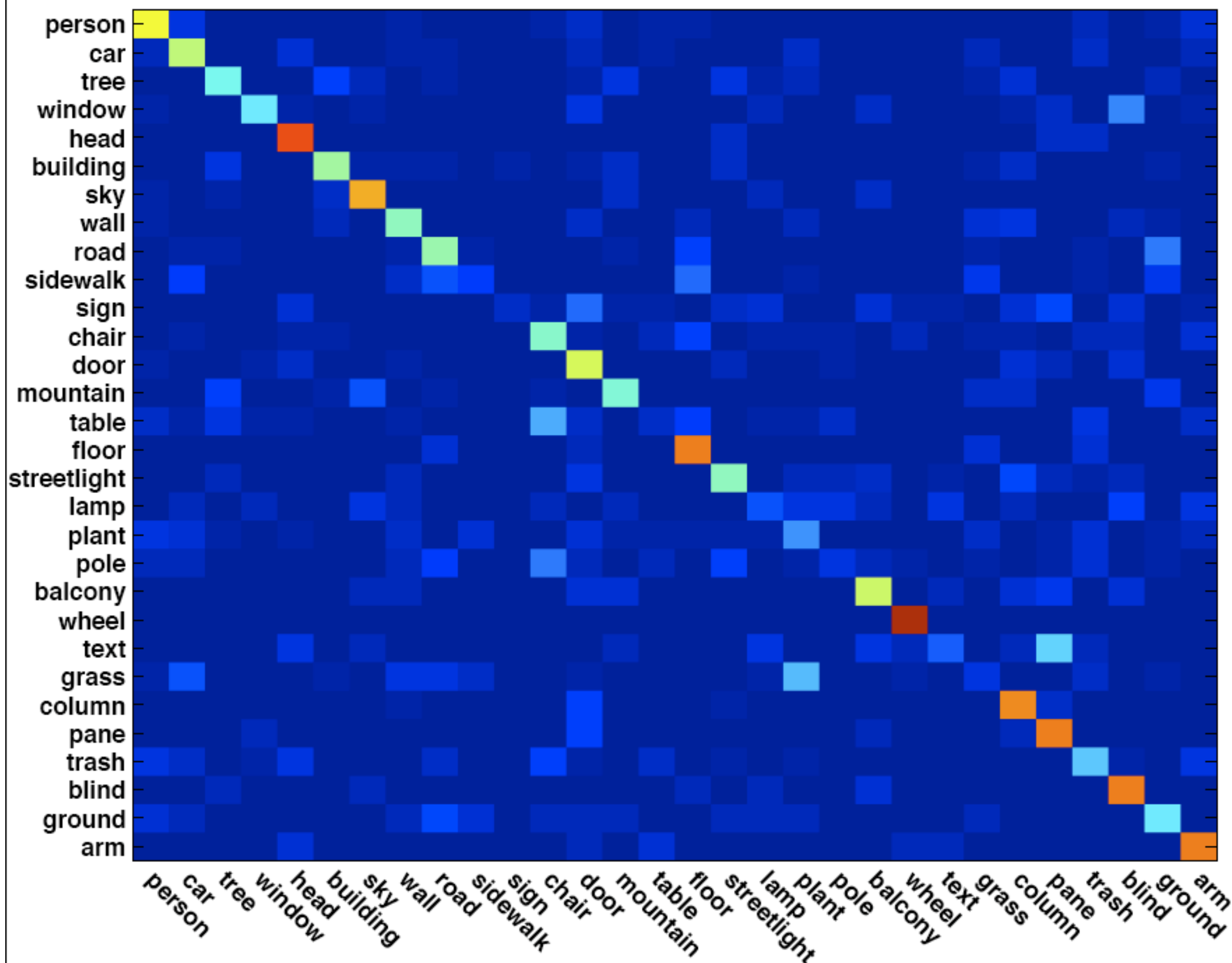




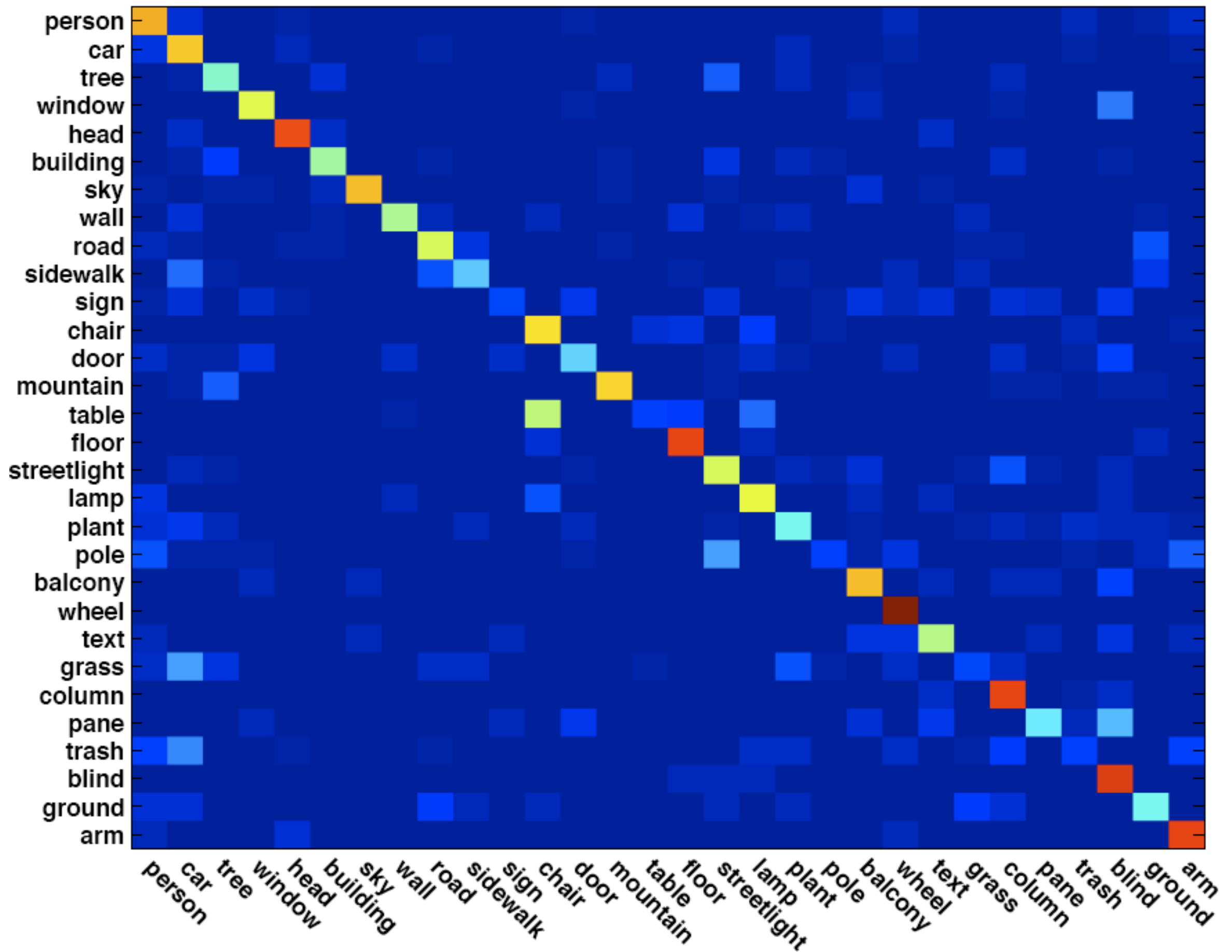
CoLA



KDE



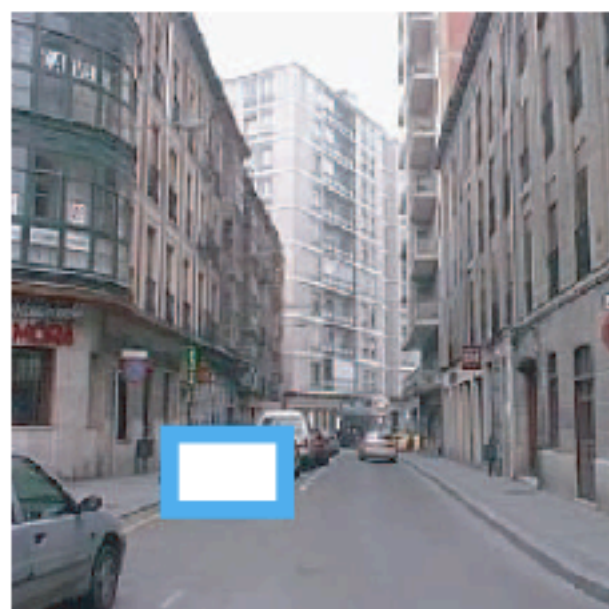
Visual Memex



Input Image + Hidden Region



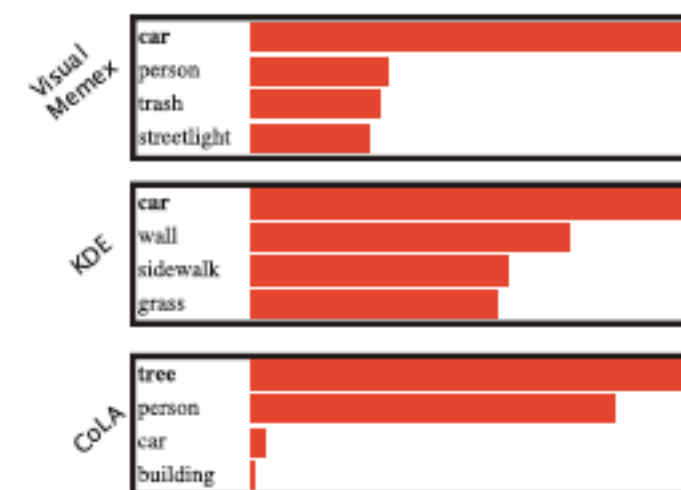
