# Exemplar-based Image Saliency and Co-saliency Detection

**Article** *in* Neurocomputing · January 2020

DOI: 10.1016/j.neucom.2019.09.011



**5 authors**, including:

Rui Huang
Civil Aviation University of China
**17** PUBLICATIONS **238** CITATIONS
SEE PROFILE

Wei Feng
Tianjin University
**105** PUBLICATIONS **1,147** CITATIONS
SEE PROFILE

Zezheng Wang
Tianjin University
**10** PUBLICATIONS **20** CITATIONS
SEE PROFILE

Yan Xing
Civil Aviation University of China
**13** PUBLICATIONS **75** CITATIONS
SEE PROFILE

Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

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

# Exemplar-based image saliency and co-saliency detection

Rui Huang[a,b,c,d], Wei Feng[c,d,*], Zezheng Wang[c,d], Yan Xing[a], Yaobin Zou[b]

[a] College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China
[b] Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002, China
[c] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[d] Key Research Center for Surface Monitoring and Analysis of Cultural Relics, State Administration of Cultural Heritage, Tianjin 300350, China

ARTICLE INFO

*Article history:*
Received 19 July 2018
Revised 18 July 2019
Accepted 8 September 2019
Available online xxx

Communicated by Dr. Zhu Jianke

*Keywords:*
Saliency detection
Co-saliency detection
Exemplar
Label propagation

ABSTRACT

Image saliency and co-saliency detection that aim to detect salient objects in an image or common salient objects in a group of images are import in computer vision. Researchers often treat saliency and co-saliency as two separate problems. In this paper, we show that these two problems can be solved in a single framework, i.e., treating saliency and co-saliency as finding suitable exemplars. Image-level and region-level exemplars are proposed to obtain the similar images and to propagate the saliency values, respectively. Our method only requires a small number of labeled images having similar appearances with a query image. The exemplars help to detect the real salient objects, which is different from the conventional heuristic methods that are fragile for the images with complex scenes. We have conducted abundant experiments on saliency and co-saliency benchmark datasets, which verifies the effectiveness of our method.



## 1. Introduction

Saliency and co-saliency that detect salient and co-salient objects have been studied for several years because of their wide applications in computer vision. Such applications include image segmentation and co-segmentation [1], content-based image compression [2], image retargeting and inpainting [3], and object tracking [4,5].

The existing saliency detectors can be classified into heuristic models and deep learning models. Heuristic saliency detectors are designed based on the contrast of features [6], light classifier (e.g., SVM) [7], graph propagation [8], reconstruction error [9,10] or other simple learning strategies [11]. The main characteristic of heuristic saliency detectors is using hand-crafted features, such as RGB, LAB, Gabor, Texton and LBP. These saliency detectors can barely achieve promising saliency detection on images with complex scenes because of the inefficient ability of hand-crafted features in distinguishing salient objects and non-salient backgrounds. Deep learning based saliency detectors [12,13] break away from conventions of using hand-crafted features and directly learn the saliency related features with multi-layer convolutional neural networks (CNN) by adding the loss function at the top of CNNs. With the development of CNNs and learning strategies, deep learning based saliency detectors show superior saliency detection performance on almost all benchmark datasets. However, training CNNs for saliency detection requires numerous of pixel-level images and high performance GPUs.

According to the above classification principle, almost all existing co-saliency detectors [14–18] are heuristic models. Co-saliency detection primarily identifies the common salient objects in a group of images. Whereas, deep learning models experience difficulty in learning inter relationships in image groups owing to the lacking of training image groups. To the best of our knowledge, the two widely used co-saliency detection datasets, iCoseg and Image pair, possess 37 and 105 groups of images, respectively. The quantity of image groups significantly deviates from learning a CNN with millions of parameters. The latest work of [19] proposes co-salient object detection based on deep learning in which the inter-relationship of two images is measured by a three layers fully connected neural network with superpixel features as inputs. Unlike the deep learning based co-saliency detector, our method does not need lots of training images, and can be easily extend to other computer vision applications, e.g., co-segmentation.

In this paper, we propose saliency and co-saliency detection with two-level exemplars, i.e., image-level and region-level exemplars. Our method requires a small number of images that having similar foreground with query image as image-level exemplars. Then, region-level exemplars are formed with foreground regions

* Corresponding author at: School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China.
*E-mail addresses:* rhuang@cauc.edu.cn (R. Huang), wfeng@ieee.org (W. Feng), zzwang@tju.edu.cn (Z. Wang), yxing@cauc.edu.cn (Y. Xing), zyb@ctgu.edu.cn (Y. Zou).

https://doi.org/10.1016/j.neucom.2019.09.011
0925-2312/© 2019 Published by Elsevier B.V.

of image-level exemplars and background regions of the query image. We use two simple SVM classifiers, i.e., two-class SVM and one-class SVM, for saliency and co-saliency evaluations, respectively. For saliency detection, we use the regions from the retrieved similar images as positive exemplars and the regions of the image boundaries of the query image as negative exemplars for two-class SVM to identify saliency parts. For co-saliency detection, the regions in each image correspond to the positive exemplars for the one-class SVM to identify common parts. Compared with conventional heuristic saliency detectors, our method learns from labeled images to recognize salient regions and robust for the images with complex scenes. Compared with deep learning based saliency detectors, our method uses less number of images and can be easier to extend to novel data, such as mural images. We compare the proposed method with state-of-the-art methods on saliency and co-saliency detection benchmark datasets. Our method shows promising performances in both saliency and co-saliency detection tasks. The main contributions of our method are twofold:

- We propose saliency and co-saliency detection methods with a single rationale, i.e., exemplar. Although we do not use end-to-end deep learning strategy, our method shows promising performances on saliency and co-saliency benchmark datasets.
- We propose two levels exemplars, i.e., image-level and region-level exemplars for saliency and co-saliency detection. Image-level exemplars are response for finding similar images. Region-level exemplars are response for propagating saliency values between superpixels.

## 2. Related work

### 2.1. Saliency detection

Saliency starts from the human attention, with the hypothesis that the human vision system only processes parts of an image in detail while leaving others nearly unprocessed. After the pioneering work of Itti et al. [20] which proposes to use centrally surrounded differences across multi-scale image features for image saliency detection, image saliency has been booming in computer vision. However, the subsequent saliency detectors are mainly based on local contrast, resulting in high saliency values near edges instead of the salient objects themselves. The work of Cheng et al. [6] proposes region-level global contrast saliency detection for achieving uniformly highlighted salient objects. Afterward, over-segmenting image into uniform regions (i.e., superpixel) as preprocessing step becomes the main stream for saliency detection [7–9]. The first branch solves saliency detection as label propagation on graph with superpixels as nodes [8,21,22] based on the assumption that the pixels belonging to image boundaries feature high background probabilities. The second branch treats salient object as large reconstruction error with sparse coding [10] and low-rank [9,23] by the assumption that salient objects lie on uniform and low-rank backgrounds. The third branch learns to distinguish salient regions from non-salient backgrounds [7,11]. All the above mentioned saliency detectors belong to heuristic methods, which exhibit high saliency detection precision and high efficiencies. However, the above saliency detectors barely produce satisfactory results for images with complex scenes.

With the development of deep learning in image classification [24], object recognition [25] and semantic segmentation [26], recent saliency detectors [27,28] are prone to using CNNs for accurate and rapid saliency detection. However, deep learning based methods require numerous labeled images and expensive GPUs for training and are lack of generalization ability for unseen scenes. The latest work of Wang et al. [29] proposes to learn CNNs for saliency detection with image-level supervision to avoid pixel-level annotations. However, their method still needs GPUs and complex training strategies, including pre-training on ImageNet object detection dataset, self-training with estimated pixel-level labels and fine-tuning with robust loss. In this paper, we show that using a few labeled images and light classifier (e.g., SVM) can generate accurate saliency detection for general images and images with complex scenes. Our method can also be easily applied for saliency detection in images captured for unusual scenes such as mural.

