Salient Object Detection in Low Contrast Images via Global Convolution and Boundary Refinement

Nan Mu[1], Xin Xu[1,2,*], Xiaolong Zhang[1,2]
[1]School of Computer Science and Technology, Wuhan University of Science and Technology, China
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, China
xuxin@wust.edu.cn

Abstract

Benefit from the powerful features created by using deep learning technology, salient object detection has recently witnessed remarkable progresses. However, it is difficult for a deep network to achieve satisfactory results in low contrast images, due to the low signal to noise ratio property, thus previous deep learning based saliency methods may output maps with ambiguous salient objects and blurred boundaries. To address this issue, we propose a deep fully convolutional framework with a global convolutional module (GCM) and a boundary refinement module (BRM) for saliency detection. Our model drives the network to learn the local and global information to discriminate pixels belonging to salient objects or not, thus can produce more uniform saliency map. To refine the localization and classification performance of the network, five GCMs are integrated to preserve more spatial knowledge of feature maps and enable the densely connections with classifiers. Besides, to propagate saliency information with rich boundary content, a BRM is embed behind each convolutional layer. Experiments on six challenging datasets show that the proposed model achieves state-of-the-art performance compared to nine existing approaches in terms of nine evaluation metrics.

1. Introduction

Salient object detection aims to locate the most distinctive objects in an image that are consistent with human visual perception. By serving as a preprocessing step, it facilitates a wide range of computer vision tasks such as image retrieval [1], semantic segmentation [2], human pose estimation [3] and person re-identification [4].

Although the deployment of deep convolutional neural network is beneficial to saliency detection and achieves significant improvements compared with the traditional hand-crafted features based approaches in recent years, it is still a challenging research problem when it comes to the low contrast images, the main reasons lie in: 1) the definition of saliency region is strongly impacted by the fuzzy visual identity and excessive noise of low contrast images, which mislead the deep models to predict the true salient object; 2) due to the repeated pooling operations in deep learning architectures, it is inevitable to lose object semantic and image structure information, which is severely missing in low contrast scenes, thus the results of deep networks usually suffer from inaccurate shape and poor localization of salient object; 3) since the pixels around the object boundaries are centered at the similar receptive fields and the deep networks only discriminate the binary labels of image pixels, it is difficult for a trained network to learn the boundary knowledge. As Fig. 1 shows, the state-of-the-art deep models [5-6] can hardly highlight the right salient objects and the real object boundaries in low contrast images.

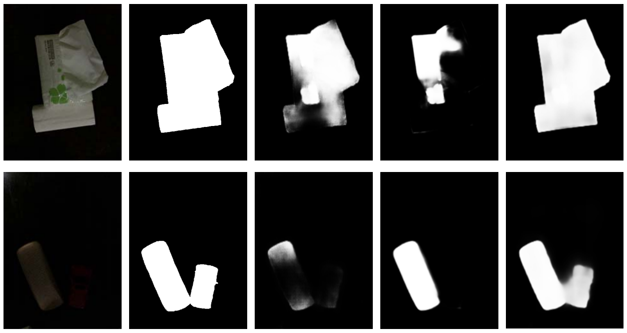

(a) Images (b) GT (c) NLDF (d) RAS (e) OURS

Figure 1. Examples of salient object detection in low contrast images. From left to right: input low contrast images, ground truth, saliency maps of two state-of-the-art models (NLDF [5] and RAS [6]), and our saliency maps.

In order to address these challenges, this paper proposes a deep fully convolutional neural network for salient object detection in low contrast images by adequately exploiting the complementary local and global information encoded in feature maps which are generated from different layers. To capture the global context, the features at multiple convolution layers are combined and compressed into a multi-level integrated global feature map. To gather the local context and promote features with strong local contrast, the multiscale local features of convolution layers, contrast features by local dissimilarity, and unpooled

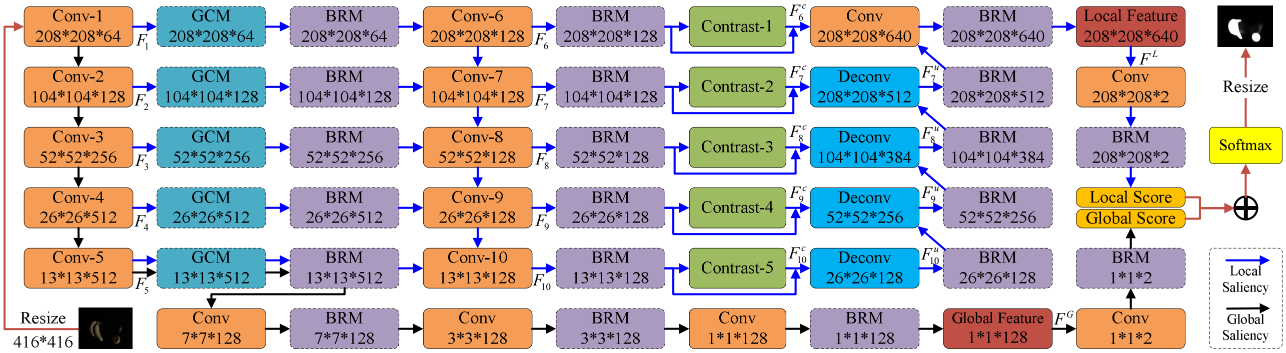


Figure 2. Overview of the proposed network for salient object detection in low contrast image.

features of deconvolution layers are concatenated into multi-level local feature maps. By fusing the local and global features via a softmax function, the saliency probability of each pixel can be computed exactly. The overview of the proposed network is illustrated in Fig. 2.

