Simultaneous Semi-Supervised Segmentation Of Category-Independent Objects From A Collection Of Images

Hager Merdassi[1], Walid Barhoumi[1,2] and Ezzeddine Zagrouba[1]

[1] Université de Tunis El Manar, Institut Supérieur d'Informatique, Research Team on Intelligent Systems in Imaging and Articial Vision (SIIVA), LR16ES06 Laboratoire de recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2 Rue Abou Raihane Bayrouni, 2080 Ariana, Tunisia
[2] Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs 2035 Charguia II, Tunis-Carthage, Tunisia
hager.merdassi@hotmail.fr, walid.barhoumi@enicarthage.rnu.tn, ezzeddine.zagrouba@fsm.rnu.tn

Abstract. This work is about simultaneous segmentation of different foreground objects from a collection of images with heterogeneous contents. Our idea is to propagate the segmentation information between images in order to detect foreground objects in all these images simultaneously, under the hypothesis of using categorized or uncategorized images, rather than resorting to image co-segmentation that forces the use of similar categorized images. In fact, given an input image, the objective is to integrate seamlessly other images in the general foreground model, in order to benefit the segmentation of the foreground objects in this image. Indeed, the proposed method aggregates general information, on foregrounds as well as on backgrounds, from a collection of images. To this end, the method is based on an energy minimization function. The linear dependence of the foreground histograms is firstly estimated to optimize the proposed energy function. Then, an iterative optimization of each image permits to remarkably optimize the final segmentation result for all images composing the input collection. Extensive experiments demonstrate that the suggested method allows full-object segmentation of the foreground from a collection of images composed of different classes of objects. Indeed, the validation of the accuracy on four challenging datasets (iCoseg, Oxford Flowers, Caltech101 and Berkeley) shows that the proposed method compares favorably with the state-of-the-art of foreground object segmentation from a collection of images. Besides, it has the challenging ability to deal accurately with uncategorized objects.

Keywords: Category-independent objects, foreground segmentation, linear dependence, co-segmentation.

1 Introduction

The current explosion in digital images led to the emergence of methods and algorithms that proposed to exploit these images, resulting in better applications for users to interact with these data. Nevertheless, seen the diversity of the used environments, these datasets are often noisy even for categorized objects' collections. In addition, these datasets are typically heterogeneous and very large. Thus, there is an urgent need for standard methods to efficiently segment objects of interest in large image collections with heterogeneous contents. In fact, object segmentation from a collection of images forms an essential step for many computer vision applications [1, 2], such as content-based image structuration, indexing and retrieval, interactive image editing, scene understanding, action recognition [3, 4], image annotation, human pose estimation [5, 6], object recognition [7, 8] and facial detection [9, 10].

Most existing relevant methods, for this purpose, mainly those based on co-segmentation, aim to extract the shared object from a set of images, having varying sizes and characteristics but related to the shared object. Recently, co-segmentation methods perform well if the appearance of the shared objects in a set of images has different degrees of variability. Nevertheless, since even a single object is often comprised of heterogeneous textures and colors, the accurate co-segmentation of foreground objects is still an open research problem. To deal with this problem, we propose in this work simultaneous segmentation method of salient objects by analyzing image collections consisting of heterogeneous images from various sources.

The main idea of the proposed method is that the visual appearances are often shared between objects of the same category, and even over disparate categories [11]. Thus, we propose a method that leads to extract foreground objects in a set of images, whether this collection of images contains similar images (*i.e.* belonging to the same semantic class) or heterogeneous ones (*i.e.* images belonging to different semantic classes). The core objective of this method is to orient the problem of salient object detection to a much easier problem of co-segmentation. Moreover, we explore the linear dependence of the foreground histograms, in order to segment accurately objects independently of their classes.

The remaining part of this paper is organized as follows. In Section 2, we briefly discuss the related work on the foreground object extraction. In section 3, we describe the proposed method for segmenting foreground objects in a set of images with heterogeneous contents. Extensive experiments, on various challenging datasets, are summarized in Section 4. Finally, we produce a conclusion with some directions for future works in Section 5.

