Exemplar Cut

Jimei Yang, Yi-Hsuan Tsai and Ming-Hsuan Yang
University of California, Merced
5200 North Lake Road, Merced CA
{jyang44, ytsai2, mhyang}@ucmerced.edu

Abstract

We present a hybrid parametric and nonparametric algorithm, exemplar cut, for generating class-specific object segmentation hypotheses. For the parametric part, we train a pylon model on a hierarchical region tree as the energy function for segmentation. For the nonparametric part, we match the input image with each exemplar by using regions to obtain a score which augments the energy function from the pylon model. Our method thus generates a set of highly plausible segmentation hypotheses by solving a series of exemplar augmented graph cuts. Experimental results on the Graz and PASCAL datasets show that the proposed algorithm achieves favorable segmentation performance against the state-of-the-art methods in terms of visual quality and accuracy.

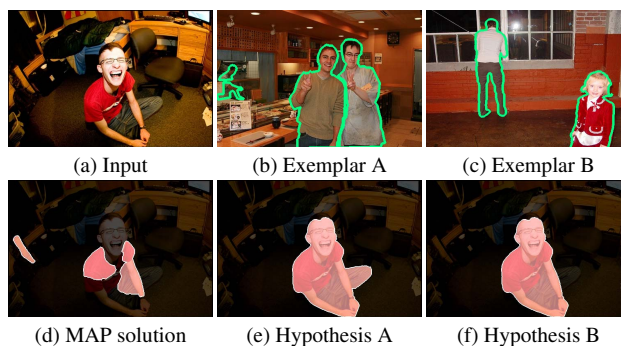| (a) Input | (b) Exemplar A | (c) Exemplar B |
| (d) MAP solution | (e) Hypothesis A | (f) Hypothesis B |

Figure 1. Generating class-specific segmentation hypotheses from exemplars (person in this example). (a) input image. (b) one exemplar image. (c) another exemplar image. (d) MAP solution for person segmentation. (e) exemplar cut using exemplar A. (f) exemplar cut using exemplar B.

1. Introduction

Category level object segmentation is one of the core problems in computer vision. Its main challenges lie in that small visual elements (pixels or superpixels) contain insufficient information that admits category level object recognition. One line of research aims at effectively propagating high level recognition results back to low level segmentation through superpixel neighborhood [10], high-order Conditional Random Fields (CRFs) [18] or object detector outputs [1, 27]. Another line makes efforts to generate object segmentation hypotheses so that recognition can be achieved more efficiently by classification or ranking [20].

Object segmentation hypotheses could be category independent or category specific. Recent work for category independent object segmentation [8, 6, 13] exploit hierarchical image segmentations, grouping strategies and cross-category shape priors in order to increase the chance of recovering true object regions. As a result, such methods are likely to generate thousands of instance-level object region hypotheses which entail laborious post-processing to filter out low-quality solutions. Category specific approaches instead, [5, 17, 4] generate single object segmentation by us-

ing efficient maximum a posteriori (MAP) inference tools (e.g., graph cut [15]), which perform well when target objects appear dominantly in the images with simple backgrounds. In real-world applications, however, target objects more often appear in cluttered backgrounds with large appearance variations and interact with the objects of other categories (e.g., PASCAL VOC datasets [9]). In these cases, the single MAP solution becomes less satisfactory (Figure 1(d)) due to the limited model capacity and training errors. A natural choice to resolve this issue is to generate multiple object segmentation hypotheses from class-specific models [3, 12] (Figure 1(e)(f)). This choice not only benefits from learning but also increases the probability of finding all the target objects.

In this paper, we propose a hybrid parametric and nonparametric model for generating a small set of highly plausible class-specific object segmentations, thereby reducing ambiguities and computational loads for sequential classification or ranking. Towards that, we first learn a pylon model [19] to obtain the parametric object segmentation energy function. Building on a bottom-up hierarchical segmentation [2], the pylon model combines a flat CRF with a region tree. The resulting energy function remains

submodular and admits efficient inference by graph cut, which brings conveniences to max-margin learning. Second, we match the test image with each exemplar by regions. For each region in the test image, we retrieve k nearest neighbors (K-NN) from the matching exemplar, so that the node potentials of the pylon model are augmented by K-NN matching scores. Therefore, an object segmentation hypothesis can be generated by solving a graph cut with the exemplar augmented energy function, which we refer as *exemplar cut*.

Our method leverages both the generalizability of parametric models and the flexibility of nonparametric models. Parametric models usually make assumptions on image segmentations. For example, CRFs and pylons assume that regions are classifiable in the node potentials, and labels between adjacent regions are consistent up to the Potts pairwise potentials. Under these assumptions, the MAP inference usually produces reasonably smooth labeling around the target (Figure 1(d)). The reason of missing some parts and predicting a false negative lies in that the node classifiers are less effective in handling heterogeneous appearance in complex background.