To effectively leverage the available visual information in low contrast image, the network is refined by two modules: a *global convolutional module* (GCM, shown in Fig. 3(a)) and a *boundary refinement module* (BRM, shown in Fig. 3(b)). GCM aims to expand the valid visual receptive field of feature maps and add dense connections with classifiers, which enables our new network to locate the most attentive regions and acquire more object semantic information with very few extra costs. BRM is a residual structure, which is embedded behind each convolutional layer of the network to keep the details of object boundaries.

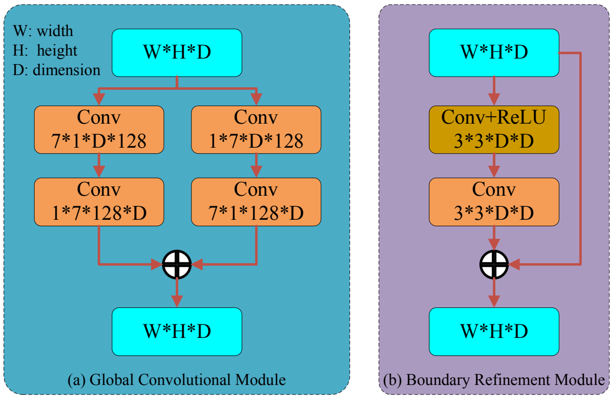


Figure 3. The structures of global convolutional module (GCM) and boundary refinement module (BRM).

To verify the effectiveness of our network, the proposed model is evaluated on six datasets and compared with nine state-of-the-art saliency models. Extensive experimental results demonstrate that our approach quantitatively and qualitatively outperforms other methods with respect to the accuracy of salient objects and the integrity of object boundaries. To summarize, the contributions of this work can be summarized as follows:

1) We found that the deep models are not enough for salient object detection task in low contrast images and formulated a deep fully convolutional encoder-decoder network in both local and global forms to learn saliency maps.

2) We embed a global convolutional module and a boundary refinement module into our network. The former focuses on the object semantic and spatial distribution of low contrast images to help better refine the structure and location information of salient objects. The latter helps maintain the completeness of object boundaries.

3) Compared with nine state-of-the-art saliency methods, the proposed model achieves superior performance on five benchmark datasets and our nighttime image dataset in terms of nine evaluation metrics.

2. Related Works

Generally, existing salient object detection methods can be mainly categorized into two streams: traditional hand-crafted models and deep learning based models.

Traditional hand-crafted models [7-12] mainly employed low-level visual features (e.g., color [9], contrast [10]) and heuristic priors [11-12] to distinguish salient objects from background. These features can be effective in simple scenarios, however they are of limited ability to represent salient objects in complex scenes and capture the object semantic knowledge, thus the traditional methods are incapable of producing satisfactory salient objects in challenging low contrast images. Therefore, it is necessary to consider high-level image information for salient object detection.

Deep learning based models [13-19] can learn the high-level semantic features from the training samples, which are more effective in locating the salient object in complex scenes and have achieved near human-level performance in various computer vision tasks. These models can be broadly classified into region-wise and pixel-wise saliency detection approaches. The region-wise methods predict saliency score by leveraging the image patch as basic processing unit. Zhao *et al.* [13] designed a multi-context deep learning framework by integrating the global and local context information of image superpixels. Li and Yu [14] estimated the saliency score of each superpixel by employing the multiscale features extracted from deep network. Wang *et al*. [15] incorporated the local estimation and global search to capture the saliency information of image patches and candidate objects. Since these methods treat each image patch as an independent unit, the spatial knowledge is easily lost in the training procedure. Moreover, to predict the saliency scores of all patches in the image, the processing procedure is repeated over and again, thus the computation is redundant and time-consuming. To overcome these drawbacks, the pixel-wise methods map an input image to a saliency map directly by exploiting a trained deep fully convolutional saliency network. Wang *et al*. [16] developed a recurrent fully convolutional network by introducing the saliency prior information to automatically refine the detection results. Hou *et al*. [17] introduced short connections to optimize the detailed structures of salient object by transforming features from deeper to shallower side-output layers. Liu *et al*. [18] presented a pixel-wise contextual attention network to select the local or global informative contexts to detect salient objects. Hu *et al*. [19] incorporated the recurrently aggregated deep features into a deep saliency network to improve the accuracy of salient objects. In general, these deep learning based saliency models can achieve desirable performance even when handling the image of complex scenarios. However, due to the lack of well-defined features to encode saliency information in low contrast images, the salient results of these models tend to lose some structure details and boundary parts of salient object and simultaneously contain many non-salient objects and background contents when directly merging multiple high-level features in the network. Inspired but differed from these deep models, we employ a global convolutional module and a boundary refinement module in a top-down manner to guide progressive saliency learning. Benefit from it, more semantic properties, accurate spatial information, and rich boundary knowledge can be learned, thus lead to significant improvement of salient object detection in low contrast images.

3. The Proposed Model

Targeting at mapping a low contrast image to a pixel-level saliency map, a deep fully convolutional framework is designed to combine the local and global saliency information. To accurately highlight the visual salient objects and further refine the object contours, five *global convolutional modules* (GCMs) and twenty *boundary refinement modules* (BRMs) are embedded into the deep network.

3.1. Network Architecture

The proposed model is built on the fully convolutional neural network as salient object detection framework and mainly utilize the pretrained VGG-16 net [20] as the feature extraction network. The whole architecture is composed of convolution and deconvolution layers with different output dimensions, which can make our model possess the capability of capturing the local and the global features from various resolutions.

Given an image I, which is resized to 416×416 as the input unit, five feature maps $\{F_1, F_2, F_3, F_4, F_5\}$ are generated from the five convolution blocks (Conv-1 to Conv-5), each block has a kernel size 3×3 and contains a max pooling operation of stride 2 to decrease the spatial resolution from 208×208 to 13×13. Inspired by global convolutional network [21], we proposed a GCM to enable densely connections between convolution blocks and features, which makes the obtained features have more abundant neural information and robust to locally disturbances. We also utilized a BRM to connect each convolution block to preserve the boundary information.

