

---

# Communal Cuts: sharing cuts across images

---

Evan Shelhamer, Stefanie Jegelka, Trevor Darrell  
UC Berkeley

## Abstract

Identifying objects that are common to a set of images is an important step in unsupervised image analysis. While existing approaches to object extraction and co-segmentation commonly focus merely on joint per-class appearance models, we propose a model that incorporates additional label and boundary co-occurrence information with little additional communication across images. This model admits efficient inference and can improve the accurate extraction of difficult, fine-structured objects.

## 1 Introduction

Understanding the content of collections of weakly-labeled, partially-labeled, or unlabeled images remains an important problem in computer vision. A substantial part of this challenge is to identify and extract objects that are simultaneously present in many images. When supervision becomes weak or unreliable, an invaluable (or perhaps the only) source of information are joint patterns and underlying structure that are shared across images. Indeed, common approaches to co-segmentation and object discovery identify joint, cross-image appearance models for each object category, but do not model boundary characteristics. While these shared models capture the coarse essentials of each object class, they necessarily leave uncertainty about detailed variations within a single image. As a result, the assignment of fine-grained structures to object or background remains challenging. In addition, many approaches trade off scalability with spatial accuracy, and tolerate coarser results. Objects such as bikes or trees appear to be particularly affected by these drawbacks [6, 7, 13].

In this work, we aim to decrease this uncertainty about fine structures by incorporating higher-order information about typical label co-occurrences, or characteristics of object boundaries. Specifically, we use submodular functions to formulate a model that shares such co-occurrence information with surprisingly little communication across images. We show an efficient, parallelizable inference algorithm and preliminary experimental results. Moreover, our model is compatible with a number of different approaches to learning joint category-wise appearance models.

**Related work.** Early approaches to co-segmentation process two images at a time [12, 3, 16]. Subsequent work used user interaction [1, 11, 13] or focused on efficiently finding commonalities in sets of images without supervision [8, 14], a problem similar to that of weakly supervised object detection. The approaches range from graph cuts [3, 16, 1, 11] to submodularity [8] to spectral clustering [7]. All of these approaches focus on obtaining better foreground models.

**Notation and setup.** Formally, we observe a collection  $\mathcal{I}$  of  $m$  images, where each image  $I \in \mathcal{I}$  is a collection of  $n_I$  pixels  $z_{Ii}$ . We aim to infer a discrete label (category)  $x_{Ii} \in \mathcal{L}$  for each  $z_{Ii}$ . Each label  $\ell \in \mathcal{L}$  defines an object class and indexes a distinct foreground (or background) model with parameters  $\theta_\ell$ . For example, the likelihood  $p(z_i|x_i = \ell; \Theta)$  may be given by a Gaussian mixture model with parameters  $\theta_\ell$ . We aim to find a MAP estimate of  $\mathbf{x}$ , i.e., a labeling of all images that (approximately) maximizes  $p(\mathbf{x}|\mathbf{z}, \Theta)$ . For ease of exposition, we will here focus on the case of binary labels, but point out that the approach easily generalizes to multiple labels.

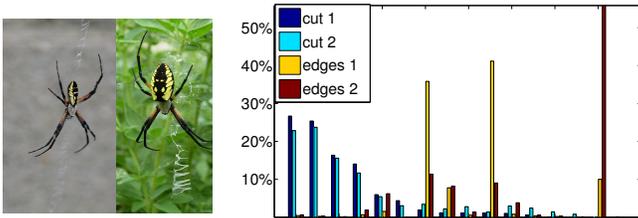


Figure 1: Edge type histograms for the segmentations of the two spiders (16 edge types), and for all edges in the two images. The correct cut consists of only a small set of shared edge types while many other edge types exist in the images. This information can be used as cue to infer correct boundaries.