On the other hand, the nonparametric segmentations [24, 21, 16, 23] are more flexible to model assumptions. These methods are able to segment an image by transferring prior knowledge (e.g., labels and shape masks) from retrieved exemplars or regions in a database of segmentation exemplars. However, considering the statistical instability of using exemplars, challenges arise from integrating the retrieved or matched segmentation results into a single solution. Our method avoids such issue and instead queries each exemplar to generate one segmentation hypothesis. By adjusting the pylon energy function by the exemplar matching score, we fuse the parametric and nonparametric classifiers [7] on the node potentials and still take advantage of the label consistency assumption and learned parameters on the pairwise potentials. Consequently, we increase the possibilities of correcting the mistakes of parametric models and prevent segmentation from noisy labeling.

We carry out experiments on the Graz-02 [22] and PASCAL VOC 2010 datasets [9]. We use the intersection/union overlap scores [9] to evaluate the upper bound performance of segmentation hypotheses. The results show that the proposed exemplar cut algorithm generates better segmentation hypotheses than the MAP solution and performs favorably against the state-of-the-art methods based on parametric min cut [14], diverse M-Best solutions [3] and multiple choice learning [12]. We also analyze the performance of hypotheses at different MAP quality levels. The results on the Graz-02 dataset suggest that exemplar cut maintains high recall rates when MAP solutions miss the target objects.

2. Related Work

As the focus of this work is generating good hypotheses for class-specific segmentation, we discuss the most related work in three aspects.

Parametric Min Cut. The parametric min cut algorithm [14] introduces a constant value to the node potentials of the graph cut energy function, which changes the decision threshold of classifying the nodes into foreground and background. By varying the constant value, a series of graph cuts are solved to produce a set of segmentation hypotheses. This technique has been used in [6, 13] for category independent object segmentation hypotheses. As the classification thresholds are changed uniformly for all the nodes, the parametric min cut usually produces noisy segmentation results where good segments are companied by false negatives. In contrast, our exemplar cut adaptively determines the decision boundary by K-NN matching scores with exemplar regions.

M-Best Solutions. When the single MAP solution becomes less satisfactory, it is beneficial to find M best solutions. A potential issue is that the top M most probable solutions may be similar to each other if many noisy local minimal solutions exist close to the MAP one [28]. To address that, Batra *et al*. [3] propose to explore different local modes of the energy function by enforcing the solution diversity. In their work, the energy function is augmented with dissimilarity constraints that isolate the current solution from previous ones by a pre-defined threshold. This strategy entails a greedy algorithm to find solutions sequentially. In contrast to the dissimilarity metric that operates as a repulsive force to push the current solution away from existing ones, the matching similarity in our work performs as an attractive force that pulls the current solution towards the exemplars.

Multiple Choice Learning. This approach aims to generate multiple structured outputs [12] by learning a set of sub-models simultaneously, rather than inferring multiple solutions from a single model. Recall that we need a diverse set of segmentation hypotheses. It is thus essential to enforce sub-models as different as possible. In fact, the multiple choice learning approach realizes this objective in the training phase by discriminative clustering. It assigns the training exemplars to sub-models by evaluating their segmentation errors so that a sub-model is eventually optimized towards a subset of exemplars. This approach is constrained by the clustering structure of training exemplars and sensitive to the initialization. When the number of sub-models is not properly chosen, the segmentation capabilities of learned sub-models may be imbalanced (some too strong and others too weak) so that the weak predictor degenerates in the training phase. In addition, since each sub-model governs a set of exemplars, we can also perform exemplar cut to each sub-model of multiple choice learning.

3. Exemplar Cut

In this section, we present the proposed exemplar cut algorithm for class-specific object segmentation in details. We first introduce the underlying segmentation model and then present our approach of generating multiple segmentation hypotheses with exemplars.

3.1. Category Specific Object Segmentation

Pylon Model. We use the two-class pylon model [19] as the underlying mechanism for category specific object segmentation. In Figure 2, we parse an image with a simplified two-class pylon model in a segmentation tree.

Figure 2. A two-class pylon model illustration. We parse an image by a segmentation tree. Each node represents a segment at a different level and its figure/ground assignment energy is given by $U(f_i)$ in (3). The edges between the leaf nodes $V(f_i, f_j)$ in (4) denotes the pairwise smoothness term. The edge between a segment and its ancestor represents the consistency constraint $C(f_i, f_{a(i)})$ in (1). A labeling of the image is visualized by colored bounding boxes around the segments. The red box indicates the figure label ($f_i = 1$) while the blue box indicates the ground label ($f_i = 2$). The dashed box indicates the segment is not being used to explain the image ($f_i = 0$) and all the solid boxes constitute a complete image.

