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/© 2019 Elsevier B.V. All rights reserved.

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

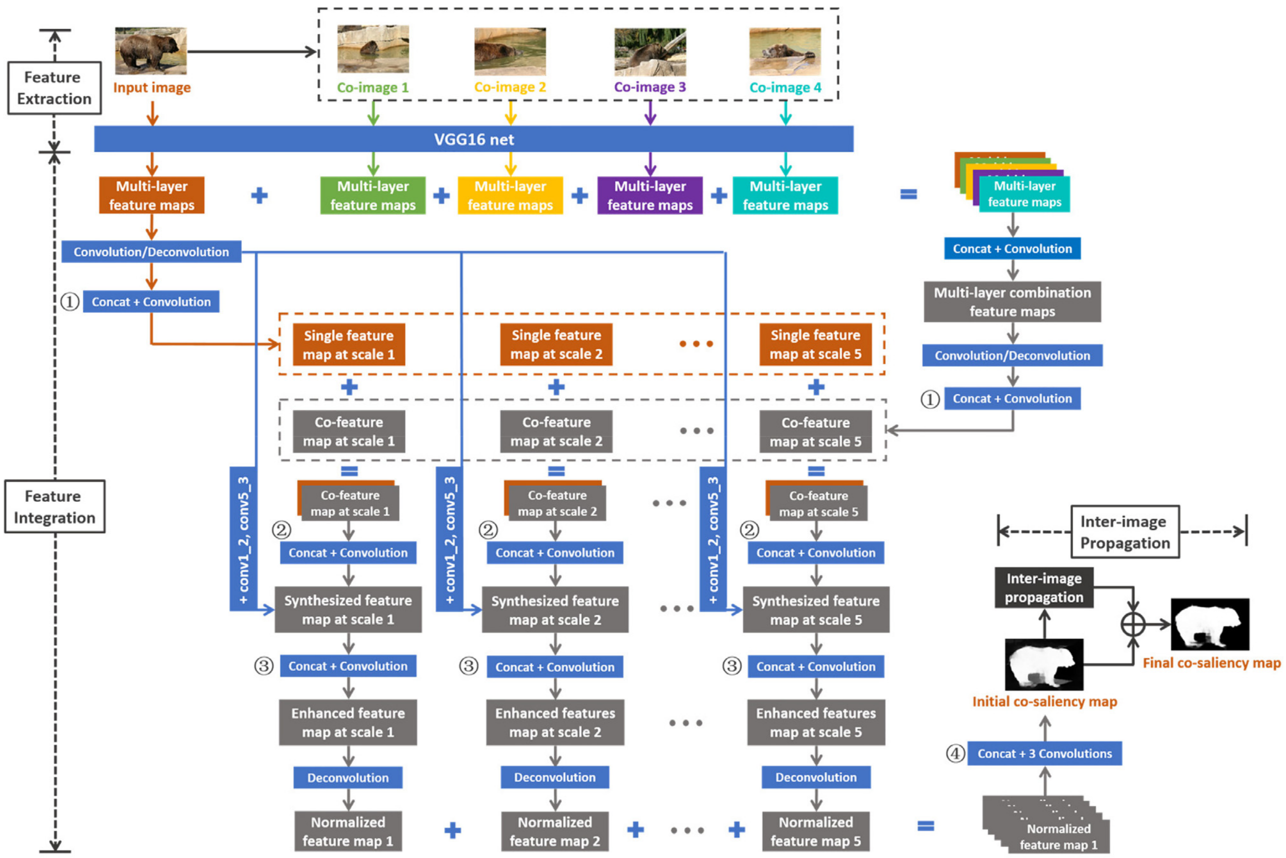


Fig. 1. Flowchart of the proposed co-saliency detection model. These colorful boxes represent feature maps or deep network layers: the orange boxes represent the feature maps extracted from the input image; the green, yellow, purple and cyan boxes represent the feature maps extracted from the four co-images, respectively; the gray boxes represent feature maps generated during the feature integration process; the blue boxes represent the deep network layers; the other operations are represents using black boxes. The colorful arrows between boxes indicate the information stream. The three steps of the proposed model, namely feature extraction, feature integration and inter-image saliency propagation, are marked using the black dotted lines. The four stages of the feature integration process are pointed by the circled numbers.

Fig. 1. The following subsections are arranged as follows: Sect. 2.1 gives a brief description of feature extraction; Sect. 2.2 elaborates the four stages of feature integration, which generates the initial co-saliency map; Sect. 2.3 presents the inter-image saliency propagation method, which leads to the improved quality of final co-saliency map.

2.1. Feature extraction

Given the input image $\mathbf{I}$, four co-images $\{\mathbf{CI}_1, \mathbf{CI}_2, \mathbf{CI}_3, \mathbf{CI}_4\}$ are first selected randomly from the image category for co-saliency detection, to which the input image $\mathbf{I}$ also belongs. The input image and the four co-images $\{\mathbf{I}, \mathbf{CI}_1, \mathbf{CI}_2, \mathbf{CI}_3, \mathbf{CI}_4\}$ are combined into a co-saliency detection group (CDG) for the input image $\mathbf{I}$. In the following step of feature integration, all images in CDG are jointly exploited to generate the initial co-saliency map corresponding to the input image $\mathbf{I}$. Then, all images in CDG are resized to a fixed size of $W \times H$, which is set to 256×256 in our work. Finally, to extract representative and effective features, we feed all images in CDG, $\{\mathbf{I}, \mathbf{CI}_1, \mathbf{CI}_2, \mathbf{CI}_3, \mathbf{CI}_4\}$, into five VGG16 models with shared weights. VGG16 model [40] is a well-known deep network for image classification and is widely used for feature extraction in many computer vision tasks. The original VGG16 model has 13 convolutional layers, 5 max-pooling layers and 3 fully connected layers. Due that the five max-pooling layers are set with a kernel size of two and a stride of two, the size of output feature maps is successively reduced layer by layer with a factor of two. For the pixel-wise co-saliency detection task, we remove the last max-pooling layer and the three fully connected layers, and extract features from the following five convolution layers of the VGG16 model: conv1_2 (the output feature map has the same spatial resolution as the input image namely $W \times H$, but with 64 channels), conv2_2 ($W/2 \times H/2$, 128 channels), conv3_3 ($W/4 \times H/4$, 256 channels), conv4_3 ($W/8 \times H/8$, 512 channels) and conv5_3 ($W/16 \times H/16$, 512 channels). The output feature maps at the five layers for all the five images in CDG are denoted as $\{\mathbf{F}^{ori}_{ij}, i, j = 1, 2, 3, 4, 5\}$, where subscript i denotes the i^{th} image in CDG ($i = 1$ indicates the input image, while $i = 2,3,4,5$ indicates the four co-images) and the subscript j denotes the feature map output by the j^{th} layer mentioned above. These feature maps will be used as the input to the next procedure, *i.e.* feature integration.

