Saliency Detection Via Similar Image Retrieval

Linwei Ye, Zhi Liu, *Senior Member, IEEE*, Xiaofei Zhou, Liquan Shen, and Jian Zhang, *Senior Member, IEEE*

***Abstract*—This letter proposes a novel saliency detection framework by propagating saliency of similar images retrieved from large and diverse Internet image collections to boost saliency detection performance effectively. For the input image, a group of similar images is retrieved based on the saliency weighted color histograms and the Gist descriptor from Internet image collections. Then, a pixel-level correspondence process between images is performed to guide the saliency propagation from the retrieved images. Both initial saliency map and correspondence saliency map are exploited to select the training samples by using the graph cut-based segmentation. Finally, the training samples are input into a set of weak classifiers to learn the boosted classifier for generating the boosted saliency map, which is integrated with the initial saliency map to generate the final saliency map. Experimental results on two public image datasets demonstrate that the proposed model can achieve the better saliency detection performance than the state-of-the-art single-image saliency models and co-saliency models.**

***Index Terms*—Boosting, image retrieval, saliency detection, saliency propagation.**

I. Introduction

HUMAN BEING is able to identify the visual object of interest in a complex scene quickly and effortlessly. Inspired by the mechanism of human visual attention, a lot of saliency models have been proposed in the last decades, and they play an important role in a number of multimedia applications including salient object segmentation [1], [2], image retrieval [3], image/video retargeting [4]–[6], and media editing [7]. Generally speaking, there are two classes of saliency models, i.e., for human fixation prediction and for salient object segmentation/detection. The former class aims to mimic human visual scanpath to predict eye fixations and a recent benchmark [8] presents a complete review and performance evaluation. In this letter, we focus on the latter class of saliency models, which aims to highlight the whole salient object with well-defined boundaries and has made significant progress in the recent years. In the following, we will introduce some representative saliency models for a single image and co-saliency models for a group of images.

By exploiting the global information over the image, kernel density estimation-based nonparametric model [1] and region contrast (RC) model [9] utilize the color distribution and spatial information of over-segmented regions to measure saliency. Hierarchical segmentation [10], [11] and superpixel segmentation [12] have been proved effective for improving the saliency detection performance. The machine learning methods have also been exploited to improve the saliency detection performance. A set of features are combined in the conditional random field (CRF) learning in [13] and a number of regional features are used to train a random forest regression model, which computes saliency scores across multiple levels to generate the saliency map in [14]. Multiple kernel boosting (MKB) [15] is introduced to measure saliency in [16] in which the training sets are obtained by thresholding the weak saliency map and used to learn a strong classifier for generating the strong saliency map.

Furthermore, co-saliency models are proposed to detect the co-salient objects in a group of images based on distinctiveness in each image and repetitiveness across the whole group [17]–[25]. Single-image saliency is generally used as a prior, and co-multilayer graph [17] as well as Markov random field (MRF) optimization [18] and rank-one constraint [19] are exploited to detect co-salient objects in image pairs as well as in a set of images. In [20], a cluster-based method using contrast cue, spatial cue, and corresponding cue is exploited to generate co-saliency maps for an image group. Multiscale region segmentation is beneficial for improving co-saliency detection performance by using region match [21], object prior with global similarity [22], and region-level fusion with pixel-level refinement [23]. In [24], the guided saliency maps are generated by queries of a single-image saliency map and then are integrated into the co-saliency map. In [25], a novel framework for co-saliency detection effectively combines co-salient object discovery and recovery. In [26], an unsupervised joint object discovery and segmentation method is exploited to cut out the common objects from a collection of Internet images. In [27], the common object contours are extracted by the random forest and then single saliency map and inter saliency map are fused on the small overlapped groups from a large image set to generate co-saliency maps.

