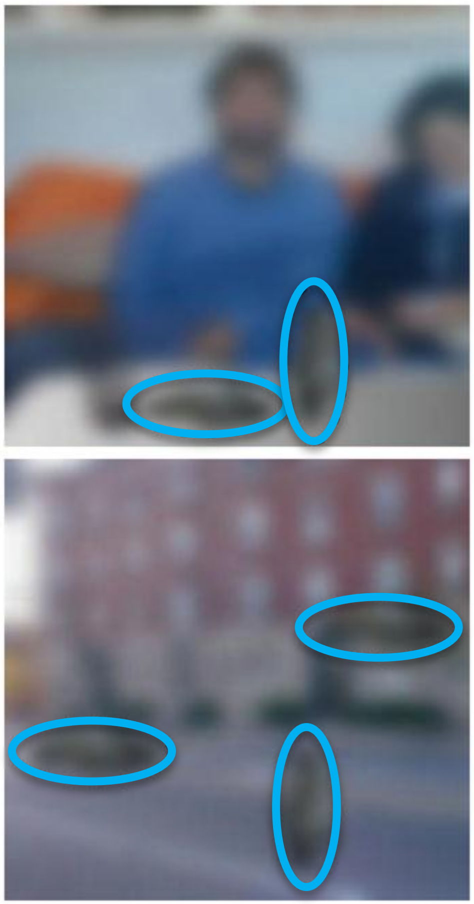
CS 395T – Visual Recognition Context and Scenes

Aron Yu

Oct 5, 2012

Context in Detection



Myopic View of
Local Objects

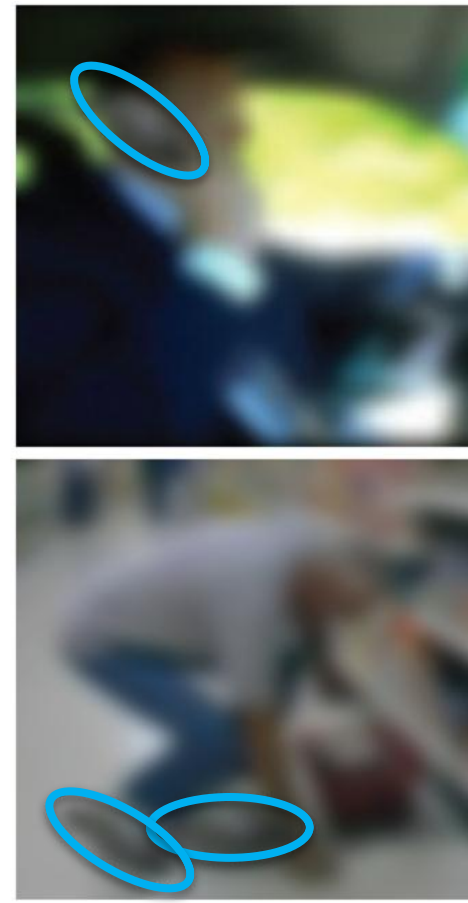
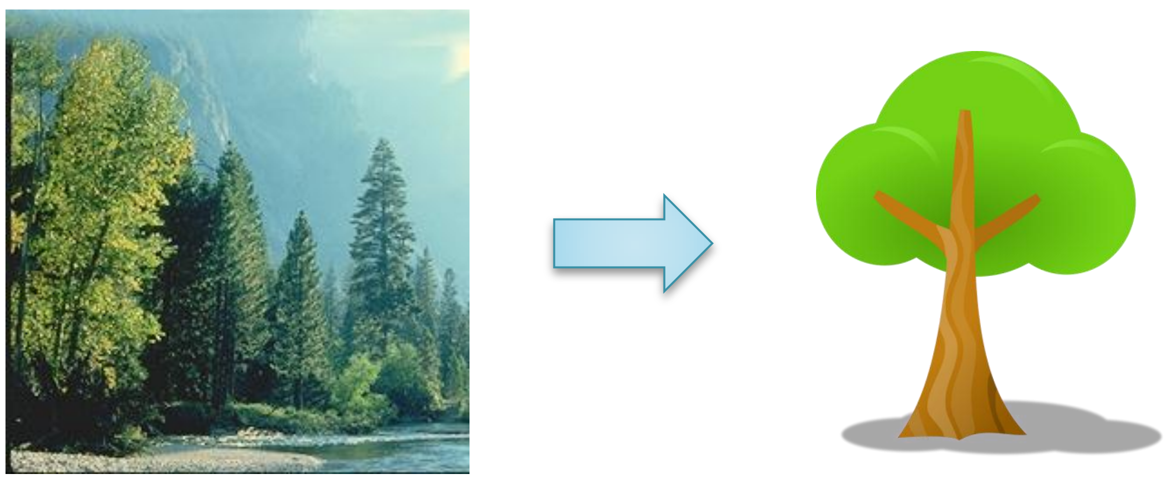


Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Using the Forest to See the Trees
Exploiting Context for Visual Object Detection and Localization

A. Torralba, K. P. Murphy, W. T. Freeman



Overview

- Past Approach
  - associate objects with other objects in the image
  - presence of pedestrians given presence of cars

- Holistic Approach
  - associate objects with the scene category as a single entity
  - presence of pedestrians given street scene

Image Credit: A. Olivia, A. Torralba. – Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope

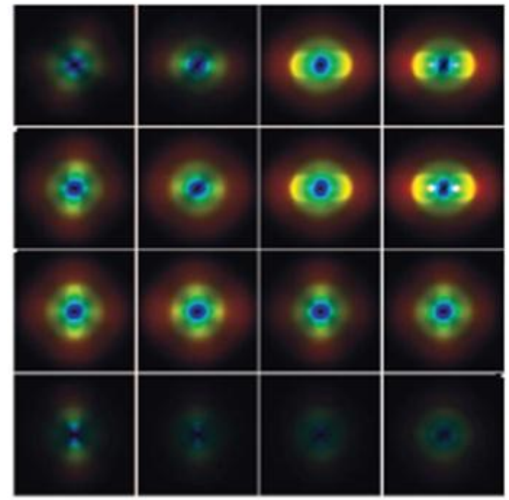
Goals

- Use global features to obtain contextual prior for object categories
- Develop a probabilistic framework for combining local (bottom-up) and global (top-down) features

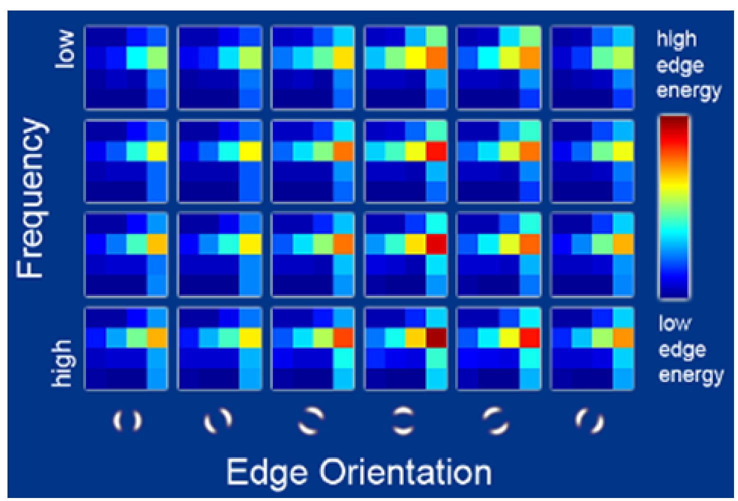
- Target Problems
  - object presence detection (is there a car?)
  - object localization (where is the car?)

