# Probabilistic Graphical Models
# Part III: Example Applications

Selim Aksoy

Department of Computer Engineering
Bilkent University
saksoy@cs.bilkent.edu.tr

CS 551, Spring 2011

# Introduction

- We will look at example uses of Bayesian networks and Markov networks for the following applications:
  - Statistical text analysis — probabilistic latent semantic analysis
  - Scene classification — probabilistic latent semantic analysis
  - Object detection — probabilistic latent semantic analysis
  - Image segmentation — Markov random fields
  - Contextual classification — conditional random fields

# Statistical Text Analysis

- T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Machine Learning*, vol. 42, no. 1–2, pp. 177–196, January–February 2001.
- The probabilistic latent semantic analysis (PLSA) algorithm has been originally developed for statistical text analysis to discover topics in a collection of documents that are represented using the frequencies of words from a vocabulary.

# Statistical Text Analysis

- PLSA uses a graphical model for the joint probability of the documents and their words in terms of the probability of observing a word given a topic (aspect) and the probability of a topic given a document.
- Suppose there are $N$ documents having content coming from a vocabulary with $M$ words.
- The collection of documents is summarized in an $N$-by-$M$ co-occurrence table $n$ where $n(d_i, w_j)$ stores the number of occurrences of word $w_j$ in document $d_i$.
- In addition, there is a latent topic variable $z_k$ associated with each observation, an observation being the occurrence of a word in a particular document.

# Statistical Text Analysis

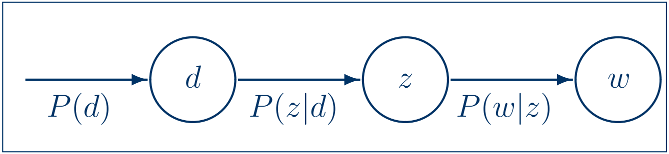



Figure 1: The graphical model used by PLSA for modeling the joint probability $P(w_j, d_i, z_k)$.

# Statistical Text Analysis

- The generative model $P(d_i, w_j) = P(d_i)P(w_j|d_i)$ for word content of documents can be computed using the conditional probability

$$P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i).$$

- $P(w_j|z_k)$ denotes the topic-conditional probability of word $w_j$ occurring in topic $z_k$.
- $P(z_k|d_i)$ denotes the probability of topic $z_k$ observed in document $d_i$.
- $K$ is the number of topics.

# Statistical Text Analysis

- Then, the topic specific word distribution $P(w_j|z_k)$ and the document specific word distribution $P(w_j|d_i)$ can be used to determine similarities between topics and documents.
- In PLSA, the goal is to identify the probabilities $P(w_j|z_k)$ and $P(z_k|d_i)$.
- These probabilities are learned using the EM algorithm.

# Statistical Text Analysis

- In the E-step, the posterior probability of the latent variables are computed based on the current estimates of the parameters as

$$P(z_k|d_i, w_j) = \frac{P(w_j|z_k)P(z_k|d_i)}{\sum_{l=1}^{K} P(w_j|z_l)P(z_l|d_i)}.$$

- In the M-step, the parameters are updated to maximize the expected complete data log-likelihood as

$$P(w_j|z_k) = \frac{\sum_{i=1}^{N} n(d_i, w_j)P(z_k|d_i, w_j)}{\sum_{m=1}^{M} \sum_{i=1}^{N} n(d_i, w_m)P(z_k|d_i, w_m)},$$

$$P(z_k|d_i) = \frac{\sum_{j=1}^{M} n(d_i, w_j)P(z_k|d_i, w_j)}{\sum_{j=1}^{M} n(d_i, w_j)}.$$

# Statistical Text Analysis

| Aspect 1 | Aspect 2 | Aspect 3 | Aspect 4 |
|---|---|---|---|
| imag | video | region | speaker |
| SEGMENT | sequenc | contour | speech |
| textur | motion | boundari | recogni |
| color | frame | descrip | signal |
| tissu | scene | imag | train |
| brain | SEGMENT | SEGMENT | hmm |
| slice | shot | precis | sourc |
| cluster | imag | estim | speakerindepend |
| mri | cluster | pixel | SEGMENT |
| algorithm | visual | paramet | sound |

Figure 2: Four aspects (topics) to most likely generate the word "segment", derived from a $K = 128$ aspects model of a document collection consisting of abstracts of $1568$ documents on clustering. The displayed word stems are the most probable words in the class-conditional distribution $P(w_j|z_k)$, from top to bottom in descending order.