2 Related work

In this section, we review related work on the general framework of foreground object extraction in digital images. In fact, many methods have been recently proposed for extracting salient objects from a clean collection of categorized images, where categorized objects are referred to objects in the same category [12]. These methods address the problem of joint segmentation of different instances of a single category of object across a collection of images [13], in order to exploit a large amount of contextual information from the image collection to optimize the separation between the foreground objects and the background. Such an issue is referred to as the co-segmentation of categorized objects, a closely related task to co-detection [14, 15], and has been actively studied in recent papers [16, 17]. In fact, co-detection and co-segmentation methods were introduced to exploit the collective power of a collection of images. In particular, most co-segmentation methods [18, 19, 20] are based on low-level image appearance information and they formulated the problem of co-segmentation as an energy minimization problem. Image co-segmentation methods started by segmenting the common objects in two images of an object [20], multiple images with the same single object [21], multiple classes in each image [22] or multiple images with more than one object in an image [23].

More recently, other methods focused on the co-segmentation of multiple image groups with different characteristics but related to the same general object [17]. Indeed, various methods [13] have focused on simultaneous segmentation of categorized objects, through either supervised learning (given user interactions) [23, 24, 25] or unsupervised learning [26, 27], from all images. Most of these methods model the appearance cues [23, 24], the object shape [27] and/or subspace structure [23] across the image collection. In fact, there exists two main families of methods. On the one hand, to benefit object segmentation, co-segmentation methods rely on shared appearance of the object of interest between views, high variability between backgrounds across views, and discrimination between foreground and background appearance distributions [28]. On the other hand, silhouette-coherent extraction methods approaches rely on geometric consistency of the segmentations, usually building some form of dense shape representation of the foreground object [28].

Since silhouette-coherent extraction methods are complex and computationally involved due to the updates of dense shape representations [28], co-segmentation methods are much more used. In general, most existing methods of segmentation of categorized objects were built on the assumption that all images in the input collection contain the target object. Thus, they may not work well when there are some noisy images (*i.e.* images that do not involve the target category of object) in the given collection, what is the case for collections gathered by a text query from image search engines [13]. To overcome this limitation, we propose in this work a method for the joint co-segmentation of foreground objects from all images in a heterogeneous collection.

The suggested work goes one step further to directly segment uncategorized object from a noisy image collection, while previous works all assumed that images from the input collection all contain the same target categorized object. There are few co-segmentation methods [22, 29] that further conduct the co-segmentation of multiple objects of multiple categories. These methods assume that each image should contain at least one object among multiple known categories. In contrast, we co-segment simultaneously a collection of uncategorized images with different object instances of unlimited number of unknown categories.

3 Proposed method

Our intuition is to obviously profit of foreground information as well as of background information, notably the spatial context, across a collection of uncategorized images in order to provide valuable information for object segmentation. In fact, the proposed method is based on linear dependence of the generated foreground histograms. This leads to an interesting optimization model and effective solutions for tandem objects segmentation of a high number of images. In fact, solving the co-segmentation problem usually returns to the minimization of an energy function that can be generally represented by:

$$\Xi(\boldsymbol{I_1}, \boldsymbol{I_2}, \dots, \boldsymbol{I_n}) = \sum_{i=1}^{n} \sum_{p} \boldsymbol{D_{i,p}} \boldsymbol{x_p} + \sum_{i=1}^{n} \sum_{p \sim q} \boldsymbol{S_{i,pq}} \left| \boldsymbol{x_p} - \boldsymbol{x_q} \right| + \boldsymbol{\alpha}.\boldsymbol{G}(\boldsymbol{F_1}, \boldsymbol{F_2}, \dots, \boldsymbol{F_n}), \quad (1)$$