2.2. Feature integration

The procedure of feature integration can be divided into four stages, as shown in Fig. 1. In the first stage, multi-layer convolutional features are integrated into multi-scale single feature maps and co-feature maps. In the second stage, at each scale, we combine the single feature map with the co-feature map via concatenation and convolution, and obtain multi-scale synthesized feature maps. In the third stage, we employ low-level boundary features and high-level semantic features to enhance the synthesized feature maps. In the last stage, the resulting multi-scale enhanced feature maps are normalized to the unified scale and fused together to generate the initial co-saliency map.

2.2.1. Multi-scale single feature maps and co-feature maps

To find correlative information among all images in CDG, we first apply the concatenation layer to cascade all convolutional features originated from the same convolution layer in VGG16 model, and a convolution layer follows for further combination, yielding the multi-layer combination feature maps, $\{\mathbf{F}_j^{com}, j = 1, 2, 3, 4, 5\}$. Next, for making a better use of these combination feature maps and catering to the uneven sizes of different co-salient objects, the combination feature map of each layer, $\mathbf{F}_j^{com}$, passes through a convolution layer or deconvolution layer to obtain its multi-scale feature maps, $\{\mathbf{F}_{jl}^{com}, l = 1, 2, 3, 4, 5\}$, where l denotes the l^{th} scale. In our implementation, the five scales are set to 16×16, 32×32, 64×64, 128×128 and 256×256, respectively. Then, using the concatenation layer and convolution layer, the above multi-scale feature maps with the same scale but from the five different layers are integrated into the multi-scale co-feature maps, $\{\mathbf{F}_l^{co}, l = 1, 2, 3, 4, 5\}$. Similarly, the same operations are also performed on the convolutional features of the input image $\mathbf{I}$, and finally, we obtain the multi-scale single feature maps, $\{\mathbf{F}_l^{single}, l = 1, 2, 3, 4, 5\}$, which contain plenty of internal information of the input image $\mathbf{I}$.

2.2.2. Multi-scale synthesized feature maps

For the co-saliency detection task, we not only need internal information of the input image for detecting the salient objects, but also require correlative information of the other images in CDG for detecting the common objects. Therefore, at each scale, the feature integration is then performed between the co-feature map $\mathbf{F}_l^{co}$ and the single feature map $\mathbf{F}_l^{single}$ via a concatenation layer and a convolution layer. The outputs are called multi-scale synthesized feature maps, $\{\mathbf{F}_l^{syn}, l = 1, 2, 3, 4, 5\}$, which contain the internal information of the input image and the correlative information of all images in CDG. Accordingly, the synthesized feature maps are suitable to basically serve as the features for detecting co-salient objects.

2.2.3. Multi-scale enhanced feature maps

It is acknowledged that the lower convolutional features in conv1_2 go through fewer convolution layers and max-pooling layers, so that they preserve more boundary information [42]. The higher convolutional features in conv5_3 go through more convolution layers and max-pooling layers, so that they contain the deeper semantic information. To further improve the co-saliency detection accuracy, we add the convolutional features from conv1_2 and conv5_3 by introducing short connection to the multi-scale synthesized feature maps, $\{\mathbf{F}_l^{syn}, l = 1, 2, 3, 4, 5\}$. In detail, at each scale, a concatenation layer is first applied to cascade the convolutional features from conv1_2, conv5_3 and the synthesized feature maps, and then a convolutional layer follows, generating the multi-scale enhanced feature maps, $\{\mathbf{F}_l^{en}, l = 1, 2, 3, 4, 5\}$. We expect that the enhanced feature maps help to obtain clear boundaries and accurate co-saliency detection. And in fact, according to the ablation study results of Fig. 6 and Table 3 in Section 3.3, the effectiveness of the enhanced feature maps is verified.

2.2.4. Initial co-saliency map

To normalize the enhanced feature maps into the same size as the input image $\mathbf{I}$, the enhanced feature maps at the five scales go through the deconvolution layers with different convolution kernel sizes, to generate the normalized feature maps, $\{\mathbf{NF}_l, l = 1, 2, 3, 4, 5\}$. The integration of the normalized feature maps is slightly different from the integration operations mentioned above. Concretely, after the routine concatenation layer, we apply three convolution layers for a deeper integration, and the resulting feature map is termed as the initial co-saliency map $\mathbf{F}_{initial}$.

The feature integration mechanism in our model considers many factors including multi-layer feature extraction, multi-scale feature integration, integration of single feature maps and co-feature maps, enhancement of low-level features and high-level features. These factors are all devoted to detecting co-salient objects effectively. Concretely, multi-layer feature extraction is first used to provide features originated from convolutional layers of different phases. Then, the multi-scale feature integration is performed on the extracted features under five scales, yielding the initial co-saliency maps. Next, it should be noted that the key point of co-saliency detection is how to extract the correlative features among all images. Here, we employ the concatenation layers and convolution layers to combine all convolutional features originated from all images. The resulting co-feature maps contain correlative information, and are integrated with single feature maps, which contain the internal information of input image. Lastly, the low-level boundary features and high-level semantic features are widely used in many deep learning based saliency models. Here, the low-level features and high-level features are all integrated with the synthesized feature maps to generate the enhanced feature maps.

In our work, the networks of feature extraction and feature integration are jointly trained in an end-to-end manner. Given the co-saliency detection training dataset $\mathbf{T} = \{(\mathbf{X}_n, \mathbf{Y}_n)\}_{n=1}^N$ with N training samples, where $\mathbf{X}_n = \{x_n^t, t = 1, 2, ..., T\}$ and $\mathbf{Y}_n = \{y_n^t, t = 1, 2, ..., T\}$ are the input image and its binary ground truth with T pixels. Specifically, $y_n^t = 1$ denotes the co-salient object pixel and $y_n^t = 0$ denotes the background pixel. For simplicity, we drop the subscript n and consider each image independently. Thus, for generating the initial co-saliency map, the loss function is defined as:

$$L(\mathbf{W}, b) = -\sum_t y^t \log P\left(y^t = 1|\mathbf{X}; \mathbf{W}, b\right) + \left(1 - y^t\right) \log P\left(y^t = 0|\mathbf{X}; \mathbf{W}, b\right), \tag{1}$$

where $\mathbf{W}$ and b are the kernel weights and bias, respectively. We use the softmax classifier to compute the normalized probabilities $P(y^t = 1|\mathbf{X}; \mathbf{W}, b) \in [0, 1]$ and$P(y^t = 0|\mathbf{X}; \mathbf{W}, b) \in [0, 1]$. The first term $P(y^t = 1|\mathbf{X}; \mathbf{W}, b)$ indicates the probability of the pixel belonging to co-salient objects according to the initial co-saliency map, and the latter term $P(y^t = 0|\mathbf{X}; \mathbf{W}, b) \in [0, 1]$ indicates the probability of the pixel belonging to background according to the initial co-saliency map.

