



A Salient Object Detection Algorithm Based on Region Merging and Clustering

Weiyi Wei, Yijing Yang, Wanru Wang, Xiufeng Zhao, Huifang Ma

► To cite this version:

Weiyi Wei, Yijing Yang, Wanru Wang, Xiufeng Zhao, Huifang Ma. A Salient Object Detection Algorithm Based on Region Merging and Clustering. 11th International Conference on Intelligent Information Processing (IIP), Jul 2020, Hangzhou, China. pp.3-13, 10.1007/978-3-030-46931-3_1 . hal-03456978

HAL Id: hal-03456978

<https://inria.hal.science/hal-03456978>

Submitted on 30 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A Salient Object Detection Algorithm Based on Region Merging and Clustering

Weiwei Wei¹, Yijing Yang^{1,*}, Wanru Wang¹, Xiufeng Zhao¹, Huifang Ma¹

¹College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu

^{1,*}Corresponding author 2190438736@qq.com

Abstract. Salient object detection has recently drawn much attention in computer vision such as image compression and object tracking. Currently, various heuristic computational models have been designed. However, extracting the salient objects with a complex background in the image is still a challenging problem. In this paper, we propose a region merging strategy to extract salient region. Firstly, boundary super-pixels are clustered to generate the initial saliency maps based on the prior knowledge that the image boundaries are mostly background. Next, adjacent regions are merged by sorting the multiple feature values of each region. Finally, we get the final saliency maps by merging adjacent or non-adjacent regions by means of the distance from the region to the image center and the boundary length of overlapping regions. The experiments demonstrate that our method performs favorably on three datasets than state-of-art.

Keywords: Salient Object Detection, Clustering, Region Merging.

1 Introduction

Salient object detection is an essential problem in computer vision which aims at highlighting the visually outstanding regions/object/structures from the surrounding background [1]. It has received substantial attention over the last decade due to its wide range of applications in image compression [2], behavior recognition [3] and co-segmentation [4].

According to the human visual systems, salient object detection methods can be divided into two categories: one is a bottom-up method; the other is the top-down methods. In the top-down approaches, the salient object obtained in a scene is always discrepant for different people. Therefore, it is more complex to construct the top-down salient object detection model, so there are few models for salient object detection. On the contrary, the bottom-up approach has attracted much attention and many salient object detection models have been proposed. In the traditional bottom-up methods, the salient object detection based on spatial and frequency domains. In spatial domains, Sun et al. [5] propose a salient object detection method for region merging. Shen et al. [6] propose a unified approach via low-rank matrix recovery.

Peng et al. [7] propose a structured matrix decomposition model to increase the distance between the background and the foreground for salient object detection. Feng et al. [20] propose a salient detection method by fusing image color and spatial information to obtain saliency values to separate foreground and background for images. Xu et al. [21] propose a universal edge-oriented framework to get saliency maps. These salient object detection methods mentioned above could basically complete the salient object detection task, but the results are not accurate. Zhao et al. [24] obtain the saliency map by combining sparse reconstruction and energy optimization. Yu et al. [25] construct the quaternion hypercomplex and multiscale wavelet transform to detect the salient object. Guo et al. [26] obtain the saliency map by combining the boundary contrast map and the geodesics-like map. Marcon et al. [27] propose an update to the traditional pipeline, by adding a preliminary filtering stage to detect the salient object. As the demand for application scenarios continues to increase, co-saliency detection method [8] and moving salient object detection method [9] also introduced into the field of salient object detection, but there are also challenges for its application.

In order to improve the accuracy of the salient object detection results, several frequency domain methods have been developed, Hou et al. [10] argue that a salient detection model for spectral residuals. Guo et al. [11] use the spectral phase diagram of the quaternion Fourier transform to detect the spatiotemporal salient object. Achanta et al. [12] exploit color and luminance features detect the salient region by frequency tuning method. Early studies of salient object detection models are designed to predict human eye fixation. It mainly detects salient points where people look rather than salient regions [13]. In [14] for each color subband, the saliency maps are obtained by an inverse DWT over a set of scale-weighted center-surround responses, where the weights are derived from the high-pass wavelet coefficients of each level. Spectrum analysis is easy to understand, computationally efficient and effective, but the relevant interpretation of the frequency domain is unclear.