where, $\boldsymbol{\zeta} = \{\boldsymbol{I_1}, \boldsymbol{I_2}, \dots, \boldsymbol{I_n}\}$ is the input collection of $\boldsymbol{n}$ images. The first two terms are the Markov random field (MRF) energy terms for each image, where $\boldsymbol{D}$ and $\boldsymbol{S}$ are a data term and a smoothness term, respectively. The variables $\boldsymbol{x_p}$ and $\boldsymbol{x_q}$ denote the pixel label, such that $\boldsymbol{x = 0}$ for background and $\boldsymbol{x = 1}$ for foreground. The last term $\boldsymbol{G}$ is a global term that penalizes the difference, or maximizes the similarity, between the foreground histograms $\boldsymbol{F_i}$, where $i \in \{\boldsymbol{1}, \dots, \boldsymbol{n}\}$, relatively to the $\boldsymbol{n}$ input images. The coefficient $\boldsymbol{\alpha}$ expresses the relative weights of the global term. Several methods [18, 19, 20, 22, 30, 31, 32, 33, 34, 35, 36] have been recently introduced to evaluate the global term. They are based on the general model of (1). Nevertheless, they only focus on similar images, where simultaneous segmentation of different images has not been explored. In this work, we tackle the above problem, in order to segment the foreground objects from a set of images with homogeneous or heterogeneous contents.

The first step of the proposed method looks to generate histograms of the input images. It consists in providing a binary segmentation of each image in order to separate the foreground from the background. In fact, given the number of bins $\boldsymbol{K}$, each pixel of each image is associated to one of the $\boldsymbol{K}$ bins of the corresponding histogram. Then, a binary segmentation of each image partitions the set of pixels into foreground and background pixels, using the decision variable for segmentation $\boldsymbol{x}$, and we define a K-dimensional vector $\overrightarrow{\boldsymbol{H_i}}$ for each image I_i, where $\overrightarrow{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{k})$ is the number of pixels in the image $\boldsymbol{I_i}$ that were associated to the k^{th} histogram bin. Then, we derive a matrix $\boldsymbol{H}$ of size $\boldsymbol{K} \times \boldsymbol{n}$ that includes all the generated histograms $\overrightarrow{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{k})$, each as a column. This matrix can be expressed by the sum of the two matrices $\boldsymbol{F}$ and $\boldsymbol{B}$ [19]:

$$\boldsymbol{H} = \boldsymbol{F} + \boldsymbol{B}, \quad (2)$$

where, $\boldsymbol{F} = [\boldsymbol{F_1}, \dots, \boldsymbol{F_n}]$ and $\boldsymbol{B} = [\boldsymbol{B_1}, \dots, \boldsymbol{B_n}]$ are the foreground and background histograms.

The purpose of the second step is to make the columns of $\boldsymbol{F}$ similar for foreground objects of the same class and dissimilar for different foreground objects, while segmenting simultaneously objects from all input images. Note that the matrix $\boldsymbol{F}$ of size

$\boldsymbol{K} \times \boldsymbol{n}$ gathers all generated histograms of the foregrounds, such that each column of the produced matrix $\boldsymbol{F}$ corresponds to the foreground histogram of an input image. To do that, we made use of linear dependence of the foreground histograms' vectors and this task returns to search the rank of $\boldsymbol{F}$. In fact, for similar objects, foreground vectors $\boldsymbol{F_i}$ must be similar or linearly dependent in the general case to assume invariance against scale variation. Thus, the rank $\boldsymbol{r}$ of $\boldsymbol{F}$ must be equal to one. Otherwise, for heterogeneous images, foreground vectors $\boldsymbol{F_i}$ must be dissimilar or linearly independent and the rank $\boldsymbol{r}$ of $\boldsymbol{F}$ should be equal to the number of foreground objects in the input collection. In cases where a precise matrix $\boldsymbol{F}$ of rank $\boldsymbol{r}$ cannot be found, a "slack" in the form of a small (sparse) residual [21] $\boldsymbol{P}$, may be permitted, where $\boldsymbol{F} = \boldsymbol{R} + \boldsymbol{P}$. Once we defined the matrix $\boldsymbol{R}$, the object segmentation model also includes the MRF segmentation terms for each image. Thus, the object segmentation energy is expressed by the following minimization problem:

$$\min_{\boldsymbol{x}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left(\sum_{\boldsymbol{p=1}}^{\boldsymbol{L}} \boldsymbol{D_{i,p} x_{i,p}} + \sum_{\boldsymbol{p=q=1,p\sim q}}^{\boldsymbol{L}} \boldsymbol{S_{i,pq} y_{i,pq}} + \boldsymbol{\alpha} \left| \boldsymbol{F_i} - \boldsymbol{R_i} \right|_2^2 \right), \tag{3}$$

