

Salient Regions Detection using Background Superpixels

Shima Foolad
Electrical & Computer Engineering Faculty
Semnan University
Semnan, Iran
Email: sh.foolad@semnan.ac.ir

Ali Maleki
Biomedical Engineering Department
Semnan University
Semnan, Iran
Email: a.maleki@semnan.ac.ir

Abstract— Detection of the most interesting region in an image has become an important subject in computer vision. Bottom-up model detects regions that differ with respect to their surrounding ones. These regions are known as salient regions. In this paper, we propose a new bottom-up model for saliency detection using the color of background regions. In the model, first, the image is segmented into superpixels. The boundary superpixels of the image are considered as background and others as uncertain superpixels. Then, the saliency is determined based on color difference between each uncertain superpixel and all background superpixels in the CIE LAB space. The proposed method can highlight the whole object regions uniformly and suppress the background regions effectively. Experimental results on the MSRA-1000 dataset demonstrate our method performs well against the state-of-the-art methods in terms of speed and accuracy.

Keywords- Boundary superpixels, bottom-up visual attention; background color feature; saliency detection.

I. INTRODUCTION

Human visual system is able to quickly identify most interest regions in a scene. The first area of an image which attracts the attention, the most salient region is called. How to identify this region has become essential for machine vision tasks, and it has many applications, such as image segmentation [1], object recognition [2], image retrieval [3] and image compression [4]. Therefore, different models have been proposed for saliency detection over the past few years.

Existing models can be classified as either bottom-up [5]–[17] or top-down [18], [19]. Bottom-up methods estimate saliency using low-level image features (e.g. color, orientation and luminance). Usually, a different region respect to its neighbors in an image will attract more attention. On the contrary, top-down approaches employ high-level features (e.g. face). The methods often require the prior knowledge of the stimuli and the learning process.

Many researchers utilize the bottom-up methods for saliency detection. Itti [5] proposed a biologically inspired model based on color, intensity and orientation information. This model only introduces local features and not considers the object scale. Ma and Zhang [6] estimate the saliency value

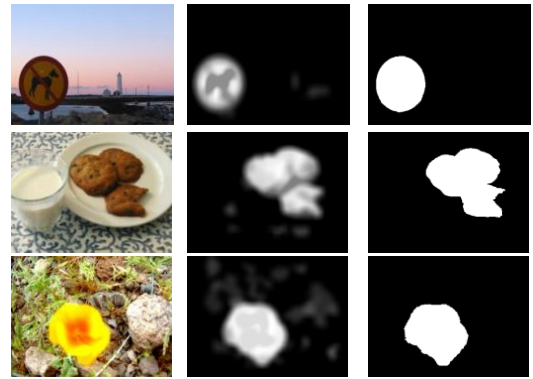


Figure 1. Original images; their saliency maps using the proposed method; and ground truth

using a fuzzy growing method. The method only highlights the object boundary regardless of its inside. Achanta [7] used center-surround feature distances to extract the salient regions. Some methods [8], [10] estimated saliency in the frequency domain that often they are not appropriate for natural images with high color complexity. Also methods such as İmamoğlu [13] extracted low-level features by wavelet transform domain. Some approaches [11], [16] applied the instance of biological evidence that shows objects near the image center are more attractive to human. The approaches often cannot identify the salient objects away from the image center.

Most mentioned methods do not detect salient regions correctly and more highlight high-contrast edges. Furthermore, the methods often estimate saliency using the center-surround contrast. Here, we propose a bottom-up saliency detection model based on the color of the background regions (See Fig. 1). In the model, a given image is segmented into superpixels. For each superpixel, its color mean is computed in the CIE LAB space. The boundary superpixels of the image are considered as background and the others as uncertain superpixels. The color distance between each uncertain superpixel with entire the background superpixels is computed, and its minimum distance is obtained. The saliency value of each uncertain superpixel is defined according to its minimum distance, and salient regions are detected.

The remainder of this paper is organized as follows. Section II describes our method for saliency detection. Section III reports the experimental results and Section IV concludes the paper.

