Received September 19, 2017, accepted October 11, 2017, date of publication October 20, 2017,
date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2764503

Two Birds With One Stone: A Unified Approach to Saliency and Co-Saliency Detection via Multi-Instance Learning

HONGLIN QUAN[1,2], SONGHE FENG[1,2], AND BAIFAN CHEN[3]

[1]The Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China
[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[3]School of Information Science and Engineering, Central South University, Changsha 410083, China

Corresponding author: Baifan Chen (chenbaifan@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61472028, Grant 61403423, and Grant 61673048, in part by the Beijing Natural Science Foundation under Grant 4162048, in part by the Fundamental Research Funds for the Central Universities under Grant 2017JBZ108, and in part by the Open Research Fund of Key Laboratory of Network Crime Investigation of Hunan Provincial Colleges under Grant 2016WLZC017.

ABSTRACT Saliency detection on an individual image as well as co-saliency detection from a group of images are currently popular topics or reflect future trends more recently due to their importance and challenging roles in computer vision. In many cases, co-saliency detection is usually dependent on the single-image saliency detection results. Nevertheless, most efforts have been made to tackle them separately and not much attention has been paid to tackling them together in a unified idea. Being aware of these two tasks are highly related, the difference from previous surveys is that this paper applies a unified framework by employing a multi-instance learning (MIL) algorithm to resolve both the issues, and formulating single-image saliency and co-saliency detection as top-down weakly supervised learning paradigm. Specifically, for single-image saliency detection, we utilize the evidence confidence-support vector machine algorithm to learn a discriminant model to predict the saliency on test images. For co-saliency detection from image group, we concatenate the EC values and saliency scores to obtain the final results of co-saliency detection. By observing the importance of the selection of negative bags in the MIL framework, we also introduce a novel selection strategy of negative bags to improve the robustness of the proposed method. Experimental results on publicly available image benchmark data sets have demonstrated that the proposed unified framework can achieve competitive performances as compared with the state-of-the-art algorithms in terms of accuracy and effectiveness.

INDEX TERMS Co-saliency analysis, diverse density, evidence confidence-support vector machine (SVM), multi-instance learning, saliency analysis, sparse representation.

I. INTRODUCTION

A. BACKGROUND

Recent years have witnessed the emergence of an explosive growth of digital information lead by huge amount of images. How to identify a subset of vital visual information from these images, i.e., visual saliency detection, has emerged as a hot topic. Considering the difference of the visual saliency computation process, we can devide the existing methods into two major classes, namely, bottom-up and top-down, respectively. The former manner is a purely data-driven method, which can be used to simulate the visual perception system without the need of combining prior knowledge for saliency detection. Most bottom-up models focused on contrast-based information that combined various different features by using the principle of feature integration(e.g., [17], [32]–[34]). However, in many cases, it can only get better results under good conditions and tends to perform poorly when the interference increase. The latter manner (e.g., [35]–[37]) belongs to task-driven model, which usually exploits high-level perception knowledge to guide the saliency detection.

Apart from aforementioned approaches of single image saliency detection, there also exist several methods of co-saliency detection in recently years. For instance, to address the problem of a large instance space, Fu et al. [26] applied adaptive instance selection strategy. Cao *et al.* [41] proposed a new method according to reconstruction error, which suppressed the background and the error foreground. Zhang *et al.* [42] used a novel framework integrated the information of deep into the information of wide that could help to locate the salient object accurately and suppress the influence of background. Han *et al.* [43] integrated deep learning into co-saliency detection to discover the consistency of internal structure of co-salient objects.

B. MOTIVATION

Although the bottom-up methods discussed above could get relatively discriminative results, they still had some drawbacks. On the one hand, they only concentrate on using single idea to address the problem of either the single image saliency detection or the co-saliency detection, rather than using a unified idea to focus on solving the two aspects of problems simultaneously. On the other hand, not much attention has been paid to formulating saliency detection from the perspective of multi-instance learning(MIL), but some efforts in this direction was proposed over recent years (e.g., [28], [27], [25]).

In addition, currently, co-saliency detection used evidence confidence evidence confidence values with saliency information to get the final results without the testing phase, which means that selecting a discriminative negative bag plays a significant role in MIL. If the selected negative bags are not discriminative, the final results are vague and the effect of method is relatively poor.

In order to solve the first problem, we propose a top-down visual saliency model combined with MIL, which can be adopted on single image saliency detection and co-saliency detection simultaneously. For single image saliency detection, we use the framework to get a discriminative classifier that is used to distinguish the tags of images in the testing phase. For co-saliency detection, we improve the initial diverse density algorithm(DD) that integrate single image saliency to fit the essence of co-saliency definition.

As for the second problem, we propose a novel selection strategy of negative bags. Specifically, we use the method of sparse representation to get a reconstruction coefficient vector, which can help us select negative bags according to the order of descending.

C. MAIN CONTRIBUTION

In summary, three major contributions can be drawn as follows:

(1) We present a unified MIL framework that can be utilized for both single image saliency detection and co-saliency detection.

(2) A novel strategy of negative bags selection is proposed in co-saliency detection to suppress the similar background information in the positive bags.

(3) A Saliency Diverse Density (SDD) algorithm derived from the DD algorithm is proposed for co-saliency detection, which perfectly fits the essence of co-saliency definition (i.e., saliency + repeatness).

The rest of the paper first describe some previous and related literature in Section 2. We then introduce the unified model framework in detail in Section 3. In Section 4, the experimental results of different methods are analyzed and compared. Finally, conclusions are drawn in the last section.

II. RELATED WORKS

In this section, we briefly review the related work and give the rich literature on three subjects, which are single image saliency detection, co-saliency detection and multi-instance learning respectively.

A. SINGLE IMAGE SALIENCY DETECTION

Single image saliency detection mainly aims at finding salient object from one image and obtaining the corresponding saliency map. To the best of our knowledge, the work of single image saliency detection adopted by Shen and Wu [11] used a new framework to detect saliency combined with low rank matrix by integrating the features of low-level and prior information. Yang and Yang [13] proposed a model according to a Conditional Random Field(CRF), and then, based on CRF and sparse coding, they obtained a classifier that was trained by max-margin approach. Achanta *et al.* [14] used color and luminance as features to detect the saliency of images by applying the proposed frequency-tuned approach. Instead of using contrast between the foreground and the background, Yang *et al.* [29] used graph-based manifold ranking method to rank the coherence of the image elements.

B. CO-SALIENCY DETECTION

The major bottleneck of single image saliency detection is the lack of ability to solve the problem of detecting the co-salient object from a group of images. As a result, co-saliency detection had became a hot topic in the field of computer vision and was widely used in many fields, such as co-segmentation, target co-recognition and so on.

Several existing methods have proposed an effective way for co-saliency detection. Huang *et al.* [3] applied multi-scale pyramid and each scale of images to detect the single image saliency. At last, the final results were obtained by fusing the single image saliency and prior of co-saliency detection. Chen and Hsu [4] improved rank-sparsity decomposition by using a rank-sparsity model that made the factors between sparse and low-rank become the score of saliency maps. Different from other methods, this model used the uncertainty of observations of people and made the factors between sparse and low-rank become the score of saliency maps. Firstly, saliency maps for each image were calculated. Secondly, for each local region, saliency values were calculated. Finally, saliency areas could be found accurately by matching other common images. The method proposed by Meng *et al.* [45] was positioned to take advantage of both geometrical relationship and graph matching.

