Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Object co-segmentation via salient and common regions discovery

Yong Li [a], Jing Liu [a], Zechao Li [b],*, Hanqing Lu [a], Songde Ma [a]

[a] National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, No. 95, Zhongguancun East Road, Beijing 100190, China
[b] School of Computer Science, Nanjing University of Science and Technology, Xiaolingwei Road 200, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 30 November 2013
Received in revised form
27 November 2014
Accepted 9 December 2014

Keywords:
Object co-segmentation
Low rank matrix recovery
Discriminative learning

ABSTRACT

The goal of this paper is to simultaneously segment the object regions in a set of images with the same object class, known as object co-segmentation. Different from typical methods, simply assuming that the common regions among images are the object regions, we additionally consider the disturbance from consistent backgrounds, and indicate not only common regions but salient ones among images to be the object regions. To this end, we propose an adaptive discriminative low rank matrix recovery (ADLRR) algorithm to divide the over-completely segmented regions (i.e., super-pixels) of a given image set into object and non-object ones. The proposed ADLRR is formulated from two views: a low-rank matrix recovery term for salient regions detection and a discriminative learning term adopted to distinguish object regions from all super-pixels. An additional regularized term is incorporated to jointly measure the disagreement between the predicted saliency and the objectiveness probability. For the unified learning problem by connecting the above three terms, we design an efficient alternate optimization procedure based on block-coordinate descent and augmented Lagrange multipliers method. Extensive experiments are conducted on three public datasets, i.e., MSRC, iCoseg and Caltech101, and the comparisons with some state-of-the-arts demonstrate the effectiveness of our work.

1. Introduction

Object segmentation is a fundamental task in computer vision and multimedia areas, which is beneficial to many applications, e.g., object retrieval, object recognition and image editing [1–6]. Some fully supervised or interactive object extraction approaches [7–9,1] have been proposed to address the object segmentation problem. Typically the interactive approaches require the user to manually indicate the location of objects in the image, then appearance models are derived from the user input. The segmentation process is often cast as minimization of a binary and pairwise energy function, which can be efficiently optimized by the standard minimum cut algorithm [8]. However the interactive approach cannot be applied to the large scale dataset due to expensive cost of human intervention. Then some fully supervised object detection approaches [10–13] are proposed to avoid the human intervention during the test process. Generally, the object detectors follow the sliding-window paradigm. A classifier is first trained based on pixel-level label to distinguish windows containing instances of a given class from all other windows, then the classifier is used to localize instances of the specific class. However, the object detectors are specialized for one object class and cannot be applied to the general classes. Moreover, most supervised methods have to be given large numbers of training images to learn those detectors. Meanwhile, the presence of common object classes in multiple images makes up for the absence of detailed supervisory information. It becomes possible to segment multiple images jointly to get the common objects, which is known as co-segmentation and has been actively studied in recent years.

To alleviate the problems above, recent researches [14–19] focus on the weakly supervised methods of object co-segmentation, which is to simultaneously segment the same or similar objects appearing in a set of images without pixel-level supervised information. Most of the previous approaches work on the assumption that the regions common among images are deemed as the object regions [14–19]. However, the truth is not the case, since the background regions may also be consistent. For example, as shown in Fig. 1, the 'grass' regions may be common within the 'baseball' images, but they are not the target objects. How to effectively resist the disturbance of such case, and further to precisely identify the true object regions become our focus in this paper.

To this end, we jointly exploit the saliency detection and common region mining from a set of images to perform the task of object co-segmentation. We use the term object co-segmentation to emphasize the fact we are interested in segmenting "objects" rather than "stuff" [20]. Namely, the object regions

* Corresponding author.
E-mail address: zechao.li@njust.edu.cn (Z. Li).

http://dx.doi.org/10.1016/j.neucom.2014.12.110
0925-2312/© 2015 Elsevier B.V. All rights reserved.

are assumed to be not only common among images but also salient in contrast with background regions. It can naturally eliminate the disturbance of those background regions consistent with each other. As shown in Fig. 1, we can easily catch the 'baseball player' regions as the true object regions because they are salient and simultaneously appear in these images, while the common but non-salient regions (e.g., baseball field) and the salient but uncommon regions (e.g., baseball referee and billboard) are deemed as background.

To discover the salient and common regions, we propose a co-segmentation framework based on adaptive low rank matrix recovery and discriminative learning, named as adaptive discriminative low rank matrix recovery (ADLRR), as shown in Fig. 2. Given a set of images of an object class, we first over-completely segment each image into super-pixels, and then employ the proposed ADLRR to identify the salient and common regions in the class specific image set. Inspired by the work in [21], we adopt the basic idea of low-rank matrix recovery to detect the salient regions of an image, i.e., decomposing the super-pixel-wise representation of each image into a low-rank matrix and a sparse matrix, and using the l_1-norm of each column in the sparse matrix to measure the saliency of the corresponding super-pixel. Besides, a class specific feature transform matrix is introduced to enhance the salient regions and ensure that the matrix representing the background has a low rank. Furthermore, discriminative learning is incorporated on the image set to learn the best hyperplane to separate the salient and common super-pixels from the non-salient regions and salient but uncommon regions. Extensive experiments on three publicly available benchmarks, i.e., MSRC, iCoseg and Caltech101, show the satisfied performance of our proposed method. Our main contributions are summarized as follows.

- To the best of our knowledge, we are the first to consider the disturbance of consistent background regions for object co-segmentation.
- To overcome the disturbance of consistent background, we propose to perform saliency detection and discriminative learning in a unified framework to identify the common and salient object regions.
- To enhance the salient regions and make sure that the background lies in a low dimensional space, a class specific feature transform matrix is introduced.

The rest of the paper is organized as follows. Related works about image co-segmentation are conducted in Section 2. Then, we elaborate our proposed model for object co-segmentation in Section 3, and its optimization algorithm is presented in Section 4. The experimental evaluation is given in Section 5 followed with the conclusion in Section 6.

2. Related work

The co-segmentation problem has been actively studied in the past few years. Rother et al. [14] first addressed the co-segmentation of image pairs by histogram matching and incorporating a global constraint into MRFs. Mukherjee et al. [15] modified the energy function in [14] with the l_2-norm instead of the l_1-norm, since such a model has some interesting properties and allows the use of alternative optimization methods. Different from [14,15], Batra et al. [22] used a single foreground and background model for all images with the help of human interaction, since such a model has the submodular property, the optimization

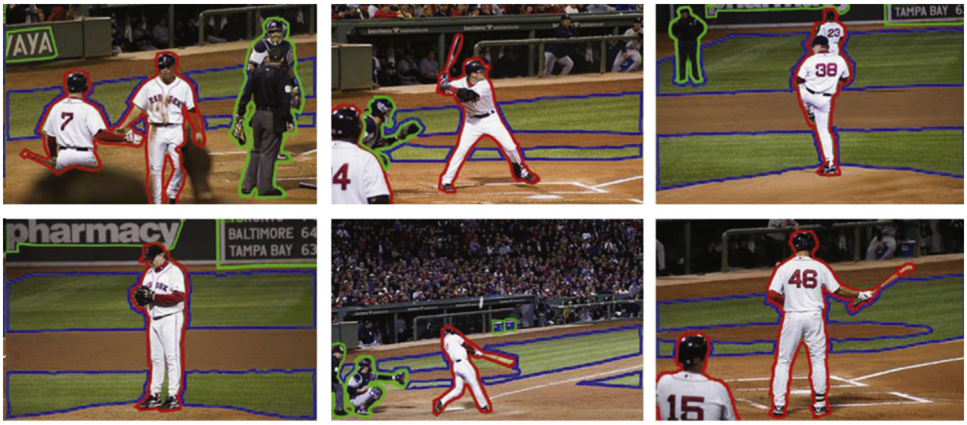


Fig. 1. Examples about the object region. The regions inside red contour are the true object regions which are salient and common, while the both types of regions inside green contours and blue contours are non-object ones, which are salient but not common, and common but not salient respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