Global Feature - GIST

Polar Form



Spatial Envelope



Vectorize

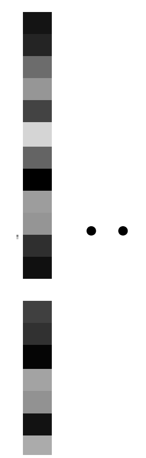


Image Idea: http://cybertron.cg.tu-berlin.de/pdci11ws/scene_completion/implementation.html

Local Features

- Preview of next week's paper on multiclass and multiview object detection
  - local feature boosting using part-model

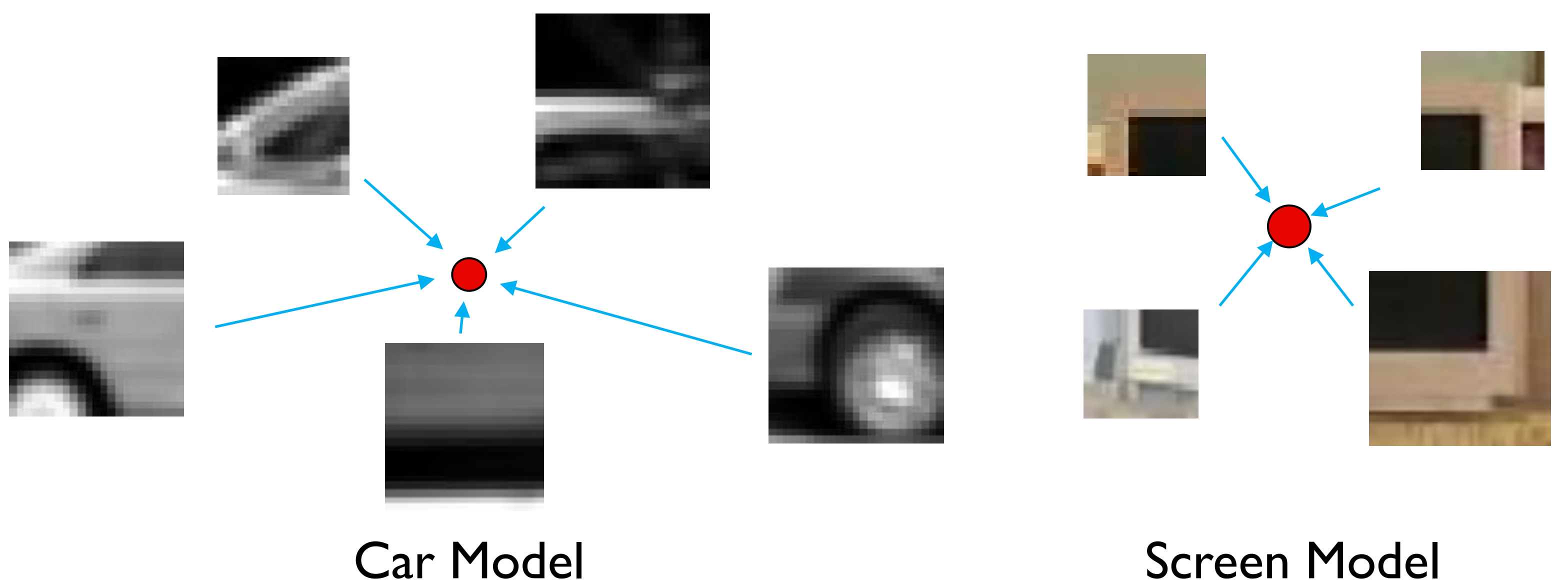


Image Credit: http://people.csail.mit.edu/torralba/shortCourseRLOC/CVPR2007_part3.ppt