2.3. Inter-image saliency propagation

The initial co-saliency map is generated by the deep networks of feature extraction and feature integration with four stages. To further improve the quality of initial co-saliency map, we adapt an inter-image saliency propagation method in our previous work [4] to the task of co-saliency detection, for a full utilization of low-level features. Concretely, after we obtain the initial co-saliency maps of all images in the same image category, the inter-image saliency propagation based on manifold ranking [35] is employed to propagate the saliency values between each pair of images in the same image category. Following this way, we can obtain the final co-saliency maps with the improved quality.

Firstly, for each input image $\mathbf{I}$, the four most similar images, $\{\mathbf{SI}_r, r = 1, 2, 3, 4\}$, are retrieved from the same image category based on the image retrieval method in [43], in which the similarity $Sim(\mathbf{I}, \mathbf{SI}_r)$ is computed by using the chi-square distance of Gist descriptor and the weighted color histogram. The input image and its corresponding similar images are segmented to superpixels of different numbers by using the SLIC algorithm [44]. We denote the number of superpixels at eight segmentation scales as

$M_k (k = 1, 2, ..., 8)$, which is set to 150, 200, 250, 300, 350, 400, 450 and 500, respectively.

Next, we construct a close-loop inter-image graph $G(V, E)$ for the input image and one of its similar image, where V represents a group of nodes, namely superpixels, and E denotes a group of undirected edges connecting the nodes. The edge connection conforms one of the following three conditions: (1) each node is connected with its nearby nodes; (2) all nodes at image borders are connected together; (3) the potential object nodes in the two images are connected mutually, where the potential object nodes are determined by the corresponding initial co-saliency maps $\mathbf{F}_{initial}$. Specifically, the mean saliency value of a superpixel is calculated by averaging the initial co-saliency values of all pixels belonging to the superpixel. The potential object seed refers to such a superpixel, of which the mean saliency value is greater than the mean saliency value of the entire initial co-saliency map $\mathbf{F}_{initial}$.

Then, at the k^{th} scale, the weights of edges are defined by an affinity matrix,$\mathbf{A} = [a_{pq}]_{2M_k \times 2M_k}$, and $\mathbf{D} = diag\{d_{11}, \ldots, d_{2M_k 2M_k}\}$ denotes the degree matrix, where $d_{pp} = \sum_{q=1}^{2M_k} a_{pq}$. If the p^{th} node is connected with the q^{th} node, their affinity is defined as follows:

$$a_{pq} = \exp\left(-\frac{||c_p - c_q||}{\lambda^2}\right), \tag{2}$$

where c_p (resp. c_q) denotes the average color of pixels covered by the superpixel of the p^{th} (resp. q^{th}) node in the *Lab* color space. λ^2 is used to control the strength of weight between a pair of nodes, and is set to 0.1 in our work. If there is no connection between the p^{th} node and the q^{th} node, a_{pq} is set to 0.

Finally, by using the similar image $\mathbf{SI}_r$, the saliency propagation map at the k^{th} scale is defined as follows:

$$\mathbf{F}^{r,k}_{prop} = (\mathbf{D} - \alpha\mathbf{A})^{-1}\mathbf{v}, \tag{3}$$

where $\mathbf{v} = [v_1, v_2, ..., v_{2M_k}]^T$ denotes the indication vector, and we use the nodes belonging to the potential object nodes in the similar image $\mathbf{SI}_r$ as the labeled nodes. The parameter α, which is in the range of (0,1) and specifies the relative contribution from the indication vector to the saliency propagation map, is set to 0.99 for the larger relative contribution. Therefore, the saliency propagation map from the similar images is defined as follows:

$$\mathbf{F}_{prop} = \frac{1}{4 \times 8} \sum_{r=1}^{4} \sum_{k=1}^{8} \left[Sim(\mathbf{I}, \mathbf{SI}_r) \cdot \mathbf{F}^{r,k}_{prop} \right]. \tag{4}$$

And the final co-saliency map is formulated as:

$$\mathbf{F}_{final} = Norm\,[\mathbf{F}_{initial} + \mathbf{F}_{prop}], \tag{5}$$

where the pixel-wise addition operation is exploited to preserve salient object regions highlighted in both initial co-saliency map and saliency propagation map, and also suppress background regions. The *Norm* operation normalizes the final co-saliency map into the range of [0, 1].

3. Experimental results

3.1. Experimental setting

3.1.1. Dataset

To train our network, we adopt the Cosal2015 dataset [16] (which contains 50 image categories and has a total of 2015 images), the PASCAL-VOC dataset [45] (20 image categories, 1037 images) and the Coseg-Rep dataset [46] (23 image categories, 572 images), and all images are manually annotated with pixel-wise binary ground truths. For each input image of the three training datasets, we randomly select four images from the same image category of the input image as the co-images. Obviously, one input image can have many groups of co-images. Besides, to augment the training samples, we select all groups of co-images where any two groups have at least two different images. Accordingly, the original three training datasets are extended to 81,103 groups, in which each of them consists of one input image and its four co-images.

To verify the effectiveness of the proposed co-saliency detection model, we choose two public co-saliency detection datasets including iCoseg dataset [47] (38 image categories, 643 images) and MSRC dataset [48] (7 image categories, 233 images) as the test datasets.

3.1.2. Implementation details

We implemented our model based on MATLAB R2014a platform with the Caffe toolbox [49]. We run our co-saliency model in a PC with an i7-4790 K CPU (32GB memory) and a NVIDIA Titan XP GPU (12GB memory). Inspired by the Siamese network [50], our feature extraction network is constructed based on five VGG16 models with shared weights, and is initialized by the pre-trained model of VGG16. In the training phase, we use the standard stochastic gradient descent (SGD) method with batch size 8, momentum 0.9 and weight decay 0.005. The learning rate is set to $1e-8$ and decreased by 10% after 50,000 iterations. We need about 250k training iterations for convergence, which takes nearly 9 days. In the test phase, the generation of initial co-saliency map takes 0.21 s and the process of inter-image propagation takes 40 s for an image with a resolution of 256×256.

