Image Cosegmentation via Multi-task Learning

Qiang Zhang, Jiayu Zhou, Yilin Wang, Jieping Ye, Baoxin Li
Computer Science and Engineering
Arizona State Uni., Tempe, AZ, USA
qzhang53,jiayu.zhou,ywang370,jieping.ye,baoxin.li@asu.edu

Abstract

Image segmentation has been studied in computer vision for many years and yet it remains a challenging task. One major difficulty arises from the diversity of the foreground, which often results in ambiguity of background-foreground separation, especially when prior knowledge is missing. To overcome this difficulty, cosegmentation methods were proposed, where a set of images sharing some common foreground objects are segmented simultaneously. Different models have been employed for exploring such a prior of common foreground. In this paper, we propose to formulate the image cosegmentaion problem using a multi-task learning framework, where segmentation of each image is viewed as one task and the prior of shared foreground is modeled via the intrinsic relatedness among the tasks. Compared with other existing methods, the proposed approach is able to simultaneously segment more than two images with relatively low computational cost. The proposed formulation, with three different embodiments, is evaluated on two benchmark datasets, the CMU iCoseg dataset and the MSRC dataset, with comparison to leading existing methods. Experimental results demonstrate the effectiveness of the proposed method.

1 Introduction

Though having been studied over the past few decades with a lot of algorithms being proposed, e.g., normalized cut [31], graph cut [11], image segmentation still remains a challenging vision task. One key difficulty arises from the ambiguity between the foreground object and the background, especially when no proper prior is given. To alleviate such ambiguity, different methods have been developed in the past. For example, [26] used human input to improve the segmentation; [23] used nonlinear shape prior to improve the segmentation. Recently, image cosegmentation [27] was proposed, where multiple images sharing the common foreground object are segmented simultaneously. The prior that a common object is shared among the images is helpful for reducing the ambiguity in the segmentation task.

Since [27], numerous cosegmentation algorithms have been proposed. [33] extended MRF for segmenting a pair of images. Co-segmenting more than two images was discussed in [16, 20]. In [17, 19] a more challenging problem called multiple foreground segmentation problem, where multiple foreground objects are presented in the images, was explored. [21, 28] proposed weakly supervised methods, where image annotation was utilized to facilitate the segmentation.

There are several limitations in the existing methods. The approach of [33] can segment only two images. Though the method of [16] is able to segment more than two images simultaneously, its complexity is too high for more than a couple of images: it took 8 minutes for segmenting two images but 6 hours for eight images. Image annotation is required in [21]. SIFT-flow was used in [29] to find the correspondences for each pair of images, which can be costly (no time information was available for that method) considering that there would be a lot of pairs of images for a big set.

Considering those limitations, in this work we propose to address the image cosegmetantion problem using the multi-task learning framework. In the proposed method, segmentation of each image is viewed as one task and multiple tasks are solved simultaneously. Each task finds a classifier segmenting the foreground from the background in an image, while the prior that the images share the common foreground object is captured by the intrinsic relatedness among the tasks. To model this relatedness (or prior), we evaluate and compare three different schemes, considering shared features cross tasks or similarity of the classifiers.

The contributions of this paper are twofold. First, we formulate the image cosegmentation problem via a multi-task learning framework, resulting in a systematic way of modeling the important prior of shared foreground among multiple images. Second, we implement and analyze three different embodiments of the multi-task learning formulation, demonstrating the flexibility of the formulation in addressing the image cosegmentation task. These contributions are supported by comparative experiments where the proposed method is shown to outperform, in terms of both accuracy and speed, other leading cosegmentation methods on two benchmark datasets: the CMU iCoseg dataset and the MSRC dataset.

In the rest of the paper, we first review some related work in Sec. 2; the proposed method is presented in Sec. 3; Sec. 4 shows the experiment results of the proposed method with comparisons to the existing methods; and the paper concludes with a discussion in Sec. 5.

2 Related Work

In this section, we will review some related works in the literature.