### 2.2. Co-saliency detection

Co-saliency detection was first introduced by [18,30] for capturing common salient objects in a group of images. This method attracts significant interest in computer vision because of wide applications in image co-segmentation [31], matching [32], and object co-recognition [33]. Modeling co-saliency as the combination of image saliency and inter-image similarity is a widely used strategy in existing co-saliency detectors [9,34–36]. Fu et al. [14] present a cluster-based co-saliency detection method using contrast, spatial and corresponding visual cues. Liu et al. [17] propose to detect co-salient objects using joint fine-scale regional similarity and coarse-scale object prior. Ye et al. [16] provide an effective co-salient object discovery and recovery strategy. Huang et al. [9] propose a Gaussian mixture model based co-saliency prior and obtain multi-image co-saliency by fusing intra-saliency maps and co-saliency priors. In their later work [15], they also propose co-saliency detection based on color feature reinforcement to remove single saliency residuals. Recent works [37,38] also pay attentions on studying how to employ the deep learning methods to detect co-saliency detection. Besides using RGB image, depth information is also used for providing additional cues in recent co-saliency detectors [39,40]. Our co-saliency detector utilizes region-level exemplars from one image to generate co-saliency identification maps for other images. Fusing co-saliency identification maps for an image can help us find co-salient regions in the corresponding image.

### 2.3. Exemplar-based methods

To the best of our knowledge, only a few saliency detection methods explore the idea of exemplar propagation. Ye et al., [41] propose saliency detection by propagating pixel-level labels from similar images from Internet image collections. Retrieval from Internet not only requires high time complexity, but also cannot guarantee the quality of retrieved images. Wang et al., [42] present saliency detection using image-level exemplars with low-level features. However, using low-level features cannot grantee semantic similarity between query image and image collections. In [43], intra-class association and discrimination from background are proposed for deep association of top-down saliency detection. Different from [43], our method uses features of deep CNN to gradually match similar images and uses SVM classier to propagate saliency values from labeled images. Further more, most of the above mentioned saliency detectors separate saliency and co-saliency detection as two independent steps. In this paper, we propose saliency and co-saliency detection based on a single rationale, i.e., exemplars.

## 3. The method

The main framework of the proposed exemplar-based saliency and co-saliency detection is shown in Fig. 1. Saliency detection consists of generating image-level exemplar set and region-level exemplar set, label transferring and refinement. Co-saliency detection includes saliency detection, exemplar label cross propagation, co-saliency fusion and refinement.

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

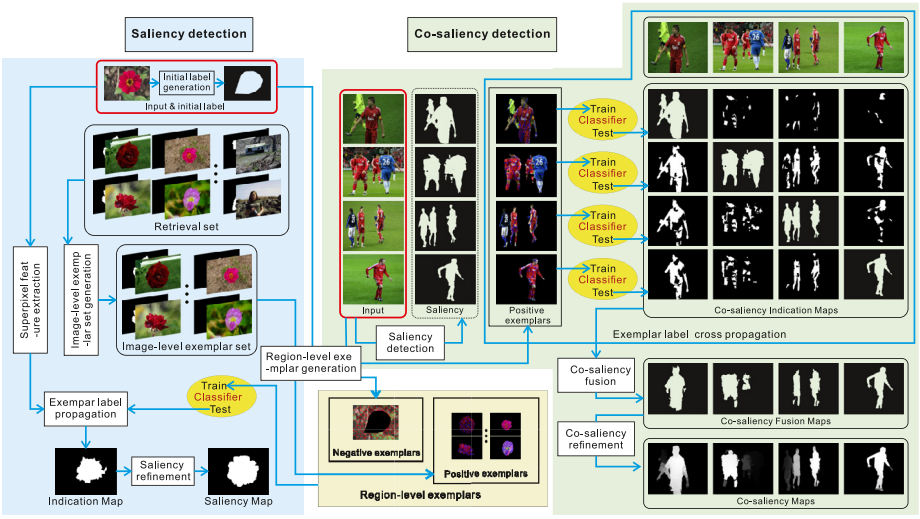



**Fig. 1.** The framework of the proposed exemplar-based image saliency and co-saliency detection model. Please see text for details.

### 3.1. Exemplar-based saliency detection

#### 3.1.1. Image-level exemplar set generation

Given a query image $I_q$, we first find an image-level exemplar set $\mathbb{S}_{\text{ImEx}}$ from a given labeled image dataset $\mathbb{S}_{\text{Retrieval}}$. In this paper, we randomly select 4000 images from MSRA5000 [11] to construct $\mathbb{S}_{\text{Retrieval}}$ because $\mathbb{S}_{\text{Retrieval}}$ should contain diverse images and small enough to decrease the query time. We empirically use 4000 images to form our retrieval set $\mathbb{S}_{\text{Retrieval}}$.

To generate $\mathbb{S}_{\text{ImEx}}$, we use saliency-related deep features as image-level features. Specifically, we crop the image regions according to the bounding box of salient objects and extract deep features of this cropped image. Because the extracted features mainly depict the salient objects, we name them saliency-related deep features. In an image, we use the smallest bounding box containing all salient objects. The initial salient region of $I_q$ is a convex hull. Then we extract convolutional features of the cropped sub image by FCN [44]. The basic architecture of FCN is VGG16, which is designed for image classification and trained with ILSVRC [45]. The network used in this paper is named FCN-8s, which is fine-tuned by the PASCAL VOC2011 segmentation set [46] for semantic segmentation. The cropped sub images are resized into $224 \times 224$ and fed into FCN network. We use feature maps of layers Conv3_3, Conv4_3, Conv5_3, FC6, and FC7 with the sizes of $256 \times 106 \times 106$, $512 \times 53 \times 53$, $512 \times 27 \times 27$, $4096 \times 8 \times 8$ and $4096 \times 8 \times 8$, respectively. Wherein the first number is channel number, the second and third numbers are the heights and widths of the feature maps. These features are used because lower layer features capture the low-level appearance information, middle layer features describe the middle-level texture information and higher layer feature introduce high-level semantic information. We adopt average pooling operation to generate a $D$ dimensional feature vector for feature maps $\mathrm{F} \in \mathbb{R}^{D \times H \times W}$ generated by a convolutional layer, where $D$, $H$ and $W$ refer to channel number, height, and width, respectively. The final saliency-related feature of each layer is denoted as $\mathbf{f}^{\text{layer}}$, where layer belongs to Conv3_3, Conv4_3, Conv5_3, FC6 and FC7.

To find semantic and appearance similar image-level exemplars, we use cascade matching strategy to gradually reduce the matched images. Specifically, we first use the features of FC7 to find 100 semantic similar image-level exemplars. Then we use the features of FC6 to find 80 image-level exemplars from the 100 exemplars. From FC7 to Conv3_3, we finally obtain 10 appearance similar image-level exemplars. We set the numbers of image-level exemplars in FC7, FC6, Conv5_3, Conv4_3 and Conv3_3 to 100, 80, 60, 40 and 10, respectively. The distance between query image $I_q$ with image-level exemplar $I_e$ is defined as

$$d(\mathrm{I}_q, \mathrm{I}_e) = \|\mathbf{f}^{\text{layer}}_{\mathrm{I}_q} - \mathbf{f}^{\text{layer}}_{\mathrm{I}_e}\|^2, \tag{1}$$

where $I_e \in \mathbb{S}_{\text{Retrieval}}$ when using features of FC7. $I_e \in \mathbb{S}^{Sub}_{\text{Retrieval}}$ when using features of other layers. $\mathbb{S}^{Sub}_{\text{Retrieval}}$ is the matched image-level exemplars obtained from previous matching step. Note that we sort the distances by ascending order and select images belonging to the first *#Num* ranked images. *#Num* is decided by the corresponding layer. The final 10 images construct our image-level exemplar set $\mathbb{S}_{\text{ImEx}}$.

