



PGCA-Net: Progressively Aggregating Hierarchical Features with the Pyramid Guided Channel Attention for Saliency Detection

Jiajie Mai¹, Xuemiao Xu², Guorong Xiao³, Zijun Deng², Jiaying Chen²

¹Department of Telecommunications Engineering and Management, Beijing University of Posts and Telecommunication, Beijing100876, China

²School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

³Key Laboratory of Science & Technology and Finance, Guangdong University of Finance, Guangzhou 510521, China

ABSTRACT

The Salient object detection aims to segment out the most visually distinctive objects in an image, which is a challenging task in computer vision. In this paper, we present the PGCA-Net equipped with the pyramid guided channel attention fusion block (PGCAFB) for the saliency detection task. Given an input image, the hierarchical features are extracted using a deep convolutional neural network (DCNN), then starting from the highest-level semantic features, we stage-by-stage restore the spatial saliency details by aggregating the lower-level detailed features. Since for the weak discriminative ability of the shallow detailed features, directly introducing them to the semantic features will only lead to sub-optimal results. Thus, we take a novel pyramid channel attention mechanism to attend to the useful detailed shallow feature channels before aggregation. The experimental results show that our proposed method outperforms its competitors on 5 benchmark testing sets.

KEY WORDS: Saliency detection; channel attention; image segmentation; computer vision; deep learning.

1 INTRODUCTION

THE salient object detection needs to highlight the most visually attractive regions in an image. Detecting the salient objects can contribute to many applications, such as the semantic segmentation with weak supervision (Lai et al. (2017) and the visual object tracking (Hong et al. (2015)). However, the saliency detection is very challenging, because finding salient objects requires high-level understanding of the whole image, while accurately segmenting out salient regions also matters.

The traditional methods (Liu et al. (2011), Yang et al. (2013), Zhu et al. (2014) for saliency detection are based on hand-crafted features, which are related to contrast, texture and color, or heuristic priors to detect the salient objects in an image. Since saliency detection requires high-level understanding of the input image, these hand-crafted features or priors usually fail for their weak representation ability degrading the practicability of the saliency detection.

In recent years, many works based on deep learning are explored for saliency detection after

witnessing the great success of the deep convolutional neural networks (DCNNs) on some mainstream computer vision works, such as object recognition (He et al. (2016), semantic segmentation (Long et al. (2015) and object detection (Ren et al. (2015)). Many methods (Zhao et al. (2015), Li et al. (2015), Long et al. (2015), Li et al. (2015) were proposed to generate accurate detection results. Despite the much more high-level image understanding of the deep convolutional neural network than the hand-crafted features or priors, the saliency detection results, or the so called saliency maps, are very coarse, due to the neglect of the saliency fine details in the process of the down-sampling of the FCNs. Later, work by (Hou et al. (2017), and Wang et al. (2017) combined the semantic coarse feature and the detailed feature together to fully exploit their complementary information to predict fine saliency maps. However, due to the discriminative ability gap of the features on different levels, directly fusing them will only lead to sub-optimal performance. The detailed features tend to have a high activation value on the non-salient regions, because of the small receptive field of neurons in the low-level of the CNN. Combining such

noisy features with the semantic coarse features will introduce much non-salient noisy features, further downgrading the detection accuracy.

In this paper, we propose the PGCA-Net with the novel pyramid guided channel attention mechanism to more accurately detect salient regions in the input images by effectively aggregating multi-level features that span across different stages of the CNN. First, we use a deep CNN to extract the hierarchical features from an input image. Then we take the highest-level features as the initial aggregated feature and stage-by-stage selectively aggregate the lower-level features to add more salient details to the current aggregated features with the pyramid guided channel attention mechanism. The discriminative ability for the salient objects is different for the features of the different levels. While the low-level features contain abundant saliency details, their discriminative ability is weak due to the small receptive field. Directly including the shallow detailed features to the refined semantic coarse features may bring a strong activation on the non-salient regions, degrading the performance on the saliency detection. To solve this issue, we adopt the channel attention mechanism to reweight the attentive degree of the detailed features, highlighting the most discriminative channels and suppress the noisy channels in the shallow detailed features. While the original channel attention mechanism is based on the global average pooling features, we extend it to a pyramid version to further consider the multi-scale global features to better exploit the high-level discriminative features as the guidance for attention. With the powerful pyramid guided channel attention mechanism, we can integrate the complementary information of the multi-level feature effectively and generate accurate saliency detection results. To strengthen the learning signal for the saliency detection, we take the deep supervision strategy, which imposes the supervision signal to the network at each feature aggregation step.