Image Cosegmentation: many different algorithms have been proposed for image cosegmentation. In [27], an extension of MRF was proposed for cosegmenting a pair of images, where histogram matching was utilized. [33] compared three different histogram matching methods. However, those methods are limited to the context of a pair of image. Later, [16] proposed discriminative clustering for cosegmenting more than two images. However it suffers from high computational cost, e.g., it took 8 minutes for segmenting two images but 6 hours for eight images. [20] proposed another method for segmenting more than two images. While earlier works focused on segmenting images which shares a unique foreground object, a more challenging version called multiple foreground segmentation problem was addressed in [17, 19], where multiple foreground objects are presented in the images. To improve the accuracy of segmentation, supervisory information was also considered, e.g., in [21, 28], where image annotation was considered to facilitate cosegmetation. This idea was further explored in [29], where the image annotation is automatically discovered from Internet.

Visual Saliency/Cosaliency: visual saliency, which predicts regions in the field of view that draw the most visual attention, has attracted a lot of interest from researchers. One early work that is widely known was proposed in [13]. Since then, a lot of different models have been proposed for computing visual saliency, e.g., graph-based visual saliency [12] and PCA saliency [24]. A recent survey was reported in [5]. Visual saliency has been

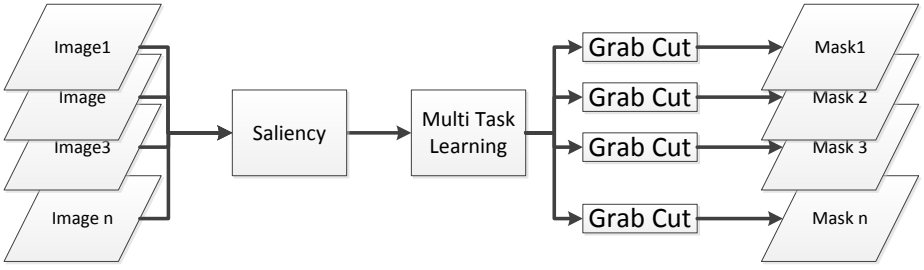

Figure 1: The overview of the proposed algorithm. We first extract the saliency map for the input images; according to the saliency map, we pick the seed regions to initialize the multi-task learning algorithm; and finally according to the output of multi-task learning algorithm, we use grab cut to obtain the final segmentation result.

used in other vision tasks including image segmentation, where visual saliency is used to initialize the foreground. For image cosegmentation, recognizing the salient object should be common in the set of images, image cosaliency has been proposed, which considers both the repeatedness and saliency of the patches. In [7], image cosaliency was considered for a pair of images. Image cosaliency for more than two images was proposed in [6].

Multi-task Learning: Multi-task learning aims at improving the generalization performance of a set of related machine learning tasks (e.g., classification and regression), leveraging the intrinsic relatedness of these tasks. Multi-task learning has been an active topic in machine learning and has found successful application in various areas especially in computer vision [36, 37, 38] and bioinformatics [4]. There are many approaches in multi-task learning for modeling the task relatedness. Regularization-based models are among the most popular since they can easily incorporate existing computational models and various assumptions of task relatedness. In [9], the model parameters from different tasks were assumed to be drawn from the same distribution and thus close to the mean. When the feature dimension is high, typical assumptions include a shared subspace [7] and a shared set of features across tasks [22]. While many multi-task learning approaches focus on learning a shared representation across all tasks, there are efforts on finding clustered structures among tasks [14, 41] and learning task relationship [39].

3 Proposed Method

In this section, we present the proposed method, an overview of which is illustrated in Fig 1. In experiments of this paper, we first over-segment the images into superpixels and then use them as basic units for subsequence processing. For obtaining the superpixels, we use SLIC [1] and set the number of superpixels for each image to 200. For notations, we use $\mathbf{X}_i^j$ to represent the descriptor of the j_{th} superpixel in i_{th} image and y_i^j as its label. In the following subsections, we will elaborate each component of the proposed method.