#### 3.1.2. Region-level exemplar set generation

We use SLIC [47] to generate superpixels to accelerate the processing speed. Superpixels of salient foreground regions of the images from $\mathbb{S}_{\text{ImEx}}$ and those in the external regions of the convex hull of $I_q$ form positive and negative region-level exemplar set $\mathbb{S}_{\text{Re}}$. We extract superpixel features from deep convolutional features. As $\mathbb{S}_{\text{ImEx}}$ guarantees that image-level exemplars are semantic similar to the query image $I_q$, we use low-level and middle-level features to guarantee similarity in appearances. Specifically, we use features of Conv2_2, Conv3_3, Conv4_3 and Conv5_3 of FCN [44]. For an image I, we first extract the features of the above layers. Then, we resize the feature maps to the size of I. All feature maps are concatenated by channels. The $i$th superpixel feature vector of image I, i.e., $\mathbf{f}_{\mathrm{I},i}$, is formed by the average values of the feature maps in the corresponding superpixel.

#### 3.1.3. Exemplar label propagation

To propagate the labels of region-level exemplars of $\mathbb{S}_{\text{Re}}$, we train a two-class SVM [48] classifier. However, as we know that the number of the positive region-level exemplars are larger than the number of the negative region-level exemplars. Unbalanced positive and negative exemplars result in bias prediction. We resample same number of positive region-level exemplars with negative region-level exemplars to balance the positive and negative region-level training exemplars. This strategy exerts little influence on performance but reduces training time. In our experiment, we use SVM with radial bias kernel. For query image $I_q$, we extract all superpixel features as mentioned in the previous section. Then, the indication map $SI_q$ for salient objects can be obtained by the results of classifying the superpixels of $I_q$.

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

#### 3.1.4. Saliency refinement

Saliency indication map $\mathrm{SI}_q$ only indicates the locations of salient objects and backgrounds, which ignores the neighborhood of the superpixels. Thus, we need further process to generate a smooth saliency map. In this paper, we apply cellular automate propagation [49] to propagate saliency values in whole image. The propagated saliency map is denoted as $\mathrm{SP}_q$. Different from [49], we use LAB color features and the probability maps of FCN [44] to increase the distinguishing ability of the features for calculating superpixel similarity for saliency propagation. Nevertheless, saliency propagation might endow high values to some background regions near salient objects. Thus, we boost the values of salient objects and suppress the values of backgrounds with the following equation:

$$\mathbf{SF}_q(i) = \begin{cases} \sigma\sqrt{2\pi}e^{\frac{\mathbf{SP}_q(i)^2}{2\sigma^2}} & if \quad \mathbf{SP}_q(i) \le 0.5, \\ 1 & else, \end{cases} \tag{2}$$

where $i$ denotes the location of a pixel, $\mathbf{SF}_q$ represents the final saliency map and $\sigma$ is set to 2.

### 3.2. Exemplar-based co-saliency detection

#### 3.2.1. Saliency generation

The rationale behind saliency and co-saliency detection is same, i.e., propagating saliency values from suitable exemplars. However, for co-saliency detection, finding co-salient regions, requires information from intra-group images. $\mathbb{G} = \{\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_K\}$ denotes an image group for co-saliency detection, where $K$ is the number of images in group $\mathbb{G}$. We can generate the saliency maps for images of $\mathbb{G}$ as mentioned in previous section. $\mathbb{G}_{\mathrm{Sa}} = \{\mathrm{SF}_1, \mathrm{SF}_2, \ldots, \mathrm{SF}_K\}$ denotes the saliency map group. We use a threshold to segment each saliency map of $\mathbb{G}_{\mathrm{Sa}}$ to obtain a salient region mask set $\mathbb{G}_{\mathrm{Mask}} = \{\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_K\}$. For simplicity, we set threshold $T_k = 1.5 * mean((\mathrm{SF}_k(i)))$, where $mean(\cdot)$ is a mean function. Thus, we can obtain $\mathbf{M}_k$ by

$$\mathbf{M}_k(i) = \begin{cases} 0 & if \quad \mathbf{M}_k(i) \le T_k, \\ 1 & else. \end{cases} \tag{3}$$

#### 3.2.2. Exemplar label cross propagation

As shown in Fig. 1, the core of co-saliency detection is generating co-saliency indication map. The salient foreground regions in each image can be considered as region-level exemplars. However, for co-saliency detection, the convolutional features of FCN cannot distinguish co-salient parts from salient ones because co-saliency relies on low-level features, e.g., color, more than semantic deep features. Thus, we use average RGB and LAB features in a superpixel to represent region-level features. In addition, we opt not to use negative exemplars in saliency detection for unsuitable negative exemplars that fail to remove unco-salient regions.

In general, the foreground superpixels from an image $\mathrm{I}_k$ form a region-level exemplar set $\mathbb{S}_{\mathrm{Re}}$. We train a one-class SVM classifier to classify the foreground regions of the remaining images $\{\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_{k-1}, \mathrm{I}_{k+1}, \ldots, \mathrm{I}_K\}$ to generate co-saliency indication maps by $\mathrm{SI}_{k,l}$, where $l = [1, K]$ and $l \neq k$. $\mathrm{SI}_{k,l}$ refers to the co-salient regions of image $\mathrm{I}_l$ to exemplar image $\mathrm{I}_k$. However, as shown in Fig. 1, we find that only a few regions are recognized as co-salient in some indication maps with whole salient objects featuring similar colors. To resolve this problem, we refine each co-saliency indication map $\mathrm{SI}_{k,l}$ as proposed in Section 3.1.4. The refined results are bi-segmented with Eq. (3) to generate an improved co-saliency indication map $\mathrm{SI}'_{k,l}$.

#### 3.2.3. Co-saliency fusion

$\mathrm{SI}'_{k,l}$ cannot indicate whether a superpixel is co-salient in image group $\mathbb{G}$. To generate group-extended co-saliency and suppress single saliency residuals [15], we use hybrid fusion strategy by combining addition and multiplication to integrate $\mathrm{SI}'_{k,l}$. The fused co-saliency map can be obtained by

$$\mathbf{CS}_k^{\mathrm{fuse}}(p) = \begin{cases} \frac{1}{K}\sum_{l=1}^{K}\mathrm{SI}'_{k,l}(p), & if \ \sum_{j=1}^{K}\mathrm{SI}'_{k,l}(p) > \frac{K}{2} \\ \prod_{l=1}^{K}\mathrm{SI}'_{k,l}, & else. \end{cases} \tag{4}$$

where $\sum_{l=1}^{K}\mathrm{SI}'_{k,l}(p)$ is the number of images voting superpixel $p$ in image $k$ to be co-saliency.

#### 3.2.4. Co-saliency refinement

The co-salient objects may be incomplete in fused co-saliency map $\mathbf{CS}_k^{\mathrm{fuse}}$. Thus, we refine the co-saliency results with the similar manner for saliency detection. However, we restrict co-saliency as a subset of saliency detection. Thus, for co-saliency refinement, we use three types of features (i.e., LAB, saliency map and the probability maps of FCN [44]) to calculate the similarity between two superpixels. Finally, we obtain the co-saliency map $\mathbf{CS}_k^{\mathrm{F}}$.

## 4. Experimental results

### 4.1. Criteria

We use average precision, recall, $\mathrm{F}_\beta$-measure, Mean absolute error (MAE) and AUC to evaluate different saliency and co-saliency detection methods. We set $\beta^2$ to 0.3 to emphasize precision.

### 4.2. Saliency detection

#### 4.2.1. Setting

*Dataset*. We use four benchmark datasets, i.e., MSRA5000 [11], MSRA1000 [50], ECSSD [51] and PASCAL-S [52] in our experiments. In addition to the 4000 images in retrieval set $\mathbb{S}_{\mathrm{Retrieval}}$, the remaining 1000 images of MSRA5000 [11] are used to evaluate the performance of different saliency detection methods.

