Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Salient Object Detection by Pyramid Networks with Gating

Mingxin Zhang[1,2], Zhijie Wang[1], Tiezhu Sun[1] and Xiaolei Li[*1]
1 School of Control Science and Engineering, Shandong University, Jinan, 250000, China
2 Institute of Space Science and Technology, Nanchang University, Nanchang 330031,China

***Abstract*— In the paper, we propose a salient object detection architecture named as Gating for Double Pyramid Networks (GDPNet). It consists of two pyramid structures: Feature Pyramid Network (FPN) and Pyramid Pooling Module (PPM). FPN has the capability of capturing the inherent multi-scale and pyramid hierarchy, while PPM can exploit the global context information by different-region-based context aggregation. It is known that the irrelevant information corresponding to non-salient objects or background may deteriorate the performance of the model. We introduce two gating strategies, i.e., Cross-Gating for FPN and Single-Gating for PPM, to suppress the incurred irrelevant information in hidden features. Our proposed approach achieves state-of-the-art performance on five benchmark datasets, which demonstrates the effectiveness and robust feature extraction capability of proposed GDPNet.**

***Index Terms*— Convolutional Neural Network; Feature Pyramid Network; Salient Object**

I. Introduction

Differing from other object detection tasks [1], salient object detection is to identify and isolate the most distinctive objects or regions from vision perspective given an input picture by mimicking the human visual system. Differing from other segmentation-like tasks, such as semantic segmentation, salient object detection focuses only on a few predominant objects of interest and attraction, which has proved valuable for various computer vision-based tasks, such as scene parsing [2], weakly supervised semantic segmentation [3], and defect detection in industrial images [4].

Traditional salient object detection methods need a variety of hand-crafted features to employ either global or local cues [5]. However, these features are usually extracted from some specified datasets, thereby hard to be extended to all cases. Recently, deep learning has made great breakthroughs in various computer vision tasks and shows much superiority over hand-crafted features [6]. Inspired by the end-to-end model FCNs [7], researchers have spent much effort to adapt it to detect salient objects and obtained significant performance gains.

However, classical FCNs normally compute features from a single scale, which may affect the accuracy of salient object detection. Recently, the investigation of exploiting multi-scale representation for salient object detection becomes more and more popular. Luo *et al.* presented a convolutional neural network called NLDF [8], which tried to combine local and global information through a multi-resolution grid structure. Hou *et al.* proposed to further improve the performance by including short connections, skip-layers and horizontal cascade [9]. Although previous work achieved impressive performance, they lack the consideration of contextual information, which has proved to be effective in semantic segmentation task [10].

Based on these considerations, we propose a new architecture named Gating for Double Pyramid Networks (GDPNet) for salient object detection. The proposed method mainly consists of two pyramid structures, the first one is Feature Pyramid Network (FPN), which can exploit the multi-level features from CNN, the second one is Pyramid Pooling Module (PPM), it can capture more global representation about the input image, we can capture more global representation and finer boundary details about the input image by concatenating these two module. Besides, to filter out the irrelevant information that may exist in FPN and PPM, we further design two gating strategies for the proposed architecture: Cross-Gating and Single-Gating. Experiments show that the gating mechanism does bring performance improvement and help our model achieve much better performance than the existing methods.

II. Proposed Method

The proposed Gating for Double Pyramid Networks (GDPNet) is illustrated in Fig. 1, The FPN and PPM are used to extract features and the gating module, called Cross-Gating and Single-Gating, will filter out the irrelevant information of salient object detection task. In the following, we will first introduce the whole network structure, then detail the gating mechanisms.

As the two backbones of GDPNet, FPN and PPM play different roles in our saliency detection model. Let us first see FPN, which aims to make use of a CNN's hierarchical pyramid feature map, which takes possession of both low-level semantics and high-level ones, and construct a pyramid of feature map with different-level semantics. The construction of the pyramidal feature involves a bottom-up flow, a top-down flow, and skip connections. The bottom-up flow is to compute the feature extracted by the backbone network, which predicts hierarchical feature maps at a variety of scales. The top-down flow hallucinates features with larger resolution from top pyramidal levels. The feature maps extracted by bottom-up flow can be refined by skip connections.

