

# Saliency detection using texture and local cues

Qiang Qi<sup>2,1</sup>, Muwei Jian<sup>1,2\*</sup>, Yilong Yin<sup>1</sup>, Junyu Dong<sup>2</sup>, Wenyin Zhang<sup>3</sup>, Hui Yu<sup>4</sup>

<sup>1</sup>School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China

<sup>2</sup>Department of Computer Science and Technology, Ocean University of China, Qingdao, China

<sup>3</sup>School of Information Science and Engineering, Linyi University, Linyi, China

<sup>4</sup>School of Creative Technologies, University of Portsmouth, Portsmouth, UK

\*(Correspondence author) E-mail: jianmuweihk@163.com

**Abstract.** In this paper, a simple but effective method is proposed for detecting salient objects by utilizing texture and local cues. In contrast to the existing saliency detection models, which mainly consider visual features such as orientation, color, and shape information, our proposed method takes the significant texture cue into consideration to guarantee the accuracy of the detected salient regions. Firstly, an effective method based on selective contrast (SC), which explores the most distinguishable component information in texture, is used to calculate the texture saliency map. Then, we detect local saliency by using a locality-constrained linear coding algorithm. Finally, the output saliency map is computed by integrating texture and local saliency cues simultaneously. Experimental results, based on a widely used and openly available database, demonstrate that the proposed method can produce competitive results and outperforms some existing popular methods.

**Keywords:** Saliency detection; Selective contrast; Locality-constrained linear coding, Texture

## 1 Introduction

In recent years, saliency detection has attracted wide attention of many researchers and become a very active topic in computer vision research. Since it plays a significant role in many computer-vision related applications such as: object detection [1, 36, 37], image resizing [2, 4], image/video quality assessment [3], etc.

Generally, saliency detection methods can be traditionally summarized into two categories: the bottom-up and top-down approaches. Bottom-up saliency detection methods mainly rely on visual-driven stimuli, which directly utilize low-level features for saliency detection. One of the pioneering computational models of bottom-up

framework was proposed by Itti et al. in [5]. Later, in order to make salient object detection more precise, a novel graph-based method was proposed to estimate saliency map in [6]. While the graph-based method usually generates low resolution saliency maps. Then, a spectral residual-based saliency detection method was proposed by Hou and Zhang in [7]. However, the early researches are hard to handle well in the cases with complex scenes. To solve this problem and acquire accurate salient regions, Ma et al. [8] presented a local contrast-based method for saliency detection, which increases the accuracy of saliency detection. From then on, many saliency detection models have been proposed and achieved excellent performance in various fields. In [9], a discriminant center-surround hypothesis model was proposed to detect salient objects. In [10], a bottom-up framework by using natural statistics was proposed to detect salient image regions. Achanta et al. [11] proposed a novel saliency detection method based on frequency tuned model to detect salient object. In [12], a new saliency measure was proposed by using a statistical framework and local feature contrast. Wang et al. [13] proposed a novel saliency detection model based on analyzing multiple cues to detect salient objects. In [14], a hierarchical graph saliency detection model was proposed via using concavity context to compute weights between nodes. Tavakoli et al. [15] designed a fast and efficient saliency detection model using sparse sampling and kernel density estimation. In [16], a context-aware based model, which aims at using the image regions to represent the scene, was proposed for saliency detection. Yang et al. [17] presented a new method based on foreground and background cues to achieve a final saliency map. In [18], a novel approach, which can combine contrast, center and smoothness priors, was proposed for salient detection. Margolin et al. [19] introduced an effective framework by involving pattern and color distinctness to estimate the saliency value in an image. In [20], a cluster-based method was proposed for co-saliency detection. In [21], a novel saliency measure based on both global and local cues was proposed. Li et al. [22] proposed a visual saliency detection algorithm based on dense and sparse reconstruction. In [23], a novel Cellular Automata (CA) model was introduced to compute the saliency of the objects. In [24], a visual-attention-aware model was proposed for salient-object detection. Recently, a bottom-up saliency-detection method by integrating Quaternionic Distance Based Weber Descriptor (QDWD), center and color cues, was presented in [25]. In [26], a novel and effective deep neural network method incorporating low-level features was proposed for salient object detection.