A frequently used approach to label a single image  $I$  is by formulating a Markov or Conditional Random Field. In this case, the posterior probability  $p(\mathbf{x}_I | \mathbf{z}_I; \Theta) \propto \exp(-E_I(\mathbf{x}_I; \mathbf{z}_I; \Theta))$  factorizes:

$$E_I(\mathbf{x}_I; \mathbf{z}_I, \Theta) = \sum_{i \in I} \psi_i(x_i; z_i, \Theta) + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j), \quad (1)$$

where  $p(z_i | x_i; \Theta) \propto \exp(-\psi_i(x_i; z_i, \Theta))$ . The pairwise terms  $\psi_{ij}$  are defined via a (grid-structured) neighborhood graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  and may depend on  $z$ . To simplify notation, we mostly drop the arguments  $\Theta, \mathbf{z}$  from  $E$ . The standard model (1) easily generalizes to a collection of images:

$$E_{\mathcal{I}}(\mathbf{x}; \mathbf{z}, \Theta) = \sum_{I \in \mathcal{I}} E_I(\mathbf{x}_I; \mathbf{z}_I, \Theta). \quad (2)$$

In this collective model, we assume the foreground models  $\Theta$  to be shared across images. The pairwise terms  $\psi_{ij}$  are restricted to pixel pairs from the same image. Hence, conditioned on the parameters  $\Theta$ , the observed images are independent, the function  $E$  decomposes into image-wise terms and we can find the MAP estimate  $\hat{\mathbf{x}}_I \in \operatorname{argmax}_{\mathbf{x}} p_I(\mathbf{x} | \mathbf{z}; \Theta)$  for each image independently.

The pairwise potentials  $\psi_{ij}$  are fully-compatible with globally shared foreground models and induce an often necessary spatial smoothness in the labeling, but they have well-known drawbacks. Modeling the prior belief that object boundaries are short – a characteristic that holds for compact objects but fails for fine-structured objects such as bikes or trees – they can lead to coarse segmentations and to short-cutting object boundaries. This effect is even more pronounced when no foreground model fits the observation well, i.e.,  $p(z_i | x_i = \ell, \Theta)$  is low for all  $\ell \in \mathcal{L}$ , as it may be the case when labeling collections of images with weak supervision and shared compact appearance models.

To better understand the mechanisms behind this “shrinking bias”, a mathematically equivalent formulation of  $E_I(x)$  as a graph cut is instructive. Let  $\mathcal{G}_I$  be the local neighborhood graph of image  $I$ , with one node  $v_i$  for each  $x_i$ . Then it is possible to find weights  $w : \mathcal{E} \rightarrow \mathbb{R}$  such that

$$E_I(\mathbf{x}; \mathbf{z}) = \sum_{i \in I} \psi_i(x_i) + \sum_{e \in \operatorname{Cut}(\mathbf{x})} w(e) + \operatorname{const}. \quad (3)$$

This is because any labeling  $\mathbf{x} \in \{0, 1\}^n$  of the nodes induces a partition of  $\mathcal{V}$  and a cut  $\operatorname{Cut}(\mathbf{x}) = \{(v_i, v_j) \mid x_i = 1, x_j = 0\} \subseteq \mathcal{E}$  in  $\mathcal{G}$ . If the pairwise potentials  $\psi_{ij}$  satisfy  $\psi_{ij}(0, 0) + \psi_{ij}(1, 1) \leq \psi_{ij}(0, 1) + \psi_{ij}(1, 0)$ , then  $w(e) \geq 0$  for all edges  $e$  and the MAP labeling that minimizes  $E(\mathbf{x}; \mathbf{z})$  defines a minimum cut. The equivalence (3) shows that  $E_I$  penalizes the cut weight, i.e., the length of the object boundary, weighted by the contrast across the boundary. Fine-structured objects do not have a short boundary and will therefore have low probability under this model. Different approaches have addressed this shrinking bias [15, 10, 4], mostly via higher-order potentials.

**Co-occurring label transitions.** We build our model on the observation that in many images the foreground is not the only form of joint structure. The transition between labels, i.e., the co-occurrence of two different labels at the object boundary, is not arbitrary. It is often very homogeneous in appearance, and this homogeneity can serve as a valuable cue. Figure 1 illustrates this observation. We clustered the edges in both images jointly into *types* of similar edges. Similarity between two edges was defined by the distance of their feature vectors (e.g., the RGB gradient  $z_i - z_j$  for edge  $e = (v_i, v_j)$ ). The histogram shows that (1) the true cuts in the two images use only few and very similar types of edges – they are *sparse* in the types; (2) the cuts make up a large fraction of the edges of their type; (3) the statistics of the cut edges (blue bars) are very different from the statistics of all edges in the two images (red and yellow bars).

