Learn Local Priors by Transferring Training Masks for Salient Object Detection

Dan Wang[1,2], Canxiang Yan[3], Quan Zhou[4]

[1]Institute of Spacecraft System Engineering, China Academy of Space Technology, 100094, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Institute of Deep Learning, Baidu, Inc., Beijing, 100085, China
[4]College of Telecom & Inf Eng, Nanjing Univ of Posts & Telecom, Nanjing, China

ABSTRACT

In this paper, we present a novel framework to incorporate high-level guidance and low-level features to automatically identify salient objects based on two ideas. The first one considers the specific location prior to encode visual saliency, while the second one estimates image saliency using contrast with respect to background regions. The proposed framework consists of the following three steps: a) Top-down process: a specific location saliency map (SLSM) is learned. Specifically, for each image window (patch), a set of image windows with similar appearances are searched from a training image set and the corresponding segmentation masks are linearly integrated to produce a prior map, which guides the saliency object detection. 2) Bottom-up process: a multi-layer segmentation framework is employed, which provides vast robust background candidate regions specified by SLSM. Then the background contrast saliency map (BCSM) is computed based on low level image stimuli features. SLSM and BCSM are finally integrated to produce a pixel-based saliency map. Extensive experiments show that our approach achieves competitive results over MSRA 1000 and SED datasets, where each image contains more than one salient object.

***Keywords*—**Saliency, Object detection, Top-down, Bottom-up.

I. INTRODUCTION

THE human visual system (HVS) has an outstanding ability to quickly detect the most interesting regions in a given scene. In last few decades, the highly effective attention mechanisms of HSV have been extensively studied by researchers in physiology, psychology, neural systems, and image processing.

Visual attention is also an important and challenging research topic in the field of computer vision [18, 27], and the computational modeling of this system enables various vision applications, e.g., object detection/recognition [4, 11], image matching [37], image segmentation [21, 10], and video tracking [30, 25, 29]. Visual saliency can be viewed from different perspectives. Top-down (supervised) and bottom-up (unsupervised) are two typical categories. The first category often describes the saliency by the visual knowledge constructed from the training process, and then uses such knowledge for saliency detection on the test images [12, 26, 36, 24, 6, 20, 31, 33, 28, 22].

On the other hand, the bottom-up models are mainly motivated from the contrast formulation [7, 13, 14, 34]. For example, Itti et al. [18, 16] proposed a set of pre-attentive features including local center-surround intensity, color and direction contrasts. These contrasts were then integrated to compute image saliency through the winner-take-all competition. Chen et al. [8] and Achanta et al. [2] utilize the global contrast with respect to the entire scene to estimate visual saliency. Recently, Borji and Itti [5] combine local and global patch rarities as contrast to measure saliency for eye-fixation task. We argue that the contrast based on background regions also plays an important role in such processes. In this paper, we propose a novel method to integrate bottom-up, lower-level features and top-down, higher-level priors for salient object detection.

In this paper, we propose a novel method to integrate bottom-up, lower-level features and top-down, higher-level priors for salient object detection. The key idea of our top-down process is inspired by [23], which transfers segmentation masks from a supervised training set to the test image. The transferred segmentation masks are then used to derive specific location prior of salient object in the test image. Fig.1 illustrates the overview of our method. We first extract candidate windows likely to contain salient objects [4], and then transfer masks from training windows that are visually similar to windows in the test image. The intuition is that visually similar windows often have similar segmentation masks. As these windows exhibit less variability than whole images and are often centered on salient objects, they form much better support regions for location transfer. Afterwards, the bottom-up saliency map is computed based on low-level image stimuli features. In nature images, although the salient regions and backgrounds may also tend to be perceptually heterogeneous, the appearance cues (e.g., color and texture) of the salient object region are still quite different from the backgrounds. As a result, different from the previous methods that mainly utilize the local central-surround contrast [18, 26, 5] and global contrast [2, 8, 15] to encode saliency, our framework estimates visual saliency using the appearance-based contrast with respect to the background candidate regions. In order to automatically abstract robust background regions, we employ the multi-layer segmentation framework, which is able to

This work is partially supported by Natural Science Foundation of China, under No.61402430 and China Postdoctoral Science Foundation.