Nowadays, Internet media is a blooming and worth exploring area with convenient and efficient access. In this letter, we aim to utilize the similar images retrieved from Internet image collections and propose a novel saliency detection framework. The main contributions of the proposed model lie in the following three aspects. First, we propose to retrieve a

Manuscript received January 28, 2016; revised April 05, 2016; accepted April 21, 2016. Date of publication April 27, 2016; date of current version May 05, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61171144 and Grant 61471230, and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhengguo Li. (*Corresponding author: Zhi Liu.*)

L. Ye, Z. Liu, X. Zhou, and L. Shen are with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (e-mail: yelinweimail@163.com; liuzhisjtu@163.com; zxforchid@outlook.com; jsslq@163.com).

J. Zhang is with the Global Data Technologies Centre, School of Computing and Communication, University of Technology, Sydney, N.S.W. 2007, Australia (e-mail: jian.zhang@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2558489

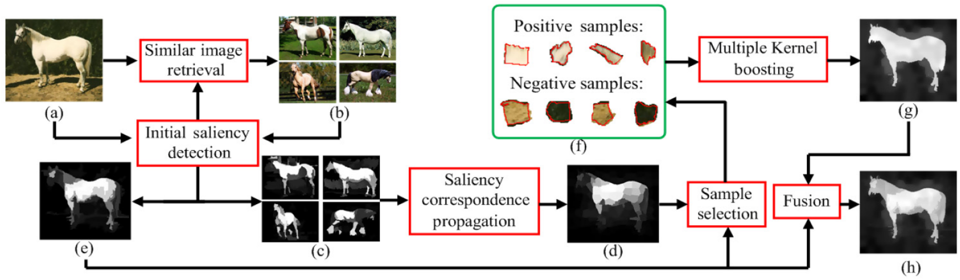

Fig. 1. Illustration of the proposed saliency model. (a) Input image. (b) Retrieved images. (c) and (e) Initial saliency maps. (d) Correspondence saliency map. (f) Training set. (g) Boosted saliency map. (h) Final saliency map.

group of similar images from Internet image collections to facilitate the saliency detection in the input image. Second, the proposed pixel correspondence-based saliency propagation method effectively transfers the original saliency values from the retrieved images to generate the correspondence saliency map. Finally, we exploit the graph cut-based object segmentation method to obtain the reliable training set for unsupervised MKB to generate the boosted saliency map. Experimental results show that the proposed model outperforms the state-of-the-art single-image saliency models and co-saliency models.

II. Proposed Saliency Model

A. Similar Image Retrieval

To effectively utilize a large set of candidate images from Internet collection, we first retrieve similar images for the input image Q to obtain a small group of images $\{R_m\}_{m=1}^{M}$ based on global feature descriptors. In our implementation, M is set to a smaller value, 4, to ensure the use of the most similar images, which are sufficient for improving the saliency detection performance with the proposed model.

For both Q and each $R_m(m = 1, \ldots, M)$, we use the RC-based saliency model [9], which can be replaced by any other single-image saliency model, to generate the initial pixel-level saliency map S^Q and $S_m^R(m = 1, \ldots, M)$, respectively. In the two color spaces, RGB and Lab, respectively, each color channel of Q and R_m is uniformly quantized into 16 bins for calculating the two color histograms over the whole image. The initial saliency values of pixels are used to weight the corresponding histogram bins, and the normalization operation is used to generate the saliency weighted color histograms within the range of [0,1]. The saliency weighted color histogram emphasizes the color appearances of salient objects rather than taking into account the whole image evenly, and thus results in a better retrieval for images with similar salient objects. The Gist descriptor [28], which consists of perceptual dimensions including naturalness, openness, roughness, expansion, and ruggedness, is extracted to represent the spatial structure of image scene.

Let $H_m^{C_1}$ and $H_m^{C_2}$ denote the saliency weighted color histograms of R_m in the RGB and Lab color space, respectively, and H_m^G denote the Gist descriptor vector. Similarly for the input image Q, let $H_Q^{C_1}$ and $H_Q^{C_2}$ denote the two saliency weighted color histograms, respectively, and H_Q^G denote the Gist descriptor. The similarity between Q and R_m is defined as

$$\mathrm{Sim}(Q, R_m) = 1 - \frac{1}{2}\left[\frac{\chi^2\left(H_Q^{C_1}, H_m^{C_1}\right) + \chi^2\left(H_Q^{C_2}, H_m^{C_2}\right)}{2} + \chi^2\left(H_Q^G, H_m^G\right)\right] \tag{1}$$

where $\chi^2(\cdot)$is the chi-square distance. Based on (1), we retrieve a total of M images with the highest similarity measures. As shown in Fig. 1(b), the retrieved images show similar object appearances with the input image in Fig. 1(a).