Top-down based saliency detection models, which consider both visual information and prior knowledge, are generally task dependent or application oriented. Thus, the top-down model normally needs supervised learning and is lacking in extendibility. Compared with the bottom-up saliency detection model, not much work has been proposed. Kanan et al. [27] presented an appearance-based saliency model using natural statistics to estimate the saliency of an input image. In [28], a novel top-down framework, which incorporated a tightly coupled image classification module, was proposed for salient object detection. In [29], a weakly supervised top-down approach was proposed for saliency detection by using binary labels.

In this paper, a novel and efficient method based on a bottom-up mechanism, by integrating texture and local cues together, is proposed for saliency detection. In con-

trast to the existing saliency detection models, we compute the texture saliency map based on selective contrast (SC) method, which can guarantee the accuracy of the detected salient regions. In addition, we incorporated an improved locality-constrained linear coding algorithm (LLC) to detect the local saliency in an image. In order to evaluate the performance of our proposed method, we carried out experiments based on a widely used dataset. Experimental results, compared with other state-of-the-art saliency-detection algorithms, show that our approach is effective and efficient for saliency detection.

The remainder of the paper is organized as follows. In Section 2, we introduce the proposed saliency-detection algorithm in detail. In section 3, we demonstrated our experimental results based on a widely used dataset and compared the results with other eight saliency detection methods. The paper closes with a conclusion and discussion in Section 4.

## 2 The proposed saliency detection method

This section presents the proposed saliency detection method by using selective contrast (SC) method for describing the texture saliency and locality-constrained method for estimating the local saliency. Texture saliency estimation method based on SC will first be described, followed by the locality-constrained method. All these different types of visual cues are fused to form a final saliency map.

### 2.1 Texture saliency detection based on selective contrast

Texture is an important characteristic for human visual perception, which is caused by different physical-reflection properties on the surface of an object with the gray level or color changes. An image is not just a random collection of texture pixels, but a meaningful arrangement of them. Different arrangements of these pixels form different textures would provide us with important saliency information. As a basic property of image, texture also affects the similarity degree among regions, which is a useful cue in saliency detection, so we take the texture cue into consideration in our designed model.

In this section, we adopt selective contrast (SC) based algorithm [30] to estimate the texture saliency. Assume that the pixels in an input image are denoted as  $X_n$ ,  $n=1,2,3,\dots,N$ , where  $N$  is the number of pixels in the image and  $\{X_n\}$  is a texture vector, which is achieved by using the uniform LBP [35]. Since the outputs of obtained textures span a very wide range of high dimensional spaces, we hope to express them in a more compact way. In order to solve this issue, we use *k-means* to cluster these texture expressions and consider the cluster centers as the representative textures. After this transformation, each texture is denoted as its nearest texture prototype, this expression is called selective contrast (SC), which explores the most distinguishable component information in texture. Thus, each pixel's texture saliency based on selective contrast (SC) can be written as follows [30]:

$$S_i(i, R_i) = \sum_{j \in R_i} d(i, j), \quad (1)$$

where  $i$  is the examined pixel,  $R_i$  denotes the supporting region for defining the saliency of pixel  $i$ , and  $d(i, j)$  is the distance of texture descriptors between  $i$  and  $j$ . The  $\ell_2$  norm can be used to define the distance measure of textures descriptors:

$$d(i, j) = \|k_i - k_j\|, \quad (2)$$

where  $k_i$  and  $k_j$  are the transformed textures expressions of pixels  $i$  and  $j$  by  $k$ -means.