3.1.3. Evaluation metrics

We evaluate the co-saliency detection performance using three metrics including precision-recall (PR) curve, F-measure and mean absolute error (MAE). The PR curve depicts the precision value and recall value of co-saliency maps at each integer threshold from 0 to 255. The F-measure is an overall metric to measure co-saliency detection performance. To calculate F-measure, the binary object mask is first obtained by adaptively thresholding each co-saliency map using [51], the precision and recall are then calculated by comparing the binary object mask with the ground truth, and the F-measure is finally calculated as follows:

$$F_\beta = \frac{\left(1 + \beta^2\right) \times Precision \times Recall}{\beta^2 \times Precision + Recall}, \tag{6}$$

where β^2 is a balance factor and set to 0.3 to indicate more importance of precision than recall as suggested in [52]. The MAE computes the difference at pixel level between the co-saliency map S and the ground truth G, and is defined as

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - G(x, y)|, \tag{7}$$

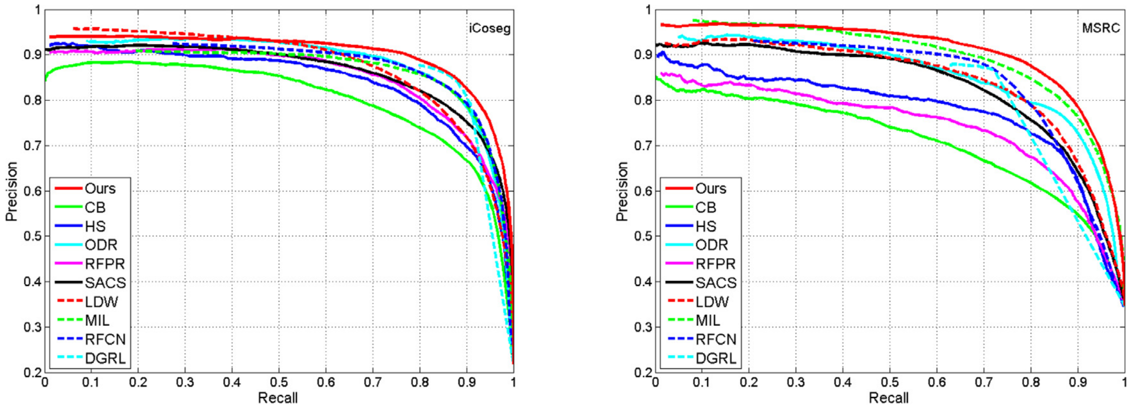
where W and H are the width and height, respectively, of the co-saliency map S.

3.2. Comparison with the state-of-the-arts

We compare our co-saliency detection model with the other seven state-of-the-art co-saliency detection models including CB [12], HS [13], ODR [15], RFPR [53], SACS [14], LDW [16], MIL [20] and two deep learning based single-image saliency detection models including RFCN [38] and DGRL [39]. For a fair comparison, we use either the implementation codes or the saliency maps provided by the authors.

3.2.1. Quantitative comparison

The comparisons of PR curves with the seven co-saliency models and two single-image saliency models on the iCoseg dataset



(a) iCoseg

(b) MSRC

Fig. 2. Precision-recall (PR) curves of seven co-saliency models and two single-image saliency models on two public datasets.

Table 1
F-measure and MAE values of seven co-saliency models and two single-image saliency models on two public datasets. The best results are shown in bold.

Dataset	Metric	Co-saliency models								Single-image saliency models	
		Ours	CB	HS	ODR	RFPR	SACS	LDW	MIL	RFCN	DGRL
iCoseg	F-measure	**.846**	.695	.702	.800	.777	.784	.605	.616	.821	.837
	MAE	**.101**	.167	.181	.107	.165	.224	.178	.159	.103	.102
MSRC	F-measure	**.851**	.573	.726	.780	.702	.769	.721	.753	.770	.759
	MAE	**.164**	.317	.281	.191	.292	.263	.257	.212	.184	.166

and the MSRC dataset are shown in Fig. 2. All the results of F-measure and MAE values for the two public datasets are listed in Table 1. According to Fig. 2 and Table 1, we can see that our co-saliency detection model consistently achieves the best performance on both datasets in terms of PR curve, F-measure and MAE. This demonstrates the effectiveness and superiority of our model compared to all the other saliency detection models.

3.2.2. *Qualitative comparison*

Figs. 3 and 4 show a qualitative comparison for the initial co-saliency maps and the final co-saliency maps generated by our model, and the saliency maps generated by using seven co-saliency models and two single-image saliency models on both datasets. Generally speaking, our model highlights co-salient objects more consistently and suppresses background more effectively when compared to all the other saliency detection models. The final co-saliency maps generated using our model show the best visual quality compared to other saliency maps. For example, in the top example of Fig. 3, the co-salient objects (*i.e.* the red soccer players) are highlighted more uniformly in our final co-saliency maps compared to other saliency maps including the initial co-saliency maps generated by our model and the saliency maps generated by other saliency models.

Further, the irrelevant regions of white soccer players and playfields are also suppressed more completely in our final co-saliency maps. In Fig. 4, compared with saliency maps generated by other saliency models and our initial co-saliency maps, our final co-saliency maps perform best. We can see that our final co-saliency maps not only highlight the salient objects uniformly, but also are with well-defined boundaries, such as the legs of cattle in the top example, and the cars in the bottom example. Furthermore, we also give a quantitative comparison for our initial co-saliency maps and our final co-saliency maps in the ablation experiments of Section 3.3.

3.3. *Ablation study*

To analyze the contributions of different components in our model, we perform the following two ablation studies on the two datasets.

Firstly, we design different variants of the proposed co-saliency detection model with different settings: (1) to validate the effectiveness of multi-scale operation, “Without-multiscale” indicates that the procedure of feature integration is executed on one scale, and we choose all possible scales, including 16 × 16, 32 × 32, 64 × 64, 128 × 128 and 256 × 256, for a comparison; (2) in order to confirm the enhancement of low-level boundary features and high-level semantic features, “Without-enhancement” discards the third stage of feature integration; (3) to verify the improvement of saliency propagation based refinement, we remove the inter-image saliency propagation, to obtain the variant “Without-propagation”, namely the initial co-saliency maps. In Fig. 5 and Table 2, the results of seven variants (note that “Without-multiscale” contains five variants with different scales) are compared with our model in terms of PR curve, F-measure and MAE value. The comparisons on both datasets show the contributions of different components, and validate the rationality of the design of our model.