where, $\boldsymbol{x_{i,p}} \in \{0, 1\}, \boldsymbol{y_{i,pq}} \in \{0, 1\}$, $\boldsymbol{D}$ and $\boldsymbol{S}$ are the data and smoothness terms of the random Markov field, $\boldsymbol{F_i} = \overrightarrow{\boldsymbol{H}}_{\boldsymbol{i}} \boldsymbol{x_i}$ and $\boldsymbol{y_{i,pq}} = \left| \boldsymbol{x_{i,p}} - \boldsymbol{x_{i,q}} \right|$. The variables $\boldsymbol{x_{i,p}}$ and $\boldsymbol{x_{i,q}}$ denote the labeling $\boldsymbol{x}$ of two neighboring pixels $\boldsymbol{p}$ and $\boldsymbol{q}$ of the i^{th} image, and $rank(\boldsymbol{R})$ is the number of the foreground objects in the studied collection. Then, the objective model penalizes $\boldsymbol{P}$ to minimize the proposed energy function and to make a small variation within the matrix $\boldsymbol{R}$. Thus, the proposed objective model is expressed as follows:

$$\min_{\boldsymbol{x}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left(\sum_{\boldsymbol{p=1}}^{\boldsymbol{L}} \boldsymbol{D_{i,p} x_{i,p}} + \sum_{\boldsymbol{p=q=1,p\sim q}}^{\boldsymbol{L}} \boldsymbol{S_{i,pq} y_{i,pq}} + \boldsymbol{\alpha} \left| \boldsymbol{P_i} \right|_2^2 \right), \tag{4}$$

where, $\boldsymbol{P_i}$ is the residual matrix of the i^{th} image. In addition, we assume that Gaussian mixture model (GMM) weight is generated and already available for each image. This is used to construct a Markov random field data term and the smoothness term that will be used for the object segmentation.

4 Results

In this section, we present the experimental results of the proposed method (PM) on various standard datasets. In fact, Figure 1.a presents a sample of object segmentation outputs for the image classes "skating", "airshows_plane" and "pandas" from the iCoseg dataset, and Figure 1.b illustrates the obtained results for a sample of 18 flower images, covering many species, from the Oxford Flowers dataset. In our experiments for both datasets, we used different groups of similar images, with varied number of images in each group, as well as varied foreground positions, shapes, locations, colors and sizes in each image. We conclude that the proposed method allows a precise segmentation of the objects of interest in all the images simultaneously. Note that in accordance with existing works [20, 21, 30], generation of histograms is based on the use of a combination of color, texture features, and SIFT. The number of bins for each color channel was between 10-20.

Moreover, Table 1 shows a comparison of the proposed method with the most relevent methods of the state-of-the-art [30, 31, 35, 37, 32, 16]. In fact, accuracy scores, on sixteen image classes from the iCoseg dataset, are computed for each method (some of the visual results are shown in Figure 1.a). We notice a high accuracy for the proposed method compared to other co-segmentation works, in terms of accuracy average, and even for each class separately, except for three classes ("taj mahal", "gymnastics" and "statue of liberty"). Indeed, the proposed method does not significantly outperform the state-of-the-art cosegmentation method [32] for two classes "taj mahal" and "statue of liberty". In fact, this method [32] is based on consistent functional maps for transporting properties between the RGB images.

Table 1. Accuracy scores of various object segmentation methods on the iCoseg dataset.

iCoseg Datasets	[30]	[31]	[35]	[37]	[32]	[16]	PM
Alaskan Brown Bear	74.8	86.4	90.0	90.0	90.4	93.5	**96.1**
Red Sox Players	73.0	90.5	90.9	90.9	94.2	96.5	**97.7**
Stonehenge1	56.6	87.3	63.3	91.3	92.5	93.0	**96.2**
Stonehenge2	86.0	88.4	88.8	84.2	87.2	83.5	**90.9**
Taj mahal	73.7	88.7	91.1	81.7	**92.6**	88.7	88.1
Elephant	70.1	75.0	43.1	86.2	86.7	90.4	**96.9**
Panda	84.0	60.0	92.7	92.2	88.6	81.2	**96.7**
Kite	87.0	89.8	90.3	94.9	93.9	96.6	**98.3**
Kite Pandas	73.2	78.3	90.2	90.9	93.1	83.8	**97.3**
Gymnastics	90.9	87.1	91.7	**97.7**	90.4	95.8	93.0
Ferrari	85.0	84.3	89.9	92.7	95.6	91.7	**95.5**
Skating	82.1	76.8	77.5	79.9	78.7	81.7	**94.8**
Women Soccer Players	76.4	82.6	87.5	86.7	89.4	93.0	**95.7**
Balloon	85.2	89.0	90.1	92.7	90.4	96.5	**99.3**
Statue of Liberty	90.6	91.6	93.8	91.1	**96.8**	92.7	91.0
Brown bear	74.0	80.4	95.3	86.2	88.1	94.8	**97.7**
Average	78.9	83.5	85.4	89.3	90.5	90.8	**95.**3