C. MULTI-INSTANCE LEARNING

Generally, the methods of MIL aim at learning the models from the data with label ambiguities (referred to as weakly supervised data), in which the training instances are from the positive and negative bags. MIT started with a paradigm in supervised learning proposed by Dietterich [31]. Several significant approaches had been developed since then. Maron and Lozano-Prez [2] proposed a MIL method based on the DD algorithm. The core idea of it was to transform the task of MIL into finding the point that had the maximum value of diversity density in the feature space. Hence, the task of MIL can be transformed into finding the point that had the maximum value of diversity density in the feature space. Zhang and Goldman [38] proposed a new method named expectation-maximization-DD(EM-DD), which transformed multi-instance data into single one. Later, Andrews *et al.* [39] put forward an approach named multiple instance-support vector machine (MI-SVM) to address the problems of MIL with the method of single instance. In order to remedy the problem of inadequate compensation of a single

object for identifying the distribution of positive class, Chen and Wang [40] formed multiple prototypes iteratively by introducing the method of DD-SVM. Recently, Zhang *et al.* used a method of co-saliency detection based on a self-spaces-MIL(SP-MIL) framework [25]. It combined MIL with self-paced learning to produce a novel framework that could select training samples theoretically. In addition, the method integrated the knowledge of sample diversity and spatial smooth to make selected samples satisfy the requirement and the approach more robust.

III. PROPOSED METHOD

In this section, we introduce single image saliency detection firstly, and then we present the co-saliency detection of group images by fusing the single image saliency. The schematic illustration of the proposed method is shown in Fig.1.

A. SINGLE IMAGE SALIENCY BASED ON MIL

Single image saliency detection contains training part and testing part in general. The images of training set are segmented and labeled by comparing with the corresponding groundtruth in the training part. Then, by learning from the training set, the EC-SVM (Evidence Confidence-SVM) [5] is utilized to obtain a discriminative classification model. In the testing part, we use this model to predict the labels of the test images and obtain the saliency maps of the test images.

1) THE DEFINITION OF BAG AND INSTANCE

We use mean-shift [1] algorithm to segment each image into several regions. Each segmented region is defined as a bag. Let the collection of positive bags be denoted by $\mathcal{B}^+ = \{B_1^+, B_2^+, \cdots, B_i^+\}$ while $\mathcal{B}^- = \{B_1^-, B_2^-, \cdots, B_i^-\}$ is denoted as the collection of negative bags. Next, one percent of the pixels are randomly selected from each segmented region, and the selected pixels are regarded as the seed pixels, which are jointly used as instances in a bag with their neighboring pixels. Each bag B_i is composed of a set of instances $\{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \cdots, \boldsymbol{x}_{ij}\}$, where j is the cardinality of instances collection in B_i. We identify the positive bags and the negative bags by comparing the selected instances with the binarization benchmark set, which are the corresponding groundtruth with two colors in black and white, according to whether the selected instances contain the object of interest. In a bag, if at least one of its instances is positive compared with the binarization benchmark, it's labeled positive; otherwise, it's labeled negative.

With the constructed bags, $\boldsymbol{x}_{ij}^{\mathcal{C}}$ is the j-th instance in the positive bag B_i^+. $\boldsymbol{x}_{ijk}^{\mathcal{C}}$ indicates the value of the k-th attribute of the vector of $\boldsymbol{x}_{ij}^{\mathcal{C}}$ instance. Similarly, denote $\boldsymbol{x}_{ijk}^{-}$ as the value of the k-th attribute of the j-th instance of the i-th negative bag in the training samples.

2) MODEL TRAINING

We extract features of each pixel by using color and texture features. The steerable pyramids of texture feature [6] can overcome the lack of translation invariance in the recursive multi-scale transformation. After the images are filtered [7] in different resolution space, they can be decomposed into a set of sub blocks in different scales and directions with translation and rotation invariance that can reduce the distortion in image processing.

The average value of the features of whole pixels in a block is denoted as the feature of the block and the average value of the features of whole pixels in an instance is denoted as the feature of the instance and the average value of the features of whole pixels in an image is denoted as the global feature of the image. If we use the vision features directly, due to the relative concept of salient, we can not get a better result of saliency. Therefore, we extract three kinds of relative characteristics. 1) The difference between an instance and the average characteristics of the adjacent blocks of the instance. 2) The difference between the average characteristic of the instance and the blocks of image edge. 3) The difference of characteristics between the instance and global image. We connect the relative characteristics of these three groups by linear combination, generating a feature vector, which is used for model training. Then, the distance between a bag and an instance is defined as the distance between this instance and the nearest instance of the bag. Therefore, for an unknown instance in a positive bag, if it is more close to the positive instances and gets far away from all the negative instances simultaneously, then it is more likely to be positive.

Suppose that $\boldsymbol{x}_{mn}$ is an instance of a bag B_i, $\mathrm{EC}(\boldsymbol{x}_{mn})$ is the evidence confidence value of the instance $\boldsymbol{x}_{mn}$ in the bag B_i, which represents the likelihood of instance $\boldsymbol{x}_{mn}$ to be a positive instance, is defined as follows:

$$\mathrm{EC}(\boldsymbol{x}_{mn}) = \prod_i \Pr\left(\boldsymbol{x}_{mn} \mid B_i^+\right) \prod_i \Pr\left(\boldsymbol{x}_{mn} \mid B_i^-\right) \tag{1}$$

