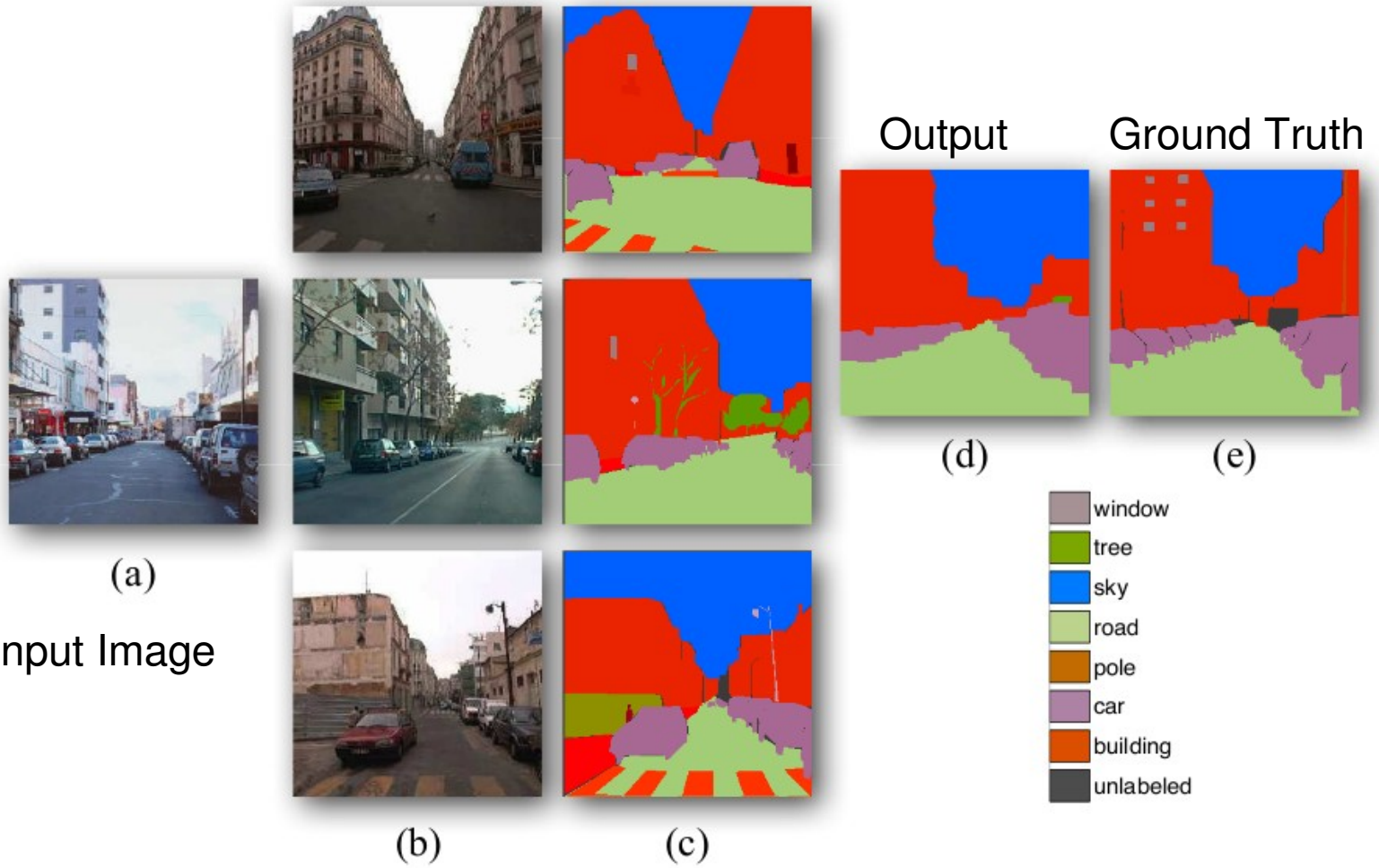
Nonparametric Scene Parsing: Label Transfer via Dense Scene Alignment

Basic Idea



Input Image

Top Matches

Annotations of top matches

From Liu, Yuen, Torralba

Previous Approaches To Object Recognition

- Bags of Features (words)
- Template Matching
- Shape models
- Cons
  - Usually deal with fixed number of object categories
  - Must retrain if adding new category