# Statistical Text Analysis

Document 1, $P\{z_k|d_1, W = \text{'segment'}\} = (0.951, 0.002, 0.001, 0.0001, \ldots)$
$P\{W = \text{'segment'}|d_1\} = 0.06$

**SEGMENT** medic imag challeng problem field imag analysi diagnost base proper **SEGMENT** digit imag **SEGMENT** medic imag need applic involv estim boundari object classif tissu abnorm shape analysi contour detec textur **SEGMENT** despit exist techniqu **SEGMENT** specif medic imag remain crucial problem [...]

Document 2, $P\{z_k|d_2, W = \text{'segment'}\} = (0.025, 0.956, 0.0002, 0.0002, \ldots)$
$P\{W = \text{'segment'}|d_2\} = 0.014$

describ new techniqu extract hierarch decomposi complex video selec brows purpos techniqu combin visual tempor inform captur import relat scene scene video allow analysi underli stori structur priori knowledg content defin gener model hierarch scene transition graph appli model implement brows video shot identifi collec kei frame repres video **SEGMENT** collec classifi accord gross visual inform [...]

Document 3, $P\{z_k|d_3, W = \text{'segment'}\} = (0.025, 0.003, 0.897, 0.016, \ldots)$
$P\{W = \text{'segment'}|d_3\} = 0.010$

paper describ contour extrac scheme refin roughli estim initi contour outlin precis object boundari author approach mixtur densiti descrip parametr descrip decompos subregion obtain region cluster descrip likelihood pixel belong object background evalu unlik activ contour extrac scheme region edgebas estim scheme integr energi minim process [...]

Document 4, $P\{z_k|d_3, W = \text{'segment'}\} = (0.025, 0.076, 0.001, 0.867, \ldots)$
$P\{W = \text{'segment'}|d_4\} = 0.010$

consid signal origin sequenc sourc specif problem **SEGMENT** signal relat **SEGMENT** sourc address issu wide applic field report describ resolu method ergod hidden markov model hmm hmm state correspond signal sourc signal sourc sequenc determin decod procedur viterbi algorithm forward algorithm observ sequenc baumwelch train estim hmm paramet train materi applic multipl signal sourc identif problem experi perform unknown speaker identif [...]

Figure 3: Abstracts of four examplary documents from the collection along with latent class posterior probabilities $P(z_k|d_i, w = \text{“segment”})$ and word probabilities $P(w = \text{“segment”}|d_i)$.

# Scene Classification

- P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, "A Thousand Words in a Scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1575–1589, September 2007.
- The PLSA model is used for scene classification by modeling images using visual words (visterms).
- The topic (aspect) probabilities are used as features as an alternative representation to the word histograms.

# Scene Classification

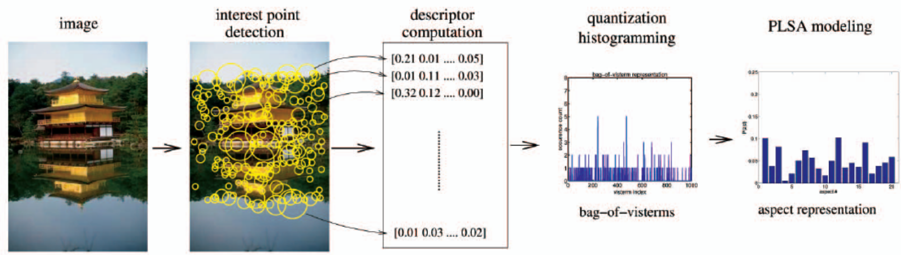



Figure 4: Image representation as a collection of visual words (visterms).

# Scene Classification

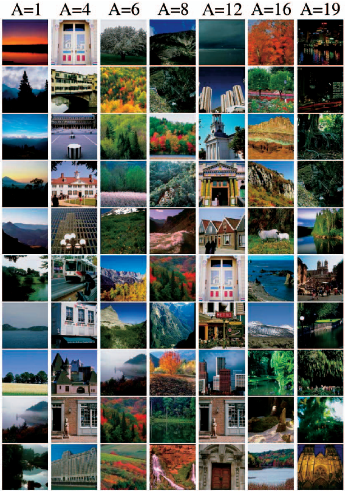



Figure 5: 10 most probable images from a data set consisting of city and landscape images for seven topics (aspects) out of 20.

