# A PRIOR-BASED GRAPH FOR SALIENT OBJECT DETECTION

*Jinxia Zhang*[⋆†] *Krista A. Ehinger*[†‡] *Jundi Ding*[⋆] *Jingyu Yang*[⋆]

[⋆] School of Computer Science and Engineering
Nanjing University of Science and Technology, Nanjing, China, 210094
[†] Brigham and Women's Hospital, Cambridge, MA, United States, 02139
[‡] Harvard Medical School, Cambridge, MA, United States, 02139

## ABSTRACT

Recently, various graph-based methods have be proposed for salient object detection. These algorithms represent image points and their similarity as nodes and edges in a graph. Although the edge structure and weighting are the heart of these methods, the graph construction has not been studied in detail. In this paper, we exploit image priors, including spatial priors, color priors, and a central bias prior, to construct the graph. We connect nodes which are spatially close in the image, nodes which have similar color features, and the boundary nodes along the borders of the image, while weighting edges according to both their color similarity and spatial proximity. Moreover, we propose a new sine spatial distance instead of the commonly-used Euclidean spatial distance, which better captures the central bias in scenes. Extensive experiments show that our method outperforms thirteen state-of-the-art methods on four different image databases.

***Index Terms***— Graph construction, Spatial prior, Color prior, Central bias prior, Salient object detection

## 1. INTRODUCTION

Since Itti et al. [1] implemented the first computational model of visual saliency, various methods have been proposed to detect salient objects in images. Graph-based algorithms, such as [2], [3], and [4], assume that nearby points with similar features are likely to have similar saliency values and model the image as a graph in which image regions are nodes and edges represent the similarity between regions. By propagating labels through the graph, these algorithms attempt to detect foreground objects while suppressing the background. Although the graph is the heart of this type of algorithm, the specifics of its construction have not been studied in much detail. In this paper, we exploit image priors to construct the graph for salient object detection, giving a performance boost over previous graph-based methods for salient object detection.

Previous methods, including [2], [3], and [4], use a spatial prior to construct the graph, so each node is connected to its neighbors in the image. This reflects the fact that spatially-neighboring pixels are likely to share similar saliency values. However, if the salient foreground object has a complex shape or pattern, or there are multiple objects, this approach may not uniformly highlight the whole foreground (see Figure 1 (B), which shows the result of [2]). Some other methods, such as [5], include a color prior, which assumes that image regions with similar colors are likely to belong to the same class (foreground or background). This can work well to detect all the salient objects which have similar colors, but when the background color is similar to the foreground, the color prior may promote false alarms. Moreover, when the foreground object is composed of regions with different colors, only part of the object will be detected

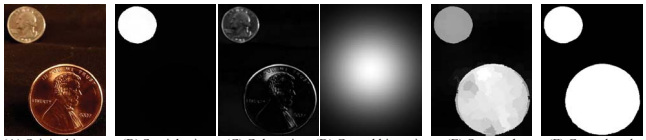


**Fig. 1**. Salient object detection results. (A) is the original image. (B), (C), and (D) are the results obtained when only the spatial prior, color prior, or central bias prior is considered. (E) is our result which uses all three priors. (F) is the ground truth.

(see Figure 1 (C), which shows the result of [5]). In addition to spatial and color priors, many models exploit the central bias in free scene viewing [6] and the related fact that photographers tend to frame the object of interest in the center of an image [7]. Hence, many saliency algorithms [7, 8] include an assumption that saliency is higher in the center of an image (Figure 1 (D)), which is beneficial when the foreground objects are near the image center but can cause false alarms when they are not.

Spatial priors, color priors, and the central bias prior all play important roles in detection of visually-salient objects, but they each have some drawbacks. Therefore, it is important to consider all priors and weight them appropriately. In this paper, we exploit all three priors to construct a suitable graph for salient object detection. We define the nodes of the graph to be the superpixels of the image. We connect nodes which are spatially close or have similar color features, as well as the boundary nodes along the borders of the image. Edges are weighted by both color and spatial proximity. Moreover, we propose a new sine spatial distance in place of the commonly-used Euclidean spatial distance, which exploits the central bias in natural images. A representative result from our method is shown in Figure 1(E). Compared to previous methods, our approach better captures the ground-truth salient object regions while suppressing the background (Figure 1(F)). We test our algorithm on four saliency databases and show that our method outperforms thirteen state-of-the-art models. The rest of this paper is organized as follows: The foundation of the graph-based manifold ranking algorithm is introduced in Section 2. Section 3 presents our prior-based graph for saliency detection. Section 4 describes our testing procedure and gives the comparison results.

