Supplementary Material of Hierarchical Dynamic Filtering Network for RGB-D Salient Object Detection

Youwei Pang[1], Lihe Zhang[1⋆], Xiaoqi Zhao[1], and Huchuan Lu[1,2]

[1] Dalian University of Technology, China
{lartpang, zxq}@mail.dlut.edu.cn, {zhanglihe, lhchuan}@dlut.edu.cn
[2] Peng Cheng Laboratory

In **Table 2** of the original paper, we show the weighted average results of each model in terms of six metrics. In Sec. 1 and Sec 2 of this document, we respectively list the results of these models across different datasets.

This supplementary document is organized as follows:

- More details about the performance contributed by different components in the proposed HDFNet.
- More detailed comparisons of RGB SOD models with the HEL and without the HEL.

1 Ablation Study

Tab. 1 shows the performance improvement contributed by different components. Note that Model 2, 4 and 6 without the DDPM directly combine the features of dense transport layer into the decoder by element-wise addition instead of using convolution operation of Model 3, 5, 7.

The experiments in Tab 1 are divided into different groups:

1. **Model 2** vs. **Model 3**: Effectiveness of DDPM using only depth features to compute dynamic filters.
2. **Model 4** vs. **Model 5**: Effectiveness of DDPM using only RGB features to compute dynamic filters.
3. **Model 6** vs. **Model 7** vs. **Model 8**: Effectiveness of DDPM using two-modality features to compute dynamic filters.
4. **Model 9** vs. **Model 10** vs. **Model 11** vs. **Model 12**: Effectiveness of three components in the HEL (L_e, L_f and L_b) and the overall HEL.

2 Effectiveness of the HEL

Tab 2 shows the performance gains of the proposed loss function in some recent RGB saliency models [3,15,10,1]. Here, "AveMetric" still denotes the weighted average results on all datasets and is consistent with the data in **Table 2** of the original paper. It is worth noting that there are some differences in our experimental settings for these models. 1) R3Net [3]: We use ResNeXt-101 [16]

⋆ Corresponding author: zhanglihe@dlut.edu.cn

as the backbone as the original paper. We only change the supervision for the final prediction to the proposed HEL. 2) CPD [15]: ResNet-50 [5] is used as the backbone. For each branch, we use the proposed HEL to supervise the prediction. 3) PoolNet [10]: The backbone network is ResNet-50. We do not use the strategy of joint training with the edge and we apply the HEL on the final prediction. 4) GCPANet [1]: We also use ResNet-50, and the HEL to supervise the final result with the same resolution as the input.

Table 1. Ablation experiments of different components. "Model ⋆" corresponds to the model "No. ⋆" in **Table 2** of the original paper. In each set of comparative experiments, we emphasize the best test results in **red**.