*Baselines*. We compare our exemplar based saliency detection method (EBIS) with 13 state-of-the-art saliency detectors, including DBS [42], DRFI [53], FT [50], GS [21], HS [51], RA [54], SF [55], DSR [10], GRSD [56], HDCT [57], MR [8], GB [58] and wCtr [22].

#### 4.2.2. Results and analysis

*Qualitative comparison.* We show two groups of images for saliency detection comparison in Fig. 2. The first group contains three images with simple scenes. From the 2nd and 3rd rows of the first group, we can find that most of saliency detectors can produce good saliency predictions for images with distinguishable foregrounds and backgrounds. However, when two or more regions with notable appearances (the 1st row) exist, the compared methods may be confused by the scenes. EBIS can identify the right salient object (the old man in the 1st row) using deep semantic features.

The second group contains three images with complex scenes, such as an image contains a green bus foreground and green trees in the background and another image shows a white-dressed man standing before a white wall. These images feature similar background and foreground. From Fig. 2, we can find that most of the compared saliency detectors fail to detect salient objects accurately. However, EBIS still generates more accurate salient objects and clearer backgrounds than the other methods.

*Quantitative comparison.* We show the detailed quantitative comparison of F-measure, MAE and AUC in Fig. 3 and Table 1. As shown in Table 1, our F-measures rank first on MSRA5000 [11] and ECSSD [51], and rank second on MSRA1000 [50], and PASCAL-S [52]. We also find that the third best saliency detector is wCtr [22]. Specially, EBIS is the only one among all saliency detectors whose F-measure reaches higher than 0.7 on ECSSD

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

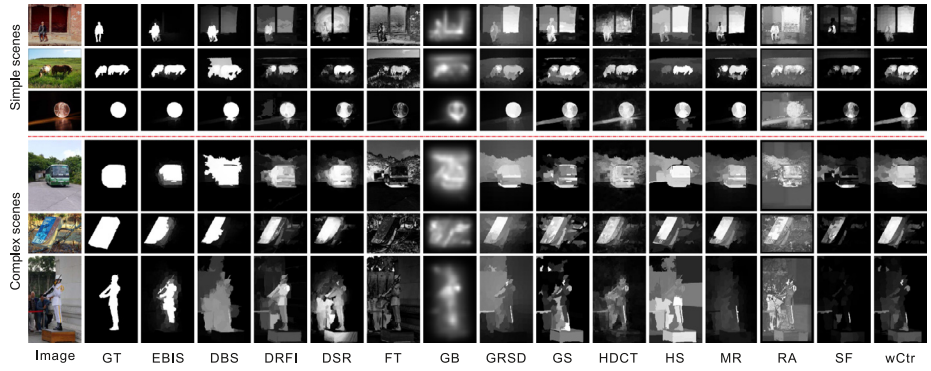



**Fig. 2.** Some examples of saliency detection results of different saliency detection methods.

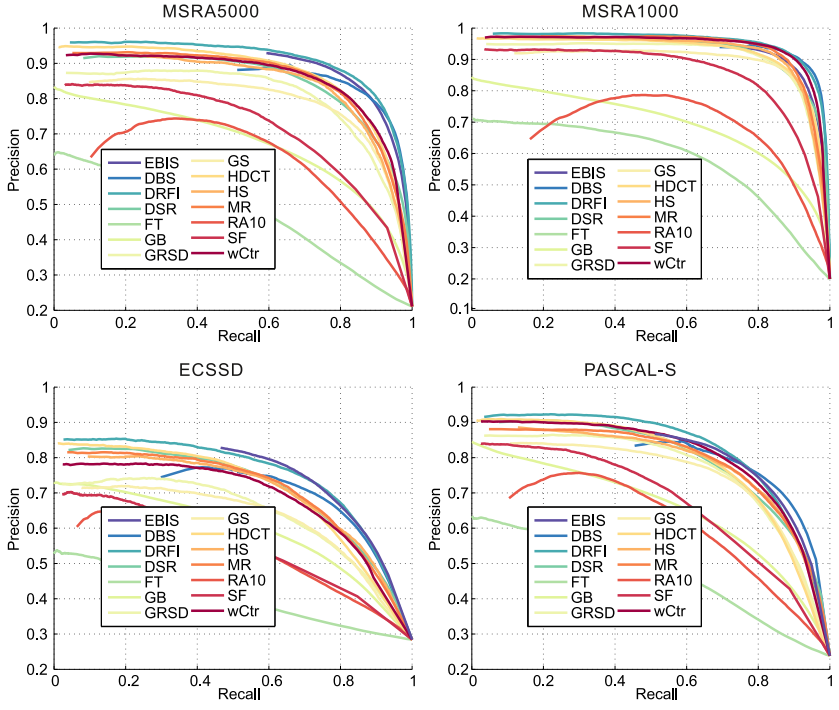



**Fig. 3.** Comparison of PR curves of different saliency methods on four saliency detection benchmark datasets.

**Table 1**
Comparison of F-measure, MAE and AUC of different saliency methods on four saliency detection benchmark datasets. Bold, italic and bold-italic fonts denote ranking first, second and third, respectively.

| Dataset | Criteria | EBIS | DBS | DRFI | DSR | FT | GB | GRSD | GS | HDCT | HS | MR | RA | SF | wCtr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | F-measure | **0.833** | ***0.825*** | *0.797* | 0.775 | 0.466 | 0.604 | 0.704 | 0.719 | 0.766 | 0.754 | 0.786 | 0.555 | 0.584 | 0.786 |
| MSRA5000 | MAE | ***0.095*** | **0.090** | 0.121 | 0.122 | 0.251 | 0.229 | 0.215 | 0.146 | 0.144 | 0.165 | 0.128 | 0.323 | 0.168 | *0.111* |
| | AUC | ***0.862*** | *0.861* | **0.875** | 0.853 | 0.680 | 0.823 | 0.849 | 0.846 | 0.864 | 0.848 | 0.851 | 0.801 | 0.776 | 0.854 |
| | F-measure | ***0.877*** | **0.902** | 0.844 | 0.850 | 0.538 | 0.617 | 0.785 | 0.796 | 0.813 | 0.830 | 0.870 | 0.581 | 0.719 | *0.870* |
| MSRA1000 | MAE | ***0.065*** | **0.047** | 0.090 | 0.080 | 0.219 | 0.215 | 0.161 | 0.107 | 0.115 | 0.111 | 0.075 | 0.308 | 0.128 | *0.065* |
| | AUC | 0.878 | ***0.887*** | **0.893** | 0.882 | 0.757 | 0.843 | 0.882 | 0.882 | *0.885* | 0.882 | 0.879 | 0.841 | 0.847 | 0.886 |
| | F-measure | **0.721** | ***0.693*** | *0.683* | 0.661 | 0.386 | 0.560 | 0.596 | 0.601 | 0.650 | 0.655 | 0.665 | 0.520 | 0.411 | 0.650 |
| ECSSD | MAE | **0.201** | ***0.208*** | 0.225 | 0.225 | 0.328 | 0.305 | 0.316 | 0.254 | 0.248 | 0.268 | 0.236 | 0.367 | 0.267 | *0.224* |
| | AUC | ***0.792*** | 0.785 | **0.801** | 0.783 | 0.603 | 0.760 | 0.762 | 0.766 | *0.786* | 0.777 | 0.778 | 0.729 | 0.679 | 0.775 |
| | F-measure | ***0.764*** | **0.769** | 0.743 | 0.733 | 0.461 | 0.599 | 0.686 | 0.694 | 0.716 | 0.717 | 0.736 | 0.560 | 0.559 | *0.748* |
| PASCAL-S | MAE | **0.114** | ***0.120*** | 0.140 | 0.133 | 0.250 | 0.233 | 0.228 | 0.161 | 0.159 | 0.191 | 0.157 | 0.326 | 0.161 | *0.128* |
| | AUC | 0.841 | 0.841 | **0.8564** | 0.840 | 0.677 | 0.822 | 0.829 | 0.838 | ***0.847*** | 0.831 | 0.826 | 0.793 | 0.758 | *0.843* |