978-1-5090-6067-2/17/$31.00 ©2017 IEEE

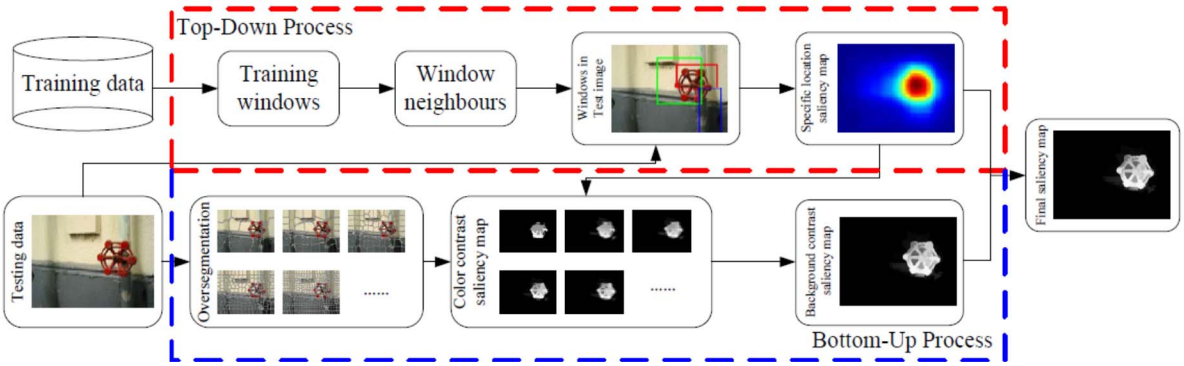

Fig.1 Our approach consists of two components: (1) Top-down process. The training data consists of images annotated with binary segmentation masks. We first detect windows likely to contain salient objects on all training images as well as on the test image using [4]. Then the binary segmentation masks are transferred from the training windows with the most similar appearance (window neighbours) to each detective windows in testing image. The transferred segmentation masks are used to derive the specific location saliency map (SLSM); (2) Bottom-up process. Using the over-segmentation technique of [3], an input testing image is first partitioned to multi-layer segmentation with coarse to fine manner. Given the SLSM as prior map, a set of robust background regions are abstracted, and then the color-based contrast saliency maps are created for each layer of segmentation. These saliency maps are combined to form our background contrast saliency map (BCSM). SLSM and BCSM are finally integrated to a pixel-accurate saliency map. (Best viewed in color)

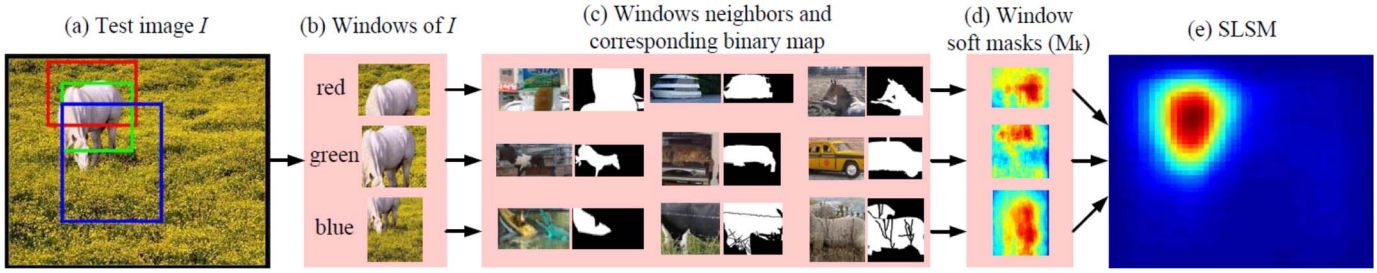

Fig.2 An example of the full pipeline for the top-down process. Given a test image *I* in (a), three windows (denoted as red, green and blue patches) are highlighted out of N windows, as shown in (b). The window neighbors are displayed in (c). Green window is tightly centered on an object and gets very good neighbors, while for red and blue windows, the neighbors are a good match for segmentation transfer, even though the window does not cover the "horse" perfectly. This results in an accurate transfer mask for each window of *I*, as illustrated in (d). On the rightmost column of (e), the Mk from all windows are integrated into a soft mask for the whole image, which is used to derive the SLSM. Note blue color denotes low saliency, while red color represents high saliency (Best viewed in color)

provide large amount of background candidates within different sizes and scales.

