Salient Object Detection: A Discriminative Regional Feature Integration Approach: A Report

Paper's Writers

Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, Shipeng Li

Advisor

Dr. Maryam Abedi

Student

Mohammad Shahpouri

September
2022

Contents

List of Figures

List of Tables

1 Image Saliency Computation

Their approach has three main steps: Multi-level segmentation, Region saliency computation, multi-level saliency fusion.

Multi-level segmentation. An image I is segmented to M-level segmentations $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_M\}$. the graph-based image segmentation approach [1] and compute the over-segmentation $\mathcal{S}_1 = \{R_1^1, R_2^1, \ldots, R_{K_1}^1\}$. Other segmentations are computed based on S_1. Each region is represented by a weighted graph. In order to decreasing the weights of edges, neighboring regions are sequentially merged by computing the similarities of the regions until the weight of two regions is greater than the threshold that is defined in [1].

Region saliency computation. Three types of features are extracted from each region: regional contrast, regional property, and regional backgroundness, which will be described in Section 2. Features, vector $\mathbf{x}$, are passed into a random forest regressor f, to compute a saliency score. The learning procedure will be given in Section 3.

Multi-level saliency fusion. M saliency maps $\{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_M\}$ are generated. To get the final saliency map $A = g(\mathrm{A}_1, \ldots, \mathrm{A}_M)$ fuse them together. g is a combinator function introduced in section 3.

2 Regional features

2.1 Regional contrast descriptor

Each region has a feature vector, including color and texture features, denoted by v. The regional contrast descriptor of R is computed as the differences $\mathrm{diff}(\mathbf{v}^R, \mathbf{v}^N)$ between its features and the neighborhood features. The difference of the histogram feature is computed by χ^2, and the differences of other features are computed as the absolute elements differences of the vectors. As a result, it yields a 26-dimensional feature vector. The details of the regional contrast descriptor are given in Table 1.

Table 1. Color and texture features describing the visual characteristics of a region which are used to compute the regional feature vector. $d(\mathbf{x}_1, \mathbf{x}_2) = (|x_{11} - x_{21}|, \ldots, |x_{1d} - x_{2d}|)$ where the d is the number of elements in the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. $\chi^2(\mathbf{h}_1, \mathbf{h}_2) = \sum_{i=1}^{b} \frac{2(h_{1i} - h_{2i})^2}{h_{1i} + h_{2i}}$ with b being the number of histogram bins.The last two columns denote the symbols for regional contrast and backgroundness descriptors. (In the definition column, S corresponds to N for the regional contrast descriptor and B for the regional backgroundness descriptor, respectively.)

Color and texture features			Differences of features		Contrast	Backgroundness
	features	dim	definition	dim		
$\mathbf{a}_1$	the average RGB values	3	$d(\mathbf{a}_1^R, \mathbf{a}_1^S)$	3	$c_1 \sim c_3$	$b_1 \sim b_3$
$\mathbf{a}_2$	the average L*a*b* values	3	$d(\mathbf{a}_2^R, \mathbf{a}_2^S)$	3	$c_4 \sim c_6$	$b_4 \sim b_6$
$\mathbf{r}$	the absolute response of LM filters	15	$d(\mathbf{r}^R, \mathbf{r}^S)$	15	$c_7 \sim c_{21}$	$b_7 \sim b_{21}$
r	the max response among the LM filters	1	$d(r^R, r^S)$	1	c_{22}	b_{22}
$\mathbf{h}_1$	the L*a*b* histogram	$8 \times 16 \times 16$	$\chi^2(\mathbf{h}_1^R, \mathbf{h}_1^S)$	1	c_{23}	b_{23}
$\mathbf{h}_2$	the hue histogram	8	$\chi^2(\mathbf{h}_2^R, \mathbf{h}_2^S)$	1	c_{24}	b_{24}
$\mathbf{h}_3$	the saturation histogram	8	$\chi^2(\mathbf{h}_3^R, \mathbf{h}_3^S)$	1	c_{25}	b_{25}
$\mathbf{h}_4$	the texton histogram	65	$\chi^2(\mathbf{h}_4^R, \mathbf{h}_4^S)$	1	c_{26}	b_{26}

2.2 Regional property descriptor

Properties of a region, including appearance and geometric features, are extracted, using feature extraction algorithm in [2]. The appearance features are the distribution of colors and textures in a region. The geometric features are size and position of a region. A 34-dimensional regional property descriptor are the result of this feature extraction. The details are given in Table 2.