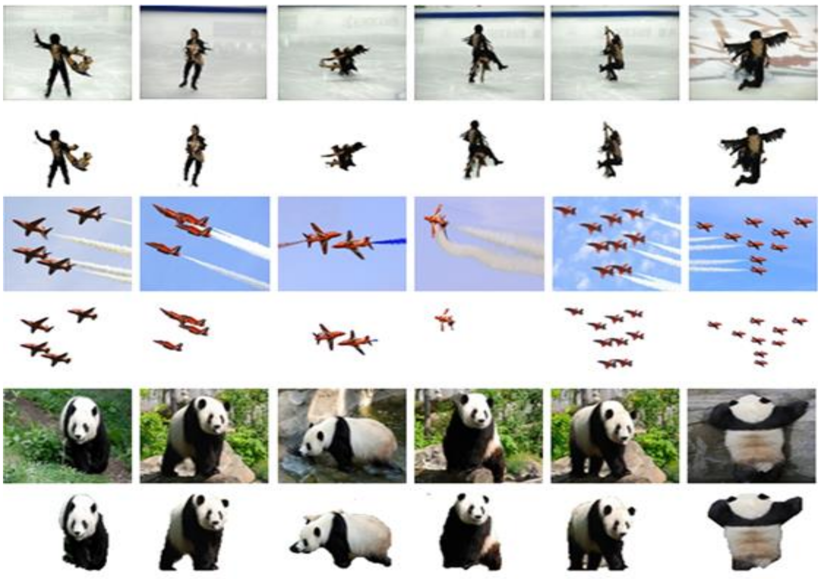
Besides, the state-of-the-art cosegmentation method [37] outperforms the proposed method for the "gymnastics" class. This method [37] uses a dense correspondences between images to capture the sparsity and visual variability of the common object over the entire database. These failures to surpass these methods on such classes can be explained by the strong edges in the background for "taj mahal" images, the view-point variations for "gymnastics" images and partial occlusions of the foreground object for "statue of lıberty" images. Note that we have restricted our selves to only 16 classes, among 38 ones, seen that accuracy scores of all compared methods are available just for the classes presented in the Table 1. Moreover the code of some methods has not been made available, thus, we directly report the accuracy scores provided in their papers.

Furthermore, Figure 2 shows object segmentation results on a challenging sample of images form the Berkeley dataset (Figure 2.a) and the Caltech101 dataset (Figure 2.b). In fact, for the first image dataset, we selected some complex images with high variablity of the foreground and backgroud. The obtained results are very encouraging, even for tiny foreground objects (e.g. the first image and the sixth one). Nevertheless, the segmentation of the foreground objects is not complete for some rare cases (e.g. the deer horns do not appear in the segmented foreground of the second image) and this is mainly due to the close homogeneity, of the missing parts of the object, with the background. For the second dataset of images, we successfully eliminate the background, however, we do not extract perfectly all the foreground objects. The challenge of this image dataset lies in the extreme changes in lighting of the foreground objects.