## 2. GRAPH-BASED MANIFOLD RANKING

Given $V = \{V(1), ..., V(q), ..., V(n)\}$, the first q points are the queries and the remaining points are to be ranked based on their

similarities to the queries [9]. $Q$ is defined such that if $V(i)$ is a query $Q(i) = 1$, and otherwise $Q(i) = 0$. Let $W : V * V \Rightarrow \mathbb{R}$ denote the weighted adjacency matrix for a graph in which the weight of each edge is defined as the similarity of the connected points. Let $F : V \Rightarrow \mathbb{R}$ denote a ranking function which assigns a ranking value $F(i)$ to each point $V(i)$. The normalized ranking value indicates the likelihood that a point belongs to the label class (foreground or background).

The manifold ranking algorithm is as follows. Step 1: Connect pairs of points. Step 2: Construct the graph with the weighted adjacent matrix $W$ defined by the similarity of $V(i)$ and $V(j)$ if there is an edge linking them. Otherwise, $W(ij) = 0$. Step 3: Normalize $W$ by $P = (D^{-1}W)^T$, in which $D$ is a diagonal matrix with the $(i, i)$ element equal to the sum of the $i$-th row of $W$. Step 4: Iterate $F_{t+1} = \alpha P F_t + (1-\alpha)Q$ until convergence, where $\alpha$ is a parameter in [0,1) which specifies the relative contribution of the scores from neighbors and the initial score. The limit of the ranking function can be computed directly without iterating [9].

$$F^* = (D - \alpha W)^{-1} Q. \quad (1)$$

Intuitively, this means that all points spread their ranking scores to their neighbors via the weighted adjacency matrix until a global stable state is achieved. The final ranking score corresponds to the saliency of a point.

The labeled queries $Q$ can be either positive (foreground) or negative (background). In this paper, we employ both positive and negative queries as in [2]. We first use each of the four image borders separately as negative queries to get four conspicuity maps and then multiply them to get a temporary saliency map. We segment the temporary saliency map into "foreground" and "background" using the mean map value as a threshold, and then use the foreground regions as positive queries to compute the final saliency map.

The weighted adjacency matrix $W$ is critical in determining how labels propagate through the graph. In the following sections, we describe how we calculate this matrix using image priors for salient object detection.

## 3. PROPOSED PRIOR-BASED GRAPH

In our approach, we first over-segment the image into homogeneous superpixels using the SLIC algorithm [10] and define these superpixels to be the nodes of a graph. The color and spatial features of a superpixel are defined as its mean in the CIELab color space and image coordinates respectively. Both the color and spatial features are normalized to be in [0, 1]. As described in Section 2, each node spreads its ranking score to its neighbors along the edges of the graph based on the edge weights. So, a well-constructed graph should connect nodes which are likely to share the same label, and weight these edges based on the similarity between two connected nodes. The following sections explain how we choose these edges and set their weights in the graph.

### 3.1. Spatial prior

Regions of the image which are spatially nearby are likely to belong to the same class. Therefore, we connect each node to its spatial neighbors (superpixels which share a boundary) and to its neighbors' neighbors (superpixels which share a boundary with any of the first set of neighbors). These edges are weighted by the color distance between the nodes.

$$W(i,j) = exp(-\frac{D_c(i,j)}{\sigma^2}) \quad (2)$$

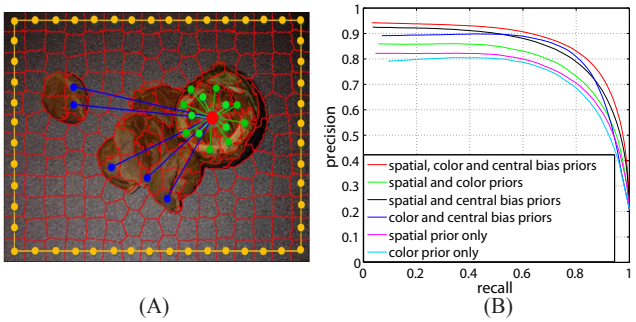


**Fig. 2**. (A) An illustration of our prior-based graph. Image superpixels (red outlines) are the nodes of the graph. A given node (red) is connected to nearby nodes according to the spatial prior (green), and to similarly-colored nodes according to the color prior (blue). In addition, nodes on the image border are connected according to the central bias prior (yellow). (B) An evaluation on the MSRA-B database which shows the effect of omitting one or more priors from the model.

