

BEYOND THE LINE OF SIGHT: LABELING THE UNDERLYING SURFACES

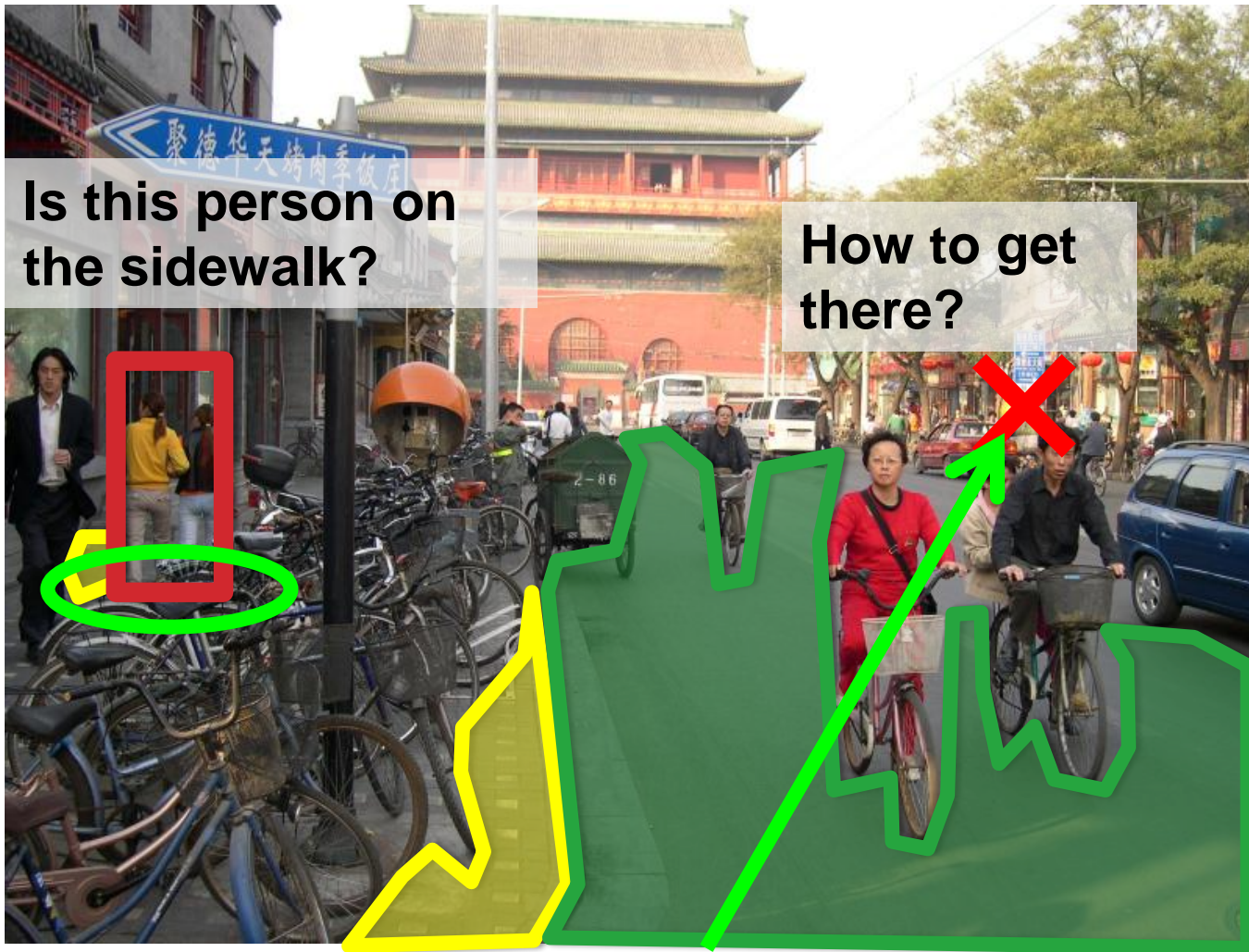
RUIQI GUO AND DEREK HOIEM

**UNIVERSITY OF ILLINOIS AT URBANA
CHAMPAIGN**

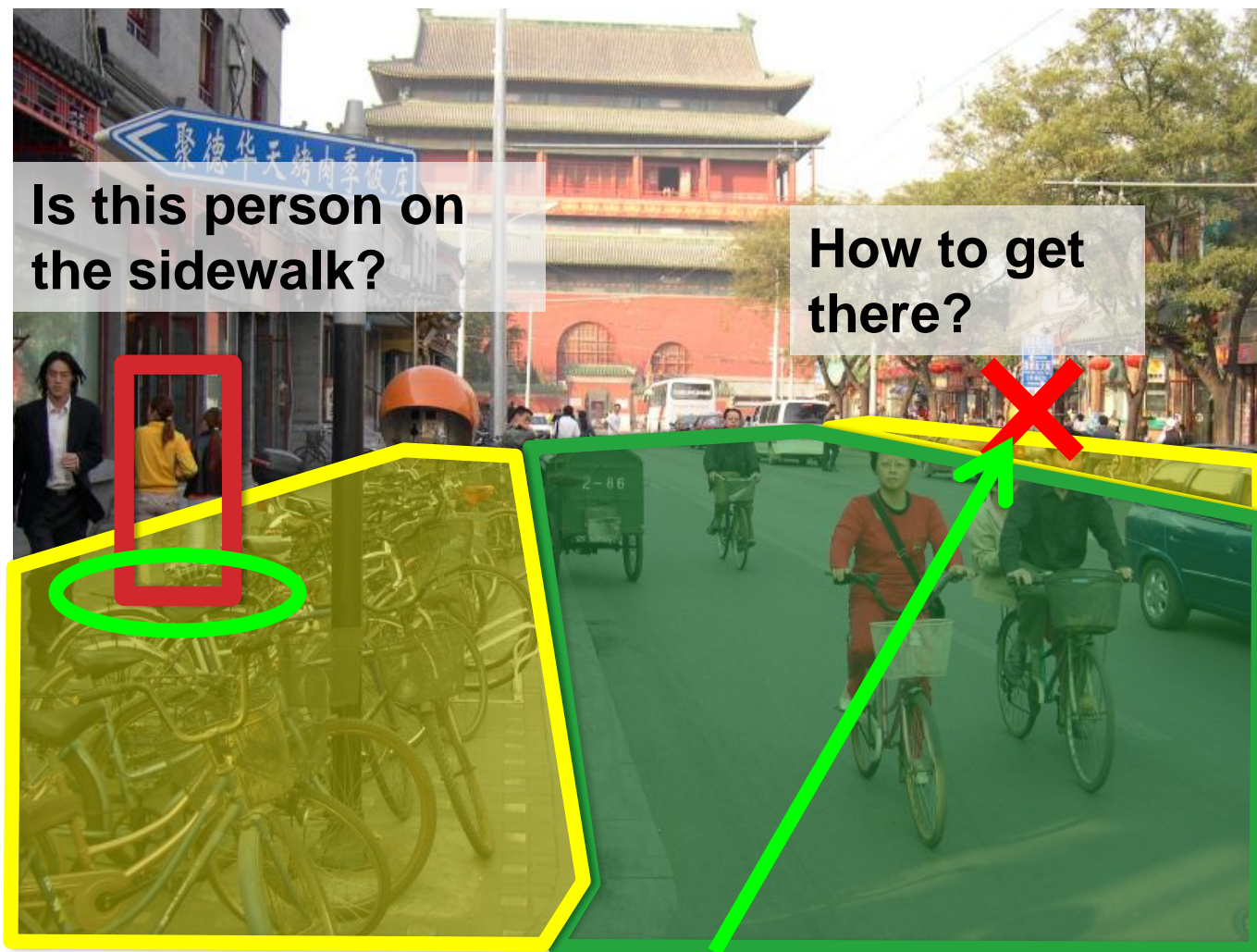
LABELING PIXELS TO OBJECT LABELS



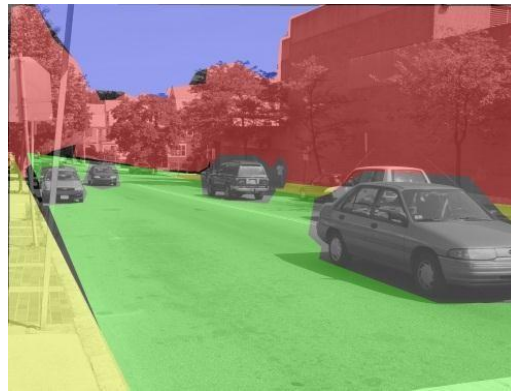
LABELING PIXELS TO OBJECT LABELS



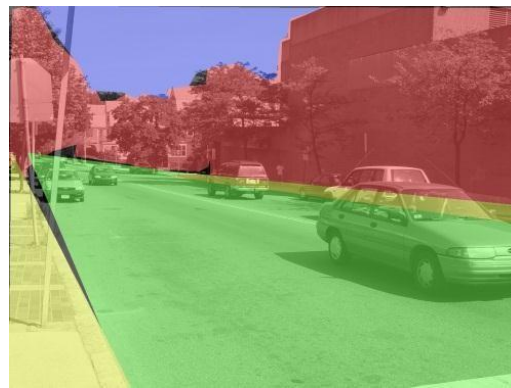
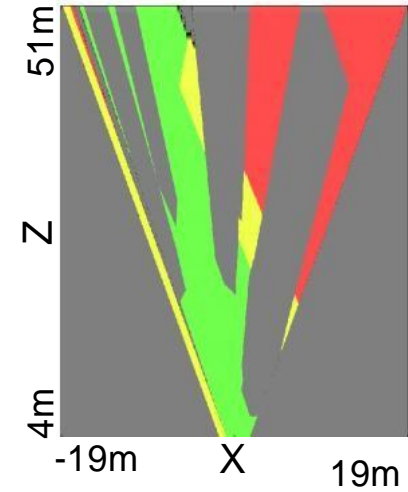
COMPLETE SURFACE LABELS



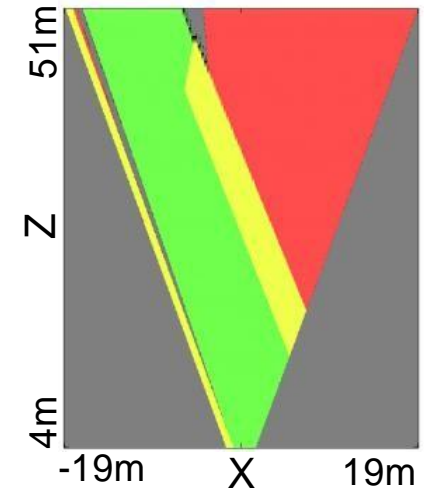
VISIBLE VS. COMPLETE SCENE