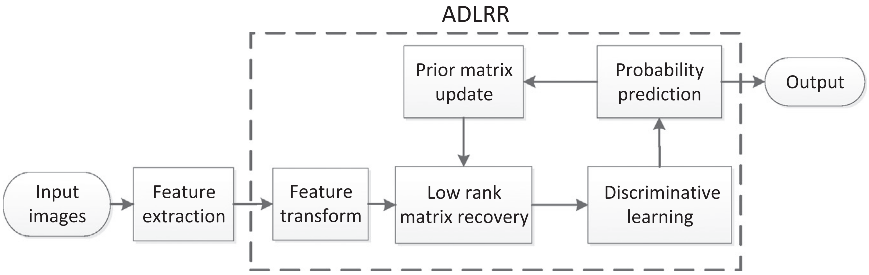


Fig. 2. The flowchart of the proposed framework.

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

process is much more efficient. Vicente [23] modified the energy function in [22] with different background appearance models among images, and it can be used in the unsupervised case. Moreover, it has a simpler optimization procedure compared with [14]. Joulin et al. [16] used a different formulation of the co-segmentation problem. They dealt with the problem in a discriminative clustering framework, in which the clustering step is used to merge image pixels into two clusters while the discriminative learning step is to maximally distinguish the two clusters. Considering the diversity of background, Joulin et al. [19] proposed an improved multi-class co-segmentation method by combining spectral clustering and discriminative clustering. Besides, Mukherjee et al. [17] adopted a direct solution to make the possible object regions in images similar to each other. Since most previous methods are limited to the small dataset, Kim et al. [18] proposed an anisotropic diffusion based method which is able to perform segmentation of a large scale dataset with multiple object classes.

However, most of the unsupervised methods suffers from the consistent background regions without human interaction. Our proposed ADLRR method eliminates the disturbance of background directly via saliency detection, and the discriminative learning step based on the saliency detection result will select common and salient regions, which is usually part of the target object. Meanwhile, our proposed ADLRR method can distinguish the target objects from the background directly rather than the clustering based methods [16,19].

3. Proposed model

We start with the saliency detection via low rank matrix recovery to eliminate the disturbance of those backgrounds consistent with each other. Then we introduce the discriminative learning term and the proposed ADLRR by integrating the two processes together into a unified framework.

The notations in this paper are demonstrated as follows. Given a set of images τ with the same class label, image over-segmentation is performed to each image i by mean-shift clustering based on extracted features including color, Gabor feature and steerable pyramid feature [24], and N_i superpixels are obtained. For the j-th superpixel in the i-th image, we use the mean of the features in this superpixel as its feature representation $f_{ij} \in R^D$, then we get the feature representation of the i-th image $\mathbf{F}_i = [f_{i1}, f_{i2}, \ldots, f_{iN_i}]$. Let $y_i \in [0,1]^{1\times N_i}$ denote the probability vector of superpixels to be foreground in the i-th image. The larger y_{ij} is, the more likely for the j-th superpixel in the i-th image to be target object.

3.1. Low rank matrix recovery

Inspired by the work in [21], we adopt the framework of low rank matrix recovery for salient object detection. For a given image, the background usually lies in a low dimensional space, while the salient regions are usually unique and quite different from the rest. An image is represented as a low-rank matrix plus a sparse noise matrix in the feature space, where the low-rank matrix explains the non-salient regions (or background), and the sparse noise matrix indicates the salient regions. Namely, $\mathbf{F}_i = \mathbf{L}_i + \mathbf{S}_i$, where $\mathbf{L}_i$ is the low rank matrix corresponding to the background and $\mathbf{S}_i$ is the sparse noise matrix corresponding to the salient regions. Since the rank norm and l_0 norm lead to an NP-hard problem and it has been shown that the nuclear norm and the l_1 norm is the tight convex approximation for the rank and the l_0 norm [25]. Thus, we obtain the following convex surrogate:

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min_{\mathbf{L}_i, \mathbf{S}_i} (\|\mathbf{L}_i\|_* + \lambda \|\mathbf{S}_i\|_1)$$
$$\text{s.t.} \quad \mathbf{F}_i = \mathbf{L}_i + \mathbf{S}_i \tag{1}$$

where $\|\cdot\|_*$ is the nuclear norm, which is the sum of the singular values of a given matrix. The l_1-norm of each column S_{ij} in $\mathbf{S}_i$ can be used to measure the saliency of the corresponding superpixel [21]. The larger $\|S_{ij}\|_1$ is, the more likely for the j-th superpixel in the i-th image to be salient.

For the task-dependent problem, it can naturally incorporate some priors to the low rank recovery framework as follows:

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min_{\mathbf{L}_i, \mathbf{S}_i} (\|\mathbf{L}_i\|_* + \lambda \|\mathbf{S}_i\|_1)$$
$$\text{s.t.} \quad \mathbf{F}_i \mathbf{P}_i = \mathbf{L}_i + \mathbf{S}_i \tag{2}$$

where $\mathbf{P}_i = diag(p_{i1}, p_{i2}, \ldots, p_{ij} \ldots, p_{iN_i})$ is a diagonal matrix corresponding to the high level prior such as color or location prior. In our proposed ADLRR method, since objects near the image center are more attractive to people, a Gaussian distribution based on the distance to the image center is chosen as a high level prior initialization to reduce small salient regions near the image edge, then the prior matrix $\mathbf{P}_i$ is updated based on the probability vector y_i of superpixels to be foreground. Since the larger value p_{ij} is, the more likely for the j-th superpixel in the i-th image to be target object, so the update rule is as follows:

$$p_{ij}^{new} = p_{ij}^{old} * \exp\left(\frac{y_{ij} - \mu}{\sigma}\right) \tag{3}$$

where μ and σ are used to control the update rate.

Furthermore, feature representation is a important issue to the performance of the model [21,26,27]. The proposed method works at the assumption that the background usually lies in a low dimensional space, while the salient regions are usually unique and quite different from the rest. However, the assumption may be not satisfied, since the low-level feature is extracted independent of the task. Therefore, a class specific transform matrix is introduced to enhance the salient regions and ensure that the matrix representing the background has low rank. The target function with feature transform is as follows,

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min_{\mathbf{L}_i, \mathbf{S}_i} (\|\mathbf{L}_i\|_* + \lambda \|\mathbf{S}_i\|_1)$$
$$\text{s.t.} \quad \mathbf{T}\mathbf{F}_i \mathbf{P}_i = \mathbf{L}_i + \mathbf{S}_i \tag{4}$$

Unlike [21], where the general feature transform matrix $\mathbf{T}$ is learnt based on the labeled images, our model learns the class specific feature transform matrix $\mathbf{T}$, based on the shared information among images of the specific class.

