FOCUS PRIOR ESTIMATION FOR SALIENT OBJECT DETECTION

Xiaoli Sun⋆ *Xiujun Zhang*† *Wenbin Zou*‡ *Chen Xu*⋆

⋆ College of Mathematics and Statistics, Shenzhen University, Shenzhen, China
† School of Electronic and Communication Engineering, Shenzhen Polytechnic, Shenzhen, China
‡ College of Information Engineering, Shenzhen University, Shenzhen, China

ABSTRACT

In the past five years, salient object detection has become one of the hot topics in the field of computer vision. Focus is a naturally strong indicator for the salient object detection task, but is not well studied. In this paper, a novel method is proposed to estimate the focus prior map for an arbitrary image. Different from the current edge density estimation based methods, the proposed method is based on the sparse defocus dictionary learning on a newly designed dataset. The focus strength is measured by the number of non-zero coefficients of the dictionary atoms. Objectness proposal method is introduced to improve the performance. Comparison with the other focusness estimation methods, the proposed focus prior map is more accurate and easier to be integrated by the other salient object detection methods. Experiments have confirmed the effectiveness and importance of the proposed focus prior.

***Index Terms*—** Salient object detection, Focus prior, Sparse representation, Objectness analysis

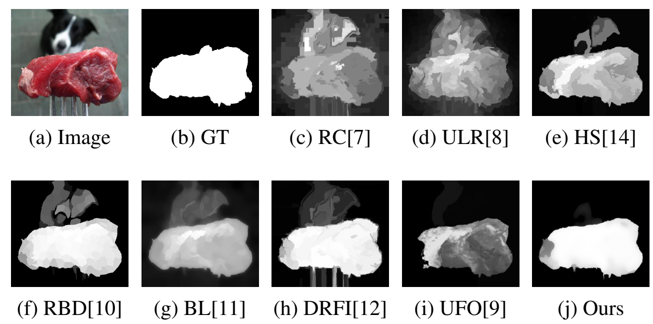


Fig. 1. An challenging example for seven state-of-the-art methods. All results of these seven methods are not satisfied, since the focus prior is not well considered. (a) Input image; (b) Ground truth for salient region; (c)∼(h) Saliency maps using [7, 8, 14, 10, 11, 12] without focus priors; (i) Saliency map using [9]; (j) Saliency map of our method. Benefiting from the focus prior proposed in this paper, our method can solve this kind of challenging problem easily.

1. INTRODUCTION

Great progress on the study of salient object detection has been made in the past five years[1]. Salient object detection aims to localize and segment the most important and attractive objects or regions from arbitrary images. Various applications have benefited from the research on this topic, such as image retargeting[2], image retrieval[3, 4], object recognition[5] and robotics[6] etc.

As is known to all, the methods of salient object detection can be roughly divided into two typical categories: the bottom-up methods and the top-down methods. The bottom-up methods generally focus on low-level features, however, the top-down methods usually involve with high level knowledge based on human perception. To the best of our knowledge, the state-of-the-art methods[7, 8, 9, 10, 11, 12] usually combine these two categories to absorb both of their merits.

A good high level prior plays an important role in the salient object detection task. In [8], Shen et al. proposed a unified low-rank model (ULR) base on the low-rank matrix recovery (LRMR) theory. In order to improve the performance, the authors integrated with three high level priors: location prior, semantic prior and color prior. In [13], Wei et al. proposed background prior for salient object detection. Furthermore, in [10], Zhu et al. proposed the boundary connectivity prior for saliency detection task. All these high level priors have been confirmed to improve the performance of salient object detection by a large margin.