	Metric	Baseline	DDPM							HEL			
			Depth Input		RGB Input		RGB-D Input			L_e	L_f	L_b	ALL
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
LFSD [8]	F_{max}	0.808	0.819	**0.832**	0.792	**0.825**	0.846	**0.851**	0.809	0.846	0.859	0.858	**0.860**
	F_{ada}	0.776	0.778	**0.788**	0.762	**0.796**	**0.816**	0.812	0.760	0.819	0.806	**0.837**	0.831
	F_β^ω	0.703	0.698	**0.716**	0.676	**0.722**	0.745	**0.747**	0.685	0.772	0.765	0.773	**0.792**
	MAE	0.116	0.117	**0.113**	0.124	**0.114**	**0.099**	0.101	0.129	0.092	0.092	0.089	**0.085**
	S_m	0.800	0.801	**0.815**	0.782	**0.816**	0.833	**0.835**	0.791	0.838	0.838	**0.847**	**0.847**
	E_m	0.846	0.845	**0.847**	0.832	**0.852**	**0.867**	0.860	0.826	0.869	0.867	0.878	**0.883**
NJUD [6]	F_{max}	0.889	0.897	**0.902**	0.891	**0.899**	0.909	**0.916**	0.892	**0.925**	0.920	0.916	0.924
	F_{ada}	0.840	0.848	**0.849**	0.851	**0.856**	0.858	**0.872**	0.843	**0.895**	0.867	0.886	0.894
	F_β^ω	0.801	0.807	**0.819**	0.809	**0.826**	0.827	**0.846**	0.808	0.876	0.858	0.857	**0.881**
	MAE	0.061	0.059	**0.055**	0.057	**0.054**	0.051	**0.047**	0.058	0.039	0.043	0.045	**0.037**
	S_m	0.880	0.887	**0.891**	0.884	**0.892**	0.898	**0.905**	0.884	**0.911**	0.903	0.906	**0.911**
	E_m	0.897	0.903	**0.905**	**0.907**	**0.907**	0.908	**0.916**	0.903	0.932	0.921	0.923	**0.934**
NLPR [12]	F_{max}	0.882	0.886	**0.889**	0.890	**0.899**	0.900	**0.908**	0.885	0.910	0.907	0.912	**0.917**
	F_{ada}	0.819	**0.810**	**0.810**	0.830	**0.845**	0.828	**0.848**	0.819	**0.882**	0.829	0.869	0.878
	F_β^ω	0.790	0.786	**0.801**	0.802	**0.824**	0.809	**0.835**	0.804	0.864	0.834	0.847	**0.869**
	MAE	0.042	0.042	**0.039**	0.038	**0.034**	0.038	**0.033**	0.039	0.029	0.032	0.030	**0.027**
	S_m	0.889	0.892	**0.899**	0.897	**0.906**	0.900	**0.915**	0.897	**0.916**	0.902	0.912	**0.916**
	E_m	0.925	**0.921**	**0.921**	0.929	**0.935**	0.927	**0.936**	0.925	0.948	0.929	**0.949**	0.948
RGBD135 [2]	F_{max}	0.892	**0.910**	0.904	0.889	**0.894**	0.917	**0.927**	0.879	0.931	0.933	0.926	**0.934**
	F_{ada}	0.834	0.864	**0.865**	**0.847**	0.841	0.883	**0.890**	0.828	0.917	0.873	0.906	**0.919**
	F_β^ω	0.770	0.807	**0.819**	0.774	**0.790**	0.835	**0.853**	0.770	0.889	0.873	0.855	**0.902**
	MAE	0.040	0.035	**0.034**	0.039	**0.037**	0.029	**0.027**	0.042	0.022	0.023	0.026	**0.020**
	S_m	0.873	0.898	**0.907**	0.873	**0.885**	0.910	**0.922**	0.878	0.925	0.925	0.915	**0.932**
	E_m	0.929	0.950	**0.959**	**0.934**	**0.934**	0.962	**0.965**	0.930	**0.974**	0.959	0.971	0.973
SIP [4]	F_{max}	0.865	0.875	**0.888**	0.875	**0.886**	0.888	**0.889**	0.861	0.897	0.899	0.897	**0.904**
	F_{ada}	0.810	0.821	**0.832**	0.832	**0.843**	**0.844**	**0.844**	0.806	**0.865**	0.840	0.863	0.863
	F_β^ω	0.753	0.761	**0.783**	0.772	**0.795**	0.788	**0.797**	0.749	0.829	0.812	0.819	**0.835**
	MAE	0.074	0.072	**0.064**	0.069	**0.063**	0.065	**0.062**	0.074	0.052	0.058	0.056	**0.050**
	S_m	0.853	0.858	**0.872**	0.858	**0.867**	0.867	**0.872**	0.849	**0.879**	0.870	0.878	0.878
	E_m	0.893	0.894	**0.903**	0.902	**0.909**	0.903	**0.906**	0.893	0.916	0.909	0.914	**0.920**
SSD [19]	F_{max}	0.844	0.859	**0.863**	0.825	**0.853**	**0.862**	0.861	0.841	0.877	**0.880**	0.875	0.872
	F_{ada}	0.770	**0.790**	0.786	0.782	**0.802**	0.799	**0.818**	0.791	0.843	0.811	0.828	**0.844**
	F_β^ω	0.730	0.744	**0.750**	0.735	**0.758**	0.759	**0.784**	0.739	0.806	0.787	0.789	**0.808**
	MAE	0.072	0.067	**0.063**	0.068	**0.059**	0.058	**0.054**	0.067	**0.047**	0.054	0.049	0.048
	S_m	0.841	0.854	**0.859**	0.840	**0.855**	0.859	**0.871**	0.846	**0.871**	0.863	0.870	0.866
	E_m	0.863	0.874	**0.876**	0.871	**0.897**	0.887	**0.902**	0.885	0.911	0.888	0.899	**0.913**
STEREO [11]	F_{max}	0.883	**0.871**	0.861	0.898	**0.906**	0.900	**0.909**	0.895	0.912	0.911	0.910	**0.918**
	F_{ada}	0.822	**0.806**	0.787	0.853	**0.863**	0.846	**0.857**	0.843	0.877	0.841	0.875	**0.879**
	F_β^ω	0.780	**0.762**	0.753	0.807	**0.829**	0.809	**0.829**	0.805	0.854	0.826	0.843	**0.863**
	MAE	0.062	**0.068**	0.071	0.055	**0.048**	0.053	**0.049**	0.055	0.042	0.048	0.045	**0.039**
	S_m	0.873	**0.865**	0.858	0.888	**0.901**	0.892	**0.901**	0.890	0.903	0.891	0.903	**0.906**
	E_m	0.901	**0.896**	0.883	0.919	**0.924**	0.918	**0.922**	0.914	0.932	0.917	0.931	**0.937**
DUTRGBD [13]	F_{max}	0.875	0.886	**0.901**	0.893	**0.916**	0.911	**0.920**	0.872	**0.926**	0.922	0.924	**0.926**
	F_{ada}	0.817	0.822	**0.846**	0.839	**0.872**	0.858	**0.871**	0.812	**0.892**	0.861	0.890	**0.892**
	F_β^ω	0.740	0.749	**0.784**	0.772	**0.817**	0.800	**0.820**	0.742	0.857	0.827	0.840	**0.865**
	MAE	0.079	0.076	**0.066**	0.067	**0.054**	0.059	**0.054**	0.077	0.044	0.052	0.050	**0.040**
	S_m	0.853	0.865	**0.877**	0.874	**0.892**	0.886	**0.895**	0.859	**0.908**	0.891	0.896	0.905
	E_m	0.893	0.899	**0.912**	0.911	**0.928**	0.915	**0.922**	0.896	0.937	0.921	0.930	**0.938**
AveMetric	F_{max}	0.875	0.879	**0.882**	0.884	**0.896**	0.898	**0.904**	0.878	0.909	0.909	0.907	**0.914**
	F_{ada}	0.819	**0.820**	**0.820**	0.839	**0.852**	0.846	**0.856**	0.823	**0.878**	0.845	0.874	**0.878**
	F_β^ω	0.768	0.768	**0.780**	0.787	**0.811**	0.803	**0.820**	0.777	0.849	0.827	0.836	**0.857**
	MAE	0.067	0.066	**0.063**	0.060	**0.054**	0.056	**0.052**	0.064	0.044	0.050	0.048	**0.041**
	S_m	0.865	0.868	**0.873**	0.874	**0.886**	0.884	**0.893**	0.871	**0.898**	0.887	0.895	**0.898**
	E_m	0.898	0.899	**0.900**	0.909	**0.916**	0.913	**0.918**	0.903	0.929	0.916	0.926	**0.933**