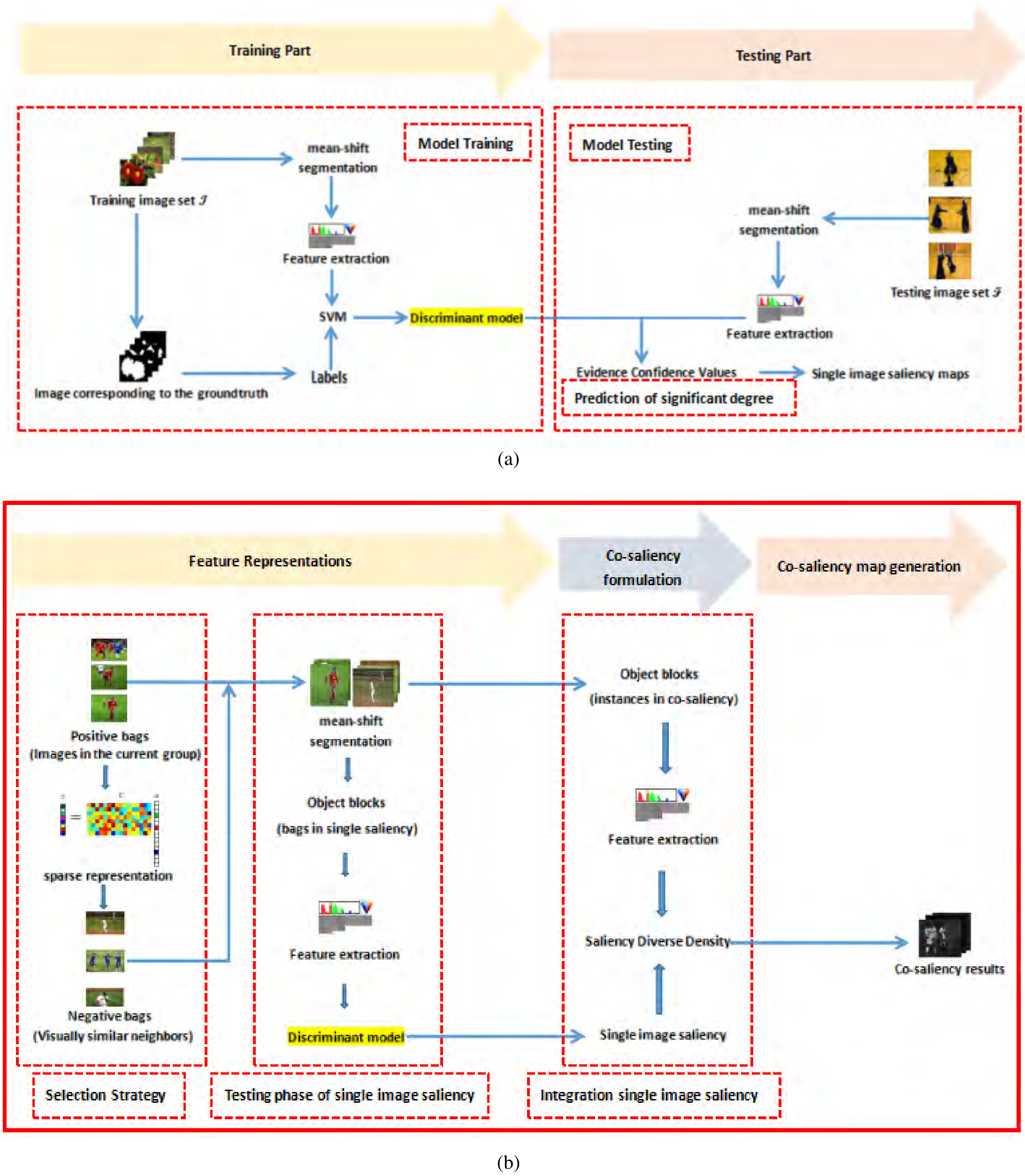


FIGURE 1. Diagrams of the proposed methods on single image saliency detection and co-saliency detection from group images. (a) Single image saliency detection. (b) Co-saliency detection from group images.

we use the model Noisy-OR [2] to estimate the above formulation:

$$\Pr\left(x_{mn} \mid B_i^+\right) = 1 - \prod_j \left[1 - \Pr\left(x_{mn} \mid x_{ij}^{C}\right)\right] \quad (2)$$

$$\Pr\left(x_{mn} \mid B_i^-\right) = \prod_j \left[1 - \Pr\left(x_{mn} \mid x_{ij}^{-}\right)\right] \quad (3)$$

where $\Pr\left(x_{mn} \,|x_{ij}\right)$ is defined as follows:

$$\Pr\left(x_{mn} \,|x_{ij}\right) = \exp\left\{-d^2\left(x_{ij}, x_{mn}\right)/\delta^2\right\} \quad (4)$$

where δ is the scaling parameter. Then, the distance between an instance and another instance is defined as the distance between the feature vectors, i.e.:

$$d^2\left(x_{ij}, x_{mn}\right) = \sum_k \left(x_{ijk} - x_{mnk}\right)^2 \quad (5)$$

where the x_{ijk} and x_{mnk} represent the values of the k-th characteristic vector in each corresponding instance.

From the formulation, we can see that if an instance in B_i^+ is close to x_{mn}, we can get a higher value of $\Pr\left(x_{mn} \mid B_i^+\right)$ and only if all the instances in B_i^- are far away from x_{mn}, we can get a higher value of $\Pr\left(x_{mn} \mid B_i^-\right)$. Hence, if all the instances in B^- are far away from x_{mn} and each B^+ includes at least one instance close to x_{mn} simultaneously, EC (x_{mn}) will be high.

As a result, the larger the EC value of the instance, the more likely it is a positive instance.

We calculate the EC value of each instance in positive bags, each instance is sorted according to the degree of EC. The greater the degree of EC value, the more likely it is positive. We put the top five instances of the maximum likelihood of each positive bag into a set of identified positive instances, which is entirely composed of positive instances to form a bag. Then, each bag B_i is mapped to the feature space based on the set of the identified positive instances. In order to describe the bag, we calculate the distance between each instance in the bag B_i and the set of the identified positive instances that is composed of a feature vector:

$$\psi\left(B_i\right)=\left(d\left(e_1^*, B_i\right), d\left(e_2^*, B_i\right), \cdots, d\left(e_{|\mathcal{E}^*|}^*, B_i\right)\right)^T \quad (6)$$

where $e_k^* \in \mathcal{E}^*$, $\mathcal{E}^*$ is the set of identified positive instances, $|\mathcal{E}^*|$ represents the cardinality of identified instances. $d\left(e, B_i\right)$ can be written as follows:

$$d\left(e, B_i\right)=\min_{\boldsymbol{x}_{ij} \in B_i}\left(\|e-\boldsymbol{x}_{ij}\|\right) \quad (7)$$

which represents the distance between an instance and a bag that is equivalent to the distance between the instance and the nearest instance in the bag.

We utilize LIBSVM [30] classifier to obtain the classification model by using the characteristics and labels of bags. In the testing part, we can use the model to predict the label of the bags and obtain the saliency maps according to the forecast results. Algorithm 1 summarizes the whole procedure of the proposed method of single image saliency detection presented above.

Algorithm 1 Single Image Saliency Detection Algorithm

Input: Training images set $\mathcal{I}$; Images corresponding to the groundtruth; Testing images set $\mathcal{F}$.
Output: Image saliency maps set $\mathcal{M}$.
Training-Part:
for *image* $I_t \in \mathcal{I}$ **do**
1. Use the algorithm of Mean-Shift to segment the image I_t and define each segmented block as a bag;
2. Combine I_t corresponding groundtruth and get the labels of bags;
3. From each bag B_i select the instances, and each instance is composed of 1% pixels randomly selected from the bag and its adjacent pixels;
4. Extract feature vector;

end
5. Utilize the framework of EC-SVM MIL to obtain the discriminative classification model.
Testing-Part:
for *image* $F_z \in \mathcal{F}$ **do**
1. Use the algorithm of Mean-Shift to segment the image F_z and define each segmented block as a bag;
2. From each bag B_i select the instances, and each instance is composed of 1% pixels randomly selected from the bag and its adjacent pixels;
3. With the EC-SVM MIL framework, the label of each bag is obtained by using the discriminative classification model to calculate the confidence degree of each bag;
4. The confidence degree of the sub block is denoted as the EC value of each bag, and the image saliency map M_z is generated.

end

B. CO-SALIENCY DETECTION FROM GROUP IMAGES BASED ON MIL

From the perspective of co-saliency detection from group images, the major point of results needs to fit the essence, which are the saliency and repeatness. To satisfy the characteristic of repeatness, the co-saliency of group images should have common region. To satisfy the characteristic of saliency, each image in the group should keep saliency in corresponding single image. As a result, we propose a unified framework based on MIL to get the saliency maps of co-saliency combined with single image saliency. In this section, we use a kind of novel selection strategy based on global features to select negative bags, which backgrounds are similar to the backgrounds of positive bags. Then, the original DD algorithm [2] is improved to give a higher weight for the foreground area of an image to reduce the probability that the background is chosen as a target instance.

1) THE DEFINITION OF BAG AND INSTANCE

Co-saliency detection aims at finding positive instances from the positive bags. Each image in a folder is denoted as a positive bag and the mean-shift algorithm [1] is used to segment the image. Then, each segmented block is denoted as an instance of a bag instead of selecting pixels in single image saliency detection. We define the blocks containing co-salient region as positive instances in a bag and define others as negative instances. In order to obtain the saliency maps accurately, we utilize the information from negative bags, which are chosen from other folders and play a vital role in suppressing background and highlighting the common region in the positive bags. Thus, we choose negative bags by introducting a novel selection strategy in the next section.