Nowadays, the deep learning based methods have been applied to saliency detection. Li et al. [22] propose a CNN based saliency detection methods by fusing deep contrast feature and handcrafted low-level features to acquire saliency maps. Li et al. [23] use optimized convolutional neural network extract the high-level features, and then the high-level features and low-level features were fused to obtain the fusion features. Finally, the fusion features were input into SVM to separate the foreground and background for images. However, these two methods have high time complexity.

This paper focuses on the incomplete salient region and unclear boundary of the salient regions in the detection results, we propose a super-pixel level salient object detection method based on clustering and region merging strategy to improve the performance of detecting the salient object in a scene. The work flow of our salient object detection framework is described in Fig1. Firstly, the SLIC [15] (Simple Linear Iterative Cluster) algorithm is used for super-pixel segmentation. Secondly, generate the initial saliency maps. This step involves two processes: the super-pixels lie in image boundaries are clustered and calculate the distance between the remaining super-pixels and the classes centers. Color background maps and spatial relationship background maps are calculated by calculating the distance between the super-pixels

and the boundary class. Then the color background maps and the spatial relation background maps are combined linearly to produce the initial saliency map. Thirdly, the final saliency maps obtained after two steps regions merging. The first stage is based on the initial saliency map where adjacent regions are merged by taking into account color features and spatial information. The second stage is to calculate the saliency value of the regions and merge the regions according to the saliency value. The main contributions are summarized as follows:

- 1) We cluster the boundary super-pixels to generate the initial saliency maps based on the prior knowledge that the image boundaries are mostly background.
- 2) A two-stage optimization strategy is proposed to further optimize the saliency. In the first stage, we merge adjacent small regions with identical or similar features. In the second stage, we merge adjacent or no-adjacent regions according to the ranking of regional saliency values.

2 Construction of Initial Saliency Map

Given an image I , we represent it by N super-pixels produced by SLIC. Then M different classes are generated by clustering boundary super-pixels in the CIELab color space with K-means algorithm. We establish an initial saliency map with three steps: (i) Constructing color background maps. (ii) Constructing spatial correlation background maps. (iii) Constructing initial saliency maps. The color background maps and spatial background maps are merged to construct initial saliency maps. The specific process is as follows.

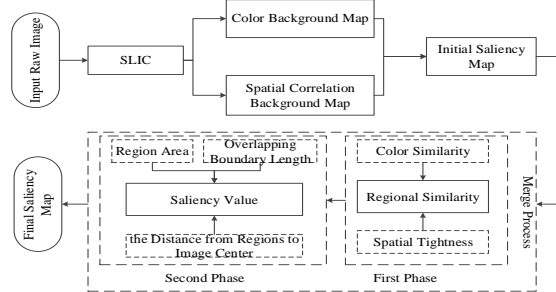


Fig. 1. The work flow of the proposed salient object detection framework.

2.1 Constructing Color Background Map

In this paper, we first use the k-means algorithm to cluster the boundary super-pixels into M classes. Next, calculate the feature gap between other super-pixels and the M different classes, this can measure the probability that the super-pixels to be the object regions. In this paper, we build M background maps C_m based on the number of M clusters. C_m is defined as the color feature difference between super-pixels $i(i=1,2,\dots,N)$ and the boundary classed $m(m=1,2,\dots,M)$. The class maps can be formulated as follows.

$$C_m = \frac{1}{p_m} \sum_{i=1}^N \frac{1}{\exp(\frac{-|c_m, c_i|}{2\alpha_1^2}) + \eta} \quad (1)$$

Where c_m and c_i denotes the color features of each boundary class and super-pixel respectively. $|c_m, c_i|$ denotes the Euclidean distance between each super-pixel and boundary classes in CIELab color space, α_1, η are balance factors, $p_m (m=1, 2, \dots, M)$ is the number of super-pixels in different boundary classes. Through lots of experiments, the detection results are more accurate when $\alpha_1 = 0.2$, $\eta = 10$, $M=3$. The three background maps generated in this paper are shown in Fig 2.



Fig .2. Color background maps. From left to right are Input image; 1st class map, 2nd class map, 3rd class map, GT.

