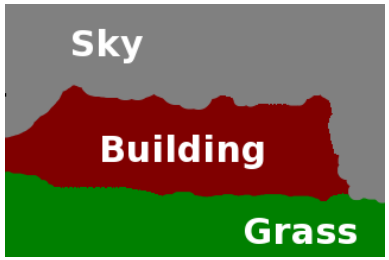
Scene Segmentation with Conditional Random Fields Learned from Partially Labeled Images

Jakob Verbeek **Bill Triggs**

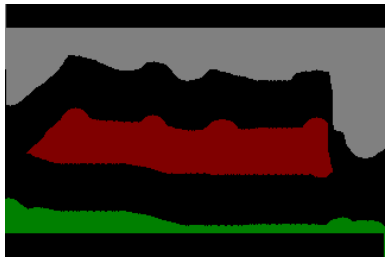
INRIA & Laboratoire Jean Kuntzmann
Grenoble, France

Learning to Segment from Partially Labeled Images

- Given a collection of images with semantic segmentations, learn a model that segments and labels new images



- Do we really need completely labeled training images?
  - tedious to produce, for diminishing returns?

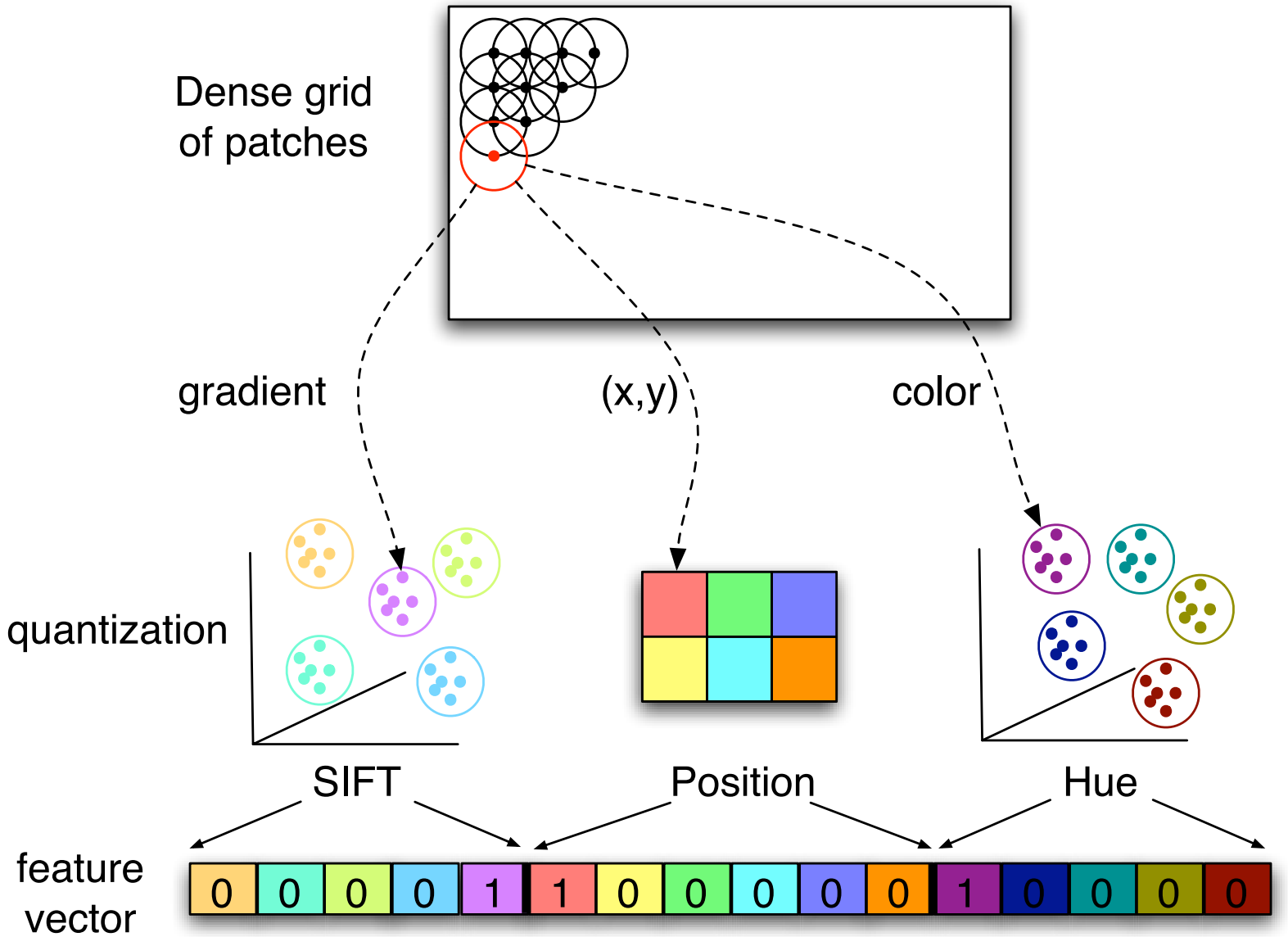


Overview

- **Image features**
  - local image descriptors