B. Saliency Correspondence Propagation

Unlike the conventional co-saliency detection in a group of images, the objects from internet image collections vary drastically such as color, size and shape, as well as various scenes. To discover the similarity of the common objects in such a situation, the dense pixel correspondence [29] between pixels in each pair of images is exploited. Let $t_m(i)$ denote the translation between the pixel i in Q and its correspondence in R_m, and Ω_C denote the pixel neighborhood on a spatial pyramid graph. So the correspondence energy function $E^C(t_m)$ is defined as follows:

$$E^C(t_m) = \sum_i \min\left(\|F^Q(i) - F_m^R[i + t_m(i)]\|_1, \lambda_c\right) + \alpha_c \sum_{i,j\in\Omega_c} \min\left(\|t_m(i) - t_m(j)\|_1, \beta_c\right) \tag{2}$$

where $F^Q(i)$ and $F_m^R[i + t_m(i)]$ denote the feature encoded by a universal dictionary extracted at the pixel i in Q and its correspondence in R_m, respectively. As suggested in [29], the two parameters, α_c and β_c, are set to 0.02 and 0.5, respectively, and λ_c is the average descriptor distance between the image pair with the truncated L1 norm. The loopy belief propagation [30] is exploited to efficiently optimize the correspondence energy function and obtain t_m to represent the translations of pixels between Q and R_m.

Based on the saliency maps of the retrieved images, $S_m^R(m = 1, \ldots, M)$, the pixel-level saliency correspondence propagation is performed to obtain the correspondence saliency map S^C for the input image Q as follows:

$$S^C(i) = \frac{\sum_{m=1}^{M}\{\omega_m(i) \cdot S_m^R[i + t_m(i)]\}}{\sum_{m=1}^{M}\omega_m(i)} \tag{3}$$

where the correspondence-based weight is defined as follows:

$$\omega_m(i) = 1 - N\left[\left\|F^Q(i) - F_m^R\left[i + t_m(i)\right]\right\|_1\right] \quad (4)$$

where $N[\cdot]$ denotes the normalization operation and $\|\cdot\|_1$ is the L1 norm. Using (4), a smaller correspondence cost in the L1-norm term indicates a higher similarity between the pixel i in Q and its correspondence in R_m, and thus a higher value of the weight $\omega_m(i)$ is used for an intense saliency propagation. For the input image in Fig. 1(a), the correspondence saliency map is shown in Fig. 1(d), which effectively suppresses background regions and better highlights some salient object regions such as the horse's body. However, some regions such as horse's head and legs are not sufficiently highlighted in Fig. 1(d) due to some inaccurate correspondences and false suppressions on salient object regions in some initial saliency maps of the retrieved images as shown in Fig. 1(c).

C. Sample Selection

Based on the initial saliency map and the correspondence saliency map, the MKB method is exploited to boost the highlighted object pixels to the complete object with well-defined boundaries, without needing extra training data for learning. Different from [16], which directly selects for MKB the positive and negative samples by thresholding the saliency map, we introduce both initial saliency and correspondence saliency into a segmentation framework to obtain more reliable positive samples for salient objects.