dataset. Note that different from other three saliency datasets, ECSSD contains images with more complex scenes. Compared with the 2nd ranked DBS [42] (0.693), EBIS achieves 4.04% relative F-measure improvement on ECSSD dataset. The average F-measures of EBIS, DBS [42] and wCtr [22] across four benchmark datasets are 0.7988, 0.7973 and 0.7635, respectively. Compared with wCtr [22], EBIS achieves 4.62% relative improvement across four benchmark datasets. Our MAE ranks first on ECSSD [51] and PASCAL-S [52] and second on MSRA5000 [11] and MSRA1000 [50]. The average MAE of EBIS, DBS [42] and wCtr [22] across four benchmark datasets reach 0.1187, 0.1162 and 0.1320, respectively. EBIS is a bit higher than DBS [42] by an absolute difference of 0.0025. Compared with

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

**Table 2**
Comparison of average running time of different saliency detectors. "-" denotes lacking of running time.

| - | EBIS | DBS | DRFI | DSR | FT | GB | GRSD | GS | HDCT | HS | MR | RA | SF | wCtr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time (s) | 78.1 | - | 19.2 | 0.42 | 0.19 | 0.31 | 0.75 | 0.61 | 3.8 | 1.07 | 1.19 | 7.89 | - | 0.47 |

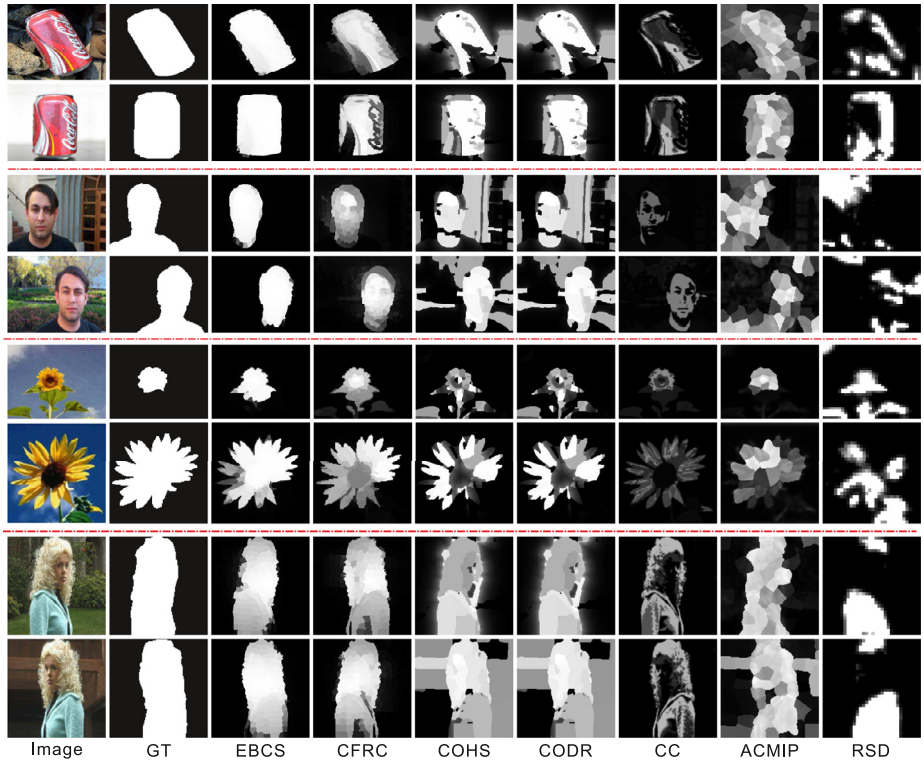



**Fig. 4.** Four examples of co-saliency detection of different co-saliency detectors on Imagepair [34] dataset.

wCtr [22], EBIS achieves 10.08% relative MAE reduction across four benchmark datasets. The AUC values of DRFI [53] rank first on all four benchmark datasets. However, its F-measure and MAE are not as good as EBIS.

The average running time of processing an image with size of $300 \times 400$ for different saliency detectors via Matlab R2016a-64bit on a laptop equipped with an i7 2.2 GHz GPU and 16 GB RAM is shown in Table. 2. EBIS takes about 78.1 s for evaluating an image, which is slower than other compared methods. Note that EBIS needs extracting deep features from FCN for query image and exemplar images. The speed of EBIS can be accelerated by saving the features of exemplar images or using GPU. To verify it, we test our method via Matlab R2018a-64bit on a computer equipped with an i7 3.6 GHz CPU, 16 GB RAM and a GTX 980 GPU. We find that (1) with pre-calculated features of region-level exemplar set and CPU, our method takes about 22.38 s to process a $300 \times 400$ image; (2) with pre-calculated features of region-level exemplar set and GPU, our method takes about 13.29 s to process the image.

### 4.3. Co-saliency detection

#### 4.3.1. Setting

*Dataset*. Two benchmark datasets, including Imagepair [34] and iCoseg [59] are used for co-saliency evaluation. Imagepair [34] contains 105 pairs of images. Both pair of images contain simple scenes and single salient object. iCoseg [59] contains 37 groups of images, in which each group includes five images having complex scenes.

*Baselines*. We compare the proposed exemplar based co-saliency detectors (EBCS) with six state-of-the-art co-saliency detectors, including CFRC [15], COHS [17], CODR [16], CC [14], ACMIP [34] and RSD [60] on two co-saliency detection benchmark datasets. As ACMIP [34] and RSD [60] are designed for detecting co-saliency for image pairs, they are excluded from comparison on the iCoseg [59] dataset.

#### 4.3.2. Results and analysis

*Qualitative comparison*. We show several typical co-saliency detection results on Imagepair [34] and iCoseg [59] in Fig. 4 and Fig. 6, respectively. As shown in Fig. 4, CFRC [15] is the best co-saliency detector among all compared co-saliency detectors. Although CFRC [15] can accurately detect co-salient objects, it cannot uniformly highlight the whole co-salient objects and cannot fully suppress the backgrounds. Co-saliency detectors COHS [17], CODR [16] and ACMIP [34] not only detect co-salient objects but also misclassify some background regions into co-salient parts. The co-salient objects of CC [14] have low co-saliency values and are incomplete. Note that higher co-saliency value refers to higher confidence of the corresponding pixels belonging to co-salient parts. RSD [60] fails to detect co-salient objects (i.e., the man's head) in the second group of images. On the contrary, our proposed co-saliency detector, EBCS, can uniformly highlight the co-salient objects and suppress backgrounds.