Local Features

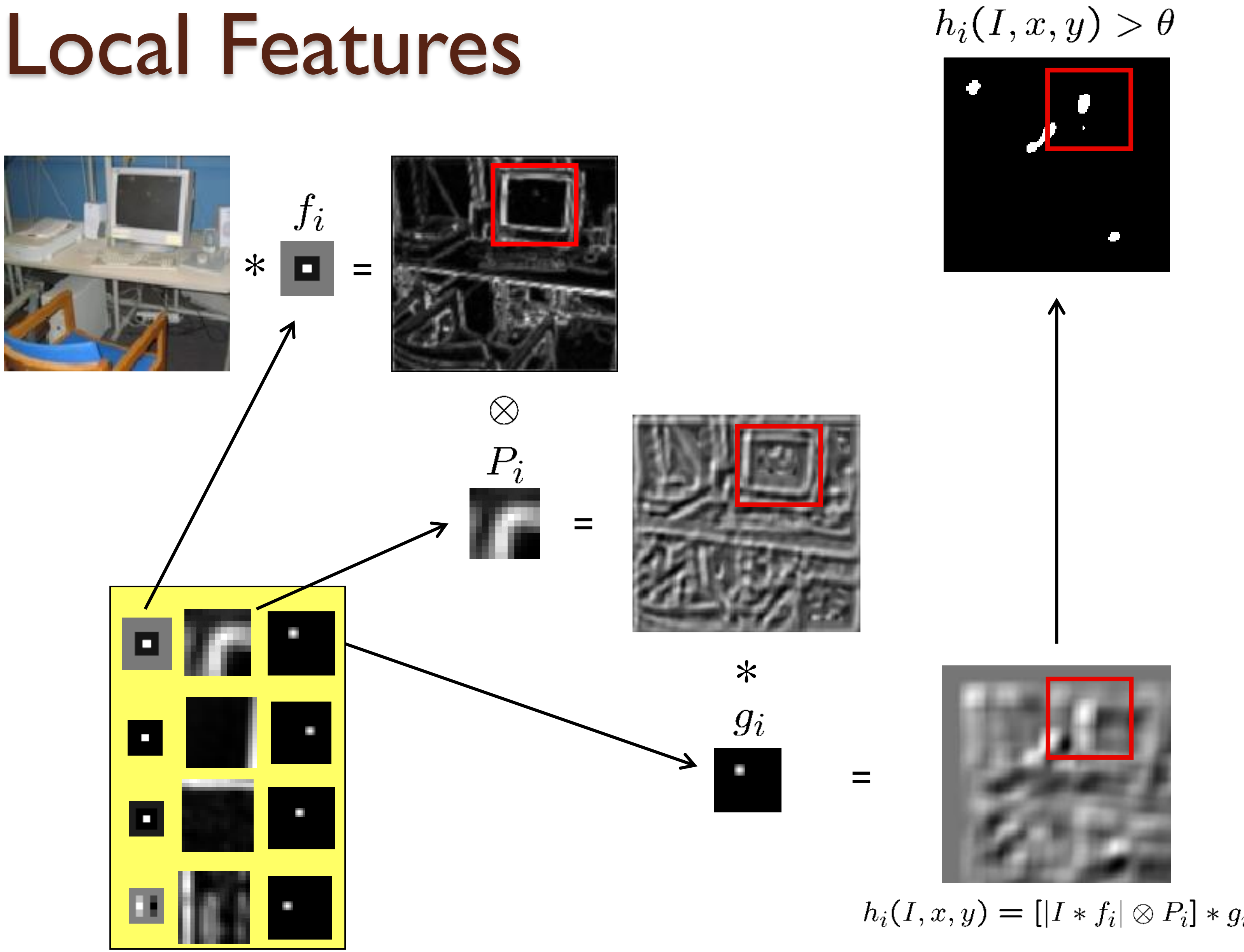


Image Credit: http://people.csail.mit.edu/torralba/shortCourseRLOC/CVPR2007_part3.ppt

8-Scene Dataset

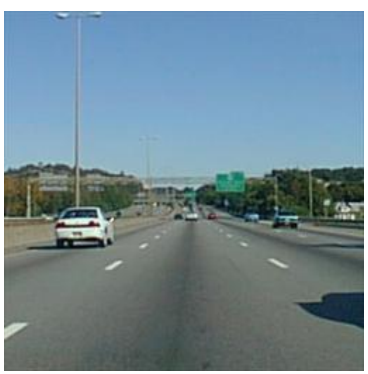
Coast

Fields

Forests

Mountains

Highways



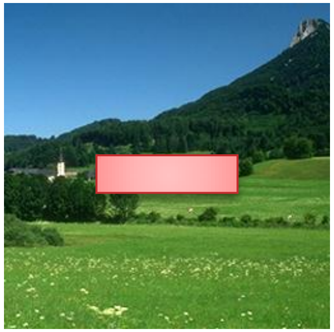
Streets

Inside City

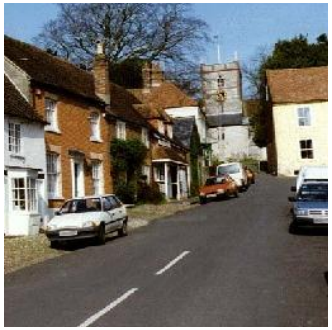
Skyscrapers

Image Credit: A. Olivia, A. Torralba. – Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope

Object Presence



- Is the object in the image?
  - binary classifier
  - prob. of object presence given gist
- How many objects are in the image?
  - categorize the scene from gist (quantization)
  - prob. of having *n* objects given scene



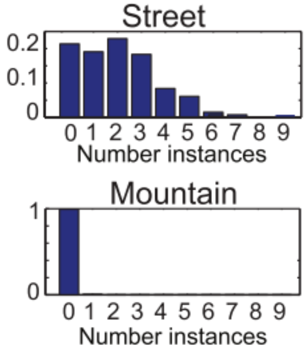
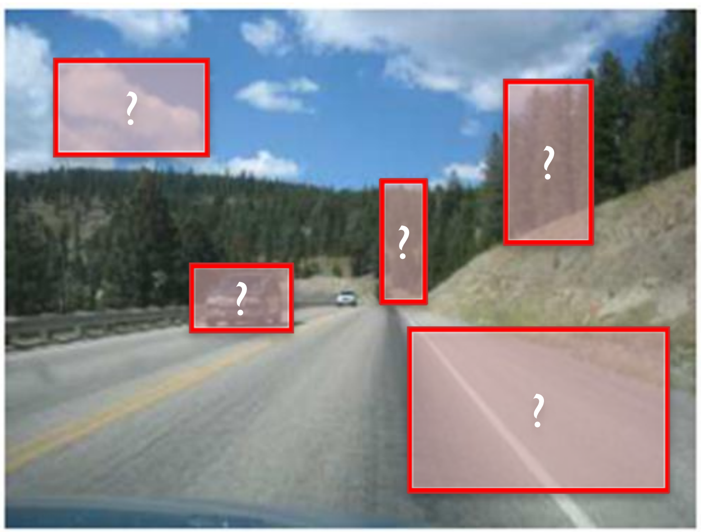


Image Credit: A. Olivia, A. Torralba. – Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope

Object Localization

- Where are the objects?
  - local feature descriptors
  - confidence score c_i^t per region i
  - use top D (~10) most confident regions for evaluation



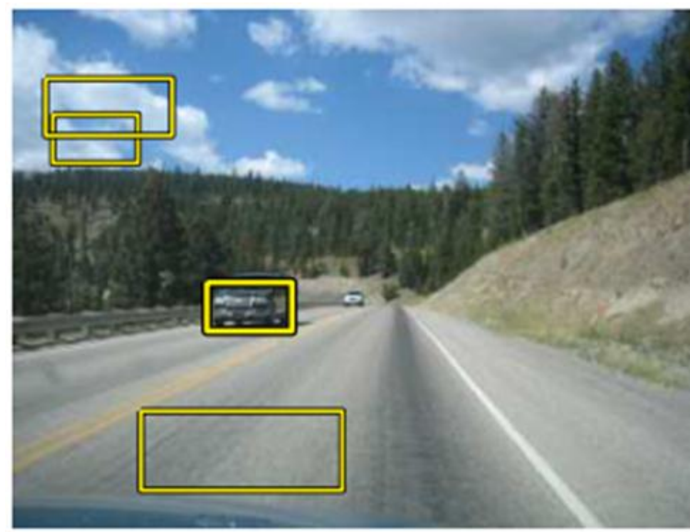


Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Object Localization

- Location trimming using gist
  - mixture of experts model
  - predict most likely vertical location
  - “mask out” unlikely regions for individual objects

Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Integrated Model

- Combine global and local features
- Without location trimming
- Prob. of object being present given confidence scores and gist
  - find the number of objects present using gist (global)
  - show that many confidence scores (local)

$$p(O^t_{1:D} | c^t_{1:D}, g) \propto \underbrace{p(O^t_{1:D} | g)}_{\text{presence of object given gist}} \prod_{i=1}^{D} \underbrace{p(c^t_i | O^t_i)}_{\text{confidence scores given presence of object}}$$