In order to further improve the efficiency by reducing the computational complexity, a limited number of textons are trained from a set of images and the textures can be quantized to  $M$  textons. By this means, the computation is reduced to looking up a distance dictionary  $D_i$  of  $M \times M$  dimensions. Thus, the Eq. (1) can be rewritten as:

$$S_i(i, R_i) = \sum_{\psi(j)} f(\psi(j)) D_i(\psi(i), \psi(j)), \quad (3)$$

where  $i, R_i$  have a similar meaning in Eq. (1),  $\psi(i)$  is the function mapping pixel  $i$  to its corresponding prototype texture and  $f(\psi(j))$  is the frequency of  $\psi(j)$  in region  $R_i$ . Readers can be referred to the [30] for more details.

Fig. 1(b) shows some texture saliency maps of the input images. From the results we can see that the texture saliency maps can exclude most of the background pixels and detect almost the whole salient objects in images, but it leads to missing some homogeneous regions which have the similar texture appearances.

## 2.2 Local saliency detection using locality-constrained method

The motivation of local estimation is the local outliers, which are standing out from their neighbors with different colors or textures and tend to attract human attention. In order to detect local outliers and get acceptable performance, local coordinate coding method [31], which described the locality is more essential than sparsity, has been used in saliency detection. Furthermore, the proposed SC model in Section 2.1 only takes the texture information into consideration, which misses some local information. Thus, we employ an improved algorithm based on locality-constrained linear coding (LLC) [32] to estimate the local saliency.

For a given image, we first over-segmented the image into  $N$  regions,  $\{r_i\}, i=1, 2, \dots, N$ . For each region  $r_i$ , let  $\mathbf{X}$  be a set of 64-dimensions local descriptors extracted from the image, and  $\mathbf{X} = [x_i^0, x_i^1, \dots, x_i^{63}]^T$ ,  $i=1, 2, \dots, N$ . Therefore, the original function of LLC method is written as follows [32]:

$$\min_B \left\{ \sum_{i=1}^N \left( \|x_i - Db_i\|^2 + \omega \|dr_i .* b_i\|^2 \right) \right\}, \quad s.t. \quad \mathbf{1}^T b_i = 1, \quad \forall_i, \quad (4)$$

where  $\mathbf{B}=[b_1, b_2, \dots, b_N]$  is the set of codes for  $\mathbf{X}$ ,  $\mathbf{D}=[d_1, d_2, \dots, d_M]$  is the codebook with  $M$  entries, and the parameter  $\omega$  is used to balance the weight between the penalty term and regularization term. The constraint  $\mathbf{1}^T b_i = 1$  follows the shift-invariant requirements of the LLC coding and  $\cdot^*$  denotes an element-wise multiplication. Here, the  $dr_i$  is the locality adaptor that gives different freedom for each codebook vector based on its similarity to the input descriptor  $x_i$ , and is defined:

$$dr_i = \exp\left(\frac{[dist(x_i, \mathbf{D})]}{\lambda}\right), \quad (5)$$

where  $dist(x_i, \mathbf{D})=[dist(x_i, d_1), dist(x_i, d_2), \dots, dist(x_i, d_M)]^T$ ,  $dist(x_i, d_i)$  denotes the Euclidean distance between  $x_i$  and the code book vector  $d_i$ ,  $\lambda$  is used to adjust the weight decay speed for the locality adaptor, and  $M$  is the number of elements in the codebook. More details about LLC can be referred to [31, 32].

In this paper, we adopted an improved LLC algorithm [32] to detect the local cue. We consider the  $K$  nearest neighbors in spatial as the local basis  $\mathbf{D}_i$  owing to the vector  $b_i$  in Eq. (4) with a few non-zero values, which means that it is sparse in some extent. It should be noted that the  $K$  is smaller than the size of the original codebook  $M$ . Thus, the Eq. (4) can be rewritten as follows:

$$\min_b \left( \sum_{i=1}^N \|x_i - \mathbf{D}_i b_i\|^2 \right), \quad s.t. \quad \mathbf{1}^T b_i = 1, \quad \forall_i, \quad (6)$$

where  $\mathbf{D}_i$  denotes the new codebook for each region  $r_i$ ,  $i=1, 2, \dots, N$  and  $K$  is the size of the new codebook and is empirically set at  $K = 2M/3$ .

