Analysis:

Objects in Context

[Rabinovich, Vedaldi, Galleguillos, Wiewiora, Belongie]

&

Object Categorization using Co-Occurrence, Location and Appearance

[Galleguillos, Rabinovich, Belongie]

Utsav Prabhu

03/18/2009

Relations between objects

(Biederman et al., 1982)

1. Interposition

   Objects interrupt their background – fire hydrant in front of a building

2. Support

   Objects tend to rest on surfaces – car on a road

3. Probability

   Objects tend to be found in some scenes but not in others – cars with buildings, trees with grass,…

4. Position

   Given an object is probable in a scene, it often is found in some positions and not others – sky towards the top, grass towards the bottom

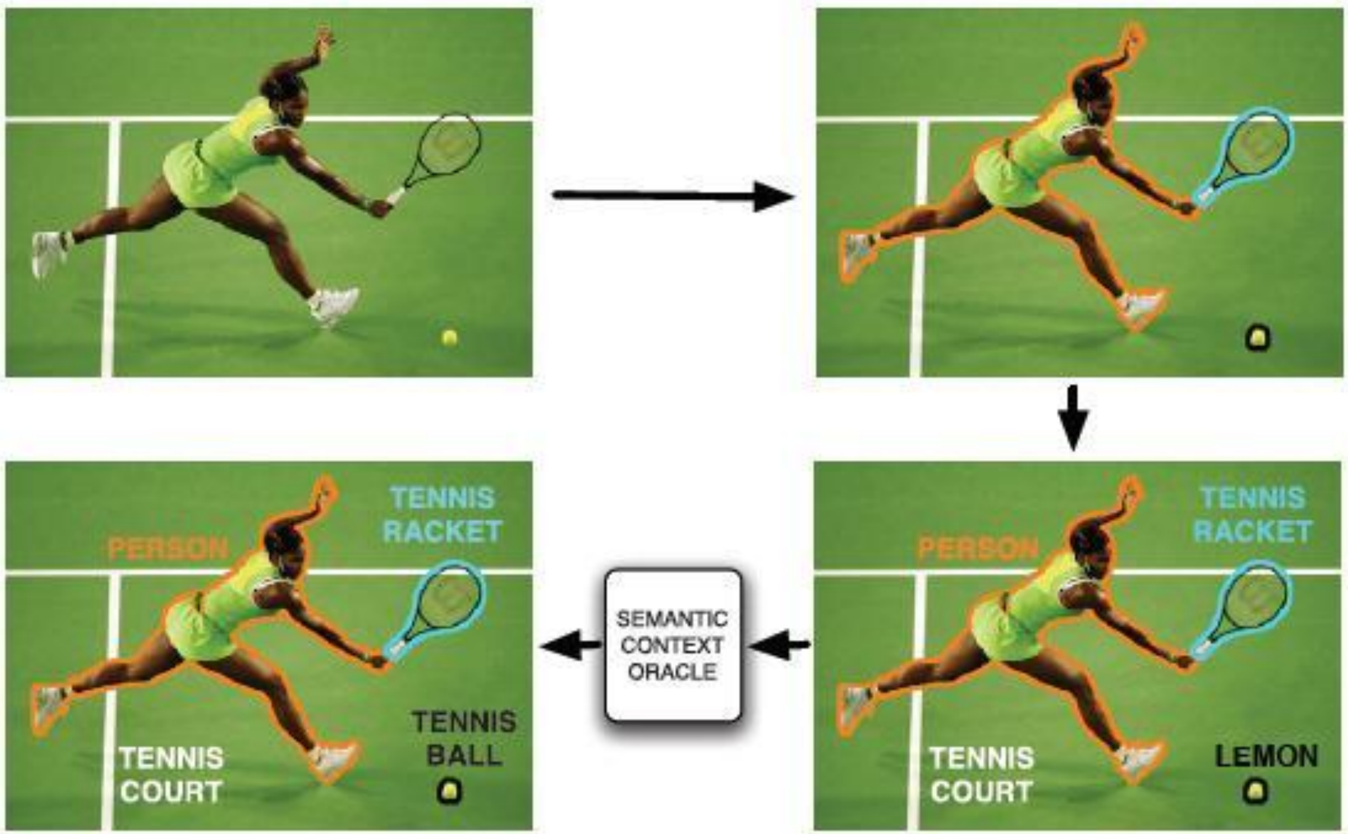
5. Familiar size

   Objects have a limited set of size relations with other objects – person larger than dog

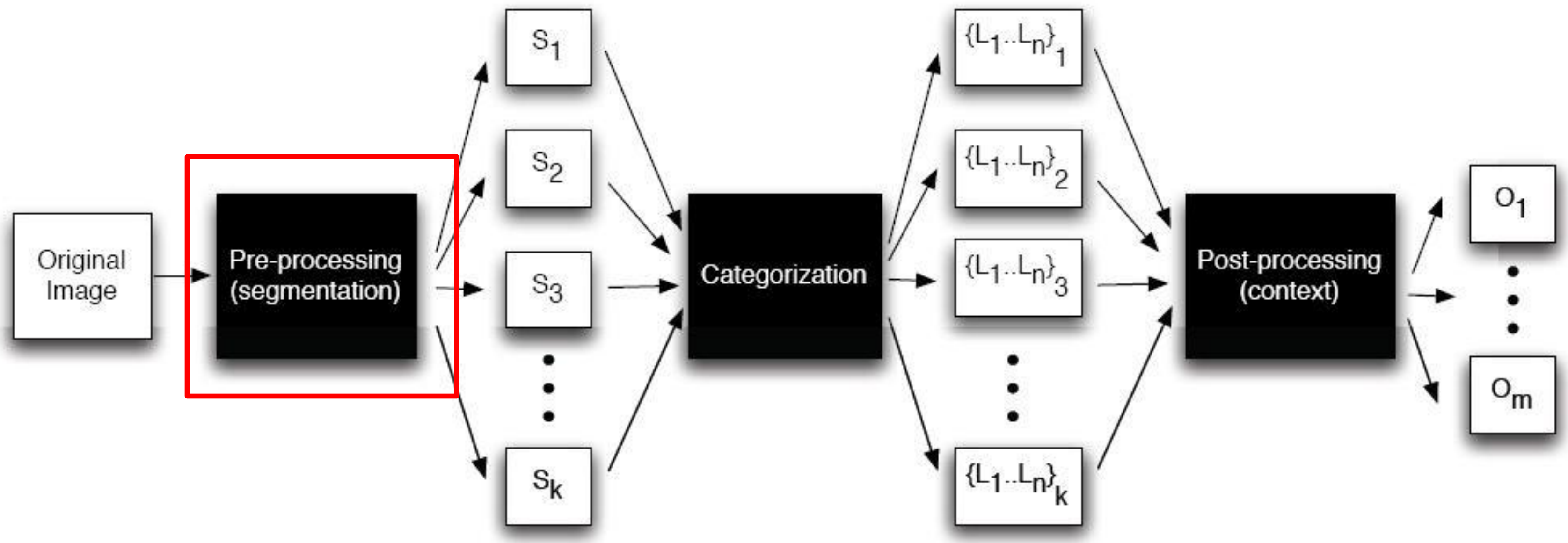
Approaches using "probability" for object recognition

- Use low-level features across the image
  - Multiscale Conditional Random Fields for Image Labeling
- Use global scene features, such as gist
  - Using the Forest to See the Trees
- Focus of attention
  - Contextual Priming for Object Detection
- Generate a context feature for *each pixel*
  - A Critical View of Context