2.2 Constructing Spatial Correlation Background Map

In salient object detection, distance is always used to measure the spatial correlation between regions. The closer the distance between the two regions is, the greater the degree of influence and association in the space, and vice versa [16]. So we construct the spatial correlation saliency maps between each super-pixel and the boundary class by means of the spatial distance relationship between two regions. We defined S_m as the spatial distance between each super-pixel and class $m (m=1, 2, \dots, M)$, so the S_m can be formulated as follows.

$$S_m = \frac{1}{p_m} \sum_{i=1}^N \exp(\frac{-|s_i, s_m|}{2\alpha_2^2}) \quad (2)$$

where s_i, s_m denotes the spatial location of super-pixel i and m respectively. α_2 is control parameter, and it is learned by experience that setting $\alpha_2 = 1$ is most suitable. We calculate the distance between the super-pixels of the first class of background area and the other super-pixels to obtain the 1st class map. In the experiment, we calculated the spatial correlation background maps of three class of background areas, in which the bright part represents the salient area and the dark area represents the no-salient area. The boundary clustering results are shown in Fig 3.

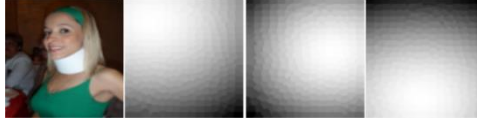


Fig .3. Spatial correlation background maps. From left to right are Input image, 1st class map, 2nd class map, 3rd class map.

2.3 Constructing Initial Saliency Map

As shown in the first map of Fig. 4, because there are two black objects, it is very difficult to distinguish their saliency by means of the color feature. But it can be seen from the map that the aircraft is closer to the center of the image than the tree, so we

can use spatial distance to enhance the saliency of the aircraft and weaken the saliency of the tree. Let S_m denotes the intensity factor to restrict the C_m . The initial saliency map can be written as follows.

$$SC = \sum_{m=1}^M C_m \cdot S_m \quad (3)$$

The merging result as shown in the third map of Fig 4. we can see from the result map that the brightness of the aircraft is higher than the tree. That is, the saliency of aircraft is strengthened and the saliency of the tree is weakened. Therefore, the image merging method is effective.



Fig .4. Initial saliency map. From left to right are Input image, GT, Initial saliency map.

3 Optimization of Merger Strategy

The object region can be revealed after merging, but we can see from the third map in Fig 4, many background regions are also enhanced as the object regions are brightened. In order to optimize the rough initial saliency maps and detect salient object more accurate, a regional merging model is established in this section, which merges those adjacent background regions with similar features. In this paper, we merge the relevant regions with two stages. Firstly, we merge adjacent small regions with identical or similar features. Secondly, we merge adjacent or non-adjacent regions according to the ranking of regional saliency values.

3.1 The First Phase of the Merging

The relationship tightness of each region with its adjacent regions be calculated first and then merging them according to the degree of the relationship tightness. In this paper, the relationship tightness between regions is represented by color similarity and spatial tightness.

Color Similarity. Color similarity is defined as the difference in color feature between the super-pixel i and j , it can be formulated as follows.

$$C(i, j) = \|c_i - c_j\| \quad (4)$$

Where c_i and c_j represent the average color value for super-pixels i and j respectively, $\|c_i - c_j\|$ is the Euclidean distance between the region i and adjacent region j .

Spatial Tightness. The spatial tightness between two regions can measure their degree of association in space. The closer the distance between the two regions is, the greater their spatial tightness. The correlation is measured by the boundary length of the overlapping regions. The longer the length of the boundaries that they co-own is, the greater the degree of association is. Let $ST(i, j)$, $S(i, j)$ and $T(i, j)$ as the spatial

tightness, the spatial distance and spatial tightness between the super-pixel i and adjacent super-pixel j . The formula can be written respectively as follows.

$$S(i, j) = \|s_i - s_j\| \quad (5)$$

$$T(i, j) = \frac{B(i, j)}{\min(L(i), L(j))} \quad (6)$$

$$ST(i, j) = (1 - S(i, j)) \cdot T(i, j) \quad (7)$$

Where s_i, s_j denote the spatial location for super-pixel i and j , $\|s_i - s_j\|$ is the Euclidean distance between region i and region j , $B(i, j)$, $L(i)$ and $L(j)$ denote the length of their overlapping boundaries, the length of the boundary of super-pixels i and j . Finally, we define $P(i, j)$ as the importance of the region j to the adjacent region i by calculating their color similarity and spatial tightness. The region i and the region j will be merged if they have higher importance. The importance $P(i, j)$ can be formulated as follows.

$$P(i, j) = \omega_1(1 - C(i, j)) + \omega_2 ST(i, j) \quad (8)$$

Some small region s_i must be merged into a large region s_j with greater relationship tightness. The merging results are shown in Fig. 5. Through experiments, the detection result is more accurate when $\omega_1 = 0.68$, $\omega_2 = 0.32$.