Sift Flow for Dense Scene Alignment

- Sift Flow-aligns an image with similar images
- General Idea
  - Use histogram intersection with bag of words to represent image
  - Find nearest neighbors to the image
  - Using dense sampling, match sift features between pairs of images

From Liu, et al. “Sift Flow: Dense Correspondence across Different Scenes”

A threshold to prevent very large discrepancies from skewing the result

Difference in sift features between Corresponding pixels of images

$$E(\mathbf{w}) = \sum_{\mathbf{p}} \min\left(\left\|s_1(\mathbf{p}) - s_2(\mathbf{p}+\mathbf{w}(\mathbf{p}))\right\|_1, t\right) +$$

$$\sum_{\mathbf{p}} \eta\left(|u(\mathbf{p})| + |v(\mathbf{p})|\right) +$$

Sum of magnitude of flow vector components

$$\sum_{(\mathbf{p},\mathbf{q})\in\varepsilon} \min\left(\alpha|u(\mathbf{p}) - u(\mathbf{q})|, d\right) +$$

$$\sum_{(\mathbf{p},\mathbf{q})\in\varepsilon} \min\left(\alpha|v(\mathbf{p}) - v(\mathbf{q})|, d\right).$$

Difference between flow vector of pixel *p* and the flow vector's of its neighbors

From Liu, Yuen, Torralba

What to Take Away From the Equation

We want:

1) Small flow vectors

2) The flow vectors of the pixels surrounding pixel *p* to all be similar

3) Pixels that are being matched to have similar sift features

The energy function is then minimized

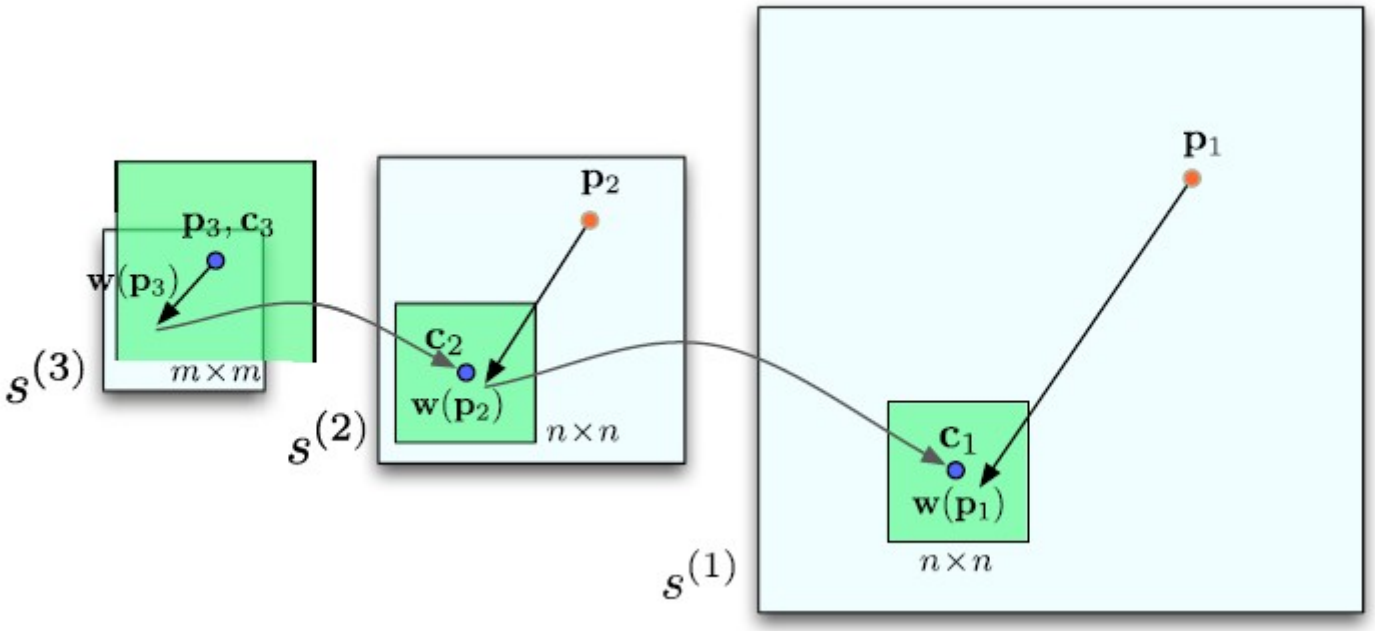
Problems With Sift Flow…

- Scalability—becomes computationally expensive with large images

  -pixel in one image can match with any pixel in the other image

  -If the image size is $n \times n$, then the algorithm would be $O(n^4)$.