In this formula, $D_c(i, j)$ denotes the color distance between the nodes $i$ and $j$, which is defined as the Euclidean distance between the color features. $\sigma$ is a constant which controls the strength of the weight.

### 3.2. Color prior

Image regions with similar colors are likely to share the same label, and this prior can help to detect multiple foreground objects in an image. Therefore, we also connect the nodes which are neighbors in color space. However, this can produce false alarms when part of the background has a similar color to the foreground object. Therefore, we modify the edge weights (Formula 2) to include both the color distance and the spatial distance.

$$W(i,j) = exp(-\frac{D_c(i,j) + D_s(i,j)}{2\sigma^2}) \quad (3)$$

$D_s(i, j)$ denotes the spatial distance of the nodes $i$ and $j$. The standard way to compute the spatial distance is the Euclidean distance on the image plane. This graph works well when the background is simple and smooth. However, when the background pattern is more complex, this graph may not fully suppress the background regions. To further improve the performance, we exploit another image prior, the central bias.

### 3.3. Central bias prior

Pixels near the center of an image are more likely to belong to a salient object than pixels on the image borders, which are likely to be background [6, 7]. As in [2], we add edges to fully connect all of the nodes on the boundary of the image, to reflect the fact that they are likely to share the "background" label. Furthermore, we notice that there is a drawback to using the Euclidean distance to compute spatial distances, namely that the Euclidean distance is largest between the opposite edges of the image (eg, the left and right borders of the image), which means there is a strong tendency for the image borders to be assigned opposite labels. In fact, the image borders are likely to have the same label ("background"), while points in the center of the image have the opposite label ("foreground"). To solve this problem, we propose a new sine spatial distance.

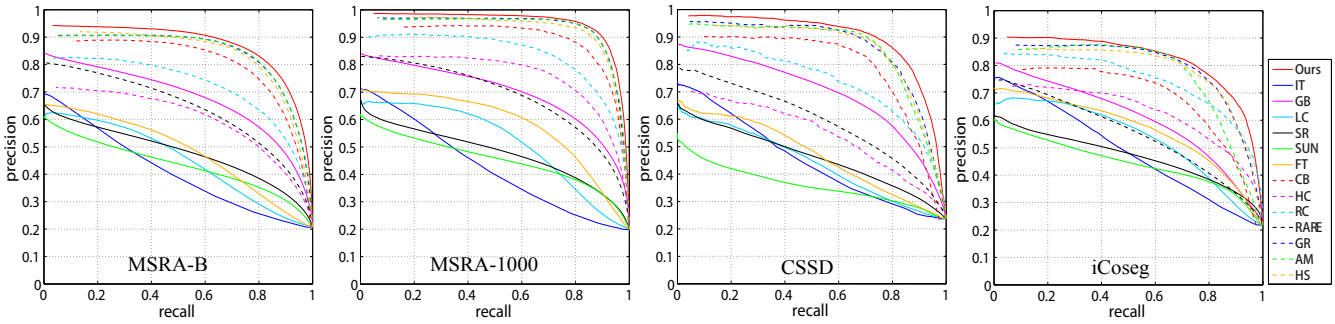


**Fig. 3**. The quantitative comparison of different methods on MSRA-B, MSRA-1000, CSSD and iCoseg databases. Our method gives the best performance on all four databases.

$$D_s(i,j) = \sqrt{(\sin(\pi \cdot |x_i - x_j|))^2 + (\sin(\pi \cdot |y_i - y_j|))^2} \quad (4)$$

$x_i$ and $y_i$ denote the horizontal and vertical coordinates of a point $i$, which have been normalized to be in [0, 1]. $\sin(\pi \cdot |x_i - x_j|)$ computes the sine spatial distance between two nodes $i$ and $j$ along the horizontal axis, while $\sin(\pi \cdot |y_i - y_j|)$ is the distance along the vertical axis. From this formula, it's clear that nodes which have a small Euclidean spatial distance also have a small sine spatial distance. Also, this sine measure means that nodes at the left and right borders of the image are considered spatially close, which reflects the fact that image borders are likely to share the same label.

