Appendix: MFFN: Multi-view Feature Fusion Network for Camouflaged Object Detection

Section A shows the experimental details of MFFN on the salient object detection (SOD) task and the test results. We further add more experimental details of MFFN on the COD task and the evaluation metrics changes during training in Section B.

A. Experiments on Salient Object Detection

To illustrate the generalizations and soundness of proposed structural design, we evaluated the proposed model on the salient object detection (SOD) task.

A.1. Datasets

Specifically, we only use SOD dataset DUTS-TR [9] for MFFN model training, the model performance is evaluated on three existing datasets: ECSSD [12], HKU-IS [2], DUTS-TE [9]. The DUTS dataset contains 10553 training images(DUTS-TR) and 5019 test images(DUTS-TE). All the training images are collected from the ImageNet DET training/validation set, while the test images are collected from the ImageNet DET test set and the SUN dataset. The ECSSD dataset contains 1 000 images obtained from the Internet. This dataset is extended by Complex Scene Saliency Dataset (CSSD). Salient objects contain complex structures, and the background has certain complexity. HKU-IS contains 4 447 images, and each image in this dataset meets one of the following three criteria :1) contains multiple scattered salient objects; 2) At least one salient object is in the image boundary; 3) The apparent similarity between the salient objects and the background.

A.2. Implementation Details

In order to better evaluate the performance of our model, the compared algorithms are also only trained with DUTS-TR [9] and adopt the same hyperparameters and training strategies [8]. Specifically, the initial learning rate is set to 0.05, and follows a linear warm-up and linear decay strategy. In addition, batchsize is set to 8 and trained for 50 epochs, and SGD optimizer is used.

A.3. Comparisons with State-of-the-arts

We compare our proposed model with 10 existing SOD models. As shown in Tab. 1, our proposed model outperforms most of the competitors in five evaluation metrics, which indicates that our multi-view strategy can be effectively and efficiently generalized to other segmentation tasks.

B. Implementation Details and Evaluation Results on COD

B.1. Implementation Details

In this section, we explain the choice of hyperparameters. The weight λ of L_{UAL} was initially set as 1.5, and then the cosine strategy is adopted for dynamic adjustment. For comparison with the SOTA model ZoomNet [7], we set the initial image size to 384×384. The ratio adopted for distance views is 1.5 and 2.0. Finally, the size of FPN output by backbone is (12×12, 24×24, 48×48, 96×96, 192×192), and the number of channels is uniformly adjusted to 64. In addition, in the CFU module, the number of interaction groups in *Channel-wise Local Interaction Process* (CLIP) part is 3, and the step of progressive iteration in the *Overall Progressive Iteration* (OPI) is 4. We also encourage readers to experiment with other parameter settings.

B.2. Early Stopping

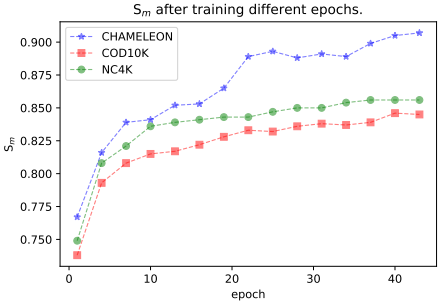
We do not focus on exploiting more epoches as there is no clear evidence that COD detectors will benefit from longer training. During our experiments, we found that the first time for the result dropping was appeared in approximate 40th epoch, as shown in Tab. 2 and illustrated in Fig. 1. We also provide the results between 40th epoch and 43rd epoch. To achieve a trade-off between performance and time consumption, we chose the results from the 40th epoch as our final evaluation results.

B.3. How to get and evaluat the results of our proposed MFFN?