  - aggregate features
- **Segmentation model**
  - conditional random field structure
  - learning from partially labelled images
- **Experimental results**
  - how much does each component of the model contribute?
  - resistance to missing labels
- **Summary**

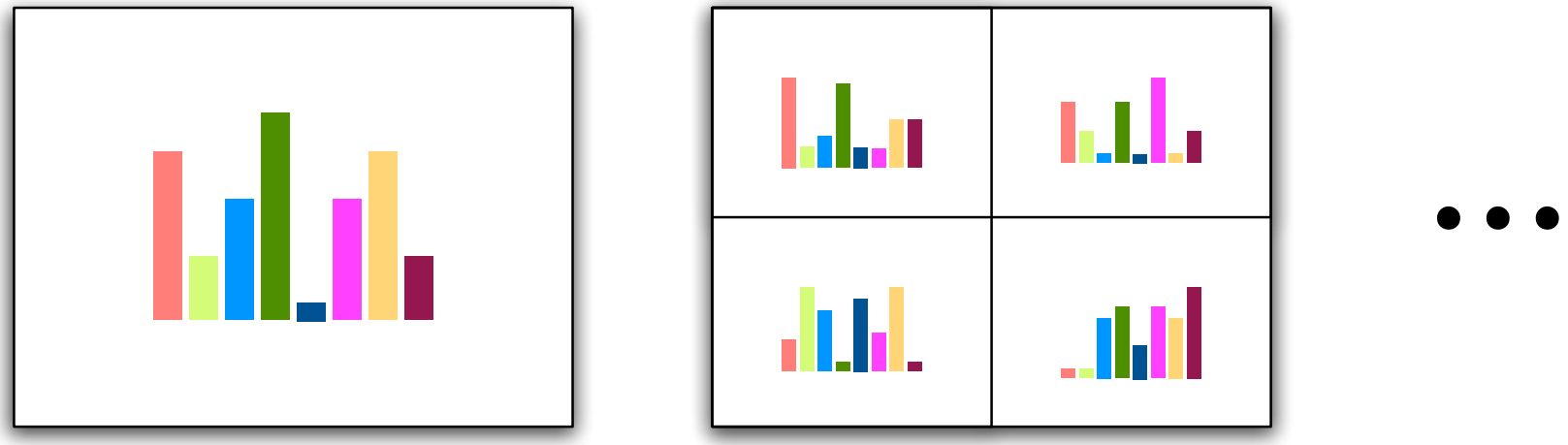
Image Features

Local Image Descriptors



Aggregate Features

- **Capture overall image context** by introducing region-level features
- **Accumulate a local feature histogram** ("bag of visual words") in each cell of a coarse grid covering the image (1×1, 2×2, ...)
- **Histogram couples to every patch label in its cell**