Secondly, on the basis of the above variant “Without-enhancement”, we design five more detailed variants to verify the effect of convolutional features from different convolution layers in Fig. 6 and Table 3: (1) “layer1” means that we only use the convolutional features from conv1_2 to enhance the synthesized feature maps in the third stage of feature integration; (2) similarly, “layer5” means that the used convolutional features come from conv5_3; (3) “layer125” means that the used convolutional features come from conv1_2, conv2_2 and conv5_3; (4) “layer1235” means conv1_2, conv2_2, conv3_3 and conv5_3; (5) “layer12345” means all five layers, namely conv1_2, conv2_2, conv3_3, conv4_3 and conv5_3. The above variant “Without-enhancement” is also shown in Fig. 6 and Table 3 for comparison. Obviously, our model,

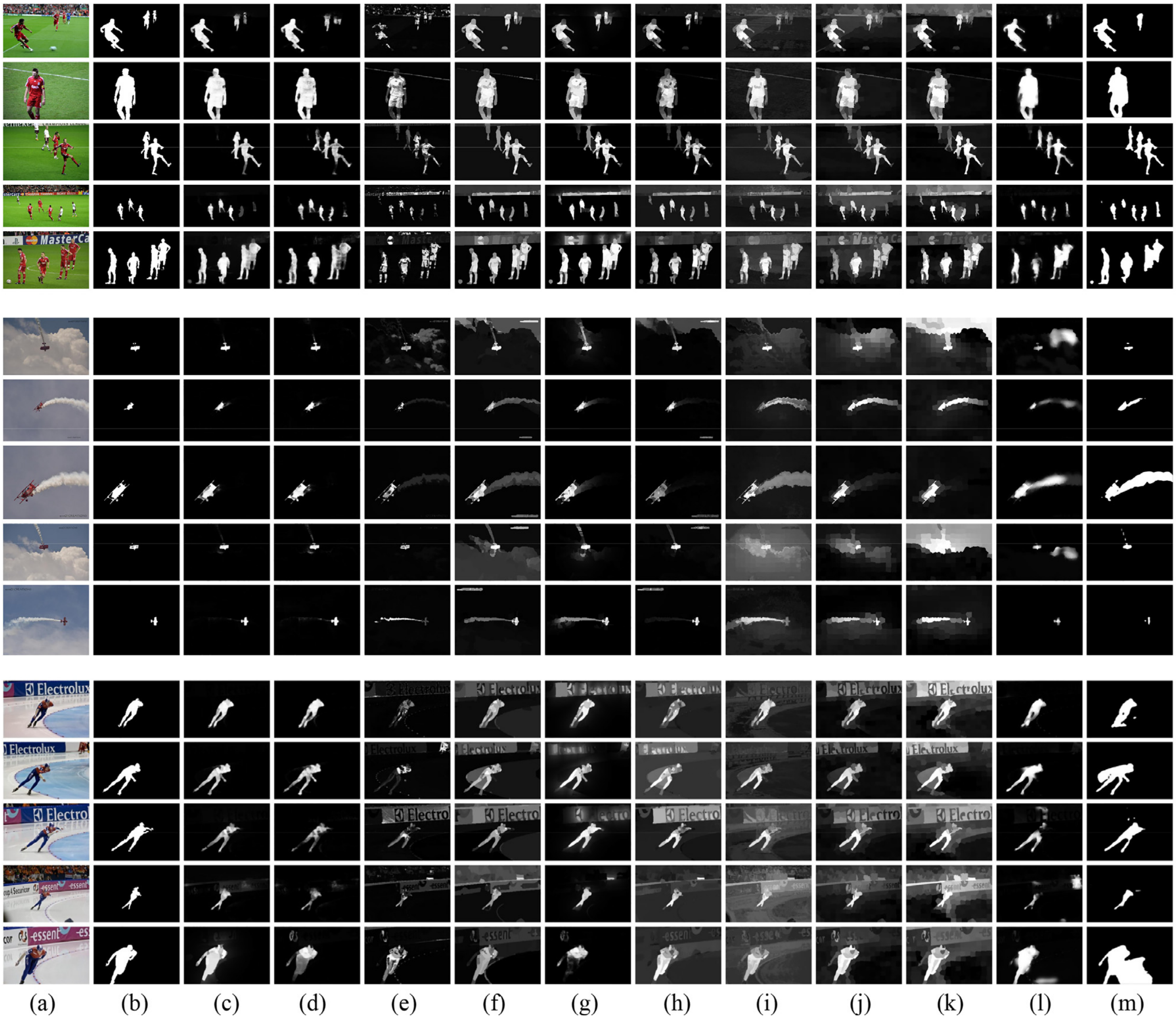


Fig. 3. Visual comparison of saliency maps on the iCoseg dataset. From top to bottom, every five images come from the same image category. (a) Images; (b) ground truths; (c) final co-saliency maps and (d) initial co-saliency maps generated by our model; saliency maps generated using (e) CB, (f) HS, (g) ODR, (h) RFPR, (i) SACS, (j) LDW, (k) MIL, (l) RFCN and (m) DGRL.

Table 2
F-measure and MAE values of different variants of our model on two datasets. The best results are shown in bold.

Dataset	Metric	Ours	Without-multiscale					Without-enhancement	Without-propagation
			16 × 16	32 × 32	64 × 64	128 × 128	256 × 256		
iCoseg	F-measure	**.846**	.802	0.798	0.791	0.797	0.792	.818	.815
	MAE	**.101**	.123	0.124	0.127	0.125	0.125	.120	.102
MSRC	F-measure	**.851**	.833	0.827	0.821	0.819	0.815	.825	.823
	MAE	**.164**	.178	0.178	0.186	0.187	0.179	.173	.165

Table 3
F-measure and MAE values of different variants of our model on two datasets. The best results are shown in bold.