# Object Detection

- H. G. Akcay, S. Aksoy, “Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, July 2008.
- We used the PLSA technique for object detection to model the joint probability of the segments and their features in terms of the probability of observing a feature given an object and the probability of an object given the segment.

# Object Detection

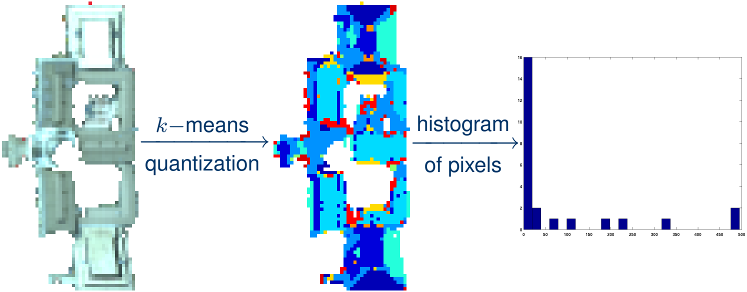



Figure 6: After image segmentation, each segment is modeled using the statistical summary of its pixel content (e.g., quantized spectral values).

# Object Detection

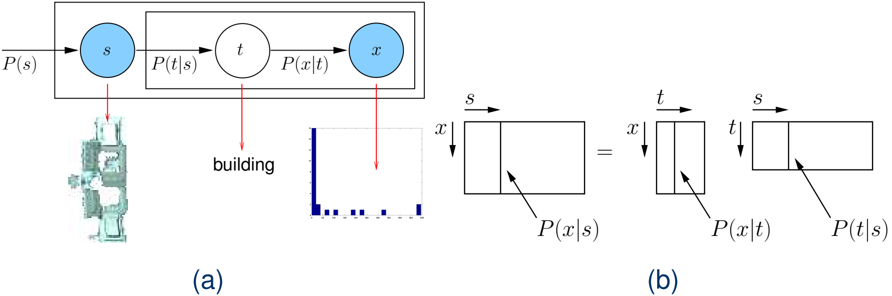
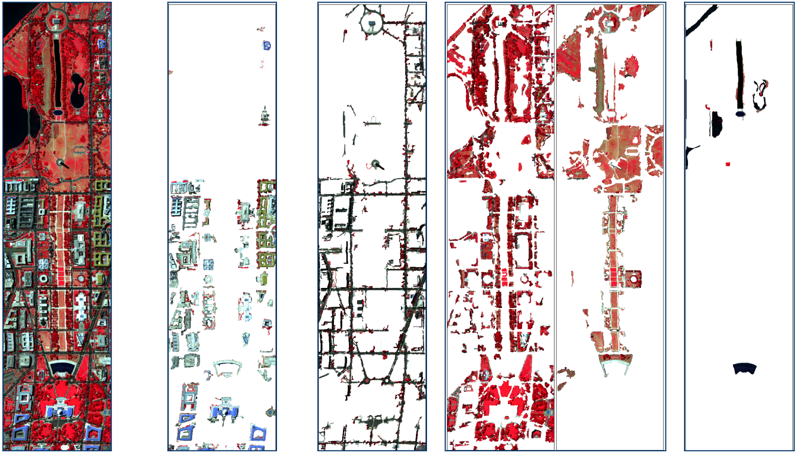



Figure 7: (a) PLSA graphical model. The filled nodes indicate observed random variables whereas the unfilled node is unobserved. The red arrows show examples for the measurements represented at each node. (b) In PLSA, the object specific feature probability, $P(x_j|t_k)$, and the segment specific object probability, $P(t_k|s_i)$, are used to compute the segment specific feature probability, $P(x_j|s_i)$.

# Object Detection

- After learning the parameters of the model, we want to find good segments belonging to each object type.
- This is done by comparing the object specific feature distribution $P(x|t)$ and the segment specific feature distribution $P(x|s)$.
- The similarity between two distributions can be measured using the Kullback-Leibler (KL) divergence $D(p(x|s)\|p(x|t))$.
- Then, for each object type, the segments can be sorted according to their KL divergence scores, and the most representative ones for that object type can be selected.