Table 2. The regional property descriptor.

description	notation	dim
the average normalized x coordinates	p_1	1
the average normalized y coordinates	p_2	1
the $10th$ percentile of the normalized x coordinates	p_3	1
the $10th$ percentile of the normalized y coordinates	p_4	1
the $90th$ percentile of the normalized x coordinates	p_5	1
the $90th$ percentile of the normalized y coordinates	p_6	1
the normalized perimeter	p_7	1
the aspect ratio of the bounding box	p_8	1
the variances of the RGB values	$p_9 \sim p_{11}$	3
the variances of the $L^*a^*b^*$ values	$p_{12} \sim p_{14}$	3
the variances of the HSV values	$p_{15} \sim p_{17}$	3
the variance of the response of the LM filters	$p_{18} \sim p_{32}$	15
the normalized area	p_{33}	1
the normalized area of the neighbor regions	p_{34}	1

2.3 Regional backgroundness descriptor

The pseudo-background region B, 15-pixel wide narrow border region of the image, is extracted and compute the backgroundness discriptor with the pseudo-background region as a reference. The backgroundness feature of the region R is then computed as the differences $\mathrm{diff}(\mathbf{v}^R, \mathbf{v}^B)$ between its features $\mathbf{v}^R$ and the features $\mathbf{v}^B$ of the pseudo-background region, resulting a 26-dimensional feature vector. See details in Table 1.

3 Learning

Learning the regional saliency regressor. A set of training examples are applied to learn a regional saliency estimator. The training examples consist of a set of confident regions $\mathcal{R} = \{R_1, R_2, \dots, R_Q\}$ and the corresponding saliency scores $\mathcal{A} = \{a_1, a_2, \dots, a_Q\}$. A region is considered to be confident if the number of the pixels belonging to the salient object or the background exceeds 80% of the number of the pixels in the region. A random forest regressor f is learnt from the training data $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Q\}$ and the saliency scores $\mathcal{A} = \{a_1, a_2, \dots, a_Q\}$.

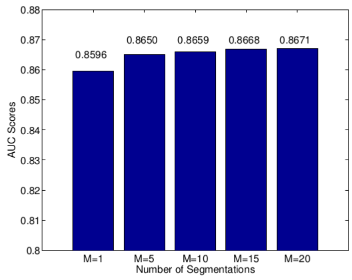
Learning the multi-level saliency fusor. To fuse the multi-level saliency maps to form the final saliency map $\mathbf{A}$, a combinator $g(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M,)$ is learnt. The conditional random field [3] has solved such a problem before. A linear combinator $\mathbf{A} = \sum_{m=1}^{M} w_m \mathbf{A}_m$ is applied to learn the weights using a least square estimator, i.e. , minimizing the the sum of the losses $(\|\mathbf{A} - \sum_{m=1}^{M} w_m \mathbf{A}_m\|_F^2)$ over all the training images.

4 Experimental results

4.1 Setup

MSRA-B[1], SED[2], SOD[3], and iCoSeg[4] are datasets which are utilized to evaluate the performance.

4.2 Empirical analysis

(a) 

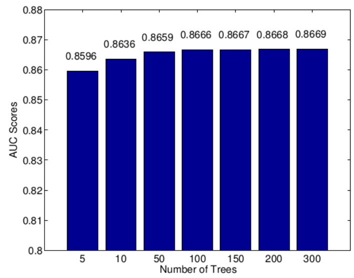
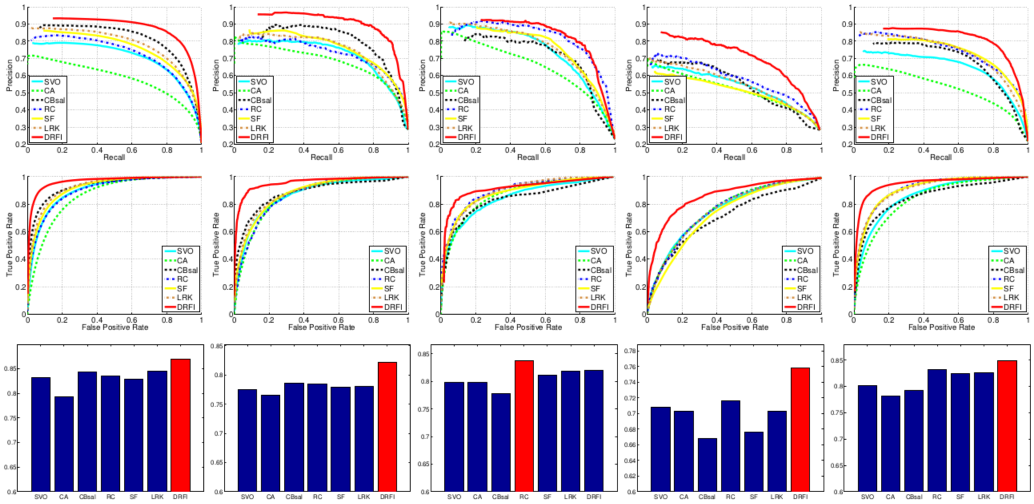
(b) 