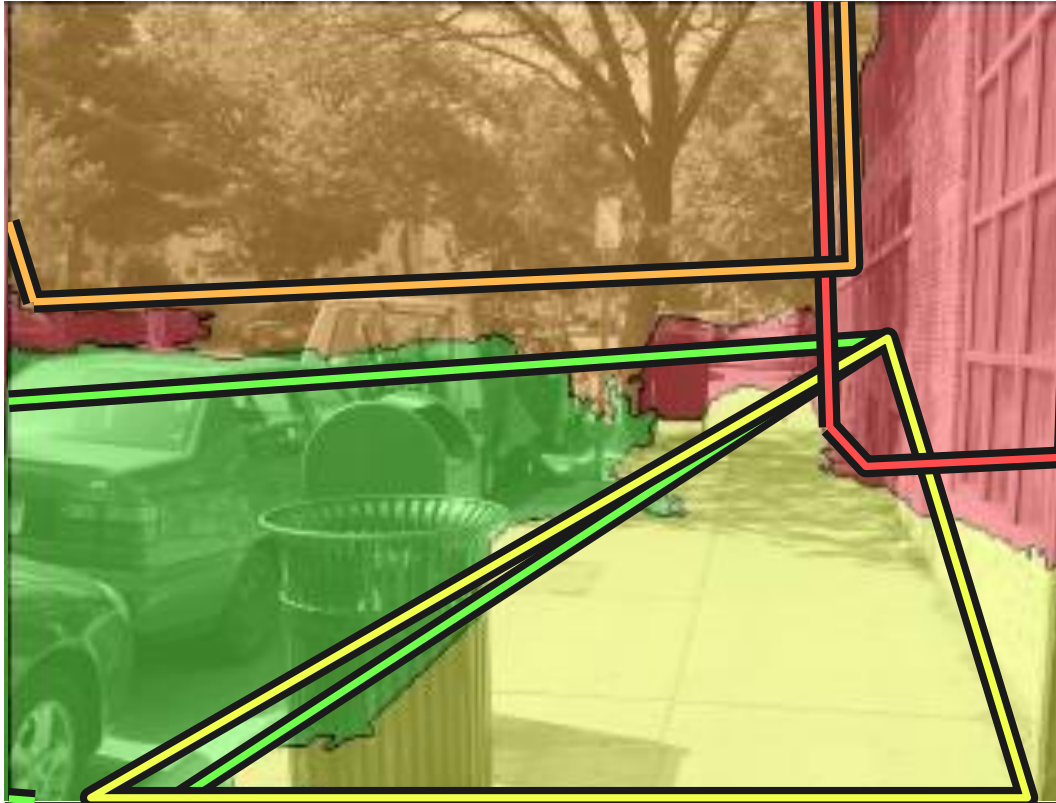
Visible Surface Labels



Complete Surface Labels



PROBLEM STATEMENT



Structural noise prediction/detection for multiple surfaces

PREVIOUS WORKS

- **Generic image parsing**
 - Geometric Context [Hoiem et al. 2007]
 - SuperParsing [Tighe and Lazebnik, 2010]
- **Scene layout**
 - Single image structure recovery [Lee et al. 2009]
 - Thinking inside of the box [Hedau, 2010]
 - Thinking Blocks World Revisited [Gupta et al. 2010]
 - 3D estimation of objects and scene layout [Greiger et al. 2011]