Feature	Normalization	Dim.	$d(\mathbf{x},\mathbf{y})$
Average RGB values	divided by 255	3	$\|\mathbf{x}-\mathbf{y}\|_2^2$
Average LAB values	divided by 100	3	$\|\mathbf{x}-\mathbf{y}\|_2^2$
Absolute responses of LM filter	untouched	15	$\|\mathbf{x}-\mathbf{y}\|_2^2$
LAB histogram	normalized to unit sum	2048	$\chi^2(\mathbf{x},\mathbf{y})$
Hue/saturation histogram	normalized to unit sum	16	$\chi^2(\mathbf{x},\mathbf{y})$
Texton histogram	normalized to unit sum	65	$\chi^2(\mathbf{x},\mathbf{y})$
Center of superpixel	divided by size of image	2	$\|\mathbf{x}-\mathbf{y}\|_2^2$
Bounding box of superpixel	divided by size of image	4	$\|\mathbf{x}-\mathbf{y}\|_2^2$
Aspect ratio of bounding box	untouched	1	$\|\mathbf{x}-\mathbf{y}\|_2^2$
Area of superpixel	divided by area of image	1	$\|\mathbf{x}-\mathbf{y}\|_2^2$

Table 1: The feature used in this paper, which is designed according to [15]. The normalization process aims to keep the data within the range of $[0,1]$. The 3^{rd} column shows the dimensions of the feature and the fourth column shows the distance functions, where $\chi^2(\mathbf{x},\mathbf{y}) = \sum_i \frac{2(\mathbf{x}_i-\mathbf{y}_i)^2}{(\mathbf{x}_i+\mathbf{y}_i)}$.

3.1 Feature Extraction

For each superpixel, we extract the feature according to [15], which includes geometry measurements, color, texture and edges. The list of features is shown in Tab. 1.

The similarity measure of the superpixels is one of the most important component for image segmentation. For image cosegmentation, we need not only the similarities measure of the superpixels within each image, but also the similarities measure of superpixels cross different images. The similarities of the superpixels within each image are computed as:

$$A(i,j;p,q) = K(\mathbf{X}_i^j,\mathbf{X}_p^q) \times e^{\frac{-|loc(\mathbf{X}_i^j)-loc(\mathbf{X}_p^q)|_2^2}{2\sigma}} \text{ if } i = p \tag{1}$$

where $A(i,j;p,q)$ measures similarity between $\mathbf{X}_i^j$ and $\mathbf{X}_p^q$, $\mathrm{loc}(\cdot)$ is the location of the mass center of the superpixels and $K(\cdot,\cdot)$ is the kernel function. This measurement assign high similarity score to superpixels which are both spatially close and feature-wise similar.

However, the similarities of superpixels cross different images cannot be computed in this way, as $e^{\frac{-|loc(\mathbf{X}_i^j)-loc(\mathbf{X}_p^q)|_2^2}{2\sigma}}$ is no longer meaningful for $i \neq p$. In [29], SIFT-flow is applied to find the correspondences of pixels/superpixels between each pair of images and the correspondences are used as the measurement of geometrical closeness. This method would be extremely costly for set with many images and would become unstable when the differences of the images are significant, e.g., when the foreground objects of different images vary too much in scale. We propose the following approach for measuring the similarity between superpixels of inter-images. For each superpixel in i_{th} image, we find top k most similar superpixels in all other images $p \neq i$, i.e.,

$$A(i,j;p,q) = K(\mathbf{X}_i^j,\mathbf{X}_p^q) \times \mathrm{KNN}(i,j;p,q) \text{ if } i \neq p \tag{2}$$

where $\mathrm{KNN}(i,j;p,q) = 1$ if superpixel $\mathbf{X}_p^q$ belongs to the top k most similar superpixels with $\mathbf{X}_i^j$; otherwise $\mathrm{KNN}(i,j;p,q) = 0$. In our experiment, we found that this scheme was able to deliver good performance.

The kernel function $K(\cdot,\cdot)$ is defined as $K(x,y) = e^{-\frac{d(x,y)}{\eta^2}}$, where $d(x,y)$ is the sum of distances of all features as described in Tab. 1 Col. 4 and η is the expected mean of $d(x,y)$.