The contributions of our approach are two-fold:

(1) In the top-down process, it proposes a specific location prior for salient object detection. Through window mask transferring, our method is able to provide more accurate location prior to detect salient regions, which results in more accurate and reliable saliency maps than the models using center-biased assumptions, such as [36];

(2) In the bottom-up process, unlike the previous approaches that utilize the local and global contrast to predict visual saliency, we attempt to estimate visual saliency using the contrast with respect to the background regions.

II. Our Approach

In this section, we elaborate on the details of our method. We first introduce how to obtain the specific location saliency map (SLSM) by transferring window masks. Given the multi-layer segmentations and SLSM on hand, we select a series of background regions that are used to compute background contrast saliency map (BCSM). Finally, two maps are incorporated to generate per-pixel saliency.

A. Specific Location Saliency Map (SLSM)

Finding Similar Windows. Given an image I, we first detect windows likely to contain an object using the "objectness" technique of [4]. It tends to return more windows covering an object with a well-defined boundary in space, rather than amorphous background elements. In our experiments, sampling only N windows per image (e.g., $N = 100$) already covers most salient objects, and the top three windows are shown in Fig.2(a). We extract these windows for all training images as well as for the test image. Putting all the training windows together, we obtain the training window set $\{W_t\}$. From Fig.2(a), it is observed that many detective windows are centered on a salient object. This leads to retrieving much better neighbors, whose segmentation masks are more suitable to transfer for test image. Fig.2(b) shows three "horse" windows in the test image. The nearest neighbor windows accurately depict similar animals in similar poses, resulting in well matching binary segmentation masks, as illustrated in Fig.2(c).

Given a test image, we compare each test window W_k; $k = \{1,2,\ldots,N\}$ to all training windows $\{W_t\}$. Thus, the set $\left\{S_k^j\right\}$; I $=\{1,2,\ldots,M\}$ containing the segmentation masks of the top M training windows most similar to W_k is passed on to the next

processing stage.

Segmentation Transfer. Let $S_T(x,y)$ be the SLSM to convey a sense of the likely segmentation of a pixel based only on its location within the image, so that $S_T(x,y)$ gives the probability of pixel at location (x,y) to be salient. We construct $S_T(x,y)$ for each pixel from the segmentation masks transferred via all windows containing it.

1) Soft masks for windows. For the k^{th} test window W_k, we have a set $\{S_k^j\}$ containing the segmentation masks of neighbors from the training set. We now compute a soft segmentation mask M_k for each W_k as the pixel-wise mean of the masks in $\{S_k^j\}$. For this, all masks in $\{S_k^j\}$ are resized to the size of W_k in both their width and height dimensions. In this aligned space, a pixel value in M_k corresponds to the probability for it to be salient object in $\{S_k^j\}$. Fig.2(d) shows the corresponding Mk for the detected windows.

2) Soft mask for the test image. We now integrate the M_k for all windows into a single soft segmentation mask $M(x,y)$ for the test image I. For each window W_k, we place its soft mask M_k at the image location defined by W_k. The soft mask $M(x,y)$ of the test image is the pixel-wise mean of these placed masks. A pixel value in $M(x,y)$ is the probability for it to be salient, according to all transferred segmentations (as illustrated in Fig.2(d)). Therefore, we define the SLSM $S_T(x,y)$ as

$$S_T(x,y) = M(x,y) \quad (1)$$

Due to the integration of soft foreground masks from the individual windows, our approach achieves even more robust results. Fig.3 exhibits some SLSMs of nature images over MSRA 1000 and SED datasets.

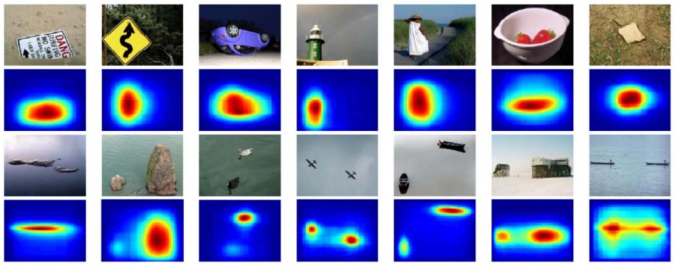

Fig.3 Illustration of SLSM. The first and third rows are the example images form MSRA 1000 and SED dataset, respectively. The second and fourth rows are the corresponding SLSM, where blue denotes low saliency and red color represents high saliency.