Their Solution…



Down Sampled Image ← Original Image

1) Find matching points in downsampled images
2) Proceed to the next image level and use prior knowledge gained to only search over a subset of the image
3) Repeat yet again…

Algorithm is now $O(n^2 \log n)$ AND performs better than ordinary matching!

From Liu, Yuen, Torralba

Label Parsing

1)Find k-nearest neighbors to the query image using GIST matching

2)Compute SIFT flow from query image to each nearest neighbor…use this to rerank images

3) Select best candidate images

4)Use Sift flow to transfer annotations from the candidate images to the query image

Sift features

annotations

Sift Flow

Likelihood term

$$-\log P\left(c|I,s,\{s_i,c_i,\mathbf{w}_i\}\right)=\sum_{\mathbf{p}}\psi\left(c(\mathbf{p});s,\{s_i'\}\right)+$$

$$\alpha\sum_{\mathbf{p}}\lambda\left(c(\mathbf{p})\right)+\beta\sum_{\{\mathbf{p},\mathbf{q}\}\in\varepsilon}\phi\left(c(\mathbf{p}),c(\mathbf{q});I\right)+\log Z,$$

Prior Term-the prior probability that object category l appears at pixel *p*…obtained from training set

Smoothness term-used to bias the neighboring pixels to have the same label as pixel *p* given no other information

Difference in sift features between Corresponding pixels of images

Training images with category l at corresponding pixel

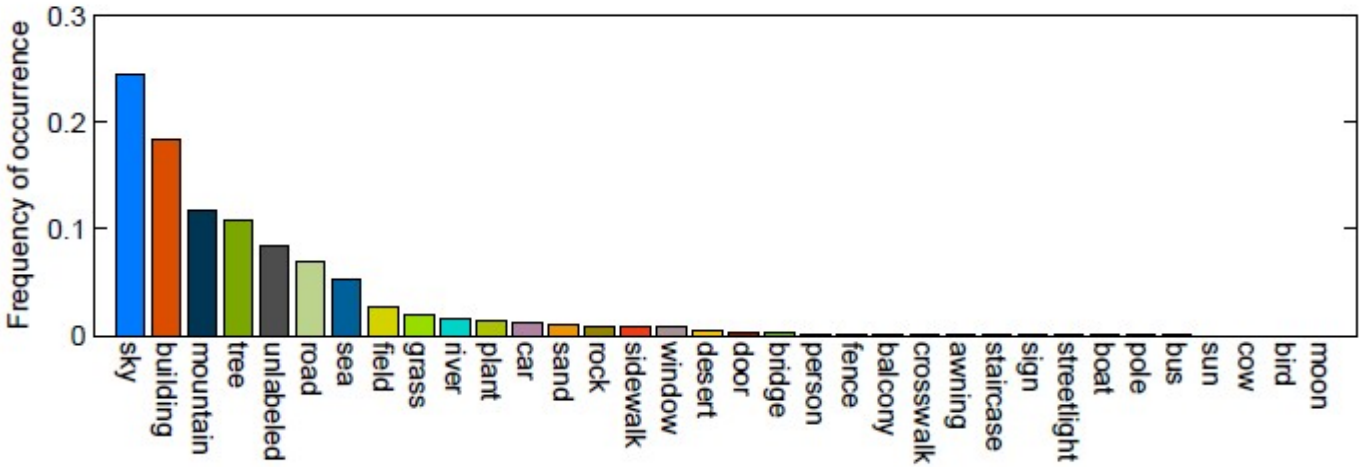
$$\psi\left(c(\mathbf{p})=l\right)=\begin{cases}\min\limits_{i\in\Omega_{\mathbf{p},l}}\|s(\mathbf{p})-s_i(\mathbf{p}+\mathbf{w}(\mathbf{p}))\|, & \Omega_{\mathbf{p},l}\neq\emptyset\\ \tau, & \Omega_{\mathbf{p},l}=\emptyset\end{cases}$$