Input Image + Hidden Region

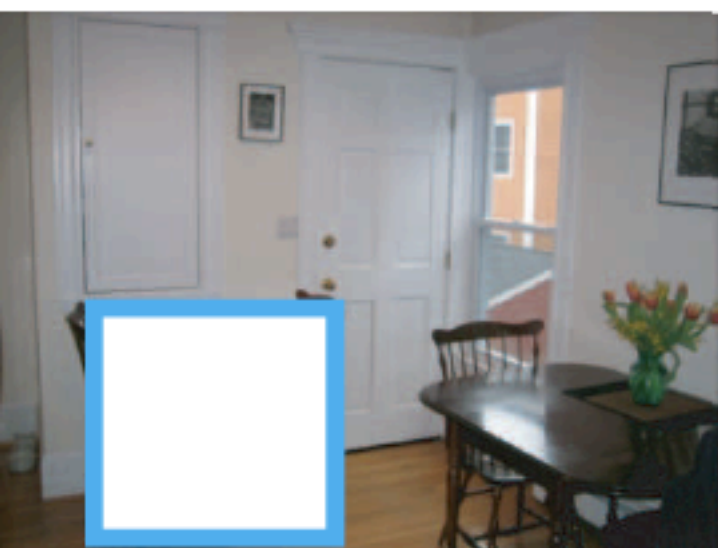


Visual Memex Exemplar Predictions

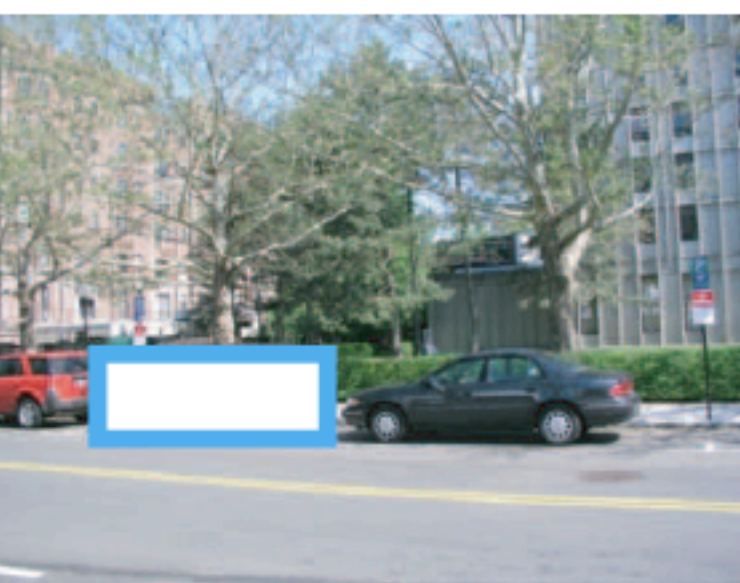


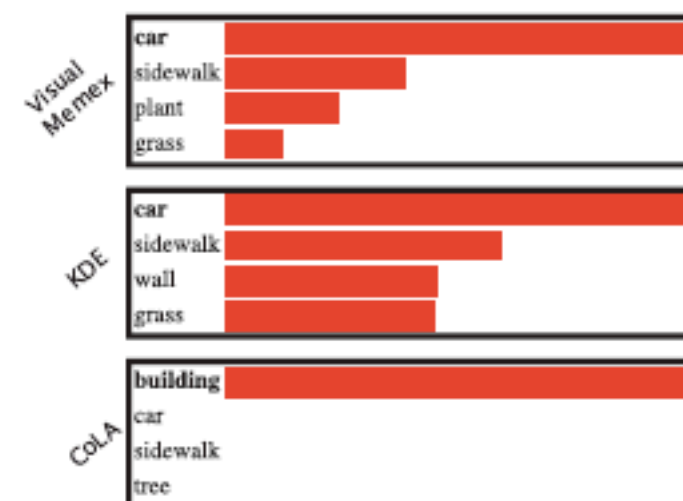
Categorization Results