FRAMEWORK

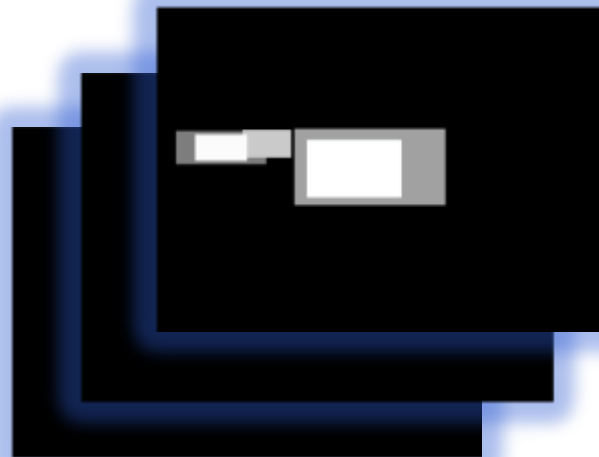
Input



FRAMEWORK

Input

Compute
Features

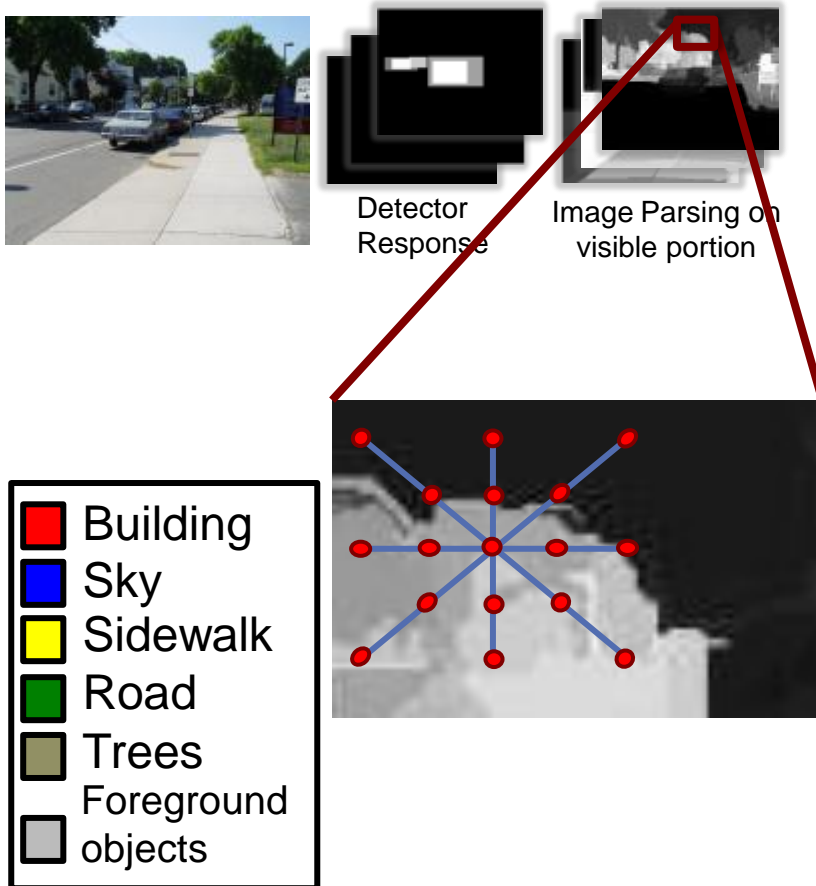
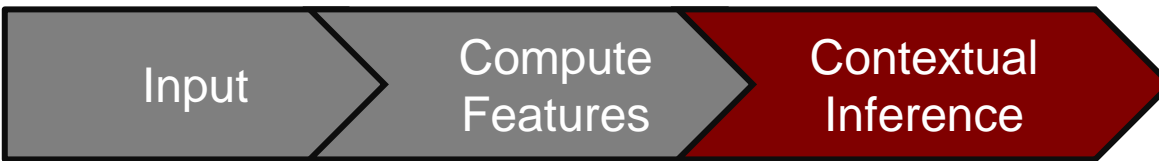


Detector
Response

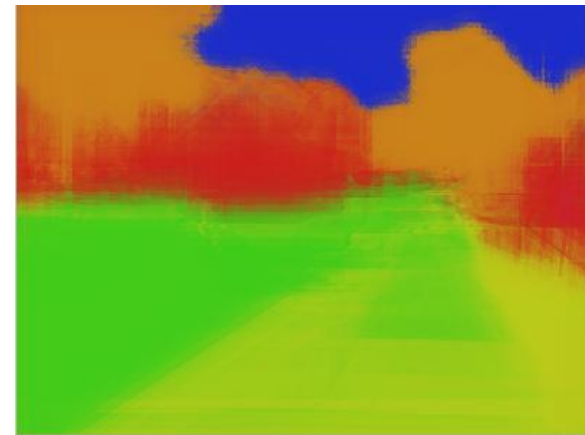


Image Parsing on
visible portion

FRAMEWORK



Idea 1:
Use surrounding context



Label Confidence Map

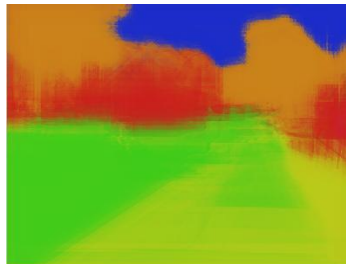
FRAMEWORK



Detector Response

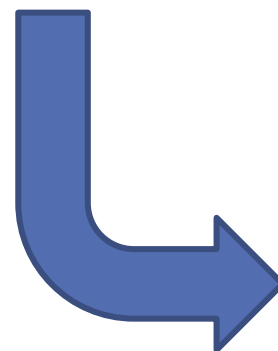


Image Parsing on visible portion



Idea 2:
Complete regions
Have shape prior

- Building
- Sky
- Sidewalk
- Road
- Trees
- Foreground
- objects



Polygon Retrieval



FRAMEWORK

Input

Compute
Features

Contextual
Inference

Retrieve
Shape

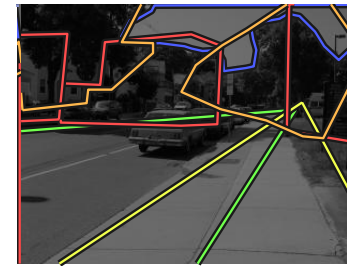
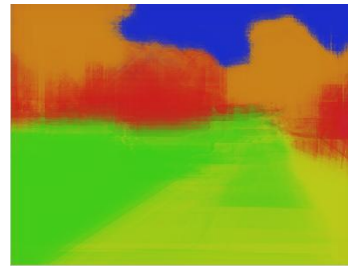
Final
Result