II. PROPOSED METHOD

We propose a saliency detection model that its flowchart is illustrated in Fig. 2. In the proposed method, a given image is segmented into N (here, $N = 200$) superpixels using the Simple Linear Iterative Clustering (SLIC) algorithm [20]. A superpixel contains similar pixels and preserves the structural information of a salient object. The superpixels in the four sides of the image (image boundary) are considered as background. They are referred as “background superpixels” and others as “uncertain superpixels.”

The given image converts from RGB space to CIE LAB. The CIE LAB color space is similar to human visual perception, and preserves luminance as well as color information simultaneously. Due to its high uniformity with human color perception, the CIE LAB space is a good choice to determine the color difference. For each superpixel, its color mean is computed in the CIE LAB space. Then, the distance of each uncertain superpixel with the background superpixel is calculated as follows:

$$dist(i, j) = \|c_i - c_j\| \quad (1)$$

Where c_i and c_j are the color means of the pixels within uncertain superpixel i and the background superpixel j respectively, $\|\cdot\|$ is the Euclidean distance. For each uncertain superpixel (i), its minimum distance with entire background superpixels is computed as follows:

$$M(i) = \min_j dist(i, j) \quad j = 1, 2, \dots, L \quad (2)$$

Where, $dist(i, j)$ is the color distance computed in (1), and L is the number of background superpixels. If the color distance between an uncertain superpixel with entire the background superpixels is the high, it means that this superpixel is not similar to any the background superpixel, and it is considered as a salient region. Using of an adaptive threshold, the mean of minimum distances (M) in (2) is determined as a threshold (T).

$$T = \text{mean}(M(i)) \quad i = 1, 2, \dots, K \quad (3)$$

Where K is the number of the uncertain superpixels. The mean of each uncertain superpixel that is bigger than the threshold, is determined as the salient region, and saliency map is created based on it.

$$S^*(i) = \begin{cases} M(i) & \text{if } M(i) > T \\ 0 & \text{if } M(i) \leq T \end{cases} \quad i = 1, 2, \dots, K \quad (4)$$

Where T is the adaptive threshold, M is minimum distances matrix, K is the number of uncertain superpixels, and S^* is a saliency map. High value of M indicates more salient of the superpixel. For the background superpixels, saliency values are