We show two groups of co-saliency detection on iCoseg [59] in Fig. 6. The first group of images features complex backgrounds and single saliency residuals. For the second image of the first group, EBCS is the only co-saliency detector that suppresses the objects located at the top-left image. CFRC [15] exhibits the best co-saliency detection among the compared methods, whose results are clean in background regions. However, the co-saliency

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

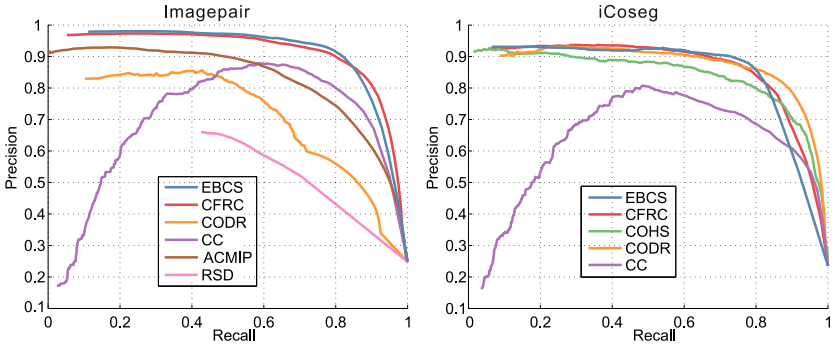



**Fig. 5.** Comparison of PR curves of different saliency methods on two co-saliency detection benchmark datasets.

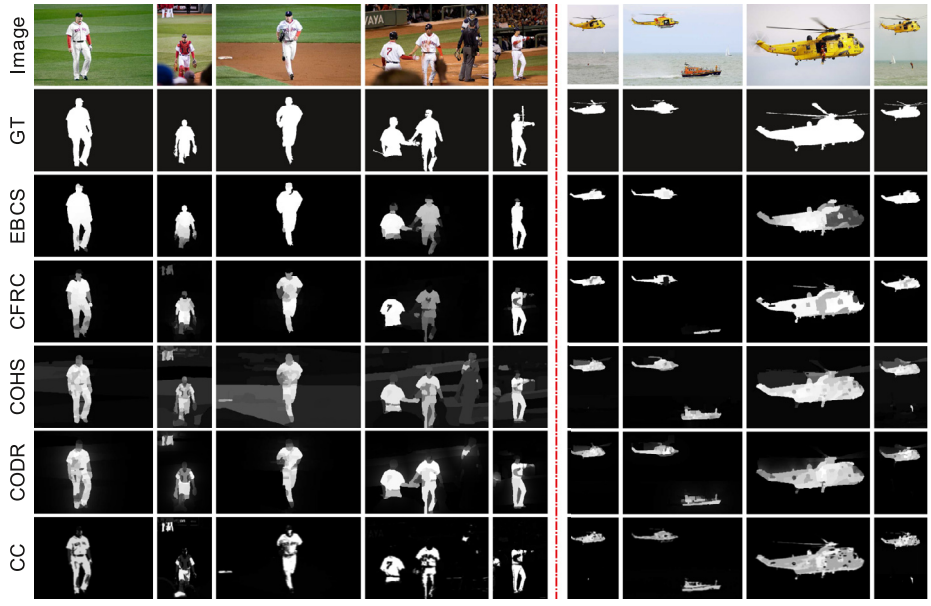



**Fig. 6.** Two examples of co-saliency detection of different co-saliency detectors on iCoseg [59] dataset.

**Table 3**
Comparison of F-measure, MAE and AUC of different co-saliency detectors on two benchmark datasets. We use bold, italic and bold-italic fonts to denote the first, second and third ranked methods, respectively.

| Dataset | Criteria | EBCS | CFRC | COHS | CODR | CC | ACMIP | RSD |
|---|---|---|---|---|---|---|---|---|
| Imagepair | F-measure | **0.867** | ***0.829*** | 0.657 | 0.657 | 0.497 | *0.715* | 0.598 |
| | MAE | **0.092** | ***0.117*** | 0.230 | 0.230 | 0.187 | *0.178* | 0.203 |
| | AUC | ***0.834*** | **0.847** | 0.767 | 0.767 | 0.809 | *0.822* | 0.724 |
| iCoseg | F-measure | **0.833** | ***0.798*** | 0.727 | *0.790* | 0.481 | – | – |
| | MAE | **0.107** | *0.132* | 0.192 | ***0.118*** | 0.189 | – | – |
| | AUC | 0.817 | *0.837* | ***0.851*** | **0.858** | 0.812 | – | – |

values of co-salient objects are lower than that of EBCS. Although COHS [17] detects co-salient objects in each image, it also detects single saliency residuals (i.e., black dressed referee) and cannot suppress the background satisfactorily. Compared with COHS [17], CODR [16] can suppress the background more effectively. Clustering-based co-saliency detector CC [14] can find the main colors of co-salient objects, thus resulting in incomplete objects featuring different colors. The second group of images presents a clear background. We can find that all co-saliency detectors can detect the co-salient helicopter. However, except for the results of EBCS, the findings of other co-saliency detectors feature single saliency residuals (i.e., boat) in the second image. To sum up, EBCS can generate precise, smooth and complete co-salient objects.

*Quantitative comparison.* We show quantitative comparison of different co-saliency detectors in Fig. 5 and Table 3. Besides AUC values of EBCS, the values of F-measure and MAE are higher than other compared methods. On Imagepair [34] dataset, the first, second and third ranked co-saliency detectors are EBCS, CFRC [15] and ACMIP [34], respectively. Specifically, compared with the second ranked CFRC [15] and third ranked ACMIP [34], EBCS achieves 4.58% and 21.26% relative F-measure improvements, and achieves 21.37% and 48.31% relative MAE reductions. On iCoseg [59] dataset, F-measure of EBCS is 0.833 and ranks first. F-measures of CFRC [15] and CODR [16] are 0.798 and 0.79, ranking second and third, respectively. Note that F-measure of EBCS is the only variable that exceeds 0.8. The MAE value of EBCS is 0.011 lower than the second ranked CODR [16]. The average F-measures

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

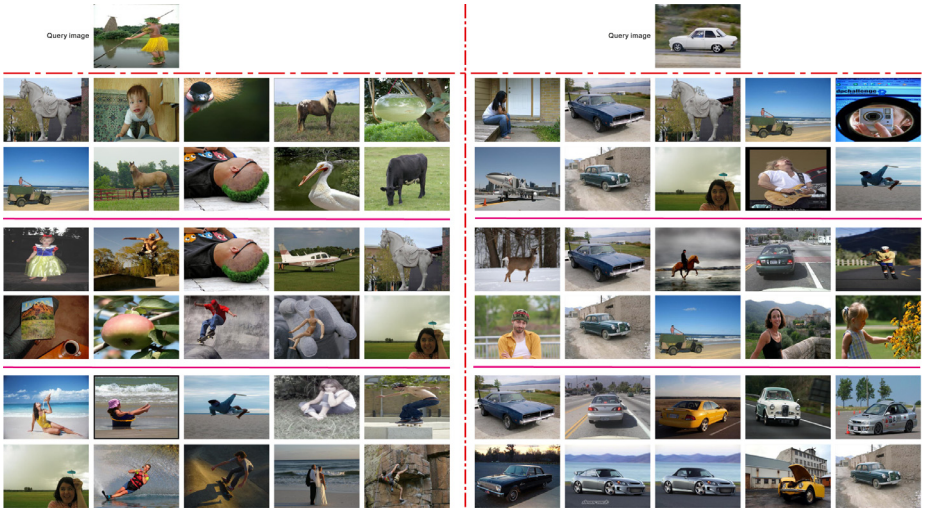



**Fig. 7.** Two examples of constructing image-level exemplar sets for two given query images. The retrieved sets by using the first, second and third strategies are separated by the pink lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 8.** An example of co-saliency detection. The images from the first row to the last row are inputs, ground truth, saliency detection maps and co-saliency detection maps, respectively.

**Table 4**
Comparison of running time for processing the image pair "Mscara" of different co-saliency detectors. "-" denotes lacking of running time.

| Criteria | EBCS | CFRC | COHS | CODR | CC | ACMIP | RSD |
|---|---|---|---|---|---|---|---|
| Time (s) | 130.47 | 3.68 | 48.41 | 53.94 | 0.79 | - | 49.7 |

on two co-saliency datasets of EBCS and CFRC [15] reach 0.85 and 0.8135, respectively. Compared with CFRC [15], EBCS achieves 4.5% relative average F-measure improvement. The average MAEs on two co-saliency datasets of EBCS and CFRC [15] reach 0.0995 and 0.1245, respectively. Compared with CFRC [15], EBCS achieves 20.1% relative average MAE reduction. Generally speaking, EBCS shows better co-saliency detection performance on both Image-pair [34] and iCoseg [59] datasets.