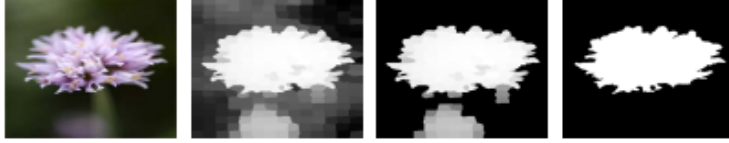


Fig .5. The first phase merges maps. From left to right are input image, initial saliency map, first phase of the merging map, GT.

3.2 The Second Phase of the Merging

Based on the above first merging, it is noted that the majority of background regions are merged. However, there are some non-adjacent or adjacent background regions with greater intensity are not belong to salient regions (as shown in the third map of Fig 5). Hence, in order to eliminate these background regions, we use Sal to denote these regional saliency value, utilize the Sal value to decide whether to merge it with background regions or not. Sal is influenced by the region area(V), the distance from regions to image center (CS) and the overlapping boundary length with image boundaries(BL). The region with a low value of Sal is merged with background regions.

Region Area. Regions with the larger area are more salient than smaller regions, so larger regions will be assigned greater saliency values. The formula can be written as

$$V = \frac{v}{\max(v)} \quad (9)$$

Where v is the area of each associated object in the image, and $\max(v)$ is the area of the largest associated object region in the first merged maps.

Center Distance. In general, the center regions of the image more attractive than others, and they are more likely to be salient regions [16]. In this paper, therefore, calculates the region Sal according to the distance between the regions and the center of the scene.

$$CS = \frac{\sum_{i=1}^{nP} \|R, PC_i\|}{areaP} \quad (10)$$

Where nP is the number of super-pixel in the region P , PC is the spatial location of the super-pixels included in the region P , R is the center position of the image, and $areaP$ is the area of the region P .

Overlapping Boundary Length. The longer the overlapped boundary between a region with the image boundaries is, the greater the possibility of being background regions, so BL can be written as follows.

$$BL = \frac{PB}{RB} \quad (11)$$

Where PB is the length of overlapped boundary that between each salient object and image boundaries, and RB is the total length value of the image boundaries. Hence, the saliency value Sal is defined as follows.

$$Sal = V + CS - BL \quad (12)$$

We merge the small salient objects into the large area by means of the Sal sorted. When some regions Sal value is very close, these regions can't be retained. The results are shown in Fig 6.

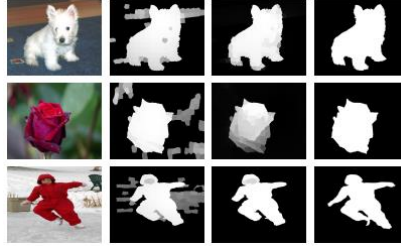


Fig .6. Second phase merge map. From left to right are Input image. The first stage merged map. The second stage merged map. GT.

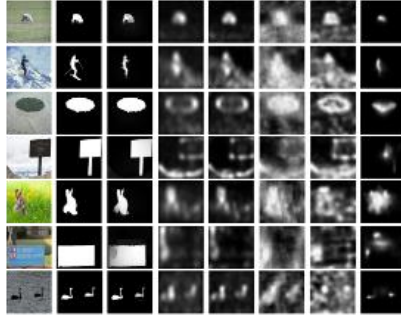


Fig .7. Visual effect map. From left to right: input image, GT, Ours, SS, SR, SIM, SER, FES.

4 Experimental Results and Analysis

We test on ASD, ECSSD and DUT-OMRON datasets. ASD and ECSSD contain 1000 images respectively, the DUT-OMRON dataset contains 5168 images and salient object maps that artificially annotated. Moreover, we compare our method with 5 classic salient detection algorithms. They are SER [17], SS [18], SR [10], SIM [14], FES [19].