The learning process is as follows. Given a set of images τ with the same class label, image saliency detection is performed to each image based on Eq. (2). For the j-th superpixel in the i-th image, we use q_{ij} to indicate whether or not the superpixel belongs to the salient regions, q_{ij} is 0 when the corresponding superpixel is salient and 1 when the corresponding superpixel is the background region. Since we do not have labeled images, so we prefer the superpixels with small saliency values to be background. We prefer the 25% percent superpixels with the smallest saliency values to be background for each image during the experiment. Let $\mathbf{Q}_i = diag(q_{i1}, q_{i2}, \ldots, q_{iN_i})$, then $\mathbf{T}\mathbf{F}_i\mathbf{Q}_i$ is corresponding to the background information in the transformed feature space and should have a low rank given a good feature transform matrix $\mathbf{T}$. Therefore, the problem of learning $\mathbf{T}$ can be formulated as follows,

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \frac{1}{\tau} \sum_{i=1}^{\tau} \|\mathbf{T}\mathbf{F}_i\mathbf{Q}_i\|_* - \gamma \|\mathbf{T}\|_*$$
$$\text{s.t.} \quad \|\mathbf{T}\|_F = 1 \tag{5}$$

where τ is the number of images of a given class, and the regularization term $-\gamma\|\mathbf{T}\|_*$ is to avoid the trivial solution where the rank of the feature transform matrix $\mathbf{T}$ becomes arbitrarily small, while the constraint $\|\mathbf{T}\|_F = 1$ is to prevent the

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

transformation **T** from being arbitrarily small. Just like [21], gradient descent approach is adopted to optimize the target function in Eq. (5).

3.2. *Discriminative learning*

The saliency detection is mainly evaluated from the view of a single image. However, it cannot exactly catch the object regions common among images. To this end, a logistic regression based discriminative learning is exploited to predict the probability of each superpixel to be the target object. The objective function is to minimize the following negative log likelihood function:

$$E_D = -\sum_{i=1}^{\tau}\sum_{j=1}^{N_i}[y_{ij}\log(h(f_{ij}))+(1-y_{ij})\log(1-h(f_{ij}))] \tag{6}$$

where $h(f_{ij})=\frac{1}{1+\exp(-\theta^T f_{ij}+b)}$ is the predictive result and θ is the model parameter to be learnt. For the convenience of implementation, we use the augmented parameter vector $\theta^T=[b,\theta^T]$ and the augmented feature vector $f_{ij}^T=[1,f_{ij}^T]$, then we have the reduced form of the logistic regression $h(f_{ij})=\frac{1}{1+\exp(-\theta^T f_{ij})}$.

3.3. *Proposed formulation*

As discussed in Section 1, the ideal object regions as a result of co-segmentation are required to be both salient and common among a given set of images. Thus, we expect, the aforementioned two parts should be learned simultaneously and promote each other. We import a regularization penalty to measure the disagreement between their predicted results. Specifically, the l_1 norm of each column S_{ij} in $\mathbf{S}_i$ stands for the salient score of the j-th superpixel and y_{ij} in discriminative learning is the probability of the j-th superpixel to be target object. Consequently, the disagreement is measured by the following equation:

$$E_R = \sum_{i=1}^{\tau}\sum_{j=1}^{N_i}(y_{ij}-\alpha_i\|S_{ij}\|_1)^2 \tag{7}$$

where α_i is the normalized weight for the superpixel saliency in the i-th image. By jointly exploiting the above aspects, the proposed model is formulated as follows:

$$\begin{aligned}\arg\min_{y_i,\mathbf{L}_i,\mathbf{S}_i,\theta} & \sum_{i=1}^{\tau}(\|\mathbf{L}_i\|_*+\lambda\|\mathbf{S}_i\|_1)-\mu_1\sum_{i=1}^{\tau}\sum_{j=1}^{N_i}[y_{ij}\log(h(f_{ij})) \\ & +(1-y_{ij})\log(1-h(f_{ij}))]+\mu_2\sum_{i=1}^{\tau}\sum_{j=1}^{N_i}(y_{ij}-\alpha_i\|S_{ij}\|_1)^2 \\ \text{s.t.}\quad & \mathbf{T}\mathbf{F}_i\mathbf{P}_i=\mathbf{L}_i+\mathbf{S}_i,\quad i\in\tau\end{aligned} \tag{8}$$

where μ_1 and μ_2 are two non-negative trade-off parameters, and $h(f_{ij})=\frac{1}{1+\exp(-\theta^T f_{ij})}$ is the logistic function.

4. Model optimization

Considering the objective function is a difference of convex functions, we propose an alternate optimization procedure and summarize it in Algorithm 1. Feature transform matrix is first learnt based on gradient descent approach, and details have been introduced in Section 3.1. Given feature transform matrix **T**, the optimization procedure of the target function in Eq. (8) can be divided into three subprocedures. The first is to perform low rank matrix recovery given the guidance information y. The second is to estimate parameter θ in the discriminative learning term based on the saliency detection result. The third is to predict the probability y to be target object based on saliency detection result and discriminative learning together. The three subprocedures are alternately optimized and the algorithm will converge to the local minimum. Details about the optimization procedure will be introduced as follows.

Algorithm 1. Object co-segmentation by ADLRR.

```
Input: Feature Matrix $\mathbf{F}$, high level Prior $\mathbf{P}_i$, $y=0$,
and the required parameters
solve Eq. (2) by the method in [25]
solve Eq. (5) by gradient descent to learn $\mathbf{T}$
while $t < maxIter$ do
    solve problem (9) in Algorithm 2
    $\alpha_i = \frac{1}{\max \|S_{ij}\|_1}$
    $y_{ij} = \alpha_i\|S_{ij}\|_1$
    while not converaged do
        $\theta^{t+1} = \theta^t - \alpha_{step1}[h(f_{ij})-y_{ij}]f_{ij}$
    end while
    while not converaged do
        $y^{t+1} = y^t - \alpha_{step2}[-\mu_1\theta^T\mathbf{F}+2\mu_2(y^t-S)]$
    end while
    update the prior matrix $\mathbf{P}_i$ with Eq. (3)
end while
Output: foreground probability $y$
```

Table 1
Results of ADLRR and some special cases on MSRC-v2 dataset.

Class	Images	ADLRR	DLRRP	DLRRT	DLRR [30]	LRRT	LRR
Bike	30	**53.9**	51.4	52.8	48.8	48.2	44.6
Bird	30	45.5	**46.3**	43.7	44.4	42.2	43.9
Car	30	**57.1**	54.1	56.2	53.3	56.6	55.2
Cat	24	55.3	59.3	56.1	58.6	55.6	**59.5**
Chair	30	**54.4**	50.6	53.5	50.5	53.2	50.4
Cow	30	**66.3**	65.6	64.0	63.8	61.7	61.8
Dog	30	47.9	49.2	46.7	**50.1**	46.1	48.1
Face	30	**53.8**	51.2	53.1	52.1	53.0	52.6
Flower	30	**62.1**	59.3	60.4	56.3	55.6	53.6
House	30	**61.8**	54.1	60.4	51.1	57.4	50.4
Plane	30	**48.0**	47.7	47.5	46.1	46.0	41.3
Sheep	30	**70.1**	68.2	68.1	64.9	66.2	62.8
Sign	30	**65.1**	63.3	64.4	62.4	63.4	61.4
Tree	30	**70.2**	62.7	69.1	57.1	64.3	54.5
Average		**58.0**	55.9	56.9	54.2	55.0	52.9

Table 2
Results of ADLRR and some state-of-the-arts on MSRC-v2 dataset.