Object segmentation can be usually formulated as a binary label assignment by solving the energy minimization problem of image pixels on an undirected graph [31]. Each pixel i in the input image Q gets the label $L_i \in \{0, 1\}$, where $L_i = 1$ denotes the object label and $L_i = 0$ denotes the background label. The segmentation energy function is defined as follows:

$$E^S(L) = \sum\nolimits_i U(i, L_i, G, S^Q, S^C) + \sum\nolimits_{(i,j)\in\Omega_s} V(i,j) \quad (5)$$

where Ω_s denotes the set of neighboring pixels. The smoothness term $V(\cdot)$ penalizes neighboring pixels with different labels based on the color contrast, and its definition is the same as [32]. The data term $U(\cdot)$ consists of both saliency term and appearance term with the following definition as:

$$U(i, L_i, G, S^Q, S^C) = \begin{cases} -\log N\left[S^Q(i) + S^C(i)\right] - \log G_O(i), & \text{if } L_i = 1 \\ -\log\left\{1 - N\left[S^Q(i) + S^C(i)\right]\right\} - \log G_B(i), & \text{if } L_i = 0. \end{cases} \quad (6)$$

In (6), the former term is the saliency term based on the initial saliency and the correspondence saliency, which indicates the location probability of salient object, and the latter term is the appearance term including $G_O(i)$ and $G_B(i)$, which are used to describe the color appearance of salient object and background, respectively, by using Gaussian mixture models (GMM). The average saliency value of S^Q is used to roughly separate salient object pixels and background pixels for initializing the parameters of the two GMMs. Meanwhile, the initial background pixels are used to constitute the negative samples.

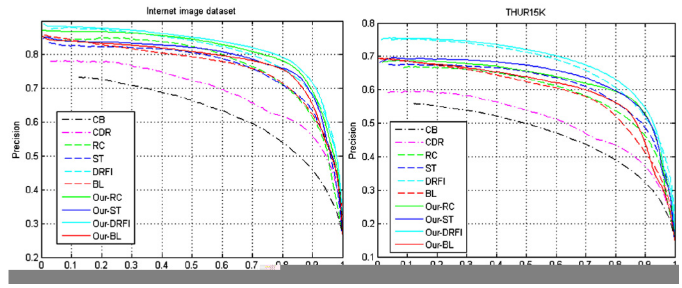

Fig. 2. (Better viewed in color). PR curves of different saliency models on the two datasets: Internet image dataset (left) and THUR15K (right).

The max-flow algorithm [33] is exploited to perform graph cut to optimize the pixel labels by updating the graph in an iterative way and obtain the final segmentation result in which the pixels labeled as object are reliable to constitute the positive samples.

For efficiency, we select training samples on the superpixel segmentation result, which is generated using the simple linear iterative clustering (SLIC) algorithm [34]. Superpixels have regular and compact shape with better boundary adherence. Then, each superpixel, which contains more than a half of pixels labeled as positive sample, is selected as the positive superpixel sample with a label of $+1$. The negative superpixel samples are selected using a similar way with a label of -1. Some superpixel samples selected by our method are shown in Fig. 1(f) in which some parts of the horse and background are accurately selected as positive and negative samples.

D. Saliency Map Generation

The MKB method combines SVMs with multiple kernels including linear, polynomial, radial basis function (RBF), and sigmoid as the weak classifiers to learn a boosted classifier by using the AdaBoost method [16]. The decision function is defined as follows:

$$F(r) = \sum\nolimits_{d=1}^{D} \beta_d h_d(r) \quad (7)$$

where $h_d(r) = \alpha_d^{\mathrm{T}} k_d(r) + b_d$ denotes an SVM classifier and D is the number of weak classifiers. Let H denote the total number of training samples, r_i and l_i denote the ith training sample and the corresponding positive/negative label assigned in Section II-C. The kernel $k_d(r)$ and the parameter vector α_d are represented as $[k_d(r, r_1), \ldots, k_d(r, r_H)]^{\mathrm{T}}$ and $[\alpha_{d,1} l_1, \ldots, \alpha_{d,H} l_H]^{\mathrm{T}}$, respectively. Then, α_d, b_d, and β_d are learned by using the AdaBoost optimization process. To represent the local features of each superpixel sample, we use the color histograms in both RGB and Lab color spaces, and the LBP histogram [35], which can better represent the local texture of superpixel, rather than using the Gist descriptor, which is a global descriptor unsuitable for representing local features.