The running time for processing the image pair "Mscara" of different co-saliency detectors is shown in Table. 4. All methods are tested on the same computer that used for saliency detection. EBCS takes about 130.5 s for computing the co-saliency maps for the image pair. Compared with other co-saliency detectors, EBCS is the slowest. However, the running time of EBCS can be reduced by employing GPU or pre-calculated features of exemplars as well as EBIS. We run our code with "Mscara" image pair by using GPU or pre-calculated features. The total running time of using GPU without pre-calculated features is 60.9 seconds. And the running time can be further reduced to 29 seconds by using GPU with pre-calculated features.

### 4.4. Discussion

#### 4.4.1. Different retrieving strategies

We show two examples of image-level exemplar sets for two given query images in Fig. 7. We use three different strategies to retrieve images. The first strategy is retrieving by concatenating all features of Conv3_3, Conv4_3, Conv5_3, FC6 and FC7 as a single feature vector. The second strategy is cascade retrieving with features of Conv3_3, Conv4_3, Conv5_3, FC6 and FC7. The third strategy is cascade retrieving with features of FC7, FC6, Conv5_3, Conv4_3 and Conv3_3. We set the numbers of image-level

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

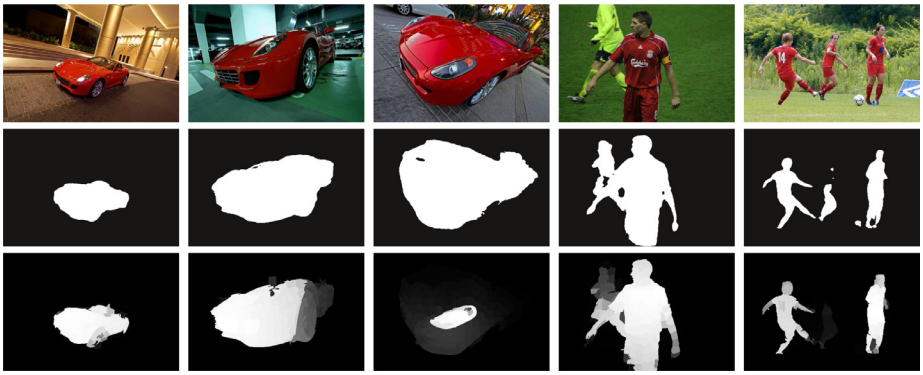


**Fig. 9.** A failure case of co-saliency detection. The images from the first row to the last row are inputs, saliency detection maps and co-saliency detection maps, respectively.

exemplars to 100, 80, 60, 40 and 10, respectively. From Fig. 7, we have three observations. (1) Retrieving by concatenating all features cannot find suitable image-level exemplars. (2) When retrieving by low-level features (e.g., Conv3_3) first and then retrieving by high-level features (e.g., FC7), lots of semantic similar images are filter out. There are only few images similar with query image. (3) Retrieving by using high-level features (e.g., FC7) first, results in high recall of semantic similar images with query image. The third strategy used in this paper is beneficial for generating better image-level exemplar set.

*4.4.2. Low-level v.s. high-level features*

The purpose of co-saliency detection is to find the salient objects that have common characteristics. High-level semantic features (e.g., FC6 and FC7) are good at distinguishing two different types of objects, which are not suitable for co-saliency detection. As shown in Fig. 8, the saliency detection computed by FCN features can detect the football players in all five images. However, we can find that the easily obtained low-level features are more efficient to distinguish the co-salient red dressed players from players dressed in blue, yellow and white. Although the proposed method is suitable for evaluating co-saliency detection in current co-saliency detection datasets, i.e., imagepair and iCoseg, it cannot deal with the co-saliency detection in image group shown in Fig. 9. We can find that the objects of images are dominated by red color and have different semantic classes. Although the red cars are co-salient objects, almost all the red colored objects are detected by the proposed method.

## 5. Conclusion

We have proposed exemplar based saliency (EBIS) and co-saliency detection (EBCS). We show that both saliency and co-saliency detection can be designed with a single core idea that exemplars can distinguish saliency and co-saliency regions from background regions. To detect saliency, we propose to find image-level exemplars first. Then, saliency values are propagate to the query image by region-level exemplars. Different from saliency detection, in co-saliency detection, we only use one-class SVM to find common regions in other images. Both EBIS and EBCS show promising detection performances on the benchmark datasets. In future work, we want to explore the ability of the deep convolutional neural network in propagating the labels of the exemplars.

## Declaration of Competing Interest

We wish to draw the attention of the Editor to the following facts which may be considered as potential conflicts of interest and to significant financial contributions to this work.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author.

Signed by all authors as follows:

Rui Huang rhuang@cauc.edu.cn
Wei Feng wfeng@ieee.org
Zezheng Wang zzwang@tju.edu.cn
Yan Xing yxing@cauc.edu.cn
Yaobin Zou zyb@ctgu.edu.cn

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61671325 and 61572354), the Natural Science Foundation of Tianjin (18JCQNJC00400), the Fundamental Research Funds for the Central Universities of Civil Aviation University of China (3122018C021 and 3122018C020), the Open Funds Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering (2018SDSJ02), and Scientific Research Foundation of Civil Aviation University of China (600001050115 and 600001050117).

## References

[1] H. Fu, D. Xu, B. Zhang, S. Lin, R.K. Ward, Object-based multiple foreground video co-segmentation via multi-state selection graph, IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 24 (11) (2015) 3415–3424.
[2] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Tran. Image Process. 13 (10) (2004) 1304–1318.
[3] J. Wu, Q. Ruan, Object removal by cross isophotes exemplar-based inpainting, in: Proceedings of the International Conference on Pattern Recognition, 2006, pp. 810–813.
[4] Q. Guo, W. Feng, C. Zhou, C.M. Pun, B. Wu, Structure-regularized compressive tracking with online data-driven sampling, IEEE Tran. Image Process. 26 (12) (2017) 5692–5705.
[5] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, S. Wang, Learning dynamic siamese network for visual object tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1781–1789.

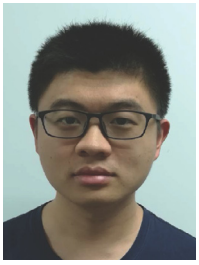
Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