Dataset	Metric	Ours	Without-enhancement	Layer1	Layer5	Layer125	Layer1235	Layer12345
iCoseg	F-measure	**.846**	.818	.820	.823	.830	.829	.833
	MAE	**.101**	.120	.113	.115	.109	.111	.106
MSRC	F-measure	**.851**	.825	.829	.830	.825	.836	.838
	MAE	**.164**	.173	.176	.174	.173	.173	.170

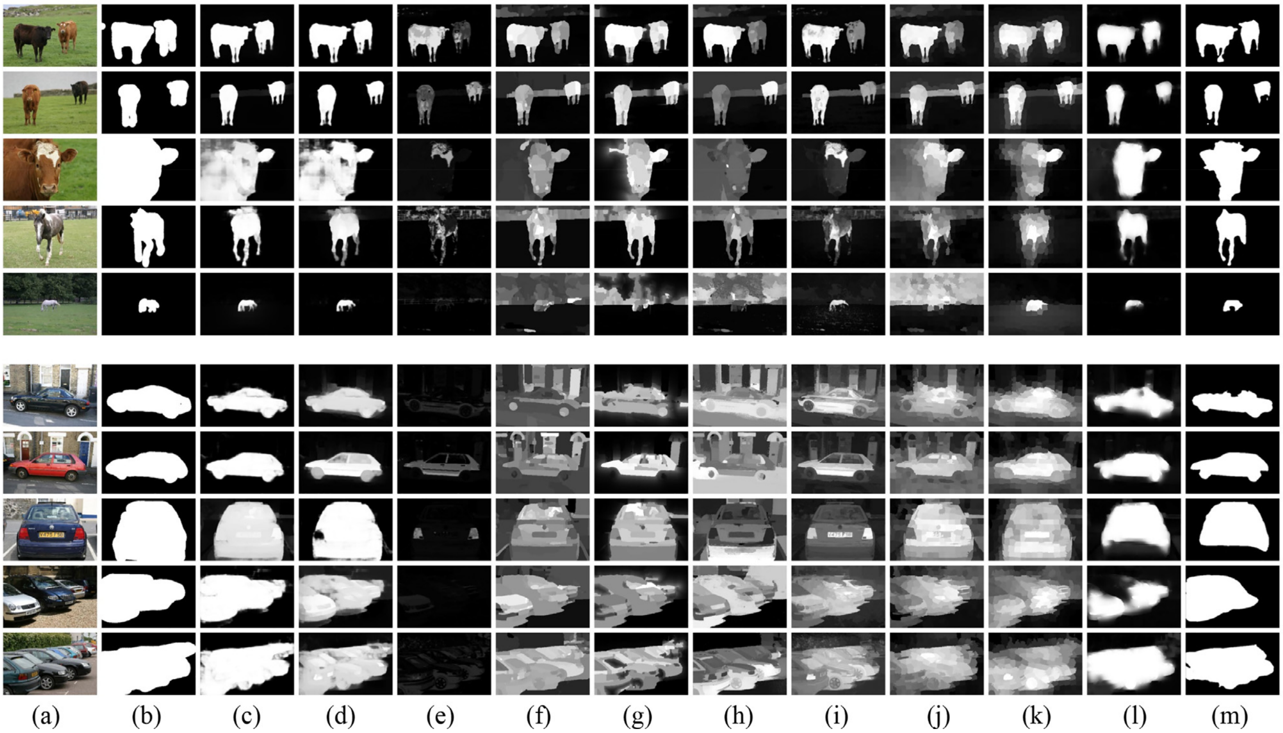
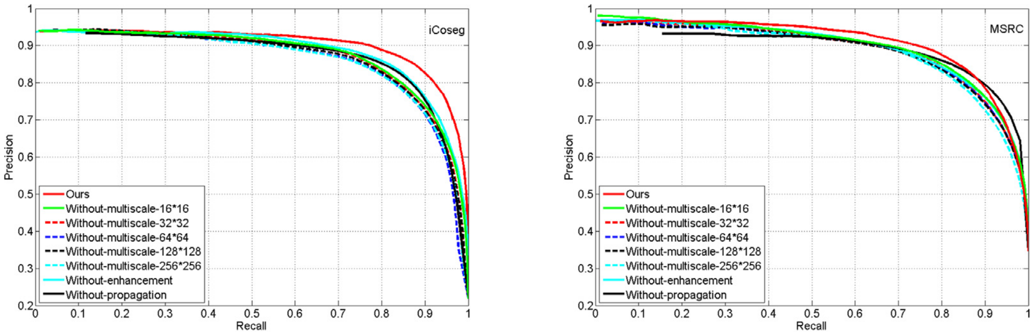


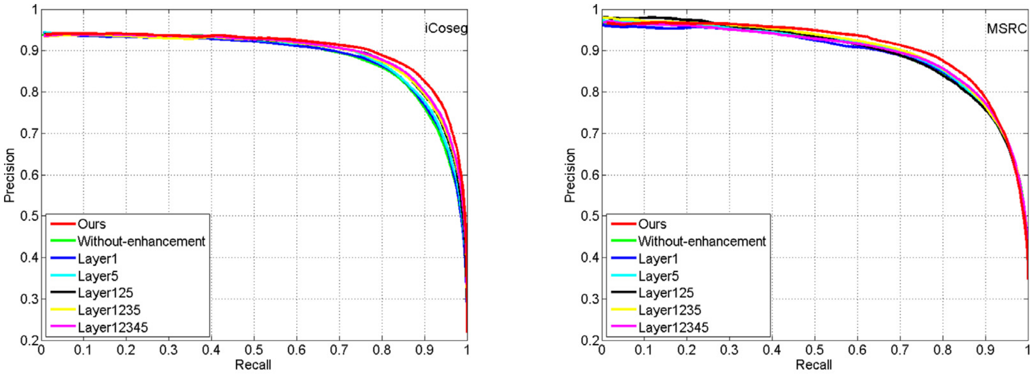
Fig. 4. Visual comparison of saliency maps on the MSRC dataset. From top to bottom, every five images come from the same image category. (a) Images; (b) ground truths; (c) final co-saliency maps and (d) initial co-saliency maps generated by our model; saliency maps generated using (e) CB, (f) HS, (g) ODR, (h) RFPR, (i) SACS, (j) LDW, (k) MIL, (l) RFCN and (m) DGRL.



(a) iCoseg

(b) MSRC

Fig. 5. Precision-recall (PR) curves of different variants of our model on two datasets.



(a) iCoseg

(b) MSRC

Fig. 6. Precision-recall (PR) curves of different variants of our model on two datasets.

which uses the convolutional features from conv1_2 and conv5_3, achieves the best performance in terms of PR curve, F-measure and MAE value. The variant "Without-enhancement" achieves the worst performance because its synthesized feature maps are not enhanced by any convolutional features, while the variants "layer1" and "layer5" improve the performance to some extent. The comparison shows that convolutional features from the two layers are helpful to enhance the feature maps. The other variants "layer125", "layer1235" and "layer12345" are worse than our model, and such a performance degradation indicates that excessive features may bring in redundant information.

