SPSN: Superpixel Prototype Sampling Network for RGB-D Salient Object Detection (Supplementary Material)

Minhyeok Lee*, Chaewon Park*, Suhwan Cho, and Sangyoun Lee

Yonsei University, Seoul, Korea
{hydragon516, chaewon28, chosuhwan, syleee}@yonsei.ac.kr

1 Visualization of our results

In this section, we supplement the qualitative results of the proposed SPSN. Fig. 1 visualizes our results in various challenging scenes. AM_{pred} is the auxiliary prediction superpixel map, which is the sampling result of PSNM, and AM_{gt} is the auxiliary ground truth superpixel map mentioned in Section 3.4 of the paper. Furthermore, $RelyW_R$ and $RelyW_D$ are values representing the reliability of RGB feature maps and depth feature maps, respectively, and are mentioned in Section 3.5 of the paper. As shown in Fig. 1, the proposed SPSN can properly sample only the salient prototypes in complex scenes or scenes that contain multiple objects. Moreover, through $RelyW_R$ and $RelyW_D$ values, our method adaptively weighs feature maps based on RGB and depth reliability. Consequently, our model is robust to low-quality depth maps and the inconsistency between RGB images and depth maps, leading to excellent performance.

2 Qualitative comparison

This section supplements the qualitative results of our model. In Fig. 2, we compare our outputs with those of the following eight state-of-the-art methods: DCF [2], D2F [7], CASG [5], PGAR [1], CMWM [4], CoN [3], CIM [8], and DMRA [6]. We selected 12 challenging scenes to validate the accuracy of our model. Generally, all the results show that our model accurately generates saliency maps, which is because our model precisely discriminates the foreground objects and the background by utilizing PSNM. Furthermore, the results in Fig. 2 show that our model is capable of selecting the more reliable modality between RGB image and depth map. Especially, our model choose to rely more on the depth map for samples with indiscriminate RGB images caused by long distance to the foreground, multiple objects, and low lighting (e.g., the fourth, the eighth, and the ninth row). Similarly, the mechanism works same for input images with unreliable depth maps, applying more weight to the RGB images when generating saliency maps (e.g., the first and the fifth row). Also, it is observed that our model is robust to samples with camouflaged objects (e.g., the first row).

* These authors contributed equally

Table 1. Statistical comparisons of the prototype sampling methods

Index	Method	NJU2K [25]				NLPR [35]			
		$E_\xi \uparrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$M \downarrow$	$E_\xi \uparrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$M \downarrow$
(a)	Baseline	.904	.863	.869	.051	.912	.877	.842	.037
(b)	Random	.922	.859	.870	.048	.923	.878	.861	.032
(c)	Grid	.923	.864	.872	.048	.938	.881	.867	.031
(d)	Superpixel	**.925**	**.870**	**.877**	**.046**	**.943**	**.887**	**.874**	**.030**

Table 2. Performance comparison according to the combination of proposed modules

Index	Method							NJU2K [25]				NLPR [35]			
	FFM	PGM	PSNM Part A	PSNM Part B	PSNM Part C	PSNM Part D	RSM	E	S	F	M	E	S	F	M
(a)		✓		MLP	✓	✓		.925	.870	.877	.046	.943	.887	.874	.030
(b)		✓		MLP		✓		.914	.865	.871	.049	.928	.881	.868	.033
(c)		✓		MLP	✓	✓	✓	.930	.878	.880	.045	.945	.899	.876	.028
(d)	✓	✓		MLP	✓	✓		.929	.881	.879	.044	.943	.898	.880	.028
(e)	✓	✓		EdgeConv	✓	✓		.938	.908	.903	.038	.951	.911	.893	.026
(f)		✓	✓	EdgeConv	✓	✓		.943	.912	.907	.036	.953	.916	.900	.025
(g)	✓	✓	✓	EdgeConv	✓	✓		.944	.912	.910	.035	.954	.919	.902	.025
(h)	✓	✓	✓	EdgeConv	✓	✓	✓	**.950**	**.918**	**.920**	**.032**	**.958**	**.923**	**.910**	**.023**