Semantic Context



Flowchart of approach used



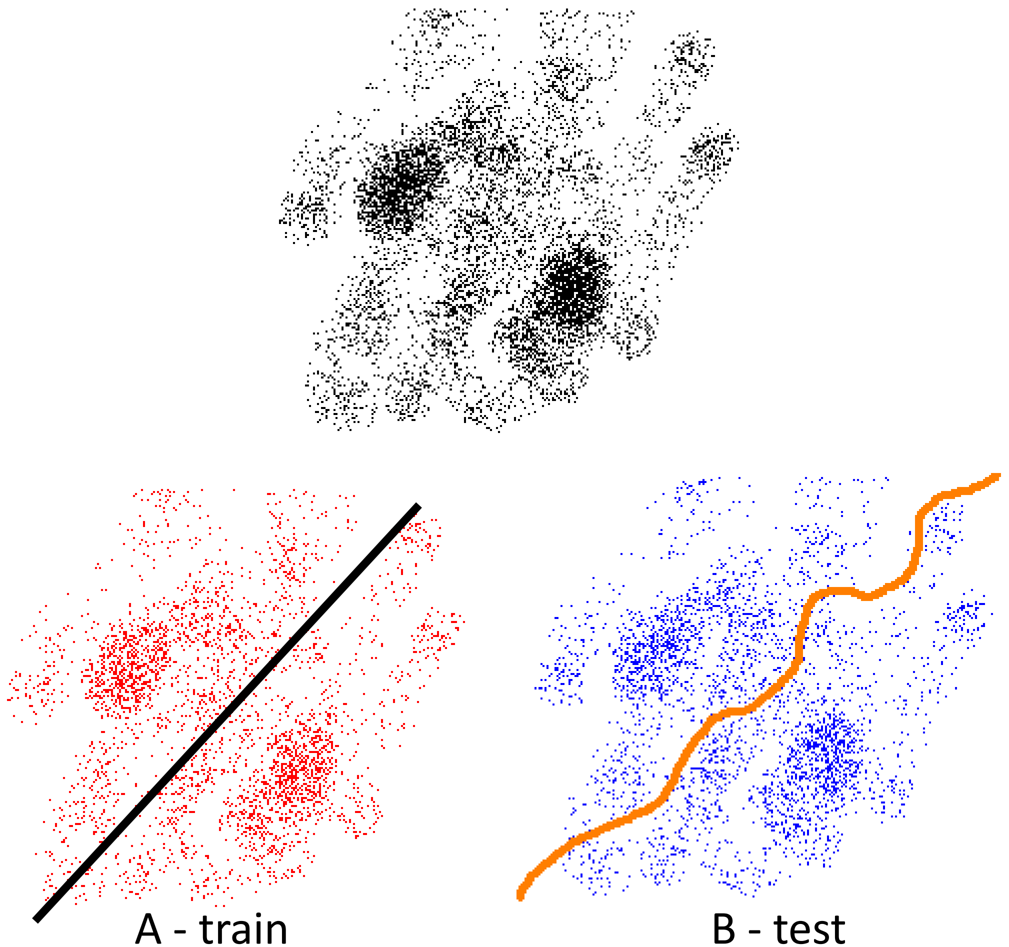
Step 1: Segmentation

- Roadblocks:
  - Number of segments
  - Cues used to segment (pixel locations, color, texture,…)
  - Combination of the above cues

- Solution: Stability based segmentation

Stability based clustering
Take 1

1. Split the dataset into 2 disjoint subsets A & B
2. Cluster A into k groups
3. Train a classifier φ using the labels from the clustering algorithm
4. Cluster B into k groups
5. Also classify data in B using the classifier φ
6. Compare the 2 results and determine a stability score
7. Repeat for a range of k



Stability based clustering
Take 2

1. Cluster the entire data into k clusters
2. Perturb the data
   - Add noise
   - Perturb the positions of each data point
3. Cluster the data again using same k
4. Repeat steps 1-3 many times
5. Permute all the labelings except one (anchor)
6. Calculate a signature based on:

No classifier!
Reduced complexity!

Indicates label agreement over all perturbations

$$S(k) = \frac{1}{n - \frac{n}{k}} \left(\sum_{i=1}^{n} s_i - \frac{n}{k} \right)$$

Normalization coefficient

Prevents bias for different values of k

7. Try all possible anchors & choose the one with highest stability

Stability based segmentation

- Cues used: Color, Texture
- 9 different cue weightings used
- Noise added 20 times
- Segmentations for k=2 through k=9