We first segment an image I into a hierarchical region tree $\mathbf{S} = \{S_1, S_2, ..., S_{2L-1}\}$ by the *gPb* contour detector [2]. We index the leaf segments from 1 to L, the intermediate segments from $L + 1$ to $2L - 2$ and the root segment (the entire image) as the last one, $2L - 1$. We also denote $a(i)$ as the ancestor of segment i and $p(i, j)$ as the shortest path from segment i to segment j. Note that segment i and its ancestor $a(i)$ are overlapped, so we only need to keep one of them to explain the image. Each segment $S_i \in \mathbf{S}$ thus could be assigned a label $f_i \in \{0, 1, 2\}$, where $f_i = 1$ indicates the foreground, $f_i = 2$ the background,

and $f_i = 0$ not being used for explaining the image. To produce a consistent labeling $\mathbf{f} = \{f_1, ..., f_{2L-1}\}$, the pylon model requires that for any leaf segment, there is only one non-zero label along its path to the root node in the tree,

$$\forall i = 1, ..., L, \forall j \in p(i, 2L-1), f_j > 0 \wedge f_j \cdot f_{a(j)} = 0. \quad (1)$$

This constraint guarantees the complete and non-overlapping labeling.

We formulate an energy function for the pylon model similar to a conventional CRF,

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

where the unary term $U(f_i)$ specifies the cost of assigning a label f_i for the segment i, and the pairwise term $V(f_i, f_j)$ instantiates the non-negative boundary cost between any two adjacent segments $(i, j) \in \mathcal{N}$. The set of adjacent segments is denoted by $\mathcal{N}$. In particular, we define the unary term as linear models,

$$U(f_i) = \begin{cases} |S_i| \cdot <\mathbf{w}_1, \mathbf{h}(S_i)>, & \text{for } f_i = 1, \\ |S_i| \cdot <\mathbf{w}_2, \mathbf{h}(S_i)>, & \text{for } f_i = 2, \\ 0, & \text{for } f_i = 0, \end{cases} \quad (3)$$

where $\mathbf{h}(S_i)$ denotes the feature vector extracted from the segment S_i and $\mathbf{w}_1, \mathbf{w}_2$ denotes the unary parameter vectors. We use the size of segment $|S_i|$ as a weighting factor to encourage the pylon inference to select larger segments. The pairwise term is defined by a weighted Potts model,

$$V(f_i, f_j) = <\mathbf{w}_3, \mathbf{b}(S_i, S_j)> \cdot \delta[f_i \neq f_j], \quad (4)$$

where $\mathbf{b}(S_i, S_j)$ is a vector of exponentiated boundary strength with different bandwidths and $\mathbf{w}_3$ is the smoothness parameter vector.

Inference. It is not trivial to infer the pylon model with both semantic labels $f_i = 1, 2$ and exclusive label $f_i = 0$. In order to leverage the strength of graph cut for optimal inference on binary Markov random fields, Lempitsky *et al.* [19] decompose the target label f_i of each segment into a pair of equivalent binary variables x_i^1 and x_i^2, which indicate whether any of the segments along the path $p(i, 2L - 1)$ falls entirely into the regions of class 1 or 2. Based on this re-parametrization technique, both the unary terms (except the root) and consistency constraint in (1) are absorbed into a new energy function $E(\mathbf{x}^1, \mathbf{x}^2)$ as pairwise terms. With some manipulations, this new energy function becomes submodular and can be minimized by graph cut. To recover the target labeling $\mathbf{f}$ from equivalent labelings $(\mathbf{x}^1, \mathbf{x}^2)$, we can simply examine if its ancestors up the tree are assigned non-zero labels. More details about this re-parametrization method can be found in [19].
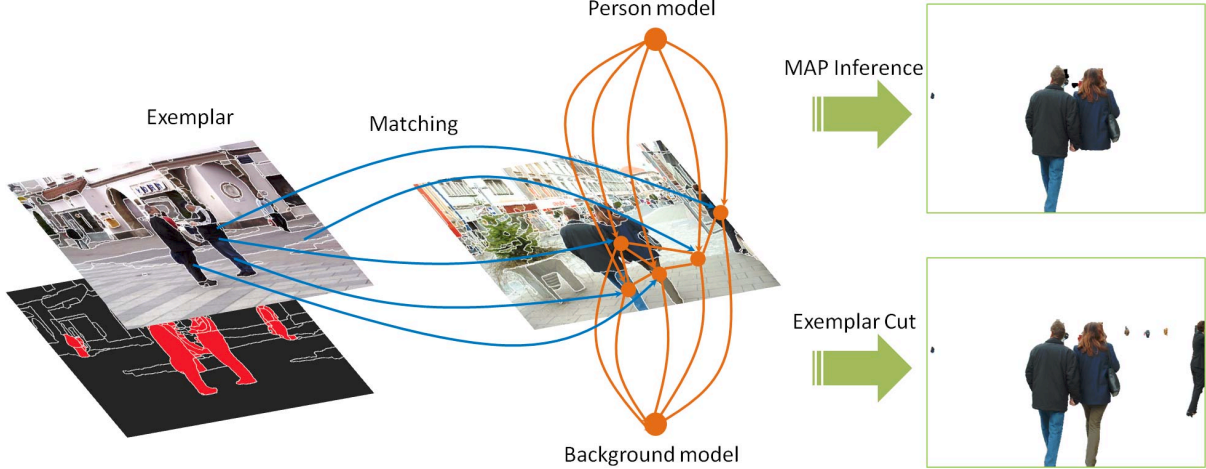
Figure 3. Exemplar cut algorithm. The goal is to segment an input image in the **center** into the foreground object and the background. We construct a pylon graph [19] on the regions generated by [2]. The node energies (orange links) are constructed by the pre-learned person and background models (orange nodes). The pairwise energies are constructed on the adjacent regions by measuring their shared boundaries (white lines in the image) and compatibilities. Note that we omit the hierarchical structure for ease of illustration. On the **left** are an exemplar image (**top**) and its segmentation mask (**bottom**). The exemplar is also segmented into regions by [2] and each region is assigned a label based on the mask. We find the best match (blue links) in the exemplar image for each region of the test image. Based on their labels, we incorporate the node energy of each region with the matching similarities. We solve a graph cut to this augmented energy function for generating an exemplar cut solution, which is shown on the **bottom right**. As the test image shares similar appearance with the exemplar image, the resulting segmentation has a very high accuracy. The original MAP solution on the **top right** instead misses many small targets and one occluded person.