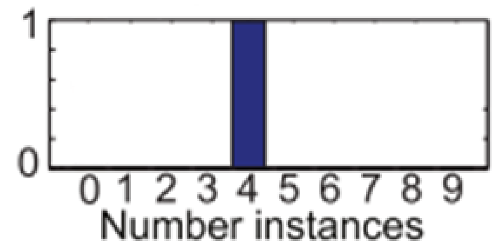
Integrated Model

- With location trimming
- Add a location term l_i^t given presence of object and gist
    - suppress or boost confidence scores according to location of confidence region

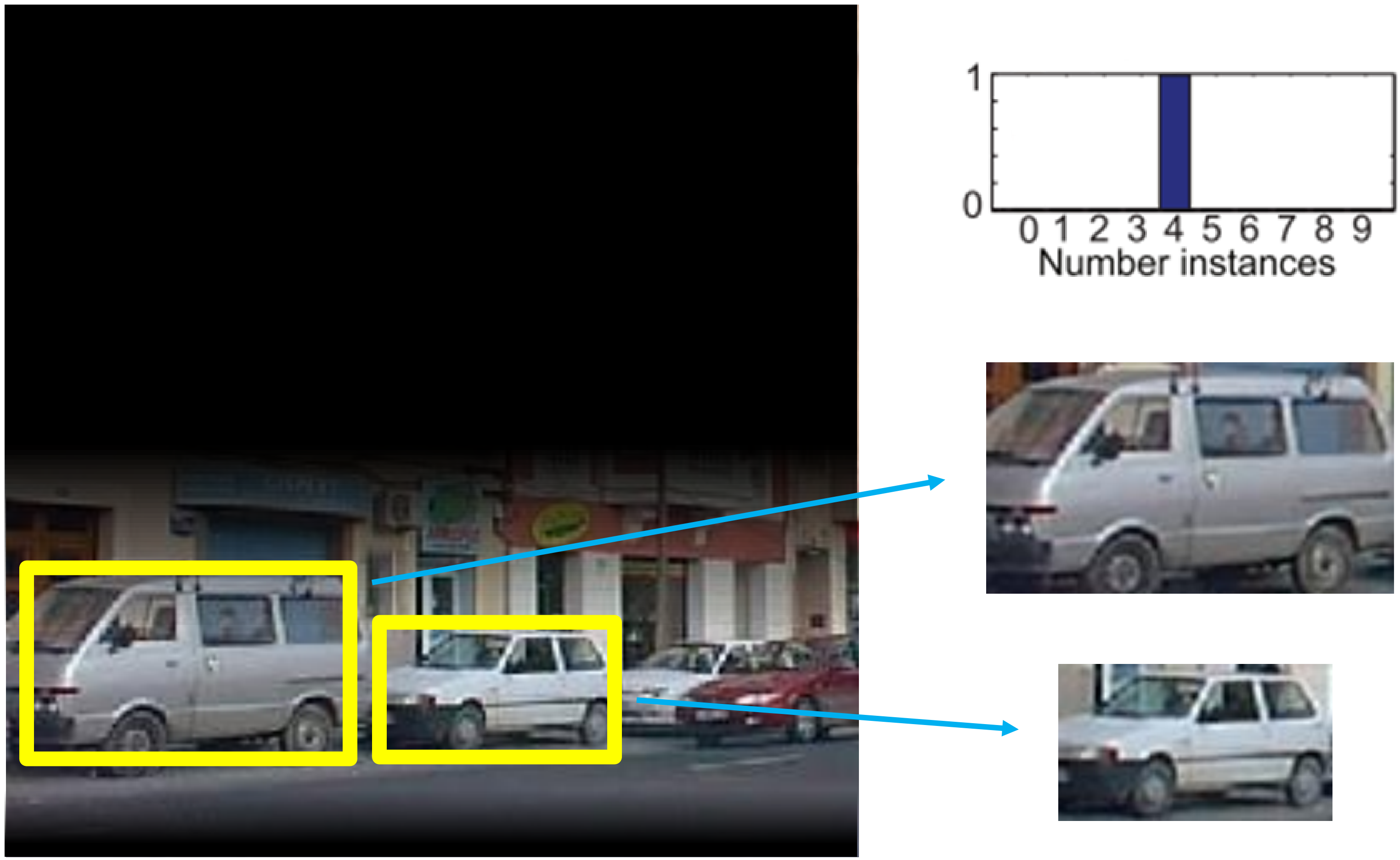
$$p(O^t_{1:D} | c^t_{1:D}, \ell^t_{1:D}, g) \propto p(O^t_{1:D} | g) \prod_{i=1}^{D} p(c^t_i | O^t_i) \underbrace{p(\ell^t_i | O^t_i, g)}_{\text{location given presence of object and gist}}$$

Toy Demo

Toy Demo



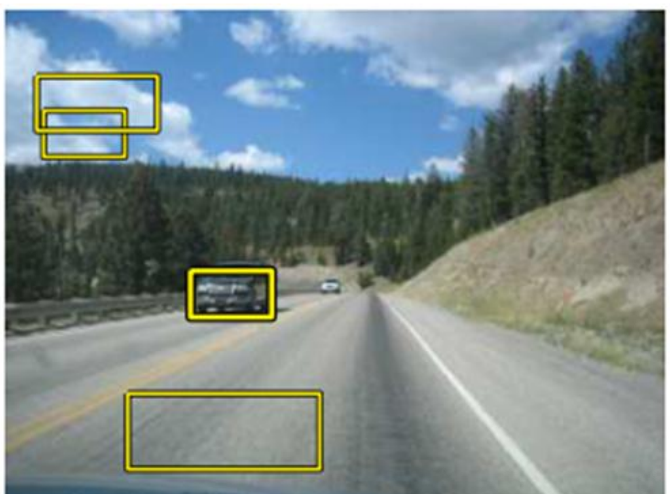
Toy Demo



Results

- 2688 images with 8 scenes
  - half for training, half for testing
- Focused solely on car identification
- Integrated model is better than local features only