Class	Images	ADLRR	DLRR [30]	MCC [19]	CoSand [18]	DCC [16]
Bike	30	**53.9**	48.8	43.3	29.9	42.3
Bird	30	45.5	44.4	**47.7**	29.9	33.2
Car	30	57.1	53.3	**59.7**	37.1	59.0
Cat	24	55.3	**58.6**	31.9	24.4	30.1
Chair	30	**54.4**	50.5	39.6	28.7	37.6
Cow	30	**66.3**	63.8	52.7	33.5	45.0
Dog	30	47.9	**50.1**	41.8	33.0	41.3
Face	30	53.8	52.1	**70.0**	33.2	66.2
Flower	30	**62.1**	56.3	51.9	40.2	50.9
House	30	**61.8**	51.1	51.0	32.2	50.5
Plane	30	**48.0**	46.1	21.6	25.1	21.7
Sheep	30	**70.1**	64.9	66.3	60.8	60.4
Sign	30	**65.1**	62.4	58.9	43.2	55.2
Tree	30	**70.2**	57.1	67.0	60.0	60.0
Average		**58.0**	54.2	50.2	36.5	46.7

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

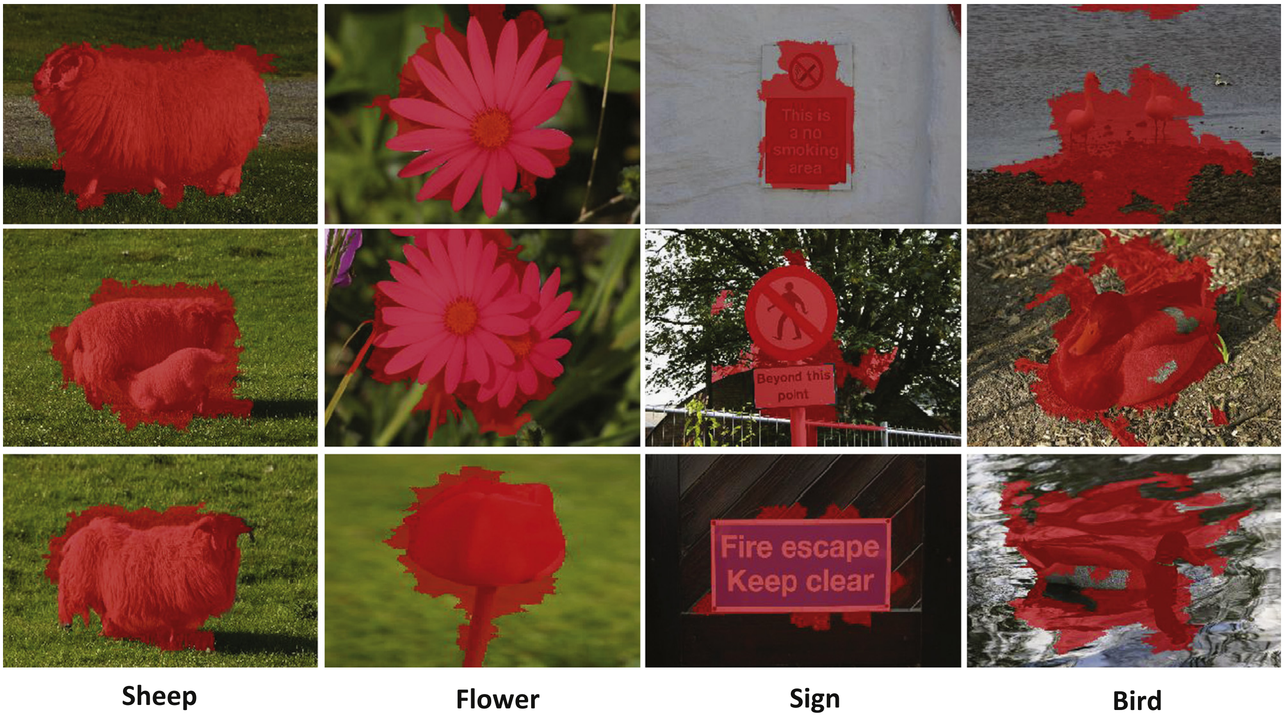


Fig. 3. Some object co-segmentation results of MSRC-v2 dataset.

4.1. *Update $\mathbf{L}_i, \mathbf{S}_i$ as given y_i*

Low rank matrix recovery given the guidance information y_i can be optimized by solving the following augmented Lagrange function:

$$(\mathbf{L}_i^*, \mathbf{S}_i^*) = \arg\min_{\mathbf{L}_i, \mathbf{S}_i} \| \mathbf{L}_i \|_* + \lambda \| \mathbf{S}_i \|_1 + \mathrm{tr}(\mathbf{Z}^T (\mathbf{TF}_i \mathbf{P}_i - \mathbf{L}_i - \mathbf{S}_i))$$
$$+ \frac{\beta}{2} \| \mathbf{TF}_i \mathbf{P}_i - \mathbf{L}_i - \mathbf{S}_i \|_F^2 + \mu_2 \sum_{j=1}^{N_i} (y_{ij} - \alpha_i \| S_{ij} \|_1)^2 \quad (9)$$

where $\mathrm{tr}(\cdot)$ is the trace of a matrix, $\mathbf{Z}$ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. The inexact ALM method in [28] is used for efficiency and outlined in Algorithm 2, and $T_\varepsilon[\cdot]$ is the soft-thresholding (shrinkage) operator defined as follows:

$$T_\varepsilon[x] = \begin{cases} x - \varepsilon & \text{if } x > \varepsilon, \\ x + \varepsilon & \text{if } x < -\varepsilon, \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

4.2. *Update θ as given $\mathbf{L}_i, \mathbf{S}_i, y_i$*

Parameters in the discriminative learning term can be estimated based on the saliency detection result. First, the probability y_i to be target object can be got by setting the regularization penalty Eq. (7) to be 0, namely $y_{ij} = \alpha_i \| S_{ij} \|_1$. Then, the parameters in logistic function can be achieved by stochastic gradient descent rule as follows:

$$\theta^{t+1} = \theta^t - \alpha_{step1} [h(f_{ij}) - y_{ij}] f_{ij} \quad (11)$$

where α_{step1} controls the update rate.

Algorithm 2. Solving problem (9) by inexact ALM.

```
Input: matrix $\mathbf{T}$, $\mathbf{F}_i$, $\mathbf{P}_i$, $y_i$; parameters $\lambda$, $\beta$, $\mu_2$, $\alpha_i$, $D$
Initialize: $\mathbf{Z}^0 = \mathbf{TF}_i\mathbf{P}_i / J(\mathbf{TF}_i\mathbf{P}_i)$; $\mathbf{S}^0 = \mathbf{0}$; $\beta^0 > 0$; $\rho = 1.5$; $k = 0$
while not converged do
    $(\mathbf{U}, \Sigma\mathbf{V}) = \mathrm{svd}(\mathbf{TF}_i\mathbf{P}_i - \mathbf{S}_i^k + (\beta^k)^{-1}\mathbf{Z}^k)$
    $\mathbf{L}_i^{k+1} = \mathbf{U} T_{(\beta^k)^{-1}} [\Sigma] \mathbf{V}^T$
    $\varepsilon(m,n) = \frac{\lambda - 2\alpha_i\mu_2 y_{in} + 2\alpha_i^2\mu_2 \sum_{t=1, \neq m}^{D} |S_i(t,n)|}{\beta + 2\alpha_i^2\mu_2}$
    $\mathbf{x} = \frac{\beta}{\beta + 2\alpha_i^2\mu_2} (\mathbf{TF}_i\mathbf{P}_i - \mathbf{L}_i^{k+1} + (\beta^k))^{-1}\mathbf{Z}^k)$
    $S_i^{k+1}(m,n) = T_{\varepsilon(m,n)}[x(m,n)]$
    $\mathbf{Z}^{k+1} = \mathbf{Z}^k + \beta^k (\mathbf{TF}_i\mathbf{P}_i - \mathbf{L}_i^{k+1} - \mathbf{S}_i^{k+1})$
    $\beta^{k+1} = \min(\rho\beta^k, \beta_{max})$
    $k \leftarrow k+1$
end while
Output: $(\mathbf{L}_i^k, \mathbf{S}_i^k)$
```

4.3. *Update y_i as given $\mathbf{L}_i, \mathbf{S}_i, \theta$*

Probability to be target object can be optimized directly by gradient descent based on saliency detection result and discriminative learning together, and for notation simplicity, we denote $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_\tau]$, $y = [y_1, \dots, y_\tau]$, and $S = [\alpha_1 \| S_{11} \|_1, \dots, \alpha_\tau \| S_{\tau N_\tau} \|_1]$. The update rule is as follows:

$$y^{t+1} = y^t - \alpha_{step2} [-\mu_1 \theta^T \mathbf{F} + 2\mu_2 (y^t - S)] \quad (12)$$

where α_{step2} controls the update rate.