Unlike the traditional LLC algorithm, solving the improved LLC algorithm is simple and the solution can be derived analytically by

$$b_i = 1 / (\mathbf{C}_i + \omega \times dig(\mathbf{C}_i)), \quad (7)$$

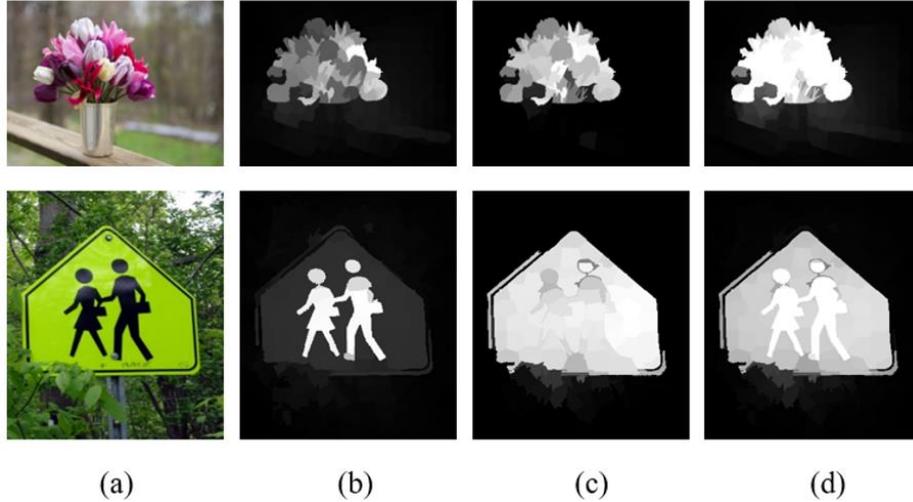
$$\square b_i = b_i / \mathbf{1}^T b_i, \quad (8)$$

where  $\mathbf{C}_i = (\mathbf{D}_i - \mathbf{I}x_i^T)(\mathbf{D}_i - \mathbf{I}x_i^T)^T$  represents the covariance matrix of the feature and  $\omega$  is a regularization parameter, which is set to be 0.1 in the proposed algorithm. As the solution of the improved LLC method is simple and fast, therefore, the local saliency value of the region  $r_i$  can be defined as follows [21]:

$$S(r_i) = \|\square x_i - \mathbf{D}_i \square b_i\|^2, \quad (9)$$

where  $\square b_i$  is the solution of Eq. (6), which is achieved by Eq. (7) and Eq. (8).

Fig. 1(c) shows some local saliency maps. We can see that the local saliency maps achieve more reliable local detailed information owing to the locality-constrained coding model.



**Fig. 1.** The proposed method for saliency detection. (a) Input images, (b) texture saliency maps, (c) local saliency maps, (d) the final saliency maps.

### 2.3 Final Saliency Fusion

The texture saliency map  $S_t$  and the local saliency map  $S_l$  are linearly combined with adaptive weights to define the final saliency map. Therefore, the final saliency map can be defined as follows:

$$Sal = \alpha S_t + \beta S_l, \quad (10)$$

where  $S_t$  and  $S_l$  are the texture saliency and the local saliency, respectively.  $\alpha$  and  $\beta$  are the weights of the texture saliency and the local saliency, accordingly. In this paper, we introduced a more effective and logical fusion method to adjust the weights between the different feature maps adaptively. We used the *DOS* (degree-of-scattering) of saliency map to determine the weighting parameters and set the  $\alpha = 1 - DOS^t$ ,  $\beta = 1 - DOS^l$  according to the method in [34].

## 3 Experimental results

We perform saliency detection experiments based on a widely used dataset: ECSSD [33], which included 1000 images acquired from the internet, and compare our method with other eight state-of-the-art methods including the Spectral Residual (SR) [7], Saliency detection using Natural Statistics (SUN) [10], Segmenting salient objects (SEG) [12], Context-aware (CA) saliency detection [16], Graph-regularized (GR) saliency detection [18], Pattern distinctness and color (PC) based method [19],

Dense and sparse reconstruction (DSR) [22], Background-Single Cellular Automata (BSCA) based saliency detection [23].