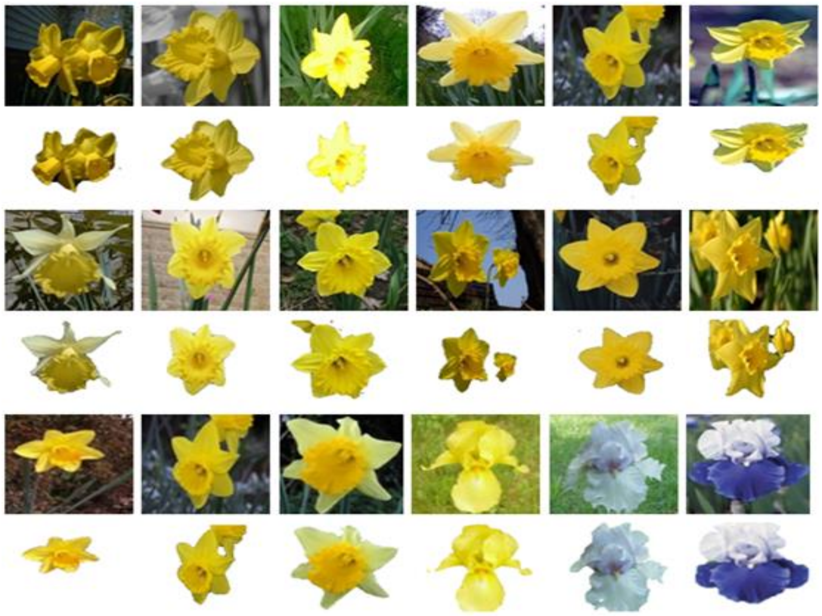
5 Conclusion

The work presented here has focused on simultaneous segmentation of categorized and uncategorized objects from an image collection. The main contribution of this work is to make co-segmentation approaches applicable to a significantly more general framework that has not been addressed before, as far as we know. In fact, the proposed method is based on energy minimization function that evaluates the linear dependence of the foreground histograms computed for each input image. The proposed method, which is applicable to a variety of object categories, will be beneficial, for segmenting foreground objects independently of their classes, especially if the object of interest is not in the center of the image or is of size different from the other objects' sizes. Indeed, experimental results show that the suggested method is able to extract the intact objects simultaneously from a large set of category-independent images, even in the case of dramatic changes of their shape and size, and in the presence of complex backgrounds.

As far as future works are concerned, we aim to compare the foreground objects rather than the totality of the images for image comparison. This task may be explored for image classification, image retrieval and key frame extraction for videos [38]. Besides, as results indicate that the suggested method scales easily to large number of images, since it is able to exploit a large amount of contextual information (on foregrounds as well as on backgrounds) from the image collection what permits more robust foreground/background segmentation, we will try to test it on very large-scale datasets.



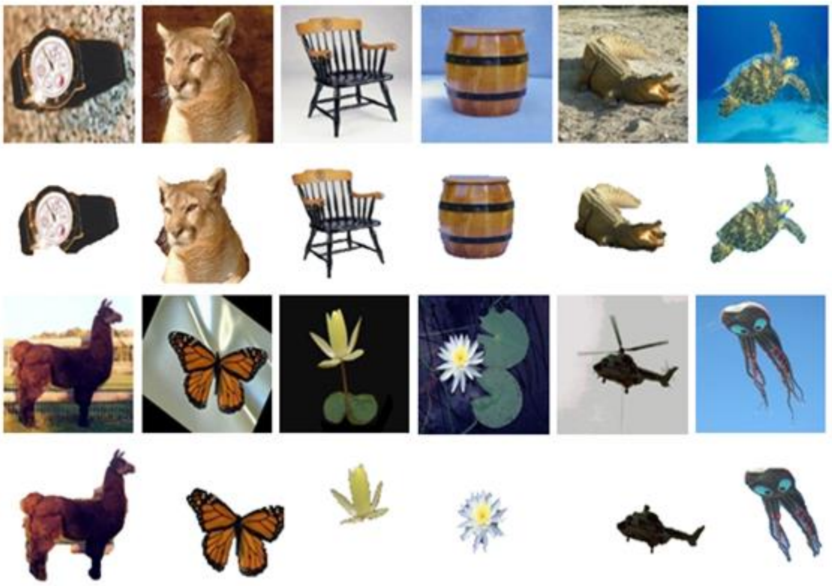
(a)



(b)

Fig. 1. Input images and resulting segmentations, while using the proposed method, for categorized objects from the iCoseg dataset (a) and the Oxford Flowers dataset (b).

(a)



(b)

Fig. 2. Input images and resulting segmentations, while using the proposed method, for uncategorized objects from the Berkeley dataset (a) and the Caltech101 dataset (b).