Jegelka and Bilmes [4] showed that this homogeneity cue can greatly benefit the segmentation of single images. Building on their model, we exploit commonalities of label co-occurrences *across* images. While we address the segmentation of foreground objects here, our joint model may have wider applicability for modeling coherent label co-occurrences across data instances.

## 2 Communal Cuts

Figure 1 suggests a model that prefers label transitions (object or segmentation boundaries) to occur at edges (pixel pairs) of the same type. Like [4], we encourage this by introducing a dependence between edges and penalizing the number of *types* in the cut. The edge dependence arises from a submodular function over sets of edges in the same neighborhood graph  $\mathcal{G} = \bigcup_I \mathcal{G}_I$  as above. A function  $F : 2^{\mathcal{E}}$  is *submodular* if it satisfies that for all sets  $A \subseteq B \subset \mathcal{E}$  and  $e \notin B$ , it holds that  $F(B \cup e) - F(B) \leq F(A \cup e) - F(A)$ . If this *diminishing marginal costs* property is restricted to a given edge type, then the penalty is reduced if the cut consists of many edges of this type. Since the cost sharing of types occurs across images, we name this model *Communal Cuts*.

Formally, let  $\mathcal{E} = \bigcup_g E_g$  be partitioned into  $k$  edge types  $E_g$  that are shared across images. We replace the sum of pairwise potentials in Eqn. (2) (or, equivalently, the graph cut in Eqn. (3)) by submodular functions  $F_g$  over pairwise potentials, one for each type  $g$ :

$$E_c(\mathbf{x}) = \left( \sum_{I \in \mathcal{I}, i \in I} \psi_{Ii}(\mathbf{x}_{Ii}) \right) + \sum_{g=1}^k F_g(\text{Cut}(\mathbf{x}_{I_1}) \cup \text{Cut}(\mathbf{x}_{I_2}) \cup \dots \cup \text{Cut}(\mathbf{x}_{I_m})), \quad (4)$$

where  $F_g$  is a submodular function restricted to  $E_g$ . We use functions of the form

$$F_g(A) = h_g\left(\sum_{e \in A \cap E_g} w(e)\right), \quad (5)$$

where  $h_g$  is a nonnegative, increasing concave function. The graph cut version corresponds to using the identity function  $h_g(y) = y$ . Hence, we could equivalently write

$$E_c(x) = \left( \sum_{I \in \mathcal{I}, i \in I} \psi_{Ii}(x_{Ii}) \right) + \sum_{g=1}^k h_g\left(\sum_{I \in \mathcal{I}} \sum_{(i,j) \in \mathcal{E}_I} \psi_{ij}(x_{Ii}, x_{Ij})\right). \quad (6)$$

The potential (4) introduces dependencies between edges (pairs of variables) across images. As opposed to  $E_{\mathcal{I}}$ , conditioned on  $\Theta$ , this function does not decompose across images. We handle this dependence via a variational coordinate descent approach and exploit that the information that needs to be communicated between different images is restricted to a small set of statistics.

### 2.1 Conditioning preserves model type

We begin by inspecting the conditional distribution  $p(\mathbf{x}_I | \mathbf{x}_{\mathcal{I} \setminus I}, \mathbf{z}, \Theta)$  for the labels of any single image  $I$ . Let  $F(A|B) = F(A \cup B) - F(B)$  denote the marginal cost of a set  $A$  with respect to  $B$  and a submodular function  $F$ . We have by Bayes' rule that