Max-Margin Learning. The pylon model can be learned in a max-margin fashion. The optimization is formulated as,

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{N}\sum_n \xi_n, \quad \text{s.t.} \quad \forall \xi_n \geq 0$$

$$\max_{\mathbf{x}}[\Delta(\mathbf{x}^{(n)},\mathbf{x}) - E(\mathbf{x},\mathbf{y}^{(n)};\mathbf{w}) + E(\mathbf{x}^{(n)},\mathbf{y}^{(n)};\mathbf{w})] \leq \xi_n$$

$$\mathbf{w_3} \geq 0$$

$$(5)$$

where $\mathbf{x}$ denotes all the binary variables $\{\mathbf{x}^1, \mathbf{x}^2\}$, $\mathbf{y}$ represents all the feature vectors $\{\mathbf{h}_i; \mathbf{b}_{ij}\}$ and $\mathbf{w}$ is the full parameter vector $[\mathbf{w}_1; \mathbf{w}_2; \mathbf{w}_3]$. We develop a stochastic gradient (sub-gradient) descent algorithm for efficiently optimizing the max-margin learning objective in (5).

3.2. Hybrid Parametric/Nonparametric Model

Pylon models have been demonstrated to be effective for figure/ground segmentation and semantic scene parsing [19], due to their ability of selecting larger regions in the segmentation tree to reduce the classification ambiguity. When the pylon models are applied to object class segmentation, their MAP solution becomes less effective in dealing with complex object appearance and their interactions. On one hand, the Markov random field assumption about image structures may be invalid for the objects with heterogeneous appearance. For example, the strong edges between different body parts and partial occlusions break the

smoothness assumption while the cluttered background violates the discontinuity assumption around object boundaries. On the other hand, the max-margin objective and the use of slack variables in (5) usually introduce bias to the process of model learning, although they improve the generalizability to unseen images and the resistance to noises.

On the contrary, exemplar based nonparametric approaches, which involve no model assumptions and training process, can generate segmentations by image matching and prior transfer. Although they are flexible of exploiting prior knowledge contained in exemplars, the nonparametric approaches still face the challenge of filtering the matching results due to the large variance of exemplars.

We herein present a hybrid method by integrating the pylon model with exemplars. Figure 3 illustrates the proposed algorithm. To segment an input image I, we first compute its parametric energy from the learned pylon model through (2), which provides a basis of generating smooth segmentations close to the ground truth. The exemplars are then used to generate segmentation hypotheses different from the MAP solution, through region matching. We represent an exemplar by an image $I^{(n)}$ and its segmentation $\mathbf{f}^{(n)}$, and denote the matching energy by $\Delta(\mathbf{f}, \mathbf{f}^{(n)}; I, I^{(n)})$. Therefore, a segmentation hypothesis can be generated by solving graph cut to the augmented energy function,

$$\tilde{E}(\mathbf{f}, \mathbf{f}^{(n)}) = (1-\lambda)U(\mathbf{f}) + \lambda\Delta(\mathbf{f}, \mathbf{f}^{(n)}) + V(\mathbf{f}), \quad (6)$$

where $\lambda \in [0,1]$ controls the tradeoff between the parametric and nonparametric energies. In this exemplar cut energy function (6), we actually resolves the learning bias of pylon model by the matching variance with exemplars.