To verify the rationality of our presented PGCA-Net, we conduct an evaluation on five benchmark saliency detection datasets, comparing it with 16 state-of-the-art methods. It shows that our method outperforms the competitors on the five benchmarks consistently. Our contributions are as follows:

- First, we develop a pyramid guided channel attention fusion block (PGCAFB) to more completely utilize the multi-scale saliency abstract coarse features to select the channels of the detailed shallow features and to add the saliency edge information to the semantic coarse features. The PGCAFB brings in only small amounts of extra parameters and computation cost but leads to a considerable performance gain.
- Second, we present an PGAC-Net equipped with the pyramid guided channel attention fusion block to conduct the saliency detection.
- Third, our presented network is validated carefully on five benchmark datasets. In all, our method achieves the best performance on all the five datasets compared to the competitors, setting a new state-of-the-art.

2 RELATED WORK

2.1 Saliency Detection

HERE we discuss the related methods for the single-RGB-image saliency detection, which can be separated into two categories, hand-designed-priors-based methods, and deep-learning-based modern methods.

The traditional saliency detection methods [] uses manually designed shallow detailed features or priors, such as color, contrast and texture, to detect the salient objects in an image. Such low-level hand-designed methods lack high-level understanding of the images and cannot generalize well on the variant real-world cases. More detailed summary of the traditional methods can be found in the survey paper by (Borji et al. (2015).

In recent years, the community started to apply the deep convolutional neural networks (DCNNs) to the conduct saliency detection. Wang et al. (2016) designed a recurrent prediction network, which recurrently takes the saliency detecting result from the last prediction stage as prior to predict a better saliency map. Zhang et al. (2017) proposed a newly designed dropout technique for saliency detecting on unseen objects. However, these early DCNN-based methods only exploited the deep semantic coarse features of the CNNs and neglected the saliency detail information, which usually spanned on the shallow detailed features from the shallow layers of the CNNs.

Recently several methods were proposed to further enhance the saliency detection results by exploiting the detailed shallow detailed features from the shallow layers of the CNN and the designed mechanism to utilize the complementarity of the detailed features and semantic features. Hou et al. (2017) proposed a network with deep supervision and detailed features for saliency detection. Wang et al. (2017) used the stage-wise mechanism to absorb the high-resolution detailed features to progressively enhance the saliency detection results in several stages. However, such methods, which exploit the low-level detailed feature without selection or filtering leads to sub-optimal detection results. We illustrate our pyramid guided channel attention mechanism to better highlight the beneficial shallow detailed features and suppress the noisy harmful shallow detailed features.

2.2 The Channel Attention Mechanism

Attention is commonly utilized in many computer vision areas for its great performance. Recently, Hu et al. (2018) proposed a very light weight but powerful

channel attention mechanism and integrated it into each residual block of the ResNet (He et al. (2016) to reach the state-of-the-art image recognition accuracy. The channel attention mechanism takes as input a feature map, then applies the squeeze operation to squeeze it to a 1×1 feature map and uses two fully connected layers to transform it to an attention vector, followed by feature channels highlighting and suppressing using the attention vector. The novel mechanism introduces a very small computation cost and memory footprint but brings observable performance gain. Our method will also take the channel attention mechanism but extends it to the pyramid guided channel attention mechanism to better utilize the semantic abstract features to guide the selection of the shallow detailed features for the salient object detection.

3 METHOD

WE show the architecture of our presented PGAC-Net framework for saliency detection in Figure 1. To begin with, we extract a multi-level hierarchical feature using the FCN. Feature maps from different levels are with different resolutions and channels. The

resolution and channel numbers are high and small for the shallow detailed feature maps and low and large for the semantic coarse feature maps. Then we develop pyramid guided channel attention fusion blocks (PGCAFBs) to intelligently integrate the feature maps from the highest-level to the lowest level. In a PGCAFB, a shallow detailed feature and a semantic coarse feature are accepted as the input, and the semantic coarse feature gives guidance to automatically select the beneficial feature channels of the shallow detailed feature, followed by using the selected shallow detailed feature channels to refine the semantic coarse feature in a residual refinement manner. We densely impose the supervision signal after each PGCAFB refinement to boost the optimization convergence and learn about more powerful saliency features. We take the final integrated features to generate the saliency maps as the output of our model. In the following section, we elaborate on how to form our proposed pyramid guided channel attention block (PGCAFB) and then utilize the PGCAFBs to build our attention-guided feature aggregation network PGCA-Net for accurate saliency detection.