Local



Integrated

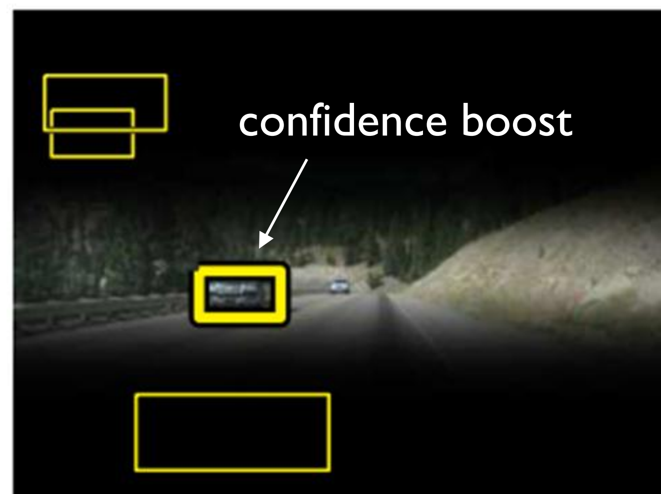
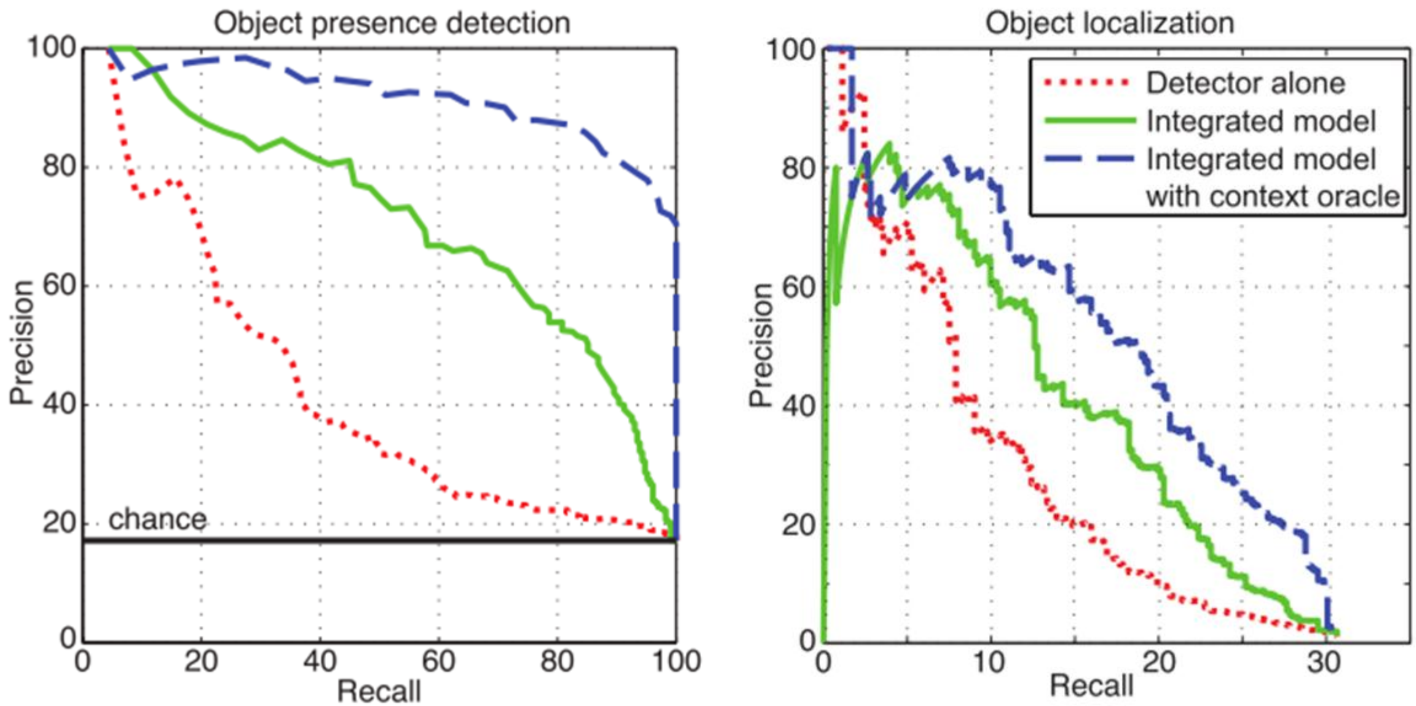


Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

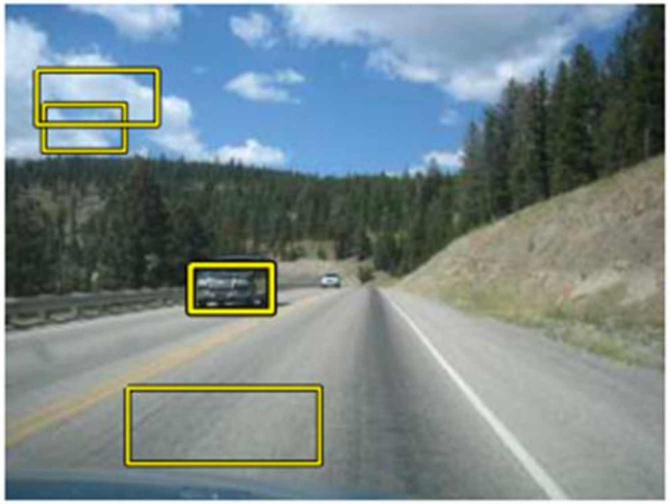
Results



- Improves precision but not recall
  - removes false-positives
- Context oracle doesn't improve the performance as much for localization

Evaluation - Strength

- Probabilistic information fusion
- Boost confidence of probable regions
    - suppress confidence of non-probable regions
- Location priming makes intuitive sense
- Better performance than with only local features



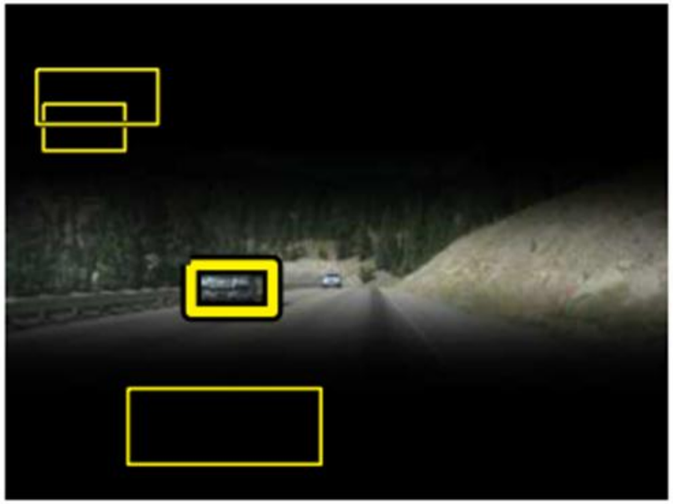


Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Evaluation - Weakness

- Tested with only cars
- Boost false positives within probable regions
- 75% accuracy on scene detector
  - better than object detector but not perfect
- Still relies heavily on object detector accuracy