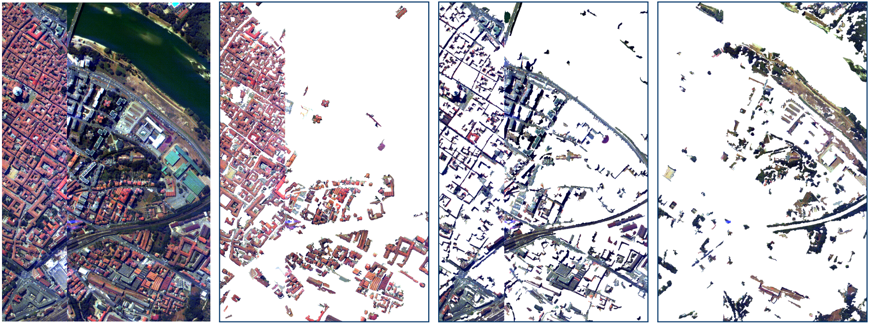
# Object Detection



(a) Image (b) Buildings (c) Roads (d) Vegetation (e) Water

Figure 8: Examples of object detection.

# Object Detection



(a) Image (b) Buildings (c) Roads (d) Vegetation

Figure 9: Examples of object detection.

# Image Segmentation

- Z. Kato, T.-C. Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, no. 10, pp. 1103–1114, October 2006.
- Markov random fields are used as a neighborhood model for image segmentation by classifying pixels into different pixel classes.

# Image Segmentation

- The goal is to assign each pixel into a set of labels $w \in \Omega$.
- Pixels are modeled using color and texture features.
- Pixel features are modeled using multivariate Gaussians, $p(\mathbf{x}|w)$.
- A first-order neighborhood system is used as the prior for the labeling process.

# Image Segmentation

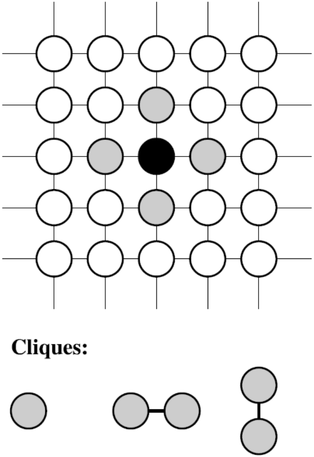



Figure 10: The Markov random field used as the first-order neighborhood model for the labeling process.

# Image Segmentation

- The prior is modeled as

$$p(w) = \frac{1}{Z} \exp\left( -\sum_{c \in \mathcal{C}} V_c(w_c) \right)$$

where $V_c$ denotes the clique potential of clique $c \in \mathcal{C}$ having the label configuration $w_c$.

- Each clique corresponds to a pair of neighboring pixels.
- The potentials favor similar classes in neighboring pixels as

$$V_c = \delta(w_s, w_r) = \begin{cases} +1 & \text{if } w_s \neq w_r, \\ -1 & \text{otherwise.} \end{cases}$$

# Image Segmentation

- The prior is proportional to the length of the region boundaries. Thus, homogeneous segmentations will get a higher probability.
- The final labeling for each pixel is done by maximizing the posterior probability

$$p(w|\mathbf{x}) \propto p(\mathbf{x}|w)p(w).$$

# Image Segmentation

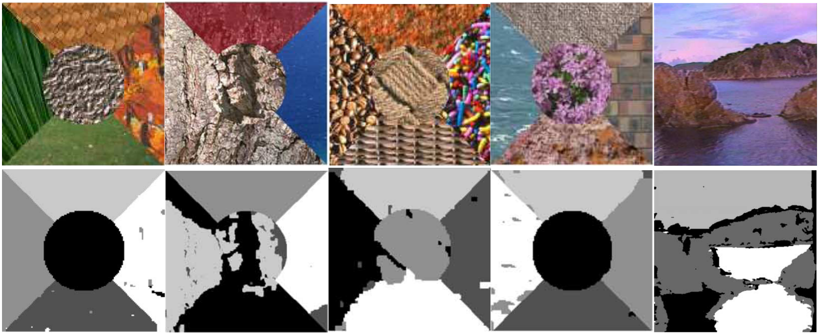


Figure 11: Example segmentation results.