Detector
Response



Image Parsing on
visible portion



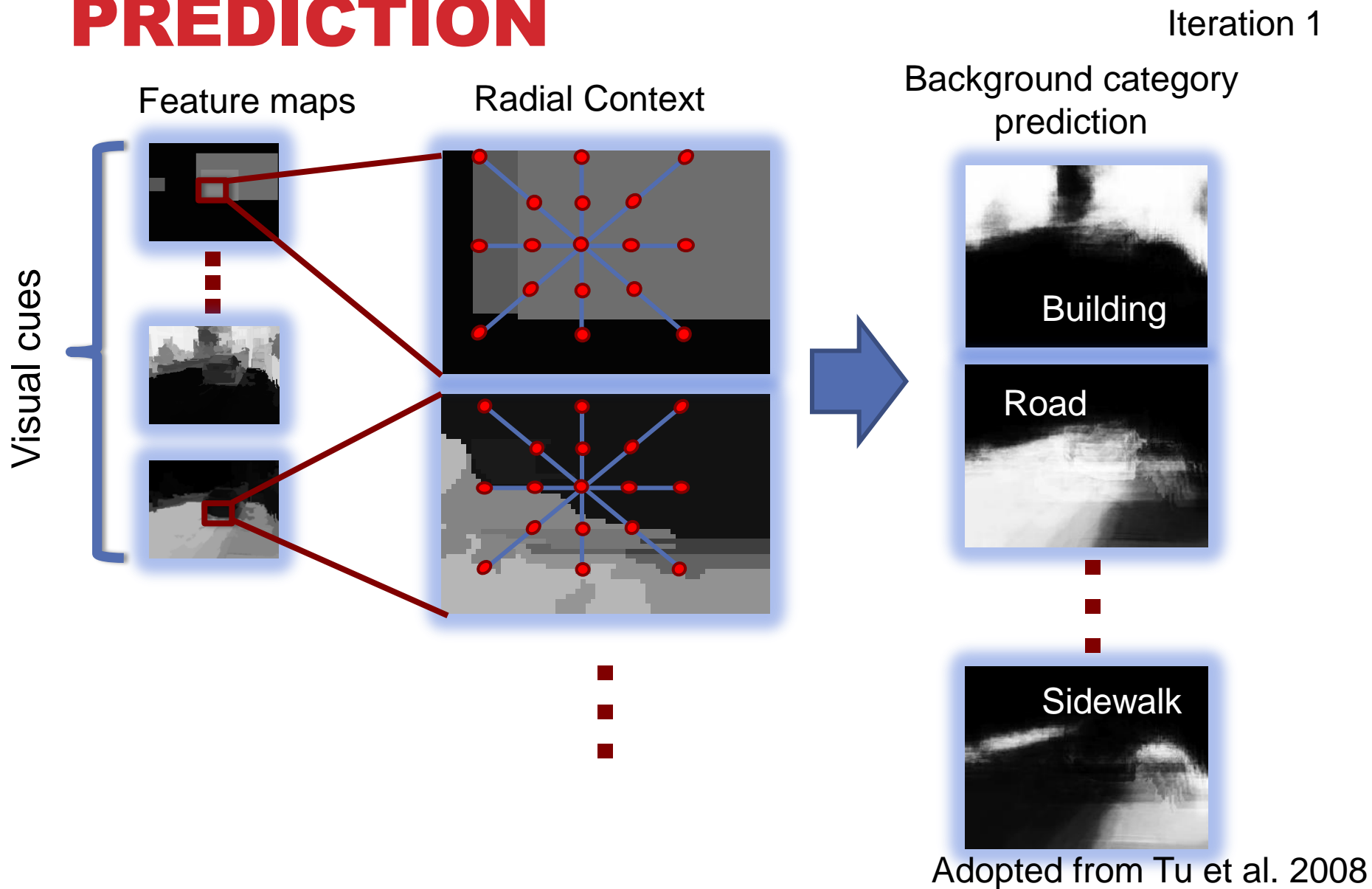
- Building
- Sky
- Sidewalk
- Road
- Trees
- Foreground
- objects

VISUAL CUES

- **Image parsing of the visible pixels**
 - GeometricContext based region classifier
 - BG and FG categories
- **Pretrained object detector (DPM)**
 - Car, pedestrian, chair, table, etc.
- **Use label confidence map as visual cues**