$\mathcal{L}$ is denoted as a set of training samples that can be categorized into a set of positive bags $\mathcal{L}^+=\left\{B_1^+, B_2^+, \cdots, B_{|\mathcal{L}^+|}^+\right\}$ and a set of negative bags $\mathcal{L}^-=\left\{B_1^-, B_2^-, \cdots, B_{|\mathcal{L}^-|}^-\right\}$. We get the set of negative bags by using the selection strategy of negative bags. $|\mathcal{L}^+|$ means the number of positive bags, and $|\mathcal{L}^-|$ means the number of negative bags, $|\mathcal{L}|=|\mathcal{L}^+|+|\mathcal{L}^-|$. B_i is composed of a set of instances $\left\{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \cdots, \boldsymbol{x}_{ij}\right\}$, where j is the cardinality of the set of instances in B_i. I is denoted as an instance in the set of positive bags $\mathcal{L}^+$. We need to compute the diversity density values of each instance $I \in \mathcal{L}^+$ and select the co-salient region according to the value of the diversity density.

2) THE SELECTION STRATEGY OF NEGATIVE BAGS

Co-saliency detection does not have the phase of testing, which indicates that the selection of the negative bags has a great impact on the effect of the detection of co-salient region. On the one hand, it is easy to view a instance of the background area in the set of positive bags as a positive instance if the positive bags and the negative bags have an excessive difference, because the positive bags have similar backgrounds and the negative bags can not suppress the interference of these similar backgrounds. On the other hand, it is easy to regard the co-salient region as a negative instance if the positive bags and the negative bags are too similar, making the selection of the target instance is not complete and causing the co-salient region is not accurate. To solve the problems, we select the negative bags by using the sparse representation method.

We extract global features and obtain the color histograms according to the quantized color space for each image. The feature vector $\boldsymbol{v_i} \in \mathbb{R}^d$ of each positive bag in $\mathcal{L}^+$ is obtained, where $\boldsymbol{v_i}$ is d dimensions, and the feature vectors of negative bags are considered as a dictionary $U \in \mathbb{R}^{d\times N}$, where U is a feature matrix of $d \times N$ and N is the size of the dictionary. We let the reconstruction coefficient denote as $\alpha^* = \{\alpha_{\mathbf{1}}, \alpha_{\mathbf{2}}, \cdots, \alpha_{\mathbf{l}}\}$, where l equals to N. According to the sparse linear combination, the feature vectors in the dictionary are used to reconstruct the feature vector of each positive bag. In general, we consider that a vector with non-zero coefficients is similar to the reconstructed vector. This indicates that the larger the non-negative coefficient of the reconstructed vector is, the more similar the original image corresponding to the coefficient is with the positive bag and the higher the likelihood of it is chosen as a negative bag. We determine the value of reconstruction coefficient by the following definition:

$$\begin{aligned} \boldsymbol{\alpha}^* = \ & \arg\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{v_i} - U\boldsymbol{\alpha_i}\|_2^2 + \lambda\|\boldsymbol{\alpha_i}\|_1 \\ & s.t.\ \boldsymbol{\alpha}_l \geq 0, \forall l = 1, 2, \cdots, N \end{aligned} \tag{8}$$

where $\lambda \geq 0$ is used to control the sparsity of the regularization parameters. $\boldsymbol{\alpha}^* \in \mathbb{R}^N$, where $\boldsymbol{\alpha}^*$ is a vector of N dimensions.

Based on the calculation of the above formulation, the reconstruction coefficient vector is obtained and the non-negative reconstruction coefficients are descended. According to the result of sorting, we select the same number of positive bags as negative bags.

In short, the negative bags are chosen by using the sparse representation algorithm. We utilize the dictionary to reconstruct a particular feature vector and use the restrictions to achieve non-negative constraints. In this way, the negative bags satisfy that the backgrounds are similar to the backgrounds of positive bags and the salient region is different from the salient region of the positive bags simultaneously. Firstly, it provides a better basis for the detection of foreground. Secondly, it helps the algorithm enhance the sensitivity and robustness to noise.

The obtained negative bags have two principles: 1) The salient region is not contained in positive bags; 2) The backgrounds are close to the backgrounds of the positive bags.

3) CO-SALIENCY DETECTION FROM GROUP IMAGES

The major drawback in the detection of co-saliency is the background of the positive bags has high consistency, which is possible to make the background areas have higher diversity density value if we use the traditional DD algorithm. As a result, we use the selection strategy to choose negative bags that satisfy the two principles proposed in the previous section. To a certain extent, it can reduce the probability that the background areas are selected as the positive instances. Nevertheless, the effect of it is limited, so we need to give a higher weight for the foreground areas by improving the DD algorithm.

According to the definition of bags and instances on the previous section, the problem of co-saliency detection is denoted as a binary classification problem with the unified framework of MIL. If $B_i \in \mathcal{L}^+$, then $y_i = 1$, otherwise, $B_i \in \mathcal{L}^-$ and $y_i = 0$. We determine the calculation of the original DD algorithm as follows:

$$\mathrm{DD}(\boldsymbol{I}, \mathcal{L}) = \sum_{i=1}^{|L|}\prod \max_j \left\{1 - |y_i - \exp\left(-d^2\left(\boldsymbol{x_{ij}}, \boldsymbol{I}\right)\right)|\right\} \tag{9}$$

The distance d between instance $\boldsymbol{x_{ij}}$ and instance $\boldsymbol{I}$ is denoted as:

$$d^2\left(\boldsymbol{x_{ij}}, \boldsymbol{I}\right) = \sum_k \left(s_k\left(\boldsymbol{x_{ijk}} - \boldsymbol{I_k}\right)^2\right) \tag{10}$$

where $\boldsymbol{x_{ijk}}$ represents the k-th attribute of the vector of instance $\boldsymbol{x_{ij}}$, $\boldsymbol{I_k}$ is the k-th attribute of the vector of instance $\boldsymbol{I}$, and s_k is the scaling factor of the k-th feature.

The above algorithm can detect the salient region of a group of images, nevertheless, it can only guarantee the repeatness of co-saliency, but neglecting the saliency of it. If the background region with high evidence confidence value is mistaken as a positive instance, the final saliency map is not ideal, which is because a group of positive bags have a similar background. Although the selection strategy of negative bags can reduce the influence of backgrounds on the final results, but if an instance in the background of positive bags has similar feature with the background of negative bags, then, the value of EC is large that the background region is considered as co-salient region and extracted. Therefore, we need to solve the problem by improving the diversity density algorithm.