# Image Segmentation

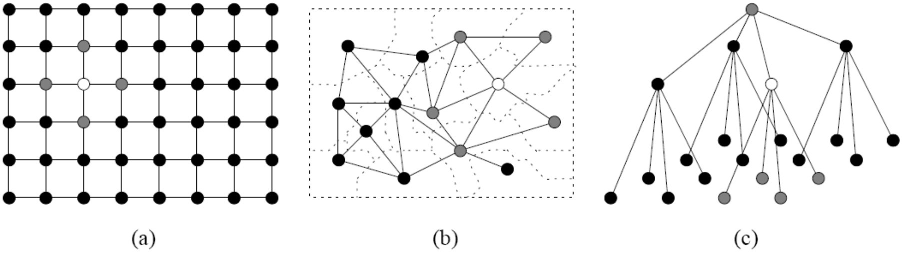



Figure 12: Example Markov random field models used in the literature. (a) First-order neighborhood system. (b) Non-regular planar graph associated to an image partition. (c) Quad-tree.

# Contextual Classification

- A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, S. Belongie, “Objects in Context,” *IEEE International Conference on Computer Vision*, 2007.
- Semantic context among objects is used for improving object categorization.

# Contextual Classification

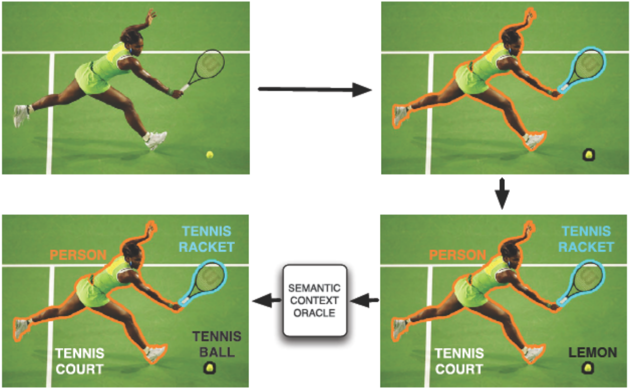



Figure 13: Idealized context based object categorization system: an original image is perfectly segmented into objects; each object is categorized; and object's labels are refined with respect to semantic context in the image.

# Contextual Classification

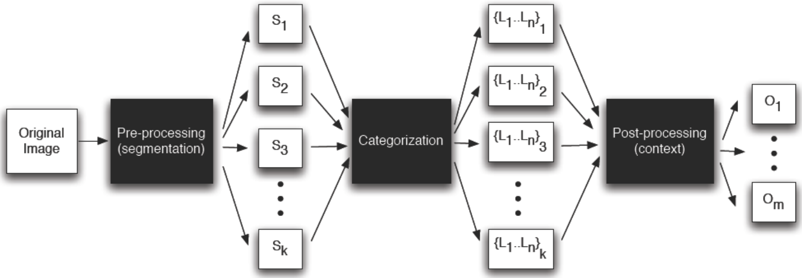



Figure 14: Object categorization framework: $S_1, \ldots, S_k$ is the set of $k$ segments for an image; $L_1, \ldots, L_n$ is a ranked list of $n$ labels for each segment; $O_1, \ldots, O_m$ is a set of $m$ object categories in the image.

# Contextual Classification

- A conditional random field (CRF) framework is used to incorporate semantic context into the object categorization.
- Given an image $I$ and its segmentation $S_1, \ldots, S_k$, the goal is to find segment labels $c_1, \ldots, c_k$ such that they agree with the segment contents and are in contextual agreement with each other.

# Contextual Classification

- This interaction is modeled as a probability distribution

$$p(c_1, \ldots, c_k | S_1, \ldots, S_k) = \frac{B(c_1, \ldots, c_k) \prod_{i=1}^{k} A(i)}{Z(\phi, S_1, \ldots, S_k)}$$

  with

$$A(i) = p(c_i | S_i) \text{ and } B(c_1, \ldots, c_k) = \exp\left( \sum_{i,j=1}^{k} \phi(c_i, c_j) \right),$$

  where $Z(\cdot)$ is the partition function.
- The semantic context information is modeled using context matrices that are symmetric, nonnegative matrices that contain the co-occurrence frequency among object labels in the training set.

