LABEL PROPAGATION BASED SALIENCY DETECTION VIA GRAPH DESIGN

Tianhao Zhang[1], *Yuan Zhou*[1,2], *Shuwei Huo*[1] *and Chunping Hou*[1]

[1] School of Electrical and Information Engineering, Tianjin University, Tianjin, China
[2] Electrical Engineering Department, Princeton University, Princeton, USA
Email: zhouyuan@tju.edu.cn

ABSTRACT

Saliency detection has been widely used as the pre-processing of the computer-vision tasks. Existing propagation based saliency detection methods simply select a $k-$regular graph for saliency propagation, which usually leads to the mistaken highlighting of the long-range smooth background regions. In this paper, we design a novel graph for label propagation based saliency detection by considering the local consistency and the global symmetry of the image scene and updating the graph model based on smoothness assumption and cluster assumption. Then, we label the reliable seeds and propagate the saliency value through the designed graph. On two widely used large open benchmark data sets, the proposed method significantly outperforms thirteen state-of-the-arts under either quantitative or qualitative evaluation.

Index Terms— Saliency detection, Label propagation, Graph construction

1. INTRODUCTION

Visual saliency detection aims at separating the most eye-catching region from the image scene. In order to process the image data efficiently, the saliency detection methods have been widely applied as the pre-processing of numerous computer vision tasks, such as visual question answering [1], image retrieval [2,3], visual tracking [4], scene classification [5], and so on.

The methods of the visual saliency detection can be divided into two categories. One is the top-down method which is task driven. The other one is the bottom-up method and is data driven. In this work, we focus on the method of bottom-up. In recent years, some of the complex deep learning methods perform well on the task. However, as no vast of hand-labeled images nor training time are required, the bottom-up visual saliency detection is still a popular research topic.

According to the research of the bottom-up saliency detection, the region that can be significantly distinguished from its surrounding regions will be salient. And the contrast is always used to evaluate the difference between regions. Depending on the size of the surrounding regions, the contrast can be classified as local [6] and global contrast [2]. And the graph based saliency propagation method [7–9] has been proposed to combine the local and global contrast.

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

Fig. 1: The performance of existing propagation based methods when facing long rang smooth background regions (a) image with long range smooth background regions near the center (b) AMC [7] (c) GMR [8] (d) MAP [9] (e) Ours.

When using the graph based saliency propagation method, the regions in an image scene are always mapped to the nodes in a graph model. And the theory of propagation is based on measuring the similarity between the nodes through the graph. Most of the existing propagation based saliency detection methods simply select a $k-$regular graph as the propagation graph. The $k-$regular graph is very useful to exploit the spatial relationship and the feature similarity of the nodes when applying the graph based propagation method into a new domain with the domain knowledge is limited. However, as the researchers have discovered sufficient prior knowledge for the domain of visual saliency detection, we should add the prior knowledge into the graph model, thereby obtaining a better result.

In this work, based on the prior knowledge of saliency detection, we propose a novel graph model for propagation based saliency detection. And the main contributions of our work are as follows: firstly, we consider the local consistency and the global symmetry of the image scene to construct the initial graph model and update the graph based on the smoothness assumption and the cluster assumption; secondly, we proposed a new model to measure the confidence value of the super-pixels to label the reliable seeds. Figure 1 shows that our method can overcome the backward, that long-range smooth regions especially near the middle part of the image are often highlighted mistakenly, which usually appears when using the graph based saliency propagation methods.

The remainder paper is organized as follows: section 2. describes the details of our method, section 3. displays the

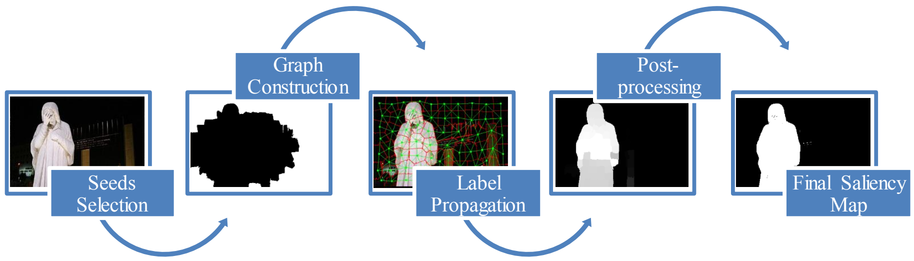


Fig. 2: Pipeline of the proposed method.

experiment results , and section 4. concludes the whole paper.