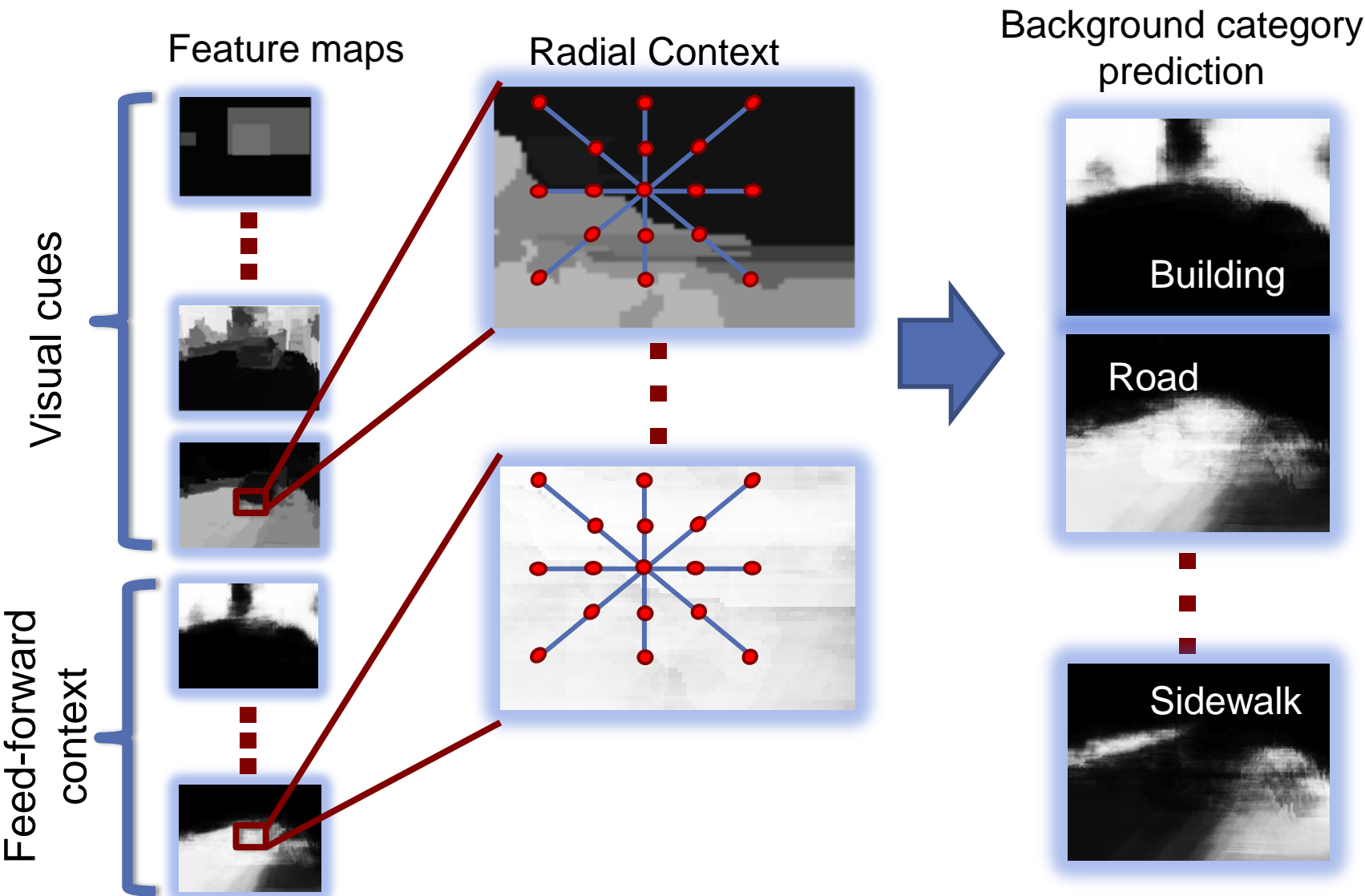


FEED-FORWARD PREDICTION



FEED-FORWARD PREDICTION

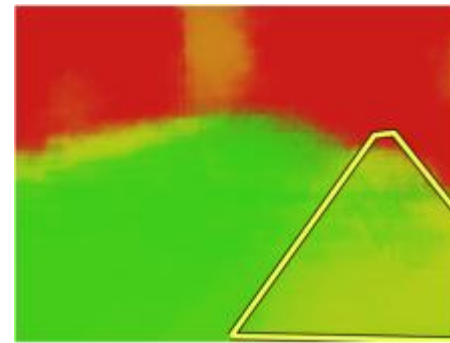
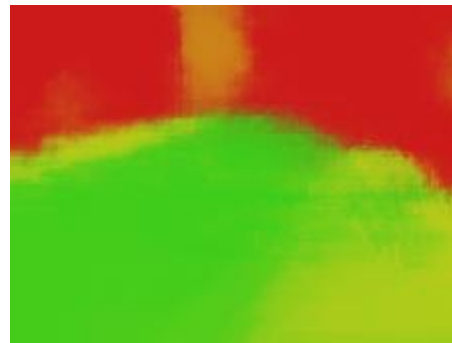
Iteration 2



SHAPE TRANSFER

- Retrieving polygons based on confidence map

$$Score(\mathbf{G}_{1,i}, \mathbf{P}_{1,k}) = \frac{\sum_x \min(G_{l,i,x}, P_{l,k,x})}{\sum_x \max(G_{l,i,x}, P_{l,k,x})}$$

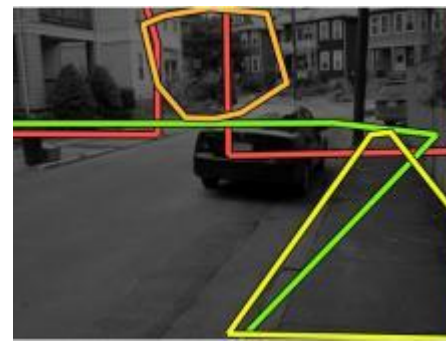
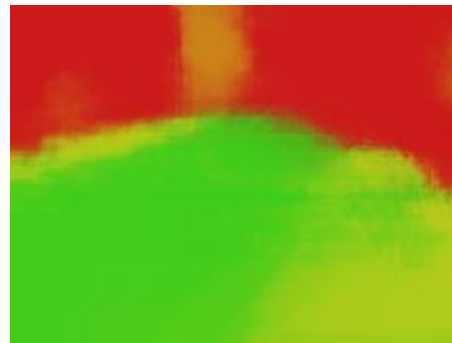


Sidewalk

SHAPE TRANSFER

- Retrieving polygons based on confidence map

$$Score(\mathbf{G}_{1,i}, \mathbf{P}_{1,k}) = \frac{\sum_x \min(G_{l,i,x}, P_{l,k,x})}{\sum_x \max(G_{l,i,x}, P_{l,k,x})}$$



Sidewalk

Building

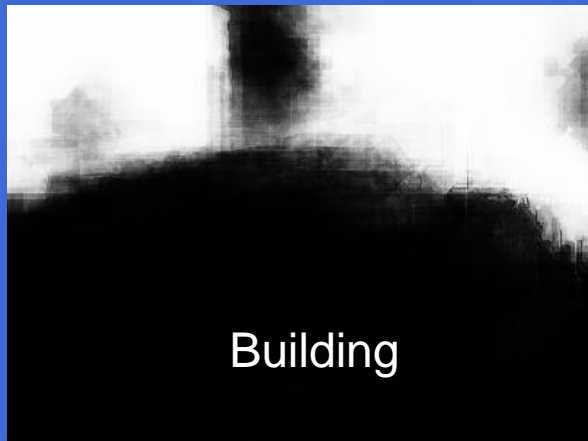
Road



CONTEXT CHANGE



Aerated shape prior



FINAL RESULT

Top 1 layout guess

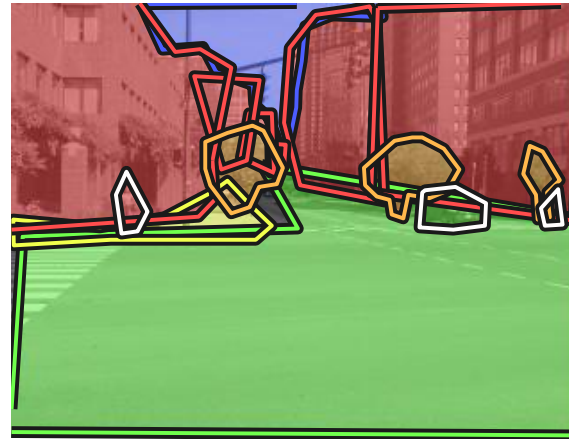


Top 2 layout guess



EVALUATION DATASETS

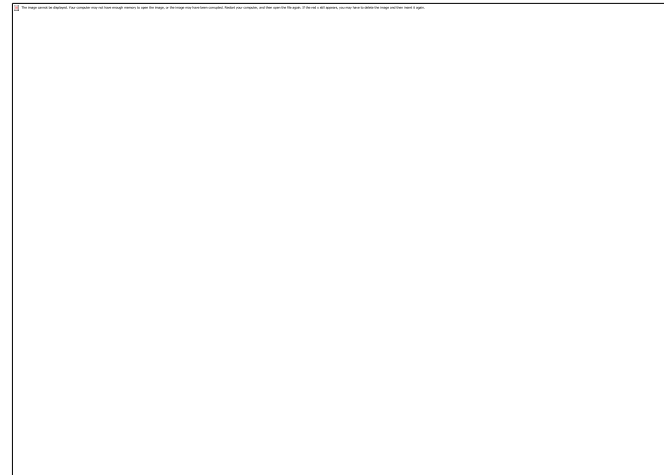
- **StreetScene (3.3K)**
 - Street scenes in Boston
- **IndoorScene (300)**
 - Indoor room layout
- **SUN09 (~10K)**
 - Part of LabelMe.



