

Article

Dual-Branch Feature Fusion Network for Salient Object Detection

Zhehan Song¹, Zhihai Xu¹, Jing Wang², Huajun Feng¹ and Qi Li^{1,*} 

¹ College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China; 21830014@zju.edu.cn (Z.S.); xuzh@zju.edu.cn (Z.X.); fenghj@zju.edu.cn (H.F.)

² Science and Technology on Optical Radiation Laboratory, Beijing 100854, China; wangjing4473@163.com

* Correspondence: liqi@zju.edu.cn

Abstract: Proper features matter for salient object detection. Existing methods mainly focus on designing a sophisticated structure to incorporate multi-level features and filter out cluttered features. We present the dual-branch feature fusion network (DBFFNet), a simple effective framework mainly composed of three modules: global information perception module, local information concatenation module and refinement fusion module. The local information of a salient object is extracted from the local information concatenation module. The global information perception module exploits the U-Net structure to transmit the global information layer by layer. By employing the refinement fusion module, our approach is able to refine features from two branches and detect salient objects with final details without any post-processing. Experiments on standard benchmarks demonstrate that our method outperforms almost all of the state-of-the-art methods in terms of accuracy, and achieves the best performance in terms of speed under fair settings. Moreover, we design a wide-field optical system and combine with DBFFNet to achieve salient object detection with large field of view.

Keywords: salient object detection; dual-branch feature fusion network; wide-field optical system



Citation: Song, Z.; Xu, Z.; Wang, J.; Feng, H.; Li, Q. Dual-Branch Feature Fusion Network for Salient Object Detection. *Photonics* **2022**, *9*, 44. <https://doi.org/10.3390/photonics9010044>

Received: 7 December 2021

Accepted: 6 January 2022

Published: 14 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Salient object detection, which aims to locate the most obvious object in an image, is widely used for object detection, visual tracking, image retrieval and semantic segmentation [1,2]. Traditional salient object detection is mainly divided into salient object detection based on the spatial domain and salient object detection based on the frequency domain. The former [3–8] is used to extract the salient objects of the image through the design and fusion of low-level features such as multi-scale contrast and color. The latter [9,10] is mainly utilized through frequency domain conversion and modification to extract the corresponding salient target. Because traditional salient object detection methods still rely on the shallow handcraft features and cannot obtain the deep semantic information of the image; therefore, they cannot accurately label salient objects in more complex scenes. Benefitting from the hierarchical structure of CNN, deep models can extract multi-level features that contain both low-level local details and high-level global semantics. Early deep saliency models include the use of convolutional neural networks to establish a multi-background convolutional neural network combining global and local background [11] and the use of multi-scale texture features in network to predict saliency map [12,13]. Although these deep models achieve significant improvements over traditional methods, the generated saliency map drops the spatial information and results in low prediction resolution. To solve this problem, many researchers use a fully convolutional network [14] to generate a saliency map of the same size as the input image resolution. Cheng et al. [15] used the different features of the shallow and deep layers and combined short-cut links to obtain a better saliency map. Liu et al. [16] combined a recurrent convolutional neural network to further refine the global structured saliency cues. Zhang et al. [17] introduced a reformulated dropout after

specific convolutional layers to construct an uncertain ensemble of internal feature units, which encouraged the robustness and accuracy of saliency detection. Li et al. [18] proposed a two-stream-deep model to solve boundary blur in saliency detection. They added a segmentation stream to extract segment-wise features and fuse with the original salient features to obtain a final saliency map. Zhang et al. [19] presented a generic aggregating multi-level convolutional feature framework for salient object detection. They integrated multi-level feature maps into multiple resolutions, but they ignored the interaction between features with different information at multi-levels. Li et al. [20] combined the segmentation method to obtain a saliency mask. However, this framework obtains a saliency map based on segmentation, which makes them unable to efficiently obtain the final result. Liu et al. [21] proposed a novel pixel-wise contextual attention network. Their goal was to learn to selectively attend to informative context locations for each pixel. Feng et al. [22] also focused on attention mechanism. They designed the attentive feedback modules to better explore the structure of objects. These attention-based methods improve the prediction results of the network by stacking attention modules layer by layer, but at the same time, they inevitably make the network more complicated.

Wide-field optical systems are also widely used in surveillance, remote sensing and other fields for object detection, visual tracking, image retrieval and semantic segmentation [23–25]. A large imaging angle of view means a large range of imaging searches, but it also brings greater aberrations.

On the basis of the above systems, we introduce an efficient and lightweight salient object detection network named the dual-branch feature fusion network (DBFFNet). Moreover, we designed a wide-field optical system with high imaging quality, which can effectively reduce aberrations. By combining DBFFNet with a wide-field optical system, we can easily obtain salient objects from a large field of view, which provides excellent clues for subsequent tasks.