In order to improve the case of diversity density values are close to 0 and avoid the impact of similar background region, the multiplication operation was converted into a more robust addition operation. Then, we combine the single image saliency with the diversity density algorithm. In this way, we can ensure that a region with high saliency value is still saliency, while the pixels of the background region are

Algorithm 2 Co-Saliency Detection From Group Images

Input: Positive bags set $\mathcal{P}$ (Images in the current folder).
Output: Co-saliency maps set $\mathcal{S}$.
Training-Part:
for *positive bag(image)* $P_i \in \mathcal{P}$ **do**
1. Sparse representation method is used to get negative bags set $\mathcal{N}$ (Images in other folders);
2. Mean-Shift algorithm is used to segment the positive bag P_i and negative bag N_i, and each segmented block is defined as an instance;
3. For positive bag P_i and the selected negative bag N_i, extract the feature of them and use the discriminant model obtained by single image saliency detection to get the corresponding single image saliency maps;
4. Integrate the result of single image saliency via SDD algorithm to get the final co-saliency map S_i;

end

suppressed in the co-saliency detection from group images. We assign the regions (instances) with a high degree of image within the positive bags a higher weight and assign the lower weight to the regions (instances) with low degree. Each instance is mapped to a single image saliency map in corresponding area, and the average value of the pixels in the corresponding region of the saliency map is used to represent the confidence value of the instance. According to this confidence value, we define the initial weight of each instance in a positive bag as a value between 0 and 1 while other instances in the negative bags are defined as 0. The improved diversity density algorithm becomes the following formulation:

$$\mathrm{SDD}(\boldsymbol{I}, \mathcal{L}) = \sum_{i=1}^{|L|} \max\Big\{1 - |y_i - \big(S_{\boldsymbol{x}_{ij}} + S_{\boldsymbol{I}}\big) \times \exp\Big(-d^2\left(\boldsymbol{x}_{ij}, \boldsymbol{I}\right)\Big)|\Big\} \tag{11}$$

where $S_{\boldsymbol{x}_{ij}}$ is the weight of the j-th instance in bag B_i, $S_{\boldsymbol{I}}$ is the weight of the instance $\boldsymbol{I}$. The distance d between the instance $\boldsymbol{x}_{ij}$ and the instance $\boldsymbol{I}$ is according to the previous definition.

The diversity density value of each instance in the set of positive bags is calculated according to the formula (11), and this value is proportional to the likelihood of a positive instance. Finally, each diversity value is mapped to [0, 255] to get the final results, which represents a confidence value of the image region corresponding to the instance that is the value of the pixels in a gray scale image. We assign all of the pixels in each corresponding instance by using these values and eventually get gray scale images to describe the co-saliency results of group images. Algorithm 2 summarizes the whole procedure of the proposed method of co-saliency detection from group images presented above.

IV. EXPERIMENTS PERFORMANCE EVALUATION

In this section, we evaluate the performance and the effectiveness of our method compared with various different methods on several datasets.

A. DATA SETS

We evaluate the accuracy of saliency maps based on three existing public image datasets, which are the ImgSal [9] dataset, the ECSSD [8] dataset, and the CMU-Cornell iCoseg [10] dataset respectively. The first two datasets are used to measure the performance of single image saliency detection. The last one is utilized to measure the performance of co-saliency detection from group images. The first dataset has a total of 235 images. Each image has a corresponding binarization reference image. These images are broken down into six categories according to the size of the object, the number of salient objects and the complexity of the background. The second dataset contains 1000 images, which is developed by the MSRC dataset and concatenated images and rich semantic information as well as complex structure for people to analyse. The last one mainly contains 38 scenes, which are stored in different folders according to the scene, and each scene contains a number of images, and all the images in the same scene contain co-salient regions. Each image contains a set of manually labeled binarization bench marks.

In the recent years, some new datasets are proposed. For instance, there are two datasets used for single image saliency detection, which are the DUTS image dataset [46] and the DUT-OMRON image dataset [29], respectively. The first one is the largest dataset currently that is designed to address the unfair comparisons and inconsistent problem among different methods. It contains both training and testing images, which are all consisted of various challenging sences. The second one is consisted of 5168 nature images, which is selected manually and has complex background as well as more than one salient object. It can make improvement for image saliency detection. There is also one dataset used for co-saliency detection, which is Cosal2015. It contains 50 different categories with various types and salient objects. Each category has some similar images deposited in one folder for co-saliency detection. These recent datasets have kinds of advantages and characteristics, we will use them as benchmark datasets to improve and extend our experiments in the future.

B. EVALUATION MEASURE

1) PR CURVE

To evaluate the performance of the proposed method, we use a widely adopted criteria: the precision recall (PR) curve. In the PR graph, the horizontal axis means recall (R), which is defined as the proportion of the true positives divided by the actual number of terms in the positive category. The vertical axis means precision (P), which is defined as the proportion of the true positives divided by all positive terms that include

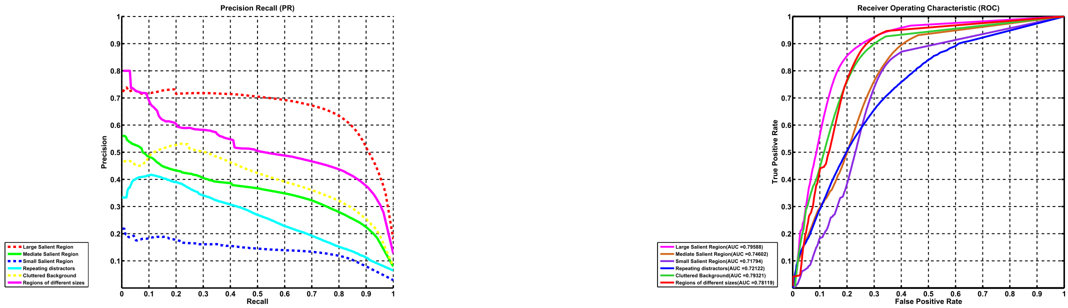


FIGURE 2. The left picture is the PR curves of six kinds of images on ImgSal database. The right one is the ROC curves of six kinds of images on ImgSal database.

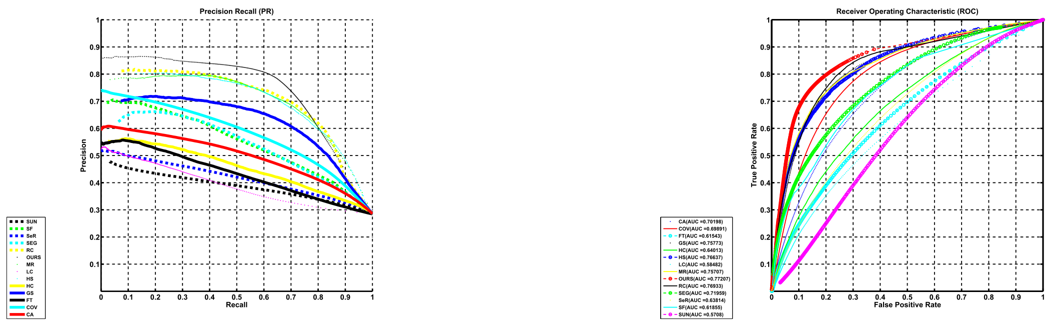


FIGURE 3. The left picture is the PR curves of the various methods on ECSSD database. The right one is the ROC curves of the various methods on ECSSD database.

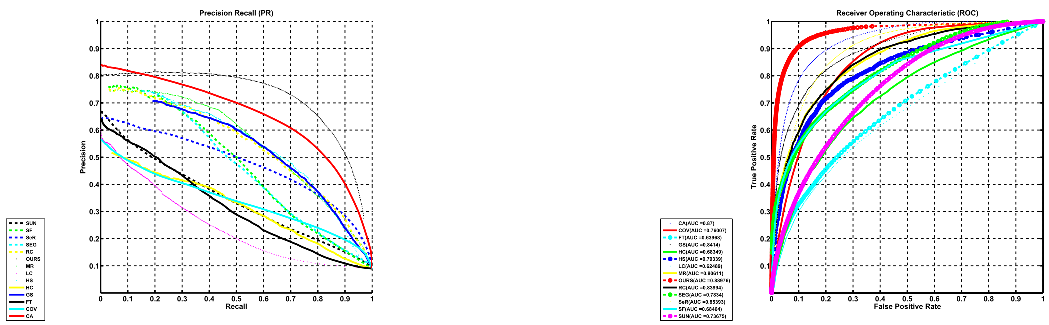


FIGURE 4. The left picture is the PR curves of the various methods on ImgSal database. The right one is the ROC curves of the various methods on ImgSal database.

the real positives and the false positives. The saliency maps are gray scale figures, which are composed of pixels between 0 and 255. Set a threshold, if the gray value of a pixel is bigger than the threshold, we regard it as significant; otherwise, we view it as insignificant. The threshold is changed from 0 to 255 to obtain the final PR curve.