4.1 Visually Compare

In Fig 7, we have selected some representative experimental results from our algorithm and comparison algorithm.

In the first map in Fig 7, the salient region is very small, and the background region is uniform. Our algorithm result is very close to the GT. There are too many backgrounds are revealed in the SS, SR, SIM, and SER result maps. In the second map, the black object is in sharp contrast with the white background. It can be seen from the map that our algorithm detects more salient object than others, and only a small part is not detected successfully, because we use super-pixel to pre-process image, this small object is segmented into the background, causing it to be viewed as background. In the third and fourth maps, the salient object detected by our algorithm is complete, the boundary is clear. In the fifth map, the background is complex. The salient object detected by our algorithm is complete and the boundary is clear. In the face of complex background, the comparison algorithm will also display the background regions. Although our algorithm will show it in the first stage of the merging, it will be merged with the background in the second stage, which improves the accuracy of our algorithm. The background of the sixth map is complex, with salient background objects and high detection difficulty. Therefore, in the detection results of SS, SR, SIM, and SER, most of the background is revealed, and the salient object is displayed inaccurate. Our algorithm detection results are similar to the GT map and superior to algorithms. The last map is multi-objective images. Therefore, our algorithm is ideal in visual effects than 5 comparison algorithms.

4.2 Objective Evaluations

We evaluate the performance using Precision-recall (P-R) curve and F-measure. The comparison result diagrams are shown as Fig 8.

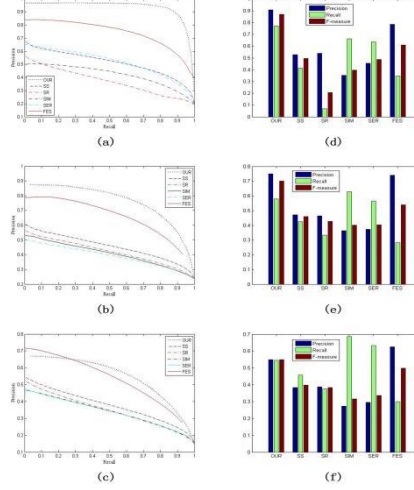


Fig .8. Contrast data diagram. The first column denotes recall and precision curves of different methods in ASD dataset, ECSSD dataset and DUT-OMRON dataset (from top to bottom). The second column denotes recall、 precision and F-measure value in ASD dataset, ECSSD dataset, DUT-OMRON dataset (from top to bottom).

From (a) and (b), we can see that the numerical line of our algorithm is above the numerical line than other algorithms. This shows that the performance of our algorithm is better than all comparison algorithms. In (d) and (e), the F-measure of our algorithm is higher. In the ECSSD dataset, although the background in the images is more complex, the accuracy and recall of algorithm are also close to 90%, which is much higher than the FES algorithm. The F-measure of the six kinds of algorithms are 0.87, 0.49, 0.20, 0.39, 0.48, and 0.60, respectively. That is, the F-measure of our algorithm is higher than other algorithms. The four data diagrams show that the detection results of our method are more accurate. The images in DUT-OMRON dataset are complex. Many salient objects are small, the background is complex, or there are more than one salient objects, so all algorithms are less effective than 70%. In (c), the curve of our algorithm is higher than other algorithms, but our algorithm curve is lower than the FES algorithm before 0.27. The main reason is that the salient objects in the image are very small and can easily be merged into the background area, which leads to the reduction of the accuracy of our algorithm. The F-measure are 0.55, 0.39, 0.38, 0.31, 0.33, and 0.49 respectively, of which the highest is 0.55 of our algorithm and higher than the FES algorithm which is 0.49. That is to say, our algorithm is superior to the FES algorithm.

In summary, our algorithm can reveal the salient object completely and showing good results.

5 Conclusion

In this paper, a salient object detection model for region merging is proposed. The initial saliency map is obtained by merging the background map, and in order to obtain the final saliency map, effective regions merging model is proposed to

optimize the rough initial saliency map. This model makes the optimized saliency map more accurate. Experimental results show that our algorithm is effective and more accurate. In the future work, better merger strategies, such as Bayesian mergers, will be considered during the merge phase.