2. Dual-Branch Feature Fusion Network

The overall architecture is shown in Figure 1. In this section, we begin by describing the whole structure of DBFFNet in Section 2.1, then introduce the adopted three main modules in Sections 2.2–2.4. Finally, we introduce the loss function in Section 2.5.

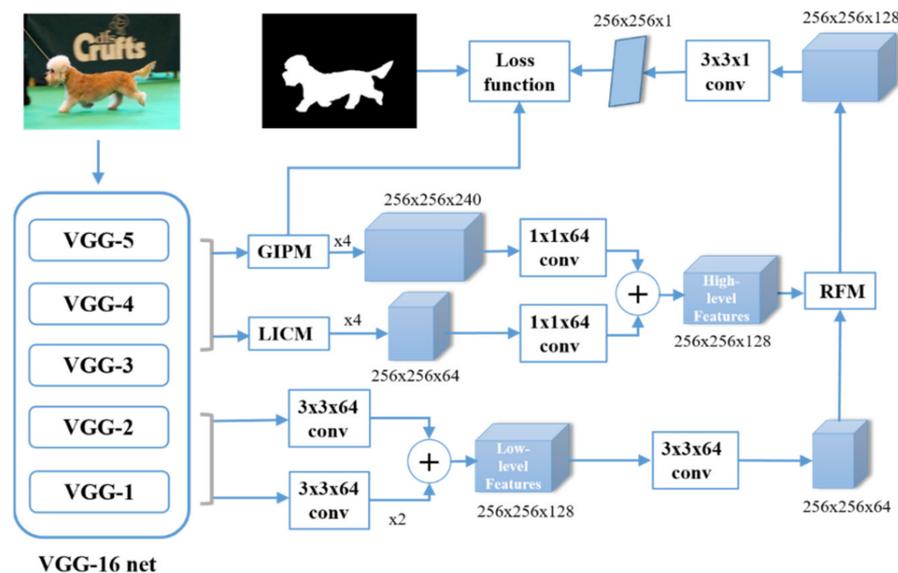


Figure 1. The pipeline of the proposed approach. LICM: local information concatenation module. GIPM: global information perception module. RFM: refinement fusion module. x2, x4: upsampling. VGG-1, VGG-2, VGG-3, VGG-4 and VGG-5 correspond to the second, fourth, seventh, tenth and thirteenth layer of VGG-16, respectively.

2.1. Overall Structure

The overall architecture based on dual-branch feature fusion and VGG-16 backbone. Because of the strong ability to combine multi-level features from classification networks, this type of architecture has been widely adopted in many vision tasks, including salient object detection. As shown in Figure 1, low-level features are concatenated from VGG-1 and VGG-2 (Branch 1). By extracting the global and local information from LICM and GIPM (Branch 2), we aim at completely distinguishing the salient objects from the background. After that, we further introduce a refinement fusion module (RFM) to ensure that high-level and low-level features from different layers can be adaptively merged together. In what follows, we describe the structures of the above mentioned three modules and explain their functions in detail.

2.2. Local Information Concatenation Module

Local information is crucial for salient object detection. Benefiting from the pyramid feature extraction method, we stack convolutional layers with different receptive fields at different scales to fully obtain local information of salient objects.

Specifically, the local information concatenation module is shown in Figure 2. High-level features separately extracted from VGG-3, VGG-4 and VGG-5 layers. These features at different scales then fed into LPM. In LPM, we adopt dilated convolution with different dilation rates, which are set to 3, 5 and 7 to capture multi-receptive-field local information. Moreover, we combine the feature maps from different dilated convolutional layers and a 1×1 convolutional feature by channel concatenation. After batch normalization and activation, we obtain features with multi-receptive-field local information in single scale. Finally, we up-sample the LPM outputs corresponding to VGG-4 and VGG-5 and combine them with the VGG-3 branch by channel concatenation.