4. Conclusion

In this paper, we propose a deep convolutional neural network based co-saliency detection model, which realizes the effective integration of multi-layer convolutional features from a group of images and inter-image saliency propagation for co-saliency detection task. Specifically, the multi-layer convolutional features of the input image and its four co-images are extracted and integrated into multi-scale synthesized feature maps containing the internal information and correlative information, which are indispensable for detecting co-salient objects. Then, with the enhancement of the low-level boundary features and high-level semantic features, the feature maps are fused together to obtain the initial co-saliency map. Finally, the inter-image saliency propagation method is utilized to propagate saliency values between images, yielding the final co-saliency map with the improved quality. Experimental results on two public datasets demonstrate the effectiveness of the proposed co-saliency detection model to boost co-saliency detection performance.

There are two directions that will be taken into account in our future work. The first one is how to modify the co-saliency model to simultaneously compute the co-saliency maps of all the five input images. Intuitively, we can extend the current co-saliency model via repeating the part of feature integration five times for each input image. Unfortunately, the network parameters will increase by nearly four times, resulting in the time-consuming training progress. Therefore, in our future work, we will try to design the co-saliency model that can output all co-saliency maps simultaneously without increasing network parameters. The second one is how to transform the current co-saliency model to deal with other related tasks, such as RGB-D co-saliency detection and video co-segmentation. Comparing to color images, RGB-D images have additional depth information and videos have temporal information. We will need to transform the current network structure to sufficiently utilize the extra information.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 61771301, 61502424, 61901145 and 61976192, Shanghai Municipal Natural Science Foundation under Grant No. 16ZR1411100, and Zhejiang Provincial Natural Science Foundation of China under Grant No. LY18F020032.

References

[1] Z. Liu, W. Zou, O. Le Meur, Saliency tree: a novel saliency detection framework, IEEE Trans. Image Process. 23 (5) (2014) 1937–1952.
[2] X. Zhou, Z. Liu, G. Sun, L. Ye, X. Wang, Improving saliency detection via multiple kernel boosting and adaptive fusion, IEEE Signal Process. Lett. 23 (4) (2016) 517–521.
[3] X. Zhou, Z. Liu, G. Sun, X. Wang, Adaptive saliency fusion based on quality assessment, Multimedia Tools Appl. 76 (22) (2017) 23187–23211.
[4] J. Ren, Z. Liu, X. Zhou, G. Sun, C. Bai, Saliency integration driven by similar images, J. Vis. Commun. Image Represent 50 (2018) 227–236.
[5] M.-M. Cheng, N.J. Mitra, X. Huang, S.-M. Hu, Salientshape: group saliency in image collections, Vis. Comput. 30 (4) (2014) 443–453.
[6] J. Chen, C. Bai, L. Huang, Z. Liu, S. Chen, Visual saliency fusion based multi--feature for semantic image retrieval, in: Proceedings of Chinese Conference on Computer Vision (CCCV), Springer, 2017, pp. 126–136.
[7] C. Bai, J. Chen, L. Huang, K. Kpalma, S. Chen, Saliency-based multi-feature modeling for semantic image retrieval, J. Vis. Commun. Image Represent 50 (2018) 199–204.
[8] L. Ye, Z. Liu, L. Li, L. Shen, C. Bai, Y. Wang, Salient object segmentation via effective integration of saliency and objectness, IEEE Trans. Multimedia 19 (8) (2017) 1742–1756.
[9] W. Zou, Z. Liu, K. Kpalma, J. Ronsin, Y. Zhao, N. Komodakis, Unsupervised joint salient region detection and object segmentation, IEEE Trans. Image Process. 24 (11) (2015) 3858–3873.
[10] Q. Hou, M.-M. Cheng, J. Liu, P.-H.-S. Torr, Webseg: learning semantic segmentation from web searches, arXiv preprint arXiv:1803.09859, 2018.
[11] F. Zhang, B. Du, L. Zhang, Saliency-guided unsupervised feature learning for scene classification, IEEE Trans. Geosci. Remote Sens 53 (4) (2015) 2175–2184.
[12] H. Fu, X. Cao, Z. Tu, Cluster-based co-saliency detection, IEEE Trans. Image Process. 22 (10) (2013) 3766–3778.
[13] Z. Liu, W. Zou, L. Li, L. Shen, O. Le Meur, Co-saliency detection based on hierarchical segmentation, IEEE Signal Process. Lett. 20 (1) (2014) 88–92.
[14] X. Cao, Z. Tao, B. Zhang, H. Fu, W. Feng, Self-adaptively weighted co-saliency detection via rank constraint, IEEE Trans. Image Process. 23 (9) (2014) 4175–4186.
[15] L. Ye, Z. Liu, J. Li, W. Zhao, L. Shen, Co-saliency detection via co-salient object discovery and recovery, IEEE Signal Process. Lett. 22 (11) (2015) 2073–2077.
[16] D. Zhang, J. Han, C. Li, J. Wang, X. Li, Detection of co-salient objects by looking deep and wide, Int. J. Comput. Vision 120 (2) (2016) 215–232.
[17] J. Han, G. Cheng, Z. Li, D. Zhang, A unified metric learning-based framework for co-saliency detection, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2018) 2473–2483.
[18] 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 Trans. Image Process. 27 (12) (2018) 5866–5879.
[19] L. Wei, S. Zhao, O. Bourahla, X. Li, F. Wu, Group-wise deep co-saliency detection, in: Proceedings of International Joint Conference on Artificial Intelligent (IJCAI), 2017, pp. 3041–3047.
[20] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, IEEE Trans. Pattern Anal. Mach. Intell. 39 (5) (2017) 865–878.
[21] H. Song, Z. Liu, H. Du, G. Sun, O. Le Meur, T. Ren, Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning, IEEE Trans. Image Process. 26 (9) (2017) 4204–4216.
[22] J. Han, H. Chen, N. Liu, C. Yan, X. Li, CNNs-based rgb-d saliency detection via cross-view transfer and multiview fusion, IEEE Trans. Cybernetics 48 (11) (2018) 3171–3183.
[23] X. Zhou, Z. Liu, C. Gong, L. Wei, Improving video saliency detection via localized estimation and spatiotemporal refinement, IEEE Trans. Multimedia 20 (11) (2018) 2993–3007.
[24] X. Zhou, Z. Liu, K. Li, G. Sun, Video saliency detection via bagging-based prediction and spatiotemporal propagation, J. Vis. Commun. Image Represent 51 (2018) 131–143.
[25] T. Wu, Z. Liu, X. Zhou, K. Li, Spatiotemporal salient object detection by integrating with objectness, Multimedia Tools Appl. 77 (15) (2018) 19481–19498.
[26] M. Sun, Z. Zhou, Q. Hu, Z. Wang, J. Jiang, Sg-fcn: a motion and memory-based deep learning model for video saliency detection, IEEE Trans. Cybernetics 49 (8) (2019) 2900–2911, doi:10.1109/TCYB.2018.2832053.
[27] Z. Wang, J. Ren, D. Zhang, M. Sun, J. Jiang, A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos, Neurocomputing 287 (2018) 68–83.
[28] L. Li, Z. Liu, J. Zhang, Unsupervised image co-segmentation via guidance of simple images, Neurocomputing 275 (2018) 1650–1661.
[29] H. Fu, D. Xu, B. Zhang, S. Lin, R. Ward, Object-based multiple foreground video co-segmentation via multi-state selection graph, IEEE Trans. Image Process. 24 (11) (2015) 3415–3424.
[30] J. Han, R. Quan, D. Zhang, F. Nie, Robust object co-segmentation using background prior, IEEE Trans. Image Process. 27 (4) (2018) 1639–1651.
[31] 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 (ECCV), Springer, 2008, pp. 144–157.
[32] D. Zhang, H. Fu, J. Han, A. Borji, X. Li, A review of cosaliency detection algorithms: fundamentals, applications, and challenges, ACM Trans. Intell. Syst. and Technol. 9 (4) (2018) 1–31.
[33] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: a discriminative regional feature integration approach, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 2083–2090.

