Neurocomputing 371 (2020) 137–146

Contents lists available at ScienceDirect

Neurocomputing

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

Co-saliency detection via integration of multi-layer convolutional features and inter-image propagation

Jingru Ren[a,b], Zhi Liu[a,b,*], Xiaofei Zhou[c], Cong Bai[d], Guangling Sun[b]

[a] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
[b] School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
[c] Institute of Information and Control, Hangzhou Dianzi University, Hangzhou 310018, China
[d] College of Computer Science, Zhejiang University of Technology, Hangzhou 311122, China

ARTICLE INFO

Article history:
Received 19 September 2018
Revised 19 August 2019
Accepted 2 September 2019
Available online 13 September 2019

Communicated by Dr. Li Bing

Keywords:
Co-saliency detection
Convolutional neural network
Feature integration
Saliency propagation

ABSTRACT

Convolutional neural networks have been successfully applied to detect salient objects in an image. However, how to better use convolutional features for co-saliency detection, which is an emerging branch of saliency detection, is not fully explored. This paper proposes a convolutional neural network based co-saliency detection model, which consists of two key parts including the integration of multi-layer convolutional features extracted from a group of images and the inter-image saliency propagation. Firstly, the input image and its four co-images belonging to the same image category are passed through the VGG16 model, to obtain the multi-layer convolutional features of these images. Secondly, multi-scale synthesized feature maps, which contain both internal features and correlative features, are generated by integrating the multi-layer convolutional features. Thirdly, via the integration of low-level boundary features and high-level semantic features, the multi-scale synthesized feature maps are enhanced and fused together to generate the initial co-saliency map. Finally, an inter-image saliency propagation method is utilized to refine the initial co-saliency map, yielding the final co-saliency map with the improved quality. Experimental results on two public datasets demonstrate that the proposed model achieves the best performance compared to the state-of-the-art co-saliency detection models.

1. Introduction

By simulating human visual attention mechanism, saliency detection [1–4] aims to identify the visual objects of interest in a natural scene automatically. It has been widely studied as a fundamental problem in many computer vision tasks, such as content-based image retrieval [5–7], salient object segmentation [8,9], semantic segmentation [10], and scene classification [11]. Besides, saliency detection has many branches like co-saliency detection [12–20], RGBD saliency detection [21,22], and video saliency detection [23–27]. As an important issue in saliency detection, co-saliency detection, which devotes to highlight the common salient objects in a group of relevant images, can be applied to many areas, such as object co-segmentation [28–30], object co-recognition [31], and weakly supervised localization [32].

Image saliency detection, generally speaking, the single-image saliency detection, has been uninterruptedly studied for decades. In [1], a novel saliency detection framework, saliency tree, is proposed to provide a hierarchical representation of saliency for generating high-quality regional and pixel-wise saliency maps. In [33], the saliency map computation is regarded as a regression problem, which uses the supervised learning approach to map the regional feature vector to a saliency score, and finally fuses the saliency scores across multiple levels. To make full use of boundary prior, a robust background measure, called boundary connectivity, is proposed in [34], which characterizes the spatial layout of image regions with respect to image boundaries.

There also exist some propagation based saliency detection models. In an ordinary saliency propagation model, the input image is first segmented to many regions. Then, these regions constitute a close-loop graph, in which the adjacent regions are connected by the weighted edges. The saliency values are finally iteratively propagated along the edges from the labeled regions to unlabeled regions. For example, Zhang et al. [35] ranked the similarity of image elements with foreground or background cues via graph-based manifold ranking, in which the saliency values of image elements are defined based on their relevance to the given seeds or queries. In [36], a propagation algorithm that employs the teaching-to-learn and learning-to-teach strategies is proposed to explicitly improve the propagation quality, and the propagation

* Corresponding author at: Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

E-mail address: liuzhisjtu@163.com (Z. Liu).

https://doi.org/10.1016/j.neucom.2019.09.010
0925-2312/

sequence is manipulated from simple regions to difficult regions. An absorbing Markov chain based saliency model is proposed in [37], which achieves a learnt transition probability matrix by the sparse-to-full method combined with multiple-layer deep features, and an angular embedding technique is exploited to refine the saliency maps.

In the past several years, numerous co-saliency detection models have been proposed. Fu et al. [12] introduced a two-layer cluster-based model, which exploits contrast cue, spatial cue and corresponding cue to cluster the pixels from single image and multiple images, respectively. In [13], the region similarity and contrast are measured on the fine segmentation result while the object prior is measured on the coarse segmentation result, and then the three measures are integrated with global similarity of each region to obtain co-saliency maps. To weight many saliency maps generated by the existing saliency models self-adaptively, Cao et al. [14] formalized the rank constraint between these saliency maps to obtain final co-saliency maps. In [15], the co-salient exemplars, structured by color and SIFT features, are propagated to perform the local and global recovery of co-salient object regions, and the foci of attention area is employed to further improve the quality of co-saliency maps. Zhang et al. [16] proposed a novel framework by introducing the deep and wide information for co-saliency detection, namely the deep information captures the concept-level properties of co-salient objects, and the wide information uses cross-group information to suppress the common background regions in the image group. In [17], a new objective function, which imposes a metric learning regularization constraint into SVM training, is optimized to jointly learn discriminative feature representation and co-salient object detector.