2. THE PROPOSED METHOD

Figure 2 shows the pipeline of our method. Firstly, we use the SLIC algorithm to segment the input image into N super-pixels. And some of the super-pixels are labeled as the seeds. Then, based on the prior knowledge of the domain of visual saliency detection, a novel graph model for propagation is constructed. After that, the saliency value is propagated from the labeled seeds to all the other nodes through the graph by using a concise propagation algorithm. Finally, we post-process the saliency map and get the final saliency map.

2.1. Background Seeds Selection

The selection of reliable seeds is essential for saliency detection. We observe that the size of the background regions is usually significantly larger than the salient region in an image scene. Therefore, instead of directly identifying the salient region, it is easier to distinguish some parts of the background regions. And the salient region can be detected indirectly.

To select the background seeds, firstly, we notate the N super-pixels as $S = \{s_1, s_2, \dots, s_N\}$. For each super-pixel s_i, we define $s_i = (s_i^{color}, s_i^{position})$, in which s_i^{color} indicates the average value of the CIE LAB color feature of all pixels in s_i and $s_i^{position}$ indicates the average value of the coordinates. Then, we compute an initial saliency map based on the spatially weighted contrast prior and the object prior.

The spatially weighted contrast prior has been widely used. For super-pixel s_i, the contrast to all the other super-pixels is defined as:

$$I(s_i) = \sum_{j=1}^{N} \exp\left(-\frac{\|s_i^{position} - s_j^{position}\|}{2\sigma_s^2}\right) \cdot \|s_i^{color} - s_j^{color}\| \quad (1)$$

With the appropriate control parameter σ_s^2, the exponential function can successfully distinguish the different contributions of the super-pixels at different positions, even if the super-pixels have the same color feature.

The object prior is always used to enlarge the contrast between the regions with and without highlighted in a saliency map. And just as being conducted in [10], the object prior is conducted on $I(s_i)$ for an efficient selection of the background seeds. And we obtain the initial saliency map.

Though the foreground regions have been highlighted accurately, some of the background regions are highlighted simultaneously. So, different from the existing methods which select the regions with high saliency value as foreground seeds, we select the background seeds from the initial saliency map. We use the OTUS method to separate the regions with low saliency value and set an adaptive threshold according to the average saliency value of the separated regions. Then, the super-pixels with saliency value lower than the threshold are labeled as background seeds.

2.2. Graph Construction

As the saliency value will be propagated along the graph, it is critical to construct a graph model. Firstly, we construct an initial graph model by considering the local consistency and the global symmetry of the image scene. Then we update the graph model based on smoothness assumption and the cluster assumption which are widely used in the domain of the semi-supervised learning.

The graph model indicates the connectivity and the similarity between nodes. We define the undirected graph $G = (V, E)$, where V denotes the node set of the graph with super-pixels as nodes and E denotes the edge set with feature similarity between the super-pixels as the weight. Then, to display the connection between the pairwise super-pixels, we define the adjacency matrix as $A = [a_{ij}]_{N \times N}$, where $a_{ij} = 1$ if there is an edge between the two super-pixels s_i and s_j, otherwise $a_{ij} = 0$. Then, we define the affinity matrix as $W = [w_{ij}]_{N \times N}$, where $w_{ij} = \exp(-\frac{\|s_i^{color} - s_j^{color}\|}{2\sigma_w^2}) \cdot a_{ij}$, to evaluate the weight of the edges.

The initial graph model is constructed based on the local consistency and the global symmetry. The local consistency assumes that the super-pixels share the common boundary tend to share the similar saliency value. Therefore, similar to the existing propagation based methods, we construct a $k-$regular graph but with $k = 1$. The global symmetry assumes that the image scenes are generally symmetrical. So, we enforce any pair of the super-pixels that on the four edges of the image are connected with each other, which reduces the geodesic distance between nodes on the edges.

