Focus Prior Estimation for Salient Object Detection

Xiaoli Sun[1], Xiujun Zhang[2], Wenbin Zou[3], Chen Xu[1]

College of Mathematics and Statistics, Shenzhen University, China[1]
School of Electronic and Communication Engineering, Shenzhen Polytechnic, China[2]
College of Information Engineering, Shenzhen University, China[3]

Introduction

The focus prior is a strong indicator for the salient object detection, because the salient objects are always photographed in focus. But the focus prior is not well considered. As shown in Fig.1, 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.

In this paper, we propose a novel method to estimate the focus prior for arbitrary images. Our method leverages the sparse representation of the blur kernel to get the focus prior, which 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.

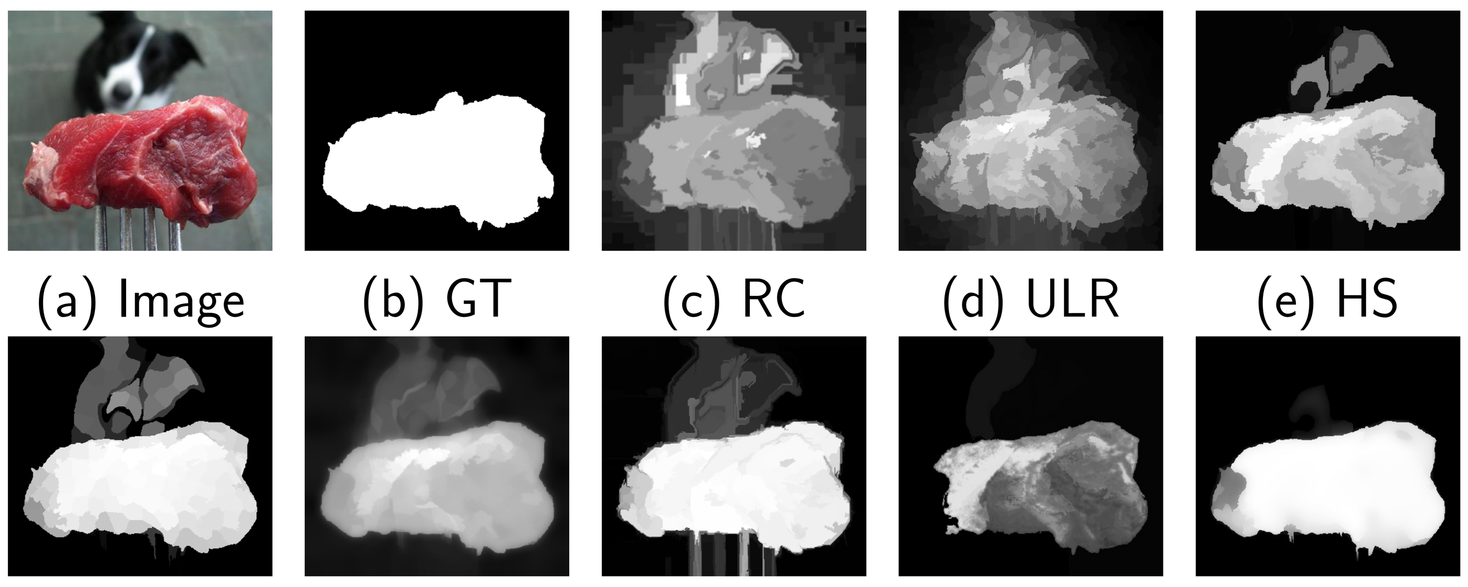

Fig.1 An challenging example for seven state-of-the-art methods. All results of these seven methods are not satisfied, since the focus prior is not well considered.

Dataset Construction for Evaluating

For estimating the focus prior maps, a new dataset has been set up in this paper. All images have the salient objects in focus and the background in blur, are shown in Fig.1.

Figure 1: Some typical examples selected from the proposed new dataset.