Figure 1. The AUC scores of the saliency maps of the validation set of MSRA-B using (a) different number of segmentations and (b) different number of trees in the random forest regressor.

[1] http://research.microsoft.com/en-us/um/people/jiansun/%20SalientObject/salient_object.htm
[2] http://www.wisdom.weizmann.ac.il/%E2%88%BCvision/%20Seg_Evaluation_DB/index.html
[3] http://elderlab.yorku.ca/SOD/index.html
[4] http://chenlab.ece.cornell.edu/projects/touch-coseg

4.3 Performance comparison

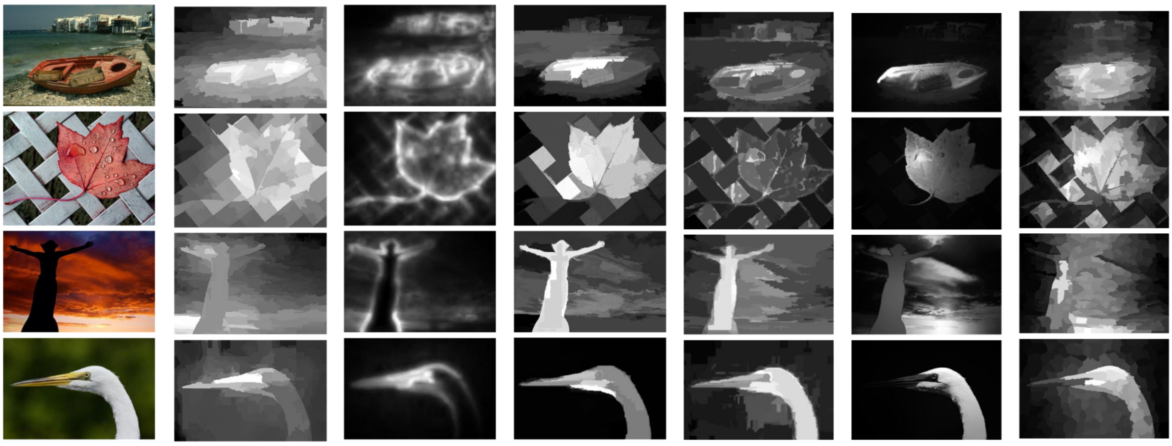
4.3.1 Quantitative comparison



(a) (b) (c) (d) (e)

Figure 2. Quantitative comparison of saliency maps produced by different approaches on different data sets. From left to right: (a) the MSRA-B data set, (b) the SED1 data set, (c) the SED2 data set, (d) the SOD data set, and (e) the iCoSeg data set. From top to bottom: the PR curves, the ROC curves, and the AUC scores.

4.3.2 Qualitative comparison



(a) input (b) SVO[4] (c) CA[5] (d) CBsal[6] (e) RC[7] (f) SF[8] (g) LRK[9]

(h) DRFI

Figure 3. Visual comparison of the saliency maps. Their method (DRFI) consistently generates better saliency maps.

References

[1] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, Sep 2004. 1

[2] D. Hoiem, A. Efros, and M. Hebert, "Geometric context from a single image," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1, pp. 654–661 Vol. 1, 2005. 2

[3] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011. 3

[4] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *2011 International Conference on Computer Vision*, pp. 914–921, 2011. 5

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2376–2383, 2010. 5

[6] Z. Y. Huaizu Jiang, Jingdong Wang, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proceedings of the British Machine Vision Conference*, pp. 110.1–110.12, BMVA Press, 2011. http://dx.doi.org/10.5244/C.25.110. 5

[7] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *CVPR 2011*, pp. 409–416, 2011. 5

[8] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–740, 2012. 5

[9] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853–860, 2012. 5