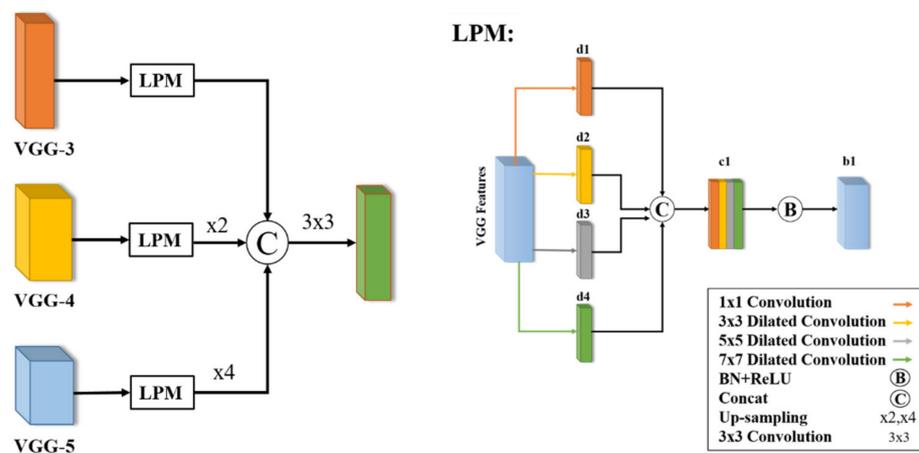


Figure 2. Detailed illustration of our local information concatenation module (LICM). It comprises three sub-branches, each of which contains a local perception module (LPM). LPM: local perception module comprises three sub-branches, each of which works with different receptive fields. After dilated convolution, all sub-branches (d1, d2, d3, d4) with the same number of channels (16) are combined to output activated features.

2.3. Global Information Perception Module

Global information can provide clearer clues for salient object detection. Combining global information can better deal with the extraction of salient objects in complex situations. For the challenging scenarios in salient object detection, such as cluttered background, foreground disturbance, and multiple salient objects, only using the local information may fail to completely detect the salient regions due to lacking the global semantic relationship among different parts of salient object or multiple salient objects.

To overcome these issues, we designed the GIPM module to capture the global information from deep features and transmit global information layer by layer. As shown in Figure 3, we firstly employ global average pooling to obtain the global information and then reassign different weights to different channels of the VGG-5 feature maps (G5) by 1×1 convolution and sigmoid function. Finally, we fed G5 into a three-layer U-Net structure. The feature maps (F3) fed into a two-layer convolutional skip link became the final GIPM outputs (F1).

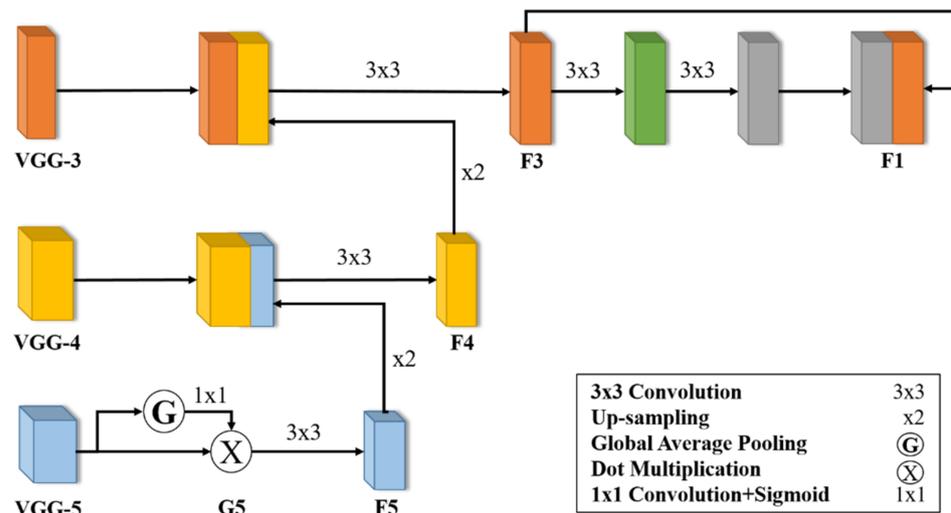


Figure 3. Detailed illustration of our global information perception module (GIPM). It contains a structure similar to U-Net to pass global information step by step and uses a short-cut at the tail to refine the outputs.

2.4. Refinement Fusion Module

Through the two branches, we obtained the high-level and low-level features. High-level features contain more concentrated object information and less background noise. Although low-level features contain more background noise, they can also supplement object information.

Thus, we introduced a refinement fusion module (RFM) in Figure 4 to better fuse the high-level and low-level features from two branches. We did not design a complex fusion structure, but adopted a classic method of concatenation to fuse the high-level and low-level features. Moreover, we introduced spatial attention to guide the low-level features for suppressing the background noise.

