Segmentation Driven Low-rank Matrix Recovery for Saliency Detection

Wenbin Zou[1]
zouwb2008@gmail.com
Kidiyo Kpalma[1]
kidiyo.kpalma@insa-rennes.fr
Zhi Liu [2]
liuzhisjtu@163.com
Joseph Ronsin [1]
joseph.ronsin@insa-rennes.fr

[1] IETR / UMR CNRS 6164,
INSA Rennes,
Université Européenne de Bretagne
Rennes, France

[2] School of Communication and Information Engineering,
Shanghai University
Shanghai, P.R. China

Low-rank matrix recovery (LRMR) model, aiming at decomposing a matrix into a low-rank matrix and a sparse one, has shown the potential to address the problem of saliency detection, where the decomposed low-rank matrix naturally corresponds to the background, and the sparse one captures salient objects. This is under the assumption that the background is consistent and objects are obviously distinctive. Unfortunately, in real images, the background may be cluttered and show low contrast with objects. Thus directly applying the LRMR model to the saliency detection has limited robustness. This paper proposes a novel approach that exploits bottom-up segmentation as a guidance cue for the matrix recovery. This method is fully unsupervised, yet obtains higher performance than the supervised LRMR model.

A key and distinguishing element of this model is the use of proposed *segmentation prior* integrating to the low-rank matrix recovery. Firstly, let us take a look at the images and their coarse-grained (CG) segmentations in Figure 1. Salient objects locate at diversity of positions: center, bottom, left, right and corner. Both background and objects are typically segmented into several regions, thus, the bottom-up segmentation can not be expected to totally separate objects from the background. However, the segmented regions of background have very high probability of connecting with the border of the image, while very few regions of objects link to it. Even if an object is truncated on the border, like the bike and the child of the two right-most images, border regions of object are small compared to the whole object in the image. In contrast, the border regions of the background are usually large, as the background appears more uniform, like sky, road, tree, wall, etc. This observation implies that objects can be roughly separated from the background by the bottom-up segmentation. Therefore, we propose the segmentation prior according to the connectivity between each region and image border. Let r_m be a segmented region of image $\mathbf{I}$, the segmentation prior of region r_m is defined as

$$h_m = \exp\left(-\frac{\|r_m \cap C\|}{\sigma \psi_m}\right) \tag{1}$$

where $\|\cdot\|$ denotes the length of intersection, C is the border of image $\mathbf{I}$, ψ_m is the outer perimeter of region r_m, and σ is a balance parameter which is set to 0.3 in our experiments. Clearly, if a region touches the image border, its prior value is in the range of $(0,1)$, otherwise it is equal to 1. In other words, the segmentation prior gives a small weight to the region touching the image border. Using (1), segmentation priors of all regions can be computed, and form the prior of the input image.

In Figure 1, one might observe that, on one hand, there are still some regions of the background without connection with the image border, on the other hand, some regions of objects are inevitably merged with the background. Indeed, such a strategy can not perfectly separate the objects from the background. However, the segmentation prior can serve as a guidance cue for LRMR model to address the task of saliency detection.

Suppose an input image $\mathbf{I}$ is segmented into N superpixels, and represented by a feature matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N]$. Let $\mathbf{H}_c = [h_1^c, h_2^c, \cdots, h_N^c]$ denote a set of CG segmentation prior values of the superpixels. In order to recover well salient objects with the LRMR model, the feature matrix $\mathbf{A}$ is firstly modulated by the CG segmentation prior $\mathbf{H}_c$

$$\mathbf{B} = [h_1^c \mathbf{a}_1, h_2^c \mathbf{a}_2, \cdots, h_N^c \mathbf{a}_N]. \tag{2}$$

Then, the modulated feature matrix $\mathbf{B}$ is used as the input of the standard LRMR model

$$\begin{aligned} \min \quad & \|\mathbf{U}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{B} = \mathbf{U} + \mathbf{E} \end{aligned} \tag{3}$$