The boosted classifier is applied to the test samples, i.e., all superpixels in the input image, to obtain the decision function values as the saliency values of superpixels, which are used to generate the boosted saliency map S^B. Then, we use S^B as the saliency term in (5) to build the graph, and perform the graph cut to obtain the segmentation mask M^B, which is further used to enhance the boosted saliency map S^B with the update as

Fig. 3. Visual comparison of saliency maps for some example images. For each example, in the top row from left to right: original image, ground truth, saliency maps generated using RC [9], ST [11], DRFI [14], and BL [16]; in the bottom row from left to right: co-saliency maps generated using CB [20] and CDR [25] and our saliency maps generated based on four different initial saliency models, i.e., Our-RC, Our-ST, Our-DRFI, and Our-BL.

follows:

$$S^B \leftarrow N\left[S^B + M^B\right]. \tag{8}$$

Based on the superpixel samples in Fig. 1(f), the boosted saliency map S^B is shown in Fig. 1(g), which can better highlight the complete salient object. Finally, by a linear combination of the initial saliency map and the boosted saliency map, the final saliency map S^F is defined as follows:

$$S^F = N\left[S^Q + S^B\right]. \tag{9}$$

As shown in Fig. 1(h), the final saliency map highlights the salient object with well-defined boundaries and suppresses the background well.

III. Experimental Results

We evaluated the proposed model on Internet image dataset [26] and THUR15K [3]. The images in both datasets were automatically downloaded through Internet, and they are suitable for validating the proposed model by similar image retrieval. Internet image dataset contains three classes of objects including car, airplane, and horse, and THUR15K are created by querying five keywords: giraffe, coffee mug, butterfly, plane, and dog jump. We evaluated the saliency detection performance on a total of 2488 images in Internet image dataset and 6233 images in THUR15K with pixel-wise binary ground truths. We compared with four state-of-the-art saliency models including RC [9], ST [11], DRFI [14], and BL [16]. Each saliency model is used for initial saliency detection in the proposed model, in order to evaluate the robustness to different initial saliency models. Furthermore, for each input image, the two state-of-the-art co-saliency models including CB [20] and CDR [25] are tested on the same group of images, which are selected using the method in Section II-A, for a fair comparison. For CB and CDR, the co-saliency map of the input image is used for comparison.

For an objective comparison, we generate the precision-recall (PR) curve by using a set of thresholds from 0 to 255 on saliency maps to obtain binary object masks, which are compared with the corresponding ground truth. The average precision and average recall are plotted to generate the PR curve on each dataset. As shown in Fig. 2, the proposed model with each of the four initial saliency models (Our-RC, Our-ST, Our-DRFI, and Our-BL) consistently outperforms the corresponding initial saliency model (RC, ST, DRFI, and BL). This clearly demonstrates that the proposed model can achieve a better saliency detection performance and show the robustness to different initial saliency models. In addition, the proposed model also significantly outperforms the two co-saliency models (CB and CDR). This indicates that the conventional co-saliency models do not show the advantage on a group of images which may have greater variations on color and shape of co-objects, while the proposed model with a variety of initial saliency models show the superiority by exploiting the co-object information and performing saliency correspondence propagation across images.

TABLE I
Average Processing Time (s) Per Image

Model	RC	ST	DRFI	BL	CB	CDR	Our
Time	0.20	28.1	5.5	20.8	5.4	28.9	18.2

For a subjective comparison, Fig. 3 shows some example saliency maps in which the leftmost three example images in the top row are from Internet image dataset and the remaining five example images are from THUR15K. We can see that saliency maps generated using the proposed model (Our-RC, Our-ST, Our-DRFI, and Our-BL) can better highlight the complete nonhomogeneous objects (the leftmost two examples in the top row), multiple objects (the leftmost three examples in the bottom row) and low-contrast objects (the third example in the bottom row), and can also suppress the complicated background regions more effectively (the rightmost two examples in the top row).

The proposed model is implemented using MATLAB on a PC with an Intel Core i7 3.5-GHz CPU and 16-GB RAM. For each model, the average processing time per image with a resolution of 320×213 is shown in Table I. Note that the average processing time of the proposed model in Table I excludes the time for generating initial saliency maps, which is dependent on different initial saliency models.