Table 2. Comparisons of RGB SOD models with the HEL (w) and without the HEL (w/o). The best result of each group is highlight in **red**.

	Metric	**R3Net**$_{18}$ [3]		**CPD**$_{19}$ [15]		**PoolNet**$_{19}$ [10]		**GCPANet**$_{20}$ [1]	
		w/o	w	w/o	w	w/o	w	w/o	w
DUTS [14]	F_{max}	0.823	**0.827**	0.858	**0.859**	0.844	**0.879**	0.856	**0.865**
	F_{ada}	0.688	**0.710**	0.786	**0.807**	0.748	**0.818**	0.759	**0.777**
	F_β^ω	0.701	**0.726**	0.774	**0.800**	0.733	**0.814**	0.744	**0.778**
	MAE	0.071	**0.069**	0.047	**0.045**	0.055	**0.040**	0.055	**0.049**
	S_m	0.821	**0.826**	0.863	**0.865**	0.849	**0.873**	0.860	**0.864**
	E_m	0.818	**0.833**	0.889	**0.903**	0.864	**0.909**	0.867	**0.881**
DUT-OMRON [18]	F_{max}	0.785	**0.794**	**0.799**	0.797	0.778	**0.807**	0.798	**0.806**
	F_{ada}	0.668	**0.685**	0.739	**0.749**	0.699	**0.754**	0.713	**0.726**
	F_β^ω	0.669	**0.693**	0.714	**0.732**	0.668	**0.736**	0.689	**0.715**
	MAE	0.079	**0.078**	0.059	**0.058**	0.066	**0.054**	0.070	**0.066**
	S_m	0.812	**0.816**	**0.828**	0.826	0.810	**0.829**	**0.826**	**0.826**
	E_m	0.804	**0.818**	0.864	**0.869**	0.839	**0.874**	0.841	**0.851**
ECSSD [17]	F_{max}	0.927	**0.932**	0.934	**0.939**	0.921	**0.943**	**0.933**	**0.933**
	F_{ada}	0.858	**0.866**	0.908	**0.918**	0.882	**0.919**	0.892	**0.896**
	F_β^ω	0.858	**0.882**	0.881	**0.906**	0.846	**0.905**	0.863	**0.882**
	MAE	0.053	**0.045**	0.044	**0.035**	0.054	**0.037**	0.049	**0.042**
	S_m	0.908	**0.910**	0.911	**0.918**	0.897	**0.918**	**0.912**	**0.912**
	E_m	0.911	**0.918**	0.941	**0.951**	0.920	**0.947**	0.930	**0.935**
HKU-IS [7]	F_{max}	0.914	**0.917**	0.922	**0.927**	0.915	**0.931**	0.923	**0.928**
	F_{ada}	0.842	**0.856**	0.886	**0.899**	0.869	**0.903**	0.878	**0.885**
	F_β^ω	0.831	**0.857**	0.864	**0.892**	0.838	**0.893**	0.851	**0.876**
	MAE	0.047	**0.040**	0.037	**0.030**	0.043	**0.029**	0.041	**0.035**
	S_m	0.890	**0.895**	0.904	**0.911**	0.896	**0.911**	0.907	**0.910**
	E_m	0.918	**0.929**	0.945	**0.956**	0.937	**0.958**	0.943	**0.947**
PASCAL-S [9]	F_{max}	**0.844**	0.836	**0.868**	0.867	0.850	**0.873**	0.856	**0.857**
	F_{ada}	0.757	**0.764**	0.819	**0.829**	0.786	**0.829**	0.798	**0.803**
	F_β^ω	0.733	**0.750**	0.784	**0.806**	0.745	**0.806**	0.764	**0.782**
	MAE	0.101	**0.091**	0.079	**0.072**	0.093	**0.073**	0.087	**0.081**
	S_m	0.813	**0.822**	0.842	**0.844**	0.822	**0.843**	**0.837**	0.835
	E_m	0.823	**0.832**	0.872	**0.889**	0.844	**0.878**	0.856	**0.863**
AveMetric	F_{max}	0.828	**0.832**	0.848	**0.849**	0.832	**0.861**	0.847	**0.854**
	F_{ada}	0.714	**0.731**	0.790	**0.804**	0.755	**0.811**	0.766	**0.779**
	F_β^ω	0.716	**0.740**	0.769	**0.792**	0.728	**0.799**	0.744	**0.773**
	MAE	0.072	**0.069**	0.052	**0.049**	0.060	**0.046**	0.061	**0.055**
	S_m	0.831	**0.835**	0.856	**0.857**	0.841	**0.862**	0.854	**0.856**
	E_m	0.830	**0.844**	0.889	**0.898**	0.865	**0.902**	0.869	**0.880**