In order to quantitatively compare the state-of-the-art saliency-detection methods, the average precision, recall, and  $F$ -measure are utilized to measure the quality of the saliency maps based on setting a segmentation threshold for binary segmentation. The adaptive threshold is twice the average value of the whole saliency map to get the accurate results. Each image is segmented with superpixels and masked out when the mean saliency values are lower than the adaptive threshold. The  $F$ -measure is defined as follows:

$$F_{\beta} = \frac{(1 + \beta) \times \text{Precision} \times \text{Recall}}{\beta \times \text{Precision} + \text{Recall}}, \quad (11)$$

where  $\beta$  is a real positive value and is set at  $\beta = 0.3$ .

We compared the performance of the proposed method with other eight state-of-the-art saliency detection methods. Fig. 2 shows some saliency detection results of different methods based on the ECSSD database. The output results show that our final saliency maps can accurately detect almost entire salient objects and preserve the salient object's contours more clearly.

We also used the precision, recall and the  $F$ -measure to evaluate the performance of different methods objectively. Fig. 3 shows the comparisons of different methods under different evaluation criterions. As can be seen from Fig. 3, our proposed method outperforms the other eight methods in terms of detection accuracy and the proposed method achieves the best overall saliency-detection performance (with precision = 79.0%, recall = 80.0%), and the  $F$ -measure is 79.2%. The experiment results show that the proposed model is efficient and effective.

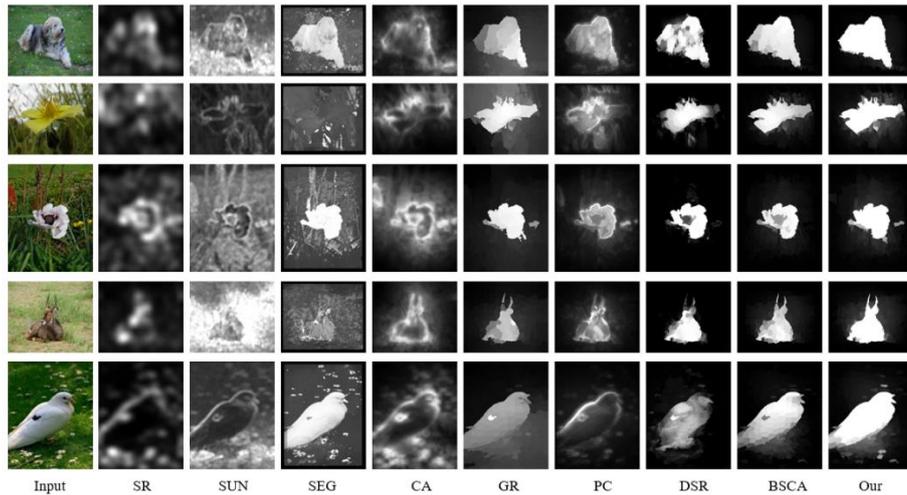


Fig. 2. Qualitative comparisons of different approaches based on ECSSD database.