For the salient object detection task, the focus prior is very important, because the salient objects are always photographed in focus. This is determined by the physiological mechanism of the human eye, so the focus prior is a strong indicator for the salient object. But in the state-of-the-art methods, the focus priors are not well considered. As shown in Fig.1, the input image has clear contrast between the object and the background. The photographer made the meat in focus and the dog in blur through setting the depth of field. It indicates that the meat is the really attractive object. But seven state-of-the-art methods have failed in this case, for not considering the focus prior, as shown in Fig.1 (c)∼(h). In [9],

This work is supported by the National Natural Science Funds of China (Grant Nos. 61472257, 61401287, 61401283).

the authors proposed the focusness for saliency detection, as shown in Fig.1 (i). It can be seen that the result of [9] has been improved with the help of focusness prior. But the focusness prior proposed in [9] is integrated in their model, and not convenient to use by the other methods. In some cases, the focusness prior maps generated by [9] are not much accurate, too. In conclusion, the focus prior is important for the salient object detection, but is not well studied. This is just our motivation in this paper.

In this paper, we propose a novel method to estimate the focus prior for arbitrary images. Different from the previous methods to estimate the density of the edges, our method leverages the sparse representation of the blur kernel to get the focus prior. Different from the focusness in [9], our focus prior is independent from any specific models and easy to be integrated by the other methods. Experiments have shown that our focus prior works well in the salient object detection task, and can help to improve the performance of different methods.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes the proposed method. The experiment results are shown in section 4. Finally, the conclusion is given in section 5.

2. RELATED WORK

Up to date, a number of salient object detection methods have been proposed. A survey and benchmark study of salient object detection is provided in [1]. The reader can refer to [1] for details. In this paper, we focus on the focus prior estimation and introduce some typical methods as below.

2.1. Defocus Map Estimation

In [15], Elder and Zucker utilized the first and second order gradient information to find the locations and the blur amount of edges. Since this work, many methods have been proposed to estimate the defocus blur, which could be roughly categorized into two kinds: gradient based methods and frequency based methods. Gradient methods[15, 16] exploit the fact that blur suppresses gradients. Therefore, the gradient distribution in a clear region should be more flat to avoid strong peaks. In the frequency point of view[17], blur suppresses high frequency components and increases low frequency ones.

In [18], Tai and Brown used local contrast prior to measure the defocus at each pixel and then applied MRF propagation to refine the defocus map. In [16], Zhuo et al. proposed an efficient defocus map estimation method based on the Gaussian gradient ratio. First, the input image is re-blurred using a known Gaussian blur kernel. Second, the ratio between the gradients of input and re-blurred images is calculated. It is observed that the blur amount at edge locations can be derived from the ratio. Base on this observation, the blur propagation is formulated as an optimization problem. By solving the optimization problem, a full defocus map is obtained. Further, in [19], Ali et al. proposed to exploit the ratio of gradient magnitude images at multiple scales, and use the scale-space theory to estimate the number of reliable scales.

2.2. Focusness Estimation

In [9], the focusness for salient object detection is firstly proposed. The authors have integrated three important visual cues : uniqueness, focusness and objectness. Uniqueness is used to capture the appearance-derived visual contrast. Objectness is adopted to keep completeness of detected salient regions. And focusness is used to reflect the fact that salient regions are often photographed in focus. In this paper, the authors have treated focusness as a reciprocal of blurness, and estimated the scale of edges using scale-space analysis. The focusness is estimated from both pixel-level and region-level.

At the pixel level, they detect the edges in an image and calculate the Differential-of-Gaussian (DOG) for each pixel at different scales. Then, calculate the differential between different scales, and find out the scale at which the differential reaches its maximum. The focusness at the pixel level is approximated as the reciprocal of this scale. At the region level, the focusness is essentially corresponded to a propagation of the focusness and/or sharpness at the boundary and interior edge pixels to the whole area of a region. The final saliency map S in [9] is defined as:

$$S = \exp(F + U) \times O \tag{1}$$

where F denotes the region-level focusness, U denotes the uniqueness and O denotes the objectness.

To the best of our knowledge, focusness for salient object detection is proposed for the first time in [9]. Since this work is seminal, the focusness is integrated into their model and is hard to be used for the other models.