Max difference of sift features at point *p*

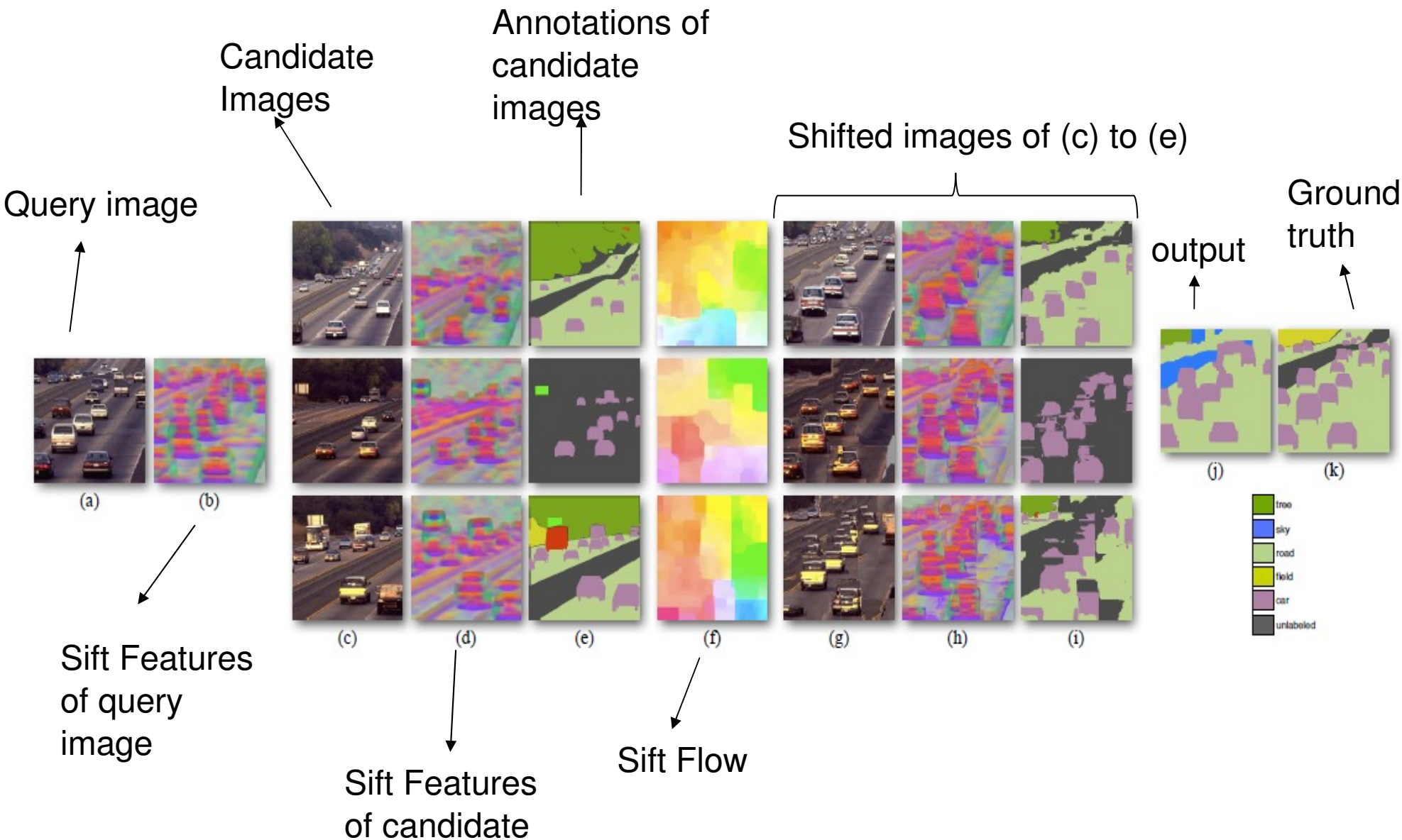
From Liu, Yuen, Torralba

Experiments

