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