B. Background Contrast Saliency Map (BCSM)

No matter where the salient object locates, it often exhibits quite different appearance cues (e.g., color and texture) within the entire scene. We thus build our background contrast saliency map (BCSM) guided by the global color-based contrast measurement [8]. Instead of computing saliency based on an entire image, here we calculate the contrast based on background candidates.

Multi-layer Segmentation. In order to make full use of background candidate regions, we employ the multi-layer segmentation framework.

Although there are many over-segmentation methods [9], the SLIC algorithm [3] is adopted in our implementation due to its efficiency. In practice, we partition test image I into J layers of segmentations. As shown in Fig.4, the advantages of using this technique are that it can often group the homogeneous regions with similar appearance and preserve the true boundary of objects.

Fig.4 Imazge representation by multi-layer segmentation. The upper panel shows examples from MSRA dataset, while the bottom panel illustrates the examples from SED dataset. From left to right are the original images and their over-segmentation results in a coarse to fine manner. Different segments are separated by white boundaries.

Computing BCSM. Denote $S_B(x,y)$ as the BCSM toconvey a sense of the dissimilarity of a pixel based on its local feature with respect to backgrounds, so that $S_B(x,y)$ also gives the probability of pixel at location (x,y) to be salient. We construct $S_B(x,y)$ for each pixel from the multi-layer segmentation via all segments containing it.

1) Color contrast saliency for each layer segmentation.

Let r_i^j be i^{th} specific segment in j^{th} layer of segmentation. According to the SLSM, we select the segments with low saliency value to be background candidates, which are ready to compute the color-based contrast saliency. Let $B^j = \{B_1^j, B_2^j, ..., B_M^j\}$ be selected background candidate regions in j^{th} layer segmentation. To measure how distinct the salient region is with respect to $B_m^j \in B^j$, we can measure the distance between r_i^j and B_m^j using various visual cues such as intensity, color, and texture/texton. In this paper, we use the inverse cosine distance between histograms of HSV space to compute the color-based contrast:

$$C_{i,m}^j\left(H(r_i^j), H(B_m^j)\right) = 1 - \frac{H(r_i^j)H(B_m^j)}{\left\|H(r_i^j)\right\| \cdot \left\|H(B_m^j)\right\|}, \quad (2)$$

where H(.) is the binned histogram calculated from all color channels of one segment, and $\|.\|$ denotes the L-2 norm.

We use histograms because they are a robust global description of appearance. They are insensitive to small changes in size, shape, and viewpoint. From (2), it is observed that the contrast between r_i^j and B_m^j is very low when they look similar, otherwise not. For the given segment r_i^j, its color contrast saliency $S_B(r_i^j)$ is computed as the mean of L smallest contrasts $\{C_{i,m}^j(.,.)\}$, m=1,2,…,M:

$$S_B(r_i^j) = \frac{1}{L}\sum_{m=1}^{L}\{C_{i,m}^j(.,.)\}. \quad (3)$$

As will be seen, when r_i^j is truly a salient region, the L smallest contrasts always get large value with respect to the background regions, resulting in high saliency for $S_B(r_i^j)$. The saliency map $S_B(r_i^j)$ is normalized to a fixed range [0; 255], and $S_B(r_i^j)$ is assigned to each image pixel belonging to r_i^j with the saliency value as $S_B{}^j(x,y)$.

2) BCSM for testing image.

We now incorporate the $S_B{}^j(x,y)$ for all segmentation layers into a single saliency map for the test image *I*. Then the BCSM $S_B(x,y)$ is defined as:

$$S_B(x,y)=\frac{1}{J}\sum_{j=1}^{J}S_B{}^j(x,y) \tag{4}$$

$S_B(x,y)$ is also normalized to a fixed range [0; 255].

C. Combined Saliency

We integrate SLSM and BCSM to produce our final saliency map $S(x,y)$ with a linearly combination model

$$S(x,y)=\eta.S_T(x,y)+(1-\eta)S_B(x,y) \tag{5}$$

where η is the harmonic parameter to balance the top-down SLSM and bottom-up BCSM. Then S(x; y) is normalized to a fixed range [0; 255].

III. Experimental Results

To validate our proposed method, we carried out several experiments on two benchmark datasets using the Precision-Recall curve and F-measure described below. The main reason behind employing several datasets is that current datasets have different image and feature statistics, stimulus varieties, and center-biases. Hence, it is necessary to employ several datasets as models leverage different features that their distribution varies across datasets.

