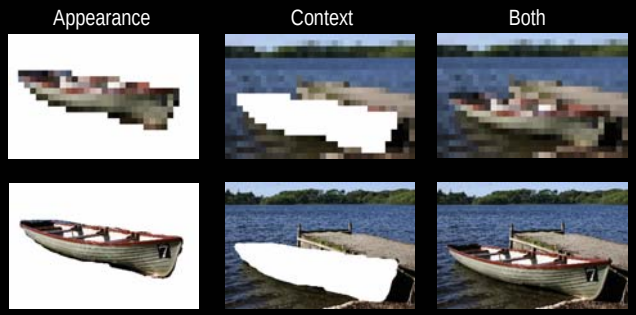
From Appearance to Context Based Recognition: Dense Labeling in Small Images

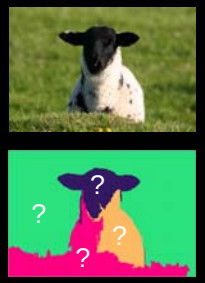
Devi Parikh[1], Larry Zitnick[2] and Tsuhan Chen[1]

[1] Carnegie Mellon University (Pittsburgh, PA) [2] Microsoft Research (Redmond, WA)

Motivation



Scenarios Studied

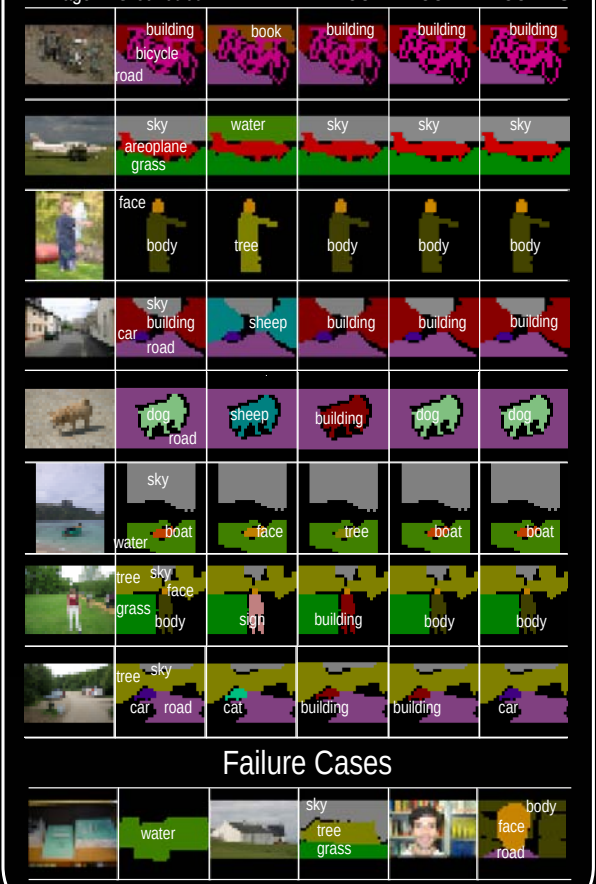


Low vs. High Resolution

Appearance vs. Context

Human vs. Machine

Image Labeling Results



Failure Cases

Approach

Segmentation:

[Felzenszwalb IJCV 2004]

Inference: Belief Propagation on fully connected CRF

$$P(\mathbf{c}\mid\mathbf{S}) = \frac{1}{Z}\prod_{i=1}^{N}\Psi_i(c_i)\prod_{i,j=1}^{N}\Phi_{ij}(c_i,c_j)$$

Appearance $\Psi_i(c_i)$

Texture + Shape [TextonBoost] [3]

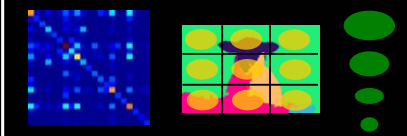
Color [GMM]

Neural Network [4]

[3] [Shotton ECCV 2006] [4] [He CVPR 2004]

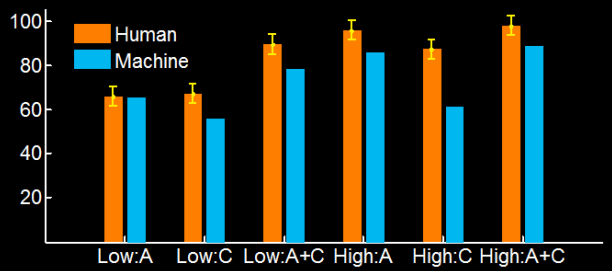
Context $\Phi_{ij}(c_i,c_j)$

Co-occurrence [5] Location Scale



[5] [Rabinovich ICCV 2007]

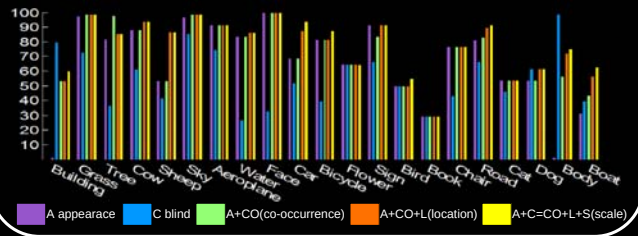
Quantitative Results



	MSRC (%)	Corel (%)
Existing (high)	75 [6]	81 [7]
Proposed (high)	91	93
Proposed (low)	83	86

[6] [Yang CVPR 2007] [7] [He ECCV 2006]

Class-wise Results



Conclusions

- Context is most useful when appearance information is weak.
- Location and scale information are useful sources of context
- Low resolution images provide an appropriate venue for studying context