[34] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2014, pp. 2814–2821.

[35] L. Zhang, C. Yang, H. Lu, X. Ruan, M. Yang, Ranking saliency, IEEE Trans. Pattern Anal. Mach. Intell. 39 (9) (2017) 1892–1904.

[36] C. Gong, D. Tao, W. Liu, S.J. Maybank, M. Fang, K. Fu, J. Yang, Saliency propagation from simple to difficult, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015, pp. 2531–2539.

[37] L. Zhang, J. Ai, B. Jiang, H. Lu, X. Li, Saliency detection via absorbing markov chain with learnt transition probability, IEEE Trans. Image Process. 27 (2) (2018) 987–998.

[38] L. Wang, L. Wang, H. Lu, P. Zhang, X. Ruan, Salient object detection with recurrent fully convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2019) 1734–1746, doi:10.1109/TPAMI.2018.2846598.

[39] T. Wang, L. Zhang, S. Wang, H. Lu, G. Yang, X. Ruan, A. Borji, Detect globally, refine locally: a novel approach to saliency detection, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, pp. 4321–4329.

[40] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of International Conference on Learning Representations (ICLR), 2015, pp. 1–14.

[41] R. Zhao, W. Ouyang, H. Li, X. Wang, Saliency detection by multi-context deep learning, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015, pp. 1265–1274.

[42] M.-D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Proceedings of the European Conference on Computer Vision (ECCV), Springer, 2014, pp. 818–833.

[43] 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.

[44] 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.

[45] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, Int. J. Comput. Vision 88 (2) (2010) 303–338.

[46] J. Dai, Y. Wu, J. Zhou, S. Zhu, Cosegmentation and cosketch by unsupervised learning, in: Proceedings of the International Conference on Computer Vision (ICCV), IEEE, 2013, pp. 1305–1312.

[47] D. Batra, A. Kowdle, D. Parikh, J. Luo, T. Chen, iCoseg: interactive co-segmentation with intelligent scribble guidance, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 3169–3176.

[48] J. Winn, A. Criminisi, T. Minka, Object categorization by learned universal visual dictionary, in: Proceedings of the International Conference on Computer Vision (ICCV), IEEE, 2005, pp. 1800–1807.

[49] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceeding of the International Conference on Multimedia, ACM, 2014, pp. 675–678.

[50] S. Chopra, R. Hadsell, Y. LeCun, Learing a similarity metric discriminatively, with application to face verification, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2005, pp. 539–546.

[51] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.

[52] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722.

[53] L. Li, Z. Liu, W. Zou, X. Zhang, O. Le Meur, Co-saliency detection based on region-level fusion and pixel-level refinement, in: Proceedings of the International Conference on Multimedia and Expo (ICME), IEEE, 2014, pp. 1–6.

Jingru Ren received the B.E. degree from Shanghai University, Shanghai, China, in 2015. She is currently pursuing the Ph.D. degree at the School of Communication and Information Engineering, Shanghai University, Shanghai, China. Her research interests include saliency detection and salient object segmentation.

Zhi Liu received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, and the Ph.D. degree from Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 1999, 2002, and 2005, respectively. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. From Aug. 2012 to Aug. 2014, he was a Visiting Researcher with the SIROCCO Team, IRISA/INRIA-Rennes, France, with the support by EU FP7 Marie Curie Actions. He has published more than 180 refereed technical papers in international journals and conferences. His-research interests include image/video processing, machine learning, computer vision and multimedia communication. He was a TPC member/session chair in ICIP 2017, PCM 2016, VCIP 2016, ICME 2014, WIAMIS 2013, etc. He co-organized special sessions on visual attention, saliency models, and applications at WIAMIS 2013 and ICME 2014. He is an area editor of *Signal Processing: Image Communication* and served as a guest editor for the special issue on *Recent Advances in Saliency Models, Applications and Evaluations* in *Signal Processing: Image Communication*. He is a senior member of IEEE.

Xiaofei Zhou received the B.E. degree from Anhui Polytechnic University, Wuhu, China, in 2012, and the M.E. and Ph.D. degrees from Shanghai University, Shanghai, China, in 2015 and 2018, respectively. He is currently a Lecturer with the Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, China. His-research interests include saliency detection and image/video segmentation.

Cong Bai received the B.E. degree from Shandong University, Jinan, China, in 2003, the M.E. degree from Shanghai University, Shanghai, China, in 2009 and the Ph.D. degree from National Institute of Applied Science of Rennes (INSA de Rennes), Rennes, France in 2013. Since 2013, he has been with the Faculty of the College of Computer Science, Zhejiang University of Technology, Hangzhou, China. He was with the School of Information Science and Engineering, Shandong Agricultural University from 2003 to 2006. His-research interests include multimedia retrieval and computer vision.

Guangling Sun received the B.S. degree in electronic engineering from Northeast Forestry University, China, in 1996 and the M.E. and Ph.D. degrees in computer application technology from Harbin Institute of Technology, China, in 1998 and 2003, respectively. Since 2006, she has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where she is currently an Associate Professor. She was with the University of Maryland, College Park as a visiting scholar from December 2013 to December 2014. Her research interests include saliency detection, face recognition, and image/video processing.