References

1. Chen, Z., Xu, Q., Cong, R., Huang, Q.: Global context-aware progressive aggregation network for salient object detection. In: AAAI Conference on Artificial Intelligence (2020) 1, 2, 3
2. Cheng, Y., Fu, H., Wei, X., Xiao, J., Cao, X.: Depth enhanced saliency detection method. In: Proceedings of the International Conference on Internet Multimedia Computing and Service. pp. 23–27 (2014) 2

3. Deng, Z., Hu, X., Zhu, L., Xu, X., Qin, J., Han, G., Heng, P.A.: R3net: Recurrent residual refinement network for saliency detection. In: International Joint Conference on Artificial Intelligence. pp. 684–690 (2018) 1, 3
4. Fan, D.P., Lin, Z., Zhao, J.X., Liu, Y., Zhang, Z., Hou, Q., Zhu, M., Cheng, M.M.: Rethinking rgb-d salient object detection: Models, datasets, and large-scale benchmarks. arXiv preprint arXiv:1907.06781 (2019) 2
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016) 2
6. Ju, R., Liu, Y., Ren, T., Ge, L., Wu, G.: Depth-aware salient object detection using anisotropic center-surround difference. Signal Processing: Image Communication **38**, 115–126 (2015) 2
7. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 5455–5463 (2015) 3
8. Li, N., Ye, J., Ji, Y., Ling, H., Yu, J.: Saliency detection on light field. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 2806–2813 (2014) 2
9. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 280–287 (2014) 3
10. Liu, J.J., Hou, Q., Cheng, M.M., Feng, J., Jiang, J.: A simple pooling-based design for real-time salient object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2019) 1, 2, 3
11. Niu, Y., Geng, Y., Li, X., Liu, F.: Leveraging stereopsis for saliency analysis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 454–461 (2012) 2
12. Peng, H., Li, B., Xiong, W., Hu, W., Ji, R.: Rgbd salient object detection: a benchmark and algorithms. In: Proceedings of European Conference on Computer Vision. pp. 92–109 (2014) 2
13. Piao, Y., Ji, W., Li, J., Zhang, M., Lu, H.: Depth-induced multi-scale recurrent attention network for saliency detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7254–7263 (2019) 2
14. Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., Ruan, X.: Learning to detect salient objects with image-level supervision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 136–145 (2017) 3
15. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3907–3916 (2019) 1, 2, 3
16. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1492–1500 (2017) 1
17. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1155–1162 (2013) 3
18. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3166–3173 (2013) 3
19. Zhu, C., Li, G.: A three-pathway psychobiological framework of salient object detection using stereoscopic technology. In: International Conference on Computer Vision Workshops. pp. 3008–3014 (2017) 2