IV. Conclusion

In this letter, we propose a novel saliency detection framework via the effective use of correspondence information from similar images and boost the saliency detection performance by unsupervised learning. The similar images are retrieved from Internet image collections and used to propagate saliency via the pixel correspondence. Then training samples are extracted to learn the boosted classifier based on MKB method for generating the boosted saliency map, which enables to improve the saliency detection performance. Experimental results on two public image datasets demonstrate that the proposed model consistently outperforms the state-of-the-art single-image saliency models and co-saliency models.

References

[1] Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1275–1289, Aug. 2012.

[2] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV'01)*, Jul. 2001, vol. 2, pp. 416–423.

[3] M. M. Cheng, N. J. Mitra, X. Huang, and S. M. Hu, "SalientShape: Group saliency in image collections," *Vis. Comput.*, vol. 30, no. 4, pp. 443–453, Apr. 2014.

[4] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch, "Retargeting images and video for preserving information saliency," *IEEE Comput. Graph. Appl.*, vol. 27, no. 5, pp. 80–88, Sep. 2007.

[5] Y. Fang, Z. Chen, W. Lin, and C. W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

[6] Y. Fang, W. Lin, B. S. Lee, C. T. Lau, Z. Chen, and C. W. Lin, "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 187–198, Feb. 2012.

[7] R. Shah and P. J. Narayanan, "Interactive video manipulation using object trajectories and scene backgrounds," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1565–1576, Sep. 2013.

[8] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 55–69, Jan. 2013.

[9] M. M. Cheng, N. J. Mitra, X. L. Huang, P. Torr, and S. M. Hu, "Salient object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[10] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'13)*, Jun. 2013, pp. 1155–1162.

[11] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1937–1952, May 2014.

[12] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'13)*, Jun. 2013, pp. 3166–3173.

[13] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[14] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'13)*, Jun. 2013, pp. 2083–2090.

[15] F. Yang, H. Lu, and Y.-W. Chen, "Human tracking by multiple kernel boosting with locality affinity constraints," in *Proc. Asian Conf. Comput. Vis. (ACCV'10)*, Nov. 2010, pp. 39–50.

[16] N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient object detection via bootstrap learning," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'15)*, Jun. 2015, pp. 1884–1892.

[17] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, Dec. 2011.

[18] K. Chang, T. Liu, and S. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'11)*, Jun. 2011, pp. 2129–2136.

[19] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.

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

[21] H. Li, F. Meng, and K. N. Ngan, "Co-salient object detection from multiple images," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1896–1909, Dec. 2013.

[22] Z. Liu, W. Zou, L. Li, L. Shen, and O. Le Meur, "Co-saliency detection based on hierarchical segmentation," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 88–92, Jan. 2014.

[23] L. Li, Z. Liu, W. Zou, X. Zhang, and O. Le Meur, "Co-saliency detection based on region-level fusion and pixel-level refinement," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME'14)*, Jul. 2014, pp. 1–6.

[24] Y. Li, K. Fu, Z. Liu, and J. Yang, "Efficient saliency-model-guided visual co-saliency detection," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 588–592, May 2015.

[25] L. Ye, Z. Liu, J. Li, W-L. Zhao, and L. Shen, "Co-saliency detection via co-salient object discovery and recovery," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2073–2077, Nov. 2015.

[26] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, "Unsupervised joint object discovery and segmentation in internet images," in *Proc. IEEE Comput. Vis. Pattern Recog. (CVPR'13)*, Jun. 2013, pp. 1939–1946.

[27] S. Du and S Chen, "Detecting co-salient objects in large image sets," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 145–148, Feb. 2015.

[28] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, May 2001.

[29] C. Zhang, C. Shen, and T. Shen, "Unsupervised feature learning for dense correspondences across scenes," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 90–107, Jan. 2016.

[30] P. Felzenswalb and D. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 41–54, Oct. 2006.

[31] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV'01)*, Jul. 2001, pp. 105–112.

[32] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[33] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.