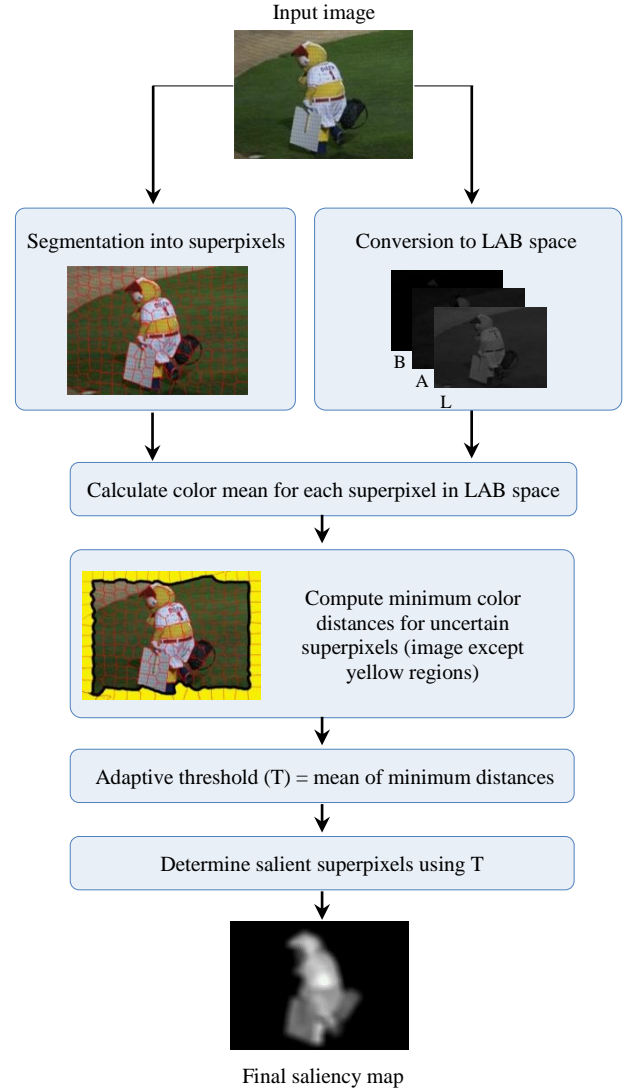


Figure 2. The flowchart of the proposed method

zero. In last, we convolve the saliency map (S^*) with a Gaussian filter ($I_{20 \times 20}$) for smoothing and noise removal.

$$S = S^* * I_{20 \times 20} \quad (5)$$

An example of generation process of the saliency map is shown in Fig. 3.

III. EXPERIMENTAL RESULTS

A. Dataset

We compare our method with nine the state-of-the-art saliency detection methods on MSRA-1000 dataset [21] comprising 1000 images with binary ground truth.

B. Evaluation Metric

Given a saliency map with values in the range $[0, 1]$, its binary saliency map is created with a fixed threshold $T \in [0, 1]$. To generate a precision-recall curve, by varying T from 0 to 1, several binary saliency maps are determined, and different

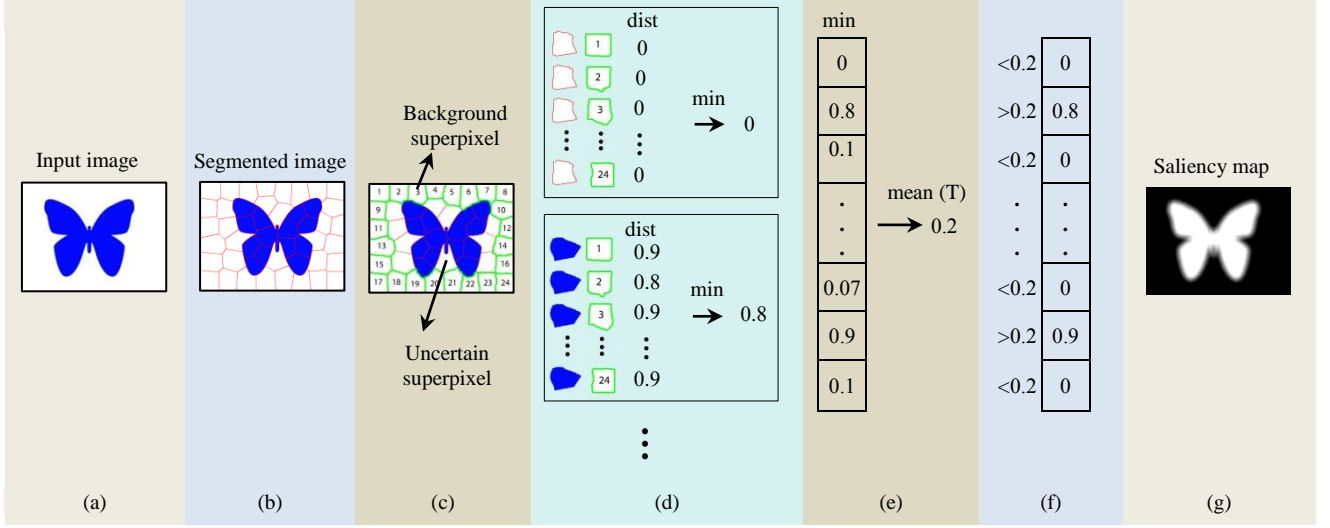


Figure 3. Example of generation process of saliency map, (a) input image, (b) the image is segmented into superpixels, (c) the type of the superpixels (background (green) or uncertain (red)) are determined, (d) for each uncertain superpixel, minimum distance is computed, (e) threshold (T) is obtained using the mean of the minimum distances, (f) saliency value is calculated for each uncertain superpixel using threshold and (g) saliency map is created (Obviously, saliency value is zero for the background superpixels).

precision-recall pairs are created. The average precision-recall curve is generated by averaging the results from all the images. Furthermore, we use adaptive threshold to generate a binary saliency map for each image. Adaptive threshold is two times the mean saliency from the binary saliency map. By averaging on entire the images, F-measure is computed as follows:

$$F - measure = \frac{(1+\beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (6)$$

Where precision is the ratio of salient pixels that are detected correctly to all the pixels of extracted regions in the saliency map, recall is the ratio of salient pixels that are detected correctly to the ground-truth, and Similar to [8, 11], $\beta^2 = 0.3$ in our experiments. Average F-measure is achieved by averaging the results from all images.