To calculate the global feature map (denoted as F^G), convolution features are gathered to capture the global context information before assigning the saliency knowledge to small region. Three convolutional layers of 128 features channels are added after Conv-5 block to change F_5 into 1×1 resolution to compute the global feature. The kernel sizes of the three layers are 7×7, 5×5, and 3×3, respectively.

To calculate the local feature maps (denoted as F^L), five convolutional blocks (Conv-6 to Conv-10), of which the kernel sizes are 3×3 and 128 channels, are connected to the processing blocks (Conv-1 to Conv-5), respectively. The multi-scale local feature maps $\{F_6, F_7, F_8, F_9, F_{10}\}$ are obtained by these convolutional layers. Since saliency value is the difference between foreground object and its surrounding background. The contrast feature (denoted as F_i^c, $i = 6,\cdots,10$) of each feature map is computed by measuring the dissimilarity between F_i and its local average [5].

$$F_i^c = F_i - F_i', \tag{1}$$

where F_i' is the output of F_i by local average pooling operation with kernel size 3×3. The deconvolution block is connected to each feature map to increase its spatial scale by upsampling of stride 2 and kernel size 5×5. The resulting unpooled feature map (denoted as F_i^u) is produced by concatenating its local feature F_i, the local contrast feature F_i^c, and the unpooled feature map F_{i+1}^u.

$$F_i^u = \mathrm{Upsamp}(\mathrm{CAT}(F_i, F_i^c, F_{i+1}^u)). \qquad (2)$$

The final local feature map F^L is generated by using a convolution layer of kernel size 1×1, which combines the information of local feature F_6, local contrast feature F_6^c, and unpooled feature map F_7^u via concatenation.

$$F^L = \mathrm{Conv}(\mathrm{CAT}(F_6, F_6^c, F_7^u)). \qquad (3)$$

3.2. Global Convolutional Module

By considering densely connections between classifiers and feature maps, GCM improves the classification ability of the proposed saliency model, which allows the network to handle various types of transformations. Meanwhile, the large kernel of GCM is helpful for the feature map to encode more spatial information, which enhances the localization precision of the salient objects.

As illustrated in Fig. 3(a), the proposed GCM has two branches, the left convolution operation consists of a 7×1 convolutional block followed by a 1×7 convolutional block, the right one employs the 1×7 and 7×1 convolutions. These two branches are combined to enable the densely connections have a large 7×7 region in the feature map, which increases the validity of receptive field. Besides, the calculation cost of GCM structure is relatively low, which is more practical.

3.3. Boundary Refinement Module

To further improve the spatial location accuracy of the salient objects, BRM is designed to optimize the localization performance near the boundaries of salient objects, which can greatly preserve the boundary information in training stage.

As shown in Fig. 3(b), BRM is modeled as a residual structure [22], it has one branch to connect the input and output layers directly without any operation. The other branch is the residual net, which contains two convolution blocks of a 3×3 kernel size. These two branches are combined by a shortcut connection, which is beneficial to learn the boundary information, thus the score of boundary pixel can be refined. The dimension of the corresponding output is the same as the input.

3.4. Salient Object Detection

The final saliency map is computed by combining the global feature map F^G and local feature map F^L. Let S^M denotes the saliency map and G^T denotes the ground truth, the probability P of a pixel p in the feature map belonging to the salient object or not can be predicted by the softmax function.

$$S^M(p) = P(G^T(p) = l) = \frac{e^{w_l^L F^L(p) + v_l^L + w_l^G F^G + v_l^G}}{\sum_{l' \in \{0,1\}} e^{w_{l'}^L F^L(p) + v_{l'}^L + w_{l'}^G F^G + v_{l'}^G}}, \qquad (4)$$

where (w^L, v^L) and (w^G, v^G) are the linear operators. The loss function of our network is the sum of cross entropy loss (denoted as $Loss^{CE}$) and boundary loss (denoted as $Loss^B$) via:

$$Loss = \sum_r \alpha_r \int_{p\in\Omega_r} (Loss^{CE}) + \sum_r \beta_r (1 - Loss^B), \qquad (5)$$

where α_r and β_r are the positive weighting constants to tune the $Loss^{CE}$ and $Loss^B$.

The cross entropy loss $Loss^{CE}$ between ground truth G^T and obtained saliency map S^M of pixels p inside region Ω_r is defined as:

$$Loss^{CE} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{l\in\{0,1\}} (G^T(p) = l)(\log(S^M(p_j) = l)). \qquad (6)$$

The boundary loss $Loss^B$ between pixels p on the true boundary map B^T and estimated boundary map B^M is computed as:

$$Loss^B = 1 - \frac{2\,|B_r^T \cap B_r^M|}{|B_r^T| + |B_r^M|}. \qquad (7)$$

The supervision of the proposed model combines the effect of $Loss^{CE}$ and $Loss^B$. Since two loss functions jointly train our model, the parameters for pinpointing the salient objects and refining the boundary can be optimized.

4. Experimental Results

Extensive experiments are conducted on six salient object detection datasets to evaluate the performance of our model against nine state-of-the-art saliency models.