3.3. K-Nearest Neighbor Region Matching

In this work, we compute the nonparametric matching energy function $\Delta(\mathbf{f}, \mathbf{f}^{(n)})$ by using the K-NN matching algorithm. We parse an exemplar image $I^{(n)}$ into a set of hierarchical regions $\mathbf{S}^{(n)}$. For each region, we extract a feature vector $\mathbf{h}_i^{(n)}$ and assign a ground truth label $f_i^{(n)}$ from its annotated object masks. The exemplar is thus represented by a set of feature-label pairs $\{(\mathbf{h}_i^{(n)}, f_i^{(n)})\}_{i=1,2,\ldots,2L-1}$. Note that we assign non-zero labels to regions that fall entirely into ground truth foreground, which increases the chance of recovering object parts. We assume that segments are independent of each other so that we approximate the matching energy by

$$\Delta(\mathbf{f}, \mathbf{f}^{(n)}) = \sum_{i=1}^{2L-1} \Delta(f_i, \mathbf{f}^{(n)}). \qquad (7)$$

For each region i in the test image, we retrieve K best segments in the exemplar based on their feature similarities,

$$(\mathbf{h}_k^{(n)}, f_k^{(n)})_{k=1,2,\ldots,K} \leftarrow \text{K-NN}(\mathbf{h}_i, \{(\mathbf{h}_j^{(n)}, f_j^{(n)})\}). \qquad (8)$$

We define the matching energy of region i by the K-NN output,

$$\Delta(f_i) = -\frac{1}{K} \sum_{k=1}^{K} <\mathbf{h}_i, \mathbf{h}_k^{(n)}> \cdot \delta[f_i = f_k^{(n)}]. \qquad (9)$$

This matching energy can be easily merged into the parametric unary term (3) so that the augmented energy function (6) remains the same form as the original energy function (2) and can be solved by the inference algorithm developed in [19]. Figure 4 shows the foreground confidence maps of pylon model and K-NN matching, and their complementary effects.

4. Experiments

We present the experimental results using the Graz-02 [22] and PASCAL VOC 2010 datasets [9] with evaluations of the proposed algorithm against several state-of-the-art class-specific methods for generating segmentation hypotheses, i.e., the parametric min cut (PMCut) method, the diverse M Best solutions (MBest) [3] and the multiple choice learning (MCL) approach [12]. We implement these three algorithms based on the learned pylon model. We use the intersection/union overlap scores to evaluate the upper bound performance of these segmentation algorithms.

(a) Input   (b) Exemplar

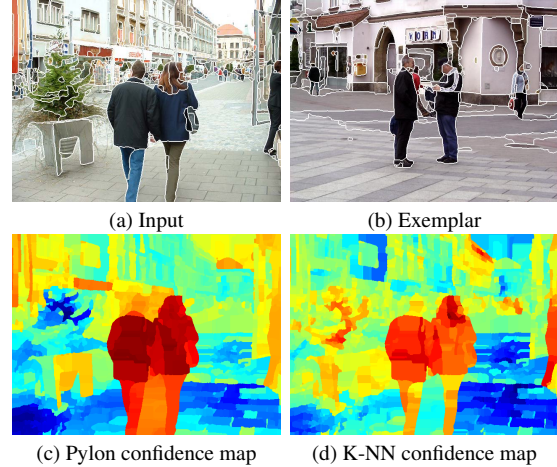(c) Pylon confidence map   (d) K-NN confidence map

Figure 4. Foreground confidence maps of the pylon model and K-NN matching method. The confidence maps are computed from the node energy. The K-NN matching recovers small targets at a distance and a partial figure near the right boundary, which is missed by the pylon model. The pylon model removes the false positives of the K-NN matching method around the tree region.

4.1. Graz-02

Setup. The Graz-02 dataset includes 3 object classes (bike, car and person) and background images, which is challenging for object segmentation due to large pose variations, scale changes and partial occlusions. Each class consists of 300 images $I^{(n)}$ of 640×480 pixels and all the images are annotated by foreground masks. The odd-numbered images per class are used for training pylon models and for exemplars, and the rest for evaluation. We set the parameter $\lambda = 0.9$ as the tradeoff of parametric and nonparametric energy, and the parameter $K = 7$ as the number of nearest neighbors in the K-NN matching method[1].

Representations. We represent an image by a bottom-up segmentation tree using the *gPb* algorithm [2]. The number of segments is around 7000 on average. Each segment S_i is represented by three kinds of features $\mathbf{h}_i = [\mathbf{h}_i^{SIFT}; \mathbf{h}_i^{color}; \mathbf{h}_i^{shape}]$. We extract dense SIFT descriptors using the VLFeat toolbox [25] and train a codebook of size 512 using all the training images of three classes. The SIFT histogram $\mathbf{h}_i^{SIFT}$ is computed using the locality constrained linear coding method [26] and max pooling. The color histogram $\mathbf{h}_i^{color}$ is computed from a color codebook of size 128 by assigning each RGB pixel to its closest codeword. The contour shape descriptor $\mathbf{h}_i^{shape}$ is extracted from a spatial pyramid of oriented *gPb* edge responses [11]. We map the concatenated feature vectors to a high-dimensional space with the explicit χ^2 kernel [25].

Ground truth labeling. To train the pylon models and use exemplars, we need to determine the ground truth labels for

---

[1] We empirically determine these values by evaluating $\lambda \in [0.7, 1.0]$ and $K = [5, 9]$.

each region in a training image, generated by the *gPb* algorithm. Due to the errors from the *gPb* contour detection and blurred boundaries of annotated masks, the average overlap score of the pylon ground truth segmentations with annotated masks is 87.8%.