3.2 Visual Cosaliency

Recently visual saliency has been used as initialization of several image segmentation algorithms. In [6], visual cosaliency was proposed and utilized to initialize the image coegmentation algorithm. However, as [6] computed co-saliency=saliency×repeatedness, it is not robust to outliers, e.g., the shared foreground object (region with high repeatedness) may be marked as non-salient by the single-view salient model for some images, then the cosaliency map for those image would be wrong. In the proposed cosaliency, a superpixel is cosalient, if it is not only salient in the corresponding image but also similar to salient superpixels of other images. More formally it can be written as:

$$s_i^j(t+1) = (1-\alpha)\hat{s}_i^j + \alpha \sum_{p,q:s_p^q(t)\geq\tau} s_p^q(t) \times A(i,j;p,q) \tag{3}$$

where $s_i^j(t+1)$ is the cosaliency score of $\mathbf{X}_i^j$ in Iteration $t+1$, $\hat{s}$ is the saliency computed from each image individually (for which we use PCA saliency in [24]), $A(i,j;p,q)$ is the similarity between superpixel $\mathbf{X}_i^j$ and $\mathbf{X}_p^q$, and τ is the threshold defining the salient superpixels (e.g., mean of saliency score of superpixels in all images).

The first part of Eqn. 3 computes the saliency for each individual image and the second part measures the similarity to the salient region of other images. α is the weight for combining those two terms. We dynamically set α according to the similarities of the salient regions of the images: for images which have similar salient regions, α will be set to a large value; otherwise, a small value will be used. More formally, we have $\alpha = \frac{\hat{\alpha}}{1+e^{\frac{\mu-\bar{s}}{\beta}}}$, where $\hat{\alpha}$ is the upper bound for α, $\bar{s}$ is the similarity of the salient region of the images via PCA saliency, μ and β are two constants. In our experiment, we set $\hat{\alpha} = 0.85$, $\mu = 0.52$ and $\beta = 0.02$.

We will repeat the iteration of Eqn. 3 until s_i^j does not change. In our experiments, we observed that typically less than 100 iterations were required before the iteration converges. After computing the cosaliency score s_i^j, we label the top 20% of the salient superpixel as the foreground and the bottom 70% ones as background (similar as what done in [18]) to initialize the image cosegmentation algorithm, which will be described below.

3.3 Multi-task Learning

As described earlier, we formulate image cosegmentation as a multi-task learning problem, where the segmentation of each image is viewed as one task. Naturally, in this formulation the prior that common foreground objects are shared among the images is assumed to be captured by the intrinsic relatedness among the tasks.

For developing a solution under this formulation, we focus on regularization-based modeling, due to its flexibility in incorporating existing computational models and supporting various assumptions of task relatedness. Mathematically, the problem is formulated as the following regularization-based multi-task learning problem:

$$\{\mathbf{W}_i\} : \arg\min_{\{\mathbf{W}_i\}} f(\{\mathbf{W}_i\}) + g(\mathbf{W}) \tag{4}$$

where $\mathbf{W}_i$ is the classifier that segments foreground object from the background in i_{th} image (and we use linear classifier in this paper), $\{\mathbf{W}_i\}$ for the set of images and $\{\mathbf{X}_i^j, \mathbf{y}_i^j\}$ for the descriptors and the labels of the superpixels in i_{th} image. The regularization term $g(.)$ encodes a specific assumption on how the tasks are related. We collectively represent the classifier parameters in a matrix form $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_N]$, where N is the number of classifiers (images).

In $f(\{\mathbf{W}_i\}) = \sum_i f_i(\mathbf{W}_i, \{\mathbf{X}_i^j, \mathbf{y}_i^j\})$, we require the classifier $\mathbf{W}_i$ both correctly classify the labeled data and also be smooth in the manifold which embeds the data. Accordingly:

$$f(\mathbf{W}_i, \{\mathbf{X}_i^j, \mathbf{y}_i^j\}) = \sum_j l(\mathbf{W}_i^T \mathbf{X}_i^j + b_i, y_i^j) + \sum_{p \neq q} |(\mathbf{W}_i^T \mathbf{X}_i^p + b_i) - (\mathbf{W}_i^T \mathbf{X}_i^q + b_i)| d(\mathbf{X}_i^p, \mathbf{X}_i^q) \quad (5)$$

where $l(x,y) = \frac{1}{1+e^{-y \times x}}$ is the logistic loss and b_i is the bias of the linear classifier. By using the Laplacian matrix, we can rewrite Eqn. 5 as

$$f(\mathbf{W}_i, \{\mathbf{X}_i^j, \mathbf{y}_i^j\}) = \sum_i l(\mathbf{W}_i^T \mathbf{X}_i^j + b_i, y_i^j) + \mathbf{W}_i^T \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T \mathbf{W}_i$$

where $\mathbf{L}_i$ is the Laplacian matrix built according to the similarity measure of superpixels as described in Sec. 3.1. In this work, since we consider foreground/background segmentation, accordingly we have $y_i^j \in \{-1, 0, 1\}$, where 1 means foreground, -1 for background and 0 for unavailable.