Standard N-cut segmentation

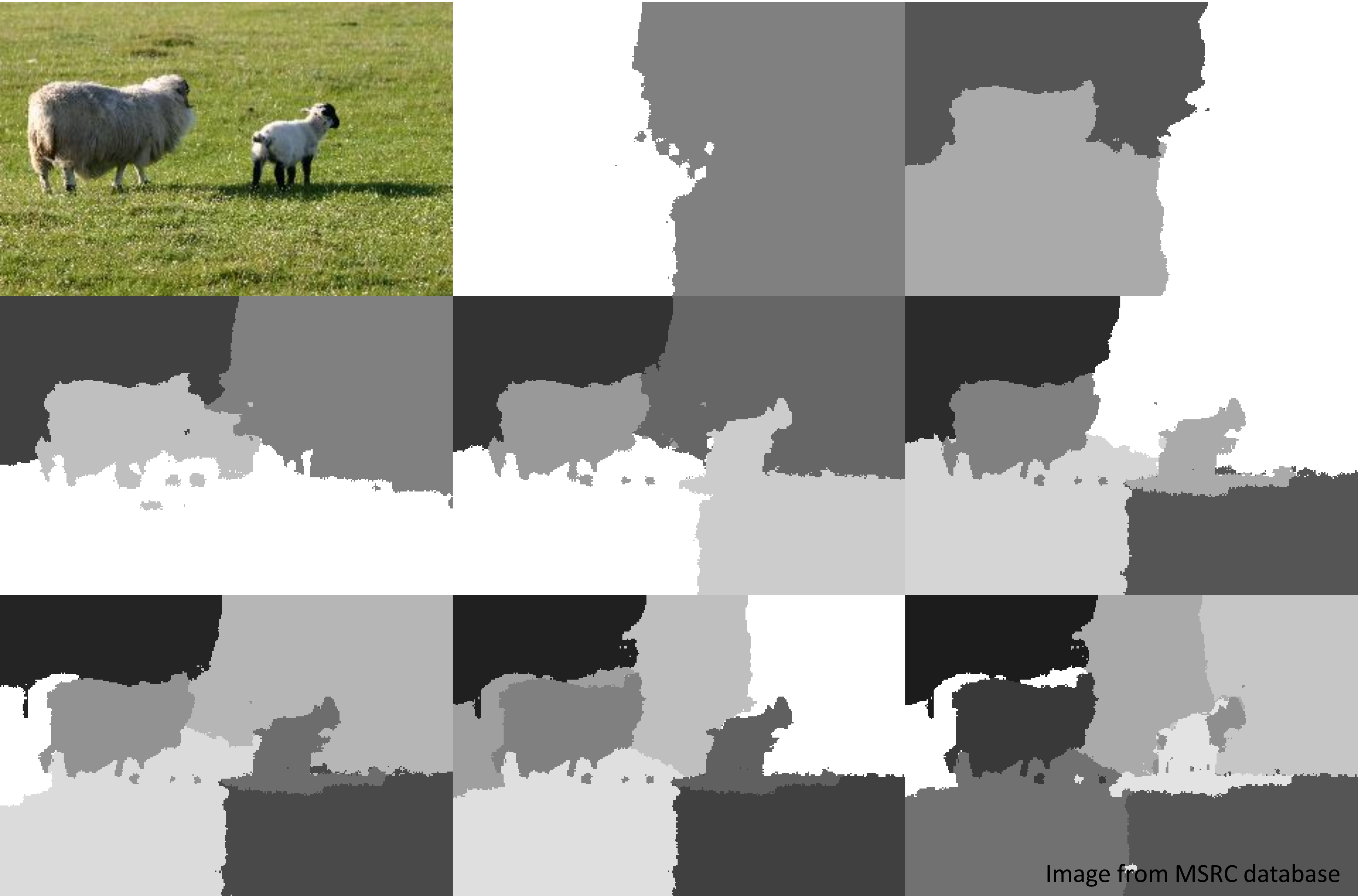


Image from MSRC database

Stability based segmentation

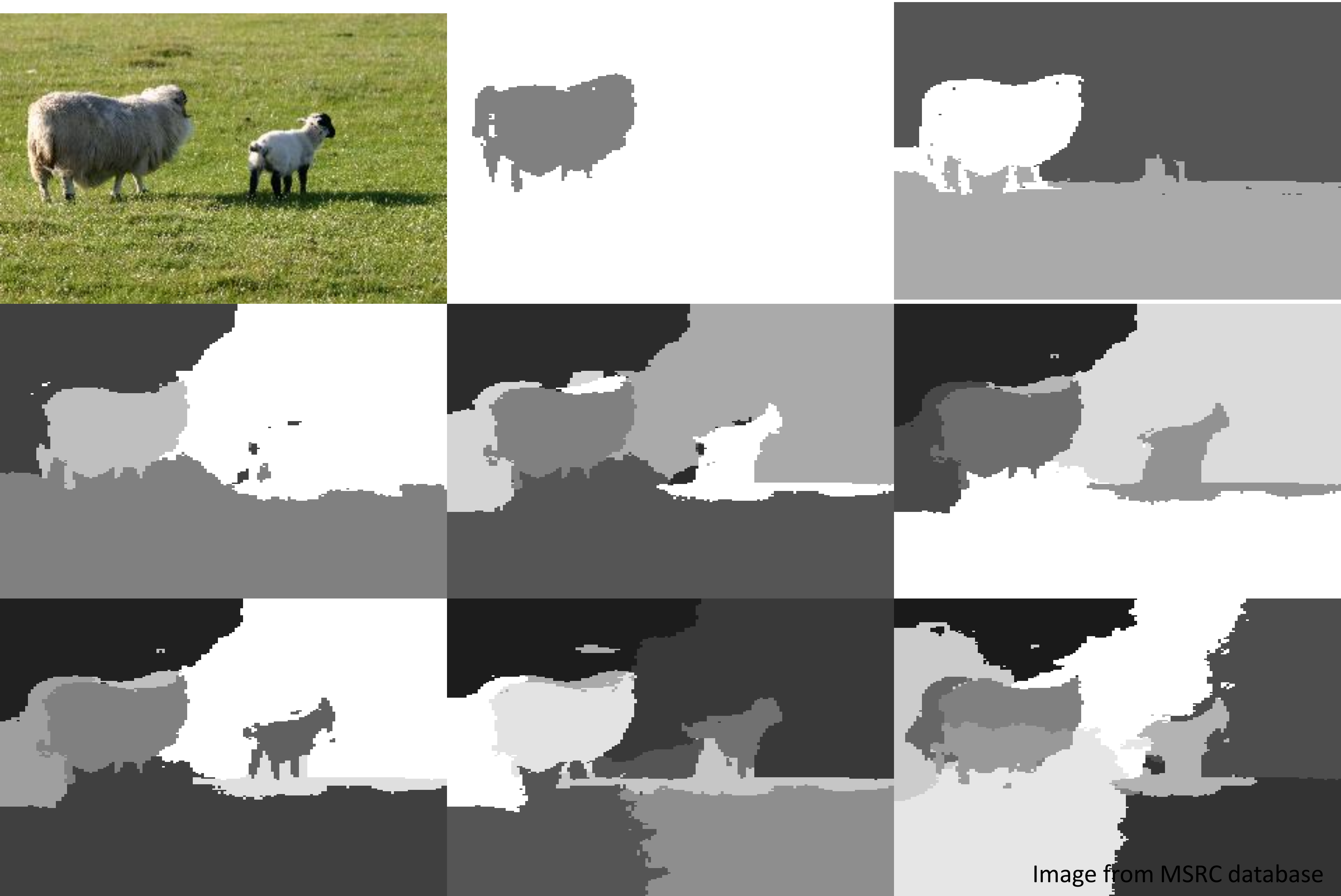
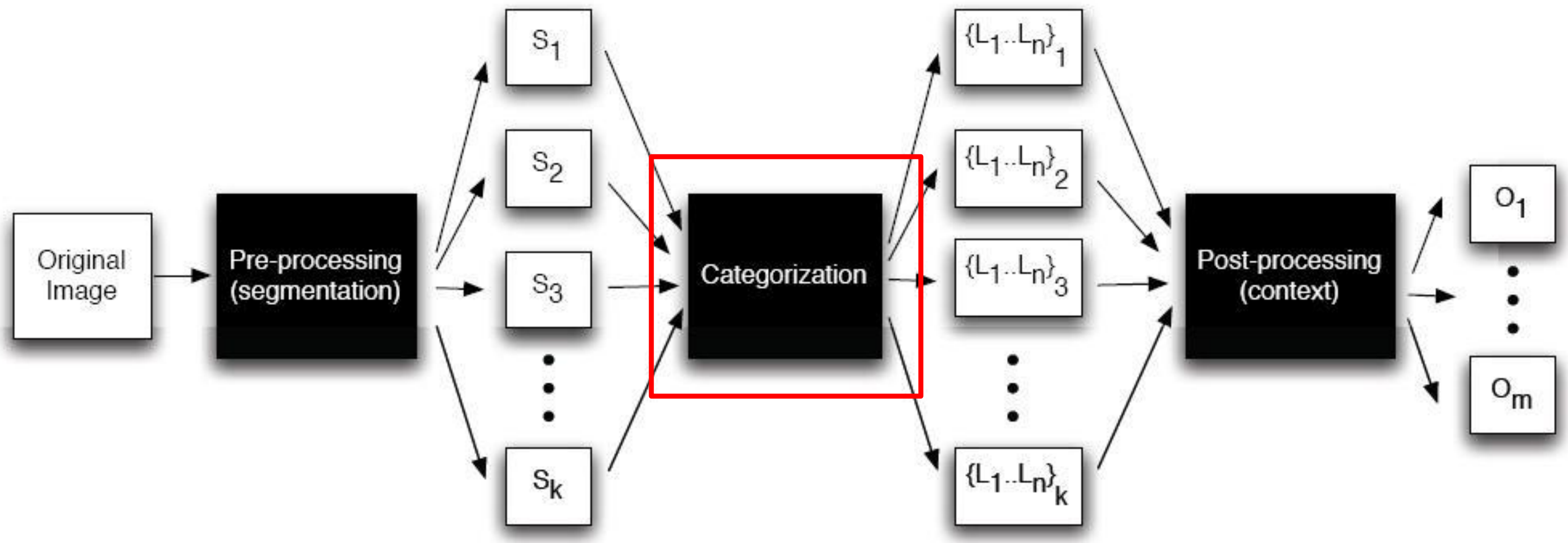


Image from MSRC database

Stability based segmentation - results

Images from MSRC database

Flowchart



Original Image
Pre-processing (segmentation)
S_1
S_2
S_3
S_k
Categorization
$\{L_1..L_n\}_1$
$\{L_1..L_n\}_2$
$\{L_1..L_n\}_3$
$\{L_1..L_n\}_k$
Post-processing (context)
O_1
O_m

Bag of Features

1. Decompose the image into a collection of *features*
2. Map the features to a finite vocabulary of *visual words*
3. Compute a *signature* of these visual words
4. Feed the signatures into a classifier for labeling

features = SIFT, *visual words* = k-means, *signature* = histogram

Integrating BoF & stable segmentation

- Each segment (of the 54) is masked & zero-padded
- Compute the signature of each segment
  - Discard features which fall outside segment boundary
- Represent the image by ensemble of segment signatures

- Reasons for doing this:
  - Clustering features in segments incorporates coarse spatial information
  - Masking makes features more shape-informative
  - Improves SNR

Image from MSRC database

Labeling segments

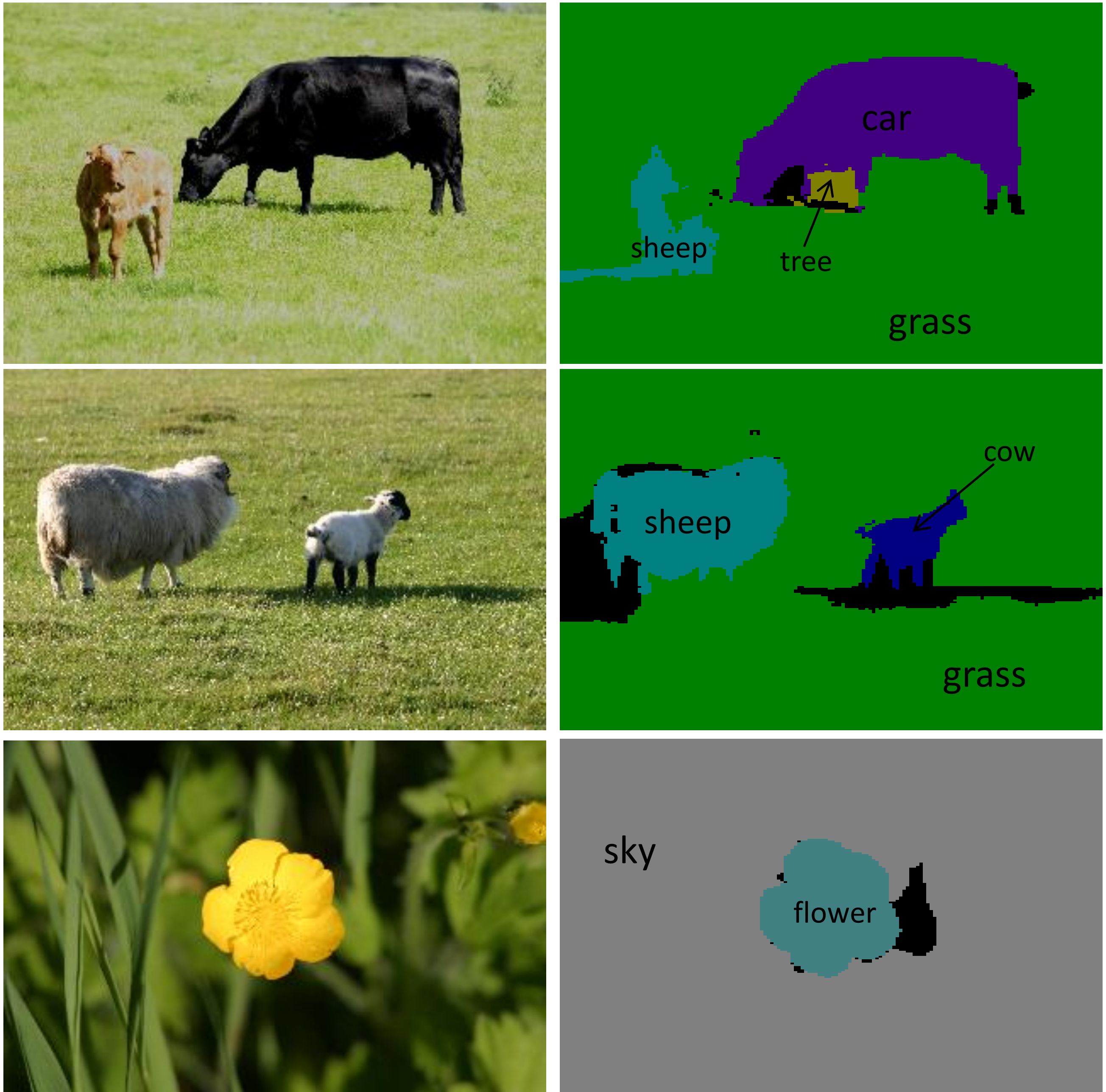
- Calculate signatures of ground truth segments of training images – $\Phi(\mathrm{I})$
- Calculate signatures of stable segments of test images - $\Phi(\mathrm{S})$
- Calculate L1 distance measure to each category:

$$d(S_q, c) = \min_i d(S_q, I_{ic}) = \min_i ||\phi(S_q) - \phi(I_{ic})||$$

- Construct a probability distribution over categories

$$p(c_i|S_q) = \left[1 - \frac{d(S_q, c_i)}{\sum_{j=1}^{n} d(S_q, c_j)}\right]$$

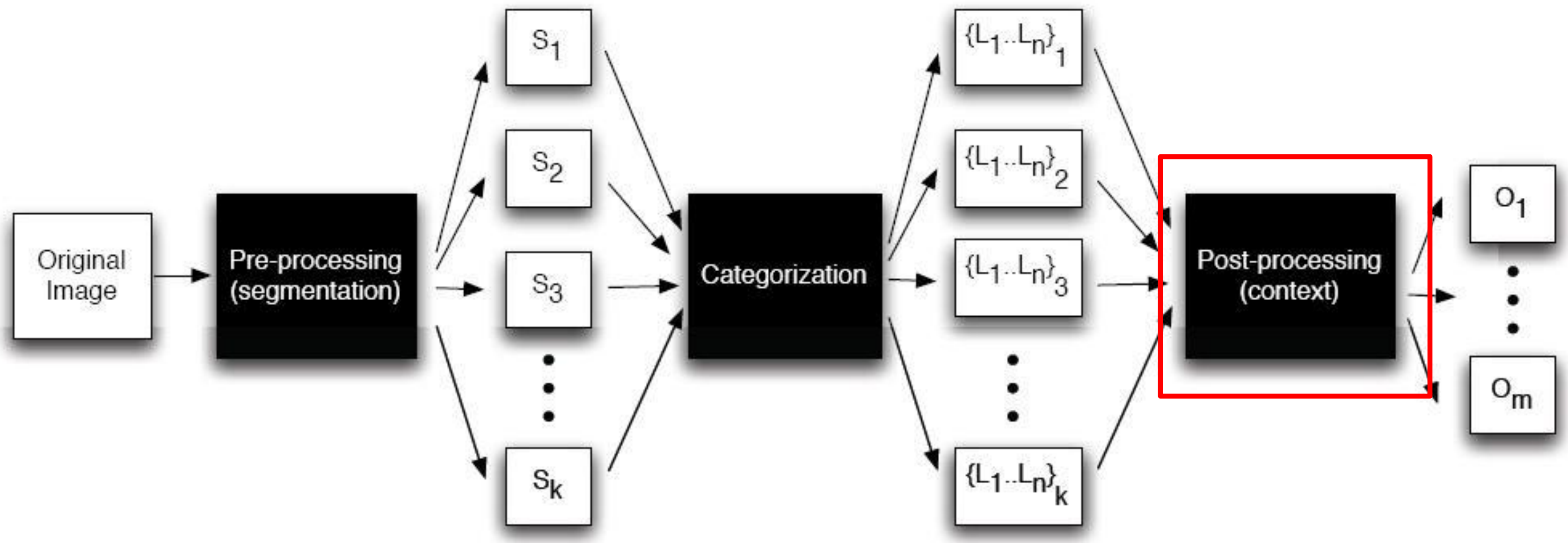
Categorization – Results



object class	R	G	B	Colour
void	0	0	0	
building	128	0	0	
grass	0	128	0	
tree	128	128	0	
cow	0	0	128	
horse	128	0	128	
sheep	0	128	128	
sky	128	128	128	
mountain	64	0	0	
aeroplane	192	0	0	
water	64	128	0	
face	192	128	0	
car	64	0	128	
bicycle	192	0	128	
flower	64	128	128	
sign	192	128	128	
bird	0	64	0	
book	128	64	0	
chair	0	192	0	
road	128	64	128	
cat	0	192	128	
dog	128	192	128	
body	64	64	0	
boat	192	64	0	

Images from MSRC database

Flowchart



Original Image
Pre-processing (segmentation)
S_1
S_2
S_3
S_k
Categorization
$\{L_1..L_n\}_1$
$\{L_1..L_n\}_2$
$\{L_1..L_n\}_3$
$\{L_1..L_n\}_k$
Post-processing (context)
O_1
O_m

Incorporating semantic context

- What we have:
  - Image I with segments $\{S_1, S_2, \ldots S_k\}$
  - Marginal probabilities $p(c_i | S_j)$
- What we want:
  - Segment labels $\{c_1, c_2, \ldots c_k\}$ for segments $\{S_1, S_2, \ldots S_k\}$ which are *in semantic contextual agreement* with each other

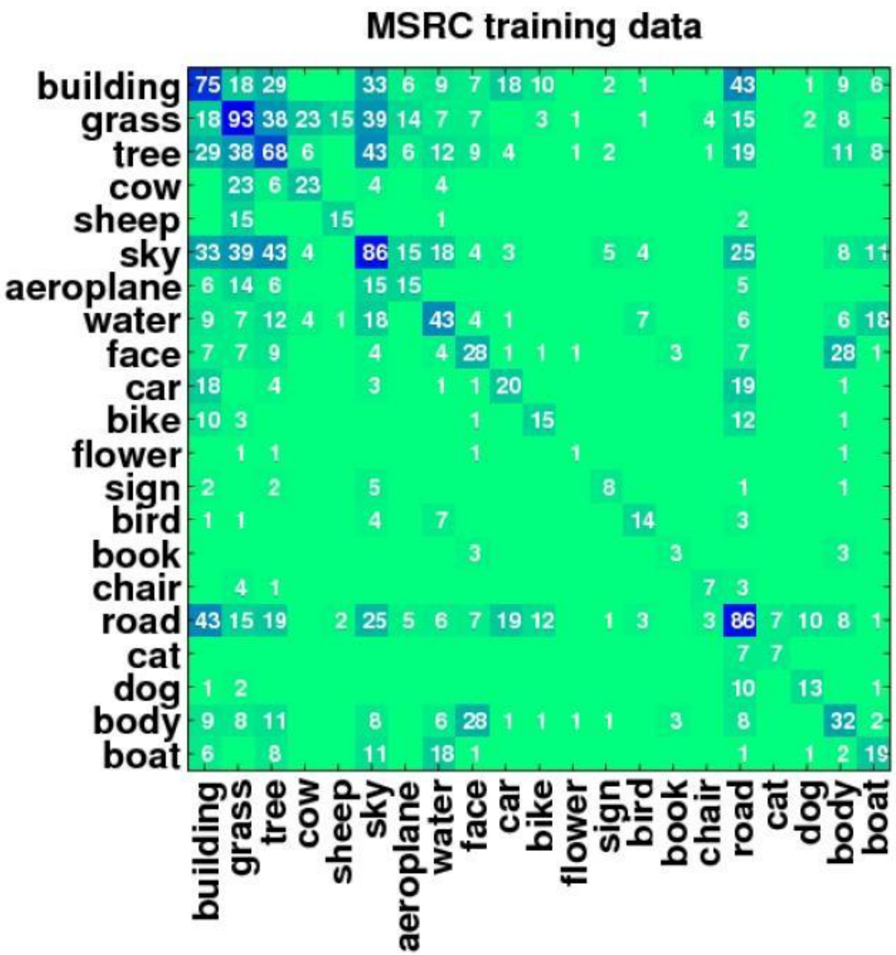
CRF framework

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

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

- Separate marginal terms from pair-wise interaction potentials $\Phi(c_i,c_j)$
- Where do we get $\Phi(c_i,c_j)$ from?
  - Co-occurrence matrix from training dataset
  - Google Sets