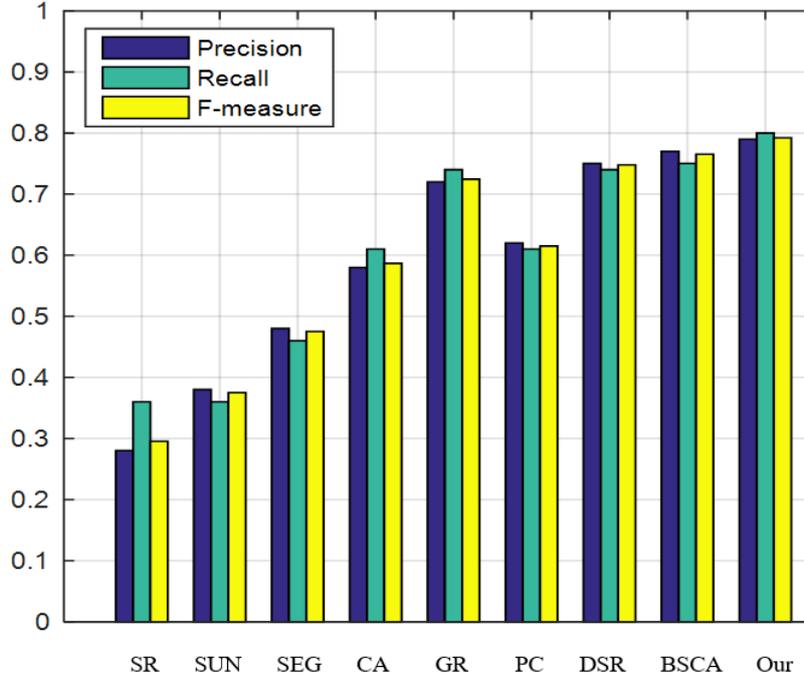


Fig. 3. Comparison of different saliency-detection methods in terms of average precision, recall, and  $F$ -measure based on ECSSD database.

## 4 Conclusion and discussion

This paper proposes a novel bottom-up method for efficient and accurate image saliency detection. This proposed approach integrated texture and local cues to estimate the final saliency map. The texture saliency maps are computed based on the selective contrast (SC) method, which explores the most distinguishable component information in texture. The local saliency maps are achieved by utilizing a locality-constrained method. We also evaluated our method based on a publicly available dataset and compared our proposed approach with other eight different state-of-the-art methods. The experimental results show that our algorithm can produce promising results compared to the other state-of-the-art saliency-detection models.

## 5 Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) (61601427); Natural Science Foundation of Shandong Province

(ZR2015FQ011); Applied Basic Research Project of Qingdao (16-5-1-4-jch); China Postdoctoral Science Foundation funded project (2016M590659); Postdoctoral Science Foundation of Shandong Province (201603045); Qingdao Postdoctoral Science Foundation funded project (861605040008) and The Fundamental Research Funds for the Central Universities (201511008, 30020084851).

## 6 References

1. Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Zhuowen Tu, Ali Borji, Deeply supervised salient object detection with short connections, *IEEE CVPR*, 2017.
2. Avidan S, Shamir A, editors. Seam carving for content-aware image resizing. *ACM Transactions on Graphics (TOG)*; 2007: ACM.
3. Ma Q, Zhang L, Wang B. New strategy for image and video quality assessment. *Journal of Electronic Imaging*. 2010; 19(1):011019–14.
4. Ming-Ming Cheng, Yun Liu, Qibin Hou, Jiawang Bian, Philip Torr, Shi-Min Hu, Zhuowen Tu, HFS: Hierarchical Feature Selection for Efficient Image Segmentation, *ECCV*, 2016.
5. L. Itti, C. Koch, E. Niebur, “A model of saliency based visual attention for rapid scene analysis,” *IEEE Trans. PAMI*, vol. 20, no. 11, pp. 1254-1259, 1998.
6. J. Harel, C. Koch, P. Perona, “Graph-based visual saliency,” *Advances in NIPS*, pp. 545-552, 2006.
7. X. Hou, L. Zhang, “Saliency detection: a spectral residual approach,” *IEEE CVPR*, pp. 1-8, 2007.
8. Y. F. Ma and H. J. Zhang, “Contrast-based image attention analysis by using fuzzy growing,” *ACM conference on Multimedia*, pp. 374-381, 2003.
9. D. Gao, V. Mahadevan, and N. Vasconcelos, “The discriminant center-surround hypothesis for bottom-up saliency,” *Advances in NIPS*, pp. 497-504, 2007.
10. L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell, “SUN: A bayesian framework for saliency using natural statistics,” *Journal of Vision*, vol. 8, no. 7, pp. 1–20, 2008.
11. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-Tuned Salient Region Detection,” *IEEE CVPR*, pp. 1517-1604, 2009.
12. E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, “Segmenting salient objects from images and videos,” in *Proc. 11th ECCV*, 2010, pp. 366–379.
13. L. Wang, J. Xue, N. Zheng, and G. Hua, “Automatic salient object extraction with contextual cue,” *IEEE CVPR*, pp. 105-112, 2011.
14. Y. Lu, W. Zhang, H. Lu, and X. Y. Xue, “Salient object detection using concavity context,” *IEEE CVPR*, pp. 233-240, 2011.
15. H. R. Tavakoli, E. Rahtu, and J. Heikkilä, “Fast and efficient saliency detection using sparse sampling and kernel density estimation,” in *Proc. 17th Scand. Conf. Image Anal.*, 2011, pp. 666–675.
16. Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 1915-1926.
17. C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, “Saliency detection via graph-based manifold ranking,” *IEEE CVPR*, pp. 3166-3173, 2013.
18. C. Yang, L. Zhang, and H. Lu, “Graph-regularized saliency detection with convex-hull-based center prior,” *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013
19. R. Margolin, A. Tal, and L. Zelnik-Manor, “What makes a patch distinct?” in *Proc. IEEE Conf. CVPR*, Jun. 2013, pp. 1139–1146.