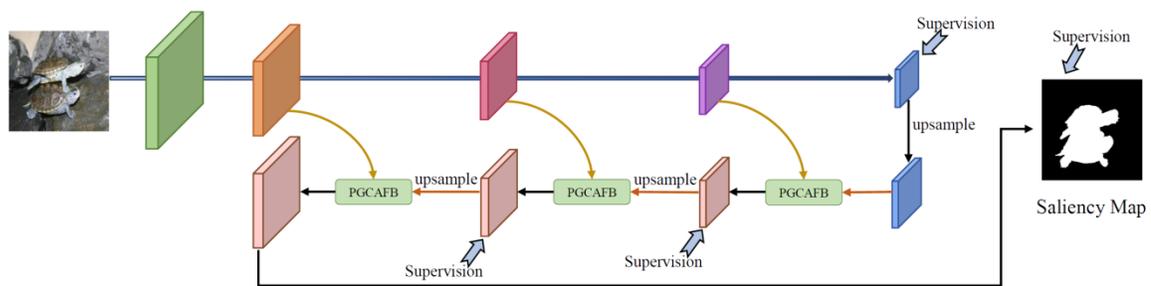


Figure 1. The Schematic Illustration of the Presented Method.

3.1 The Pyramid Guided Channel Attention Fusion Block (PGCAFB)

Due to the small receptive fields and weak discriminative ability of neurons in the shallow layers of the CNN, the shallow detailed features are easy to mistake to show strong activation on the non-salient regions, although they contain abundant detailed information. Directly using such noisy shallow detailed features to introduce the details will lead to the sub-optimal saliency detection results. Since not all channels of the shallow detailed feature maps are helpful, it is necessary to distinguish the beneficial channels, which are relatively robust from all the channels of the shallow detailed features before learning the complementary features from the semantic coarse features and the shallow detailed features. To achieve this, we apply the channel attention mechanism as our basis to selectively utilize the shallow detailed feature. The channel attention mechanism learns the dependency of the different

channels and discriminatively attends to the useful channels and neglects the noisy channels. Note that the semantic coarse features have strong discriminative abilities, because they are from deeper layers. We use the semantic coarse features as guidance to guide the shallow detailed feature selection, towards more robust results, thus, based on our basis channel attention mechanism, we simultaneously take as input the shallow detailed features and the semantic coarse features, and apply the channel attention to predict attention to impose on the shallow detailed features. Realizing that the original channel attention is based on the spatial aggregation features of only a single global scale (note that the input feature maps are first squeezed to 1×1 feature maps by using the global average pooling), we argue that taking into consideration more global scales will further boost the detection accuracy with only small amounts of extra computation cost, because more information with different global degrees are introduced. In this way, the original channel attention

is extended to the pyramid version with the multi-branch. In the different branch, input feature maps are globally aggregated in the spatial dimension to different spatial sizes (we empirically set the set of spatial size as 1×1 , 2×2 , 3×3 and 6×6).

Based on the discussion above, we develop a pyramid guided channel attention fusion block (PGCAFB) to better learn the complementary information of the shallow detailed features and the semantic coarse features. See Figure 2 for details. The PGCAFB begins with using the semantic coarse feature as guidance to discriminatively attend to the useful shallow detailed feature channel and suppress the harmful shallow detailed feature channels by the pyramid channel attention mechanism, then takes the semantic coarse features and the attentive shallow detailed features as input to learn a residual to add to the semantic coarse features.

3.2 The Attention-guided Feature Aggregation Network (PGCA-Net)

To fully exploit the multi-level features to generate the satisfactory saliency detection results, we developed an attention-guided feature aggregation network to progressively aggregate features of each level by utilizing our proposed PGCA. As shown in Figure 1, given an input image, we first use a ResNeXt50-32x4d network (Xie et al. (2017)) to extract the hierarchical multi-level features, the multi-level features are complementary. For the lower-level features, they are in a higher resolution and contain more saliency detail information but have weaker semantic discriminative ability due to the shallower layers. For the higher-level features, they are in a

lower resolution and the lack of spatial details are due to the down-sampling operation during the forward process, but they are more discriminative for roughly finding where the salient objects are. To achieve more precise saliency detection, we need to integrate the detail information of the shallow detailed features into the semantic coarse features to help locate the boundary of the salient objects more clearly, resulting in better saliency results.

We progressively aggregate the features from the highest level to the lowest level by using our presented PGCAFB. After considering the balance between the performance gain and computation cost, we neglect the features of the level at the beginning. Thus, we finish the top-down features aggregation using the 3 steps. The PGCAFBs used in all aggregation steps are not shared for stronger learning ability.