Co-occurrence matrix

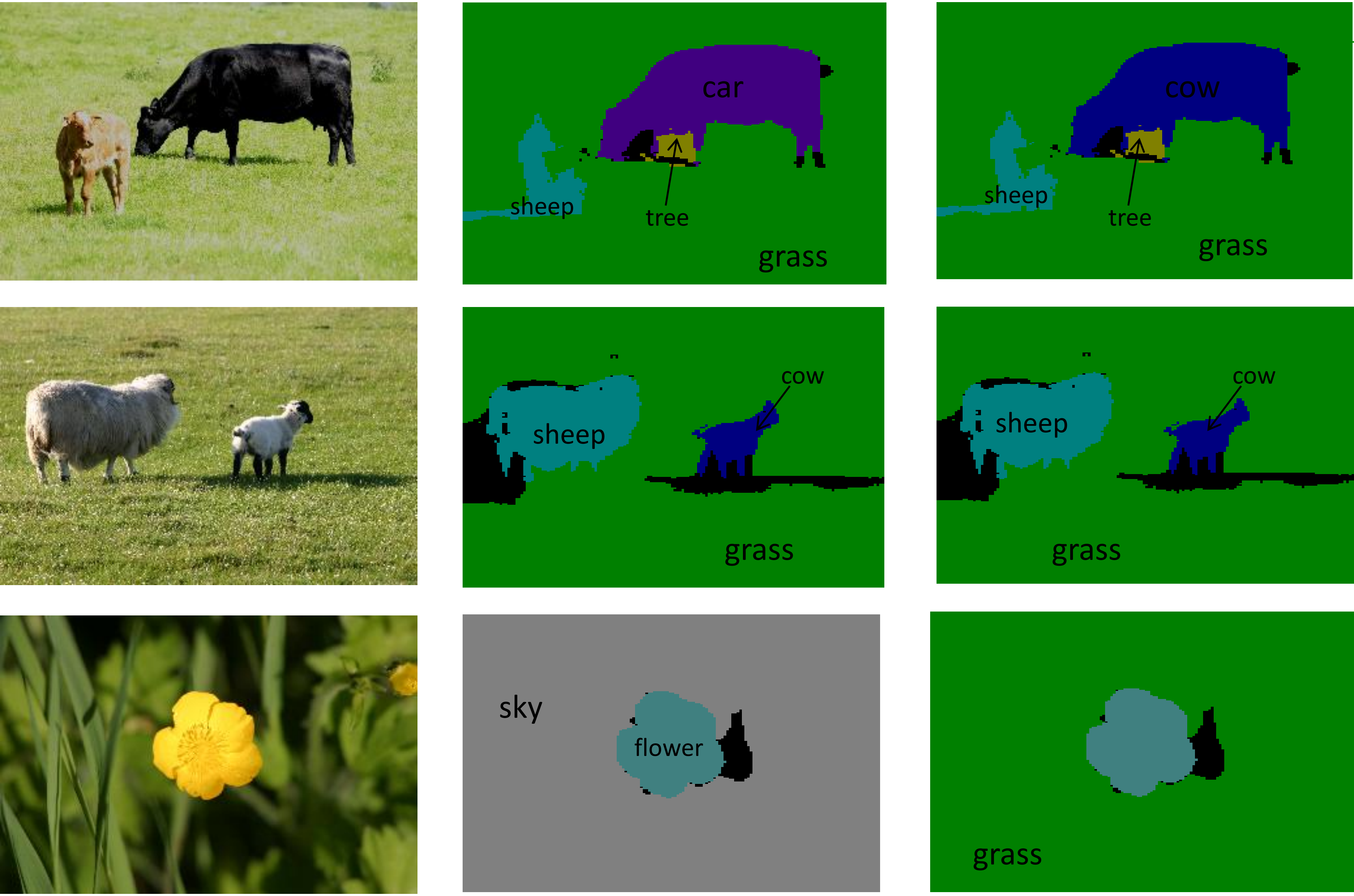


Diagonal entries = frequency of object in training set

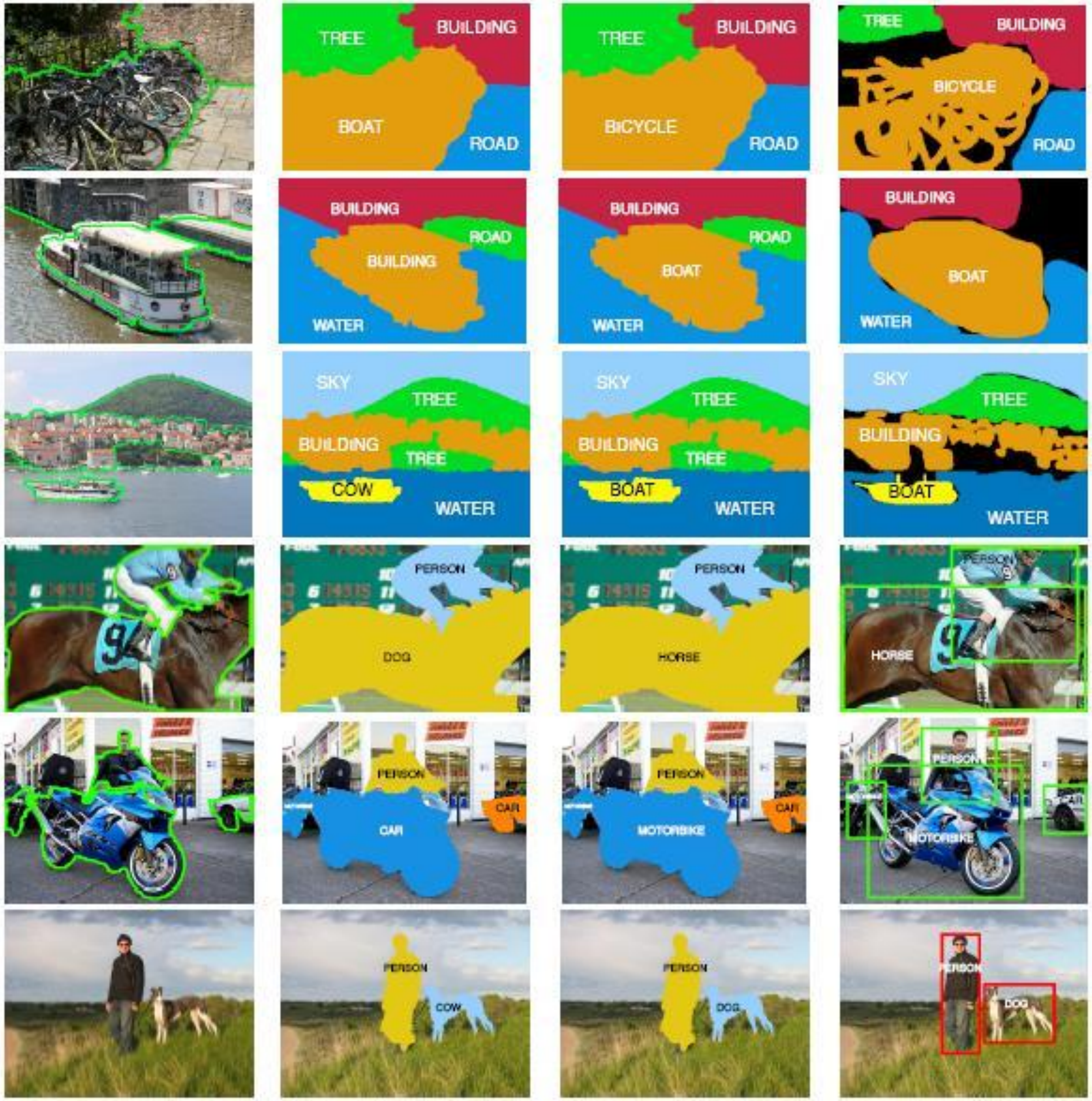
Off-diagonal entries = label co-occurrence counts

$\Phi(c_i, c_j)$ is learned from this data using MLE, gradient descent, importance sampling, monte carlo integration, …

Can we use values from the co-occurrence matrix directly?



Images from MSRC database

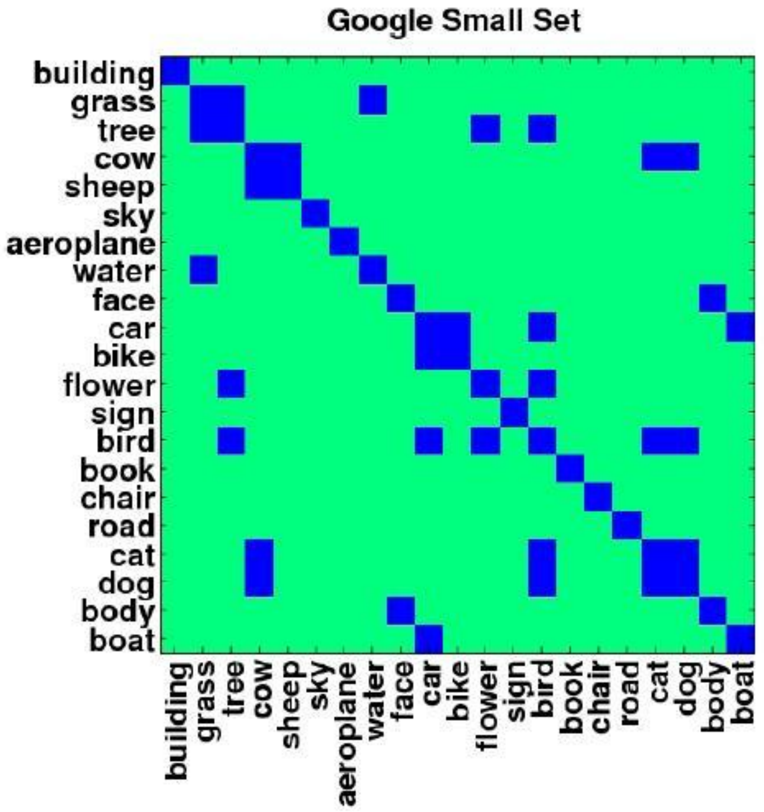


(a) (b) (c) (d)

Figure 6. Examples of MSRC (first 3) and PASCAL (last 3) test images, where contextual constraints have improved the categorization accuracy. Results are shown in two different ways, one for each dataset. In MSRC, full segmentations of highest average categorization accuracy are shown; in PASCAL individual segments of highest categorization accuracy are shown. (a) Original Segmented Image. (b) Categorization without contextual constraints. (c) Categorization with co-occurence contextual constraints derived from the training data. (d) Ground Truth.