$$\begin{aligned} p(\mathbf{x}_I | \mathbf{x}_{\mathcal{I} \setminus I}, \mathbf{z}, \Theta) &= \frac{p(\mathbf{x}_{\mathcal{I}} | \mathbf{z}, \Theta)}{p(\mathbf{x}_{\mathcal{I} \setminus I} | \mathbf{z}, \Theta)} = \frac{p(\mathbf{x}_{\mathcal{I}} | \mathbf{z}, \Theta)}{\sum_{\mathbf{x}_I \in \{0,1\}^{n_I}} p(\mathbf{x}_{\mathcal{I} \setminus I}, \mathbf{x}_I | \mathbf{z}, \Theta)} \\ &= \frac{\exp\left\{\Psi_u(\mathbf{x}_{\mathcal{I}}) + \sum_g F_g\left(\bigcup_{J \in \mathcal{I}} \text{Cut}(\mathbf{x}_J)\right)\right\}}{\exp\left\{\Psi_u(\mathbf{x}_{\mathcal{I} \setminus I}) + F_g\left(\bigcup_{J \in \mathcal{I} \setminus I} \text{Cut}(\mathbf{x}_J)\right)\right\} \sum_{\mathbf{x}_I} \exp\left\{\Psi_u(\mathbf{x}_I) + \sum_g F_g(\text{Cut}(\mathbf{x}_I) | \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I}))\right\}} \\ &= \frac{\exp\left\{\Psi_u(\mathbf{x}_I) + \sum_g F_g(\text{Cut}(\mathbf{x}_I) | \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I}))\right\}}{\sum_{\mathbf{x}_I} \exp\left\{\Psi_u(\mathbf{x}_I) + \sum_g F_g(\text{Cut}(\mathbf{x}_I) | \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I}))\right\}}. \end{aligned} \quad (8)$$

In this derivation, we used the shorthand  $\mathbf{x}_{\mathcal{I} \setminus I}$  for the set of all variables  $x_i$  with  $i \in \mathcal{I} \setminus I$ , the joint cut  $\text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I}) = \bigcup_{J \in \mathcal{I} \setminus I} \text{Cut}(\mathbf{x}_J)$ , and  $\Psi_u(\mathbf{x}_I) = \sum_{i \in I} \psi_{Ii}(x_{Ii})$ . Since the marginal cost  $F(A|B)$  is a submodular function in  $A$ , the conditional distribution is a cooperative cut model as defined in [4], and of the same type as the full posterior.

Moreover, we observe that the marginal cost

$$F_g(\text{Cut}(\mathbf{x}_I) | \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I})) = h_g\left(\sum_{e \in \text{Cut}(\mathbf{x}_I)} w(e) + \sum_{e' \in \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I})} w(e')\right) - h_g\left(\sum_{e' \in \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I})} w(e')\right)$$

only depends on the scalar statistic  $s_{g,I} = \sum_{e' \in \text{Cut}(\mathbf{x}_{\mathcal{I} \setminus I})} w(e')$ . Knowing the weights of the given image  $I$ ,  $s_{g,I}$  can easily be computed from  $s_g = \sum_{e \in \text{Cut}(\mathbf{x}_{\mathcal{I}})} w(e)$ . This means that to compute the conditional distribution  $p(\mathbf{x}_I | \mathbf{x}_{\mathcal{I} \setminus I}, \mathbf{z}, \Theta)$  we only need to exchange  $\Theta$  and the  $k$  statistics  $s_{g,I}$  between images. Figure 2 illustrates this observation.

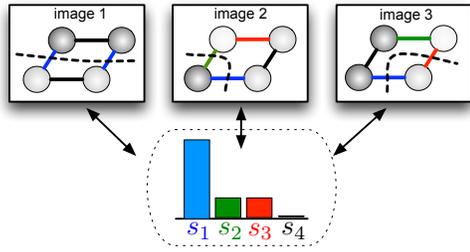


Figure 2: Sharing of the statistics  $s_g$  across images is the only required communication across images, apart from the foreground models  $\Theta$ . Edge types  $E_g$  are indicated by different colors (blue, green, red, black).

**Appearance models.** We learn joint models of the foreground classes and a separate background model for each image. The appearance models are learned via linear support vector machine (SVM) classifiers on color and kernel descriptors [2] on neighborhood patches. For a completely unsupervised approach, we initialize with CoSand [8]; but other unsupervised methods are also suitable.