Datasets. We test our proposed model on two datasets: (1) Microsoft Research Asian (MSRA) 1000 dataset [2] is the most widely used for model comparison. It contains 1000 images with resolution of approximate 400× 300 or 300 ×400 pixels, which provides accurate object-contour based ground truth. (2) The SED [6] dataset is a smaller dataset only containing 100 images with resolution ranged from 300×196 to 225×300 pixels. It is more challenging, however, since there are two salient objects in each image.

Baselines. To show the advantages of our method, we selected 12 state-of-the-art models as baselines for comparison, which are spectral residual saliency (SR [15]), spatiotemporal cues (LC [39]), visual attention measure (IT [18]), graph-based saliency (GB [19]), frequency-tuned saliency (FT [2]), salient region detection (AC [1]), contextaware saliency (CA [12]), global-contrast saliency (HC and RC [8]), saliency filter (SF [35]), low rank matrix recovery (LR [36]), and geodesic saliency (SP [38]). In practice, we implemented all the 12 state-of-the-art models using a Dual Core 2.6 GHz machine with 4GB memory over two datasets to generate saliency maps.

Evaluation Metrics. In order to quantitatively evaluate the effectiveness of our method, we conducted experiments based on the following widely used criteria. The precision recall curve (PRC) is used to evaluate the similarity between the predicted saliency maps and the ground truth. Precision corresponds to the percentage of salient pixels correctly assigned, while recall corresponds to the fraction of detected salient pixels in relation to the ground truth number of salient pixels. Another criterion to evaluate the overall performance is the F-measure [2, 8], which is used to weight harmonic mean measurement of precision and recall. The F-measure is defined as $F_\beta=\frac{(1+\beta^2)\times\text{Precision}\times\text{Recall}}{\beta^2\times(\text{Precision}+\text{Recal})}$, β^2=0.3 following [2, 8].

Implemented Details. In all experiments, we compute GIST [32] inside each window to describe its appearance and we compare GIST descriptors with the L2 distance to find widow neighbors. The parameter settings are: N = 100 for sampling windows per image, M = 100 for window neighbors for one sampling window, J = 5 for segmentation layers in coarse to fine manner, L = 5 for computing color contrast saliency involved in Equation (3), (rgnSize, regularizer) are initialized as {25, 10}, and rgnSize is updated as {25, 50, 100, 200, 400} with fixed regularizer, η = 0.6 to balance SLSM and BCSM for producing final saliency map. We follow two widely used methodologies [2, 36] to implement our experiments. In the first implementation, we adopt the scheme that segments image according to the saliency values with a fixed threshold. Given a

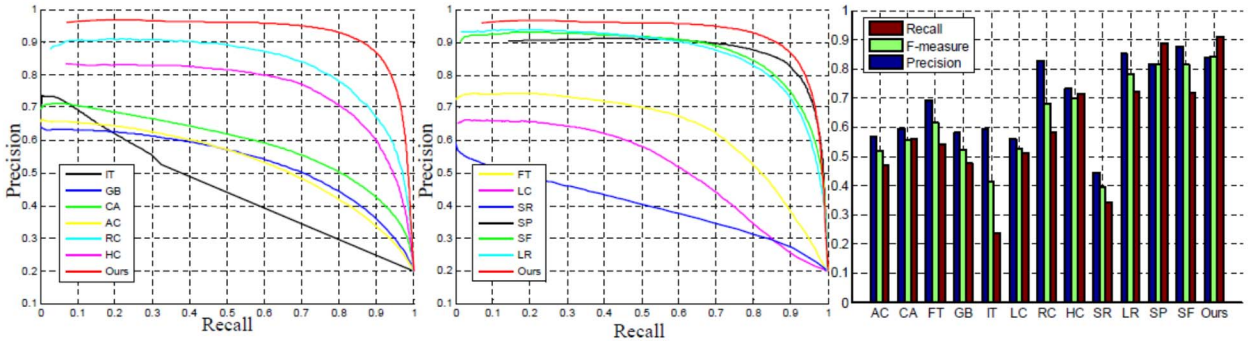

Fig.5 Quantitative comparison for all algorithms with thresholding of saliency maps using 1000 publicly available benchmark images. Left and middle: PRC of our method compared with CA [12], AC [1], IT [18], LC [45], SR [15], GB [19], SF [35], LR [36], FT [2], SP [38], HC and RC [8]. Right: Average precision, recall and F-measure with adaptive-thresholding segmentation. (Best viewed in color)