*Corresponding author: Xiaolei Li(qylxl@sdu.edu.cn).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

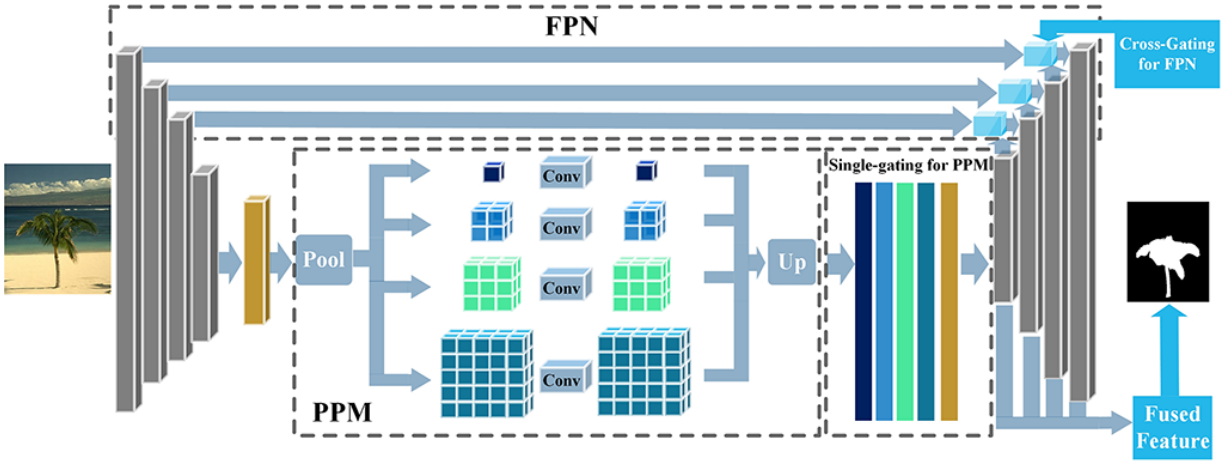

Fig. 1. The proposed GDPNet mainly consists of two pyramid structures: Feature Pyramid Network (FPN) and Pyramid Pooling Module (PPM), which adopt two gating mechanisms, Cross-Gating and Single-Gating, to filter out the irrelevant information.

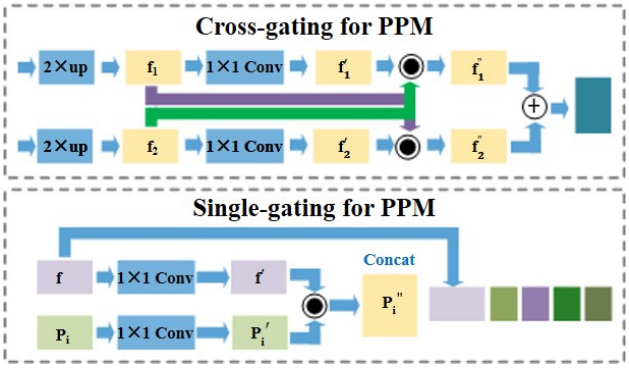

Fig. 2. The diagrams of Cross-Gating and Single-Gating.

The features outputted from the bottom-up flow and top-down flow with equal dimension will be fused together. The feature from the bottom-up flow contains elementary semantics. However, the activation values of it are more precisely computed for the reason that fewer subsampled operations are implemented.

PPM merges feature maps at four diverse pyramidal scales. The roughest scale is global pooling which outputs a single bin output. The next pyramidal scale splits the feature into diverse sub-regions and constitutes pooled feature map for diverse locations. The results of diverse scales in the PPM include the features that have different sizes. To retain the weights of global features, PPM uses convolutional layers after pyramidal pooling to reduce the dimension of context representation. By means of bilinear interpolation, the low-dimension features are directly upsampling to obtain the feature that has the same size with the input feature. Last, varied scales of feature maps are combined as the final feature map of PPM.