StreetScene



IndoorScene



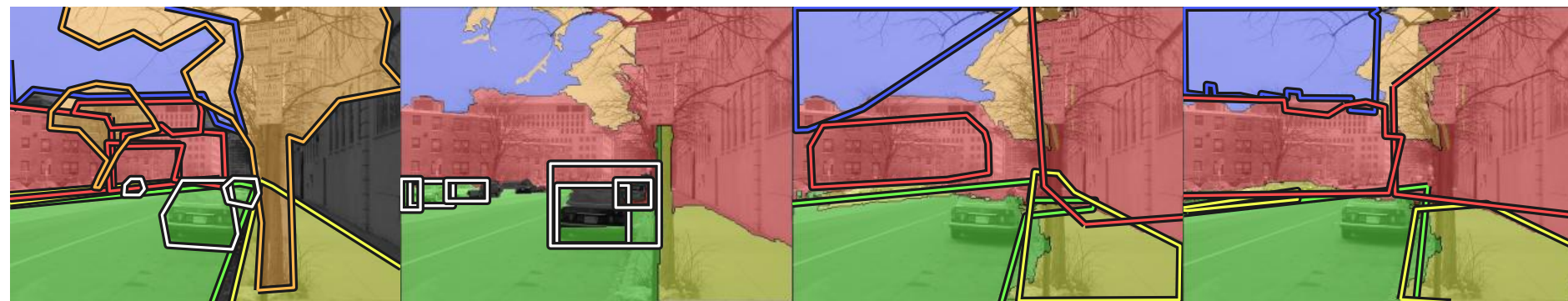
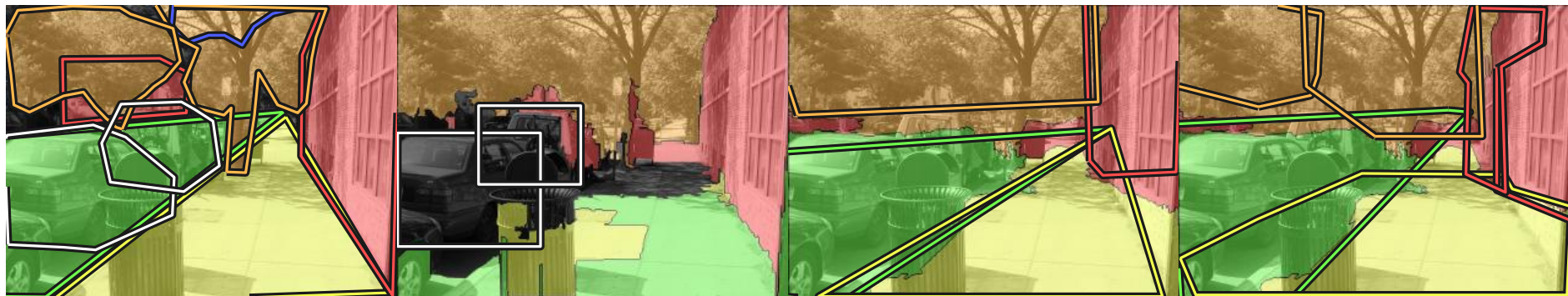
SUN09

Ground truth

Visible Surface

Completed surfaces
w/ first polygon guess

Completed surfaces
w/ second polygon guess



 Building
 Sky

 Sidewalk
 Road

 Trees
 Store

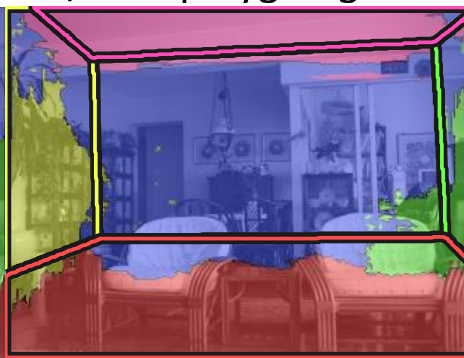
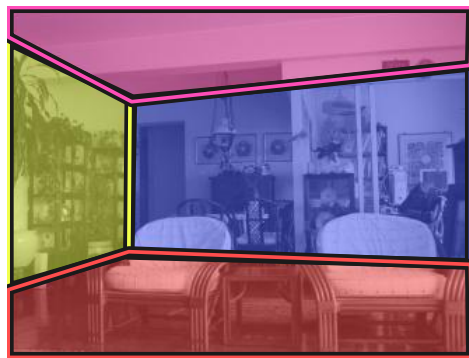
 Foreground objects

Ground truth



Visible Surface

Completed surfaces
w/ first polygon guess

Completed surfaces
w/ second polygon guess



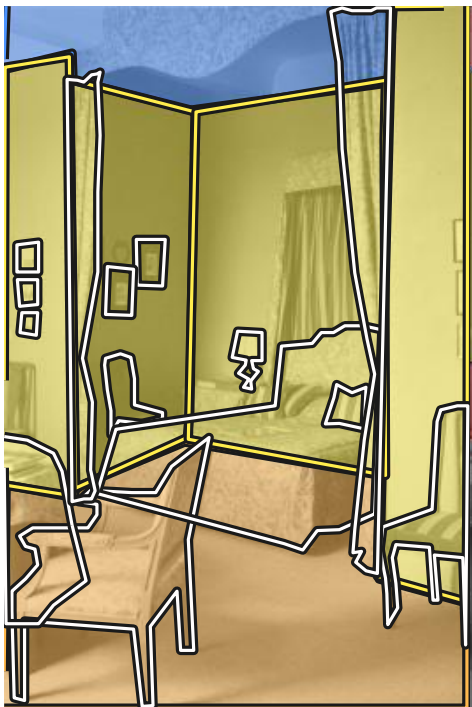
 Floor
 Center wall

 Left wall
 Right wall

 Ceiling
 Foreground objects

Ground truth

Visible Surface

Completed surfaces
w/ first polygon guessCompleted surfaces
w/ second polygon guess

Building

Ground

Sky

Water

Chair

Ceiling

Field

Tree

Sidewalk

People

Floor

Road

Wall

Vehicle

Object

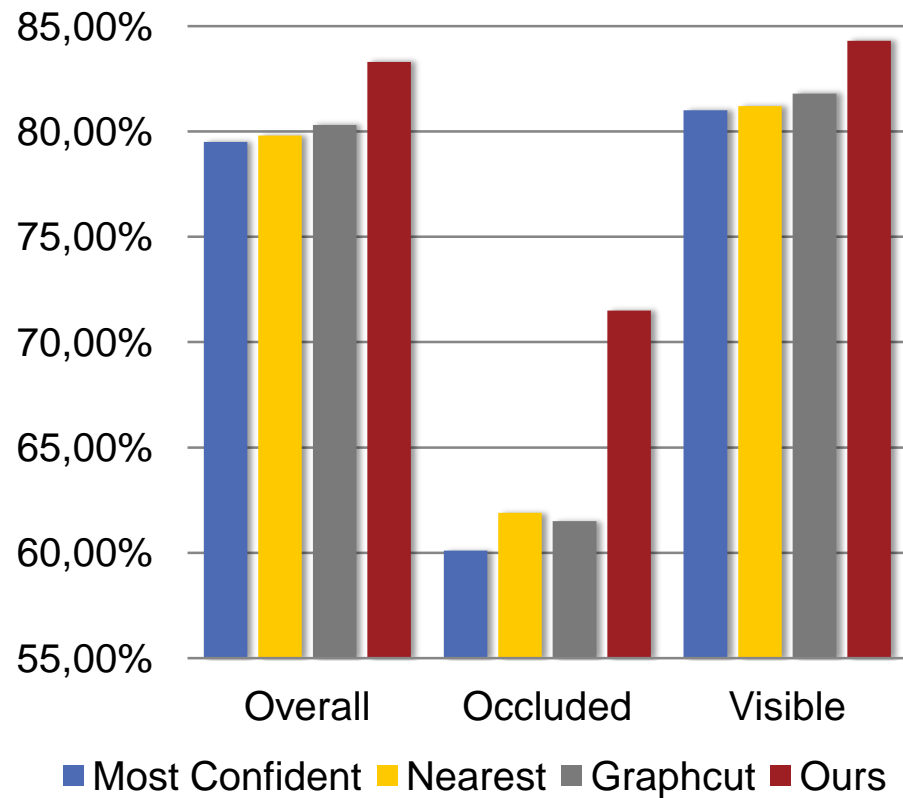
BASELINES

- **Most confident**
 - The most confident background label
 - Relies on visual appearance
- **Nearest**
 - Fills in using nearest background label
- **Graphcut**
 - Contrast sensitive smoothing