**Defining edge types.** Partitioning the edges in all images jointly into edge types  $E_g$  is a large clustering problem. For efficiency at scale and robustness to local differences, the clustering is carried out hierarchically: edges in each image are first over-clustered, then the local edge types are unified by a global joint clustering of the per-image centroids, and finally the joint types are propagated back to the local edges through the global clusters. This scalable approach allows for differing distributions of edge types in which some types may not be present in every image.

### 3 Inference

For inference, we iteratively update  $\mathbf{x}$  via MAP inference and then update  $\Theta$  given the updated  $\mathbf{x}$ . MAP inference for  $\mathbf{x}$  is hard in general [5]. If  $k$  is small and  $h_g$  is piecewise linear with few breakpoints, one may use the exact method in [9]. Here, we describe a more general approximation.

The assignments  $\mathbf{x}_I$  can be updated in parallel or sequentially. In both cases, we approximate  $p(\mathbf{x}|\mathbf{z}; \Theta)$  by approximating the functions  $F_g$  that induce dependencies across images. Once  $\mathbf{x}$  has been updated, the approximations  $\hat{F}_g$  are adjusted to be tight at the current assignment. Replacing  $F_g$  by  $\hat{F}_g$  in  $E(\mathbf{x})$  results in an upper bound  $\hat{E}(\mathbf{x}) \geq E(\mathbf{x})$  that satisfies  $\hat{E}(\mathbf{x}^t) = E(\mathbf{x}^t)$  at the current labeling  $\mathbf{x}^t$ . Hence, minimizing  $\hat{E}(\mathbf{x})$  must lead to progress or convergence. The approximations rely on the following approximation of a submodular function.

**Lemma 3.1** ([4]). *Let  $F$  be monotone submodular and  $F \geq 0$ . Then*

$$\hat{F}(A|B) \triangleq F(B) + \sum_{e \in A \setminus B} F(e|B) - \sum_{e \in B \setminus A} F(e|\mathcal{E} \setminus e) \geq F(A)$$

for all  $A \subseteq \mathcal{E}$ , and  $\hat{F}(B|B) = F(B)$ .

Given any current labeling  $\mathbf{y}$ , we use Lemma 3.1 to define upper bounds

$$\hat{E}(\mathbf{x}; \mathbf{y}) = \Psi_u(\mathbf{x}) + \sum_{g=1}^k \hat{F}_g(\text{Cut}(\mathbf{x}) | \text{Cut}(\mathbf{y})) \geq E(\mathbf{x}). \quad (9)$$

**Parallel (Jacobi) updates.** At closer inspection, we see that  $\hat{F}$  in Lemma 3.1 is a sum of weights:

**Lemma 3.2.** *Using the approximation in Lemma 3.1, it holds that  $\hat{F}_g(C; \text{Cut}(\mathbf{y})) = \sum_{e \in C} \nu_g(e; \mathbf{y}) + \text{const}$ , where*

$$\nu_g(e; \mathbf{y}) = \begin{cases} F(e | \text{Cut}(\mathbf{y})) & \text{if } e \notin \text{Cut}(\mathbf{y}) \\ F(e | \mathcal{E} \setminus e) & \text{otherwise.} \end{cases}$$

The weights  $\nu_g(e; \mathbf{y})$  can be computed from the summary statistic  $s_g$  of  $\mathbf{y}$  instead of  $\mathbf{y}$ .

The last statement follows from the observations in Section 2.1 and implies that all we need to share across images are the  $s_g$ . We may hence write

$$\widehat{E}(\mathbf{x}; \{s_g\}) \triangleq \widehat{E}(\mathbf{x}; \mathbf{y}) = \Psi_u(\mathbf{x}) + \sum_{e \in \text{Cut}(x)} \sum_{g=1}^k \nu_g(e; s_g) + \text{const.} \quad (10)$$

The function  $\widehat{E}(\mathbf{x}; \{s_g\})$  is a pairwise potential like the CRF model (3) and decomposes across images. Hence, we can compute  $\mathbf{x}^{t+1} = \text{argmax}_{\mathbf{x}} \widehat{E}(\mathbf{x}; \mathbf{x}^t)$  by updating each  $\mathbf{x}_I$  separately in parallel via a minimum cut in the image graph  $\mathcal{G}_I$  with weights  $\nu_g$  restricted to  $\mathcal{G}_I$ . To compute  $\widehat{E}$  in the next step, we only need to update the statistics  $s_g$ .