To further strengthen the gradient for optimizing and to learn more about the robust salient features, we apply the deep supervision mechanism to impose the supervision signal to each aggregation step during the training process. By directly giving the supervision signal to each step, the PGCAFB can try its best to aggregate all the useful features to reach accurate detection results. During testing, we only take the saliency maps predicted using the final aggregated features as the output results.

4 EXPERIMENTS

WE will first illustrate the datasets and the evaluation metrics, then introduce the training and testing strategies of our presented framework, and lastly report the numerical results.

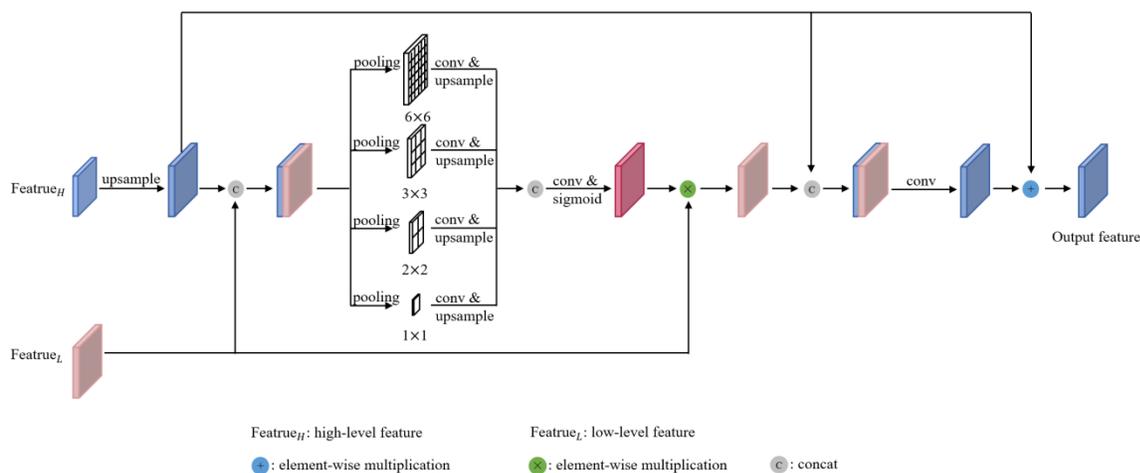


Figure 2. The Schematic Illustration of the Presented Pyramid Guided Channel Attention Fusion Block (PGCAFB) to Merge Features at Two Adjacent Levels (Relatively, We Call the Feature from the Deeper Layer as the Semantic Coarse Feature and the Feature from the Shallower Layer as the Shallow Detailed Feature).

4.1 Datasets and Evaluation Metrics

4.1.1 Datasets

We adopt the commonly used MSRA10K dataset (Cheng et al. (2015)) as our training set, which contains 10,000 image pairs. To fully evaluate the performance of the saliency methods, we take five widely used saliency benchmark datasets for testing, including ECSSD (Yan et al. (2013)) with 1,000 image pairs, HKU-IS (Li et al. (2015)) with 4,447 image pairs, PASCAL-S (Zhang et al. (2017)) with 850 image pairs, SOD (Hou et al. (2017)) with 300 image pairs, and DUT-OMRON (Yang et al. (2013)) with 5,168 image pairs.

4.1.2 Evaluation Metrics

We adopt two widely used saliency evaluation metrics, mean absolute error (MAE) and F-measure (F_β), in our experiment. MAE directly compares the predicted results with the corresponding ground truth using the absolute distance. F_β has been proven to be consistent with the visual perception of humans (Yan et al. (2013)). The MAE and the F_β are formulated as follows:

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

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

where S and G denotes the predicted saliency map and the ground truth saliency map, H and W denotes the height and width of the saliency map, and β^2 is set as 0.3 to emphasize the precision.

4.2 The Training and Testing Strategies

4.2.1 The Training Parameters

We implement our presented method based on the PyTorch framework. To boost the optimization convergence and reduce over-fitting the risk, we use the ResNeXt50-32x4d well trained on the ImageNet (Deng et al. (2009)) classification task to initialize the feature extraction network of our framework, and initialize other parameters using the msra initializer (He et al. (2015)). Our presented method is trained on the commonly used MSRA10K dataset. The stochastic gradient descent (SGD), is utilized to adjust the network for 10,000 times. On each training step, we bi-linearly scale each input image to the size of 416×416 and conduct a random horizontal flipping before feeding it to the network. During the training process, we tune the learning rate (Liu et al. (2015)) with the initial step size of 0.001 and the power is 0.9. It takes approximately one and half hours to train our proposed model using one GTX 1080Ti GPU.