5. Experiments and results

To validate the effectiveness of the proposed method, we conduct experiments on three publicly available benchmarks, i.e., MSRC-v2,[1] iCoseg [22] and Caltech101 [29]. In our experiments, to validate the effectiveness of the proposed ADLRR, we compare it with its special cases and a number of related state-of-the-art approaches, which are enumerated as follows.

[1] http://research.microsoft.com/en-us/projects/ObjectClassRecognition/

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

- *LRR*: It is to use the saliency detection result as the final object probability, and the saliency detection process is based on low rank matrix recovery [21].
- *LRRT*: It is to enhance LRR with feature transform.
- *DLRR* [30]: It stands for the discriminative low rank matrix recovery method, which integrates saliency detection and discriminative learning into a unified framework. It was proposed in our previous work [30].
- *DLRRT*: It is to enhance DLRR with feature transform.
- *DLRRP*: It is to enhance DLRR with prior matrix update.
- *ADLRR*: The proposed ADLRR is to enhance DLRR with feature transform and prior matrix update.
- *DCC* [16]: It deals with the co-segmentation problem in a discriminative clustering framework.
- *MCC* [19]: It proposes a multi-class co-segmentation method by combining spectral clustering and discriminative clustering.
- *CoSand* [18]: It proposes an anisotropic diffusion based method which is able to perform segmentation of a large scale dataset with multiple object classes.

Table 3
Results of ADLRR and some special cases on iCoseg dataset.

Class	Images	ADLRR	DLRRP	DLRRT	DLRR [30]	LRRT	LRR
Baseball	25	**68.3**	64.9	63.8	62.8	57.6	55.5
Football	33	38.6	38.5	**39.4**	39.3	34.5	35.0
Monk	17	**44.8**	44.1	44.5	40.4	36.8	35.7
Brown bear	5	42.1	43.2	43.2	39.5	**43.4**	43.3
Ferrari	11	**59.9**	57.4	58.3	54.5	52.3	49.7
Skating	11	56.6	**59.8**	51.6	54.2	42.9	46.5
Alaskan bear	19	52.1	**52.4**	50.3	41.6	44.7	44.5
Taj Mahal	5	**53.3**	47.6	52.2	46.6	48.5	41.2
Helicopter	12	64.3	64.3	63.0	62.4	64.0	**64.4**
Kite	18	46.5	**51.5**	45.2	45.8	48.2	48.1
Average		**52.6**	52.4	51.2	48.7	47.3	46.4

Table 4
Results of ADLRR and some state-of-the-arts on iCoseg dataset.

Class	Images	ADLRR	DLRR [30]	MCC [19]	CoSand [18]	DCC [16]
Baseball	25	**68.3**	62.8	13.6	56.7	31.4
Football	33	38.6	39.3	38.7	**40.0**	14.9
Monk	17	44.8	40.4	**73.8**	70.5	68.4
Brown bear	5	42.1	39.5	**57.5**	39.2	49.4
Ferrari	11	59.9	54.5	38.7	**61.2**	26.4
Skating	11	56.6	54.2	**72.7**	32.1	38.1
Alaskan bear	19	52.1	41.6	41.6	31.9	46.1
Taj Mahal	5	53.3	46.6	37.1	**72.8**	38.4
Helicopter	12	**64.3**	62.4	33.3	12.3	61.0
Kite	18	46.5	45.8	22.1	9.3	**57.8**
Average		**52.6**	48.7	42.9	42.6	43.2

The visual features used for over-segmentation include color features, Gabor features, and steerable pyramid features, which are same to the work in [21]. The segmentation performance is measured by the *intersection-over-union* score and defined by $\frac{1}{|\tau|}\sum_{i\in\tau}\frac{R_i\cap GT_i}{R_i\cup GT_i}$, where R_i is the segmentation result of image i and GT_i is the ground truth. This evaluation metric is standard in PASCAL challenges. Generally, the trade-off parameters in target formulation are set as follows, $\mu_1=0.05$, $\mu_2=1$, and the parameters in update rule are set as follows, $\mu=1$, $\sigma=1$. Specially, since the iCoseg data has relatively high class coherence and small number of images, we choose $\mu_1=0.08$ to increase the effect of the discriminative learning term and $\sigma=2$ to make the system more robust.

5.1. MSRC-v2 dataset

Extensive experiments are conducted on the MSRC-v2 dataset for 14 classes, and each class contains 30 images, except that the 'cat' class contains 24 images. Results of the proposed ADLRR and the special cases are shown in Table 1, We can find that LRRT outperforms LRR by 3.1%, and DLRRT outperforms DLRR by 2.7%, so feature transform is effective to enhance the salient regions and ensure that the background has low rank. We can also find that DLRR outperforms LRR by 1.3%, and DLRRT outperforms LRRT by

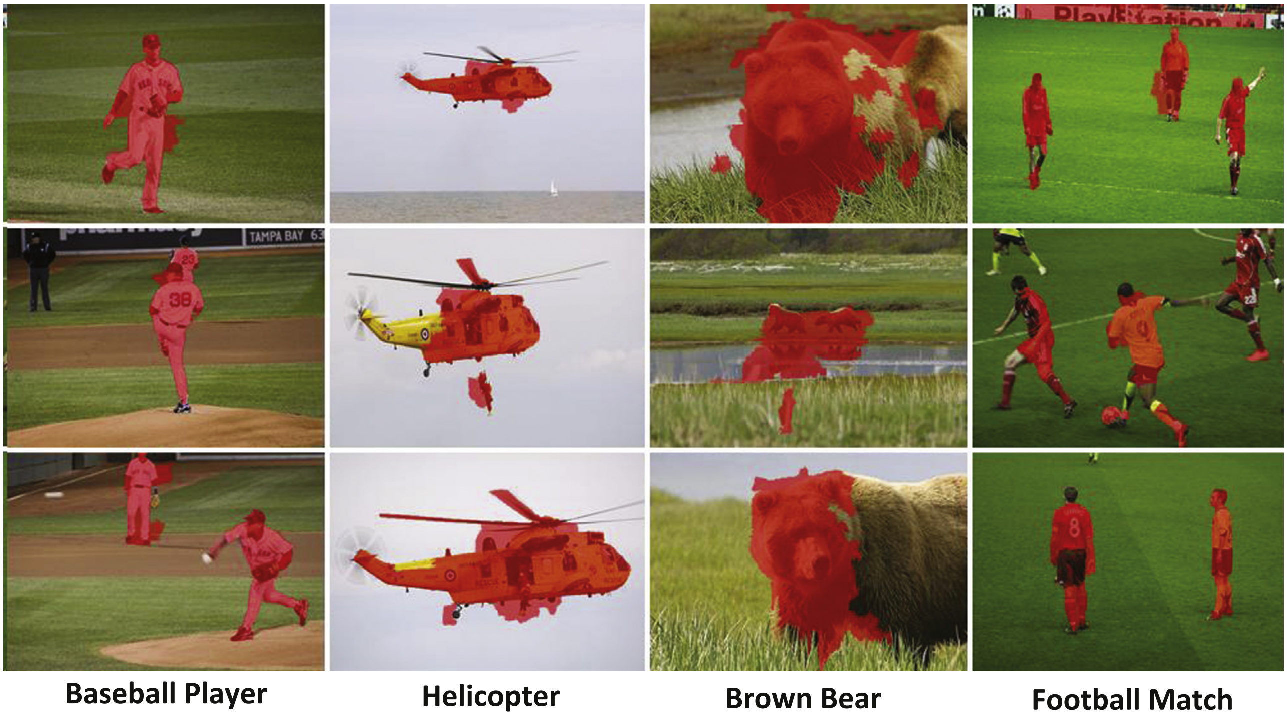

Baseball Player
Helicopter
Brown Bear
Football Match

Fig. 4. Some object co-segmentation results of iCoseg.