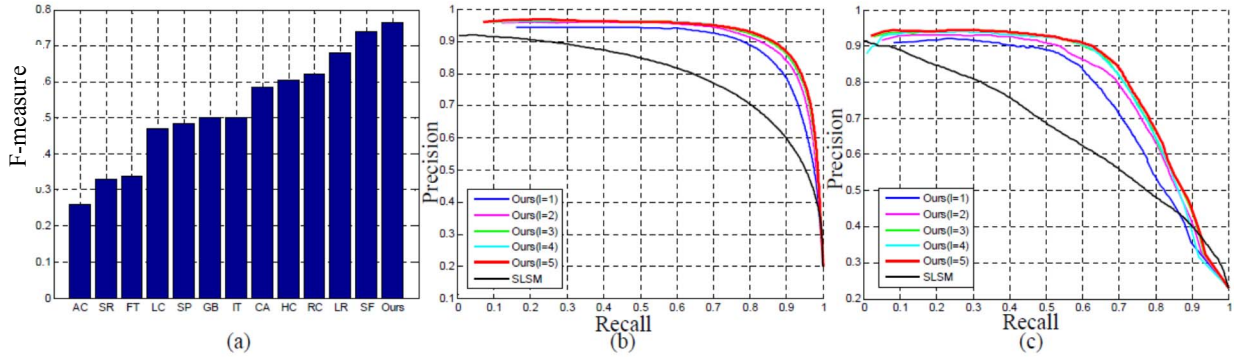

Fig. 6 (a) F-measure of the different saliency models to ground truth over SED. (b) and (c) The comparison of PRC by gradually increasing the layers of segmentation over MSRA 1000 and SED dataset, respectively. The performance of individual SLSM is also included.'