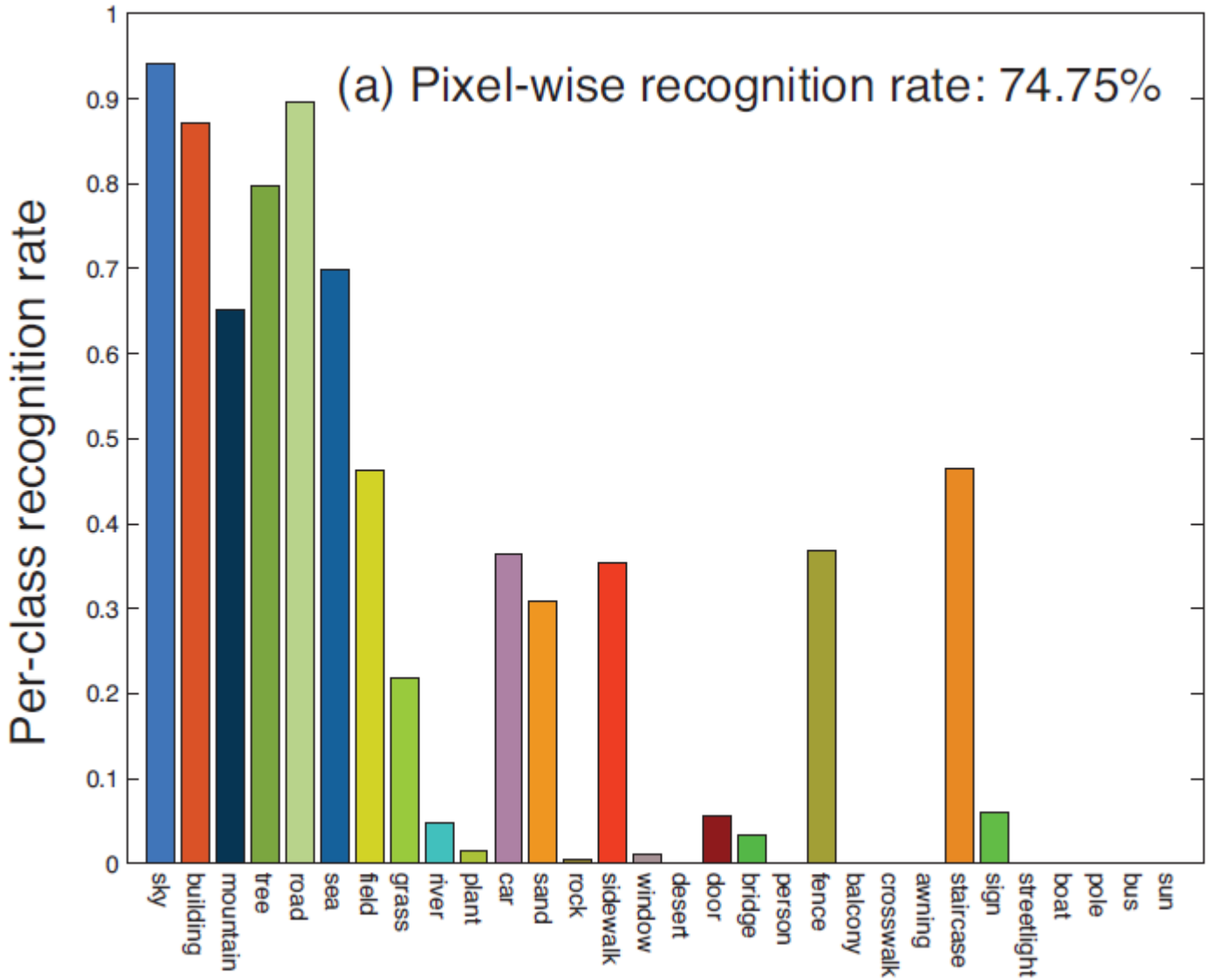
- LabelMe dataset
  - 2488 for training, 200 for test
  - Top 33 object categories (34th category="other")