Though the $k-$regular graph can exploit the spatial relationship of the nodes, the sparsity and the homogeneity of the connection will increase the geodesic distance when facing the long-range smooth regions. So, we increase the ratio of the connection between the super-pixels in such regions. Based on the affinity matrix of the initial graph model, we pick out 60% of the edges with large weight. And some of the edges connect the super-pixels at both ends of each edge into lines. We observe that these lines have good connectivity in the long-range smooth regions with a similar performance as the geodesic line but without calculating all the geodesic lines between any two nodes. So, we respectively segment all the super-pixels on the different lines into the different clusters. Based on the cluster assumption which indicates that the nodes in the same cluster are likely to have the similar saliency value, we enforce the pairwise super-pixels contained in the same cluster to be connected with each other.

Super-pixels with different visual features may have the same saliency value. And the background seeds we have successfully labeled may have different features in color. However, as the label propagation theory is based on the smooth assumption which indicates that the super-pixels with similar color supposed to have the similar saliency value, the background seeds with different colors will have an effect with each other during the propagation. Therefore, we pick out the clusters that contain the background seeds and enforce the super-pixels between these clusters cannot be connected with each other.

2.3. Saliency Propagation

A concise and meaning clear label propagation algorithm has been proposed by [11] and been widely used in various research domains. With an affinity matrix $W = [w_{ij}]_{N\times N}$, a degree matrix can be defined as $D = diag\{d_{11}, d_{22}, \ldots, d_{NN}\}$, where $d_{ii} = \sum_{j=1}^{N} W_{ij}$. And a row normalized affinity matrix can be defined as:

$$P = D^{-1} \cdot W \tag{2}$$

The p_{ij} shows the probability of the saliency value to propagate from super-pixel s_i to s_j. After renumbering the order of super-pixels with the labeled super-pixels coming first, the matrix P can be expressed as:

$$P = \begin{pmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{pmatrix} \tag{3}$$

The propagation result is defined as $f = (f_L, f_U)^T$, where f_L means the results of the seeds and is fixed as "1", and f_U means the results of the other nodes and is initialized as "0". The label propagation algorithm can be defined as:

$$f_U^{t+1} = P_{UU} \cdot f_U^t + P_{UL} \cdot f_L \tag{4}$$

Based on the background selected seeds and the constructed graph, we implement the label propagation algorithm for 40 iterations empirically, and obtain the propagation results f^*. After the normalization of f^*, the saliency map obtained from the background propagation can be defined as:

$$S_B = 1 - f^*_{normalized} \tag{5}$$

In order to further highlight the salient region uniformly, similar as the propagation process of background seeds, we select the foreground seeds based on the saliency map S_B and propagate the saliency value after the construction of the novel propagation graph. And we obtain a saliency map S_F after the normalization of the saliency value.

2.4. Post-processing

To improve the application efficiency of the saliency map, firstly, we map the saliency value from the super-pixel level to the pixel level just like [12] to reduce the segmentation error caused by SLIC algorithm. Then, as the application usually requires the clearly separating between the salient region and the background, we apply a sigmoid function just as [13] to enlarge the contrast.

3. EXPERIMENTAL EVALUATION

We compare our method with thirteen state-of-the-art saliency detection methods. They are CA10 [6], SVO11 [14], SF12 [15], GC13 [16], AMC13 [7], GMR13 [8], RBD14 [17], RC15 [2], BSCA15 [18], MAP15 [9], HS16 [19], DSR16 [10], and MST16 [13]. Among them, AMC, GMR and MAP are representative graph based saliency propagation methods which are closely related to our method.

3.1. Evaluation Metrics

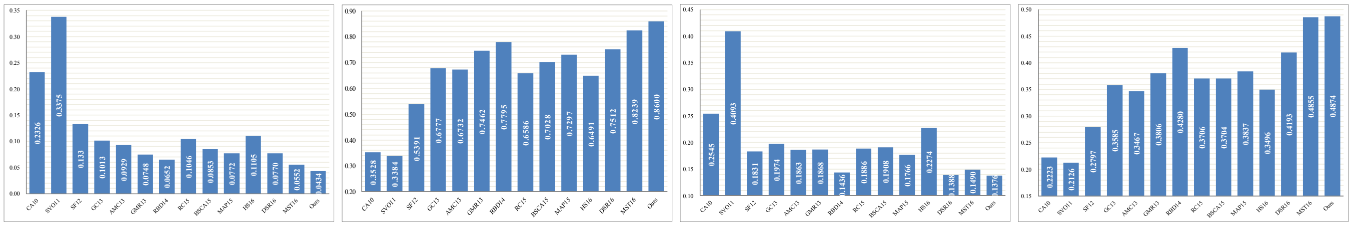
We apply weighted F_β^ω-measure and Mean Absolute Error (MAE) to evaluate our method and compare with others.