3. OUR METHOD

3.1. Dataset Construction for Evaluating

For estimating the focus prior maps, a new dataset has been set up in this paper, which contains 231 images. All im-

Fig. 2. Some typical examples selected from the proposed new dataset.

ages have the salient objects in focus and the background in blur. Some typical examples are shown in Fig.2. The source images and the ground truth labels can be found at `http://www.vbig.org/sunxl/index.html`.

3.2. Sparse Dictionary Learning for Defocus

According to [20], the sparse representation can capture the basic information of the input matrix by using dictionary atoms. Given an input matrix $Y = \{y_1, \dots, y_n\} \in R^{d\times n}$, each vector y_i can be represented by a set of dictionary atoms as:

$$\begin{aligned} &\min_{x_i} \|x_i\|_0 \\ &s.t. \|y_i - Dx_i\|_2^2 \le \varepsilon \end{aligned} \tag{2}$$

where $D \in R^{d\times n}$ is an over-complete dictionary learned from Y. x_i is the coefficient to reconstruct y_i. The l_2 norm forces the representation error small enough to accurately recover the original signal. The l_0 norm of x_i forces a few number of dictionary atoms in D which has be used to reconstruct y_i, namely indicating sparsity.

In our method, a slight blurred procedure by Gaussian kernel with $\sigma = 2$ is imposed on images of the proposed dataset. Then we extract overlapped image patches from them. Size of each patch is 8×8, forming an input vector y_i of $d = 64$. At last, our sparse dictionary D with 64 atoms is trained according to Eq.2. Based on our sparse dictionary D, each image patch can be reconstructed by a few atoms together with their non-zero coefficients via Eq.2. Fig.3 have shown the visualization of our sparse dictionary learned from the proposed dataset.

3.3. Focus Prior Map Estimation

According to the signal recovery theory for 1D signal, a square wave can be recovered by a series of sine waves. For the 2D images, a clear edge is corresponding to the square wave of 1D signal. It means that clear edges of an image can be recovered by a lot of sparse dictionary atoms learned from the previous subsection. In contrast, blur edges can be

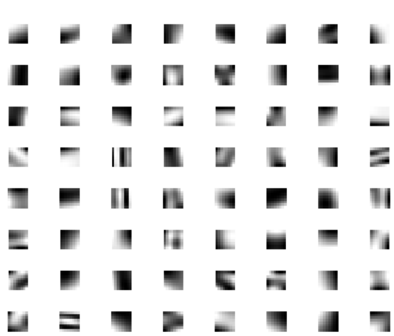


Fig. 3. Sparse dictionary learned from the proposed dataset.

(a) (b) (c)

Fig. 4. An illustration of our method. (a) original images, (b) raw results of focus prior maps, (c) enhanced results by objectness analysis.

recovered by a few of sparse dictionary atoms. So we define the following equation:

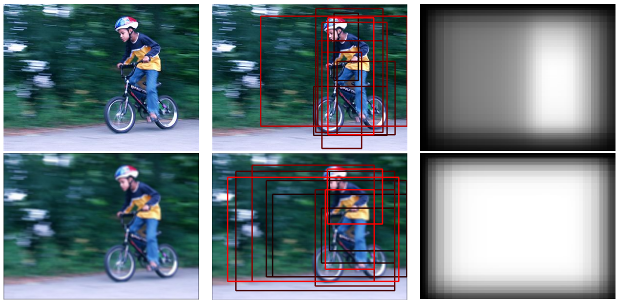
$$f(P_i) = \|x_i\|_0 \tag{3}$$

Here, P_i denotes the i-th patch. The focus strength for the i-th patch, denoted as $f(P_i)$, is defined as the number of non-zero elements in x_i. An illustration have been given in Fig.4 (b).