Google sets

- Automatically create sets of *possibly related* items from a few examples
- Based on search statistics, trends, web page content, dictionary / thesaurus, wikipedia, …



Can Google Sets provide a true semantic context based grouping criterion?

Google sets – sanity check

- Query #1: “dog”
  - Results: “dog” “cat” “trackbacks 0” “ティムティム” “canine” “canid” “bird” “pets” “dogs” “horse” “edit” “comments 0” “puppy”
  - Categories found in the results: “cat”, “bird”

Q: How often do dogs and {cats, birds} appear in the same image?

Lets look at the largest annotated database we have: LabelMe.

- Number of images containing dogs = 223
- Number of images containing dogs and cats = 0
- Number of images containing dogs and birds = 0

Google sets – sanity check

- Query #2: “cow”
  - Results: “cow” “pig” “horse” “dog” “cat” “bear” “sheep” “duck” “rabbit” “chicken” “goat” “cash” “animal” “calf”
  - Categories found in the results: “dog” “cat” “sheep” “bird”

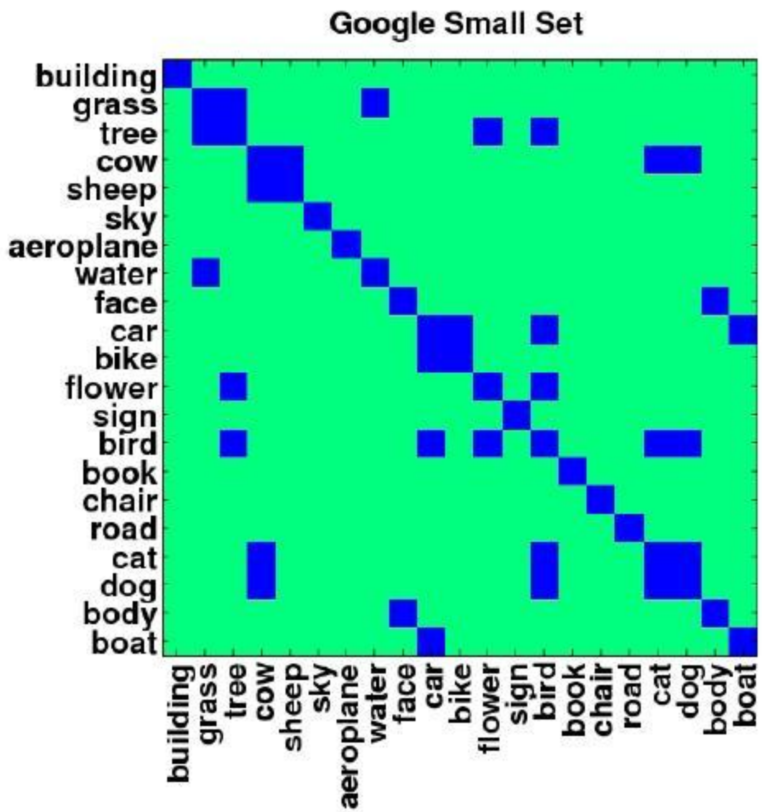
  - Number of images containing cows = 33
  - Number of images containing cows and dogs = 0
  - Number of images containing cows and cats = 0
  - Number of images containing cows and sheep = 0
  - Number of images containing cows and birds = 0

Google sets – sanity check

- Query #3: “car”
  - Results: “car” “truck” “auto” “train” “parking” “cars” “boat” “suv” “bus” “motorcycle” “hotel”
  - Categories found in the results: “boat” “bike”

  - Number of images containing cars = 6600
  - Number of images containing cars and boats = 0
  - Number of images containing cars and bikes = 1

Live Demo - Flickr



Conclusion: Google sets is not really a good source for a semantic context based grouping criterion

Experimental Results

- MSRC & PASCAL datasets

	No Context	Google Sets	Using Training
MSRC	45.0%	58.1%	68.4%
PASCAL	61.8%	63.4%	74.2%

Table 1. Average Categorization Accuracy.

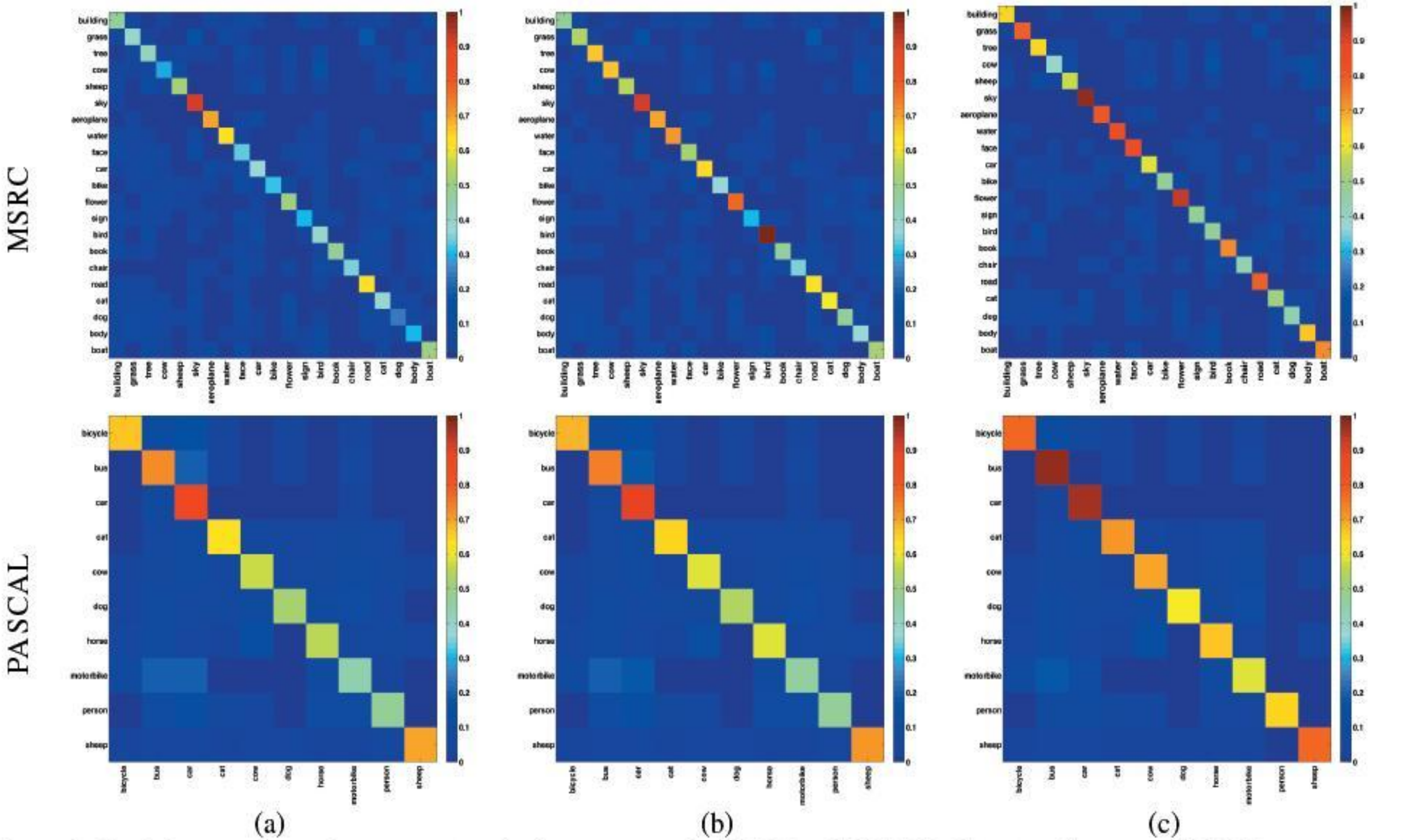
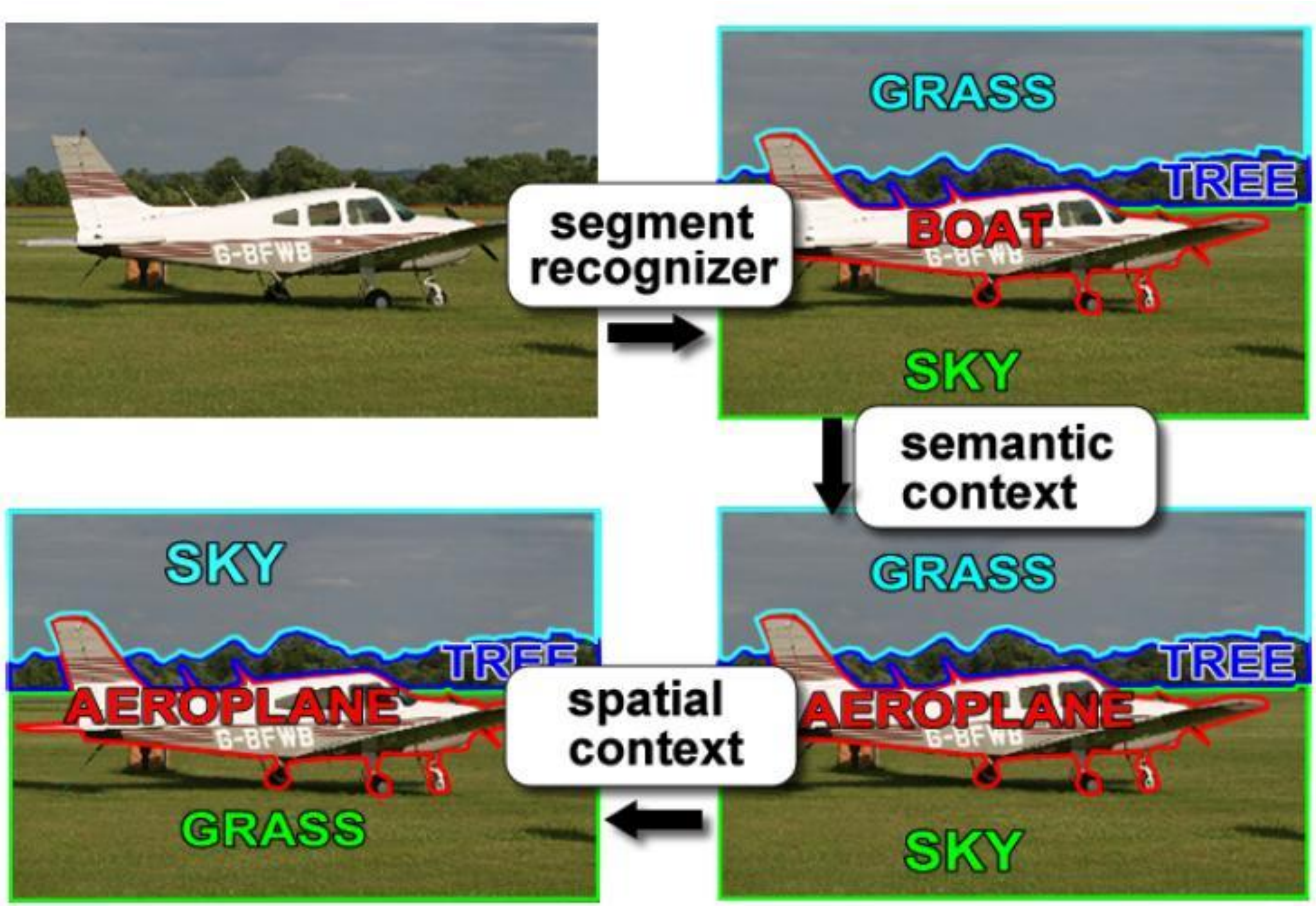