3 Supplementation of ablation study

Effects of superpixel Table 1 shows the performance of different prototype extraction methods. Index (a) in the table is the baseline model without using the prototype sampling method. The random sampling method (b) considers random pixels in an image as superpoints and generates a prototype from that coordinate. The grid method (c) indicates that the prototypes are generated using evenly divided square masks from an image. In addition, we remove ASPP and multi-scale fusion architecture of FFM, Part A of PSNM, and RSM to compare only the effect of differences in the sampling method. For a fair comparison, we set the number of prototypes used in methods (b), (c), and (d) to the same value, $N_S = 100$. As shown in Table 1, our proposed superpixel-based component sampling method outperformed the other methods, demonstrating its strong ability to capture the common properties of a group of images.

Effects of the proposed modules Table 2 shows the performance results of various combinations of the proposed modules. Table 1 (d) and Table 2 (a) refer to identical experiments. Furthermore, we replaced the EdgeConv layer of Part B with the MLP layer in some cases.

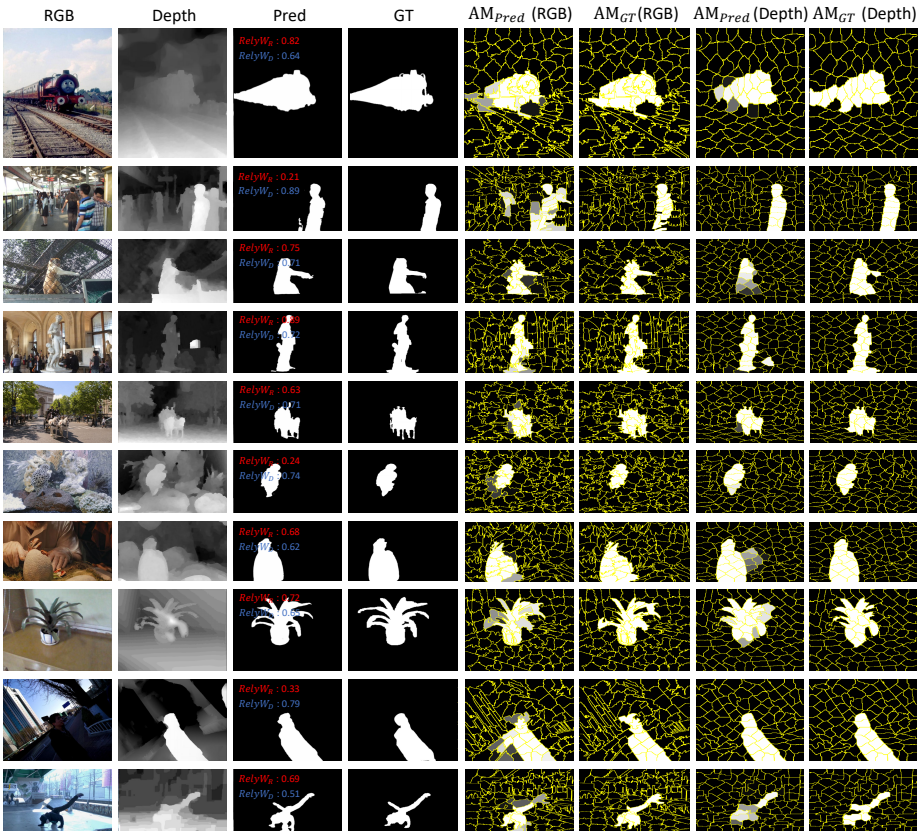

Fig. 1. Visualization of our results in challenging situations. AM_{pred} and AM_{GT} are described in Section 3.4, and $RelyW_R$ and $RelyW_D$ are described in Section 3.5 in the paper.