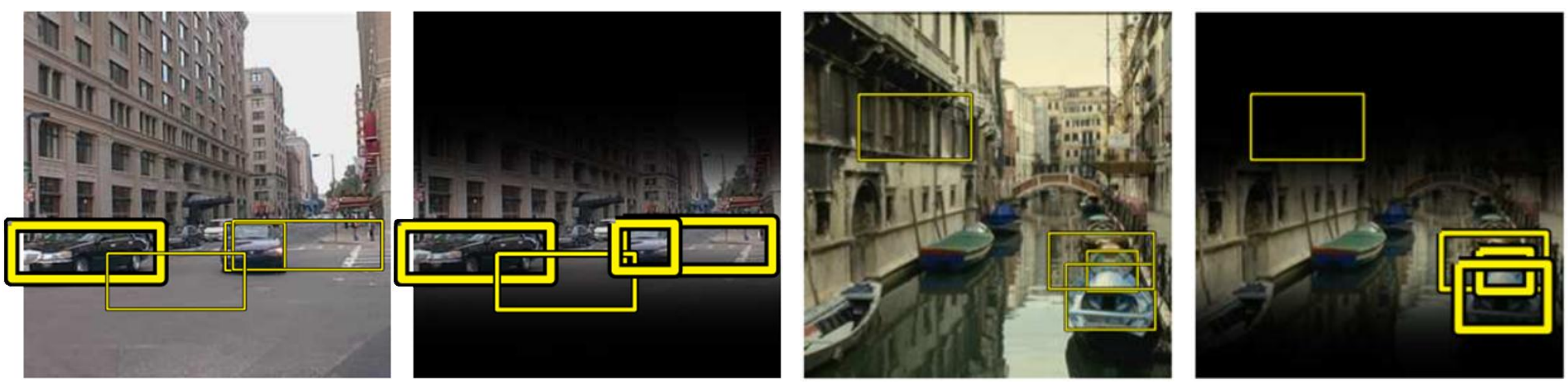


Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Evaluation - Weakness

- Scenes with less spatial regularity
  - suppress true positives within non-probable regions

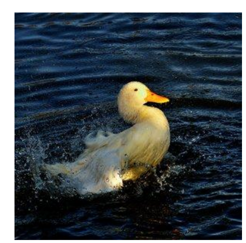


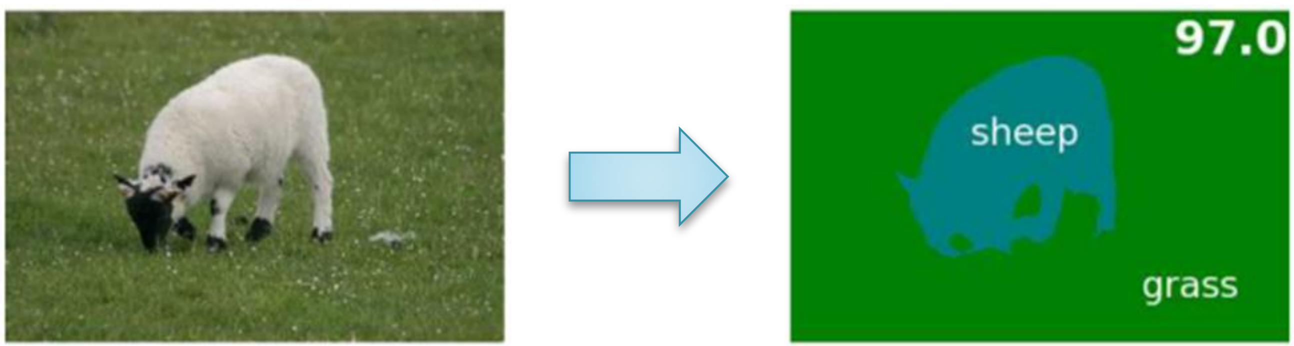
Image Credit: A. Torralba, K.P. Murphy, W.T. Freeman – Using Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization

Summary

- Overall thoughts
  - successfully incorporated scene information into a probabilistic model
  - scene context helps to reduce false-positives
  - localization is much harder than presence detection
  - object detector accuracy is still crucial
- Extension
  - more datasets, more objects
  - multiple objects in the same image

Multi-Class Segmentation with Relative Location Prior

S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller



Overview

- Multiclass image segmentation
  - classify all pixels in an image
  - leverage context information for spatial relationships

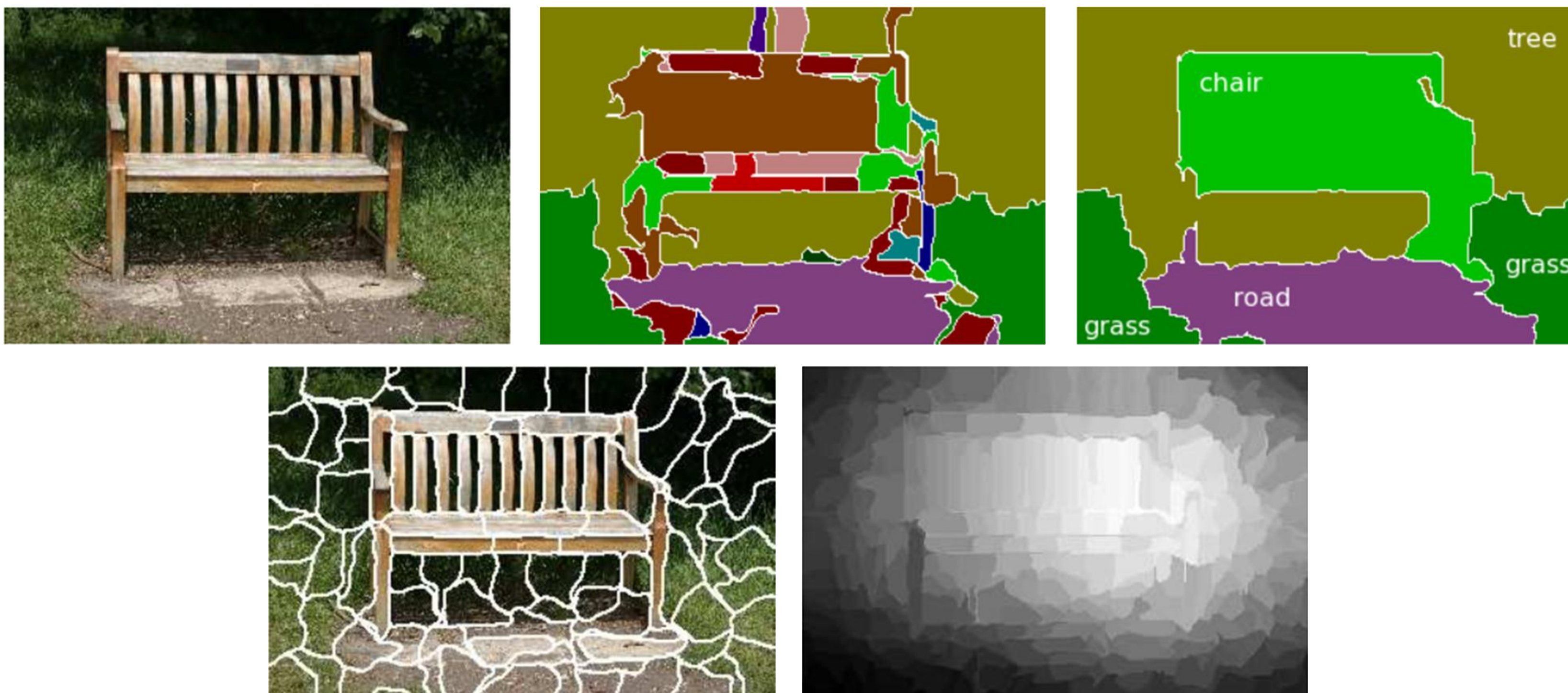
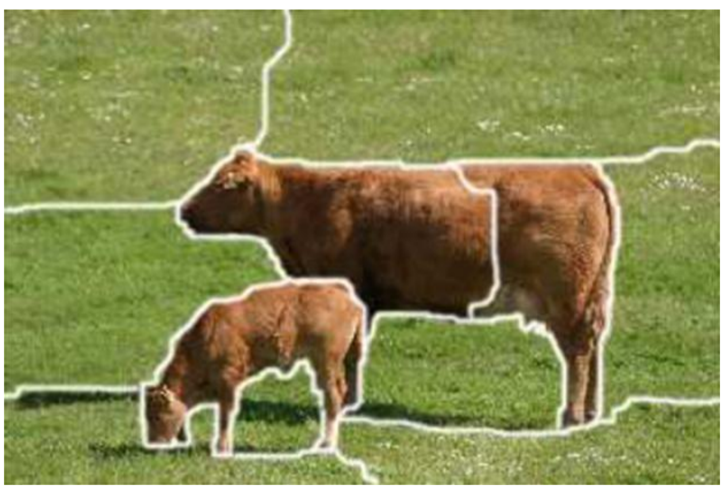