For $g(\mathbf{W})$, we consider the following three multi-task learning assumptions:

1) the task parameters are drawn from the same distribution (short as "mean"). Since cosegmentation is supposed to capture a common foreground object among a set of images, it is reasonable to assume that the parameters of the classifiers are drawn from the same distribution, and thus the parameters are close to the mean value $\sum_i \mathbf{W}_i$, leading to

$$\{\mathbf{W}_i\} : \arg\min_{\{\mathbf{W}_i\}} f(\{\mathbf{W}_i\}) + \lambda \|\mathbf{W}_i - 1/N \sum_{j=1}^{N} \mathbf{W}_j\|_2^2. \quad (6)$$

2) the models of the tasks share a common low-rank subspace (short as "low"). When the images are too different from each other, assuming the models are from the same distribution is sometimes too restrictive. An alternative approach is to assume that the models are from the same low-dimensional subspace, which means the model vectors $\mathbf{W}_i$ are formed by linear combinations of a few shared basis vectors. In terms of model matrix $\mathbf{W}$, we expect a low-rank structure. To encourage a low-rank structure on the model matrix, we penalize the convex surrogate of the rank function – trace norm of $\mathbf{W}$, which gives us

$$\{\mathbf{W}_i\} : \arg\min_{\{\mathbf{W}_i\}} f(\{\mathbf{W}_i\}) + \lambda \|\mathbf{W}\|_* \quad (7)$$

where $\|\mathbf{W}\|_* = \sum_{i=1}^{\min(d,N)} \sigma_i(W)$ is the trace norm of $\mathbf{W}$. The limitation of the low-rank subspace assumption is that when either we have few images (N) or the feature dimension (d) is very small i.e., $\min(d,N)$ is small, the low-rank assumption is of little use because the upper bound of the rank of $\mathbf{W}$ is already small.

3) the models of the tasks share the same small subset of features (short as "$\ell_{2,1}$"). For different images, depending on the content of the image, chances are that the prediction models might be very different from each other. However, since they are for segmentation

of the same object, the relevant features may be similar. The group Lasso technique can be used to enable joint feature selection across tasks, giving us

$$\{\mathbf{W}_i\} : \arg\min_{\{\mathbf{W}_i\}} f(\{\mathbf{W}_i\}) + \lambda \|\mathbf{W}\|_{2,1} \tag{8}$$

Optimization. The optimization problem in Eqn.6 is smooth and thus can be solved efficiently by solving the gradient equation. In Eqn.7 and Eqn.8, however, the regularization terms are non-smooth. Accelerated projected gradient can be used to solve them. In this paper we use the implementation in the multi-task learning package MALSAR [40]. As the three problems are all convex, thus the convergence to global optimal would be obtained.

After we find the classifiers with the multi-task learning methods, we apply the classifiers to each superpixel of the images. The classifiers return a score within $[0,1]$ via logistic function. We then label a superpixel as background if its response is smaller than 80% of the mean response of all of the superpixels of all images, which will be used for initialization of Grabcut algorithm [26] to obtain the final segmentation, as in [29].

4 Experiment

We now report experiments evaluating the proposed method on two widely-used datasets: CMU iCoseg (37 sets of images, 4 to 41 images per class) [3] and MSRC (14 sets of images, around 30 images for each set) [29]. We compared with several existing methods, some of them being the current state-of-art. For performance metric, we compute the accuracy of the segmentation result over the manually labeled mask, which includes both foreground and background $p = \frac{\text{true postive+true negative}}{\text{area of image}}$. Since the proposed method is unsupervised, the existing methods which required training stage, e.g., [32][34], are not included for comparisons. Note, we don't include the results from [33][29] for iCoseg dataset, because their results are only computed with 30 classes instead of the total 37 classes, which makes the comparison unfair. For all experiments, we use $\lambda = 0.01$ for all three methods.