The proposed salient object detection architecture is built +on FPN, and PPM embedded in FPN. PPM is connected to the last layer of the bottom-up pathway and then is fed into the top-down branch in FPN. If different layers produce same size features, we call these layers in the same stage. We denote the final features of each stage in ResNets as $\{C_2, C_3, C_4, C_5\}$ and the features output by FPN as $\{P_2, P_3, P_4, P_5\}$, where P_5 is the result after pyramid pooling module. In the salient object detection task, we find merging total features in FPN is superior to just utilizing the feature that has the highest resolution. Therefore, there is a fusion process before the model predicts the final saliency result. In order to merge the various features such as $\{P_2, P_3, P_4, P_5\}$, we resize them to the size of P_2 and combine these features. We then adopted a convolution layer to merge the features with varied scales and to reduce channel dimensions at the same time.

As mentioned before, to filter out the irrelevant information from the non-salient objects or backgrounds, we propose a Cross-Gating for FPN and a Single-Gating for PPM, which are detailed as follows.

A Cross-Gating mechanism (as shown in Fig. 2) is introduced into FPN, to filter out the irrelevant information and emphasize the relevant parts. In Cross-Gating, the gate for the feature f_1 depends on the feature f_2. Meanwhile, the feature f_2 is also gated by f_1. Specifically, f_1 goes through a 1×1 convolutional layer to be transformed to $f_1^{'}$. The 1×1 convolutional layer learns the gating parameters and $f_1^{'}$ is the learned gate for gating f_2. Then $f_1^{'}$ times f_2 with dot product to accomplish the final gating operation. The f_1 should be gated by f_2 in the same way. After Cross-Gating, the gated features should only maintain the relevant information.

In the original FPN, the upsampling map P_i is directly merged with the corresponding bottom-up map C_i by

element-wise addition. In our proposed architecture, we add the Cross-Gating operation before the element-wise addition. With Cross-Gating, multi-scale features in shallow and deep layers can be gated by each other. Benefited by Cross-Gating, the final fused feature maintains the key relevant information in multi-scale features so that the predicted saliency map can classify the salient object and background accurately.

In Cross-Gating, the two features should share the same dimension so that they can be operated with dot product. In PPM, the final features have different dimensions, which indicates that Cross-Gating is not compatible with them anymore. However, there is still irrelevant information in these features, which needs to be gated out before the concatenation. Hence, Single-Gating is designed for PPM, which aims to utilize C_5 to gate the output features of different levels in pyramid pooling module.

As shown in Fig. 2, the feature f and pooling output p_i both undergo a 1×1 convolutional layer, and these two 1×1 convolutional layers share the same dimension. These two 1×1 convolutional layers are employed to learn gating parameters and make the convoluted f and p_i have the same dimension. As the convoluted results, $f^{'}$ times $p_i^{'}$ with a dot product to get the final gated pooling results $p_i^{''}$. Finally, feature f ($i.e.$ C_5 in PPM) concatenates with all 4 gated pooling results.

III. Experimental Results

In this section, we evaluate our approach on five benchmark datasets, including MSRA-B [11], ECSSD [12], HKU-IS [13], DUT-OMRON [14] and SOD [15], [16], all of which are publicly available. MSRA-B contains 5000 images, which have been widely used for visual salient object detection. The majority of the pictures have one salient object with a corresponding ground truth. ECSSD includes 1000 pictures which have complicated architecture collected from Internet. Its target ground truth contains 5 themes. HKU-IS includes 4447 pictures, and the majority of images have low contrast and various objects. It contains 2500 pictures for training, 500 pictures for validation and 1447 pictures for testing. DUT-OMRON consists of 5168 challenging images, and each of them includes various salient objects. SOD includes 300 pictures which contain various salient objects with low contrast and overlapping boundaries.

To well evaluate the proposed method, we adopt 3 widely used metrics, including P-R curve, F-Measure (MaxF) and Mean Absolute Error (MAE). Given a binary prediction B, we can compute the precision and recall as $precision = |B \cap Z|/|B|$ and $recall = |B \cap Z|/|Z|$, where $|Z|$ denotes the ground truth and $|\cdot|$ denotes the operation of summing up the non-zero values. We can obtain the P-R curve by computing the mean of the precision and recall over the prediction on a specific dataset.