References

1. Li, Y., Hou, X., Koch, C., Rehg, J. M., & Yuille, A. L. (2014). The Secrets of Salient Object Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 4321–4328. Doi: 10.1109/CVPR.2014.43.
2. Xiao, J., Ehinger, K. A., Hays, J., Torralba, A., & Oliva, A. (2016). SUN Database: Exploring a Large Collection of Scene Categories. *International Journal of Computer Vision*, *119*(1), 3–22. Doi: 10.1007/s11263-014-0748-y.
3. Zhang, H., Zhou, W., Reardon, C., & Parker, L. E. (2014). Simplex-based 3D spatio-temporal feature description for action recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2067–2074). Doi: 10.1109/CVPR.2014.265.
4. Tian, Y., Luo, P., Wang, X., & Tang, X. (2015). Pedestrian detection aided by deep learning semantic tasks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 5079–5087). Doi: 10.1109/CVPR.2015.7299143.
5. Cherian, A., Mairal, J., Alahari, K., & Schmid, C. (2014). Mixing Body-Part Sequences for Human Pose Estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, 2361–2368. Doi: 10.1109/CVPR.2014.302.
6. Dantone, M., Gall, J., & Leistner, C. (2013). Human Pose Estimation Using Body Parts Dependent Joint Regressors. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13)*, 3041–3048. Doi: 10.1109/CVPR.2013.391.
7. Zhu, F., Jiang, Z., & Shao, L. (2014). Submodular object recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2457–2464). Doi: 10.1109/CVPR.2014.315.
8. Lo Presti, L., & La Cascia, M. (2016). 3D skeleton-based human action classification: A survey. *Pattern Recognition*, *53*, 130–147. Doi: 10.1016/j.patcog.2015.11.019.
9. Ding, L., & Martinez, A. M. (2008). Precise detailed detection of faces and facial features. In *26th IEEE Conference on Computer Vision and Pattern Recognition*, 1-7. Doi : 10.1109/CVPR.2008.4587812.
10. Haoxiang, L., Zhe, L., Xiaohui, S., Brandt, J., & Gang, H. (2015). A convolutional neural network cascade for face detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 5325–5334. Doi: 10.1109/CVPR.2015.7299170.
11. Kim, J. & Grauman, K. (2012). Shape Sharing for Object Segmentation. *European Conference on Computer Vision*, 7578, 444-458. Doi: 10.1007/978-3-642-33786-4_33.
12. Li, K., Zhang, J., & Tao, W. (2016). Unsupervised Co-Segmentation for Indefinite Number of Common Foreground Objects. *IEEE Transactions on Image Processing*. Doi: 10.1109/TIP.2016.2526900.
13. Wang, L., Hua, G., Xue, J., Gao, Z., Zheng, N. (2014). Joint Segmentation and Recognition of Categorized Objects from Noisy Web Image Collection. In *IEEE Transactions on Image Processing*, 23(9), pp. 4070-4086. Doi: 10.1109/TIP.2014.2339196.
14. Bao, S. Y., Xiang, Y., & Savarese, S. (2012). Object co-detection. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (7572 LNCS, pp. 86–101). Doi: 10.1007/978-3-642-33718-5_7.
15. Shi, J., Liao, R., & Jia, J. (2013a). CoDeL: A human co-detection and labeling framework. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2096–2103). Doi: 10.1109/ICCV.2013.262.
16. Fu, H., Xu, D., Lin, S., & Liu, J. (2015). Object-based RGBD image co-segmentation with mutex constraint. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 4428–4436). Doi: 10.1109/CVPR.2015.7299072.