The results of the proposed methods with comparison to the existing methods are summarized in Tab. 2 (a) and (b), where the per-set accuracy is also illustrated in Fig. 2. From the results, we can find that the proposed methods significantly outperform most of the compared existing methods. The average performances are about 88% on iCoseg dataset and 81% on MSRC dataset, where the lowest accuracy (around 0.6) is reported for Set 17 from iCoseg dataset. An explanation to this could be that the images in this set are less homogeneous and enforcing the classifiers to be similar is less beneficial. The proposed methods have better performances on iCoseg dataset than on MSRC dataset, which could be explained by that (as shown in Fig. 4) the foreground objects in MSRC dataset are more heterogeneous compared with those in iCoseg dataset. In addition, out of the three methods, the "mean" works best on iCoseg dataset but outperformed by "$\ell_{2,1}$" and "low" in MSRC, which can be explained as follows":"mean" assumes that the classifiers are under the same distribution, while "$\ell_{2,1}$" and "low" have more relaxed assumptions; accordingly, "$\ell_{2,1}$" and "low" work better on dataset with images having larger variations.

We also evaluated the computation time and accuracy of the proposed methods with different number of images to be segmented. To this end, we used two sets of [29], "Airplane100" and "Car100", with each set containing 100 images. In our experiment, we first excluded the outliers (the images do not share common foreground) and then applied the proposed methods to different numbers of images, from 5 to 90 with a step-size 10. The results

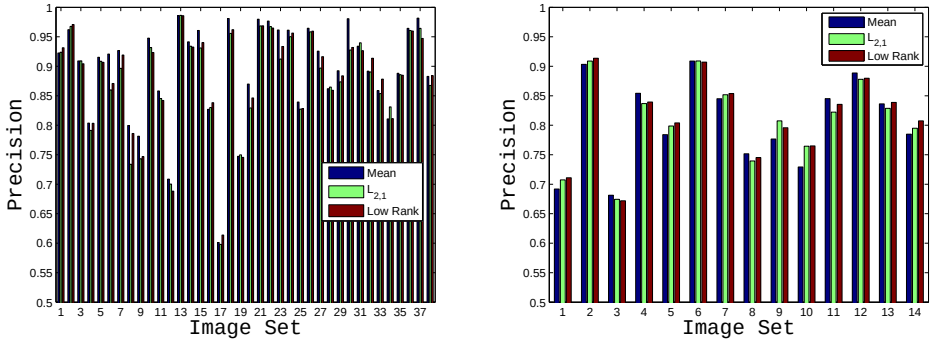

Figure 2: The precision of the proposed algorithms for each set of images of iCoseg (left) and MSRC (right) dataset, where three multi-task learning methods are compared. The x-axis is the index of the image sets, and y-axis is the precision.

Method	Accuracy
[16]	78.9%
[26]*	82.4%
[30]*	83.9%
[33]	60.2%
Mean	88.67%
$\ell_{2,1}$	87.81%
Low	88.33%

(a)

Method	Accuracy
[16]	46.70%
[17]	50.20%
[20]	36.60%
[25]	40.90%
[29]	87.66%
[8]	63.00%
Mean	80.58%
$\ell_{2,1}$	80.87%
Low	81.20%

(b)

Table 2: (a) The result on iCoseg dataset. In this experiment, the proposed methods outperform all the compared approaches. (b) The result on MSRC dataset.

are shown in Fig. 3. From these results, we can see that, the computation times increase linearly with the number of images to be segmented. The performances of the proposed approach become stable with more than 20 image. For “horse 100" set, the performance decreases when the number of images increases, which might be explained by that the images in this set are less homogeneous. As the number of images increases, the classifiers become diverse and have less shared information, then multi-task learning approach creates less benefits.