2) ROC CURVE

We also use the ROC curve to evaluate the performance of the proposed method. In the ROC graph, the horizontal axis is false positive rate (FPR), which is denoted as the proportion of actual negative instances divided by all the negative instances in the positive class predicted by the classifier. The vertical axis is true positive rate (TPR), which is denoted as the proportion of the actual positive instances divided by all positive instances in the positive class predicted by the classifier. AUC stands for the area under the ROC curve, and we can compare the size of AUC value to determine the quality of the classifier. The greater the AUC value is, the better the quality of classifier is.

3) AVERAGE PRECISION, AVERAGE RECALL, AVERAGE F-MEASURE

In the experiment, we first adopt an adaptive threshold to compute the salient parts in which the adaptive threshold is defined as twice the mean saliency. F-measure is denoted as follows:

$$F-measure = \frac{\left(1+\gamma^2\right) Precision \times Recall}{\gamma^2 \times Precision + Recall}$$

where γ is set as 0.3 in our experiment.

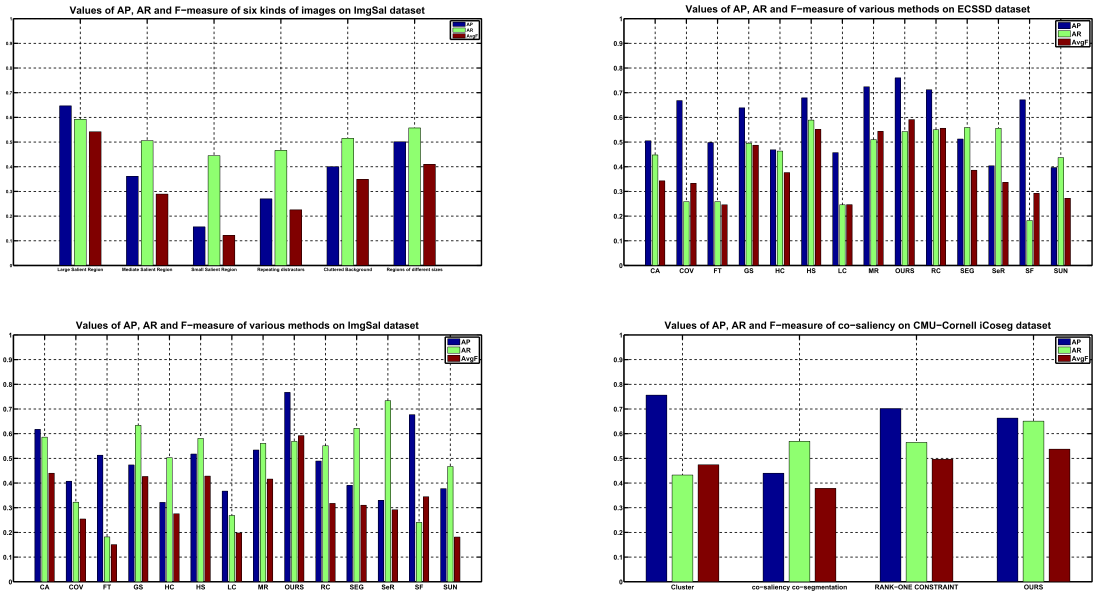
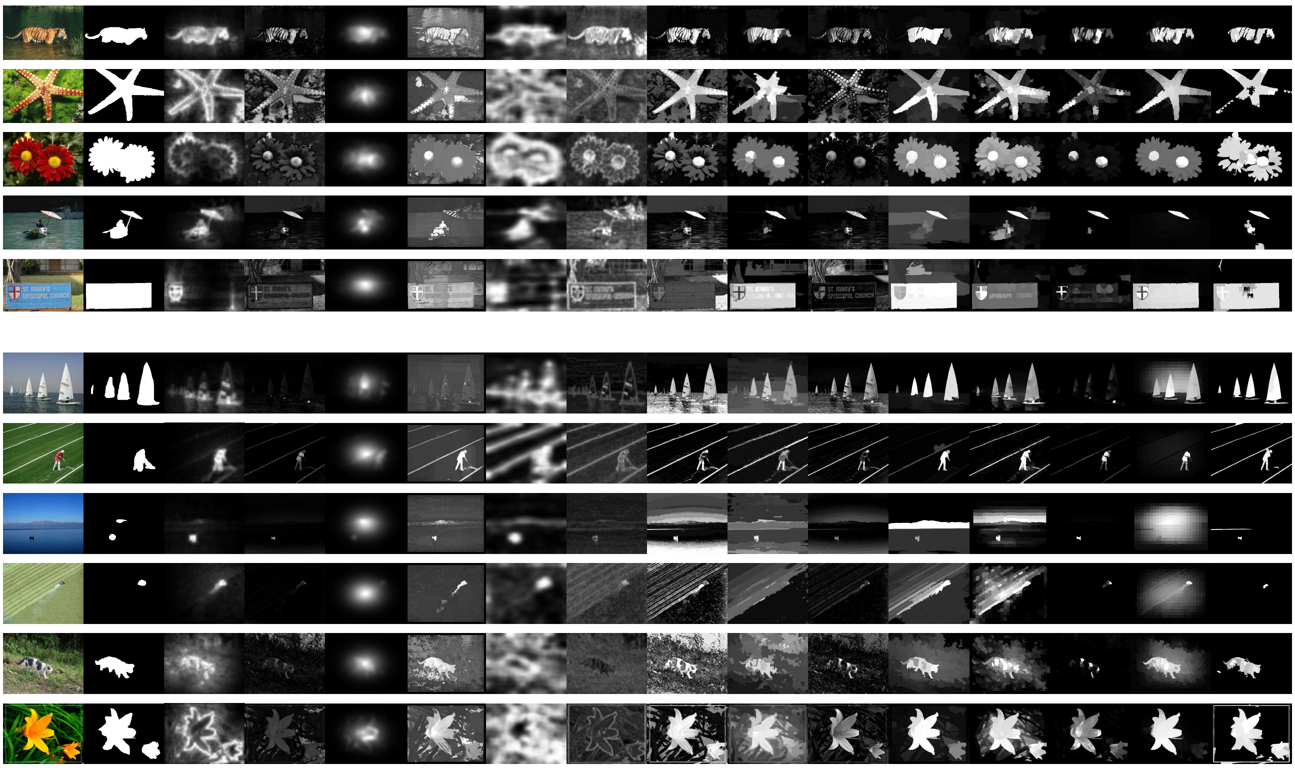


FIGURE 5. **The left picture of the first line is the AP, AR and average F-measure values of our algorithm on six kinds of images on ImgSal database. The right picture of the first line is the AP, AR and average F-measure values of various methods on ECSSD database. The left picture of the second line is the AP, AR and average F-measure values of various methods on ImgSal database. The right picture of the second line is the AP, AR and average F-measure values of Co-saliency on CMU-Cornell iCoseg database.**