4.1. Experimental Setup

Evaluation datasets. We test the proposed model on five public benchmarks datasets and a low contrast image dataset. 1) MSRA-B dataset [23], which contains 5000 images, and most of the images only have one salient object. 2) DUT-OMRON dataset [24], which includes 5168 images of complex background. 3) PASCAL-S dataset [25], which has 850 challenging natural images. 4) HKU-IS dataset [14], which provides 4447 images of multiple salient objects with overlapping boundaries and of low contrast. 5) DUTS dataset [26], which has a training set of 10553 images and a test set of 5019 images. Both these images are of complex scenarios. 6) *Nighttime Image* (NI) dataset, which is proposed by us and contains 1000 images captured in the dark evening with a stand camera. The resolution of these image is 500×667.

Evaluation models. The proposed salient object detection framework is compared with five traditional hand-crafted saliency models and four deep learning based saliency models, including: *context-aware* (CA) model [27], *saliency optimization* (SO) model [28], *bootstrap learning* (BL) model [29], *structured matrix decomposition* (SMD) model [30], *multiple instance learning* (MIL) model [31], *non-local deep features* (NLDF) model [5], *learning to promote saliency* (LPS) model [32], *contour to saliency* (C2S) model [33], and *reverse attention saliency* (RAS) model [6].

Evaluation criteria. To estimate the performance of the proposed model with other models, nine evaluation metrics are utilized, including:

1) The *true positive rates* and *false positive rates* (TPRs-FPRs) curve. $TPR = TP/(TP+FN)$ corresponds to the ratio of salient pixels which are correctly detected to all salient pixels, and $FPR = FP/(FP+TN)$ is computed as ratio of falsely detected salient pixels to all true non-salient pixels, where TP (true positive) and FN (false negative) are the sets of correctly detected salient pixels and falsely detected non-salient pixels, respectively. FP (false positive) and TN (true negative) are the falsely detected salient object pixels and correctly detected non-salient pixels, respectively.

2) The *precision-recall* (PR) curve. $P = TP/(TP+FP)$ is defined as the ratio of correctly detected salient pixels to all detected salient pixels, $R = TP/(TP+FN)$ is the same as TPR, which measures the comprehensiveness of the detected salient pixels.

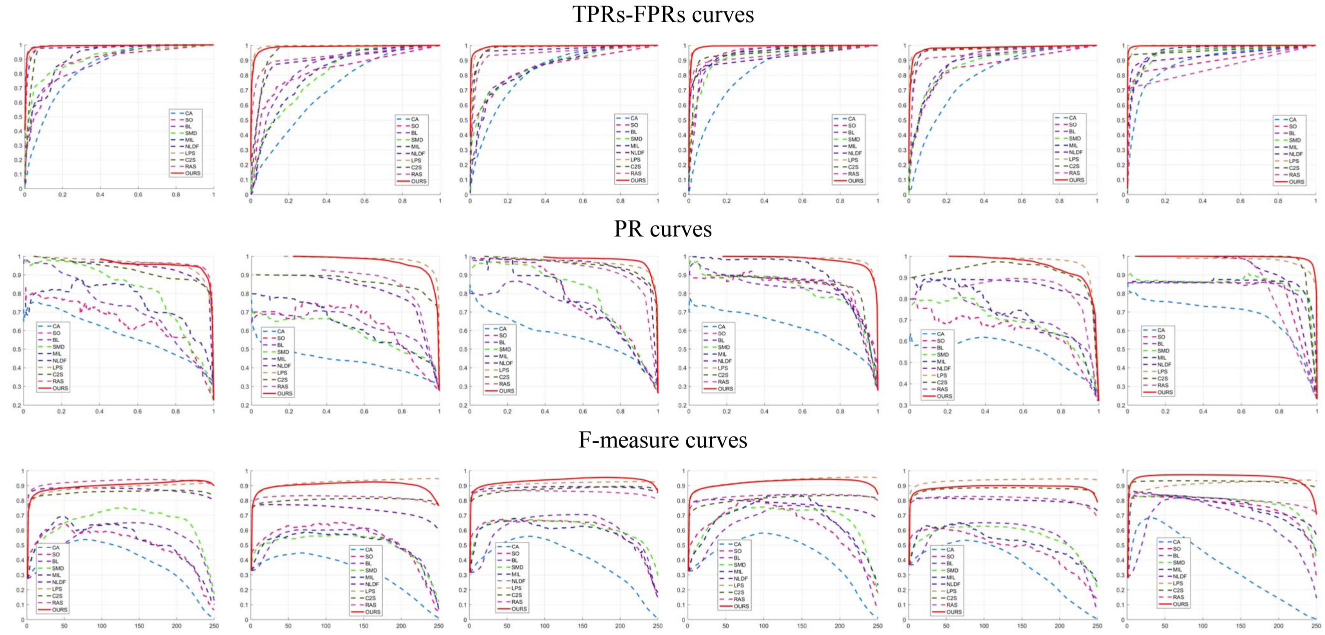
3) F-measure curve. $F_\beta = (1+\beta^2)P \cdot R/(\beta^2 \cdot P + R)$ is computed as a weighted harmonic mean of P and R, where β^2 is set to 0.3 to emphasize the effect of P. The F-measure curve is created by comparing the ground truth with the binary saliency maps which are obtained by varying the threshold to determine whether a pixel belongs to salient object.

4) The *area under the curve* (AUC) score, which is defined as the percentage of areas under the TPRs-FPRs curve, it gives an intuitive indication of how well a saliency map predicts the true salient objects.

5) The *mean absolute error* (MAE) score, which is computed as the average absolute difference between the resulting saliency map S^M and the ground truth G^T as: $\text{MAE} = \text{mean}(|S^M - G^T|)$. The smaller MAE value indicates higher similarity between S^M and G^T.

6) The *weighted F-measure* (WF) score [34], which is calculated by introducing a weighted P to measure the exactness and a weighted R to measure the completeness.