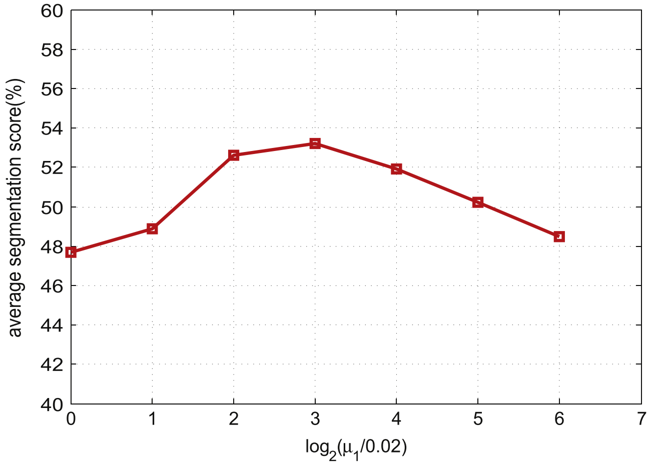


Fig. 5. Effect of discriminative learning with different values of the tradeoff parameter μ_1 on the iCoseg dataset.

1.9%, so the unified framework combining saliency detection with discriminative learning is useful to discover the common and salient regions. Furthermore, DLRRP outperforms DLRR by 1.7%, which demonstrates the reasonability to update the weight of superpixels based on the foreground probability during the low rank recovery process. Generally, ADLRR with feature transform and prior matrix update, shares the advantages of DLRRT and DLRRP. The proposed ADLRR outperforms all the special cases in terms of the average performance, and achieves the best performance for 11 of 14 classes.

The comparison of different methods is listed in Table 2, where we directly cite the results of [16,19,18] reported in [19]. From the results in Table 2, the proposed ADLRR achieves the best performance for 9 of 14 classes and outperforms all the compared methods in terms of the average performance. The average segmentation score increases 7.8% relatively to MCC [19], 11.3% relatively to DCC [16] and 21.5% relatively to CoSand [18]. In addition, our method is more robust than the others to the class changes. Compared with DLRR proposed in our previous work [30], ADLRR improves obviously due to feature transform and prior matrix update.

Fig. 3 presents some object co-segmentation results of the proposed method in the MSRC-v2 dataset, in which the successful and unsuccessful examples are both shown. ADLRR improves obviously for the classes with high class coherence (e.g. tree, flower, house, sheep), and the failed results (e.g. bird) may due to the high diversity of the object appearance.

5.2. iCoseg dataset

The iCoseg dataset is first collected to do image co-segmentation with human interaction with 38 groups, 643 images [22]. Since we mainly focus on the segmentation of the common object regions from the background, the class number in [19,16] is set to 2, and we use the publicly versions of [19,16] to get the co-segmentation results and prefer the regions with larger overlap with ground truth as the target object regions. Furthermore, we vary the class number from 2 to 5 and report the best result for the multiple foreground segmentation method [18].

From the results in Table 3, the same conclusion can be got just like the MSRC-v2 dataset. LRRT outperforms LRR by 0.9% and DLRRT outperforms DLRR by 2.5%, which demonstrate the effectiveness of feature transform. DLRRP outperforms DLRR by 3.7% which demonstrates the effectiveness of prior matrix update. Generally, ADLRR with feature transform and prior matrix update shares the advantages of DLRRT and DLRRP. The proposed ADLRR achieves the best performance for 5 of 10 classes.

Table 4 gives a quantitative comparison with [16,19,18,30] on the iCoseg dataset. We can find that ADLRR achieves the best performance for 4 of 10 classes, and the average segmentation score increases 9.7% relatively to MCC [19], 9.4% relatively to DCC [16] and 10.0% relatively to CoSand [18]. Conclusions conducted on MSRC-v2 dataset are further verified, ADLRR outperforms all the compared methods in terms of the average performance. Compared with DLRR proposed in our previous work [30], ADLRR improves obviously due to the feature transform and prior matrix update.

Fig. 4 presents some object co-segmentation results of the proposed method in the iCoseg dataset, in which the successful and unsuccessful examples are both shown. The brown bear class fails, since the target object of the brown bear class is not salient enough. Meanwhile, the football match class fails due to the fact of

Table 5
Results of ADLRR and some special cases on Caltech101 dataset.

Class	Images	ADLRR	DLRRP	DLRRT	DLRR [30]	LRRT	LRR
Ant	42	47.9	**48.9**	46.3	46.5	46.0	45.9
Bass	54	54.7	53.5	**55.3**	54.1	54.6	53.5
Butterfly	91	**61.8**	60.2	61.8	60.6	61.1	60.0
Cannon	43	54.1	52.0	**54.3**	53.1	54.2	52.6
Cellphone	59	**49.0**	43.8	48.9	42.6	47.1	41.3
Chair	62	59.5	58.7	60.4	60.4	**60.6**	60.0
Cougar body	47	**53.7**	52.7	52.8	52.2	49.7	48.8
Dolphin	65	44.2	**44.3**	43.0	42.6	42.7	42.2
Ferry	67	**53.5**	53.1	51.7	51.4	51.1	50.4
Flamingo	67	**48.2**	47.9	45.9	45.7	45.1	44.7
Hedgehog	54	**54.1**	47.8	52.8	48.4	52.7	48.9
Helicopter	88	42.2	**43.3**	40.1	40.2	40.2	40.3
Ibis	80	39.1	**40.4**	37.5	37.7	38.0	37.7
Joshua tree	64	**53.3**	52.9	52.2	51.3	50.9	48.9
Lotus	66	**53.4**	51.6	52.4	51.4	51.6	50.3
Okapi	39	53.9	**55.4**	53.1	54.5	52.8	53.8
Rhino	59	47.7	45.6	47.7	46.6	**49.4**	48.3
Schooner	63	56.4	54.6	58.0	57.4	**59.6**	58.6
Sea horse	57	**40.8**	40.3	39.5	38.9	38.8	38.4
Sunflower	85	**57.5**	55.3	56.7	54.8	54.7	52.8
Tick	49	**62.3**	**62.3**	60.9	60.9	60.3	60.3
Water lilly	37	**52.6**	51.9	48.5	48.0	45.8	45.0
Wild cat	34	**54.4**	51.5	53.5	51.6	50.4	47.6
Average		**51.9**	50.8	51.0	50.0	50.3	49.1

Table 6
Results of ADLRR and some state-of-the-arts on Caltech101 dataset.