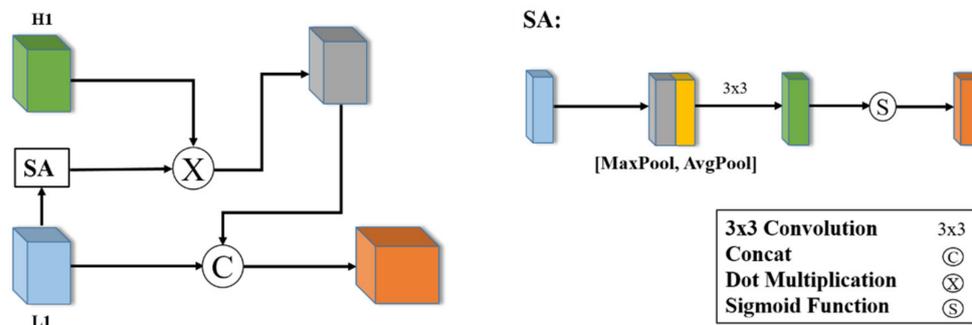


Figure 4. Detailed illustration of the refinement fusion module (RFM). In order to better integrate the high-level (H1) and low-level (L1) features, we introduced a spatial attention (SA) mechanism. By focusing on the spatial attention of the high-level features, we reduced the cluttered background influence in the low-level features.

2.5. Loss Function

In salient object detection, binary cross-entropy loss is often used as the loss function to measure the difference between the generated saliency map and the ground truth, which can be formulated as below:

$$L_{bce} = -\frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W (G_{ij} \log(P_{ij}) + (1 - G_{ij}) \log(1 - P_{ij})) \quad (1)$$

where W, H denote the width and height of the input image, respectively, G_{ij} is the ground truth label of the pixel (i, j) and P_{ij} represents the corresponding predicted results in position (i, j) .

In order to accelerate the network convergence, we also added auxiliary loss for F1, F3, F4 and F5 in Figure 3. The total loss calculated by Equation (2) contains the main loss and the auxiliary loss

$$L_{total} = L_{main} + \theta_i \sum_{i=1,3,4,5} L_{aux} \quad (2)$$

where θ_i means the weights of the auxiliary loss. F1, F3, F4 and F5 are upsampled to the same size as the ground truth via bilinear interpolation. The main loss and the auxiliary loss are calculated using Equation (1).

3. Experiment Settings

3.1. Datasets

We carried out experiments on two public salient object detection datasets, which are ECSSD [26] and DUTS [27]. Our model is trained on the training set (10,553 images) from DUTS and tested on its test set (5019 images) along with ECSSD. We evaluated the performance using the mean absolute error (MAE) and F-measure score. F-measure, denoted as F_β , is an overall performance measurement and is computed by the weighted harmonic mean of the precision and recall:

$$F_\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (3)$$

where β^2 is set to 0.3 as performed in previous work to weight precision more than recall. The MAE score indicates how similar a saliency map S is compared with the ground truth G :

$$MAE = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W |S(i, j) - G(i, j)| \quad (4)$$

where W and H denote the width and height of S , respectively.

3.2. Training and Testing

We carried out data augmentation by horizontal and vertical flipping, image rotating and random cropping. When fed into the DBFFNet, each image is warped to size 256×256 and subtracted using a mean pixel provided by VGG net at each position. The coefficients of auxiliary loss equal 0.2. The initial learning rate is set to 2×10^{-4} and the overall training procedure takes about 90 epochs. For testing, the images were scaled to 256×256 to feed into the network and then the predicted saliency maps were bilinearly interpolated to the size of the original image.

4. Results and Discussion

4.1. Comparison with the State-of-the-Art

The saliency maps for visual comparisons are provided below. We compared DBFFNet with UCF [17], DHS [16], DCL [18], DSS [15], Amulet [19], MSR [20], PiCANet [21] and

AFNet [22]. Figure 5 shows some example results of our model, along with another eight state-of-the-art methods for visual comparisons.

Our method gives superior results in low contrast (rows 1–2) and complex background scenes (rows 3–4). Additionally, it recovers more complete details (row 5). From the comparison, we can see that our method performs robustly when facing these challenges and produces better saliency maps.

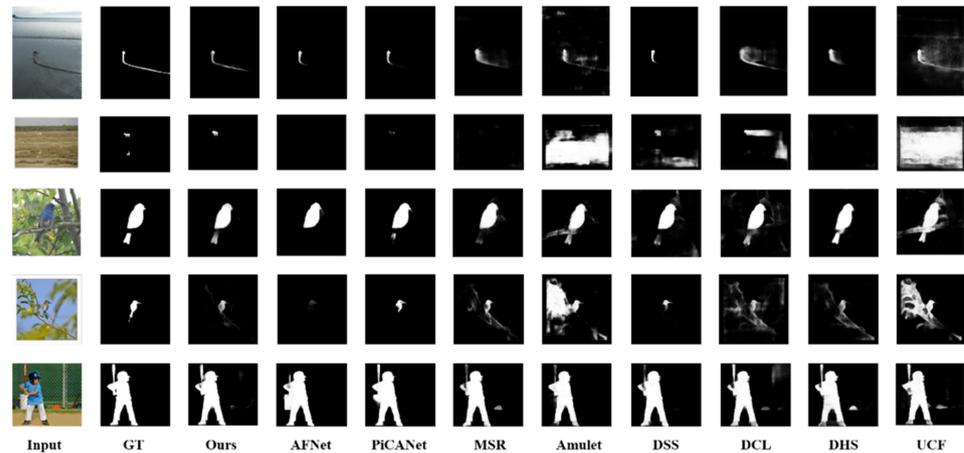


Figure 5. Visual comparison with different methods in various scenarios.