Results. We use the learned pylon models of each class to generate MAP segmentations of test images. For the diverse MBest approach, we observe that its performance improves slightly when the number of solutions is close to 30, thereby its overlap scores are reported by using 30 solutions per image. For the MCL method, we train up to 5 sub-models for each class and observe some sub-models are underutilized (as no training exemplars belong to some sub-models). For the PMCut, we follow the method in [13] and draw 41 samples by varying the constant parameter from -2.0 to 2.0 with the increment 0.1. By using exemplar cuts, we generate 150 raw segmentation hypotheses per test image. Note that all these methods generate redundant segmentations. We evaluate their performance from raw results without considering post-processing to merge the duplicates and remove the low quality ones. Table 1 shows the overlap scores of segmentation approaches by using the "oracle" evaluation protocol [3]. The exemplar cut algorithm outperforms the MAP solution by 15.3% on average and the second best MCL approach by 5.1%. We present some qualitative results in

Figure 5. Segmentation results on the Graz-02 test set. We present two representative images per class for comparisions. Bikes, cars and people are highlighted by green, blue and red masks, respectively. From left to right, the segmentations are generated by MBest, MCL, PMCut and ExemplarCut. Best viewed in color.

Table 1. "Oracle" overlap scores in the Graz-02 dataset.

overlap	Bicycle	Car	Person	mean
MAP	66.4	61.4	60.6	62.8
MBest	69.8	67.3	66.0	67.7
PMCut	73.9	70.6	70.7	71.7
MCL	73.1	75.1	71.0	73.0
ExemplarCut	**77.4**	**78.9**	**78.0**	**78.1**

Figure 5, and more results can be found in the supplementary material. It shows that the segmentation hypotheses generated by exemplar cut are usually able to deal with occlusions, suppress the background clutters and recover the missing targets.

We generate up to 150 segmentations per test image. Moreover, we are interested in the performance of exemplar cut segmentations when the MAP solutions fail to detect the targets. We consider the hypotheses as true positives if their overlap with ground truth are greater than 50%; otherwise as false positives. In Figure 6, we compare the recall rates of segmentation hypotheses at different MAP quality levels. When the MAP overlap score is smaller than 10%, almost losing the targets, the hypotheses generated by exemplar cut (red curves) achieve recall rates 75.0%, 52.6% and 46.7% for bike, car and person datasets. The overall recall rates are 97.3%, 88.7% and 90.0% for bike, car and person datasets, respectively.
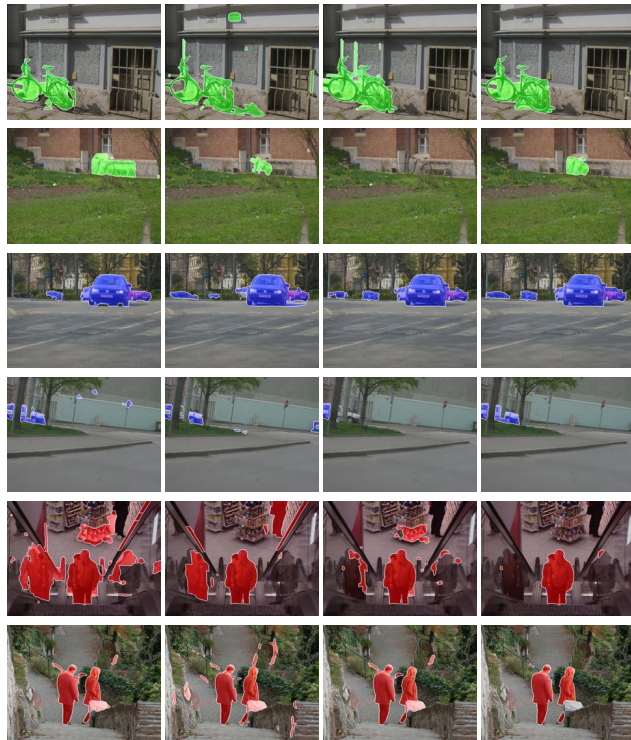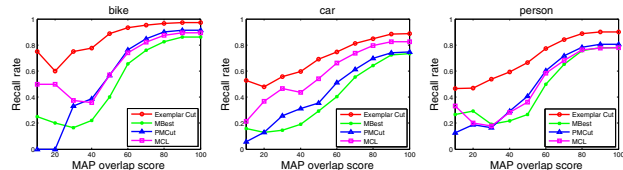

Figure 6. Recall rates at different MAP quality levels on the Graz-02 datasets. From the left to the right, we compare the evaluated four algorithms for bike, car and person, respectively.

4.2. PASCAL VOC 2010

Setup. In the PASCAL VOC 2010 dataset, each of the training, validation and test sets includes 964 annotated images from 20 object categories and one background class. The images in this dataset may include multiple objects from several categories. We use the training set to learn pylon models, and evaluate the proposed algorithm on the validation set. For class-specific object segmentations, we first parse the multi-class segmentation masks into class-specific object masks, and then train a pylon model on the positive exemplars for each class.