Image Credit: S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller – Multi-class Segmentation with Relative Location Prior

Conditional Random Fields

- Discriminative undirected probabilistic model
  - pair-wise neighbors, no long range dependencies
- Node potentials
  - object class likelihood
- Pair-wise potentials
  - label smoothness preference



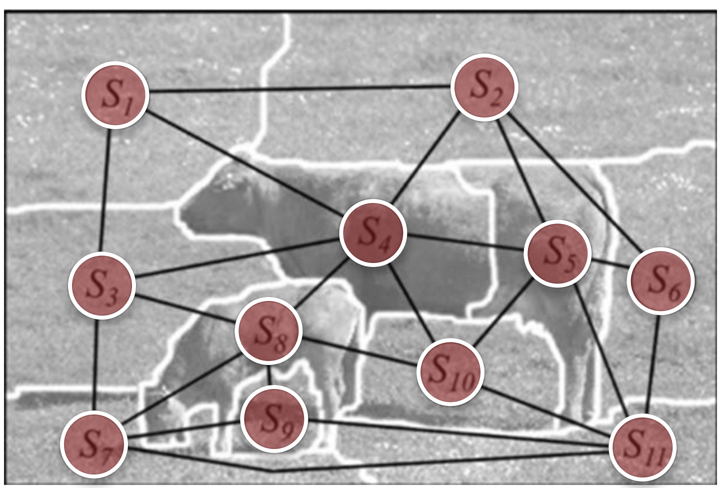


Image Credit: S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller – Multi-class Segmentation with Relative Location Prior

Relative Location Prior

- Encodes relative location between object classes
  - conditioned on offset $p(c_i = car | c_j = road, D(S_i, S_j))$
- Each superpixel votes for the most likely label for all other superpixels

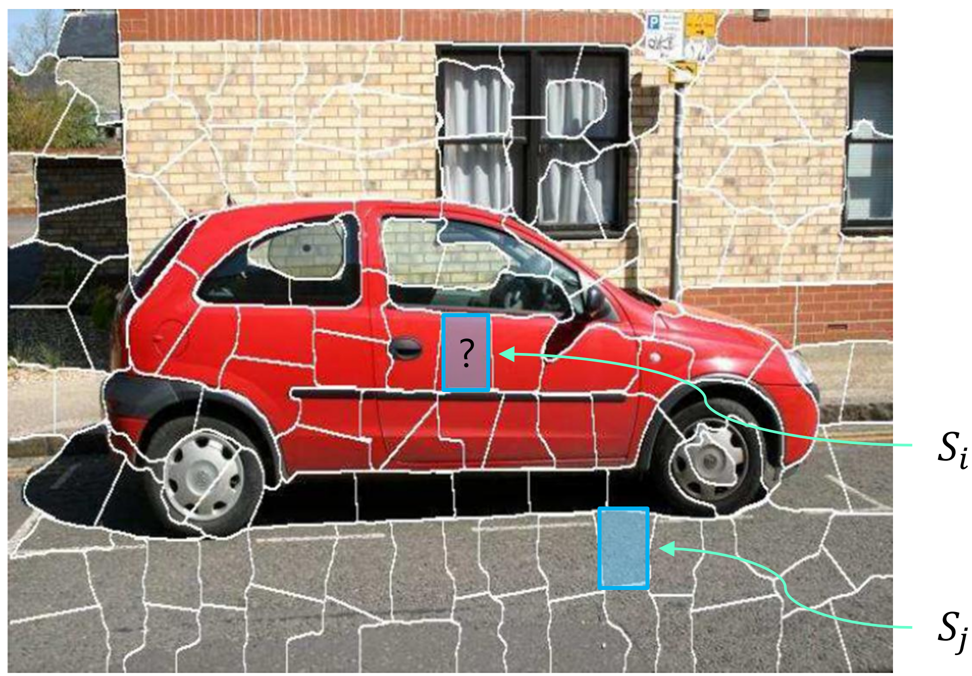


Image Credit: S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller – Multi-class Segmentation with Relative Location Prior

Relative Location Prior

- Encodes relative location between object classes
  - conditioned on offset $p(c_i = car | c_j = road, D(S_i, S_j))$
- Each superpixel votes for the most likely label for all other superpixels

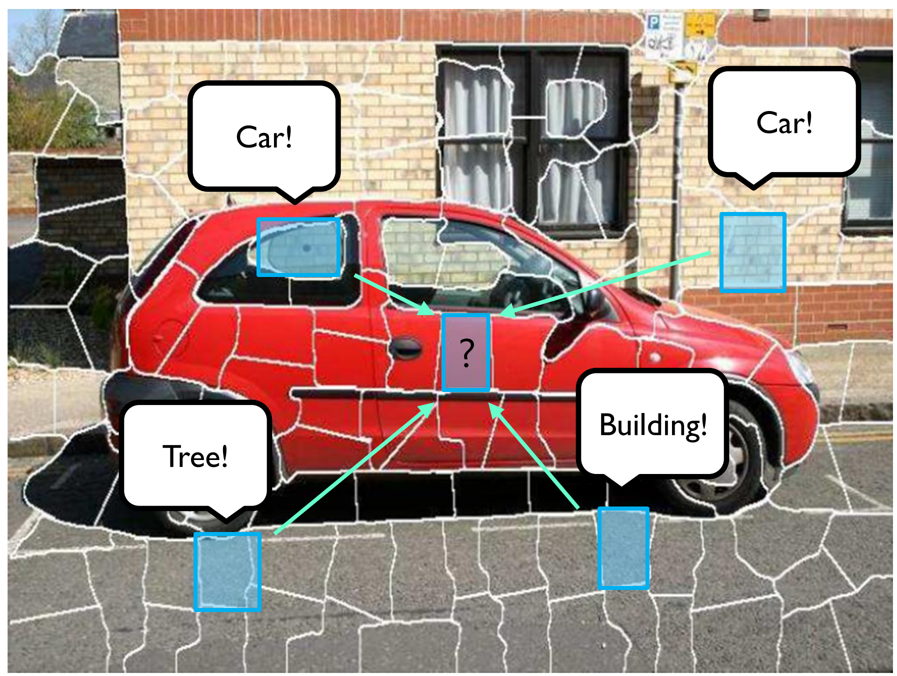


Image Credit: S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller – Multi-class Segmentation with Relative Location Prior