Figure 5. Confusion matrices of average categorization accuracy for MSRC and PASCAL datasets. First row: MSRC dataset; second row: PASCAL dataset. (a) Categorization with no contextual constraints. (b) Categorization with Google Sets context constraints. (c) Categorization with Ground Truth context constraints.

Discussion

- Does co-occurrence truly represent the semantic context of an object?
- Does masking and zero-padding each segment incorporate any kind of shape-information about the segment?
- Should context have the last say in a feed-forward model?

Incorporating spatial context



Spatial context descriptor

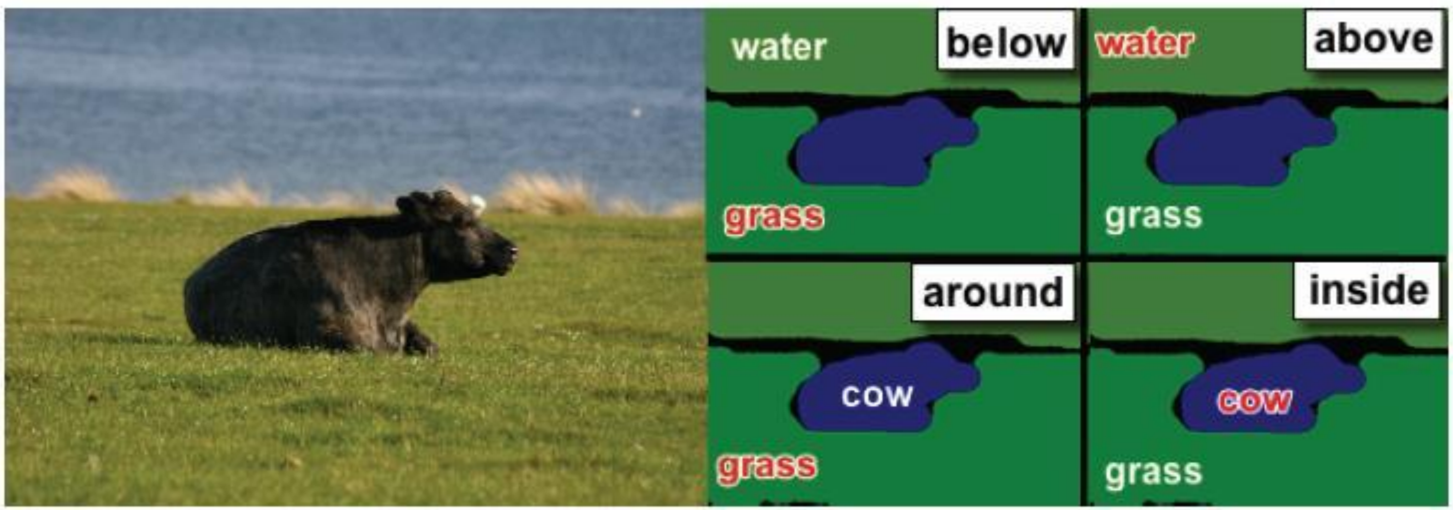
- Pair-wise feature
- 3-dimensional descriptor:

$$Fij = \left(\mu_{ij}, O_{ij}, O_{ji}\right)^T \; \forall \, i, j \in \mathbb{C}, i \neq j$$

$$O_{ij} = \frac{\beta_i \cap \beta_j}{\beta_i} \qquad \mu_{ij} = \mu_{yi} - \mu_{yj}$$

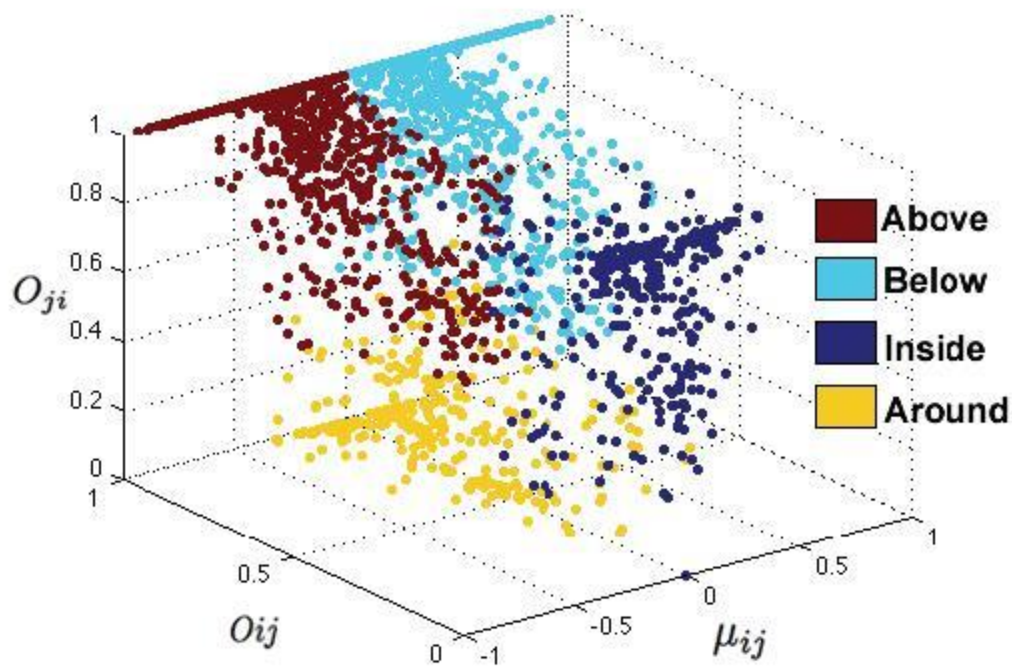
  - μ_{ij} is the difference in y components of centroids of the 2 objects
  - β_i is the bounding box / pixel mask of object i