Implementations. Similar to the Graz-02 experiments, we represent an image with a bottom-up segmentation tree using the *gPb* method and represent each segment by SIFT,

Table 2. "Oracle" class-wise overlap scores on the VOC 2010 validation set.

	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dtable	dog	horse	motorbike	person	plant	sheep	sofa	train	tvmonitor	average
MAP	52.9	19.1	35.6	39.3	23.3	49.9	43.1	47.4	11.8	56.9	36.8	45.1	50.6	44.3	37.9	20.9	58.1	34.5	48.9	33.1	39.5
MCL	57.9	**29.3**	45.7	36.1	32.8	57.4	50.9	61.0	23.9	68.8	49.7	60.1	58.5	55.1	50.0	20.9	65.4	45.9	61.1	35.7	48.3
PMCut	60.3	24.3	47.9	51.0	39.2	67.1	51.6	66.4	22.7	69.3	50.7	55.9	56.7	55.7	50.3	29.6	**71.6**	45.2	62.7	43.9	51.1
ExemplarCut	**66.8**	28.0	**66.6**	**60.7**	**44.1**	**71.1**	**59.8**	**72.7**	**34.5**	**76.4**	**71.4**	**68.9**	**68.2**	**64.9**	**63.9**	**39.5**	71.0	**62.1**	**70.4**	**55.2**	**60.8**

color and shape features. Considering the large appearance variations, we train a SIFT codebook of size 8096. We use the same method in the Graz-02 experiments to obtain ground truth pylon labelings and use the same values for the parameter λ and K. For exemplar cut of each class, we use the positive training images as exemplars and thus generate about 50 segmentation hypotheses per class on average. We implement the PMCut method by varying the parameter from -2 to 2 by increasing 0.1 each step. For the MCL algorithm, we choose to initialize one predictor per 10 images since each class has different training images. We segment one test image using the models from 20 classes, because we have no information about the object categories before running segmentation algorithms.

Results. We evaluate the quality of segmentation hypothesis sets by "oracle" overlap scores in Table 2. In the top panel, we present the overlap scores for class-specific object segmentation hypotheses. The exemplar cut performs the best on 18 of 20 classes, and achieves 60.8% average overlap score, with an improvement over the second best, PMCut, by 9.7%. To elucidate the quality of our segmentation hypotheses, we present one image per class and compare its exemplar cut and the second best segmentation in Figure 7. In order to see the potentials of class-specific hypotheses for category level object segmentation, we compute the upper bound overall overlap score in Table 3. For a

Table 3. "Oracle" multi-class overlap scores on the VOC 2010 validation set.

	MAP	MBest [3]	MCL	PMCut	ExemplarCut
overlap	44.86	48.0	54.3	58.8	68.7

test image, we select the best segmentation from each class as its category confidence map. The class segmentation is thus given by the maximum confidence score from all the classes at pixels. Intuitively, if there is an oracle segmentation selector, we can achieve 68.7% overlap accuracy on the PASCAL VOC 2010 validation set. This result suggests the potential of exemplar cut for category level object segmentation. Note that the MBest result [3], is computed from a multi-class CRF model [18]. We do not present the MBest result by the pylon model because we find it difficult to tune a single dissimilarity parameter for all the classes.

4.3. Discussions

The number of segmentation hypotheses for exemplar cut equals the number of exemplars (i.e., 50). In both Graz-02 and VOC 2010 datasets, exemplars are limited for each class. Thus it is convenient to use all of them to generate hypotheses. When the number of exemplars is large, it will be infeasible to use all the exemplars. Similar to the algorithms for scene parsing [21, 24], we can also use image retrieval techniques to find the most relevant exemplars for hypothesis generation. The segmentation results in this work may include some isolated false negatives in the background area and duplicate segments from different hypotheses. It will be worthwhile developing segment filters to remove noise and redundancy [6] before using them for classification purposes.

5. Conclusions

We present a novel exemplar cut algorithm for generating class-specific object segmentation hypotheses. It combines a learning based parametric segmentation model and a matching based nonparametric segmentation algorithm in a principled way. Experimental results on the Graz-02 and PASCAL VOC 2010 datasets demonstrate that the proposed exemplar cut algorithm achieves favorable results in terms of visual quality and accuracy. In addition, the results show that the proposed algorithm is especially effective when the MAP approaches fail to generate good segmentation results on images with complicated scenes.

References

[1] P. Arbelaez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik. Semantic segmentation using regions and parts. In *CVPR*, 2012.