With the development of deep learning technique, deep neural network, especially the convolutional neural network yields unusually brilliant results in the field of computer vision. Many researchers have applied deep neural networks to saliency detection. In [38], the recurrent architecture is designed to automatically learn to refine the saliency map by correcting its previous errors. In [39], the localization to refinement network recurrently focuses on the spatial distribution of various scenarios and helps to refine the saliency map by the relations between each pixel and their neighbors. By using deep neural networks, the performance of saliency detection has been pushed forward significantly. Certainly, deep learning has also been applied to co-saliency detection. Jeong et al. [18] utilized deep saliency networks to transfer co-saliency prior knowledge and capture high-level semantic information, and the obtained co-saliency maps are further improved by seed propagation over an integrated graph. This work directly feeds the high-level convolutional features and low-level handcrafted features of each segment into a fully-connected network, which may ignore the position information between pixels for effective co-saliency detection. In contrast, the fully convolutional network can overcome the shortcoming and is more suitable for handling co-saliency detection task. Wei et al. [19] proposed an end-to-end group-wise deep co-saliency detection model based on the fully convolutional network, but this model just exploits the features from the last convolution layer and lacks sufficient utilization of all convolutional features from the whole group of images.

Based on the above analysis, the convolutional neural network based co-saliency detection model is underexplored. Therefore, in this paper, we propose a novel co-saliency detection model, which fully exploits multi-layer convolutional features of a group of images and effectively performs inter-image saliency propagation to achieve the better co-saliency detection performance. As shown in Fig. 1, the input image, together with its four images selected from the same image category with the input image, form an image group for co-saliency detection. These images in the image group are first fed into the VGG16 model [40] to obtain the multi-layer convolutional features. Second, these convolutional features are integrated to multi-scale synthesized feature maps, which contain the internal features of the input image and the correlative features of the whole image group. Third, in order to further utilize boundary features and semantic features of the input image, the low-level and high-level convolutional features are blended with the synthesized feature maps, and the resulting multi-scale enhanced feature maps are fused together to generate the initial co-saliency map. Finally, an inter-image saliency propagation method is exploited to improve the quality of the initial co-saliency map, yielding the final co-saliency map.

The feature integration is an important part in our co-saliency model. Comparing with the existing feature integration mechanisms adopted by some deep learning based saliency models, the feature integration mechanism proposed in our model is better suited for the co-saliency detection task. For example, in [41], the global context features and the local context features are both taken into account, and are integrated in a unified multi-context deep learning framework. This feature integration mechanism is effective for single-image saliency detection, but as for the loss of correlative information of co-salient objects, it is inappropriate for co-saliency detection. Besides, in our co-saliency model, the synthesized feature maps contain the internal features of the input image and the correlative features of the whole image group. In contrast to the deep learning based co-saliency models, our feature integration mechanism sufficiently utilizes the convolutional features. For example, in [19], the convolutional features are extracted from the single layer, and the feature integration is performed under single scale, where the extracted convolutional features discard many helpful cues from other convolution layers, and the single-scale feature integration may bring out the insufficient integration results. In stark contrast, our model integrates the multi-layer convolutional features under multiple scales for detecting co-salient objects. Further, the low-level features and high-level features are also used for the enhancement of the synthesized feature maps.

Overall, the main contributions of our co-saliency detection model are summarized in the following four aspects:

(1) We propose a convolutional neural network based co-saliency detection model, which consists of two key parts including the integration of multi-layer convolutional features from a group of images and the inter-image saliency propagation based refinement.
(2) To take full advantage of the convolutional features, we extract features of a group of images from multiple convolution layers, and these features go through four stages of feature integration to obtain the initial co-saliency map.
(3) The inter-image saliency propagation method in this paper is adapted from our previous work [4] to handle the co-saliency detection task, and obtain the refined final co-saliency map.
(4) We performed extensive experiments on two public benchmark datasets. The results demonstrate the effectiveness and superiority of our model when compared with the state-of-the-art co-saliency detection models.

The rest of this paper is organized as follows. The proposed co-saliency model is detailed in Section 2. Experimental results and analysis are shown in Section 3, and Section 4 presents the conclusion.

2. Proposed co-saliency model

An overview of the proposed co-saliency model is illustrated in Fig. 1. There are three steps including feature extraction, feature integration and inter-image saliency propagation. The feature integration consists of four main integration stages, as shown in