To evaluate the model prediction, we selected F-measure metric, which is defined as

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 Precision + Recall}, \tag{1}$$

where β^2 is set to 0.3 to emphasize the $precision$ over recall as suggested in [17].

Mean Absolute Error is computed as the mean of pixel-wise absolute difference between the estimated saliency map S and its corresponding ground truth Z, as below

$$MAE = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} |S(i,j) - Z(i,j)|, \tag{2}$$

where W and H denote the width and height of a given image, respectively.

The weights of the Encoder of GDPNet were initialized with the weights ResNet-101 [18] pretrained on ImageNet [19]. We initialize the weights of ohter layers with a random way. The other layers have batch normalization [20]. ReLU [21] is applied after batch normalization. Similar to previous work, the "poly" learning rate policy is adopted. The initial learning rate and power are set to 0.02 and 0.9, respectively. We adopt SGD as the optimizer of the model.

In this section, the proposed method is compared with state-of-the-art salient object detection methods, including two conventional methods GC [22] and DRFI [23], and eleven CNN-based methods (LEGS [24], MC [2], DCL [25], RFCN [26], DHSNET [27], ELD [28], DISC [29], NLDF [8], UFNet [30] and DSS [9]). It is worth noting that our results are obtained by training on 2500 images from MSRA-B dataset.

Herein, we first evaluate the proposed method based on F-measure and MAE scores. As shown in Table I, our model trained on the MSRA-B dataset outperforms the other methods in terms of both F-measure and MAE score. On the MSRA-B, ECSSD, HKU-IS, DUT and SOD datasets, the proposed model promotes the state-of-the-art F-measure by 1.4%, 2.6%, 3.1%, 3.4% and 4.4%, which are large margins as the values are already very close to 100%. Regard the MAE scores, our approach also produced the lowest MAE score, which indicates that false cases predicted by our model are smaller than other models. It is known that HKU-IS and SOD often include plenty of pictures possessing various objects. The proposed method still showed better detection performance than the others in terms of MaxF and MAE.

In Fig. 3, we compare our saliency maps with the top three methods in Table I. For generality, we choose a variety of representative images in different circumstances, including simple scene, complex scene, multiple objects, large objects, small objects, low contrast and complex texture.

Taking all circumstances into account, it is easy to see that the proposed model can both emphasize the salient regions and predict consistent boundaries. Owing to the GDPNet structure, the proposed model provides higher scores for

TABLE I

QUANTITATIVE SALIENT OBJECT DETECTION RESULTS OVER FIVE WIDELY USED DATASETS. THE BEST RESULTS IN EACH COLUMN ARE HIGHLIGHTED IN **bold**. IT IS OBSERVED THAT OUR APPROACH PRODUCED THE BEST RESULTS ON ALL DATASETS IN TERMS OF F-MEASURE AND ON FOUR DATASETS IN TERMS OF MAE.

Model	Training		MSRA-B		ECSSD		HKU-IS		DUT		SOD	
	#Images	Dataset	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE
GC	-	-	0.719	0.159	0.597	0.233	0.588	0.211	0.495	0.218	0.526	0.284
DRFI	2,500	MB	0.845	0.112	0.782	0.170	0.776	0.167	0.664	0.150	0.699	0.223
LEGS	3,340	MB+P	0.870	0.081	0.827	0.118	0.770	0.118	0.669	0.133	0.732	0.195
MC	8,000	MK	0.894	0.054	0.837	0.100	0.798	0.102	0.703	0.088	0.727	0.179
MDF	2,500	MB	0.885	0.066	0.847	0.106	0.861	0.076	0.694	0.092	0.785	0.155
DCL	2,500	MB	0.916	0.047	0.901	0.068	0.904	0.049	0.757	0.080	0.832	0.126
RFCN	10,000	MK	-	-	0.899	0.091	0.896	0.073	0.747	0.095	0.805	0.161
DHSNET	6,000	MK	-	-	0.905	0.061	0.892	0.052	-	-	0.823	0.127
ELD	9,000	MK	-	-	0.865	0.098	0.844	0.071	0.719	0.091	0.760	0.154
DISC	9,000	MK	0.905	0.054	0.809	0.114	0.785	0.103	0.660	0.119	-	-
NLDF	2,500	MB	0.911	0.048	0.905	0.063	0.902	0.048	0.753	0.080	0.810	0.143
UFNet	2,500	MB	0.926	0.039	0.915	0.060	0.910	0.044	0.776	0.069	0.844	0.124
DSS(vgg)	2,500	MB	0.927	0.028	0.915	0.052	0.913	0.039	-	-	0.842	**0.118**
DSS(resnet)	2,500	MB	0.936	0.030	0.928	0.048	0.920	0.035	-	-	0.850	0.119
GDPNet	2,500	MB	**0.950**	**0.028**	**0.954**	**0.046**	**0.951**	**0.034**	**0.810**	**0.054**	**0.900**	0.120