5 Conclusion and Discussion

In this paper, we presented a novel image cosegmentation approach based on multi-task learning, where segmenting of each image was viewed as a task and the prior that common objects are shared in images is modeled as the intrinsic relatedness of the tasks. In implementing the solutions under this formulation, we evaluated and compared three types of schemes. In experiments, we evaluated the proposed methods on two widely used bench-

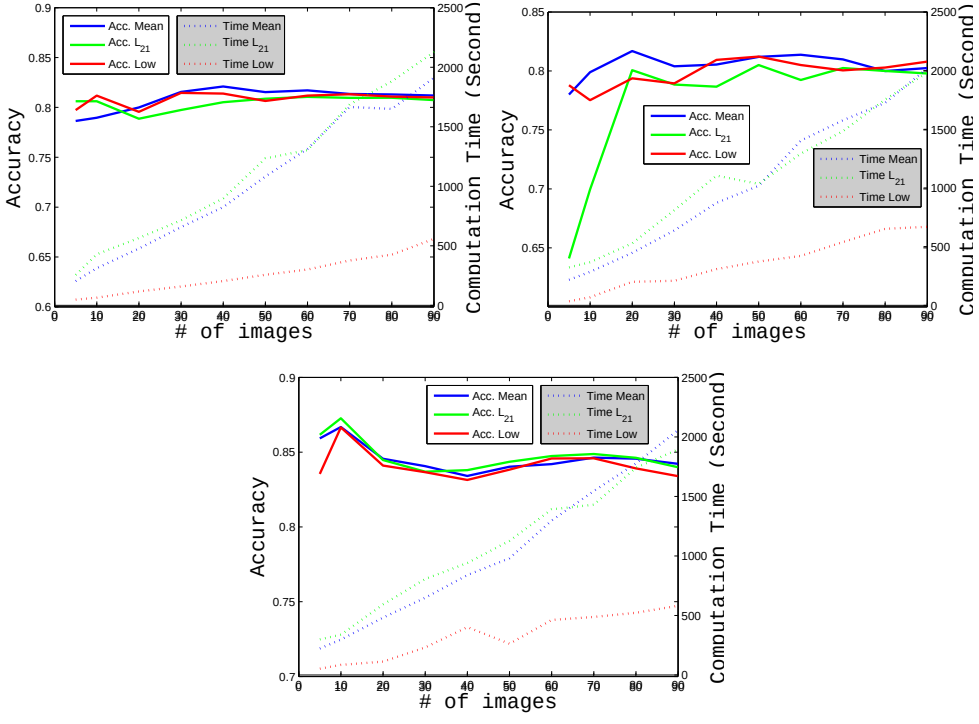

Figure 3: The accuracy and computation time of the proposed methods on “Airplane100" and “Car100" set with varying number of input images.

mark datasets, iCoseg and MSRC, and competitive results were achieved. The proposed methods were also demonstrated to be efficient, where segmenting 5 to 30 images costed less than three minutes. For future work, we plan to investigate the replacement of grab-cut by joint-grab-cut [21] as suggested by [10].

6 Acknowledgment

The work was supported in part by a grant (#1135616) from the National Science Foundation.
Any opinions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

iCoseg Dataset

MSRC dataset

Figure 4: Example of image segmentation on iCoseg dataset and MSRC dataset, where the green contour shows the segmentation results.

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 34(11):2274–2282, 2012.

[2] Rie Kubota Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *The Journal of Machine Learning Research*, 6: 1817–1853, 2005.

[3] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3169–3176. IEEE, 2010.

[4] Steffen Bickel, Jasmina Bogojeska, Thomas Lengauer, and Tobias Scheffer. Multi-task learning for hiv therapy screening. In *Proceedings of the 25th international conference on Machine learning*, pages 56–63. ACM, 2008.

[5] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *PAMI*, PP(99):1, 2012. ISSN 0162-8828. doi: 10.1109/TPAMI.2012.89.

[6] Kai-Yueh Chang, Tyng-Luh Liu, and Shang-Hong Lai. From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2129–2136. IEEE, 2011.

[7] Hwann-Tzong Chen. Preattentive co-saliency detection. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1117–1120. IEEE, 2010.

[8] Jifeng Dai, Ying Nian Wu, Jie Zhou, and Song-Chun Zhu. Cosegmentation and cosketch by unsupervised learning. In *14th International Conference on Computer Vision*, 2013.

[9] Theodoros Evgeniou and Massimiliano Pontil. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117. ACM, 2004.

[10] A. Faktor and M. Irani. Co-segmentation by composition. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1297–1304, Dec 2013. doi: 10.1109/ICCV.2013.164.

[11] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[12] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.

[13] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254 –1259, nov 1998.

[14] L. Jacob, F. Bach, and J.P. Vert. Clustered multi-task learning: A convex formulation. *Advances in Neural Information Processing Systems*, 2008.