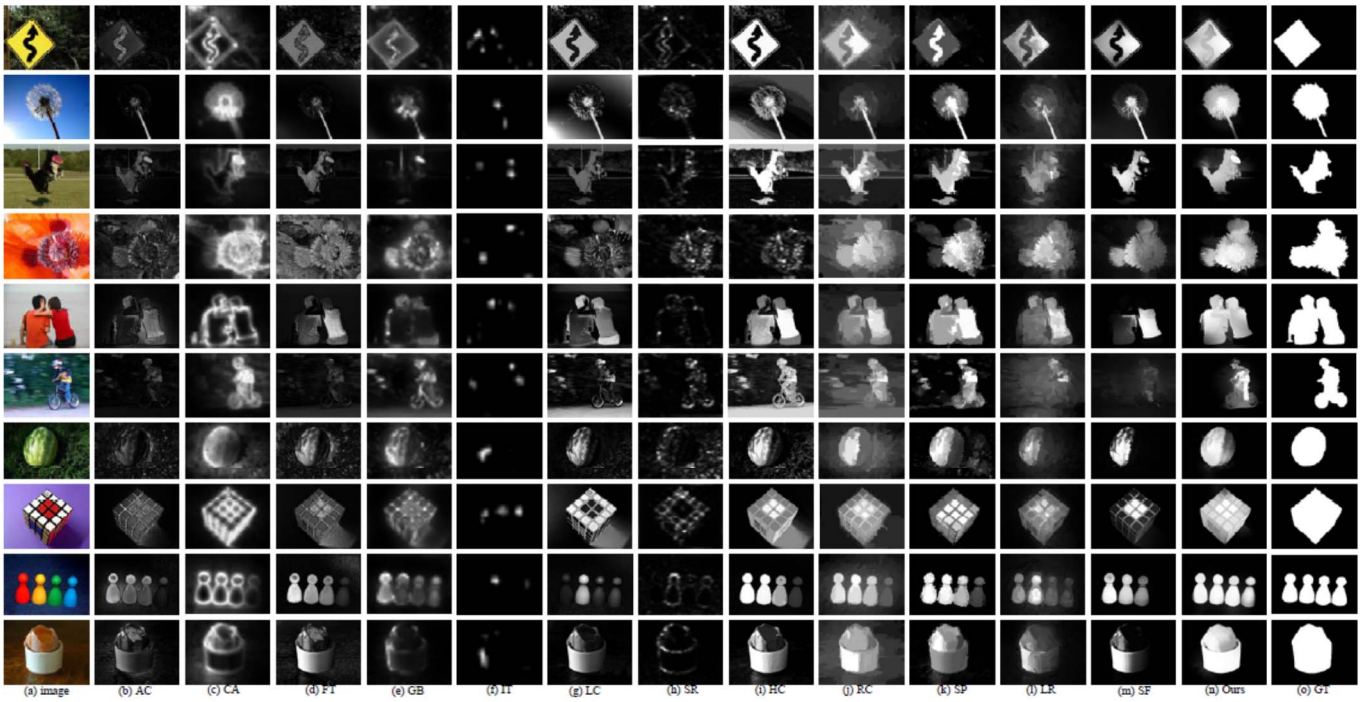

Fig.7 Visual comparison of previous approaches with our method. See the legend of Fig.5 for the references to all methods.

threshold T∈ [0, 255], the regions whose saliency values are higher than threshold are marked as a salient object. The segmented image is then compared with the ground truth to obtain the precision and recall. We draw the PRC using a series of precision-recall pairs by varying T from 0 to 255.

In the second implementation, the test image is segmented by an adaptive threshold method [36]. Given the over-segmented image, an average saliency is calculated for each segment. Then an overall mean saliency value over the entire image is obtained as well. If the saliency in this segment is larger than twice of the overall mean saliency value, the segment is marked as foreground. Precision and recall values are sequentially calculated, and F-measure is finally computed for evaluation.

Overall Results. The average PRC and F-measure over MSRA 1000 dataset are illustrated in Fig. 5. It clearly shows that our method outperforms other approaches. It is interesting to note that the minimum recall value of our methods starts from 0.08, and the corresponding precision is higher than those of the other methods, probably because the saliency maps computed by our methods contain more pixels with the saliency value 255. The improvement of recall over other methods is more significant, which means our method are likely to detect more salient regions, while keeping a high accuracy.

We also evaluate our method on SED dataset and compare it with other 12 models. Fig. 6(a) reports the comparison results in terms of F-measure. Our method achieves competitive results compared to the state-of-the-art and higher F-measure value (ours = 0.763) than other competitive models (SF = 0.739, LR = 0.68, RC = 0.62, and HC = 0.60), which clearly shows the validity of our approach in the case of more than one salient object within each image. Visual comparison with different methods over MSRA 1000 dataset are shown in Fig.7, and some qualitative results over SED dataset are displayed in Fig.8. Compared with other models, our method is very effective in eliminating the cluttered backgrounds, and uniformly highlighted salient regions with well-defined object shapes, no matter whether salient objects locate in image center, or far away from image center, even in the image boundary.

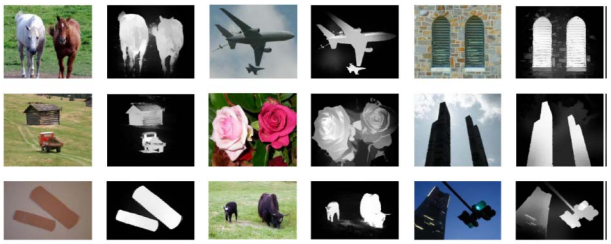

Fig.8 Some visual examples over SED dataset. The first, third and fifth columns are original images, and the second, fourth and sixth columns are the corresponding saliency maps.

Analysis of Parameter Setting and Individual Saliency Map. One factor affecting the performance is the layers of over-segmentation. Fig. 6(b) and (c) exhibit the plot of PRC with different number of segmentation layers over the MSRA 1000 and SED datasets, respectively. It is observed that better performance can be achieved along with the increasement of segmentation layers, and no further improvement after 5 layer segmentations. This demonstrates that our method performs robustly over a wide range of segmentation layers. In order to evaluate the contributions of each individual saliency map, the PRC of SLSM is also included for comparison in Fig. 6(b) and (c). The performance of SLSM is already better than most of the competitive models, whose results are shown in Fig. 5. Using BCSM noticeably improves the performance for both two datasets, which indicates the importance of measuring visual saliency using contrast with respect to background regions.