4.2.2 Testing

During the testing stage, our method generates the corresponding saliency map. To further refine the spatial coherence of the predicted saliency maps and lead to more satisfactory results, we apply the conditional fully connected field (CRF) (Krähenbühl et al. (2011)) to the detection results.

4.3 The Ablation Study

To verify the effectiveness of the method, the ablation study is executed on the major components of the PGAC-Net. The numerical results can be seen in Table 1. The “w/o PGCAFB” in the first row is the performance of the method similar to our presented PGCA-Net but replaces all of the PGCAFBs with the simple block without the attentional selection of the shallow detailed features before fusion of the semantic coarse and shallow detailed feature. We also compared our method to the “w PGCAFB_1” whose PGCAFBs aggregating the input feature maps to only one spatial size of 1×1 (original channel attention design) to show the usefulness of the integrate information of more global scales to conduct the channel attention (pyramid channel attention).

From the results in Table 1 and the visual comparisons in Figure, we find that: (1) the introducing channel attention mechanism is helpful to the multi-level feature integration, because of the discriminative ability gap of features in different levels, by comparing the “w PGCAFB_1” to “w/o PGCAFB”, and (2) considering the multiple global scale to the conduct channel attention (pyramid channel attention) could further enhance the detection accuracy.

4.4 The Comparison with the State-of-the-arts

Furthermore, our method compares our presented PGCA-Net to the 16 state-of-the-art saliency detectors, which is shown in Table 2. The methods in the first three rows, “wCtr” (Zhu et al. (2014)), and “BSCA” (Qin et al. (2015)) are traditional saliency detection methods, which are based on the hand-crafted features or priors. The other methods are based on the deep convolutional neural networks. To conduct fair comparison, we obtained the saliency detection results of other methods by downloading the results offered by authors or retrained their models using the same training set with the hyper-parameters reported in their papers.

Our methods show the MSRA10K dataset and is tested on the other five datasets. From Table 2, our PGCA-Net outperforms all its competitors on all the five benchmark datasets, and from view of the F_β and MAE metrics showing the great generalization effect of the PGCA-Net.