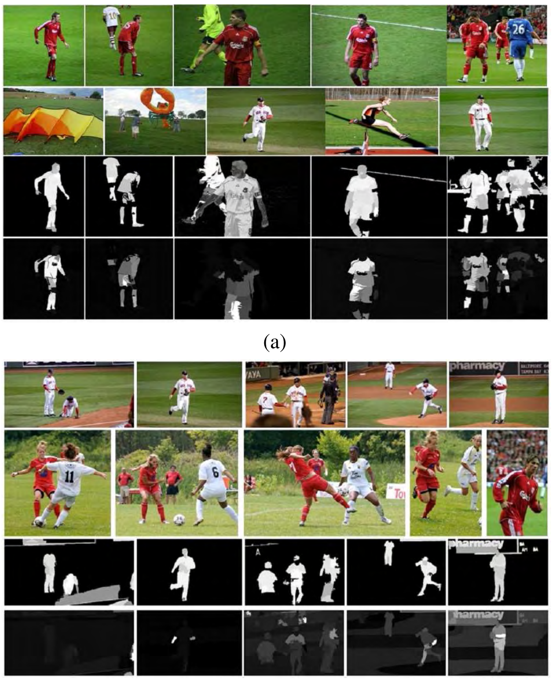
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o) (p)

FIGURE 6. **Comparison of single image saliency detection results for various approaches on the ECSSD database and the ImgSal database. The first block is the results on the ECSSD dataset. The second block is the results on the ImgSal dataset. From left to right is the method of: (a) Original, (b) Groundtruth, (c) CA, (d) FT, (e) COV, (f) SEG, (g) SeR, (h) SUN, (i) HC, (j) RC, (k) LC, (l) HS, (m) GS, (n) SF, (o) MR,(p) OURS.**

Next, we use the average precision(AP), average recall(AR) and Average F-measure(AvgF) to evaluate the obtained results by variety of approaches.

C. PERFORMANCE COMPARISON

We compare our proposed method with the most classic or the newest methods including the CA [15], COV [20],



(a)

(b)

FIGURE 7. **Results of co-saliency detection maps in two different scenes. The first line is the original pictures. The second line is the negative bags selected by the novel strategy. The third line is single image saliency detection results obtained by our proposed framework of MIL. The last line is the co-saliency detection results obtained by fusing the single image saliency. (a) The selection strategy of negative bags is used. (b) The selection of negative bags is randomly.**

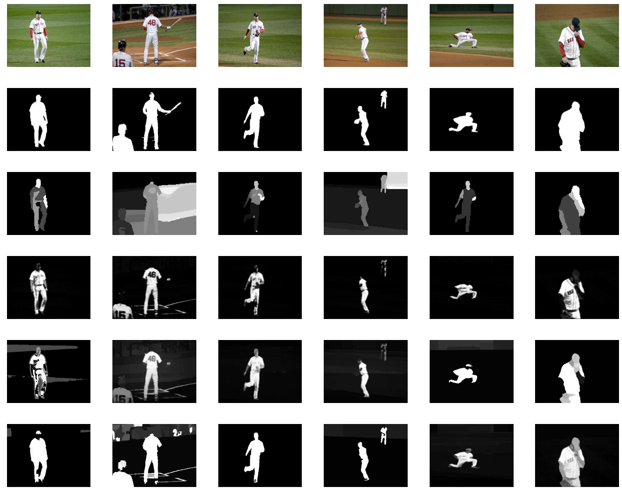


FIGURE 8. **The co-saliency detection maps of various methods on the iCoseg dataset. The first line is the original pictures. The second line is the corresponding groundtruth. The third line is the results obtained by the method of Meng [22]. The fourth line is the results obtained by the method of Cao [23]. The fifth line is the results obtained by the method of Fu [24]. The last line is the results obtained by our method.**

FT [14], GS [18], HC [17], RC [17], HS [21], LC [11], MR [29], SEG [16], SeR [44], SF [19], SUN [12] on the ECSSD, ImgSal datasets for single image saliency detection and some co-saliency detection methods including the method of Meng [22], Cao [23], Fu [24] on the CMU-Cornell iCoseg dataset. For fair comparison, the results of these various different methods are obtained by running available software or codes.

D. ANALYSIS ON SINGLE IMAGE SALIENCY DETECTION

At first, we evaluate the performance of the proposed method on the six kinds of images contained in the ImgSal database by using the PR curve, as is shown in Fig.2(left). For each category, we evaluate the performance by using the strategy of "one-versus-the-rest". As can be seen, images with large salient region have a better effect. Next, we further compare the six kinds of images by ROC curve (see the graph in Fig.2(right)). From the picture, we can see that the AUC value of images with large salient region equals to 0.79588, which is the largest compared with other kinds of images and the AUC value of images with small salient region equals to 0.71794, which is the lowest compared with other kinds of images. The result shows that the algorithm can achieve better accuracy on the images containing large salient objects. In addition, we use average precision, average recall and average F-measure to evaluate the accuracy of the obtained results. As is shown in the first line of Fig.5, the left one is the values of AP, AR and Average F-measure of six kinds of images on ImgSal dataset. We can clearly see that images with large salient region obtains better effectiveness in terms of AP, AR and Average F-measure, which indicates that the algorithm has better performance on images with large salient region.

Furthermore, in order to evaluate the effectiveness of the proposed algorithm, we select several typical algorithms and compare with our algorithm. We still use the PR curve to evaluate these algorithms (See Fig.3(left) and Fig.4(left)). As can be seen, our method has a better effect on both ECSSD and ImgSal datasets compared with other 13 methods. Then, we evaluate the performance of various methods on these two databases by adopting ROC curve (See Fig.3(right) and Fig.4(right)). The curves show that the AUC values of our method are all higher than other methods. Specifically, we can believe that our method is more robust compared to the previous methods. Moreover, we also use average precision, average recall and average F-measure to compare these different methods. The comparison results are shown in Fig.5, the right one of the first line is the values of AP, AR and Average F-measure of various methods on ECSSD dataset and the left one of the second line is the values of AP, AR and Average F-measure of various methods on ImgSal dataset. From these two figures, we can see that our method has competitive results compared with other methods about the AP scores and the Average F-measure values, which proves that our method yields state-of-the-art methods on the two kinds of datasets. The overall effect of the various methods are directly shown in Fig.6. Our algorithm gives the same degree for each segmented block in an image and uses the relative characteristics of the region to describe instances, which can guarantee the obvious object more complete. Thus, our method obtains a better result than other methods.

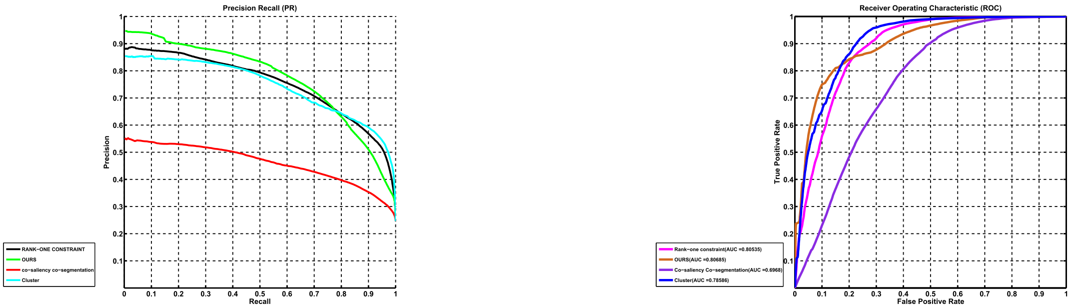


FIGURE 9. The left picture is the PR curves of co-saliency detection of various methods. The right one is the ROC curves of co-saliency detection of various methods.

In short, our method can achieve a better performance for most types of images through the above analysis, but it is not ideal for the detection of images containing small salient objects. This is due to over-segmentation. The proposed method takes the saliency analysis of each partition area as an unit, and after the segmentation of images containing small objects, the sub blocks of small objects are often included in non-salient parts of the image, leading to the effect of detection is not ideal. If we can find a better segmentation algorithm to segment the small salient objects, the effect will be significantly improved. Our future work is to find a better segmentation algorithm to improve our method.