Relative Location Prior

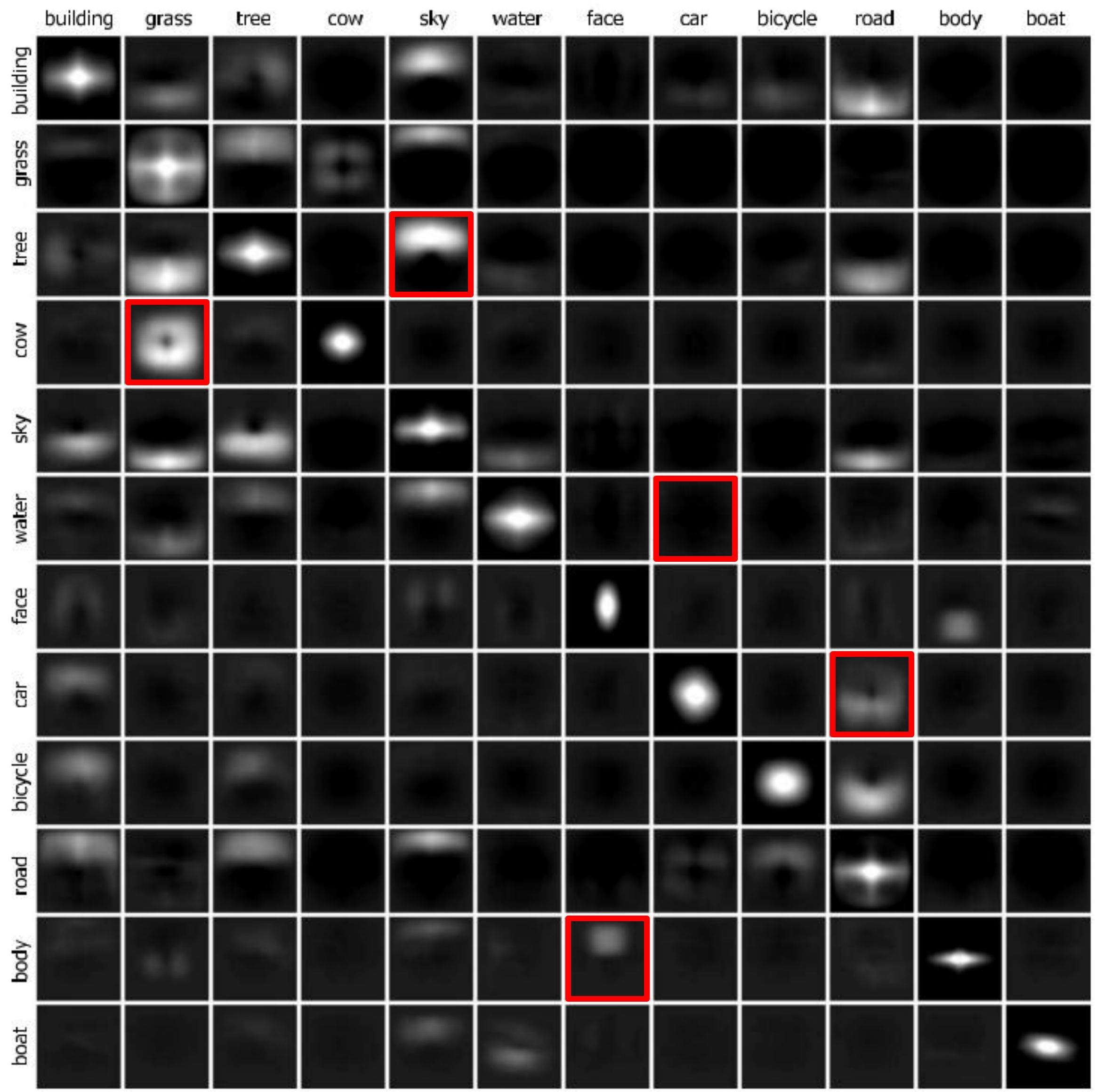
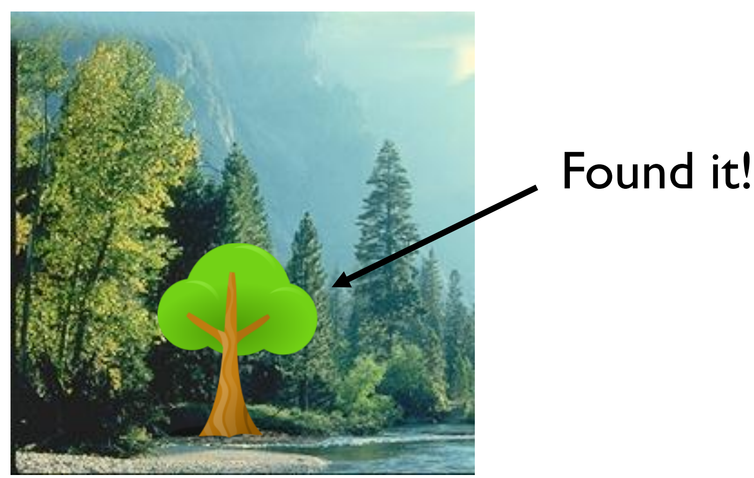


Image Credit: S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller – Multi-class Segmentation with Relative Location Prior

Connections

- Probabilistic framework
- Leveraging contextual relationships
  - location priming vs. relative location prior
- Sensitive to first stage errors
  - scene detector vs. appearance classifier



References

[1] "Using the Forest to see the Trees: Exploiting Context for Visual Object Detection and Localization" A. Torralba, K.P. Murphy, W.T. Freeman (CACM 2009)

[2] "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope" A. Oliva, A. Torralba (IJCV 2001)

[3] "Sharing Visual Features for Multiclass and Multiview Object Detection" A. Torralba, K.P. Murphy, W.T. Freeman (PAMI 2007)

[4] "Multi-class Segmentation with Relative Location Prior" S. Gould, J. Rodgers, D. Cohen, G. Elidan, D. Koller (IJCV 2008)

[5] "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data" J. Lafferty, A. McCallum, F. Pereira. (ICML 2001)