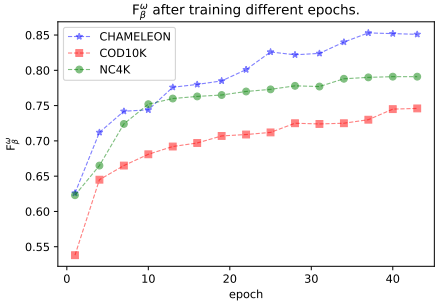
We use the open source COD evaluation tool to evaluate our prediction results, and we have submitted the test results of the COD10K dataset together with the supplementary material (due to size limitation, we cannot submit the test results of CHAMELEON and NC4K together). You can use the open source tool `https://github.com/DengPingFan/CODToolbox` for evaluation. The changes of five metrics (F_m, F_β^ω, MAE, F_β, E_m) over time (epoch) and **early stopping** are illustrated in Fig. 1a, 1b, 1c, 1d, 1e.

Table 1: Comparison of evaluation results of different Salient object detection(SOD) models on ECSSD [12], HKU-IS [2] and DUTS-TE [9]. The best results are highlighted in red, green and blue

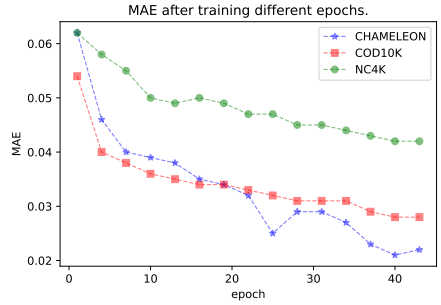
Model	Backbone	ECSSD					HKU-IS					DUTS-TE				
		$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$	$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$	$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$
PAGEnet [10]	Vgg16	0.912	0.886	0.042	0.904	0.947	0.903	0.865	0.037	0.884	0.948	0.854	0.769	0.052	0.793	0.896
PiCANet [4]	ResNet50	0.917	0.867	0.046	0.890	0.952	0.904	0.840	0.043	0.866	0.950	0.869	0.755	0.051	0.791	0.920
PoolNet [3]	ResNet50	0.926	0.904	0.035	0.918	0.956	0.919	0.888	0.030	0.903	0.958	0.887	0.817	0.037	0.840	0.926
HRS [13]	ResNet50	0.883	0.859	0.054	0.894	0.934	0.882	0.851	0.042	0.883	0.941	0.829	0.746	0.051	0.791	0.899
GCPANet [1]	ResNet50	0.927	0.903	0.035	0.916	0.955	0.920	0.889	0.031	0.901	0.958	0.891	0.821	0.038	0.841	0.929
SAMNet [6]	Handcraft	0.907	0.858	0.050	0.883	0.945	0.898	0.837	0.045	0.864	0.946	0.849	0.729	0.058	0.768	0.901
VST [5]	T2T-ViTt-14	0.932	0.910	0.033	0.920	0.964	0.928	0.897	0.029	0.907	0.968	0.896	0.828	0.037	0.845	0.939
Auto-MSFNet [14]	ResNet50	0.914	0.916	0.033	0.927	0.954	0.908	0.903	0.027	0.912	0.959	0.877	0.841	0.034	0.855	0.931
SGL-KRN [11]	ResNet50	0.923	0.910	0.036	0.924	0.954	0.921	0.904	0.028	0.915	0.961	0.893	0.847	0.034	0.865	0.939
CTDNet [15]	ResNet50	0.925	0.915	0.032	0.927	0.956	0.921	0.909	0.027	0.918	0.961	0.893	0.847	0.034	0.862	0.935
MINet [8]	ResNet50	0.925	0.911	0.033	0.923	0.957	0.919	0.897	0.029	0.909	0.960	0.884	0.825	0.037	0.844	0.927
MFFN(ours)	ResNet50	0.929	0.917	0.032	0.927	0.959	0.921	0.903	0.028	0.913	0.959	0.888	0.833	0.038	0.850	0.924



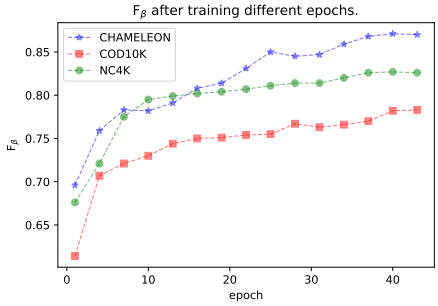
(a) The change of F_m with the increase of training epochs before the earlystopping.