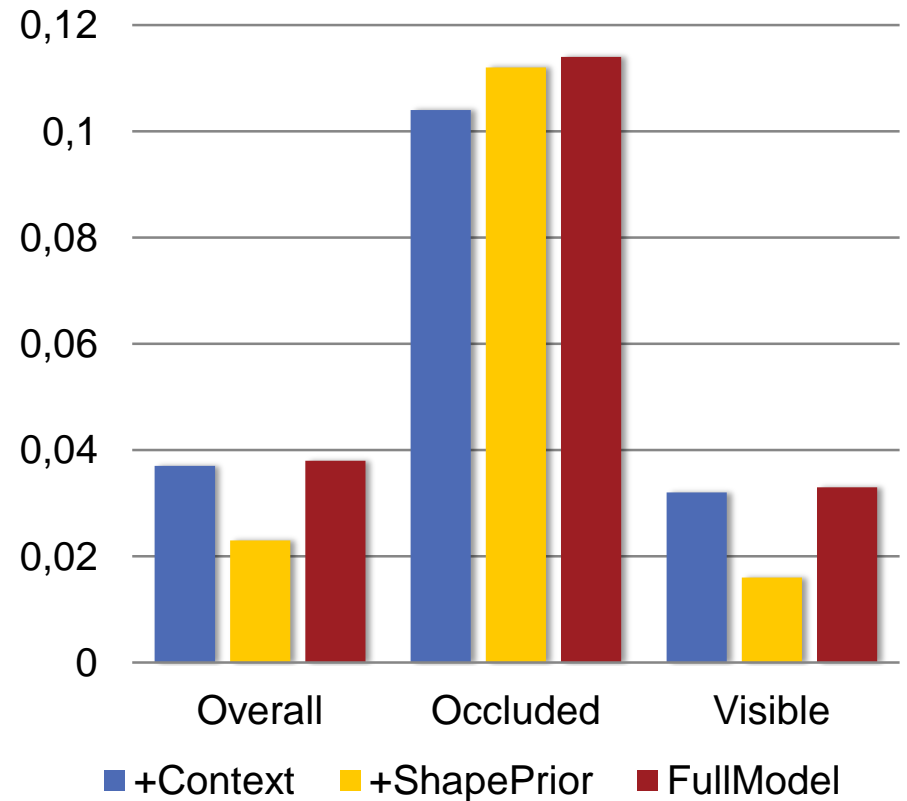
$$\psi_{pairwise}(x_1, x_2, l_1, l_2) = \mathbf{1}(l_1 \neq l_2) [\lambda_1 P(fg|x_1) + \lambda_2 (1 - P(fg|x_1)) e^{\frac{(I(x_1) - I(x_2))^2}{\sigma^2}}]$$

QUANTITATIVE RESULTS

Pixel Accuracy on StreetScene

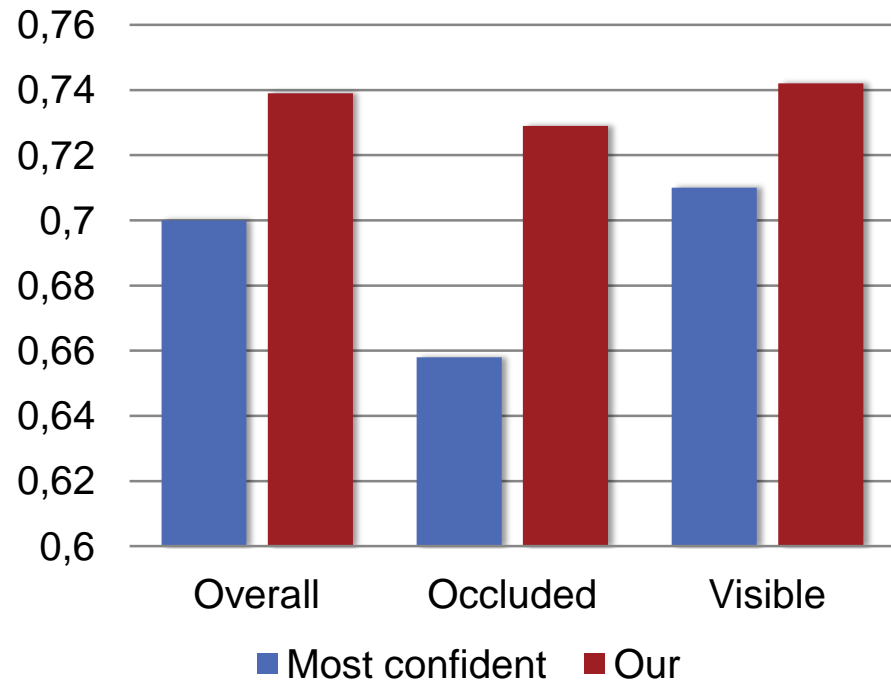


Feature Effectiveness (Improvement)