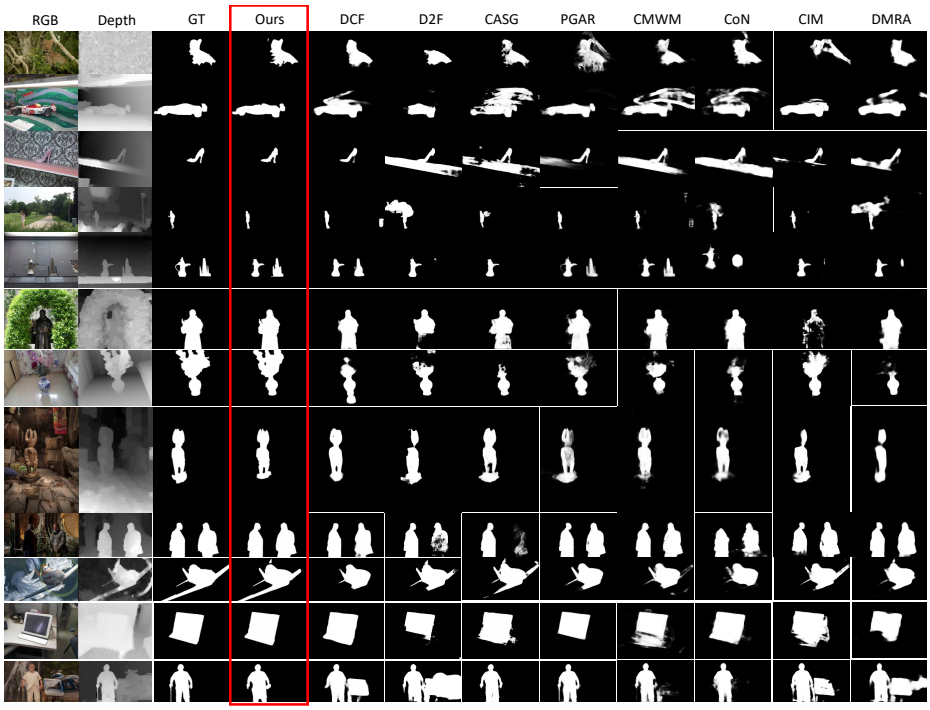

Fig. 2. Qualitative comparison with eight advanced networks in 12 challenging situations.

References

1. Chen, S., Fu, Y.: Progressively guided alternate refinement network for rgb-d salient object detection. In: European Conference on Computer Vision. pp. 520–538. Springer (2020)
2. Ji, W., Li, J., Yu, S., Zhang, M., Piao, Y., Yao, S., Bi, Q., Ma, K., Zheng, Y., Lu, H., et al.: Calibrated rgb-d salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9471–9481 (2021)
3. Ji, W., Li, J., Zhang, M., Piao, Y., Lu, H.: Accurate rgb-d salient object detection via collaborative learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16. pp. 52–69. Springer (2020)
4. Li, G., Liu, Z., Ye, L., Wang, Y., Ling, H.: Cross-modal weighting network for rgb-d salient object detection. In: European Conference on Computer Vision. pp. 665–681. Springer (2020)
5. Luo, A., Li, X., Yang, F., Jiao, Z., Cheng, H., Lyu, S.: Cascade graph neural networks for rgb-d salient object detection. In: European Conference on Computer Vision. pp. 346–364. Springer (2020)
6. 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/CVF International Conference on Computer Vision. pp. 7254–7263 (2019)
7. Sun, P., Zhang, W., Wang, H., Li, S., Li, X.: Deep rgb-d saliency detection with depth-sensitive attention and automatic multi-modal fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1407–1417 (2021)
8. Zhang, M., Ren, W., Piao, Y., Rong, Z., Lu, H.: Select, supplement and focus for rgb-d saliency detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3472–3481 (2020)