Weighted F_β^ω-measure: Previous works commonly use the Area Under the Curve (AUC) measure, the Average Precision (AP) measure and the F_β-measure to evaluate the performance of methods. However, [20] notated that the traditional evaluation methods have the limitation to reliably evaluate the performance. Therefore, they propose the weighted F_β^ω-measure which has been widely used in recent years. The weighted F_β^ω-measure is defined as:

$$F_\beta^\omega = (1+\beta^2)\frac{Precision^\omega \cdot Recall^\omega}{\beta^2 \cdot Precision^\omega + Recall^\omega} \tag{6}$$

The parameter β_2 is set to be 0.3. And the method with high value will be ranked as higher.

Mean Absolute Error (MAE): To consider the true negative assignments, we apply the MAE to evaluate the saliency



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

Fig. 3: Quantitative evaluation (a) ASD, MAE (b) ASD, F_β^ω-measure (c) DUT-OMRON, MAE (d) DUT-OMRON, F_β^ω-measure.

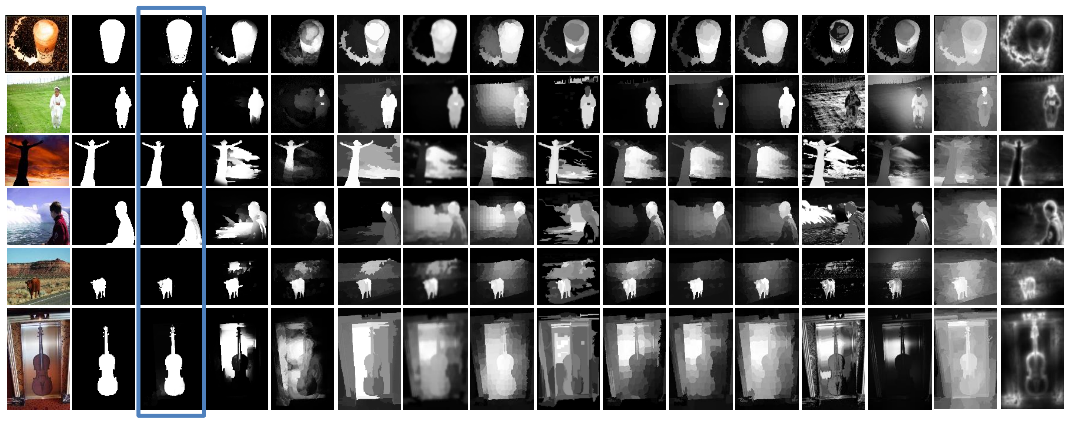


Fig. 4: Visual comparison with the state-of-the-arts. From left to right are input image, ground truth, Ours, MST16, DSR16, HS16, MAP15, BSCA15, RC15, RBD14, GMR13, AMC13, GC13, SF12, SVO14, and CA10.

map. And the method that suppresses the background uniformly while highlights the salient objects successfully will have lower MAE value.

3.2. Quantitative Evaluation

We evaluate the proposed method on two large open benchmark data sets. The first one is the ASD data set which contains 1000 images. Due to the ground truth masks are accurate and without ambiguous, the ASD has been the most widely used data set. The second one is the DUT-OMRON data set. It contains 5,168 complex images with different sizes of the salient region and complex background.

ASD: The value of MAE is showed in Figure 3(a). It shows that our method achieves the lowest value compared with all the other methods. Figure 3(b) shows the value of all evaluated methods under the weighted F_β^ω-measure. Compared with the other methods, our method is higher than all the other methods significantly, and achieves a 3.61% higher than the second best method which is a huge improvement.