3.4. Enhancement by Objectness Analysis

It is observed that when glancing at a scene, objects are much easier to abstract the attention of human than the flat regions. Meanwhile, it is more difficult to detect a blurred object by the objectness proposal algorithm than a clear one. Fig.5 has given an example. In Fig.5 (a), the top row is an original image with depth of field, and the bottom row is its blurred result with a Gaussian kernel. In Fig.5 (b), results of the objectness proposal algorithm[21] are shown. It is obvious that the bounding boxes results in the blurred image are more disperse than in the clear image. As seen from Fig.5 (c), in the clear image, the position of object can be detected, while it is failed in the blurred image. Therefore, with the help of objectness analysis, the focus prior map estimation can be more accurate. In the other view, there should be an object firstly in an image, then the focus prior map estimation is meaningful. Otherwise, the result of focus prior map may be seriously wrong.

In this paper, we use the objectness proposal method in [21] to enhance the focus prior map obtained in the previous subsection. First, the objectness proposal method [21] is executed to get the bounding boxes. Second, the pixel level results are computed by summing up the number of covered bounding boxes for each pixel. Denote the values of focusness and objectness at point $\mathbf{q}$ as $f(\mathbf{q})$ and $O(\mathbf{q})$, respectively.



(a) (b) (c)

Fig. 5. Objectness analysis. (a) the first row is an original image with field of depth, and the second row is its blurred result by a Gaussian kernel; (b) bounding box results for objectness analysis using [21]; (c) pixel level results.

Then the enhanced focus prior value at point $\mathbf{q}$ is obtained as follows:

$$Focus(\mathbf{q}) = f(\mathbf{q}) \times O(\mathbf{q}) \tag{4}$$

4. EXPERIMENTS

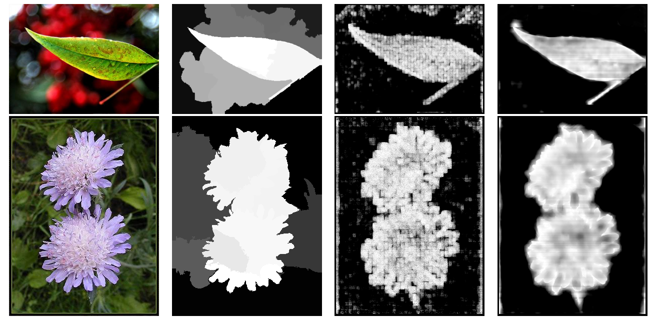
4.1. Comparisons with the Other Focus Map for Salient Object Detection

To the best of our knowledge, the work in [9] is the first method of estimating the focusness for salient object detection. In our experiments, we have shown the comparison results with [9] in Fig.6. As can be seen, our method has more accurate focus prior map than [9]. Especially, benefiting from the object-level analysis, the final results in Fig.6 (d) is much cleaner. In some cases, the final results are much closer to the ground truths, even without introducing the contrast features in color, shape, etc. This comparison experiment has confirmed the effectiveness and importance of our focus prior map.

4.2. Improvement to State-of-the-art Methods for Salient Object Detection

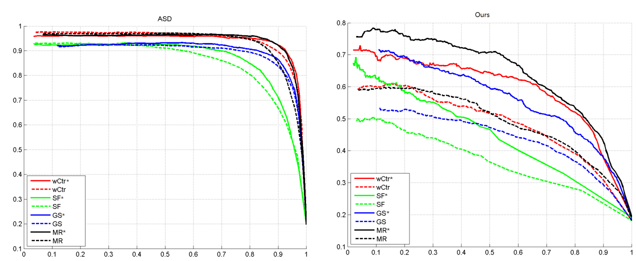
Our focus prior map is an estimation of the object in focus to the given image. It can be easily integrated into the other methods and help to improve their performances. We have integrated our focus prior map in four state-of-the-art methods: wCtr[10], SF[22], GS[13], MR[23]. The experiment results on the ASD dataset[1] have been shown in Fig.7 (a). The performances of all methods have been improved. In order to better show the effect of our method, in Fig.7 (b), we have also shown the comparison results on the proposed datasets. It is more obvious on the proposed dataset, the performances of all methods have been improved by a large margin due to integrating our focus prior map.