C. Comparisons

We compare our method with other nine state-of-the-art methods included CS [12], SDSP [11], WV [13], FT [8], AC [7], SR [10], GB [9], MZ [6] and IT [5] on MSRA-1000 dataset. Fig. 5 shows a visual comparison from the saliency maps of the methods. As seen in Fig. 5, our method creates saliency maps close to the ground truth.

In our experimental, we evaluate the methods with precision-recall curve and F-measure. Precision-recall curves and also, precision, recall and F-measure using an adaptive threshold for the state-of-the-art methods on MSRA-1000 dataset are illustrated in Fig. 4. Furthermore, average F-measure for each saliency detection method is listed in Table I. We note that the proposed method generates more discriminative saliency maps with higher precision and recall than the other nine methods.

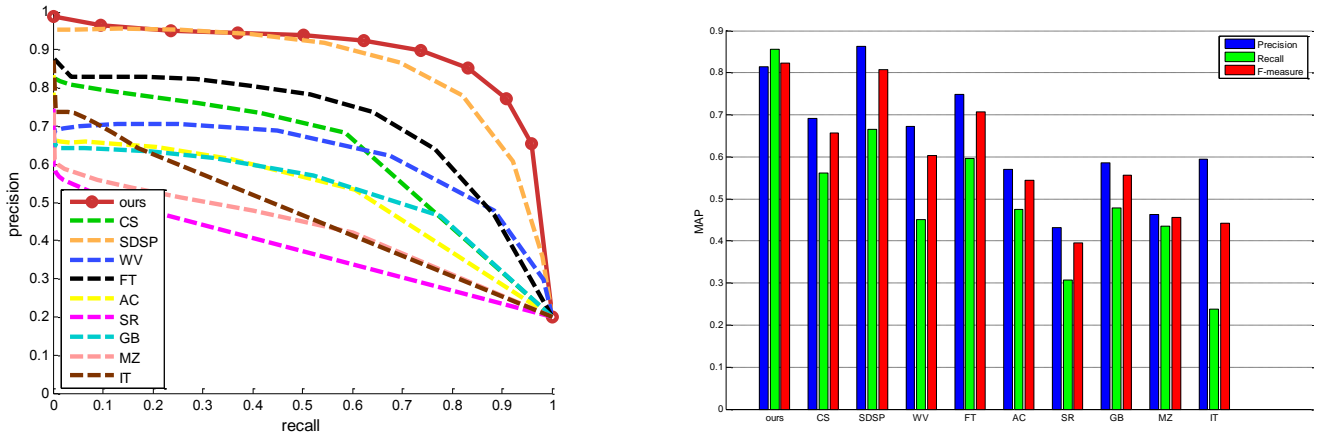


Figure 4. Left: precision-recall curves of different methods. Right: precision, recall and F-measure using an adaptive threshold. All results are computed on the MSRA-1000 dataset for comparison our method (ours) with methods included CS [12], SDSP [11], WV [13], FT [8], AC [7], SR [10], GB [9], MZ [6], IT [5].

