Decomposed Learning for Joint Object Segmentation and Categorization

Yi-Hsuan Tsai
ytsai2@ucmerced.edu
Jimei Yang
jyang44@ucmerced.edu
Ming-Hsuan Yang
mhyang@ucmerced.edu

Electrical Engineering and Computer Science
University of California
Merced, USA

We present a learning algorithm for joint object segmentation and categorization that decomposes the original problem into two sub-tasks and admits their bidirectional interaction. In the first stage, in order to decompose the output space, we train category-specific segmentation models to generate multiple figure-ground hypotheses. In the second stage, by taking advantage of object figure-ground information, we train a multi-class segment-based categorization model to determine the object class. A re-ranking strategy is then applied to classified segments to obtain the final category-level segmentation results (see Figure 1).

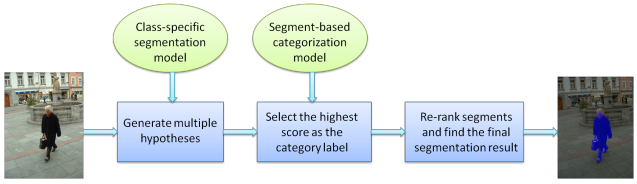

Figure 1: Overview of the algorithm.

Category-specific segmentation. The first step is to train the two-class pylon models [5] based on the segmentation tree generated by the gPb method [1]. Suppose the image I can be partitioned into hierarchical regions $\mathbf{S} = \{S_1, S_2, ..., S_{2N-1}\}$ and a figure-ground label, $f_i = 1$ or 2, respectively is assigned to each region, we formulate the conventional CRF energy function:

$$E(\mathbf{f}) = \sum_{i=1}^{2N-1} U(f_i) + \sum_{(i,j)\in\mathcal{N}} V(f_i, f_j), \tag{1}$$

where $U(f_i)$ indicates the cost of assigning f_i to the segment S_i and $V(f_i, f_j)$ is the smoothness term of the boundary cost for two neighboring segments S_i and S_j. Furthermore, we define the unary energy:

$$U(f_i) = \begin{cases} |S_i| \cdot <\mathbf{w}_1, \mathbf{h}(\mathrm{S_i})>, & \text{for } f_i = 1, \\ |S_i| \cdot <\mathbf{w}_2, \mathbf{h}(\mathrm{S_i})>, & \text{for } f_i = 2, \end{cases} \tag{2}$$

where $\mathbf{h}(S_i)$ denotes the feature vectors and the weighting factor $|S_i|$ is the size of the segment, which encourages the model to prefer larger regions.

In the testing stage, with the learned energy function for $E(\mathbf{f})$, instead of finding only the MAP solution, we introduce parametric min cut [4] into our unary function in Equation 1:

$$U(f_i, \lambda) = \begin{cases} U(f_i) + |S_i| \cdot \lambda, & \text{for } f_i = 1, \\ U(f_i) - |S_i| \cdot \lambda, & \text{for } f_i = 2. \end{cases} \tag{3}$$

Different values of λ provide our model a bias to generate parametrized results, so we can adjust the hyperplane $\mathbf{w}$ and generate multiple segmentation hypotheses by solving a series of graph cuts.

Segment-based categorization. Given an image, we can apply each category-specific model to obtain a set of segmentation hypotheses $B_i, i = 1, 2, ..., K$ for K classes. We first divide our hypothesis set into positive and negative bags, denoted by B_i^+ and B_i^- respectively. The positive bag consists of the hypotheses generated with the positive segmentation classifier w_i^+, and likewise negative bags contain examples from negative segmentation classifiers w_i^-.

For learning the categorization model, we use standard SVM model and select the best segmentation among a positive bag and the ground truth segmentation as positive samples x^+. In the meanwhile, we use all negative samples x^- from all the negative bags to reduce the chances of false positives (see Figure 2).

Given a test image, a bag of segmentation hypotheses from each segmentation model is generated as the training process. We then first determine the image category label by selecting the highest classification score.

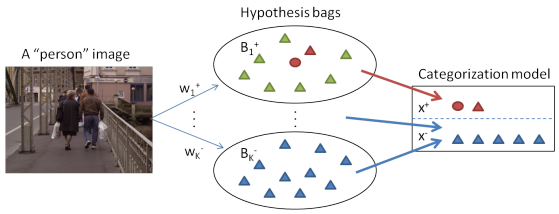

Figure 2: Training for object categorization. The red circle is the ground truth and the red triangle is the best segmentation hypothesis. All samples from B_i^- are negative samples, denoted as blue triangles.

To produce the final segmentation result, we re-rank all the hypotheses in the predicted category bag. The ranking process can be carried out by the class-wise Support Vector Regression and their scores measured by the overlapping ratio between the segment and the ground truth.

Our algorithm enjoys bidirectional interactions between segmentation and categorization. In the segmentation phase, category information facilitates breaking down the multi-class segmentation problem into class-wise sub-problems such that high-quality figure-ground separation can be generated in a reduced labeling space. In the categorization phase, segmentation information helps identifying object locations, shapes as well as context, and hence objects can be precisely represented in the feature space and improve the categorization performance. For concreteness, we demonstrate the merits of the proposed algorithm on the Graz-02 and Caltech 101 data sets.

Table 1: Graz-02 segmentation results using intersection/union overlap metric.

Method	Background	Bicycle	Car	Person	mean
[7]	82.32	46.18	36.49	38.99	50.99
[3]	77.97	55.60	41.51	37.26	53.08
Proposed	91.20	64.95	59.60	60.49	69.06

Table 2: Part of results for Caltech-101 classification. MFea denotes multiple features. Geo denotes the geometric information.

	Method	30 training
MFea + Geo	Gu et al. [2]	77.7
	SvrSegm [6]	82.3
	Proposed	84.2 ± 0.3

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 33(5):898–916, 2011. ISSN 0162-8828.

[2] C. Gu, J. Lim, P. Arbelaez, and J. Malik. Recognition using regions. In *CVPR*, 2009.

[3] A. Jain, L. Zappella, P. McClure, and R. Vidal. Visual dictionary learning for joint object categorization and segmentation. In *ECCV*, 2012.

[4] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. In *ICCV*, 2007.

[5] V. Lempitsky, A. Vedaldi, and A. Zisserman. A pylon model for semantic segmentation. In *NIPS*, 2011.

[6] F. Li, J. Carreira, and C. Sminchisescu. Object recognition as ranking holistic figure-ground hypotheses. In *CVPR*, 2010.

[7] D. Singaraju and R. Vidal. Using global bag of features models in random fields for joint categorization and segmentation of objects. In *CVPR*, 2011.