We also compare DBFFNet with eight state-of-the-art methods using quantitative evaluation. The quantitative performances of all methods are shown in Tables 1 and 2.

Table 1 shows the comparisons of MAE for two datasets. For DUTS-TE, DBFFNet ranks second. Additionally, for ECSSD, our model ranks first together with AFNet. Table 2 shows the comparisons of the F-measure score. Among the best three models, ours ranks second.

Moreover, we test the running speed of DBFFNet with another eight state-of-the-art methods. Average speed (FPS) comparisons among different methods (tested in the same environment) are reported in Table 3.

Table 1. Quantitative comparisons with different methods on 2 datasets with MAE. The best three results are shown in red, green and blue.

Method	ECSSD	DUTS-TE
UCF [17]	0.080	0.111
DHS [16]	0.062	0.067
DCL [18]	0.082	0.081
DSS [15]	0.064	0.065
Amulet [19]	0.062	0.075
MSR [20]	0.059	0.062
PiCANet [21]	0.049	0.055
AFNet [22]	0.044	0.046
Ours	0.044	0.048

Table 2. Quantitative comparisons with different methods on 2 datasets with F-measure score. The best three results are shown in red, green and blue.

Method	ECSSD	DUTS-TE
UCF [17]	0.904	0.771
DHS [16]	0.905	0.815
DCL [18]	0.896	0.786
DSS [15]	0.906	0.813
Amulet [19]	0.911	0.773
MSR [20]	0.903	0.824
PiCANet [21]	0.930	0.855
AFNet [22]	0.935	0.862
Ours	0.933	0.860

Table 3. Average speed (FPS) comparisons between our approach and the previous state-of-the-art methods. The best three results are shown in red, green and blue.

Model	FPS
UCF [17]	21
DHS [16]	20
DCL [18]	7
DSS [15]	12
Amulet [19]	16
MSR [20]	4
PiCANet [21]	10
AFNet [22]	18
Ours	43

We can clearly find from Table 3 that the average speed of our model is twice that of the model in second place. The average speed of AFNet, whose quantitative evaluation results are close to our model, is only 18 frames.

4.2. Ablation Studies

To investigate the importance of different modules in our method, we conducted ablation studies on a DUTS-test dataset. From Table 1, it can be seen that the proposed model contains all components necessary to achieve the best performance, which demonstrates that all components are necessary for the proposed method in order to obtain the best salient object detection result. Moreover, in Table 4 we adopted the model using only high-level features (VGG-3, VGG-4, VGG-5 concatenation) as a basic model without other modules; the base MAE is 0.112. First, we added LICM to the basic model and obtained a decline of 27% in MAE compared with the basic model. Then, we added GIPM to high-level features and proved the effectiveness of the global information. On this basis, we added low-level features and obtained a decline of 55% in MAE compared with the basic model. Finally, we added RFM in the model and obtained the best result, which obtained a decline of 58% in MAE compared with the basic model.

Table 4. Ablation evaluations using different components combinations. LL and HL represent low-level features and high-level features, respectively.

Method	MAE
HL	0.112
HL + LICM	0.082
HL + GIPM	0.078
HL + LICM + GIPM	0.060
HL + LL + LICM + GIPM	0.052
HL + LL + LICM + GIPM + RFM	0.048

In addition, we visualized the combined effects of different modules. As shown in Figure 6, the network can output an optimal saliency map by adding global information and low-level supplement information.

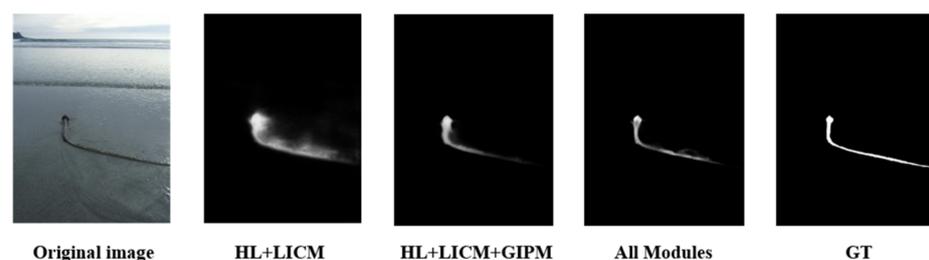


Figure 6. Saliency maps predicted by our proposed DBFFNet with different modules.