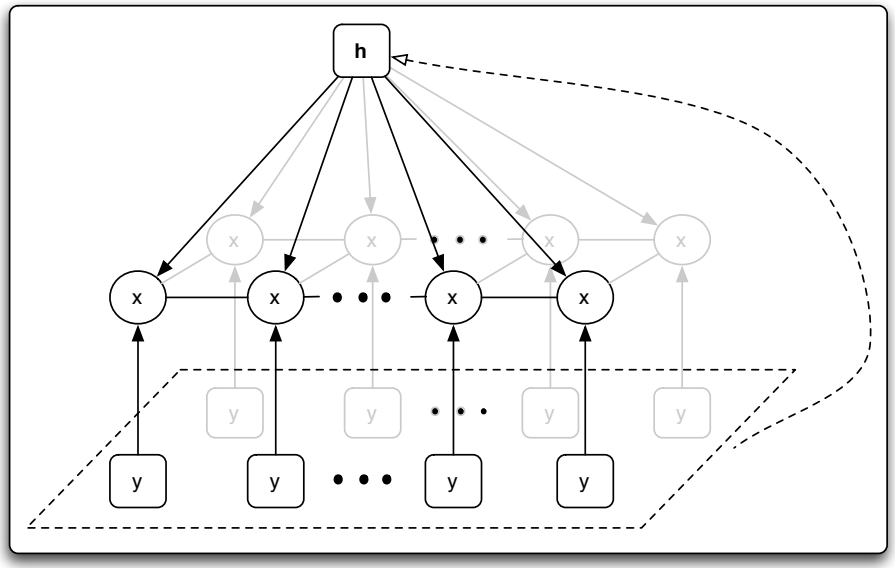


Segmentation Model

Conditional Random Field Structure

$$p(X|Y) \propto \exp -E(X|Y)$$

$$E(X|Y) = \sum_i \alpha_i(x_i) + \sum_i \beta_i(x_i) + \sum_{i \sim j} \phi_{ij}(x_i, x_j)$$



- **CRF energy function combines**
  - local image features α: quantized SIFT, color, position
  - aggregate features β: over whole image and/or coarse image grid
  - neighboring labels ϕ: Potts potential prefers equal labels

Learning from Partially Labelled Images

- **True labeling X is in a given set of allowed labelings A**
  - unlabeled sites can have any label, e.g. near object boundaries
  - weak labeling: constrain some sites to have the same label

- **Principle: Maximize the probability of the compliant label set A**
  - gradient descent based learning

$$L = \log p(X \in A|Y)$$

- **Approximate objective and gradient by running Loopy BP twice**
  - model prediction $p(X|Y)$
  - label completion $p(X|Y, X \in A)$

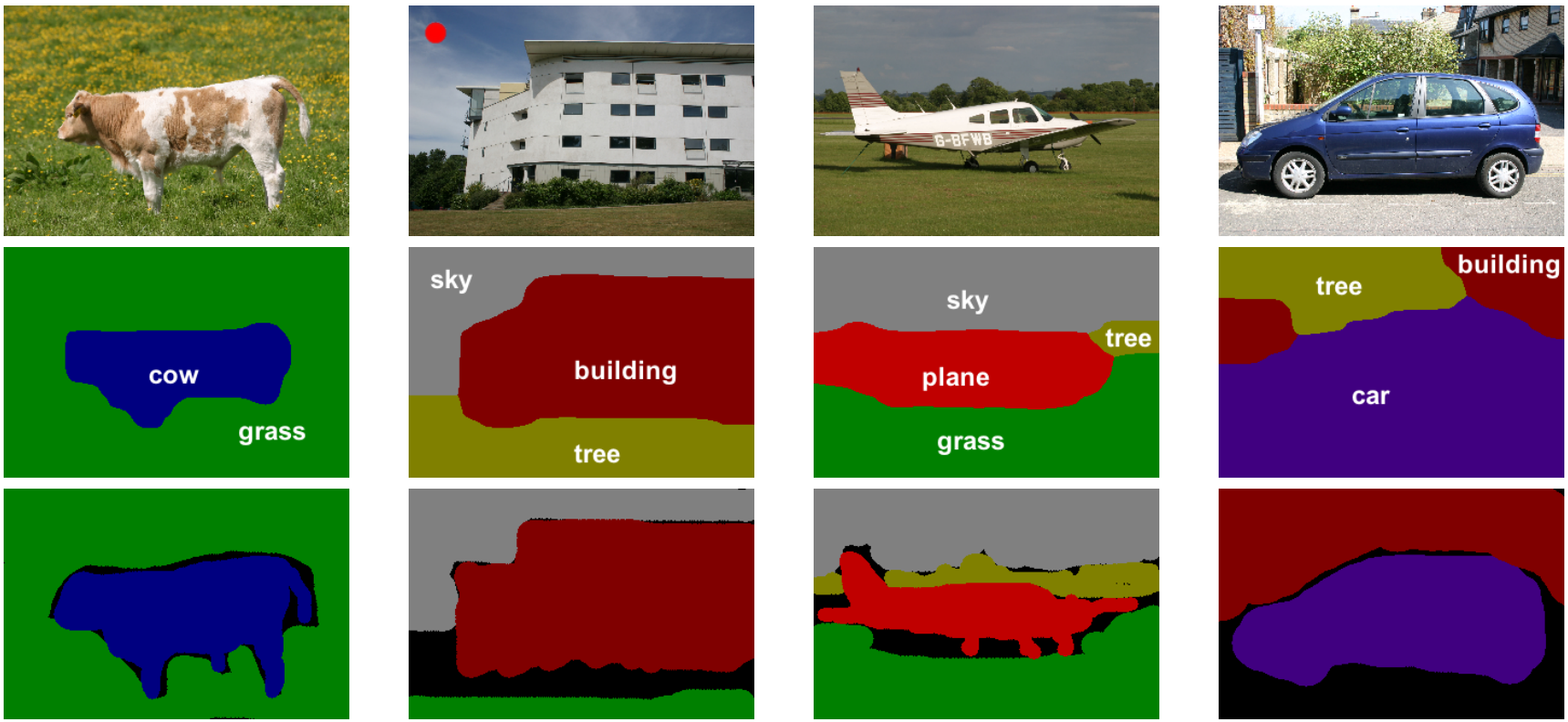
$$L \approx F_{Bethe}\big(p(X|Y)\big) - F_{Bethe}\big(p(X|Y, X \in A)\big)$$