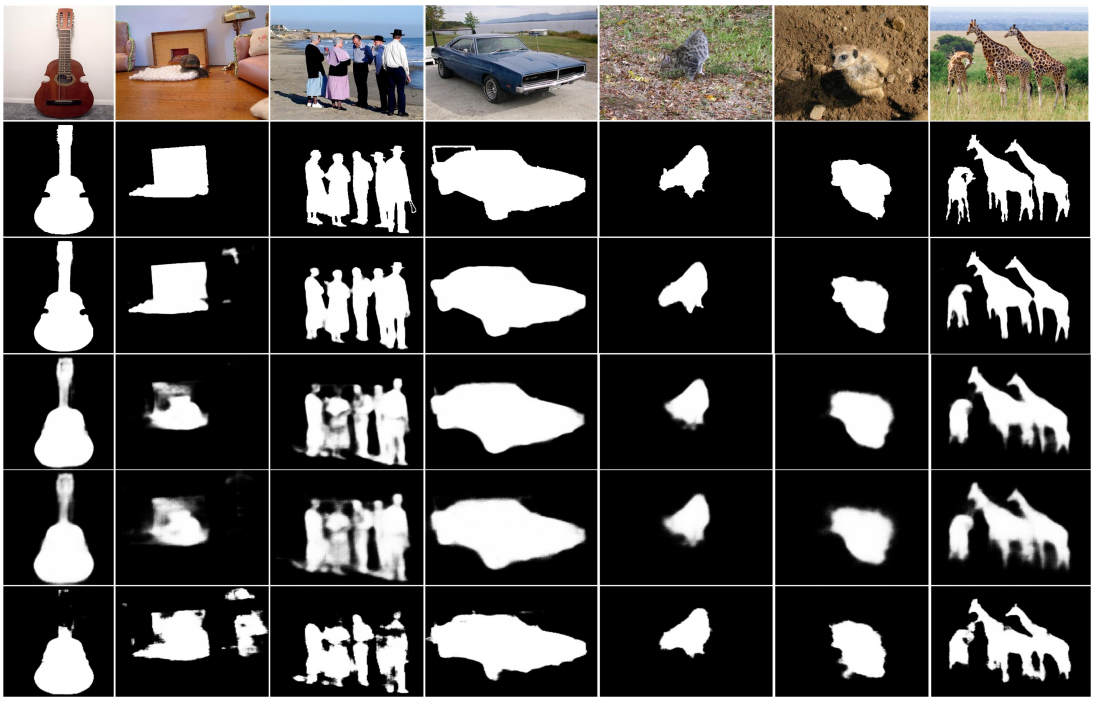

Fig. 3. To exhibit the superiority of our proposed approach compared against the existing methods, we choose various representative images from different datasets which incorporate a variety of circumstances, including simple scenes, complex scenes, multiple objects, large objects, small objects, low contrast and complex textures (as shown from 1 $\sim$ 7 columns). From top to bottom, the rows contain the input images, GT, GDPNet, DSS, UFNet, NLDF, respectively.

salient regions, which brings more obvious contrast between salient objects and the background. From Fig. 3, we can see that our results verge on the ground truth further so that we outperform other approaches in the majority of circumstances.

In Fig. 4, we compare the PR curves of our method and

TABLE II

PERFORMANCE COMPARISON WITH STATE-OF-THE-ART MODELS. THE BEST RESULTS IN EACH COLUMN ARE HIGHLIGHTED IN **BOLD**.