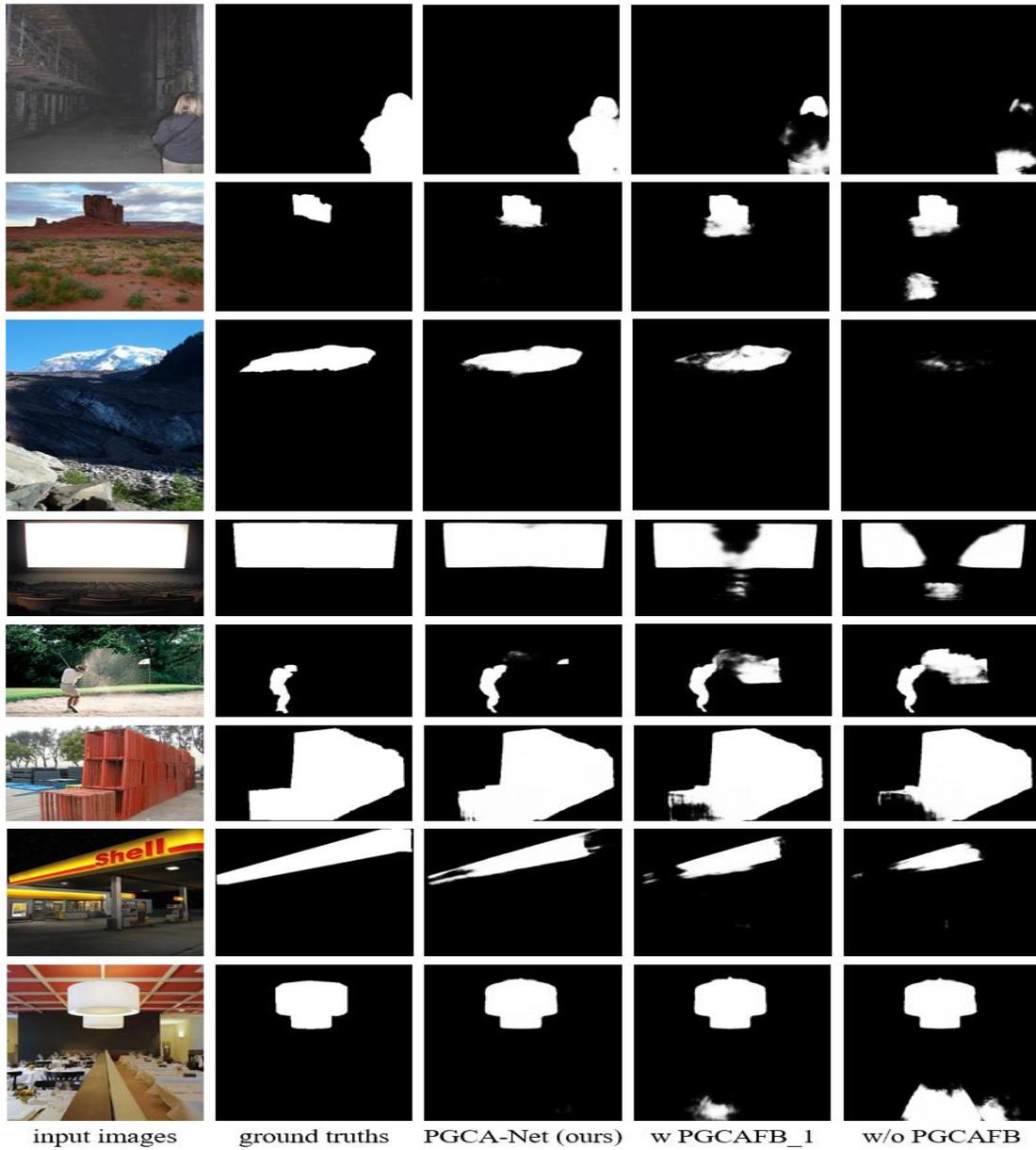


Figure 3. The Visual Comparisons on Difficult Cases for the Ablation Study.

Table 1. The F_β and MAE of Different Configurations of the Framework on the Five Saliency Detection Benchmarks for the Ablation Study.

| Method | ECSSD | | HKU-IS | | PASCAL-S | | SOD | | DUT-OMRON | |
|------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | F_β | MAE |
| w/o PGCAFB | 0.923 | 0.049 | 0.912 | 0.040 | 0.840 | 0.097 | 0.843 | 0.120 | 0.784 | 0.066 |
| w PGCAFB_1 | 0.927 | 0.047 | 0.914 | 0.038 | 0.843 | 0.094 | 0.849 | 0.117 | 0.789 | 0.065 |
| PGCA-Net (Ours) | 0.930 | 0.044 | 0.916 | 0.036 | 0.845 | 0.092 | 0.852 | 0.114 | 0.792 | 0.063 |

Table 2. The Comparisons with the State-of-the-art on the Five Widely used Datasets.

| Method | ECSSD | | HKU-IS | | PASCAL-S | | SOD | | DUT-OMRON | |
|-------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | F_β | MAE |
| MR (Yang et. al. (2013)) | 0.900 | 0.040 | 0.900 | 0.040 | 0.900 | 0.040 | 0.619 | 0.273 | 0.900 | 0.040 |
| wCtr (Zhu et. al. (2014)) | 0.900 | 0.040 | 0.900 | 0.040 | 0.900 | 0.040 | 0.632 | 0.245 | 0.900 | 0.040 |
| BSCA (Qin et. al. (2015)) | 0.900 | 0.040 | 0.900 | 0.040 | 0.900 | 0.040 | 0.634 | 0.266 | 0.900 | 0.040 |
| MC (Zhao et. al. (2015)) | 0.822 | 0.106 | 0.798 | 0.102 | 0.740 | 0.145 | 0.688 | 0.197 | 0.703 | 0.088 |
| LEGS (Wang et. al. (2015)) | 0.827 | 0.118 | 0.770 | 0.118 | 0.756 | 0.157 | 0.707 | 0.215 | 0.669 | 0.133 |
| MDF (Li et. al. (2015)) | 0.831 | 0.108 | 0.860 | 0.129 | 0.759 | 0.142 | 0.785 | 0.155 | 0.694 | 0.092 |
| ELD (Lee et. al. (2016)) | 0.867 | 0.080 | 0.844 | 0.071 | 0.771 | 0.121 | 0.760 | 0.154 | 0.719 | 0.091 |
| DS (Li et. al. (2016)) | 0.882 | 0.123 | - | - | 0.758 | 0.162 | 0.781 | 0.150 | 0.745 | 0.120 |
| RFCN (Wang et. al. (2016)) | 0.898 | 0.097 | 0.895 | 0.079 | 0.827 | 0.118 | 0.805 | 0.161 | 0.747 | 0.095 |
| DCL (Li and Yu (2016)) | 0.898 | 0.071 | 0.904 | 0.049 | 0.822 | 0.108 | 0.832 | 0.126 | 0.757 | 0.080 |
| DHSNet (Liu et. al. (2016)) | 0.907 | 0.059 | 0.892 | 0.052 | 0.827 | 0.096 | 0.823 | 0.127 | - | - |
| NLDF (Luo et. al. (2017)) | 0.905 | 0.063 | 0.902 | 0.048 | 0.831 | 0.099 | 0.810 | 0.143 | 0.753 | 0.080 |
| UCF (Zhang et. al. (2017)) | 0.910 | 0.078 | 0.886 | 0.073 | 0.821 | 0.120 | 0.800 | 0.164 | 0.735 | 0.131 |
| DSS (Hou et. al. (2017)) | 0.916 | 0.053 | 0.911 | 0.040 | 0.829 | 0.102 | 0.842 | 0.118 | 0.771 | 0.066 |
| Amulet (Zhang et. al. (2017)) | 0.913 | 0.059 | 0.887 | 0.053 | 0.828 | 0.095 | 0.801 | 0.146 | 0.737 | 0.083 |
| SRM (Wang et. al. (2017)) | 0.917 | 0.056 | 0.906 | 0.046 | 0.844 | 0.087 | 0.843 | 0.126 | 0.769 | 0.069 |
| PGCA-Net (Ours) | 0.930 | 0.044 | 0.916 | 0.036 | 0.845 | 0.092 | 0.852 | 0.114 | 0.792 | 0.063 |