**Sequential (Gauss-Seidel) updates.** Instead of updating the image-wise labels  $\mathbf{x}_I$  in parallel, we can update them sequentially. After updating each  $\mathbf{x}_I$ , we already update the approximation of the energy too. In this case, we may slightly tighten the edge weights to

$$\nu_g(e; \mathbf{y}) = \begin{cases} F(e | \text{Cut}(\mathbf{y})) & \text{if } e \notin \text{Cut}(\mathbf{y}) \\ F(e | \mathcal{E}_I \setminus e) & \text{otherwise.} \end{cases}$$

In fact, this slightly changed approximation corresponds to approximating the posterior  $p(\mathbf{x}_I | \mathbf{x}_{\mathcal{I} \setminus I}, z, \Theta)$  given in Equation (8). Algorithm 1 shows the sequential updates; this algorithm is a block coordinate descent. We alternate Algorithm 1 with updating  $\Theta$ .

---

**Algorithm 1** CommunalBCD( $\mathcal{G}, F$ )

---

**Input:**  $\mathcal{G}, F$   
**Output** labelings  $\{\mathbf{x}_I\}_{I \in \mathcal{I}}$   
 set  $\mathbf{x}^t = 0, s_g^0 = 0$  and compute  $w(\mathcal{E}_g)$  for all  $g$  and  $t = 0$   
**for**  $I \in \mathcal{I}$  **do**  
   compute  $\mathbf{x}_I^1 \in \text{argmin}_{\mathbf{x} \in \{0,1\}^{n_I}} \widehat{E}(\mathbf{x}_I; \{s_g^0\})$   
**end for**  
**repeat**  
    $t = t + 1$   
   **for**  $j = 1$  to  $|\mathcal{I}|$  **do**  
     compute  $s_g^t = \sum_{i < j} w(\text{Cut}(\mathbf{x}_{I_i}^t)) + \sum_{i \geq j} w(\text{Cut}(\mathbf{x}_{I_i}^{t-1})) \quad \forall 1 \leq g \leq k$   
     compute  $\mathbf{x}_{I_j}^t \in \text{argmin}_{\mathbf{x}_{I_j}} \widehat{E}(\mathbf{x}_{I_j}; \{s_g^t\})$   
   **end for**  
**until**  $E(\mathbf{x}^t) \geq E(\mathbf{x}^{t-1})$

---

## 4 Experiments

We evaluate Communal Cuts on a variation of the MSRC data set, which we augment with finer, more accurate ground truth labels and call MSRC-fine<sup>1</sup>. We show preliminary results for an unsupervised figure-ground segmentation task on image pairs from the same class. Twenty random pairs are drawn from each class and the initial unsupervised segmentation is established by CoSand [8]. If a degenerate initialization produces whole-image segments the pairs are discarded from evaluation. The model parameters (coefficient for  $\psi_{ij}$  and a thresholding parameter of  $h_g$  as in [4]) were chosen by cross-validation. Our results differ from those on MSRC in the literature due to the new fine-structured annotations.

Table 3 shows results for (i) CoSand (shared foregrounds, no iteration, superpixels), (ii) learning a joint foreground model and using Cooperative Cuts [4] on each image separately (joint foreground, iterations, no interaction of the cut terms across images), and (iii) Communal Cuts. The table includes two numbers: the intersection-over-union metric, and the same metric only evaluated at a thin band around the correct boundary. We observe that including edge interactions (per image (ii) or across images (iii)) does improve the accuracy of the segmentations compared to the initial CoSand results. In particular for fine-structured objects such as trees and bikes it helps to share cut statistics across images. Only in some cases, if the images are fairly different, edge sharing may be less appropriate.