Sparse Dictionary Learning for Defocus

Given an input matrix $Y = \{y_1, \ldots, y_n\} \in R^{d\times n}$, each vector y_i can be represented by a set of dictionary atoms as:

$$\min_{x_i} \|x_i\|_0 \quad s.t.\ \|y_i - Dx_i\|_2^2 \leq \varepsilon \tag{1}$$

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. 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.1. Fig.2 have shown the visualization of our sparse dictionary learned from the proposed dataset.

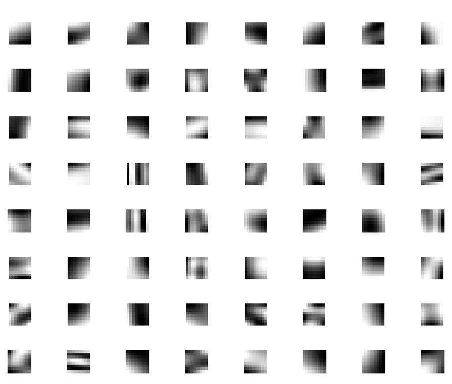

Figure 2: Sparse dictionary learned from the proposed dataset.

Focus Prior Map Estimation

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

$$f(P_i) = \|x_i\|_0 \tag{2}$$

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.3 (b).

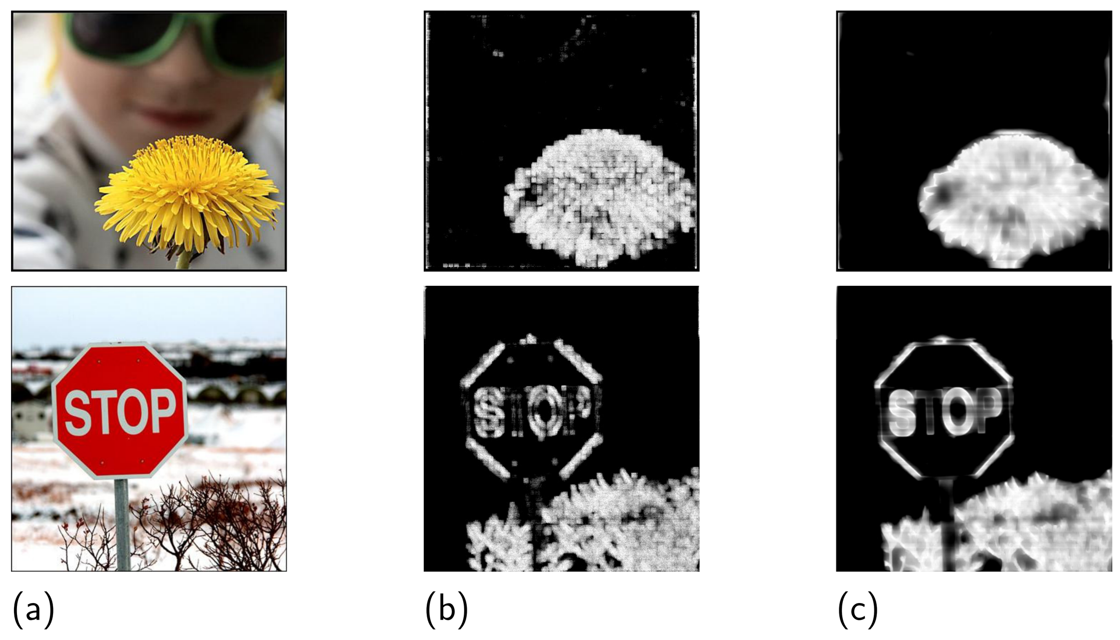

Figure 3: An illustration of our method. (a) original images, (b) raw results of focus prior maps, (c) enhanced results by objectness analysis.

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. Fig.4 has given an example. It is obvious that the bounding boxes results in the blurred image are more disperse than in the clear image. Therefore, with the help of objectness analysis, the focus prior map estimation can be more accurate.