7) The *overlapping ratio* (OR) score, which is defined as the ratio of overlapping salient pixels between the binary saliency map S^{BM} and ground truth G^T via: $\text{OR} = |S^{BM} \cap G^T| / |S^{BM} \cup G^T|$. The OR score considers the completeness of salient pixels and the correctness of non-salient pixels.



(a) MSRA (b) DUT-OMRON (c) PASCAL-S (d) HKU-IS (e) DUTS (f) NI

Figure 4. Performance comparisons of the proposed model (the red solid line) with other state-of-the-art saliency models on six datasets.

8) The *structure-measure* (S-M) score, which measures the region-aware and object-aware structural similarity between the saliency map S^M and ground truth G^T [35].

9) The average execution time per image (in second). The experiments of the first five traditional models are executed using MATLAB software on an Intel Core i5-5250 CPU (1.6GHz) PC with 8 GB RAM. The last four deep learning models and the proposed model are tested on a NVIDIA TITAN Xp GPU and Intel Xeon E5-2620 CPU (2.4GHz) processor with 64 GB RAM.

Baseline methods. To demonstrate the advantages of our configuration and evaluate the contribution of different terms, three baselines are designed to train the saliency model respectively. 1) Baseline 1, which only use the full convolutional encoder-decoder network for salient object detection by incorporating the local and global cues. We remove the GCM and BRM from our model and simply train a VGG16-based network to generate saliency maps. This baseline test will reflect the saliency performance without GCM and BRM. 2) Baseline 2, which embeds GCM into the base network and does not take the boundary refinement into account. This baseline can represent the importance of GCM to locate the true salient object. 3) Baseline 3, which embeds BRM into the base network and does not include the global convolutional module. This baseline is used to show the benefit of BRM to the saliency detection result.

The performance of three baselines is compared with the proposed model. We train the three baselines and our model on MSRA-B dataset. The training set contains 2500 images

Table 1. Quantitative results of various saliency models on six datasets. The best three scores are shown in red, blue and green colors, respectively. The up-arrow ↑ indicates the lager value achieved, the better performance is. The down-arrow ↓ has the opposite meaning.

Dataset	Criteria	Traditional hand-crafted saliency models					Deep learning saliency models				The proposed saliency model			
		CA	SO	BL	SMD	MIL	NLDF	LPS	C2S	RAS	Baseline1	Baseline2	Baseline3	OURS
(a) MSRA-B	AUC↑	0.7635	0.7687	0.7861	0.8017	0.7943	0.8486	0.8638	0.8584	***0.8737***	0.8655	**0.8671**	0.8565	**0.8746**
	MAE↓	0.2546	0.1780	0.2391	0.1686	0.1842	0.0675	**0.0422**	0.0662	**0.0311**	0.0544	0.0610	0.0644	***0.0373***
	WF↑	0.3492	0.4678	0.4319	0.5435	0.4770	0.8246	**0.8786**	0.8309	**0.9149**	0.8482	0.8335	0.8262	***0.8904***
	OR↑	0.3740	0.4765	0.5333	0.5825	0.6080	0.7610	**0.8428**	0.7966	**0.8883**	0.8195	0.7921	0.7969	***0.8599***
	S-M↑	0.5967	0.6644	0.6822	0.7496	0.6952	0.8687	**0.9081**	0.8744	**0.9359**	0.8953	0.8905	0.8704	***0.9196***
	TIME↓	66.6317	0.5899	34.4369	6.9224	110.019	0.0444	0.0413	**0.0327**	0.0390	0.0602	0.0689	0.1421	0.1523
(b) DUT-OMRON	AUC↑	0.6510	0.7147	0.7175	0.6794	0.6790	0.7480	**0.8513**	0.7985	0.7762	**0.8138**	0.7924	0.7850	***0.8427***
	MAE↓	0.3253	0.2118	0.2639	0.2544	0.2544	0.1459	**0.0364**	0.1168	**0.0962**	0.1216	0.1078	0.1143	***0.0573***
	WF↑	0.3025	0.4793	0.4534	0.4401	0.4429	0.6590	**0.9107**	**0.7661**	0.7639	0.7100	0.7374	0.7241	***0.8701***
	OR↑	0.2401	0.4679	0.4275	0.4038	0.4213	0.5917	**0.8842**	**0.7151**	0.6997	0.6625	0.6946	0.6872	***0.8411***
	S-M↑	0.5032	0.6487	0.6556	0.6238	0.6214	0.7475	**0.9300**	**0.8250**	0.8116	0.8032	0.8130	0.8040	***0.9082***
	TIME↓	61.0279	0.5512	25.6283	5.9705	110.360	0.0454	0.0420	**0.0352**	0.0354	0.0600	0.0694	0.1432	0.1507
(c) PASCAL-S	AUC↑	0.7251	0.7371	0.7486	0.7421	0.7368	0.8239	***0.8500***	0.8435	0.8010	0.8353	**0.8487**	0.8421	**0.8517**
	MAE↓	0.2726	0.1866	0.2272	0.1852	0.2045	0.0640	***0.0413***	0.0704	0.0984	0.0701	**0.0506**	0.0562	**0.0356**
	WF↑	0.3522	0.5464	0.4840	0.5368	0.5102	0.8375	***0.8989***	0.8441	0.7865	0.8264	**0.8747**	0.8545	**0.9090**
	OR↑	0.2912	0.5300	0.5305	0.5411	0.5103	0.7750	***0.8757***	0.8168	0.6956	0.7867	**0.8415**	0.8275	**0.8886**
	S-M↑	0.5882	0.6924	0.7011	0.6955	0.6724	0.8683	***0.9146***	0.8799	0.8160	0.8709	**0.9077**	0.8937	**0.9322**
	TIME↓	73.7946	1.0266	35.7323	8.7189	169.560	0.0449	0.0431	**0.0343**	0.0477	0.0603	0.0692	0.1433	0.1509
(d) HKU-IS	AUC↑	0.7608	0.8126	0.8177	0.8036	**0.8199**	0.7589	***0.8505***	0.8099	0.8013	0.7917	0.8081	0.7922	**0.8509**
	MAE↓	0.2900	0.1615	0.2245	0.1687	0.1665	0.1025	**0.0281**	0.0794	**0.0615**	0.0875	0.0717	0.0892	***0.0432***
	WF↑	0.3767	0.5677	0.4850	0.5758	0.5655	0.7030	**0.9231**	0.7878	**0.8086**	0.7693	0.7881	0.7705	***0.8988***
	OR↑	0.2853	0.5324	0.4787	0.5652	0.5765	0.6216	***0.7955***	0.7054	**0.7534**	0.7158	0.7388	0.7017	**0.8399**
	S-M↑	0.5859	0.7448	0.7232	0.7522	0.7734	0.7759	**0.9365**	0.8378	**0.8585**	0.8274	0.8547	0.8194	***0.9250***
	TIME↓	60.0477	0.5421	22.9599	7.2008	103.780	0.0441	0.0417	**0.0340**	0.0375	0.0601	0.0683	0.1428	0.1496
(e) DUTS	AUC↑	0.7201	0.7123	0.7747	0.7489	0.7403	0.7913	***0.8128***	**0.8117**	0.7754	0.8101	0.8015	0.8070	**0.8130**
	MAE↓	0.3307	0.2438	0.2636	0.2386	0.2465	0.1096	**0.0263**	**0.0817**	0.1211	0.0905	0.1037	0.0828	***0.0595***
	WF↑	0.3208	0.4322	0.4660	0.4778	0.4433	0.7412	**0.9203**	**0.8386**	0.7390	0.7852	0.7740	0.7922	***0.8521***
	OR↑	0.2559	0.3709	0.3966	0.3647	0.3713	0.5518	**0.5911**	***0.5918***	0.5747	0.5361	0.5097	0.5425	**0.5974**
	S-M↑	0.5202	0.5821	0.6514	0.6490	0.6127	0.7899	**0.9082**	0.8472	0.7917	0.8364	0.8170	**0.8477**	***0.8761***
	TIME↓	85.6488	0.5493	31.8590	5.4786	131.010	0.0444	0.0418	**0.0344**	0.0362	0.0604	0.0693	0.1435	0.1516
(f) NI	AUC↑	0.7470	0.7730	0.8224	0.7784	0.7856	0.7677	**0.8744**	**0.8435**	0.7422	0.8336	0.8407	0.8251	***0.8711***
	MAE↓	0.2002	0.0851	0.1569	0.0923	0.1075	0.1030	**0.0372**	**0.0237**	0.1057	0.0791	0.0739	0.0703	***0.0297***
	WF↑	0.3461	0.7059	0.5513	0.7240	0.7126	0.6659	**0.8916**	***0.9005***	0.6916	0.7668	0.7772	0.7678	**0.9219**
	OR↑	0.4860	0.6905	0.6741	0.6659	0.7352	0.6241	***0.8807***	**0.8778**	0.5976	0.7637	0.7896	0.7661	**0.9168**
	S-M↑	0.5624	0.7804	0.7681	0.7873	0.7876	0.7584	**0.9257**	***0.9285***	0.7330	0.8553	0.8638	0.8394	**0.9519**
	TIME↓	95.8000	2.4464	29.7269	11.8980	437.993	0.0439	0.0528	**0.0342**	0.0704	0.0608	0.0687	0.1408	0.1510