4.3. Discussion with Wide-Field Optical System

We further combined DBFFNet with a wide-field optical system. The wide-field optical system has a large field of view and excellent imaging quality. Through the detection of salient objects, it can provide good prior information for subsequent object tracking, image segmentation, etc., so as to achieve multi-task collaboration ability.

The wide-field optical system has a 28-degree field of view, and the focal length of the system is 15 mm. The image size and pixel size of the wide-field optical system are 2000×2000 and $5.5 \mu\text{m} \times 5.5 \mu\text{m}$, respectively. The optical structure is shown in Figure 7a.

Since the wide-field optical system outputs color images and the pixel arrangement is RGB, the pixel size after the combined arrangement is $11 \mu\text{m}$. Figure 7b shows the MTF curve. For the wide field of view image surface (spatial frequency 45.4 lp/mm), the MTF is all greater than 0.5.

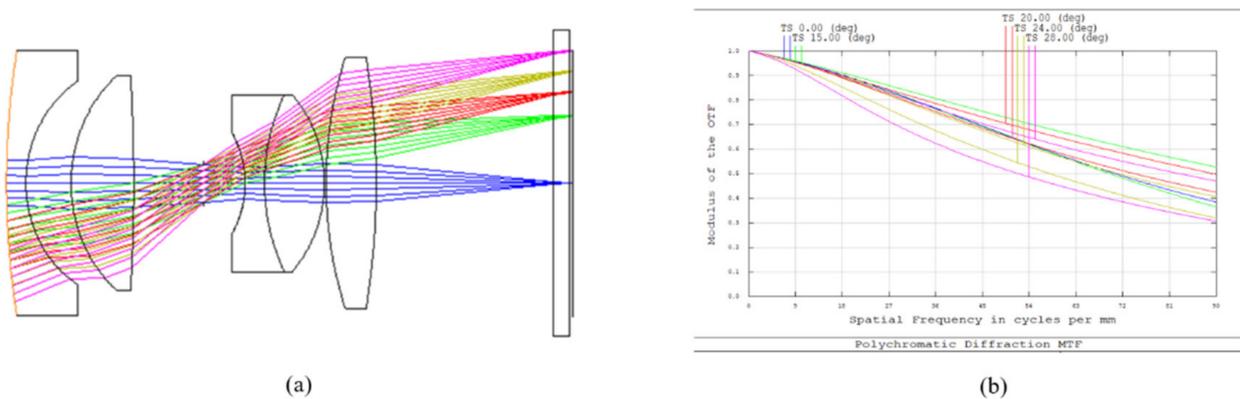


Figure 7. Wide-field optical system design and MTF curve. (a) Optical design; (b) MTF curve.

Figure 8 shows the wide-field optical system spot diagram together with field curvature and distortion. The point diagram in Figure 8a indicates that wide-field optical system has excellent imaging effects in the visible light band. The RMS radius corresponding to the maximum angle of view is only $4.877 \mu\text{m}$. As shown in Figure 8b, the maximum field curvature and distortion produced by the wide-field optical system are only 0.14 mm and 2%, respectively.

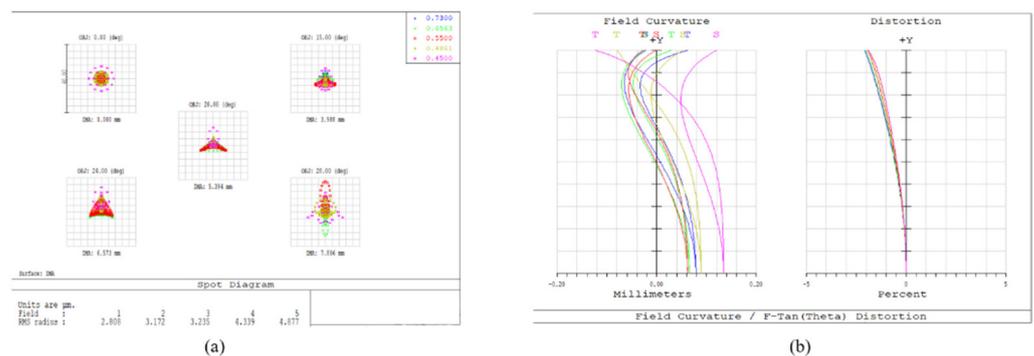


Figure 8. Wide-field optical system spot diagram together with field curvature and distortion. (a) Spot diagram; (b) field curvature and distortion.