20. Fu H, Cao X, Tu Z. Cluster-based co-saliency detection[J]. *IEEE Transactions on Image Processing*, 2013, 22(10): 3766-3778.
21. Tong N, Lu H, Zhang Y, et al. Salient object detection via global and local cues. *Pattern Recognition*, 2015, 48(10): 3258-3267.
22. X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE ICCV*, Dec. 2013, pp. 2976–2983.
23. Qin, Y., Lu, H., Xu, Y., et al.: Saliency detection via cellular automata. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 110-119. (2015)
24. Jian, M., Lam, K.M., Dong, J.J., Shen, L.L.: Visual-Patch-Attention-Aware Saliency Detection. *IEEE T. Cybernetics* vol. 45, no. 8, pp. 1575-1586. (2015)
25. Jian, M.M., Qi, Q., Sun, Y., Lam, K.M., et al: Saliency Detection Using Quaternionic Distance Based Weber Descriptor and Object Cues. *APSIPA ASC'2016*, Korean. (2016)
26. Li, H., Chen, J., Lu, H., et al.: 'CNN for saliency detection with low-level feature integration', *Neurocomputing*, 2017, 226, pp.212-220.
27. Kanan C, Tong M H, Zhang L, et al. SUN: Top-down saliency using natural statistics[J]. *Visual cognition*, 2009, 17(6-7): 979-1003.
28. Cholakkal H, Johnson J, Rajan D. A classifier-guided approach for top-down salient object detection. *Signal Processing: Image Communication*, 2016, 45: 24-40.
29. Cholakkal H, Johnson J, Rajan D. Weakly Supervised Top-down Salient Object Detection. *arXiv preprint arXiv:1611.05345*. 2016.
30. Wang Q, Yuan Y, Yan P. Visual saliency by selective contrast. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(7): 1150-1155.
31. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In *Advances in Neural Information Processing Systems*, pp. 2223–2231. (2009)
32. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367. (2010)
33. Yan, Q., Xu, L., Shi, J., Jia, J., 'Hierarchical Saliency Detection', *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
34. Tian H, Fang Y, Zhao Y, et al. Salient region detection by fusing bottom-up and top-down features extracted from a single image. *IEEE Transactions on Image processing*, 2014, 23(10): 4389-4398.
35. Guo Z, Zhang L, Zhang D. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern recognition*, 2010, 43(3): 706-719.
36. Muwei Jian, Qiang Qi, et al. "Saliency Detection Using Quaternionic Distance Based Weber Local Descriptor and Level Priors", *Multimedia Tools and Applications* (2017). DOI: 10.1007/s11042-017-5032-z
37. Muwei Jian, Qiang Qi, et al. "The OUC-Vision Large-Scale Underwater Image Database". *Proceedings 2017 IEEE International Conference on Multimedia and Expo (ICME 2017)*, Hong Kong, July, 2017.