$$\frac{\partial L}{\partial \theta} = \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{p(X|Y)} - \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{p(X|Y, X \in A)}$$

Experimental Results

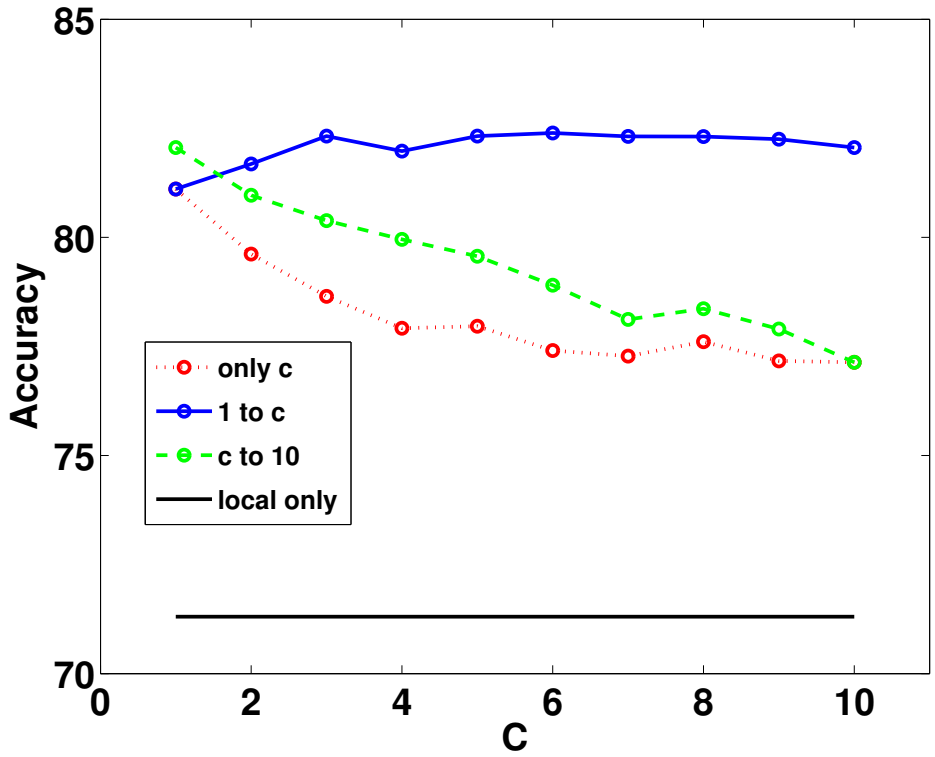
Data Set and Experimental Setup

- **Microsoft Research Cambridge data set**
  - 240 images of 320×213 pixels, 70% of pixels labeled
  - 9 classes: *building, grass, tree, cow, sky, plane, face, car, bike.*
- **Experimental setup**
  - 120 images to train model, 120 to evaluate model, average over 20 trials
  - 20x20 pixel patches, spaced 10 pixels