and the validation set contains 500 images, which are combined to train our four models. For the proposed model, it takes about eight hours to finish the whole training procedure for 10 epochs. The trained model was used to obtain saliency maps of all the six datasets. In comparison, the training sets of NLDF and RAS models are same as us, the training sets of LPS and C2S models contain 10K and 30K images, respectively.

4.2. Experimental Analysis

The proposed saliency model is compared with the mentioned nine state-of-the-art models on the six datasets in terms of nine metrics. The quantitative comparison results are demonstrated in Fig. 4 and Table 1, which show that the proposed model achieves the best or second best performance in most cases, since our saliency maps are much closer to the ground truth. Because of our local-global strategy, the salient objects have high contrast between the background. The GCM and BRM also refine the structure and boundary information effects of the salient regions. Thus, our model is efficient to the salient object detection task in low contrast images.

On MSRA-B dataset (Fig. 4(a) and Table 1(a)), most images have single object and simple background, the proposed model obtains the best performance on TPRs-FPRs curve and AUC score, while RAS model performs the best in term of other metrics (except Time). Our Baseline 2 is superior, which achieves the second best AUC score.

On DUT-OMRON dataset (Fig. 4(b) and Table 1(b)), the images have large complexity and diversity, which leads to poor performance of all the models in comparison with MSRA-B. The proposed model performs the second best in terms of all the metrics, with a small gap to best results of LPS model. This is because LPS takes a large number of images (three times over ours) as training to improve its robustness. Our Baseline 1 can achieve comparable performance on AUC score.

On PASCAL-S dataset (Fig. 4(c) and Table 1(c)), the proposed model has competitive performance compared with other models, which performs the best in terms of all the metrics (except Time). Beyond that, our Baseline 2 ranks the third on these criteria.

On HKU-IS and DUTS datasets ((d) and (e) of Fig. 4 and Table 1), most of the images have relative complex background, the proposed model achieves the best performance on TPRs-FPRs, AUC, and OR metrics. Besides, the PR curve, F-measure curve, MAE score, WF score and S-M score of our model rank the second, which are slighter than the best results obtained by LPS model. Our Baseline 3 ranks the third in terms of S-M score on DUTS dataset.