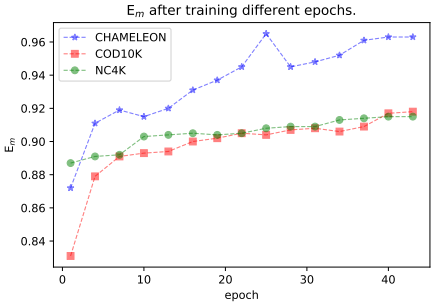
(b) The change of F_β^ω with the increase of training epochs before the earlystopping.



(c) The change of MAE with the increase of training epochs before the earlystopping.



(d) The change of F_β with the increase of training epochs before the earlystopping..



(e) The change of E_m with the increase of training epochs before the earlystopping.

Figure 1: Metrics evaluation and early stopping

Table 2: Our model performs earlystopping at epoch 43, and finally we choose the 40th epoch as our final result, and we provide the following table of the evaluation results between 40th and 43rd epoch.

epoch	CHAMELEON					COD10K					NC4K				
	$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$	$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$	$S_m \uparrow$	$F_\beta^\omega \uparrow$	$MAE \downarrow$	$F_\beta \uparrow$	$E_m \uparrow$
40	0.905	0.852	0.021	0.871	0.963	0.846	0.745	0.028	0.782	0.917	0.856	0.791	0.042	0.827	0.915
41	0.906	0.850	0.021	0.872	0.965	0.841	0.744	0.030	0.783	0.919	0.854	0.793	0.044	0.825	0.913
42	0.907	0.851	0.023	0.871	0.964	0.844	0.745	0.029	0.782	0.917	0.855	0.790	0.043	0.824	0.916
43	0.907	0.851	0.022	0.870	0.963	0.845	0.746	0.028	0.783	0.918	0.856	0.791	0.042	0.826	0.915

References

[1] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):10599–10606, Apr. 2020. 2

[2] Guanbin Li and Yizhou Yu. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 1, 2

[3] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[4] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2

[5] Nian Liu, Ni Zhang, Kaiyuan Wan, Ling Shao, and Junwei Han. Visual saliency transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4722–4732, October 2021. 2

[6] Yun Liu, Xin-Yu Zhang, Jia-Wang Bian, Le Zhang, and Ming-Ming Cheng. Samnet: Stereoscopically attentive multi-scale network for lightweight salient object detection. *IEEE Transactions on Image Processing*, 30:3804–3814, 2021. 2

[7] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan Lu. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022. 1

[8] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2

[9] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 2

[10] Wenguan Wang, Shuyang Zhao, Jianbing Shen, Steven C. H. Hoi, and Ali Borji. Salient object detection with pyramid attention and salient edges. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[11] Binwei Xu, Haoran Liang, Ronghua Liang, and Peng Chen. Locate globally, segment locally: A progressive architecture with knowledge review network for salient object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4):3004–3012, May 2021. 2

[12] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013. 1, 2

[13] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu. Towards high-resolution salient object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2

[14] Miao Zhang, Tingwei Liu, Yongri Piao, Shunyu Yao, and Huchuan Lu. Auto-msfnet: Search multi-scale fusion network for salient object detection. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 667–676. Association for Computing Machinery, 2021. 2

[15] Zhirui Zhao, Changqun Xia, Chenxi Xie, and Jia Li. Complementary trilateral decoder for fast and accurate salient object detection. In *Proceedings of the 29th acm international conference on multimedia*, pages 4967–4975, 2021. 2