The summary of our prior-based graph is as follows: Each node is connected to its spatial and color space neighbors. Furthermore, all border nodes are connected to each other as in [2]. Edge weights are based on both color distance and the sine spatial distance between connected nodes. An illustration of our graph construction is shown in Figure 2 (A).

## 4. EXPERIMENTAL COMPARISONS

We compare our method with thirteen state-of-the-art saliency detection methods: IT [1], GB [11], LC [12], SR [13], SUN [14], FT [5], CB [15], HC [16], RC [16], RARE [17], GR [2], AM [3] and HS [18] on the MSRA-B, MSRA-1000, CSSD and iCoseg databases.

**Experimental Setup:** The number of superpixels in each image is set to 200 for all experiments. There are two parameters to set in our algorithm: $\alpha$ in Formula 1, which balances the contributions of the original ranking score and the scores from neighboring nodes, and $\sigma$ in Formula 3, which determines the strength of the weight between connected nodes. These parameters are empirically chosen to be $\alpha = 0.99$ and $\sigma^2 = 0.1$ on the CSSD database and used for all experiments.

**Evaluation Metrics:** We evaluate the algorithms using precision-recall curves. Precision indicates the percentage of detected salient pixels that are correct and recall is the percentage of salient pixels which are successfully detected. The precision-recall curves are calculated by binarizing the saliency map at each threshold in the range [0:1:255] and computing the precision and recall at each threshold.

**Quantitative Comparison:** We first examine our proposed method in detail to understand the roles of the three different priors. We compare saliency detection results using various subsets of priors on the MSRA-B database; the results are shown in Figure 2 (B). It can be seen that our method requires all three prior for best performance, which demonstrates that each of the priors contributes to salient object detection.

Next, we compare our method to thirteen state-of-the-art saliency detection methods on four databases with ground-truth salient object annotations: the 5000-image MSRA-B database, the 1000-image MSRA-1000 database (a subset of the MSRA-B), the 200-image Complex Scene Saliency Database (CSSD), and the 643-image iCoseg database. The CSSD was chosen because it contains complex foreground and background patterns, and the iCoseg database was chosen because many images include multiple foreground objects. The performance of various salient object detection algorithms on these databases is shown in Figure 3. Our method outperforms thirteen state-of-the-art methods on all four databases.

**Visual Comparison:** A visual comparison of the various algorithms can be seen in Figure 4. Representative images have been chosen to highlight the differences between algorithms. The first row shows an image with a complex foreground object: our method highlights the building fairly uniformly, while most methods only detect part of it. The second row shows an image with a complex background. In this case, our method highlights the foreground object well while suppressing the background. GR and AM, which are also graph based methods, only highlight part of the salient object and incorrectly suppress object pixels which are near the border of the image. Our method works better here because our graph includes edges between similarly-colored regions, which can help to highlight the whole salient object. The third row shows an image with a complex foreground and background. Our result is more consistent with the ground truth than the other methods, which tend to highlight only part of the object or include false alarms from the background.

Despite the central bias prior, our method still performs well when the salient object is off- center, as in the fourth and fifth example images. In the fourth row, where the color difference between the foreground and the background is large, LC, FT, CB, HC and HS also perform well. However they are likely to fail when the color difference between the foreground and background is small, as in the fifth row. The last row shows an image with multiple salient objects. Our method can detect all the salient objects while the other graph-based methods, GR and AM, only detect a subset of the objects. In short, the saliency maps generated by our method are more consistent with the ground truth annotation.