Model	Training Set	MSRA-B		ECSSD		HKU-IS		DUT	
		MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE
FPN	SOD	0.860	0.103	0.853	0.095	0.851	0.074	0.680	0.117
PPM	SOD	0.827	0.105	0.851	0.093	0.834	0.093	0.655	0.137
FPN+PPM	SOD	0.866	0.101	0.855	0.113	0.857	0.083	0.684	0.150
C-G+FPN+PPM	SOD	0.870	0.089	0.903	0.074	0.901	0.065	0.724	0.130
FPN+S-G+PPM	SOD	0.874	0.080	0.905	0.070	0.905	0.060	0.726	0.118
GDPNet	SOD	**0.881**	**0.077**	**0.911**	**0.064**	**0.913**	**0.054**	**0.732**	**0.111**

TABLE III

PERFORMANCE EVALUAITON ON BENCHMARK DATASETS. THE BEST RESULTS IN EACH COLUMN ARE HIGHLIGHTED IN **BOLD**.

Training Set	MSRA-B		ECSSD		HKU-IS		DUT		SOD	
	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE
MSRA-B(2500)	-	-	**0.954**	**0.046**	**0.951**	0.034	**0.810**	**0.054**	**0.900**	0.120
ECSSD(1000)	0.923	0.044	-	-	**0.951**	**0.029**	0.774	0.058	0.893	**0.102**
HKU-IS(2500)	0.921	**0.043**	0.942	0.054	-	-	0.777	0.066	0.876	0.145
DUT-OMRON(5168)	**0.930**	0.044	0.921	0.067	0.927	0.050	-	-	0.858	0.144
SOD(300)	0.881	0.078	0.911	0.064	0.913	0.054	0.732	0.111	-	-

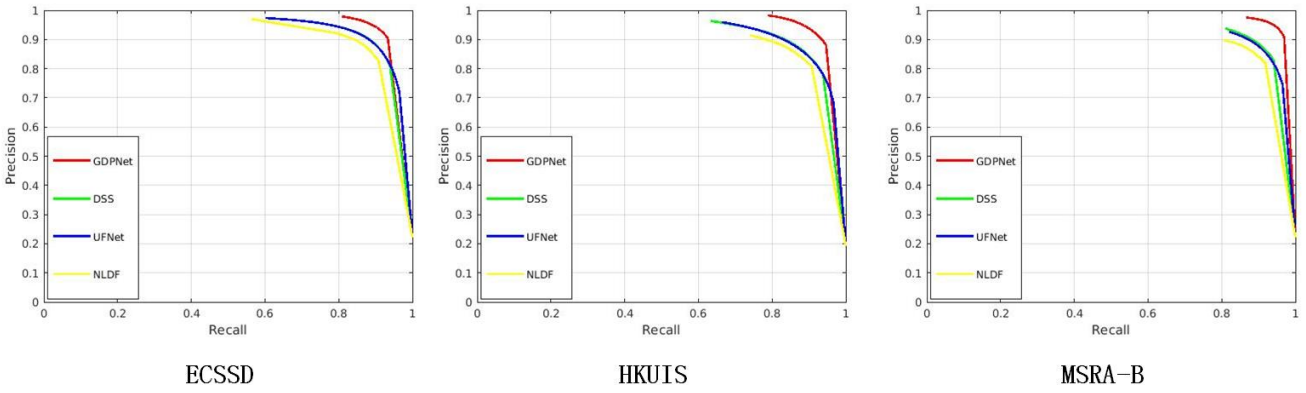

Fig. 4. Precision and recall curves on three benchmark salient object datasets.

other three best methods on three popular datasets. Among these approaches, the PR curve of the proposed model is outstanding in the upper when the recall is less than 1. We can also see that the precision of the proposed model is much higher when the recall score is close to 1, reflecting that our false positives are much fewer than the other methods.

During the study, it is interesting to find that a dataset with more training images does not necessarily lead to better performance than the one with less training images. This implies that the current datasets may differ in quality. To examine this problem, we further study the performance varying of our approach on different training datasets.