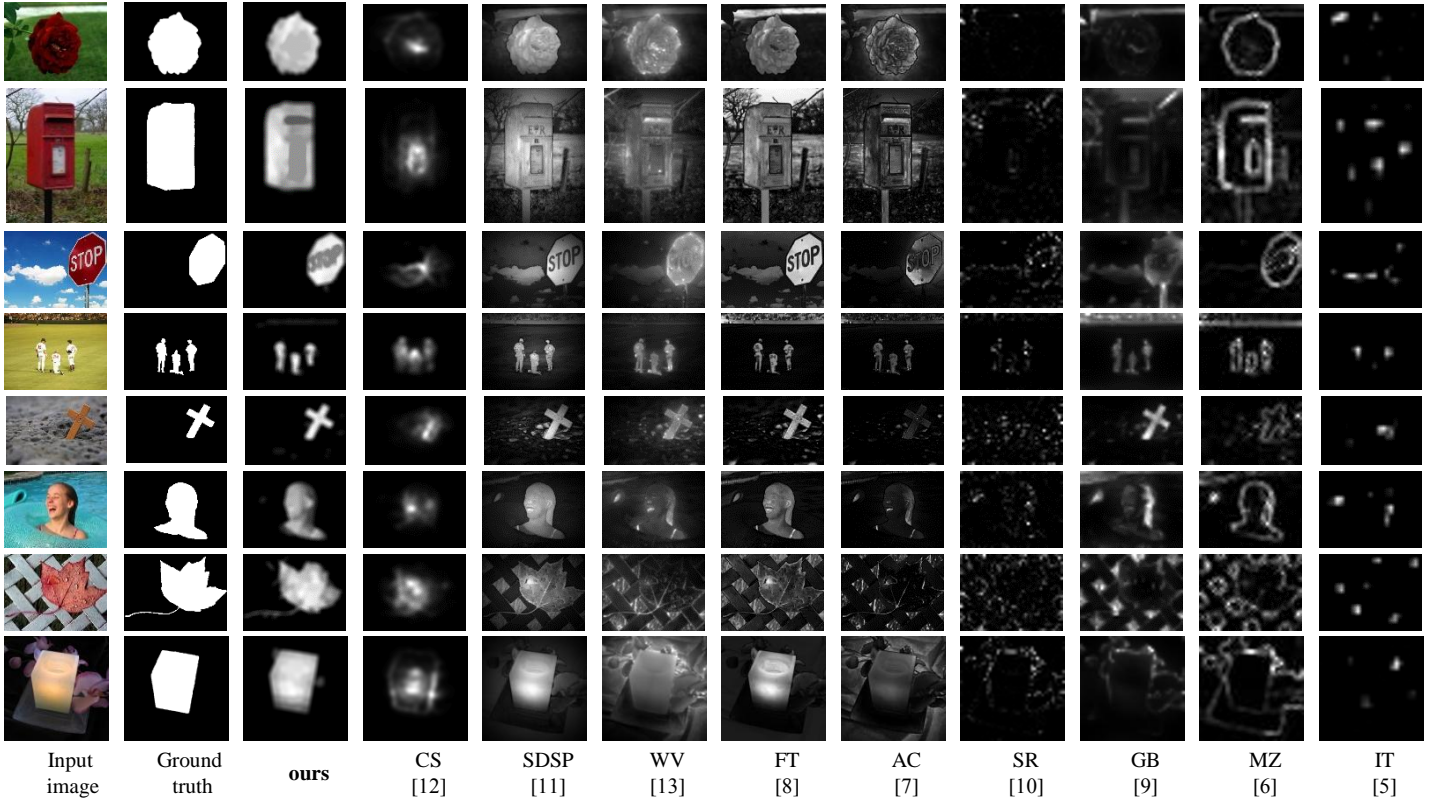


Figure 5. Visual comparison of saliency maps. Our method (ours) generates saliency maps close to the ground truth compared to the state-of-the-art methods.