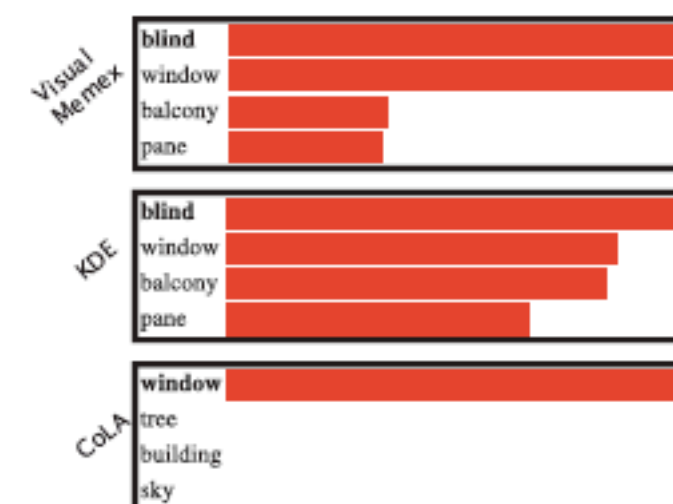


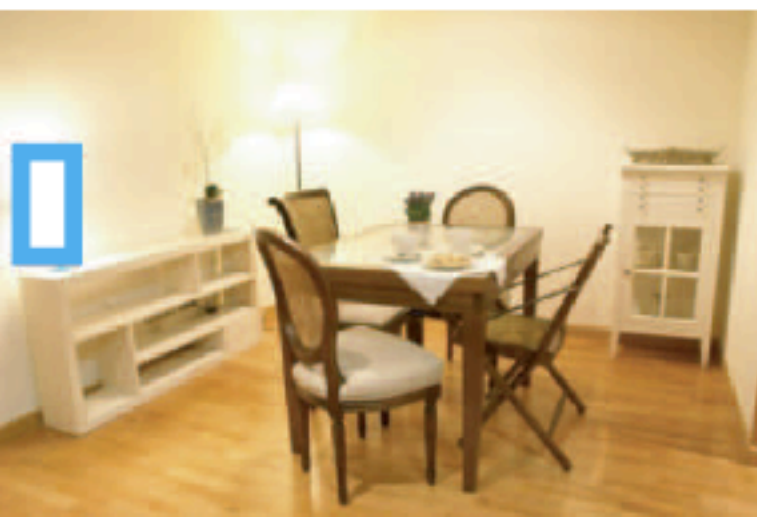


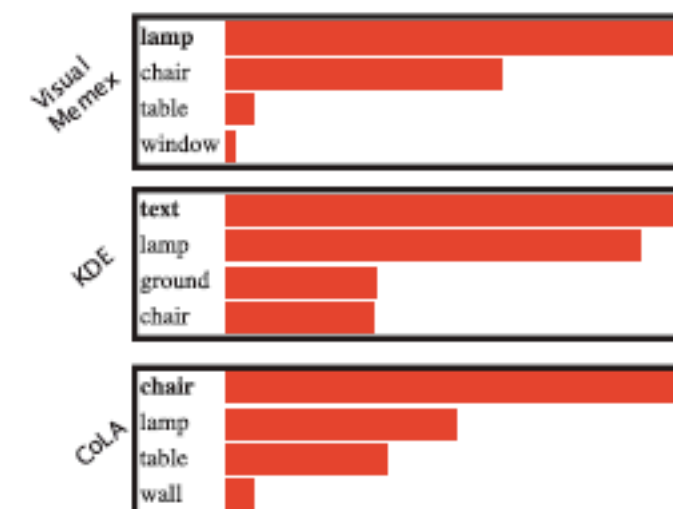
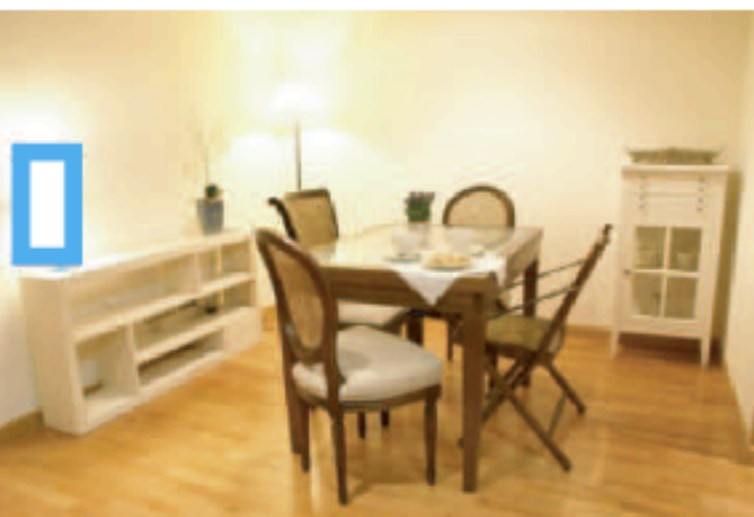












Conclusion

- Visual Memex is a Category-free approach to Modeling Object Relationships
- Is distance function learning without labels possible? (Then the entire approach would be category-free)
- Can the Visual Memex be applied to scene parsing where only an image is given?