The top view and left view of the wide-field optical system are shown in Figure 9a,b.

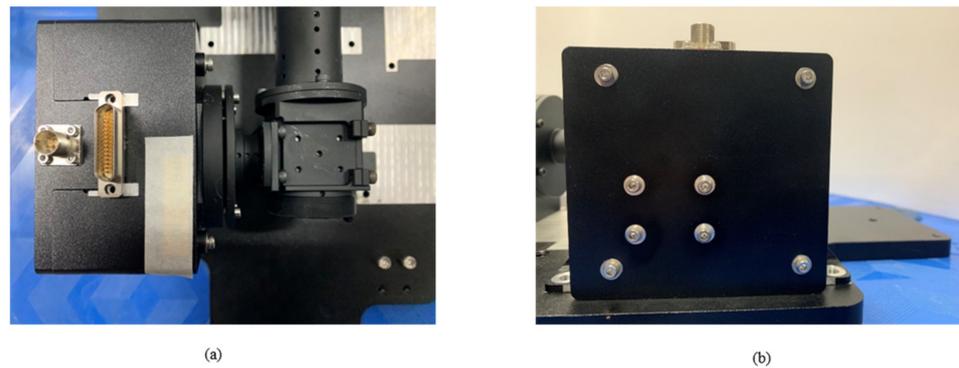


Figure 9. Wide-field optical system. (a) Top view; (b) left view.

We performed salient object detection on the images collected by the wide-field optical system. All images were taken in indoor (first two lines) and outdoor (last two lines) conditions according to the types of dolls, sculptures, landscapes and people. Figure 10 shows the results of our model along with another eight state-of-the-art methods [13–20] for visual comparisons. It can be seen from Figure 10 that for indoor scenes, the ability to separate salient targets and retain complete targets of our model is optimal. Moreover, in outdoor scenes with cluttered backgrounds, our model has shown a superior ability to extract salient objects.

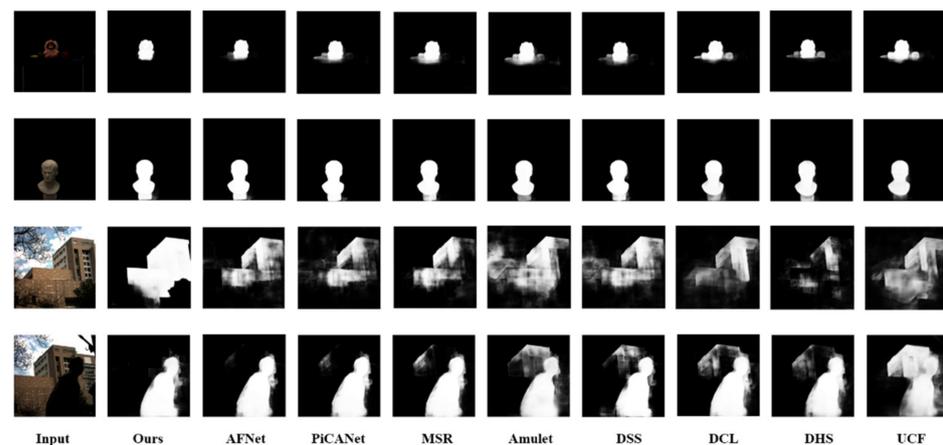


Figure 10. Results of salient object detection.

Overall, it can be seen from Figure 10 that our model is well equipped with a wide-field optical system and provides good prior information for subsequent tracking, segmentation and other tasks.

5. Conclusions

In this paper, we have presented a dual-branch feature fusion network for salient object detection. Our network uses a dual-branch structure to process low-level features and high-level features separately. The low-level features can effectively retain the supplementary information of the target through simple convolution and aggregation. By employing the global information perception module and local information concatenation module, the network can fully extract the local and global information of the target from the high-level features. Finally, the refinement fusion module merges the features of the two branches. The whole network can learn to capture the overall shape of objects, and experimental results demonstrate that the proposed architecture achieves state-of-the-art performance on two public saliency benchmarks. Moreover, we have developed a wide-field optical system with high imaging quality. Our model equipped with the wide-field optical system can achieve salient object detection with a large field of view and provide excellent prior

information for subsequent object tracking and image segmentation. We will try to combine salient object detection with subsequent computer vision tasks carried out by the wide-field optical system in the future.