From Liu, Yuen, Torralba



From Liu, Yuen, Torralba

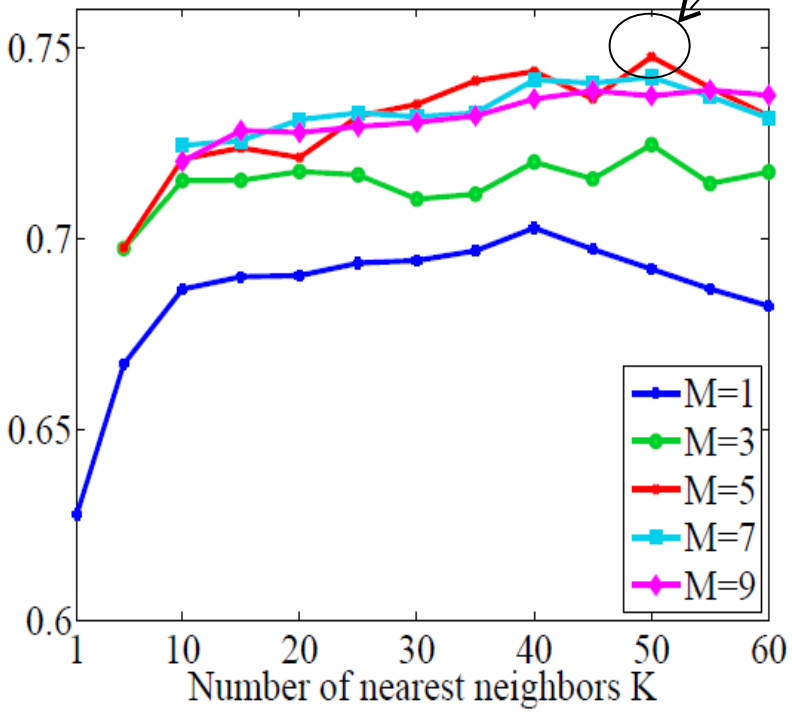


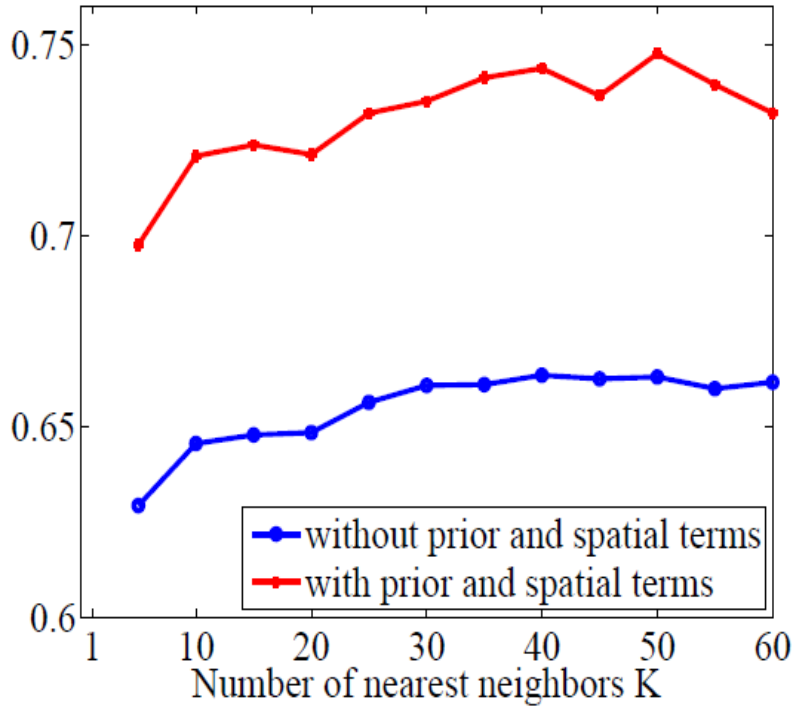
- Pixel wise recognition rate=74.75% (excluding the “unlabeled class”)
- Recognition rate for top 7 categories=82.72%

From Liu, Yuen, Torralba

More Results

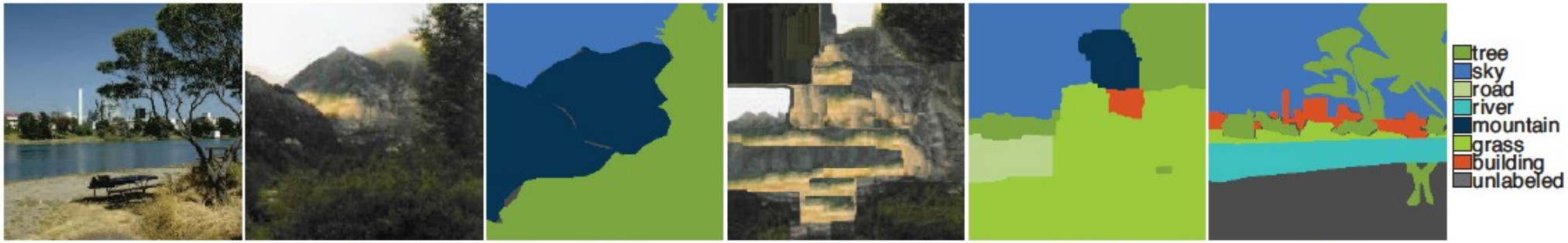
Best Performance: K=50, M=5





From Liu, Yuen, Torralba

It’s still not perfect…



From Liu, Yuen, Torralba