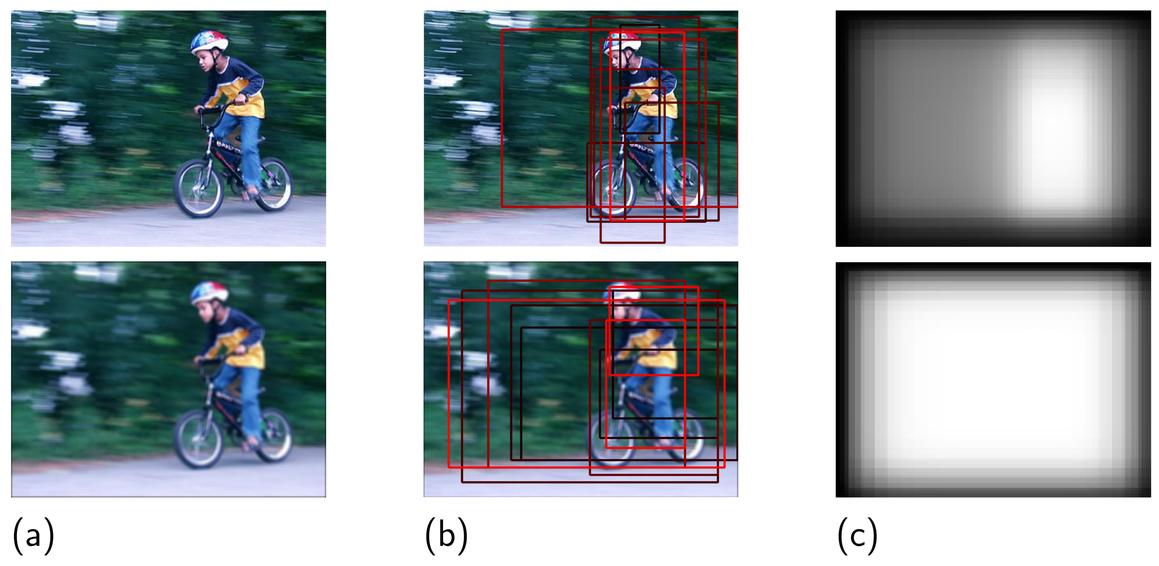

Figure 4: Objectness analysis. (a) the first row is an original image with field of depth, and the second row is its blurred result by a Gaussian kernel; (b) bounding box results for objectness analysis; (c) pixel level results.

Denote the values of focusness and objectness at point $\mathbf{q}$ as $f(\mathbf{q})$ and $O(\mathbf{q})$, respectively. Then the enhanced focus prior value at point $\mathbf{q}$ is obtained as follows:

$$Focus(\mathbf{q}) = f(\mathbf{q}) \times O(\mathbf{q}) \tag{3}$$

Experiments

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, SF, GS, MR. The experiment results on the ASD dataset have been shown in Fig.5 (a). The performances of all methods have been improved. Especially on the proposed dataset, the performances of all methods have been improved by a large margin due to integrating our focus prior map.

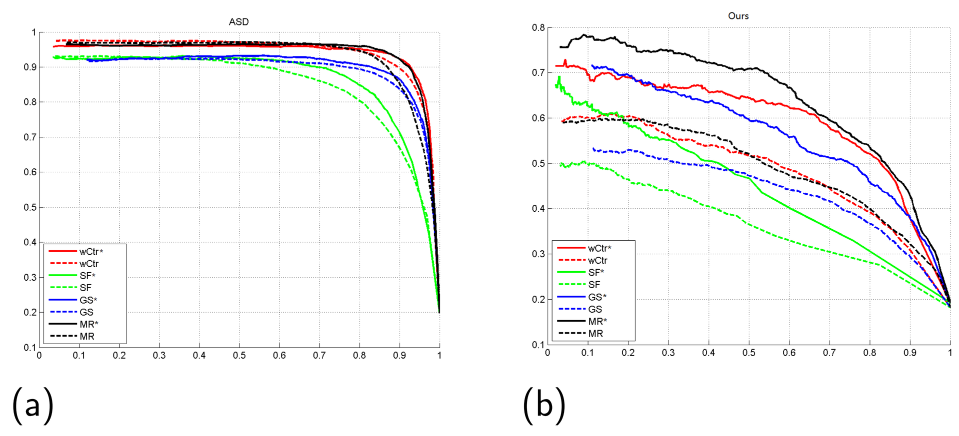

Figure 5: Precise-recall curves comparison.

Conclusion

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 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.