# Contextual Classification

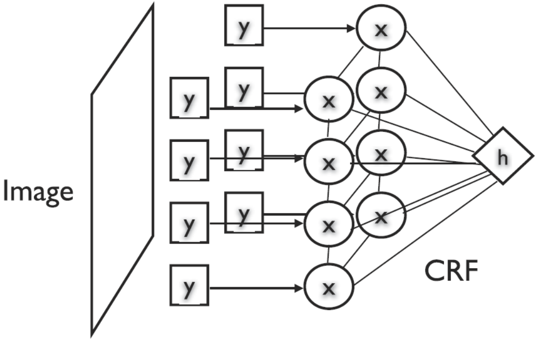



Figure 15: An example conditional random field. Squares indicate feature functions and circles indicate variable nodes. Arrows represent single node potentials due to feature functions, and undirected edges represent pairwise potentials. Global context is represented by $h$.

# Contextual Classification

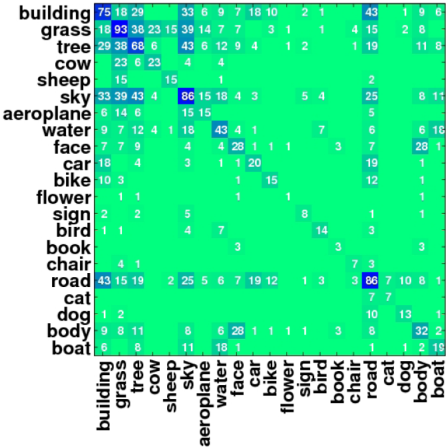


Figure 16: An example context matrix.

# Contextual Classification

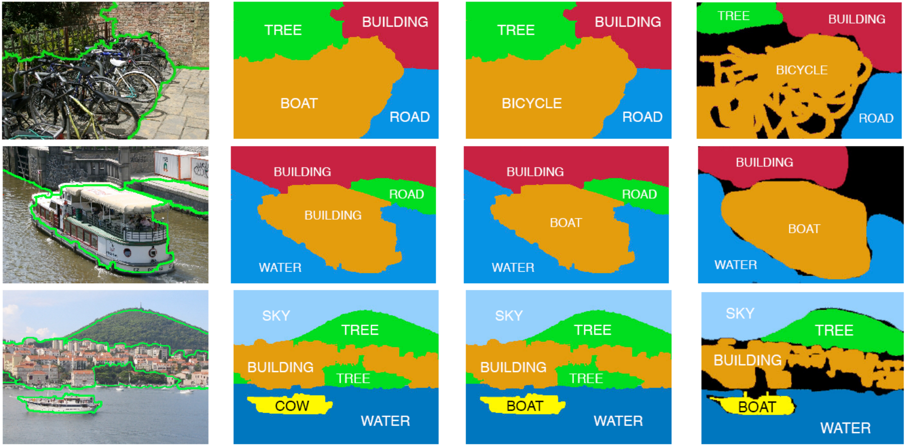



Figure 17: Example results where context improved the categorization accuracy. Left to right: original segmentation, categorization w/o contextual constraints, categorization w/ contextual constraints, ground truth.

# Contextual Classification

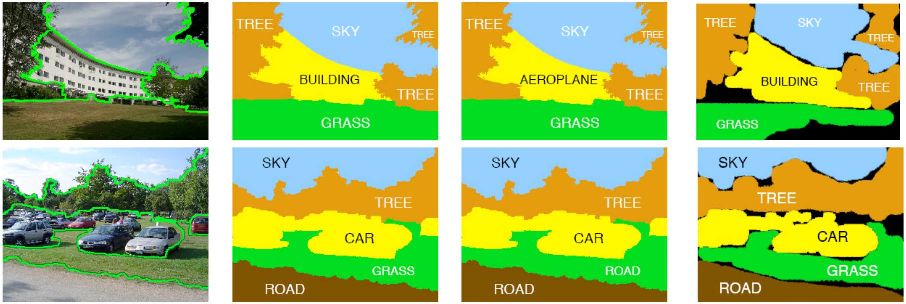



Figure 18: Example results where context reduced the categorization accuracy. Left to right: original segmentation, categorization w/o contextual constraints, categorization w/ contextual constraints, ground truth.