Class	Images	ADLRR	DLRR [30]	MCC [19]	CoSand [18]	DCC [16]
Ant	42	**47.9**	46.5	33.3	19.4	32.5
Bass	54	**54.7**	54.1	27.1	23.4	29.1
Butterfly	91	**61.8**	60.6	44.3	23.5	32.5
Cannon	43	**54.1**	53.1	39.3	35.9	33.0
Cellphone	59	**49.0**	42.6	26.2	19.9	29.4
Chair	62	59.5	**60.4**	42.6	23.9	40.6
Cougar body	47	**53.7**	52.2	33.3	25.4	32.6
Dolphin	65	**44.2**	42.6	28.6	13.9	31.8
Ferry	67	**53.5**	51.4	18.4	24.2	19.2
Flamingo	67	**48.2**	45.7	39.8	14.6	43.6
Hedgehog	54	**54.1**	48.4	33.8	35.0	30.2
Helicopter	88	**42.2**	40.2	22.3	15.9	21.9
Ibis	80	**39.1**	37.7	23.4	14.9	24.8
Joshua tree	64	**53.3**	51.3	29.8	29.2	29.2
Lotus	66	53.4	51.4	63.9	44.8	**64.6**
Okapi	39	53.9	**54.5**	40.2	23.0	40.2
Rhino	59	**47.7**	46.6	36.1	30.0	35.9
Schooner	63	56.4	**57.4**	35.3	23.3	26.5
Sea horse	57	**40.8**	38.9	25.3	16.9	31.1
Sunflower	85	57.5	54.8	**73.1**	31.7	72.1
Tick	49	**62.3**	60.9	44.9	23.2	41.8
Water lilly	37	52.6	48.0	47.8	12.8	**68.7**
Wild cat	34	**54.4**	51.6	25.3	38.5	25.1
Average		**51.9**	50.0	36.3	24.5	36.4

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

multiple foreground objects, and the multiple foreground objects occur among numbers of images. So it is difficult to separate the special kind of object from multiple foreground objects without further supervised information, and the multiple foreground methods cannot work well either, without a exact foreground number which needs to be specially chosen by human. We can also find that our proposed ADLRR can exactly extract all the common and salient objects from the background for the football match class.

Furthermore, the effect of discriminative learning with different values of the tradeoff parameter is shown in Fig. 5. The range of μ_1 is set with exponential growth. We can find that the performance of the proposed method gets better as the influence of discriminative learning term becomes larger, which demonstrates the effectiveness of discriminative learning term. When μ_1 is above a certain threshold, the performance decreases since discriminative learning term dominates final object function. While, the low rank matrix recovery term and the discriminative learning term will benefit from each other when μ_1 is in a reasonable range. Generally, the proposed method is robust, and the average segmentation accuracy is larger than 50% in a wide change of μ_1.

5.3. Caltech101 dataset

The Caltech101 database contains over 9000 images from 102 classes. 101 classes are of animals, flowers, trees, etc., and there is a background class. The number of images in each class is between 31 and 800. we perform experiments on 23 classes with relatively complex background. The comparison methods [16,19,18] are performed with the same set as the iCoseg dataset.

From the results in Table 5, the same conclusion can be drawn just like the above two datasets. LRRT outperforms LRR by 1.2% and DLRRT outperforms DLRR by 1.0%, which demonstrate the effectiveness of feature transform. DLRRP outperforms DLRR by 0.8%, which demonstrates the effectiveness of prior matrix update. Generally, ADLRR with feature transform and prior matrix update shares the advantages of DLRRT and DLRRP. The proposed ADLRR achieves the best performance for 13 of 23 classes.

Table 6 gives a quantitative comparison with [19,18,16,30] on the caltech101 dataset. We can find that ADLRR achieves the best performance for 17 of 23 classes, and the average segmentation score increases 15.6% relatively to MCC [19], 15.5% relatively to DCC [16] and 27.4% relatively to CoSand [18]. Conclusions drawn from MSRC-v2 dataset and iCoseg dataset are further verified, and ADLRR outperforms all the comparison methods in terms of the average performance. Compared with DLRR proposed in our previous work [30], ADLRR improves obviously due to feature transform and prior matrix update.

The proposed ADLRR improves obviously for the classes with high class coherence (e.g., butterfly, tick, joshua tree), while high diversity or target object with low saliency may lead to the failure of ADLRR (e.g., ibis, rhino and schooner). Fig. 6 presents some object co-segmentation results.

6. Conclusions

In this paper, a novel adaptive discriminative low rank matrix recovery (ADLRR) algorithm is proposed to perform object co-segmentation. Our method works on the assumption that object

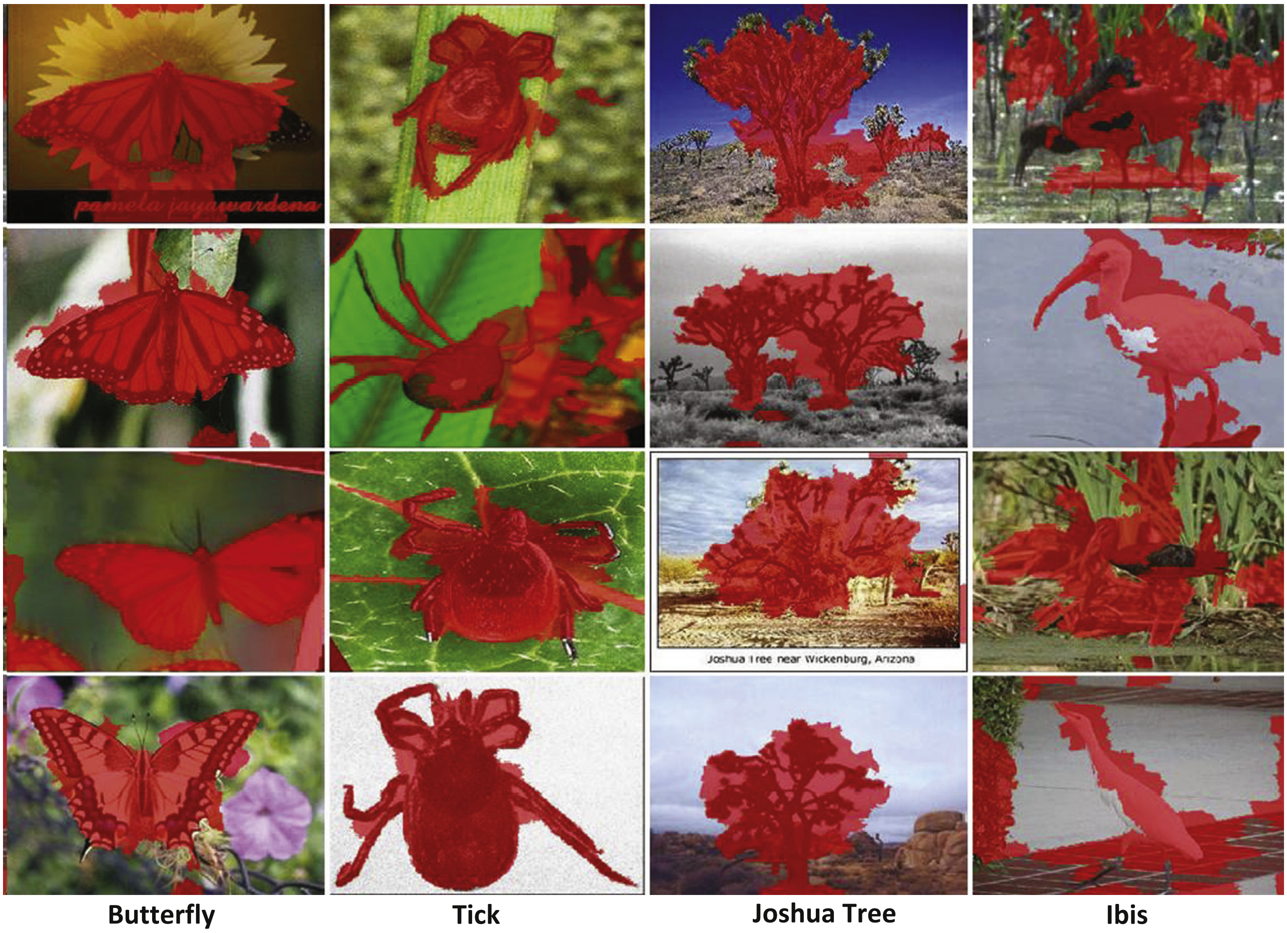


Fig. 6. Some object co-segmentation results of Caltech101.