Spatial context feature - example

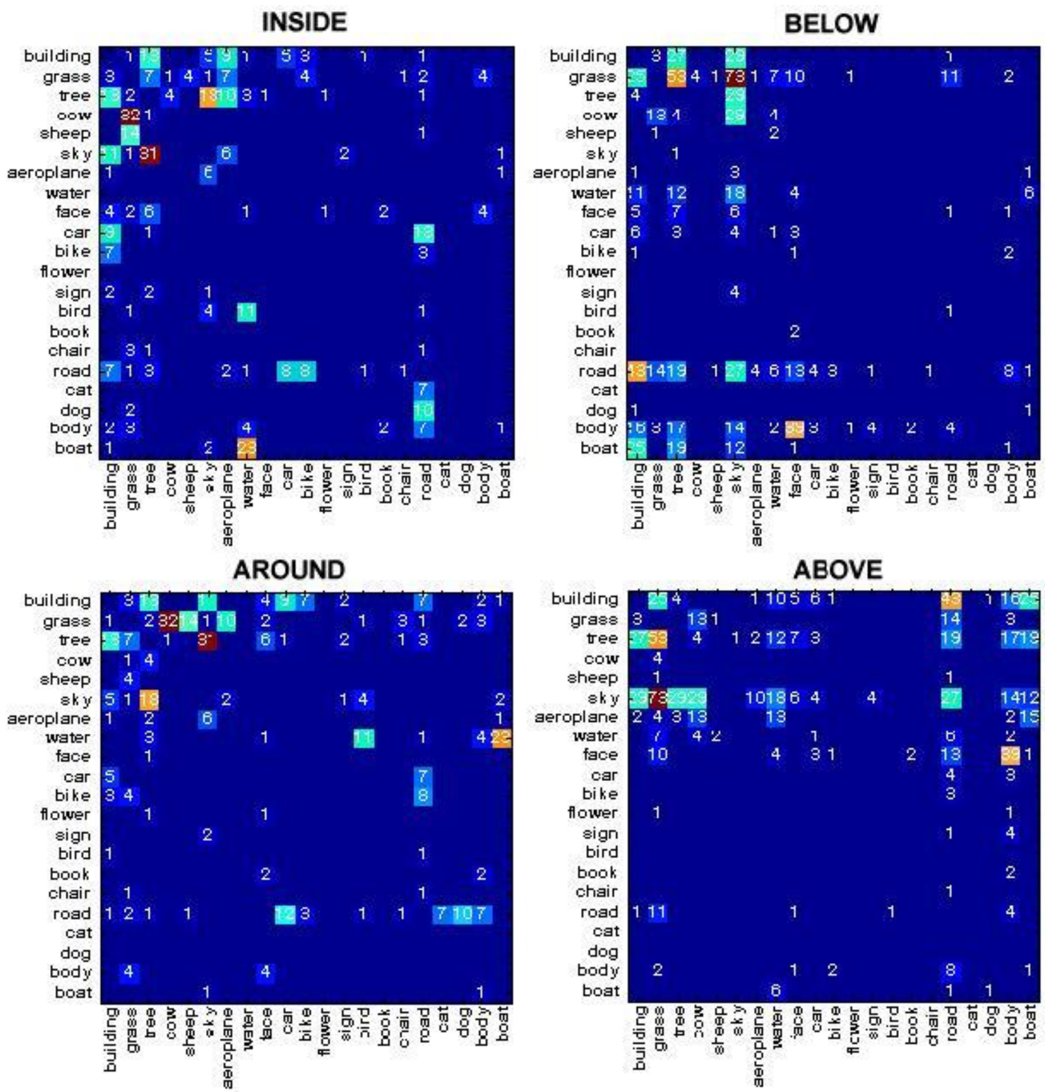


Spatial context feature

- Vector quantize this descriptor into four groups: *above, below, inside, around*



Locations and co-occurrences



Updated CRF model

$$p(c_1 \dots c_k | S_1 \dots S_k) = \frac{B(c_1 \dots c_k) \prod_{i=1}^{k} p(c_i | S_i)}{Z(\phi_0, \dots \phi_r, S_1 \dots S_k)}$$

$$\text{with } B(c_1 \dots c_k) = \exp\Big(\sum_{i,j=1}^{k} \sum_{r=0}^{q} \alpha_r \phi_r(c_i, c_j) \Big)$$

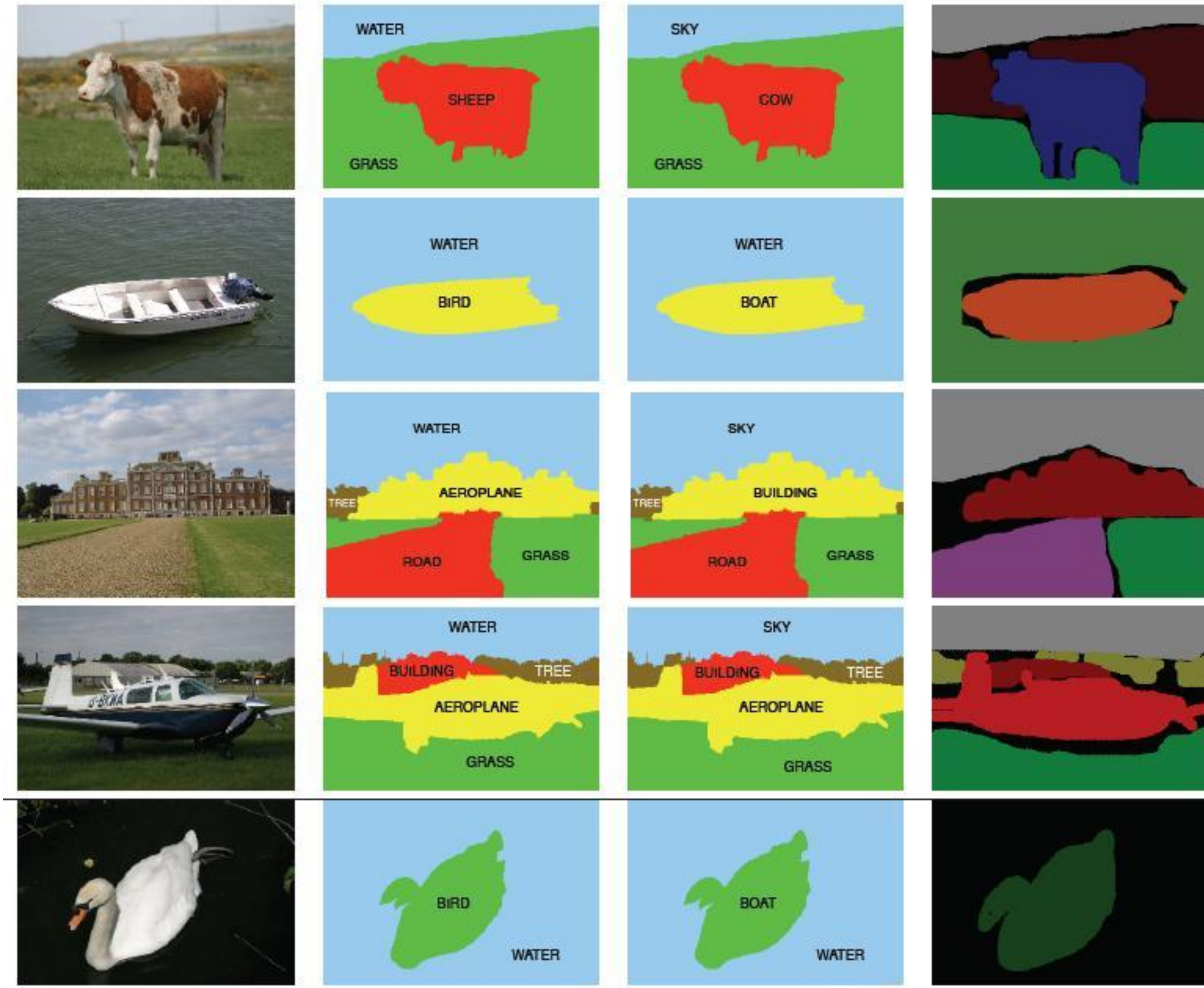
Experimental Results

Categories	Semantic Context [18]	CoLA
building	0.85	**0.91**
grass	0.94	**0.95**
tree	0.78	**0.80**
cow	0.36	**0.41**
sheep	0.55	0.55
sky	0.89	**0.97**
aeroplane	0.73	0.73
water	0.95	0.95
face	0.80	**0.81**
car	0.57	0.57
bike	0.59	**0.60**
flower	0.65	0.65
sign	0.54	0.54
bird	0.54	*0.52*
book	0.56	0.56
chair	0.42	0.42
road	0.94	**0.96**
cat	0.42	0.42
dog	0.46	0.46
body	0.75	**0.77**
boat	0.76	**0.81**

Categories	Semantic Context [18]	CoLA
aeroplane	0.63	0.63
bicycle	0.22	0.22
bird	0.18	*0.14*
boat	0.28	**0.42**
bottle	0.43	0.43
bus	0.46	**0.50**
car	0.62	0.62
cat	0.32	0.32
chair	0.37	0.37
cow	0.19	0.19
dining table	0.30	0.30
dog	0.32	*0.29*
horse	0.12	**0.15**
motorbike	0.31	0.31
person	0.43	0.43
potted plant	0.33	0.33
sheep	0.41	0.41
sofa	0.37	0.37
train	0.29	0.29
tv monitor	0.62	0.62

Table 1. Comparison of recognition accuracy between the models for MSRC and PASCAL categories. Results in **bold** indicate an increase in performance by our model. A decrease in performance is shown in *italics*.

Original image | Categorization + co-occurrence | + spatial context | Ground truth



Images from MSRC and PASCAL databases

Thank You