[6] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, S.M. Hu, Global contrast based salient region detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 409–416.
[7] N. Tong, H. Lu, R. Xiang, M.H. Yang, Salient object detection via bootstrap learning, in: Proceedings of the Computer Vision and Pattern Recognition, 2015, pp. 1884–1892.
[8] C. Yang, L. Zhang, H. Lu, R. Xiang, M.H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
[9] R. Huang, W. Feng, J. Sun, Saliency and co-saliency detection by low-rank multiscale fusion, in: Proceedings of the IEEE International Conference on Multimedia and Expo, 2015.
[10] H. Lu, X. Li, L. Zhang, X. Ruan, M.H. Yang, Dense and sparse reconstruction error based saliency descriptor., IEEE Tran. Image Process. 25 (4) (2016) 1592–1603.
[11] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2010) 353–367.
[12] N. Liu, J. Han, DHSNET: deep hierarchical saliency network for salient object detection, in: Proceedings of the Computer Vision and Pattern Recognition, 2016, pp. 678–686.
[13] L. Wang, L. Wang, H. Lu, P. Zhang, R. Xiang, Saliency detection with recurrent fully convolutional networks, in: Proceedings of the European Conference on Computer Vision, 2016, pp. 825–841.
[14] H. Fu, X. Cao, Z. Tu, Cluster-based co-saliency detection, IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 22 (10) (2013) 3766–3778.
[15] R. Huang, W. Feng, J. Sun, Color feature reinforcement for co-saliency detection without single saliency residuals, IEEE Signal Process. Lett. 24 (5) (2017) 569–573.
[16] L. Ye, Z. Liu, J. Li, W.L. Zhao, L. Shen, Co-saliency detection via co-salient object discovery and recovery, IEEE Signal Process. Lett. 22 (11) (2015) 2073–2077.
[17] Z. Liu, W. Zou, L. Li, L. Shen, O.L. Meur, Co-saliency detection based on hierarchical segmentation, IEEE Signal Process. Lett. 21 (1) (2013) 88–92.
[18] D.E. Jacobs, B.G. Dan, E. Shechtman, Cosaliency: where people look when comparing images, in: Proceedings of the ACM Symposium on User Interface Software and Technology, New York, Ny, Usa, October, 2010, pp. 219–228.
[19] D.J. Jeong, I. Hwang, N.I. Cho, Co-salient object detection based on deep saliency networks and seed propagation over an integrated graph, IEEE Tran. Image Process. PP (99) (2017) 5866–5879.
[20] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[21] Y.C. Wei, F. Wen, W.J. Zhu, J. Sun, Geodesic saliency using background priors, in: Proceedings of the European Conference on Computer Vision, 2012, pp. 29–42.
[22] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.
[23] Y. Wu, X. Shen, A unified approach to salient object detection via low rank matrix recovery, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2012, pp. 853–860.
[24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[25] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[26] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.
[27] J. Pan, K. McGuinness, E. Sayrol, N. O'Connor, X. Giro-i Nieto, Shallow and deep convolutional networks for saliency prediction, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2016, pp. 598–606.
[28] L. Wang, H. Lu, R. Xiang, M.H. Yang, Deep networks for saliency detection via local estimation and global search, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2015, pp. 3183–3192.
[29] L. Wang, Y. Wang, H. Lu, X. Ruan, M. Feng, D. Wang, B. Yin, Learning to detect salient objects with image-level supervision, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2017, pp. 3796–3805.
[30] H.T. Chen, Preattentive co-saliency detection, in: Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 1117–1120.
[31] L. Wang, G. Hua, R. Sukthankar, J. Xue, N. Zheng, Video object discovery and co-segmentation with extremely weak supervision, in: Proceedings of the European Conference on Computer Vision, 2014, pp. 640–655.
[32] J. Xue, L. Wang, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting, Pattern Recognit. 46 (11) (2013) 2874–2889.
[33] M. Cho, Y.M. Shin, K.M. Lee, Co-recognition of image pairs by data-driven monte carlo image exploration, in: Proceedings of the European Conference on Computer Vision, 2008, pp. 144–157.
[34] H. Li, K.N. Ngan, A co-saliency model of image pairs, IEEE Trans. Image Process. 20 (12) (2011) 3365–3375.
[35] L. Li, Z. Liu, W. Zou, X. Zhang, O.L. Meur, Co-saliency detection based on region-level fusion and pixel-level refinement, in: Proceedings of the IEEE International Conference on Multimedia and Expo, 2014.
[36] Y. Li, K. Fu, Z. Liu, J. Yang, Efficient saliency-model-guided visual co-saliency detection, IEEE Signal Process. Lett. 22 (5) (2015) 588–592.
[37] L. Wei, S. Zhao, O.E.F. Bourahla, X. Li, F. Wu, Y. Zhuang, Deep group-wise fully convolutional network for co-saliency detection with graph propagation, IEEE Tran. Image Process. (2019).
[38] K. Zhang, T. Li, B. Liu, Q. Liu, Co-saliency detection via mask-guided fully convolutional networks with multi-scale label smoothing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[39] R. Cong, J. Lei, H. Fu, W. Lin, Q. Huang, X. Cao, C. Hou, An iterative co-saliency framework for rgbd images, IEEE Trans. Cybern. PP (99) (2017) 1–14.
[40] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, C. Hou, Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation, IEEE Tran. Image Process. PP (99) (2017) 1-1.
[41] L. Ye, Z. Liu, X. Zhou, L. Shen, J. Zhang, Saliency detection via similar image retrieval, IEEE Signal Process. Lett. 23 (6) (2016) 838–842.
[42] Z. Wang, R. Huang, L. Wan, W. Feng, Exemplar based image salient object detection, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2017.
[43] S. He, R.W.H. Lau, Exemplar-driven top-down saliency detection via deep association, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5723–5732.
[44] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. (IJCV) 115 (3) (2015) 211–252, doi:10.1007/s11263-015-0816-y.
[46] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results, 2011(http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html).
[47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods., IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[48] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[49] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2015, pp. 110–119.
[50] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.
[51] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.
[52] Y. Li, X. Hou, C. Koch, J.M. Rehg, A.L. Yuille, The secrets of salient object segmentation, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2014, pp. 280–287.
[53] H. Jiang, J. Wang, Z. Yuan, Y. Wu, Salient object detection: a discriminative regional feature integration approach, in: Proceedings of the Computer Vision and Pattern Recognition, 2013, pp. 2083–2090.
[54] E. Rahtu, J. Kannala, M. Salo, Segmenting salient objects from images and videos, in: Proceedings of the European Conference on Computer Vision, 2010, pp. 366–379.
[55] P. Krahenbuhl, Saliency filters: contrast based filtering for salient region detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2012, pp. 733–740.
[56] C. Yang, L. Zhang, H. Lu, Graph-regularized saliency detection with convex-hull-based center prior, IEEE Signal Process. Lett. 20 (7) (2013) 637–640.
[57] J. Kim, D. Han, Y.W. Tai, J. Kim, Salient region detection via high-dimensional color transform, in: Proceedings of the Computer Vision and Pattern Recognition, 2014, pp. 883–890.
[58] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Proceedings of the International Conference on Neural Information Processing Systems, 2006, pp. 545–552.
[59] 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.
[60] Y.L. Chen, C.T. Hsu, Implicit rank-sparsity decomposition: applications to saliency/co-saliency detection, in: Proceedings of the International Conference on Pattern Recognition, 2014, pp. 2305–2310.

**Rui Huang** received the Ph.D. from the Tianjin university under the supervised of Professor Wei Feng and Professor Jizhou Sun. Currently, he is a lecturer at the Department of Computer Science and Technology in the Civil Aviation University of China. His research interests include computer vision, machine learning and visual surveillance.

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011

**Wei Feng** received the B.S. and M.Phil. degrees in computer science from Northwestern Polytechnical University, China, in 2000 and 2003, respectively, and the Ph.D. degree in computer science from City University of Hong Kong in 2008. From 2008 to 2010, he was a research fellow at the Chinese University of Hong Kong and City University of Hong Kong, respectively. He is currently a full professor in the College of Intelligence and Computing, Tianjin University, Tianjin University. His major research interest is active robotic vision and visual intelligence, specifically including active camera relocalization and lighting recurrence, general Markov Random Fields modeling, discrete/continuous energy minimization, image segmentation, active 3D scene perception, SLAM, and generic pattern recognition. Recently, he also focuses on solving preventive conservation problems of cultural heritages via computer vision and machine learning. He got the support of the Program for New Century Excellent Talents in University, China, in 2011. He is an associate editor of Journal of Ambient Intelligence and Humanized Computing and a member of the IEEE.



**Zezheng Wang** received the B.S. and M.S. degrees from Tianjin University, Tianjin, China, in 2015 and 2018, respectively. He is currently an Engineer with JD Finance. His current research interests include pattern recognition and intelligent systems, machine learning, signal processing, and face anti-spoofing.

**Yan Xing** received the Ph.D. from the China University of Mining and Technology under the supervised of Professor Fanrong Meng. Currently, she is a lecturer at the Department of Computer Science and Technology in the Civil Aviation University of China. Her research interests include data mining and machine learning.

**Yaobin Zou** is an associate professor at the Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering and College of Computer and Information Technology, China Three Gorges University.He obtained his Ph.D. degree in Computer Science and Technology at Huazhong University of Science and Technology in 2011. His research interest covers digital image processing and pattern recognition.

Please cite this article as: R. Huang, W. Feng and Z. Wang et al., Exemplar-based image saliency and co-saliency detection, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.09.011