References

1. Cheng M M, Mitra N J, Huang X, et al., "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 37, no. 3, pp. 569-582, 2015.
2. L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing.*, vol. 13, no. 10, pp. 1304-1318, 2004.
3. Wang X F, Qi C, "A behavior recognition method using salient object detection," *Journal of Xi'an Jiaotong University.*, 2018.
4. Li R, Li J P, Song C, "Research on co-segmentation of image based on salient object detection," *Modern Computer.*, no. 16, pp. 19-23, 2017.
5. Sun F, Qing K H, Sun W, et al., "Image saliency detection based on region merging," *Journal of Computer Aided Design and Graphics.*, vol. 28, no. 10, pp. 1679-1687, 2016.
6. Shen X, Wu Y, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition, IEEE*, 2012, pp. 853-860.
7. Peng H, Li B, Ling H, et al., "Salient object detection via structured matrix decomposition," *IEEE transactions on pattern analysis and machine intelligence.*, vol. 39, no. 4, pp. 818-832, 2017.
8. Zhang D, Fu H, Han J, et al., "A Review of Co-Saliency Detection Algorithms: Fundamentals, Applications, and Challenges," *ACM Transactions on Intelligent Systems and Technology (TIST).*, vol. 9, no. 4, pp. 38, 2018.
9. Yazdi M, Bouwmans T, "New trends on moving object detection in video images captured by a moving camera: A survey," *Computer Science Review.*, no. 28, pp. 157-177, 2018.
10. Hou X, Zhang L, "Saliency detection: A spectral residual approach," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
11. Guo C, Ma Q, Zhang L, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *Proceedings of IEEE Conferencen on Computer Vision and Pattern Recognition.*, pp.1-8, 2008.
12. Achanta R, Hemami S, Estrada F, et al., "Frequency-tuned salient region detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597-1604.
13. Zhang Q, Lin J, Li W, et al., "Salient object detection via compactness and objectness cues," *The Visual Computer.*, vol. 34, no. 4, pp. 473-489, 2018.
14. Murray N, Vanrell M, Otazu X, et al., "Saliency estimation using a non-parametric low-level vision model," in *Computer vision and pattern recognition, IEEE conference on Computer vision and pattern recognition*, 2011, pp. 433-440.
15. Achanta R, Shaji A, Smith K, et al., "Slic superpixels," *EPF, L Tech. Rep* 149-300, Nov. 2010.
16. Rahtu E, Kannala J, Salo M, "Segmenting salient objects from images and videos," in *European Conference on Computer Vision*, 2010, pp. 366-379.
17. HJ Seo, P Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of Vision.*, vol. 9, no. 12, pp. 1-27, 2009.

18. X Hou, J Harel, C Koch, "Image Signature: Highlighting sparse salient regions," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 34, no. 1, pp. 194-201, 2012.
19. HR Tavakoli, E Rahtu, J Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in Scandinavian Conference on Image Analysis, 2011, pp. 666-675.
20. Feng L , Wen P , Miao Y , et al. An image saliency detection algorithm based on color and space information[C]// 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). IEEE, 2017.
21. Qingzhen X , Fengyun W , Yongyi G , et al. An Edge-oriented Framework for Saliency Detection[C]// 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, 2017.
22. Li G , Yu Y . Visual Saliency Detection Based on Multiscale Deep CNN Features[M]. IEEE Press, 2016.
23. Li H , Chen J , Lu H , et al. CNN for saliency detection with low-level feature integration[J]. Neurocomputing, 2017, 226:212-220.
24. 赵恒, 安维胜, 田怀文. 结合稀疏重构与能量方程优化的显著性检测[J]. 计算机应用研究, 2019(6).
25. 余映, 吴青龙, 邵凯旋, et al. 超复数域小波变换的显著性检测[J]. 电子与信息学报, 2019, 41(9).
26. Yingchun Guo, Yi Liu, Runxin Ma. Image saliency detection based on geodesic-like and boundary contrast maps[J]. Etri Journal, 2019(1).
27. Marlon Marcon, Riccardo Spezialetti, Samuele Salti. Boosting Object Recognition in Point Clouds by Saliency Detection[M]// New Trends in Image Analysis and Processing – ICIAP 2019. 2019.