<sup>1</sup>For Communal Cuts code and MSRC-fine data, see <http://coopcut.berkeleyvision.org>.

Class	CommCut sharing across	CommCut sharing within	CoSand
Aeroplane	17.36 / 38.77	<b>18.14 / 38.85</b>	14.85 / 30.10
Bike	<b>26.31 / 34.29</b>	23.55 / 29.17	16.91 / 21.70
Bird	26.01 / <b>32.95</b>	<b>26.21 / 32.28</b>	13.64 / 22.65
Boat	<b>10.92 / 33.79</b>	10.48 / 29.78	4.63 / 26.35
Car	<b>26.60 / 37.97</b>	26.44 / 34.83	18.79 / 26.88
Cat	<b>18.18 / 29.60</b>	17.20 / 25.79	17.66 / 20.35
Chair	17.82 / 28.58	<b>18.30 / 29.50</b>	14.88 / 23.63
Cow	28.71 / 33.93	<b>33.39 / 35.47</b>	22.34 / 20.90
Dog	24.39 / 28.47	<b>26.09 / 30.06</b>	19.41 / 17.72
Face	12.44 / 27.57	13.05 / <b>24.62</b>	<b>15.84 / 24.38</b>
Flowers	60.35 / 50.36	<b>65.19 / 51.68</b>	43.48 / <b>52.33</b>
Sheep	26.06 / 25.87	<b>32.02 / 26.24</b>	23.11 / 13.14
Sign	17.15 / 27.25	18.14 / 27.51	<b>28.02 / 29.69</b>
Tree	<b>15.01 / 35.18</b>	14.91 / 32.90	3.60 / 22.29

Figure 3: Average per-class accuracies of figure-ground segmentations on MSRC-fine: intersection-over-union metric percentage on the full image (left) and on a band around the ground truth border (right). Communal Cuts help capture the details of the difficult bicycle, tree, and chair classes. Figure: example visual results. CoSand misses parts of the cow’s ear or the entire object. It also excludes the (uncommon) black parts of the flowers. Boundary interactions remedy these effects.

## References

- [1] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. Interactively cosegmentating topically related images with intelligent scribble guidance. *IJCV*, 2011.
- [2] L. Bo, X. Ren, and D. Fox. Kernel descriptors for visual recognition. In *NIPS*, 2010.
- [3] D. Hochbaum and V. Singh. An efficient algorithm for co-segmentation. In *ICCV*, 2009.
- [4] S. Jegelka and J. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *CVPR*, 2011.
- [5] S. Jegelka and J. Bilmes. Graph cuts with interacting edge costs – examples, approximations, and algorithms. *CoRR*, abs/1402.0240, 2014. URL <http://arxiv.org/abs/1402.0240>.
- [6] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *CVPR*, 2010.
- [7] A. Joulin, F. Bach, and J. Ponce. Multi-class cosegmentation. In *CVPR*, 2012.
- [8] G. Kim, E.P. Xing, L. Fei-Fei, and T. Kanade. Distributed cosegmentation via submodular optimization on anisotropic diffusion. In *ICCV*, 2011.
- [9] P. Kohli, A. Osokin, and S. Jegelka. A principled deep random field for image segmentation. In *CVPR*, 2013.
- [10] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *ICCV*, 2009.
- [11] L. Mukherjee, V. Singh, and J. Peng. Scale-invariant cosegmentation for image groups. In *CVPR*, 2011.
- [12] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching incorporating a global constraint into MRFs. In *CVPR*, 2006.
- [13] M. Rubinstein, C. Liu, and W.T. Freeman. Annotation propagation in large image databases via dense image correspondences. In *ECCV*, 2012.
- [14] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu. Unsupervised joint object discovery and segmentation in internet images. In *CVPR*, 2013.
- [15] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *CVPR*, 2008.
- [16] S. Vicente, V. Kolmogorov, and C. Rother. Cosegmentation revisited: Modes and optimization. In *ECCV*, 2010.