[1]`http://ivrgwww.epfl.ch/supplementary_material/RK_CVPR09/`.



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

Fig. 6. Comparison on the focus prior maps. (a) original images; (b) results of focusness estimation in [9]; (c) original focus prior maps of our proposed method; (d) enhanced results by objectness analysis.



(a) ASD dataset (b) Our dataset

Fig. 7. Precise-recall curves comparison. In both subfigures, wCtr*, SF*, GS*, MR* are the results of wCtr[10], SF[22], GS[13] and MR[23] methods integrating the proposed focus prior map, respectively.

5. CONCLUSIONS

In this paper, we have proposed a novel method of estimating the focus prior map for any given images. With the proposed focus prior map, the objects in focus can be found and highlighted automatically. The sparse representation has been used to learn the defocus dictionary and the non-zero number of the coefficients is used to describe the focus strength of each patch. Object level analysis is introduced to boost the performance, since the focus prior is meaningful to the objects in focus. The proposed focus prior map can be used as the focus feature for image analysis or used as an effective prior to improve the performance of salient object detection. Experiments have confirmed the effectiveness and importance of the proposed focus prior map.

6. REFERENCES

[1] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li, "Salient object detection: A survey," *ArXiv e-prints*, 2014.

[2] Yuming Fang, Zhenzhong Chen, Weisi Lin, and Chia-Wen Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Processing,*, vol. 9, pp. 3888–901, 2012.

[3] Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, and Shi-Min Hu, "Sketch2photo: internet image montage," *Acm Transactions on Graphics*, vol. 28, pp. 89–97, 2009.

[4] Peng Wang, Jingdong Wang, Gang Zeng, Jie Feng, Hongbing Zha, and Shipeng Li, "Salient object detection for searched web images via global saliency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2012, pp. 3194–3201.

[5] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona, "Is bottom-up attention useful for object recognition?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2004, pp. 37–44.

[6] Yuanlong Yu, G.K.I. Mann, and R.G. Gosine, "An object-based visual attention model for robotic applications," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, pp. 1398–1412, 2010.

[7] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2011, pp. 409–416.

[8] Xiaohui Shen and Ying Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2012, pp. 853–860.

[9] Peng Jiang, Haibin Ling, Jingyi Yu, and Jingliang Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2013, pp. 1976–1983.

[10] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2014, pp. 2814–2821.

[11] Na Tong, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Salient object detection via bootstrap learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2015, pp. 1884–1892.

[12] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, and Shipeng Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2013, pp. 2083–2090.

[13] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. IEEE, 2012, pp. 29–42.

[14] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2013, pp. 1155–1162.

[15] J. H. Elder and S. W. Zucker, "Local scale control for edge detection and blur estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 699–716, 1998.

[16] Shaojie Zhuo and Terence Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 62, pp. 1852–1858, 2011.

[17] Y. Zhang and K. Hirakawa, "Blur processing using double discrete wavelet transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2013, pp. 1091–1098.

[18] M.S. Brown Y.-W. Tai, "Single image defocus map estimation using local contrast prior," in *Proc. IEEE Conf. Image Process. (ICIP)*. IEEE, 2009, p. 1797C1800.

[19] Ali Karaali and Claudio Rosito Jung, "Adaptive scale selection for multiresolution defocus blur estimation," in *Proc. IEEE Conf. Image Process. (ICIP)*. IEEE, 2014, p. 4597C4601.

[20] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, pp. 4311–4322, 2006.

[21] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 2189–2202, 2012.

[22] F. Perazzi and P. Kr, "Saliency filters: Contrast based filtering for salient region detection,," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2012, pp. 733–740.

[23] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. IEEE, 2013, pp. 3166–3173.