5 CONCLUSION

TO finalize, we put forward the PGCA-Net with the novel pyramid guided channel attention fusion block (PGCAFB) to robustly and accurately segment out the salient objects from an input image. The high-level features are used as guidance to dynamically select the lower-level features, which contain abundant salient detail features. We adopt the channel attention mechanism to conduct the feature selection process for the purpose of obtaining performance gain while increasing only small amounts of extra computation cost. To exploit the multi-scale spatial context in the features and further boost the performance, we extended the original channel attention to the pyramid channel attention with ignorable extra parameters and computation amounts. The presented network achieves performance on the salient object detection.

6 ACKNOWLEDGMENT

THIS work was supported by Key-Area R&D Program of Guangdong Province, China (2020B010166003, 2020B010165004, 2018B010107003), Guangdong High-level personnel program (2016TQ03X319), and Guangzhou key project in industrial technology (201802010027), the Guangdong Provincial Key Laboratory of Technology and Finance & Big Data Analysis (Grant No. 2017B030301010), the National Natural Science Foundation of China (61772206, U1611461, 61472145, 61603420), the Guangdong Key Research Base of Technology and Finance (2014B030303005), the Construction on New technology Credit Service Platform Based on O2O Model (2017B080802004), the Platform of Credit Financing and Trade for Guangdong Technological Enterprises (2014B080807035).

7 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

8 REFERENCES

- Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). *Salient object detection: A benchmark*. *IEEE transactions on image processing*, 24(12), 5706-5722.
- Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H., & Hu, S. M. (2015). *Global contrast based salient region detection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), 569-582.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). *ImageNet: A large-scale hierarchical image database*. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.
- Gorgel, P., Kilic, N., Ucan, B., Kala, A. and Ucan, O.N., 2009. A Backpropagation Neural Network Approach For Ottoman Character Recognition. *Intelligent Automation & Soft Computing*, 15(3), pp.451-462.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification*. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hong, S., You, T., Kwak, S., & Han, B. (2015, June). *On-line tracking by learning discriminative saliency map with convolutional neural network*. In *International conference on machine learning* (pp. 597-606).