Author Contributions: Methodology, Z.S.; validation, Z.S.; writing—original draft preparation, Z.S.; writing—review and editing, Q.L.; supervision, Q.L., Z.X. and H.F.; project administration, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Equipment Pre-Research Key Laboratory Fund Project grant number 61424080214. And the APC was funded by Zhejiang University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mittal, M.; Arora, M.; Pandey, T.; Goyal, L.M. Image segmentation using deep learning techniques in medical images. In *Advancement of Machine Intelligence in Interactive Medical Image Analysis*; Springer Nature: London, UK, 2020; pp. 41–63.
- Verma, O.P.; Roy, S.; Pandey, S.C.; Mittal, M. *Advancement of Machine Intelligence in Interactive Medical Image Analysis*; Springer Nature: London, UK, 2019.
- Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [[CrossRef](#)]
- Tie, L.; Jian, S.; Nanning, Z.; Xiaou, T.; HeungYeung, S. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 353–367.
- Goferman, S.; Zelnik-Manor, L.; Tal, A. Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1915–1926. [[CrossRef](#)] [[PubMed](#)]
- Mingming, C.; GuoXing, Z.; Mitra, N.J.; XiaoLei, H.; ShiMin, H. Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 569–582.
- Federico, P.; Krahenbuhl, P.; Pritch, Y.; Hornung, A. Saliency filters: Contrast based filtering for salient region detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Providence, RI, USA, 16–21 June 2012; pp. 733–740.
- Mingming, C.; Warrell, J.; WenYan, L.; Shuai, Z.; Vineet, V.; Crook, N. Efficient salient region detection with soft image abstraction. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 2–8 December 2013; pp. 1529–1536.
- Xiaodi, H.; Liqing, Z. Saliency detection: A spectral residual approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.
- Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S. Frequency-tuned salient region detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 22–24 June 2009; pp. 1597–1604.
- Li, G.; Yu, Y. Visual saliency based on multiscale deep features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5455–5463.
- Zhao, R.; Ouyang, W.; Li, H.; Wang, X. Saliency detection by multi-context deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1265–1274.
- Shengfeng, H.; Rynson, L.H.W.; Wenxi, L.; Zhe, H.; Qingxiong, Y. Supercnn: A superpixelwise convolutional neural network for salient object detection. *Int. J. Comput. Vis.* **2015**, *115*, 330–344.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- Qibin, H.; Mingming, C.; Xiaowei, H.; Borji, A.; Zhuowen, T.; Torr, P.H.S. Deeply supervised salient object detection with short connections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 3203–3212.
- Nian, L.; Junwei, H. Dhsnet: Deep hierarchical saliency network for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 678–686.
- Pingping, Z.; Dong, W.; Huchuan, L.; Hongyu, W.; Baocai, Y. Learning uncertain convolutional features for accurate saliency detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–27 October 2017; pp. 212–221.
- Guanbin, L.; Yizhou, Y. Deep contrast learning for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 478–487.
- Pingping, Z.; Dong, W.; Huchuan, L.; Hongyu, W.; Xiang, R. Amulet: Aggregating multi-level convolutional features for salient object detection. In Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22–27 October 2017; pp. 202–211.

20. Guanbin, L.; Yuan, X.; Liang, L.; Yizhou, Y. Instance level salient object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 2386–2395.
21. Nian, L.; Junwei, H.; Ming-Hsuan, Y. Picanet: Learning pixel-wise contextual attention for saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3089–3098.
22. Mengyang, F.; Huchuan, L.; Errui, D. Attentive feedback network for boundary-aware salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, NY, USA, 15–20 June 2019; pp. 1623–1632.
23. Kim, H.; Chae, E.; Jo, G.; Paik, J. Fisheye lens-based surveillance camera for wide field-of-view monitoring. In Proceedings of the IEEE International Conference on Consumer Electronics ICCE, Las Vegas, NV, USA, 9–12 January 2015; pp. 505–506.
24. Zhilai, L.; Donglin, X.; Xuejun, Z. Optical and mechanical design for long focal length and wide-field optical system. *Opt. Precis. Eng.* **2008**, *2008*, 12.
25. Kashima, S.; Hazumi, M.; Imada, H.; Katayama, N.; Matsumura, T.; Sekimoto, Y.; Sugai, H. Wide field-of-view crossed dragone optical system using anamorphic aspherical surfaces. *Appl. Opt.* **2018**, *57*, 4171–4179. [[CrossRef](#)] [[PubMed](#)]
26. Qiong, Y.; Li, X.; Jianping, S.; Jiaya, J. Hierarchical saliency detection, in CVPR. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–26 June 2013; pp. 1155–1162.
27. Lijun, W.; Huchuan, L.; Yifan, W.; Mengyang, F.; Dong, W.; Baocai, Y.; Xiang, R. Learning to detect salient objects with image-level supervision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 3796–3805.