On NI dataset (Fig. 4(f) and Table 1(f)), the proposed model obtains the best results on PR curve, F-measure curve, WF score, OR score and S-M score. In terms of AUC and MAE scores, our model achieves the second best, which only have small differences (0.0033 and 0.006) to the best results of LPS and C2S models. Specifically, the C2S model takes about 0.03 second to generate a saliency map, which is the most efficient one on the six datasets.

These objective performance comparisons indicate that the method has strong potential in detecting the most salient object of complex environments. Visual comparisons of saliency maps using different baselines on NI dataset are provided in Fig. 5. It can be seen that the salient objects of Baseline 2 have accurate shape than Baseline 1, meanwhile the Baseline 3 preserves the smooth boundaries of the salient objects. Comparing with the three Baselines, the proposed model accurately detects the complete salient objects and produces coherent boundaries, it can be learnt that the local-global measure, GCM and BRM bring lots of advantages to the saliency results.

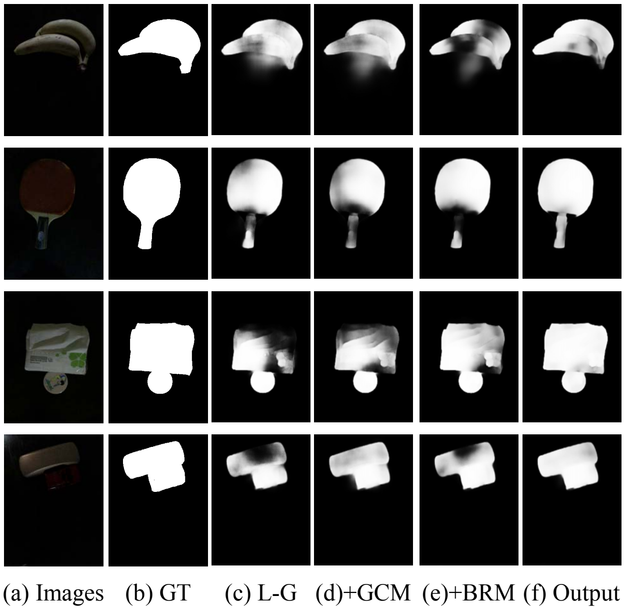


Figure 5. Saliency results of NI dataset. (a) input images, (b) ground truth, (c) Baseline 1, which only computes the local-global saliency, (d) Baseline 2, which embeds GCM into the network, (e) Baseline 3, which embeds BRM into the network, (f) the saliency maps of the proposed model.

The qualitative comparisons between the proposed model and other saliency models on six datasets are shown in Fig. 6. It can be seen that the deep learning based models consistently outperform the traditional models, the saliency maps of deep models are closer to the ground truth. For the images of simple scenarios, most approaches can achieve desirable results, while the proposed model has the best result which suppresses most of the background. For the complex images, some competing deep learning methods