Similar to evaluating an inference model, the quality of a dataset can also be evaluated by measuring the generalization capability. Good generalization capability does not necessarily means containing more training samples, but rich object features. The dataset with more images may contain relatively few kinds of objects, while the dataset with fewer images may contain more kinds of objects.

To study the difference of different datasets, we first train the proposed model on one of the five datasets, and then perform salient object detection on the other four datasets. The results are shown in Table III. It can be seen that the method trained on MSRA-B produced the best MaxF on all the other datasets and best MAE score on the datasets ECSSD and DUT, though MSRA-B is not the largest dataset. The training on ECSSD presents superior MaxF performance than on the datasets HKU-IS and SOD, although ECSSD

is the second smallest dataset and has only 1000 samples. The above fact implies that MSRA-B and ECSSD have less redundancy and can better cover the diversity of the salient objects in different scenarios. This property is beneficial to network training.

IV. Conclusions

In the paper, we designed a new framework named Gating for Double Pyramid Networks (GDPNet) for salient object detection. The proposed GDPNet mainly consists of two submodules, FPN and PPM, which are used to capture the hierarchical details and global information, respectively. Furthermore, we introduce a Cross-Gating for FPN and a Single-Gating for PPM to filter out the irrelevant information from the non-salient object or background. Experimental results on five benchmark datasets showed that the proposed GDPNet outperforms the current state-of-the-art methods by a large margin. Besides, we empirically compare the quality of the five datasets we used, which could be a reference for future researchers.

References

[1] X. Cao, L. Yang, and X. Guo, "Total variation regularized rpca for irregularly moving object detection under dynamic background," *IEEE Trans. cybernetics*, vol. 46, no. 4, pp. 1014–1027, 2016.

[2] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1265–1274, IEEE, 2015.

[3] Y. Wei, X. Liang, Y. Chen, Z. Jie, Y. Xiao, Y. Zhao, and S. Yan, "Learning to segment with image-level annotations," *Pattern Recognition*, vol. 59, pp. 234–244, 2016.

[4] X. Bai, Y. Fang, W. Lin, L. Wang, and B. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Trans. Industrial Informatics*, vol. 10, pp. 2135–2145, Nov 2014.

[5] L. Zhang, Y. Xia, R. Ji, and X. Li, "Spatial-aware object-level saliency prediction by learning graphlet hierarchies," *IEEE Trans. Industrial Electronics*, vol. 2, no. 62, pp. 1301–1308, 2015.

[6] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, IEEE, 2015.

[8] Z. Luo, A. K. Mishra, A. Achkar, J. A. Eichel, S. Li, and P.-M. Jodoin, "Non-local deep features for salient object detection.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, p. 7, IEEE, 2017.

[9] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5300–5309, IEEE, 2017.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, IEEE, 2017.

[11] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern analysis and machine intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[12] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1155–1162, IEEE, 2013.

[13] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5455–5463, IEEE, 2015.

[14] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, IEEE, 2013.

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision*, vol. 2, pp. 416–423, IEEE, 2001.

[16] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 49–56, 2010.

[17] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, IEEE, 2009.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, June 2016.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE, 2009.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

[22] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[23] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *Proceedings of the IEEE international conference on computer vision*, pp. 1976–1983, 2013.

[24] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3183–3192, IEEE, 2015.

[25] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 478–487, IEEE, 2016.

[26] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *European Conference on Computer Vision*, pp. 825–841, Springer, 2016.

[27] N. Liu and J. Han, "Dhsnet: Deep hierarchical saliency network for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 678–686, IEEE, 2016.

[28] G. Lee, Y.-W. Tai, and J. Kim, "Deep saliency with encoded low level distance map and high level features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 660–668, IEEE, 2016.

[29] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, "Disc: Deep image saliency computing via progressive representation learning.," *IEEE Trans. Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1135–1149, 2016.

[30] Q. Hou, J. Liu, M.-M. Cheng, A. Borji, and P. H. Torr, "Three birds one stone: A unified framework for salient object segmentation, edge detection and skeleton extraction," *arXiv preprint arXiv:1803.09860*, 2018.