E. ANALYSIS ON CO-SALIENCY DETECTION FROM GROUP IMAGES

For the experiment of co-saliency detection, we use the selection strategy of negative bags to select negative bags, which contain the most similar backgrounds with positive bags. For the parameter in the selection of negative bags, we empirically set $\lambda = 2$ in (8). The results of the co-saliency detection of two different scenes are shown in Fig.7, which show the differences in applying the selection strategy of negative bags and selecting randomly. The second line in the (a) are negative bags selected by the selection strategy of negative bags, which almost have the similar backgrounds and the different objects while the second line in the (b) are negative bags selected randomly, which have some different backgrounds and some similar objects that affect the final co-saliency detection results.

Moreover, we make quantitative comparisons with the method of Meng [22], Cao [23] and Fu [24] on the CMU-Cornell iCoseg dataset. Fig.8. shows the overall effect of the various methods, and we can see that the final results obtained by our method have a clearer outline around the co-salient object. In order to make a direct comparison between our algorithm and other algorithms, we still use the PR curve to evaluate it. As is shown in Fig.9(left), our method has a better appearance compared with others. Then, we use ROC curve to further evaluate these methods. As is shown in Fig.9(right), we can see that the AUC value of our proposed method equals to 0.80685, which is the better one compared with other methods. Furthermore, we apply the average precision, average recall and average F-measure to evaluate the results of co-saliency detection. As is shown in the last picture of Fig.5, our method has a higher Average F-score compared with other methods. Note the effectiveness of the proposed method, it can highlight most of the co-salient region that is believed that our method achieves better effectiveness compared with other methods .

In the experiment of co-saliency detection from group images, the foreground region and the background region are considered separately, and the backgrounds of negative bags are similar to the backgrounds of positive bags, which means that as long as the salient regions of negative bags have some big difference with positive bags, it can significantly improve the effect of the co-saliency detection from group images. Thus, our future work is to find a better way to further improve the selection strategy of negative bags.

V. CONCLUSION

In this paper, we have presented a unified MIL framework to detect both single image saliency detection and co-saliency detection from group images. For single image saliency detection, we use EC-SVM (Evidence Confidence-SVM) MIL algorithm to learn a discriminant model for saliency prediction on unknown test images, which empirically proved that using MIL for single image saliency detection is feasible and effective. For co-saliency detection from group images, we concatenate the EC values and single image saliency to obtain the final co-saliency maps. We also improve the traditional DD algorithm and propose a principled selection strategy to choose the negative bags, which can further improve the performance of co-saliency detection from group images. Experiments on three public datasets have shown that our method achieves encouraging performance compared to the state-of-the-arts approaches.

REFERENCES

[1] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[2] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Aug. 1998, vol. 10. no. 2, pp. 570–576.
[3] R. Huang, W. Feng, and J. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Turin, Italy, Jun./Jul. 2015, pp. 1–6.

[4] Y.-L. Chen and C.-T. Hsu, "Implicit rank-sparsity decomposition: Applications to saliency/co-saliency detection," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 2305–2310.

[5] W.-J. Li and D.-Y. Yeung, "Localized content-based image retrieval through evidence region identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 1666–1673.

[6] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. Int. Conf. Image Process.*, Washington, DC, USA, 1995, pp. 444–447.

[7] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[8] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended CSSD," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 717–729, Apr. 2016.

[9] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.

[10] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "iCoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 3169–3176.

[11] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 853–860.

[12] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "SUN: Top-down saliency using natural statistics," *Vis. Cognit.*, vol. 17, nos. 6–7, pp. 979–1003, Aug. 2009.

[13] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2296–2303.

[14] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 1597–1604.

[15] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[16] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 6315, K. Daniilidis, P. Maragos, N. Paragios, Eds. Berlin, Germany: Springer, 2010.

[17] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[18] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, vol. 7574. no. 1, pp. 29–42.

[19] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 733–740.

[20] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vis.*, vol. 13, no. 4, pp. 1–20, Mar. 2013.

[21] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1155–1162.

[22] F. Meng, H. Li, G. Liu, and K. N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," *IEEE Trans. Multimedia*, vol. 14, no. 5, pp. 1429–1441, Oct. 2012.

[23] X. Cao, Z. Tao, B. Zhang, H. Fu, and X. Li, "Saliency map fusion based on rank-one constraint," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Jose, CA, USA, Jul. 2013, pp. 1–6.

[24] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[25] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.

[26] Z. Fu, A. Robles-Kelly, and J. Zhou, "MILIS: Multiple instance learning with instance selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 958–977, May 2011.

[27] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1911–1922, Apr. 2017.

[28] Q. Wang, Y. Yuan, P. Yan, and X. Li, "Saliency detection by multiple-instance learning," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 660–672, Apr. 2013.

[29] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 3166–3173.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.

[31] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, nos. 1–2, pp. 31–71, 1997.

[32] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 2214–2219.

[33] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 914–921.

[34] J. Zhang and S. Sclaroff, "Saliency detection: A Boolean map approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 153–160.

[35] P. Mehrani and O. Veksler, "Saliency segmentation based on learning and graph cut refinement," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 1–12.

[36] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 883–890.

[37] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 2790–2797.

[38] Q. Zhang and S. A. Goldman, "EM-DD: An improved multiple-instance learning technique," in *Proc. Conf. Neural Inf. Process. Syst.*, 2002, pp. 1073–1080.

[39] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Proc. Conf. Neural Inf. Process.*, 2002, pp. 561–568.

[40] Y. Chen and J. Z. Wang, "Image categorization by learning and reasoning with regions," *J. Mach. Learn. Res.*, vol. 5, pp. 913–939, Aug. 2004.

[41] X. Cao, Y. Cheng, Z. Tao, and H. Fu, "Co-saliency detection via base reconstruction," in *Proc. ACM Int. Conf. Multimedia*, Orlando, FL, USA, 2014, pp. 997–1000.

[42] D. Zhang, J. Han, C. Li, and J. Wang, "Co-saliency detection via looking deep and wide," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 2994–3002.

[43] D. Zhang, J. Han, J. Han, and L. Shao, "Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1163–1176, Jun. 2016.

[44] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, pp. 1–27, Nov. 2009.

[45] F. Meng, H. Li, and G. Liu, "A new co-saliency model via pairwise constraint graph matching," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Taipei, Taiwan, 2012, pp. 781–786.

[46] L. Wang *et al.*, "Learning to detect salient objects with image-level supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 136–145.

HONGLIN QUAN received the B.S. degree from Beijing Information Science and Technology University, Beijing, China, in 2016. She is currently pursuing the Ph.D. degree with the School of Computer and Information Technology, Beijing Jiaotong University. Her research interests include computer vision and machine learning.

SONGHE FENG received the Ph.D. degree from the School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, in 2009. He was a Visiting Scholar with the Department of Computer Science and Engineering, Michigan State University, USA, from 2013 to 2014. He is currently an Associate Professor with the School of Computer and Information Technology, Beijing Jiaotong University. His research interests include computer vision and machine learning.

BAIFAN CHEN received the Ph.D. degree from the School of Information Science and Engineering, Central South University, Changsha, China, in 2005. She was a Visiting Scholar with the Department of Computer Science and Engineering, TAMU, USA, from 2014 to 2015. She is currently an Associate Professor with the School of Information Science and Engineering, Central South University. Her research interests include machine vision and robot intelligence.