[15] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, and Shipeng Li. Salient object detection: A discriminative regional feature integration approach. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013.

[16] Armand Joulin, Francis Bach, and Jean Ponce. Discriminative clustering for image co-segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1943–1950. IEEE, 2010.

[17] Armand Joulin, Francis Bach, and Jean Ponce. Multi-class cosegmentation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 542–549. IEEE, 2012.

[18] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.

[19] Gunhee Kim and Eric P Xing. On multiple foreground cosegmentation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 837–844. IEEE, 2012.

[20] Gunhee Kim, Eric P Xing, Li Fei-Fei, and Takeo Kanade. Distributed cosegmentation via submodular optimization on anisotropic diffusion. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 169–176. IEEE, 2011.

[21] Daniel Kuettel, Matthieu Guillaumin, and Vittorio Ferrari. Segmentation propagation in imagenet. In *Computer Vision–ECCV 2012*, pages 459–473. Springer, 2012.

[22] Jun Liu, Shuiwang Ji, and Jieping Ye. Multi-task feature learning via efficient l 2, 1-norm minimization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 339–348. AUAI Press, 2009.

[23] James Malcolm, Yogesh Rathi, and Allen Tannenbaum. Graph cut segmentation with nonlinear shape priors. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 4, pages IV–365. IEEE, 2007.

[24] Ran Margolin, Ayellet Tal, and Lihi Zelnik-Manor. What makes a patch distinct? In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1139–1146, 2013. doi: 10.1109/CVPR.2013.151.

[25] Lopamudra Mukherjee, Vikas Singh, and Chuck R Dyer. Half-integrality based algorithms for cosegmentation of images. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2028–2035. IEEE, 2009.

[26] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[27] Carsten Rother, Tom Minka, Andrew Blake, and Vladimir Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 993–1000. IEEE, 2006.

[28] Michael Rubinstein, Ce Liu, and William T Freeman. Annotation propagation in large image databases via dense image correspondence. In *Computer Vision–ECCV 2012*, pages 85–99. Springer, 2012.

[29] Michael Rubinstein, Armand Joulin, Johannes Kopf, and Ce Liu. Unsupervised joint object discovery and segmentation in internet images. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1939–1946. IEEE, 2013.

[30] Jose C Rubio, Joan Serrat, Antonio López, and Nikos Paragios. Unsupervised co-segmentation through region matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 749–756. IEEE, 2012.

[31] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[32] Jian Sun, Jean Ponce, et al. Learning discriminative part detectors for image classification and cosegmentation. In *International conference on computer vision*, 2013.

[33] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Cosegmentation revisited: Models and optimization. In *Computer Vision–ECCV 2010*, pages 465–479. Springer, 2010.

[34] Sara Vicente, Carsten Rother, and Vladimir Kolmogorov. Object cosegmentation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2217–2224. IEEE, 2011.

[35] Zhengxiang Wang and Rujie Liu. Semi-supervised learning for large scale image cosegmentation. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 393–400, Dec 2013. doi: 10.1109/ICCV.2013.56.

[36] Yan Yan, Elisa Ricci, Ramanathan Subramanian, Oswald Lanz, and Nicu Sebe. No matter where you are: Flexible graph-guided multi-task learning for multi-view head pose classification under target motion. *ICCV*, 2013.

[37] Xiao-Tong Yuan and Shuicheng Yan. Visual classification with multi-task joint sparse representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3493–3500. IEEE, 2010.

[38] Tianzhu Zhang, Bernard Ghanem, Si Liu, and Narendra Ahuja. Robust visual tracking via multi-task sparse learning. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2042–2049. IEEE, 2012.

[39] Y. Zhang and D.Y. Yeung. A convex formulation for learning task relationships in multi-task learning. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 733–742, 2010.

[40] J. Zhou, J. Chen, and J. Ye. *MALSAR: Multi-tAsk Learning via StructurAl Regularization*. Arizona State University, 2011. URL http://www.public.asu.edu/~jye02/Software/MALSAR.

[41] Jiayu Zhou, Jianhui Chen, and Jieping Ye. Clustered multi-task learning via alternating structure optimization. *Advances in Neural Information Processing Systems*, 2011.