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

[3] D. Batra, P. Yadollahpour, A. Guzman-Rivera, and G. Shakhnarovich. Diverse m-best solutions in Markov random fields. In *ECCV*, 2012.
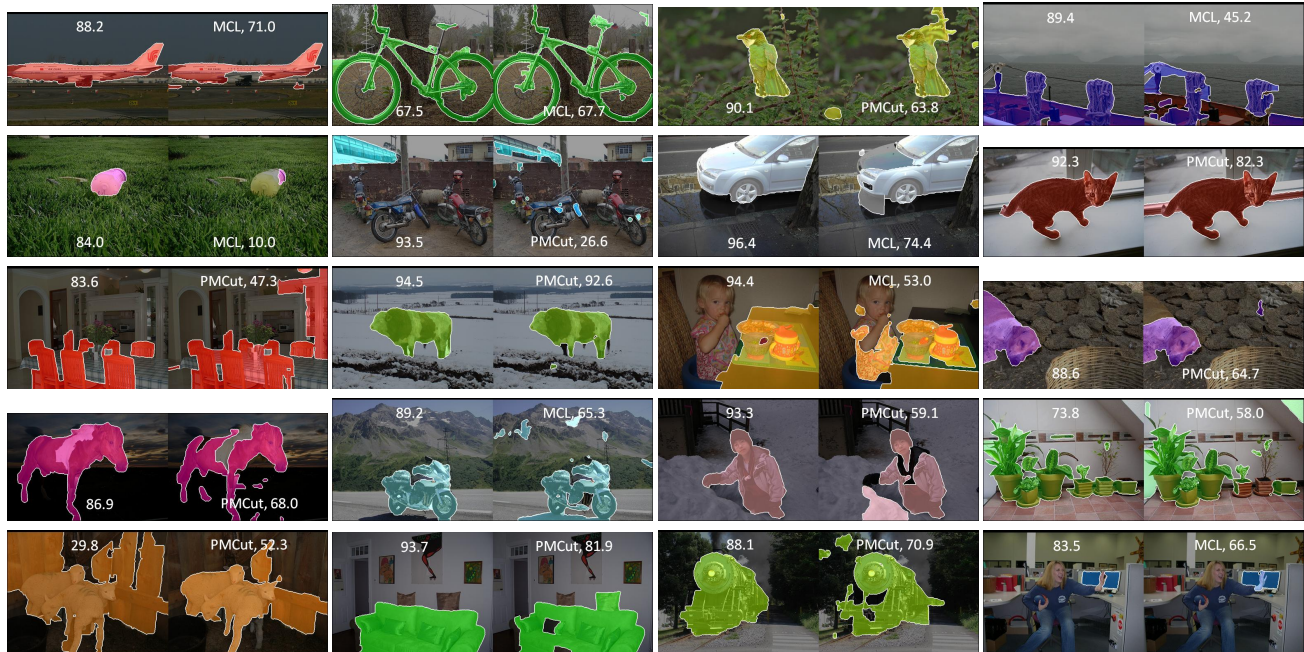
Figure 7. Comparing segmentations for 20 classes on the VOC 2010 evaluation datasets. In each panel, the left image is produced by exemplar cut and the right image is produced by the most competing algorithm (MCL or PMCut). For the bicycle image, we present a case that exemplar cut is slightly worse than MCL. For the sheep image, we present a failure case of exemplar cut where PMCut performs the best. The masks for different classes follow the VOC color codes. Best viewed in color.

[4] L. Bertelli, T. Yu, D. Vu, and B. Gokturk. Kernelized structural svm learning for supervised object segmentation. In *CVPR*, 2011.

[5] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, 2002.

[6] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *CVPR*, 2010.

[7] P. Chaudhuri, A. K. Ghosh, and H. Oja. Classification based on hybridization of parametric and nonparametric classifiers. *PAMI*, 31(7):1153 – 1164, July 2009.

[8] I. Endres and D. Hoiem. Category independent object proposals. In *ECCV*, 2010.

[9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. "http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html".

[10] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *ICCV*, 2009.

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

[12] A. Guzman-Rivera, D. Batra, and P. Kohli. Multiple choice learning: Learning to produce multiple structured outputs. In *NIPS*, 2012.

[13] J. Kim and K. Grauman. Shape sharing for object segmentation. In *ECCV*, 2012.

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

[15] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *PAMI*, 26:65–81, 2004.

[16] D. Kuettel and V. Ferrari. Figure-ground segmentation by transferring window masks. In *CVPR*, 2012.

[17] M. P. Kumar, P. Torr, and A. Zisserman. Obj cut. In *CVPR*, 2005.

[18] L. Ladicky, C. Russell, P. Kohli, and P. Torr. Associative hierarchical crfs for object class image segmentation. In *ICCV*, 2009.

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

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

[21] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing via label transfer. *PAMI*, 33(12):2368–2382, 2011.

[22] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. Generic object recognition with boosting. *PAMI*, 28(3):416–431, 2006.

[23] A. Rosenfeld and D. Weinshall. Extracting foreground masks towards object recognition. In *ICCV*, 2011.

[24] J. Tighe and S. Lazebnik. Superparsing: Scalable nonparametric image parsing with superpixels. In *ECCV*, 2010.

[25] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2010.

[27] W. Xia, Z. Song, J. Feng, L. F. Cheong, and S. Yan. Segmentation over detection by coupled global and local sparse representations. In *ECCV*, 2012.

[28] C. Yanover and Y. Weiss. Finding the m most probable configurations using loopy belief propagation. In *NIPS*, 2003.