IV. Conclusion and Future Work

In this paper, we propose a novel framework for salient object detection based on two key ideas: (1) using specific location information as top-down prior by transferring segmentation masks from windows in the training images that are visually similar to windows in the test image; (2) using the contrast based on the background candidates in bottomup process makes our method more robust than methods estimating contrast on the entire image. Compared with existing competitive models, the extensive experiments show that our approach achieves competitive results compared to the state-of-the-arts over MSRA 1000 and SED datasets. In the future, we would like to combine two saliency maps, SLSM and BCSM, with adaptive weights using learning technique, as well as [17] does.

V. References

[1] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk. Salient region detection and segmentation. Computer Vision Systems, pages 66–75, 2008.

[2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In CVPR, pages 1597–1604, 2009.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. S¨usstrunk. Slic superpixels. EPEL, Tech. Rep, 149300, 2010.

[4] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In CVPR, pages 73–80, 2010.

[5] A. Borji and L. Itti. Exploiting local and global patch rarities for saliency detection. In CVPR, pages 478–485, 2012.

[6] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In ECCV, pages 414–429. 2012.

[7] N. Bruce and J. Tsotsos. Saliency based on information maximization. In NIPS, pages 155–162, 2006.

[8] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In CVPR, pages 409–416, 2011.

[9] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. TPAMI, 24(5):603–619, 2002.

[10] Y. Fu, J. Cheng, Z. Li, and H. Lu. Saliency cuts: An automatic approach to object segmentation. In ICPR, pages 1–4,2008.

[11] D. Gao, S. Han, and N. Vasconcelos. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. TPAMI, 31(6):989–1005, 2009.

[12] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In CVPR, pages 2376–2383, 2010.

[13] V. Gopalakrishnan, Y. Hu, and D. Rajan. Random walks on graphs for salient object detection in images. TIP, 19(12):3232–3242, 2010.

[14] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In CVPR, pages 1–8, 2008.

[15] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, pages 1–8, 2007.

[16] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research, 40(10-12):1489–1506, 2000.

[17] L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. Journal of Electronic Imaging, 10(1):161, 2001.

[18] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. TPAMI, 20(11):1254–1259, 1998.

[19] H. J., K. C., and P. P. Graph-based visual saliency. In NIPS, pages 545–552, 2006.

[20] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In ICCV, pages 2106–2113,2009.

[21] C. Jung and C. Kim. A unified spectral-domain approach for saliency detection and its application to automatic object segmentation. TIP, 21(3):1272–1283, 2012.

[22] C. Kanan, M. Tong, L. Zhang, and G. Cottrell. Sun: Topdown saliency using natural statistics. Visual Cognition, 17(6-7):979–1003, 2009.

[23] D. Kuettel and V. Ferrari. Fig.-ground segmentation by transferring window masks. In CVPR, pages 558–565, 2012.

[24] C. Lang, G. Liu, J. Yu, and S. Yan. Saliency detection by multi-task sparsity pursuit. TIP, 21(3):1327–1338, 2012.

[25] J. Li, Y. Tian, T. Huang, and W. Gao. Probabilistic multitask learning for visual saliency estimation in video. IJCV, 90(2):150–165, 2010.

[26] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. TPAMI, 33(2):353–367, 2011.

[27] Y. Ma and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. In ACMMM, volume 2, pages 374–381, 2003.

[28] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang. A generic framework of user attention model and its application in video summarization. TMM, 7(5):907–919, 2005.

[29] V. Mahadevan and N. Vasconcelos. Saliency-based discriminant tracking. In CVPR, pages 1007–1013, 2009.

[30] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. TPAMI, 32(1):171–177, 2010.

[31] V. Navalpakkam and L. Itti. Search goal tunes visual features optimally. Neuron, 53(4):605–617, 2007.

[32] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV,42(3):145–175, 2001.

[33] A. Oliva, A. Torralba,M. Castelhano, and J. Henderson. Topdown control of visual attention in object detection. In ICIP, volume 1, pages 253–256, 2003.

[34] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. Vision research, 42(1):107–124, 2002.

[35] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In CVPR, pages 733–740, 2012.

[36] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In CVPR, pages 853–860, 2012.

[37] A. Toshev, J. Shi, and K. Daniilidis. Image matching via saliency region correspondences. In CVPR, pages 1–8, 2007.

[38] Y.Wei, F.Wen,W. Zhu, and J. Sun. Geodesic saliency using background priors. In ECCV, 2012.

[39] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In ACMMM, pages 815–824, 2006.