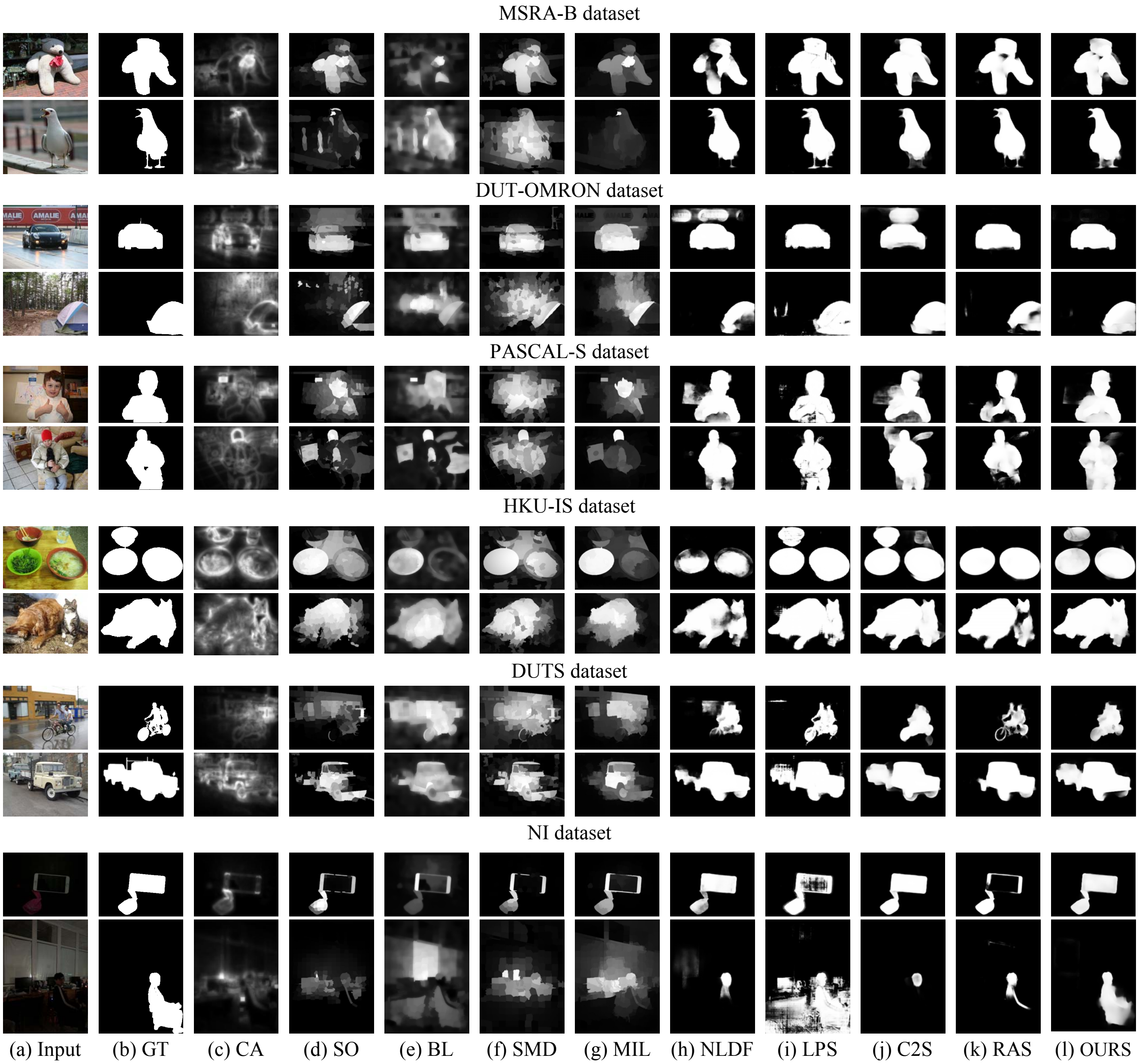


Figure 6. Visual comparisons of saliency maps produced by various models on six datasets.

fail to identify the entire salient object. By contrast, our model can pop out the whole salient object accurately. For the low contrast images, most models can hardly locate the right salient objects, while the proposed model captures the true salient regions successfully. These results illustrate the effectiveness and robustness of the proposed salient object detection network in different challenging scenes.

5. Conclusions

In this paper, a deep fully convolutional network is proposed to integrate local and global features for salient object detection in low contrast images. A global convolutional module and a boundary refinement module are developed and embedded into our base network to gradually generate finer details and completeness boundary information. As a result, more discriminative features can be obtained for more accurate salient object detection. Experimental results on six datasets show that the proposed model outperforms the state-of-the-art approaches and has great potential for other computer vision tasks in low contrast images.

Acknowledgements. This work was supported by the Natural Science Foundation of China (61602349, U1803262, 61440016 and 61273225).

References

[1] X. Yang, X. Qian, and Yao Xue. Scalable mobile image retrieval by exploring contextual saliency. *TIP*, 24(6):1709–1721, 2015.

[2] W. Shimoda and K. Yanai. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In *ECCV*, pages 218–234, 2016.

[3] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang. Multi-context attention for human pose estimation. In *CVPR*, pages 1831–1840, 2017.

[4] R. Zhao, W. Oyang, and X. Wang. Person re-identification by saliency learning. *TPAMI*, 39(2):356–370, 2017.

[5] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, and P.-M. Jodoin. Non-local deep features for salient object detection. In *CVPR*, pages 6593–6601, 2017.

[6] S. Chen, X. Tan, B. Wang, and X. Hu. Reverse attention for salient object detection. In *ECCV*, pages 236–252. 2018.

[7] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, pages 1155–1162, 2013.

[8] J. Kim, D. Han, Y.-W. Tai, and J. Kim. Salient region detection via high-dimensional color transform. In *CVPR*, pages 883–890, 2014.

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

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

[11] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, pages 2083–2090, 2013.

[12] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu. Global contrast based salient region detection. *TPAMI*. 37(3):569–582, 2015.

[13] R. Zhao, W. Ouyang, H. Li, and X. Wang. Saliency detection by multi-context deep learning. In *CVPR*, pages 1265–1274, 2015.

[14] G. Li and Y. Yu. Visual saliency based on multiscale deep features. In *CVPR*, pages 5455–5463, 2015.

[15] L. Wang, H. Lu, X. Ruan, and M.-H. Yang. Deep networks for saliency detection via local estimation and global search. In *CVPR*, pages 3183–3192, 2015.

[16] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan. Saliency detection with recurrent fully convolutional networks. In *ECCV*, pages 825–841, 2016.

[17] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr. Deeply supervised salient object detection with short connections. In *CVPR*, pages 5300–5309, 2017.

[18] N. Liu, J. Han, and M.-H. Yang. PiCANet: Learning pixel-wise contextual attention for saliency detection. In *CVPR*, pages 3089–3098, 2018.

[19] X. Hu, L. Zhu, J. Qin, C.-W. Fu, and P.-A. Heng. Recurrently aggregating deep features for salient object detection. In *AAAI*, pages 6943–6950, 2018.

[20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arxiv:1409.1556*, 2014.

[21] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun. Large kernel matters-improve semantic segmentation by global convolutional network. In *CVPR*, pages 1743–1751, 2017.

[22] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[23] 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.

[24] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, pages 3166–3173, 2013.

[25] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *CVPR*, pages 280–287, 2014.

[26] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, pages 136–145, 2017.

[27] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *TPAMI*, 34 (10): 1915–1926, 2012.

[28] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *CVPR*, pages 2814–2821, 2014.

[29] N. Tong, H. Lu, and M. Yang. Salient object detection via bootstrap learning. In *CVPR*, pages 1884–1892, 2015.

[30] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank. Salient object detection via structured matrix decomposition. *TPAMI*, 39(4):818–832, 2017.

[31] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan. Salient object detection via multiple instance learning. *TIP*, 26(4):1911–1922, 2017.

[32] Y. Zeng, H. Lu, Li. Zhang, M. Feng, and A. Borji. Learning to promote saliency detectors, In *CVPR*, pages 1644–1653, 2018.

[33] X. Li, F. Yang, H. Cheng, W. Liu, and D. Shen. Contour knowledge transfer for salient object detection. In *ECCV*, pages 355–370, 2018.

[34] R. Margolin, L. Zelnik-Manor, and A. Tal. How to evaluate foreground maps? In *CVPR*, pages 248–255, 2014.

[35] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, pages 4548–4557, 2017.