## 5. CONCLUSION

In this paper we have described an improved prior-based graph for salient object detection. To construct the graph, we exploit prior knowledge about natural images, including spatial priors, color priors, and a central bias prior. Moreover, we propose a new sine spa-

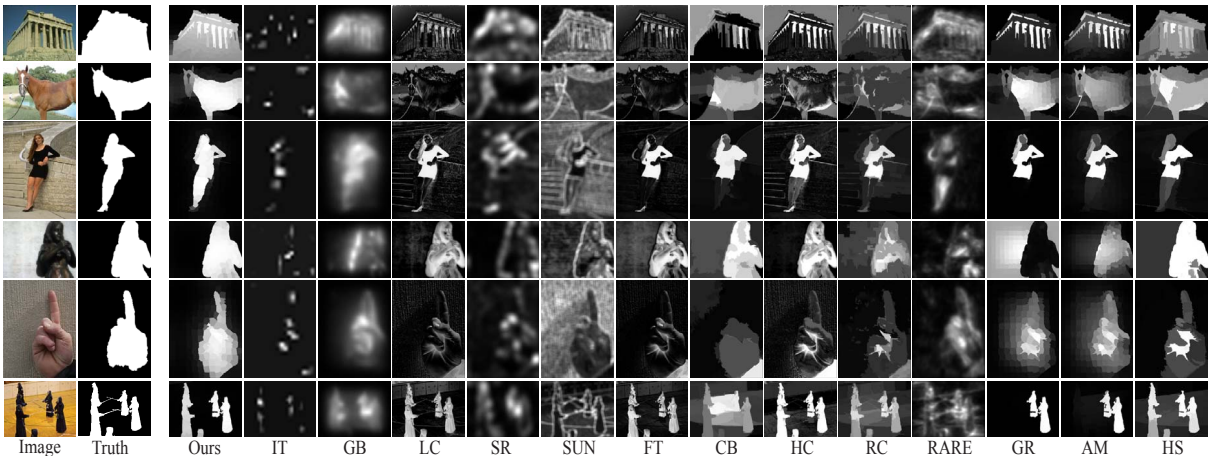


**Fig. 4**. Comparison of different salient object detection methods. The first column is the original image, the second is the ground truth, and the remaining columns are results of the evaluated algorithms. Our method is the first column of results.

tial distance which exploits the central bias in images. Experimental comparisons demonstrate that our method outperforms state-of-the-art salient object detection methods on a variety of images.

## 6. ACKNOWLEDGEMENT

This research was supported by the National Natural Science Fund of China (Grant nos. 61233011, 61103058).

## 7. REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[2] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *International Conference on Computer Vision and Pattern Recognition*, 2013.

[3] 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.

[4] Viswanath Gopalakrishnan, Yiqun Hu, and Deepu Rajan, "Random walks on graphs for salient object detection in images," *IEEE Transactions on Image Processing*, vol. 19, pp. 3232–3242, 2010.

[5] R. Achanta, S. Hemami, F. Estrada, and S. Ssstrunk, "Frequency-tuned salient region detection," in *International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[6] B. W. Tatler, "The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions," *Journal of Vision*, vol. 7, pp. 1–17, 2007.

[7] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *International Conference on Computer Vision*, 2009.

[8] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol. 11, pp. 1–15, 2011.

[9] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Scholkopf, "Learning with local and global consistency," in *IEEE Conference on Neural Information Processing Systems*, 2003.

[10] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, 2012.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Neural Information Processing Systems*, pp. 545–552, 2006.

[12] Yun Zhai and Mubarak Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM Multimedia*, 2006.

[13] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *International Conference on Computer Vision and Pattern Recognition*, 2007.

[14] L. Zhang, M. H. Tong, T. K. Marks, Honghao Shan, and G. W. Cottrell, "Sun: a bayesian framework for saliency using natural statistics," *Journal of Vison*, vol. 0, no. 0, pp. 1–20, 2008.

[15] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li, "Automatic salient object segmentation based on context and shape prior," in *British Machine Vision Conference*, 2011.

[16] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *International Conference on Computer Vision and Pattern Recognition*, 2011.

[17] Nicolas Riche, Matei Mancas, Bernard Gosselin, and Thierry Dutoit, "Rare: A new bottom-up saliency model," in *International Conference on Image Processing*, 2012.

[18] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *International Conference on Computer Vision and Pattern Recognition*, 2013.