- Hou, Q., Cheng, M. M., Hu, X., Botji, A., Tu, Z., & Torr, P. H. (2017). *Deeply supervised salient object detection with short connections*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3203-3212).
- Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-excitation networks*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).
- Krähenbühl, P., & Koltun, V. (2011). *Efficient inference in fully connected crfs with gaussian edge potentials*. In *Advances in neural information processing systems* (pp. 109-117).
- Lai, B., & Gong, X. (2017, August). *Saliency guided end-to-end learning for weakly supervised object detection*. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 2053-2059). AAAI Press.
- Lee, G., Tai, Y. W., & Kim, J. (2016). *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).
- Li, G., & Yu, Y. (2015). *Visual saliency based on multiscale deep features*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5455-5463).
- Li, G., & Yu, Y. (2016). *Deep contrast learning for salient object detection*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 478-487).
- Li, X., Zhao, L., Wei, L., Yang, M. H., Wu, F., Zhuang, Y., & Wang, J. (2016). *Deep saliency: Multi-task deep neural network model for salient object detection*. *IEEE Transactions on Image Processing*, 25(8), 3919-3930.
- Liu, N., & Han, J. (2016). *Dhsnet: Deep hierarchical saliency network for salient object detection*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 678-686).
- Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., & Shum, H. Y. (2011). *Learning to detect a salient object*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2), 353-367.
- Liu, W., Rabinovich, A., & Berg, A. C. (2015). *ParseNet: Looking wider to see better*. *arXiv preprint arXiv:1506.04579*.
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- Luo, Z., Mishra, A., Achkar, A., Eichel, J., Li, S., & Jodoin, P. M. (2017). *Non-local deep features for salient object detection*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6609-6617).
- Nagaraj, Balakrishnan, and Ponnusamy Vijayakumar. "Tuning of a PID controller using soft computing methodologies applied to basis weight control in paper machine." *Journal of Korea Technical Association of The Pulp and Paper Industry* 43, no. 3 (2011): 1-10.
- Nisi, K., B. Nagaraj, and A. Agalya. "Tuning of a PID controller using evolutionary multi objective optimization methodologies and application to the pulp and paper industry." *International Journal of Machine Learning and Cybernetics* (2018): 1-11.
- Qin, Y., Lu, H., Xu, Y., & Wang, H. (2015). *Saliency detection via cellular automata*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 110-119).
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster r-cnn: Towards real-time object detection with region proposal networks*. In *Advances in neural information processing systems* (pp. 91-99).
- Roopaei, M., Rad, P., and Jamshidi, M., 2017. Deep learning control for complex and large scale cloud systems. *Intelligent Automation & Soft Computing*, 23(3), pp. 389-391.
- Wang, L., Lu, H., Ruan, X., & Yang, M. H. (2015). *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).
- Wang, L., Wang, L., Lu, H., Zhang, P., & Ruan, X. (2016, October). *Saliency detection with recurrent fully convolutional networks*. In *European conference on computer vision* (pp. 825-841). Springer, Cham.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). *Aggregated residual transformations for deep neural networks*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).
- Yan, Q., Xu, L., Shi, J., & Jia, J. (2013). *Hierarchical saliency detection*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1155-1162).
- Yang, C., Zhang, L., Lu, H., Ruan, X., & Yang, M. H. (2013). *Saliency detection via graph-based manifold ranking*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3166-3173).
- Zhang, P., Wang, D., Lu, H., Wang, H., & Ruan, X. (2017). *Amulet: Aggregating multi-level convolutional features for salient object detection*. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 202-211).
- Zhao, R., Ouyang, W., Li, H., & Wang, X. (2015). *Saliency detection by multi-context deep learning*. In *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition (pp. 1265-1274).

Zhu, W., Liang, S., Wei, Y., & Sun, J. (2014). *Saliency optimization from robust background detection* In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2814-2821).

9 NOTES ON CONTRIBUTORS



vision.

Jiajie Mai received bachelor's degree in Telecommunications Engineering with Management from Beijing University of Posts and Telecommunications, China in 2018. His research interests include artificial intelligence and computer



Xuemiao Xu received B.S. and M.S. degrees in Computer Science and Engineering from South China University of Technology in 2002 and 2005 respectively, and Ph.D. degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2009. She is currently a professor in the School of Computer Science and Engineering, South China University of Technology. Her research interests include object detection, tracking, recognition, and image, video understanding and synthesis particularly based on the deep learning technique.



Guorong Xiao received B.S. degree in 2001 and M.S. degree in 2004 from South China University of Technology in Computer Science and Technology. He is an associate professor of Computer Application Technology. Currently he serves as the Director of Guangdong Provincial Science and Technology Finance Big Data Engineering Technological Research Center and the Deputy Director of Guangdong Provincial Key Laboratory of Technology and Finance & Big Data Analysis, and the Chief Technical Officer of the Guangdong Province Science & Technology Financial Comprehensive Information Service Platform, and the Director of Guangdong Provincial Science and Technology Finance Big Data Engineering Technological Research Center. His main research is big data analysis, intelligence information processing and computer application. He is co-first author.



University of Technology. His research interests are computer vision.

Jiaying Chen received B.Eng. degree in Computer Science and Technology from South China University of Technology, China, in 2018. He is currently working towards a master's degree in the School of Computer Science and Engineering, South China



Zijun Deng received B.Eng. degree in Computer Science and Technology from South China University of Technology, China, in 2017. He is currently working towards a master's degree in the School of Computer Science and Engineering, South China University of Technology. His research interests include computer vision and deep learning.