TABLE I. F-MEASURE FOR EACH METHOD

Method	F-Measure
ours	0.8231
CS [12]	0.6564
SDSP [11]	0.8080
WV [13]	0.3188
FT [8]	0.7072
AC [7]	0.5444
SR [10]	0.3954
GB [9]	0.5569
MZ [6]	0.4561
IT [5]	0.4426

TABLE II. COMPARISON OF AVERAGE RUN TIME

Method	Time (s)	Code
ours	0.485	Matlab
CS [12]	21.223	Matlab
SDSP [11]	0.087	Matlab
WV [13]	6.777	Matlab
FT [8]	0.024	C++
AC [7]	0.127	C++
SR [10]	0.061	Matlab
GB [9]	1.563	Matlab
MZ [6]	0.07	C++
IT [5]	0.411	Matlab

D. Run Time

We evaluate the execution time of our method with other methods. Experiments are performed using Intel Core i5 2.53 GHz processor and 4 GB memory. The average run times of the methods on the MSRA-1000 dataset are presented in Table II. Our run time is much faster than methods CS [12], WV [13] and GB [9].

IV. CONCLUSION

In this paper, we proposed a bottom-up method that uses the background superpixels color for detecting salient regions. The boundary superpixels of the image were considered as

background. The saliency of the other superpixels was determined based on color difference between each uncertain superpixel and all the background superpixels. The proposed method suppresses the background superpixels effectively, and can successfully detect the whole object superpixels uniformly, even when the object is not near the image center.

We have compared the results of our method with the state-of-the-art methods on the MSRA-1000 public dataset. The results indicate our method can achieve the best saliency detection accuracy against the other methods while it has a quite low computational complexity. Therefore, the method can be used for real-time applications.

REFERENCES

- [1] Y. Tian, J. Li, S. Yu, and T. Huang, "Learning Complementary Saliency Priors for Foreground Object Segmentation in Complex Scenes," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 153–170, 2015.
- [2] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition*, 2004. CVPR 2004., vol. 2, 2004.
- [3] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, 2012.
- [4] C. Christopoulos, "The jpeg2000 still image coding system: an overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, 2000.
- [5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [6] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *Proc. Elev. ACM Int. Conf. Multimed. Multimed.* 03, vol. 102, pp. 374–381, 2003.
- [7] R. Achanta, F. Estrada, P. Wils, and S. Sabine, "Salient Region Detection and Segmentation," *Science (80-.)*, vol. 5008, no. 2008, pp. 66–75, 2008.
- [8] R. Achanta, S. Hemamiz, F. Estraday, and S. Süssstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, 2009, pp. 1597–1604.
- [9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Adv. Neural Inf. Process. Syst.*, pp. 545–552, 2006.
- [10] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.
- [11] L. Zhang, Z. Gu, and H. Li, "SDSP: A Novel Saliency Detection Method by Combining Simple Priors," in *20th IEEE international conference on image processing (ICIP)*, 2013, pp. 171–175.
- [12] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vis.*, vol. 13, no. 4, p. 11, 2013.
- [13] N. İmamoğlu, W. Lin, S. Member, Y. Fang, and N. İmamoğlu, "A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform," *IEEE Trans. Multimed.*, vol. 15, no. 1, pp. 96–105, 2013.
- [14] Y. Xie, H. Lu, and M. Yang, "Bayesian Saliency via Low and Mid Level Cues," *IEEE Transactions on Image Processing*, pp. 1–1, 2012.
- [15] 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, 2013.
- [16] H. Wang, C. Zhang, and X. Wang, "Real-Time Visual Saliency Detection Using Gaussian Distribution," in *2014 International Conference on Medical Biometrics*, 2014, pp. 200–203.
- [17] Q. Zhou, "Object-based attention: saliency detection using contrast via background prototypes," *Electron. Lett.*, vol. 50, no. 14, pp. 997–999, 2014.
- [18] S. Zhang, M. Xu, T. Kamigaki, J. P. Hoang Do, W. C. Chang, S. Jenvay, K. Miyamichi, L. Luo, and Y. Dan, "Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing," *Science (80-.)*, vol. 345, no. 6197, pp. 660–665, 2014.
- [19] A. Borji, D. N. Sihite, and L. Itti, "What/Where to Look Next? Modeling Top-Down Visual Attention in Complex Interactive Environments," *IEEE Trans. Syst. man Cybern. Syst.*, vol. 44, pp. 523–238, 2014.
- [20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels," ... *Fédéral Lausssanne* ..., no. June, p. 15, 2010.
- [21] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, 2011.