DUT-OMRON: The MAE values and the weighted F_β^ω-measure of the evaluated methods are respectively showed in the Figure 3(c), (d), though the performance of all methods has declined, our method still outperform all the methods in the recent seven years under the two evaluation metrics.

Figure 4 lists the saliency maps generated by the evaluated saliency detection methods. It shows that the saliency maps generated by our method can highlight the salient regions while suppressing the background regions uniformly.

4. CONCLUSION

In this paper, we propose a novel graph based saliency propagation method. The consideration of the existing prior knowledge during the graph construction can solve the backward that the long-range smooth regions are usually highlighted mistakenly when using the graph based saliency propagation methods, and can suppress the background regions efficiently and highlight the salient regions uniformly. Besides, the seeds selection mechanism can successfully evaluate the confidence value of the super-pixels and label the seeds reliably. And the post-processing can improve the efficiency for the application. The proposed method outperforms the state-of-the-arts on the widely used open benchmark data sets.

5. ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (No. 61571326, 61471262, 61520106002) and National Natural Science Foundation of Tianjin (No. 16JCQNJC00900).

6. REFERENCES

[1] Kevin J. Shih, Saurabh Singh, and Derek Hoiem, "Where to look: Focus regions for visual question answering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4613–4621.

[2] Ming Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi Min Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[3] Andrea Manno-Kovacs, "Content based image retrieval using salient orientation histograms," in *IEEE International Conference on Image Processing*, 2016, pp. 2480–2484.

[4] Cong Ma, Zhenjiang Miao, and Xiao Ping Zhang, "Saliency prior context model for visual tracking," in *IEEE International Conference on Image Processing*, 2016, pp. 1724–1728.

[5] Mengshi Qi and Yunhong Wang, "Deep-cssr: Scene classification using category-specific salient region with deep features," in *IEEE International Conference on Image Processing*, 2016.

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

[7] Bowen Jiang, Lihe Zhang, Huchuan Lu, Chuan Yang, and Ming Hsuan Yang, "Saliency detection via absorbing markov chain," in *IEEE International Conference on Computer Vision*, 2013, pp. 1665–1672.

[8] Chuan Yang, Lihe Zhang, Huchuan Lu, Ruan Xiang, and Ming Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *Computer Vision and Pattern Recognition*, 2013, pp. 3166–3173.

[9] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on markov absorption probabilities.," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1639–1649, 2015.

[10] Huchuan Lu, Xiaohui Li, Lihe Zhang, and Ruan Xiang, "Dense and sparse reconstruction error based saliency descriptor," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 25, no. 4, pp. 1592, 2016.

[11] Xiaojin Zhu, Zoubin Ghahramani, and Tommi Jaakkola Mit, "Semi-supervised learning with graphs," in *International Joint Conference on Natural Language Processing*, 2005, pp. 2465 – 2472.

[12] H. Li, H. Lu, Z. Lin, X. Shen, and B Price, "Inner and inter label propagation: salient object detection in the wild.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 24, no. 10, pp. 3176–86, 2015.

[13] Wei-Chih Tu, Shengfeng He, Qingxiong Yang, and Shao-Yi Chien, "Real-time salient object detection with a minimum spanning tree," in *Computer Vision and Pattern Recognition*, 2016, pp. 2334–2342.

[14] Kai Yueh Chang, Tyng Luh Liu, Hwann Tzong Chen, and Shang Hong Lai, "Fusing generic objectness and visual saliency for salient object detection," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November*, 2011, pp. 914–921.

[15] Philipp Krahenbuhl, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition*, 2012, pp. 733–740.

[16] Ming Ming Cheng, Jonathan Warrell, Wen Yan Lin, and Shuai Zheng, "Efficient salient region detection with soft image abstraction," in *IEEE International Conference on Computer Vision*, 2013, pp. 1529–1536.

[17] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2814–2821.

[18] Yao Qin, Huchuan Lu, Yiqun Xu, and He Wang, "Saliency detection via cellular automata," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 110–119.

[19] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended cssd.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 717–729, 2016.

[20] Margolin Ran, Lihi Zelnikmanor, and Ayellet Tal, "How to evaluate foreground maps," in *CVPR*, 2014, pp. 248–255.