Local & Aggregate Features

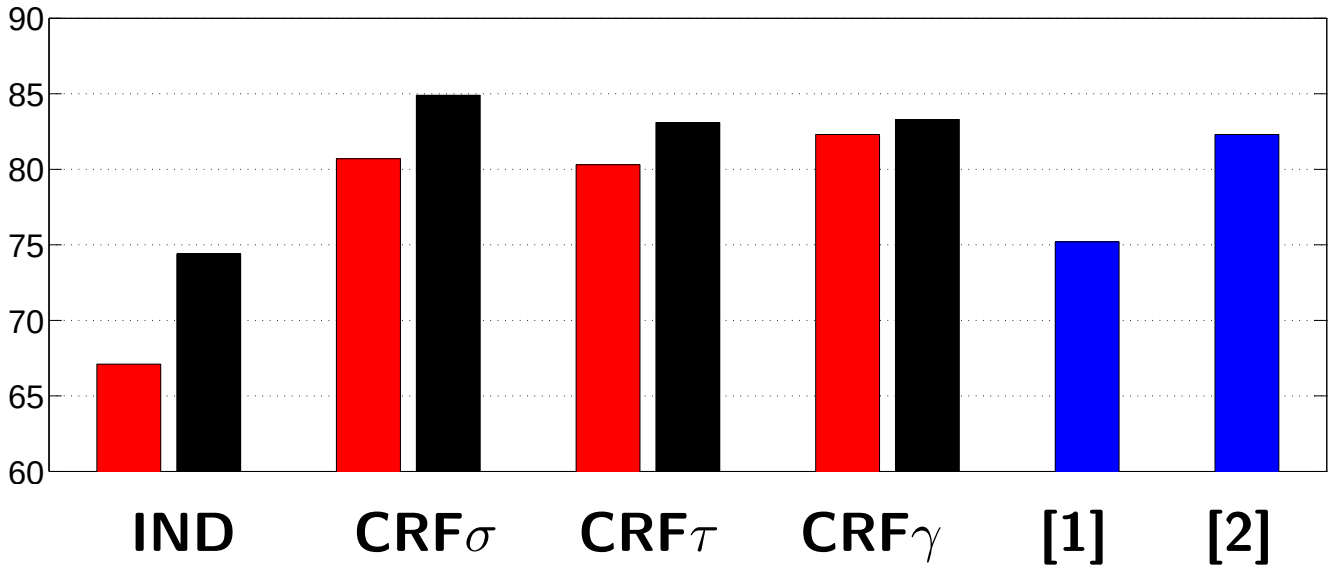
- **Performance using patch-level features, without neighbour coupling**
  - without aggregate features,
  - aggregates at single scale, or multiple scales



- **Large-scale aggregates are most informative**
  - especially image-wide ones – c.f. topic models for image classification
  - including additional aggregate scales improves results slightly

Full CRF Model

- **Effect of including pairwise CRF potential**
  - **IND:** no coupling, **CRFσ:** Potts, **CRFτ:** contrast Potts, **CRFγ:** class based
  - local features only (**red**); including global aggregate (**black**)

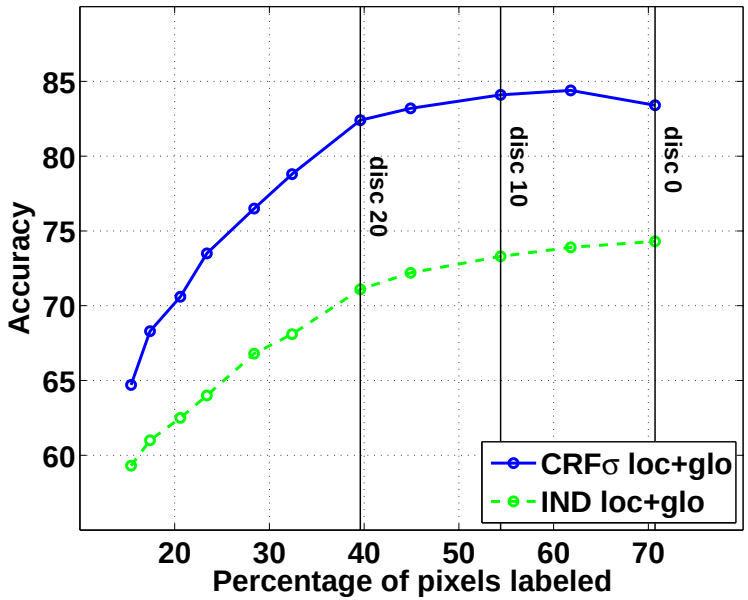


  - **[1]** Schroff et al. ICVGIP'06: optimized aggregation window, no coupling
  - **[2]** our PLSA-MRF model CVPR'07: generative, cross-validation for σ

- **Both random field spatial coupling and image-wide context are useful**
- **Exact choice of pairwise potential is less important**

Amount of Labeling

- **Recognition as a function of the amount of labeling**
  - Decimate training labels using morphological erosion filters of increasing size



- **Good performance with CRF when** 30–60% **of labels are missing**
- **Applying small erosion improves the model – due to label errors**

Summary

- **Good CRFs can be learned from partially labelled training images**
  - use model to marginalize over possible label completions
  - good segmentation results even with 60% missing labels
  - works even if label transitions are completely unobserved
- **Including aggregate features significantly improves performance**
  - image-wide aggregates are the most informative
  - adding finer aggregates gives a modest improvement
- **Pairwise potential is important**
  - different form yield comparable performance