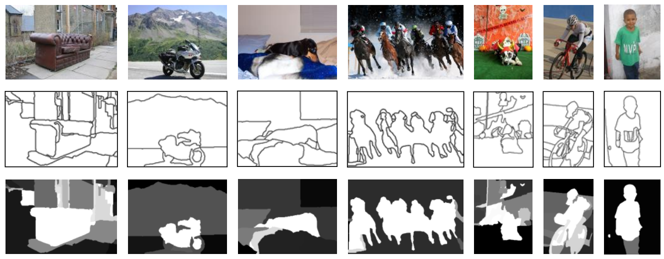

Figure 1: Examples of segmentation prior. First row: input images; second row: segmentation results; last row: segmentation prior.

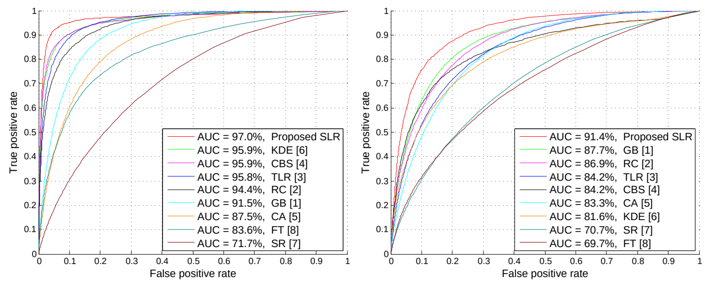

Figure 2: ROC curves and AUC scores of different models on MSRA-1000 (left) and PASCAL-1500 (right) datasets.

where λ is a coefficient to balance $\mathbf{U}$ and $\mathbf{E}$, $\|\cdot\|_*$ denotes the nuclear norm of matrix $\mathbf{U}$ (the sum of singular values of $\mathbf{U}$), and $\|\cdot\|_1$ indicates l_1-norm which ensures to produce a sparse matrix $\mathbf{E}$. With the optimal sparse matrix $\mathbf{E}$, the saliency of a superpixel is given by the l_1 energy of corresponding vectors in $\mathbf{E}$.

The success of the proposed segmentation prior modulated low-rank matrix recovery model (SLR) is mainly due to two reasons. On one hand, real images typically possess high redundancy in the feature space, and object pixels tend to be salient compared to the background, which enables to discover objects from the sparse matrix $\mathbf{E}$. On the other hand, as the segmentation prior assigns small weights to most of background feature vectors in $\mathbf{B}$, the l_1 energies of the corresponding vectors in the recovered matrix $\mathbf{E}$ are inclined to be small. Therefore, objects can be captured more effectively from the matrix $\mathbf{E}$.

The proposed SLR model is evaluated on two datasets (2500 images in total): the widely used MSRA-1000 dataset and the newly introduced but more challenging PASCAL-1500 dataset. The performance of saliency detection is measured by receiver operator characteristic (ROC) curve and the area under the curve (AUC). As shown in Figure 2, the proposed model outperforms 8 state-of-the-art models [1-8] on both datasets, with 1.1% and 3.7% improvement in terms of AUC score on MSRA-1000 and PASCAL-1500, respectively, compared to the best one among the reference models.

[1] Harel et al. Graph-based visual saliency. *Adv. Neural Inf. Process. Syst.*, 2007.
[2] Cheng et al. Global contrast based salient region detection. *CVPR*, 2011.
[3] Shen and Wu. A unified approach to salient object detection via low rank matrix recovery. *CVPR*, 2012.
[4] Jiang et al. Automatic salient object segmentation based on context and shape prior. *BMVC*, 2011.
[5] Goferman et al. Context-aware saliency detection. *CVPR*, 2010.
[6] Liu et al. Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. *TMM*, 2012.
[7] Hou and Zhang. Saliency detection: A spectral residual approach. *CVPR*, 2007.
[8] Achanta et al. Frequency-tuned salient region detection. *CVPR*, 2009.