17. Meng, F., Cai, J., & Li, H. (2016). Cosegmentation of multiple image groups. In *Computer Vision and Image Understanding* (146, pp. 67–76). Doi: 10.1016/j.cviu.2016.02.004.
18. Rother, C., Kolmogorov, V., Minka, T., & Blake, A. (2006b). Cosegmentation of image pairs by histogram matching - Incorporating a global constraint into MRFs. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1, pp. 994–1000). Doi: 10.1109/CVPR.2006.91.
19. Mukherjee, L., Singh, V., & Dyer, C. R. (2009a). Half-integrality based algorithms for cosegmentation of images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops,* (pp. 2028–2035). Doi: 10.1109/CVPRW.2009.5206652.
20. Hochbaum, D. S., & Singh, V. (2009). An efficient algorithm for Co-segmentation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 269–276). Doi: 10.1109/ICCV.2009.5459261.
21. Mukherjee, L., Singh, V., & Peng, J. (2011b). Scale invariant cosegmentation for image groups. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1881–1888). Doi: 10.1109/CVPR.2011.5995420.
22. Joulin, A., Bach, F., & Ponce, J. (2012b). Multi-class cosegmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 542–549). Doi: 10.1109/CVPR.2012.6247719.
23. Mukherjee, L., Singh, V., Xu, J., Collins, M. D. (2012c). Analyzing the Subspace Structure of Related Images: Concurrent Segmentation of Image Sets. In *Proceedings of the 12th European conference on Computer Vision* (pp. 128-142). Doi: 10.1007/978-3-642-33765-9_10.
24. Cui, J., Yang, Q., Wen, F., Wu, Q., Zhang, C., Van Gool, L., & Tang, X. (2008). Transductive object cutout. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR'08*. Doi: 10.1109/CVPR.2008.4587589.
25. Law, Y. N., Lee, H. K., Ng, M. K., & Yip, A. M. (2012). A semisupervised segmentation model for collections of images. *IEEE Transactions on Image Processing*, *21*(6), 2955–2968. Doi: 10.1109/TIP.2012.2187670
26. Lee, Y. J., & Grauman, K. (2010). Collect-cut: Segmentation with top-down cues discovered in multi-object images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3185–3192). Doi: 10.1109/CVPR.2010.5539772.
27. Alexe, B., Deselaers, T., & Ferrari, V. (2010). ClassCut for unsupervised class segmentation. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (6315 LNCS, pp. 380–393). Doi: 10.1007/978-3-642-15555-0_28.
28. Djelouah, A., Franco, J.-S., Boyer, E., Le Clerc, F., & Perez, P. (2015). Sparse Multi-View Consistency for Object Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 890-903. Doi: 10.1109/TPAMI.2014.2385704.
29. Kim, G., & Xing, E. P. (2012b). On multiple foreground cosegmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 837–844). Doi: 10.1109/CVPR.2012.6247756.
30. Joulin, A., Bach, F., & Ponce, J. (2010a). Discriminative clustering for image co-segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1943–1950). Doi: 10.1109/CVPR.2010.5539868.
31. Rubio, J. C., Serrat, J., Lopez, A., & Paragios, N. (2012). Unsupervised co-segmentation through region matching. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 749–756). Doi: 10.1109/CVPR.2012.6247745.

32. Wang, F., Huang, Q., & Guibas, L. J. (2013). Image co-segmentation via consistent functional maps. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 849–856). Doi: 10.1109/ICCV.2013.110.
33. Vicente, S., Kolmogorov, V., & Rother, C. (2010a). Cosegmentation revisited: Models and optimization. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (6312 LNCS, pp. 465–479). Doi: 10.1007/978-3-642-15552-9_34.
34. Kim, G., Xing, E. P., Fei-Fei, L., & Kanade, T. (2011a). Distributed cosegmentation via submodular optimization on anisotropic diffusion. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 169–176). Doi: 10.1109/ICCV.2011.6126239.
35. Vicente, S., Rother, C., & Kolmogorov, V. (2011b). Object cosegmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2217–2224). Doi: 10.1109/CVPR.2011.5995530.
36. Ma, T., & Latecki, L. J. (2013). Graph transduction learning with connectivity constraints with application to multiple foreground cosegmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1955–1962). Doi: 10.1109/CVPR.2013.255
37. Rubinstein, M., Joulin, A., Kopf, J., & Liu, C. (2013). Unsupervised joint object discovery and segmentation in internet images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1939–1946). Doi : 10.1109/CVPR.2013.253.
38. Barhoumi, W., & Zagrouba, E. (2013). On-the-fly Extraction of Key Frames for Efficient Video Summarization. *AASRI Conference on Intelligent Systems and Control* (pp. 78 – 84). Doi : 10.1016/j.aasri.2013.10.013.