Unsupervised Patch-based Context from Millions of Images

Santosh K. Divvala
PhD student

Carnegie Mellon
The Robotics Institute
What is Region Classification?

- Given an image, label every pixel within an image with one of $K$ labels
Region Classification

8 classes: sky, road, grass, water, building, tree, mountain, foreground

7 classes: sky, ground, frontal /\, left \rightarrow, right \leftarrow, solid X, porous O

Gould et al., 2009

Hoiem et al., 2007
Existing Approaches

- **Local patch-based models**
  - Enough data but poor discrimination

- **Higher-order region-based models**
  - Good discrimination but not enough data
Semi-supervised Approach

Labeled Data

Image with labels

Labeled Data

Classifier

Output

? 

Input

Lots of Unlabeled Data
**Approach**

Input Image → Local Classifier → Matches from Unlabeled Data → Patch-based contextual Prior → Final Result
Approach Overview

- Given: Few labeled images, and Millions of Unlabeled images

1. Train a local classifier using the labeled images
2. Retrieve matches from the unlabeled set using long-range neighborhood
3. Compute contextual prior and update local classifier result
Step 1: Local Classifier

- Multiple Segmentation based Approach
  - Simple, fast yet powerful
- Gather good spatial support for classification
- Discourages non-homogenous segments!
Getting good matches: Challenge 1

- What features to use?

Local classifier (confident-only) predictions

Hays & Efros 2007
Getting good matches: Challenge 2

• What is the right spatial support?
Step 2: Nearest Neighbor Matching

- L1 metric for appearance and semantic features
  - Weighted combination
- Sliding window based search
- 6 hours per query image on a single CPU
  - Cluster of 800 CPUs
Step3: Using Retrieved Matches

• Contextual Prior: Marginalization of nearest neighbor outputs

• Include the prior as a feature and retrain the local classifier
Semantic Region Classification

Ground Truth    Classifier Output    Final Result    Result of [1]
Geometric Region Classification
Quantitative Results

- Metric: Pixel-based average accuracy
- Semantic classification (Stanford)
  - Improves by 2.4% over [1] (75.6% to 78.0%)
- Geometric classification (CMU)
  - Main-class by 1.2% (87.2% to 88.4%) & Sub-class by 2.6% (59.3% to 61.9%) over [2]

Matching Results
Matching Results

Global Matches

Sub-Image Matches
Conclusion

• An effective (semi-supervised) approach for using millions of web images to improve region classification

• Good nearest neighbor matches from noisy web images
  – Sub-image matching & semantic features
Thank You
Problem with off-the-shelf SSL

- High appearance ambiguity prevents clustering assumption to hold
- Having a large patch would indeed lead to well-separated clusters but with little/no labeled samples!
Results
Results