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110

regions should be not only common but also salient ones among images. This is the first to be used in the task of object co-segmentation. We import the low rank matrix recovery term to measure the saliency of super-pixels so as to eliminate the disturbance from those consistent backgrounds. While a discriminative learning term is used to model the true object regions simultaneously. Besides, a regularized penalty is employed to promote both terms each other. Furthermore, class specific feature transform is imported to enhance the salient regions and ensure the matrix representing the background has low rank, and prior matrix is updated to heighten the regions with high probability to be foreground. an efficient alternate optimization procedure is designed to solve the proposed formulation. Extensive experiments have shown the outperforming performance compared with some state-of-the-arts.

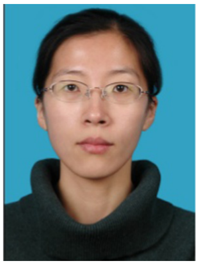
Acknowledgements

This work was supported by 863 Program (2014AA015104), National Natural Science Foundation of China ((61272329), (61472422) and (61402228)) and Open Projects Program of National Laboratory of Pattern Recognition.

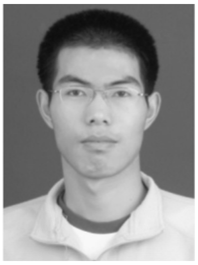
References

[1] C. Rother, V. Kolmogorov, A. Blake, Grabcut–interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (3) (2004) 309–314.
[2] Y. Liu, J. Liu, Z. Li, J. Tang, H. Lu, Weakly-supervised dual clustering for image semantic segmentation, in: CVPR, 2013.
[3] J. Liu, M. Li, Q. Liu, H. Lu, S. Ma, Image annotation via graph learning, Pattern Recognit. 42 (2) (2009) 218–228.
[4] J. Liu, B. Wang, M. Li, Z. Li, W. Ma, H. Lu, S. Ma, Dual cross-media relevance model for image annotation, in: ACM Multimedia, 2007, pp. 605–614.
[5] Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, X.-S. Hua, Visual query suggestion: towards capturing user intent in internet image search, ACM Trans. Multimed. Comput. Commun. Appl. 6 (3) (2010) 13:1–13:19.
[6] Z.-J. Zha, M. Wang, Y.-T. Zheng, Y. Yang, R. Hong, T.-S. Chua, Interactive video indexing with statistical active learning, IEEE. Trans. Multimed. 14 (1) (2012) 17–27.
[7] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: CVPR, 2010, pp. 73–80.
[8] Y.Y. Boykov, M. P. Jolly, Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images, in: ICCV, 2001, pp. 105–112.
[9] D. Kuettel, V. Ferrari, Figure-ground segmentation by transferring window masks, in: CVPR, 2012, pp. 558–565.
[10] C. Desai, D. Ramanan, C. Fowlkes, Discriminative models for multi-class object layout, in: ICCV, 2009, pp. 229–236.
[11] H. Harzallah, F. Jurie, C. Schmid, Combining efficient object localization and image classification, in: ICCV, 2009, pp. 237–244.
[12] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
[13] C.-F. Juang, W.-K. Sun, G.-C. Chen, Object detection by color histogram-based fuzzy classifier with support vector learning, Neurocomputing 72(10–12) (2009) 2464–2476.
[14] C. Rother, T. Minka, A. Blake, V. Kolmogorov, Cosegmentation of image pairs by histogram matching – incorporating a global constraint into mrfs, in: CVPR, 2006, pp. 993–1000.
[15] V. Mukherjee, L. Singh, C. R. Dyer, Half-integrality based algorithms for cosegmentation of images, in: CVPR, 2009, pp. 2028–2035.
[16] A. Joulin, F. Bach, J. Ponce, Discriminative clustering for image co-segmentation, in: CVPR, 2010, pp. 1943–1950.
[17] L. Mukherjee, V. Singh, J. Peng, Scale invariant cosegmentation for image groups, in: CVPR, 2011, pp. 1881–1888.
[18] G. Kim, E. Xing, L. Fei-Fei, T. Kanade, Distributed cosegmentation via submodular optimization on anisotropic diffusion, in: ICCV, 2011, pp. 169–176.
[19] A. Joulin, F. Bach, J. Ponce, Multi-class cosegmentation, in: CVPR, 2012, pp. 542–549.
[20] S. Vicente, C. Rother, V. Kolmogorov, Object cosegmentation, in: CVPR, 2011, pp. 2217–2224.
[21] X. Shen, Y. Wu, A unified approach to salient object detection via low rank matrix recovery, in: CVPR, 2012, pp. 853–860.
[22] D. Batra, A. Kowdle, D. Parikh, J. Luo, T. Chen, Interactively co-segmentating topically related images with intelligent scribble guidance, Int. J. Comput. Vis. 93 (3) (2011) 273–292.
[23] S. Vicente, V. Kolmogorov, C. Rother, Cosegmentation revisited: models and optimization, in: ECCV, 2010, pp. 465–479.
[24] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
[25] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma, Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization, in: NIPS, 2009, pp. 2080–2088.
[26] Z. Li, J. Liu, H. Lu, Sparse constraint nearest neighbour selection in cross-media retrieval, in: ICIP, 2010, pp. 1465–1468.
[27] C. Li, Q. Liu, J. Liu, H. Lu, Ordinal regularized manifold feature extraction for image ranking, Signal Process. 93 (6) (2013) 1651–1661.
[28] Z. Lin, M. Chen, Y. Ma, The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices, UIUC Technical Report UILU-ENG-09-2214.
[29] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in: Computer Vision and Pattern Recognition Workshop, 2004, pp. 178.
[30] Y. Li, J. Liu, Z. Li, Y. Liu, H. Lu, Object co-segmentation via discriminative low rank matrix recovery, in: ACM Multimedia, 2013, pp. 749–752.

Yong Li received the B.E. degree from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011. He is currently pursuing the Ph.D. degree at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include image content analysis, multimedia, and subspace learning. He received the Microsoft young fellowship in 2010.



Jing Liu received the B.E. and M.E. degrees from Shandong University, Shandong, in 2001 and 2004, respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2008. She is an Associate Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Her current research interests include machine learning, image content analysis and classification, multimedia. She published over 30 papers in those areas.



Zechao Li received the B.E. degree from University of Science and Technology of China (USTC), Anhui, China, in 2008, and the Ph.D. degree from National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences in 2013. He is an Assistant Professor with School of Computer Science, Nanjing University of Science and Technology. His current research interests include multimedia analysis and understanding, subspace learning, correlation mining, etc. He received the 2013 President Scholarship of Chinese Academy of Science.

Hanqing Lu received his B.S. and M.S. from Department of Computer Science and Department of Electric Engineering in Harbin Institute of Technology in 1982 and 1985. He got his Ph.D. from Department of Electronic and Information Science in Huazhong University of sciences and Technology. He is a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Current research interests include Image similarity measure, Video Analysis, Multimedia Technology and System. He published over 300 papers in those areas.

Songde Ma received his B.S. in Automatic Control from the Tsinghua University in 1968, Ph.D. degree in University of Paris in 1983 and "Doctor at d'Etat es Science" in France in 1986 in image processing and computer vision. He is a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, image understanding and searching, robotics and computer graphics, etc.

Please cite this article as: Y. Li, et al., Object co-segmentation via salient and common regions discovery, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.12.110