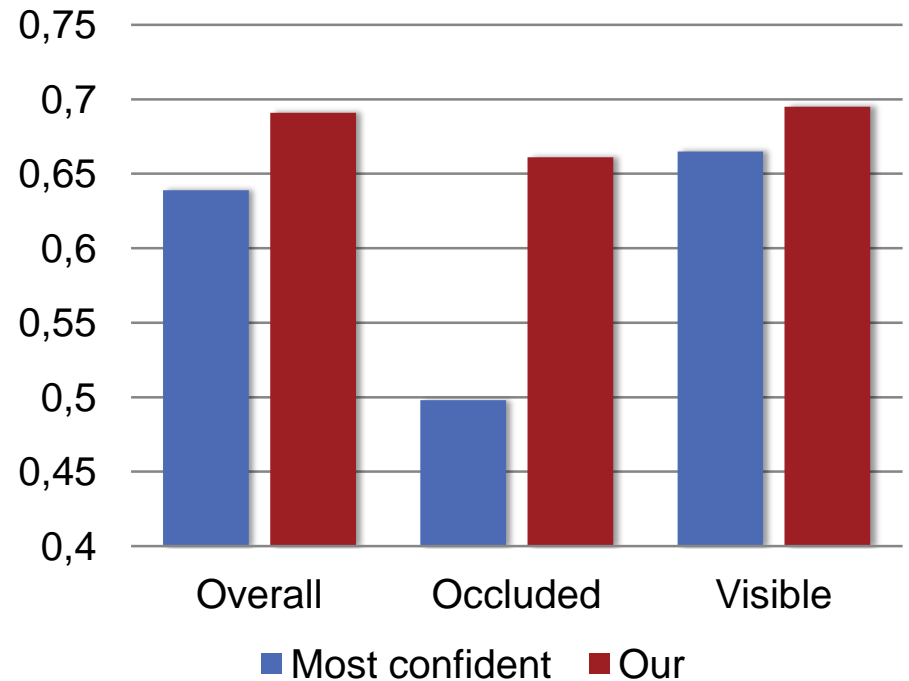


QUANTITATIVE RESULTS

Performance on IndoorScene

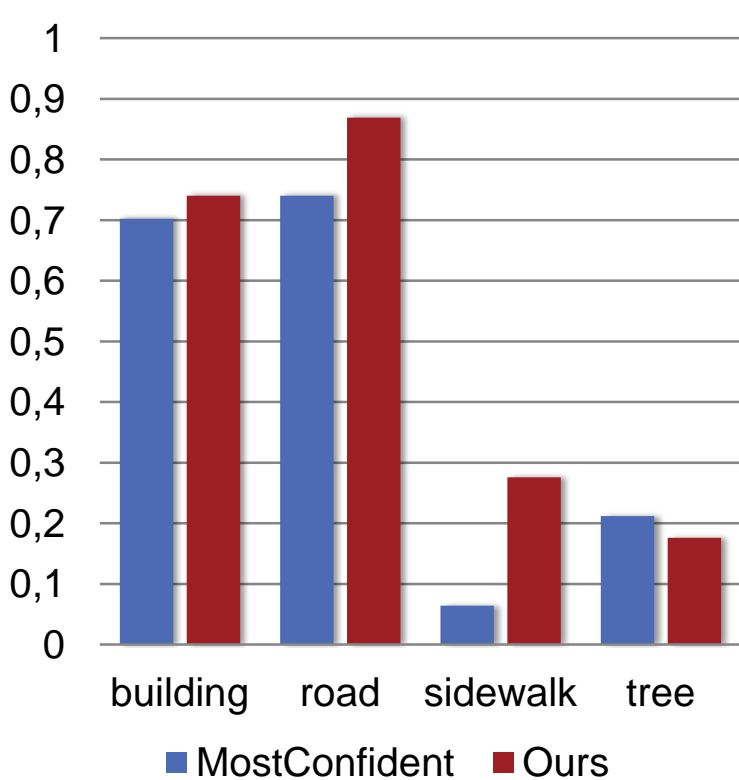


Performance on SUN09

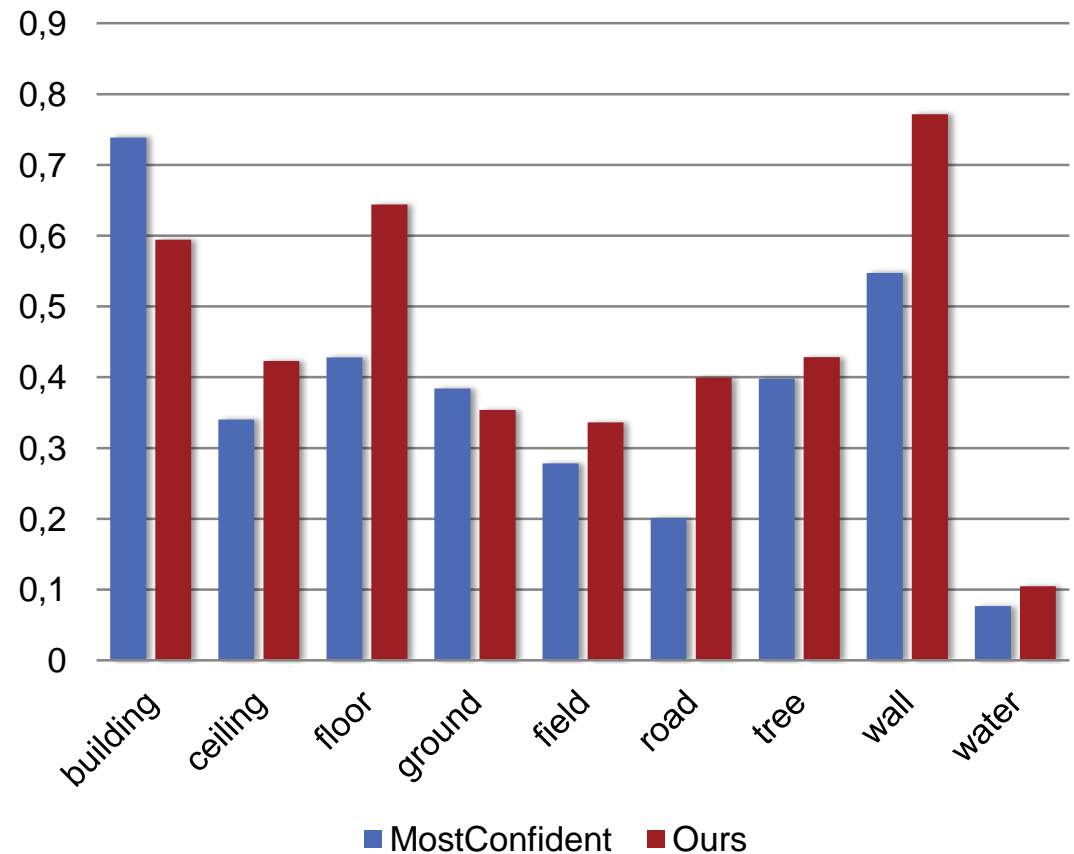


QUANTITATIVE RESULTS

Perclass Accuracy on StreetScene (occluded)



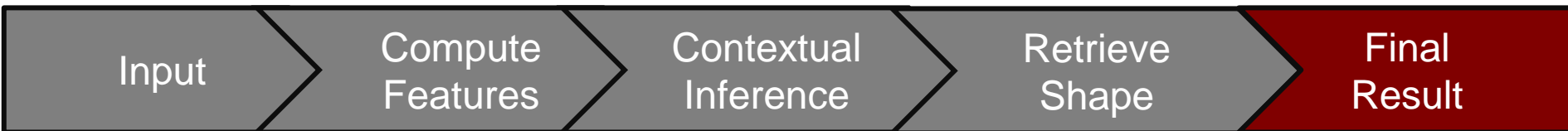
Perclass Accuracy on SUN09 (occluded)



CONCLUSIONS

- **New task: Labeling underlying surfaces**
- **General algorithm for**
 - Visible pixel maps and object detection
 - Feed-forward context
 - Simple shape prior
- **Evaluated on 3 datasets**
- **Code is available, now!**
 - <http://web.engr.illinois.edu/~guo29/projects/surface.html>

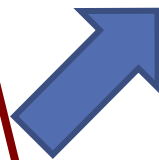
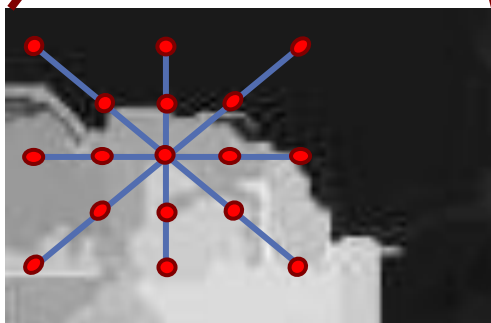
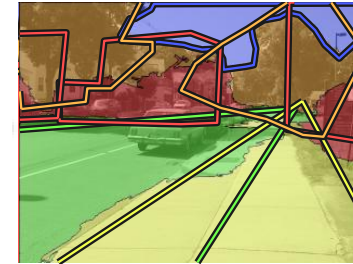
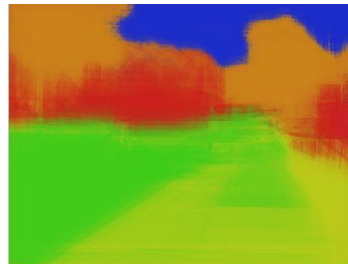
THANKS



Detector Response



Image Parsing on visible portion



- Building
- Sky
- Sidewalk
- Road
- Trees
- Foreground
- objects