Source-free Depth for Object Pop-out
– Supplementary Material –

Abstract

In this supplementary material, we provide more details and experimental results to complement the manuscript. We provide in Section 1 the details of the evaluation metrics, in Section 2 cross-dataset cross-task analysis on the performance in max F-measure, in Section 3 more comparisons against SOTA SOD models, in Section 4 more comparisons against SOTA COD models, in Section 5 the details of our Poping network, in Section 6 more qualitative comparisons, in Section 7 the fair comparison with SOTA methods across different resolutions, in Section 8 sensitivity analysis on the hyperparameters, and in Section 9 evaluation with the same backbone.

1. Evaluation Metrics

We evaluate our performance with four generally-recognized metrics: F-measure is a region-based similarity metric that takes into account both Precision (P) and Recall (R). Mathematically, we have F-measure: $F_\beta = \frac{(1+\beta^2)\cdot P\cdot R}{\beta^2\cdot P+R}$. The value of β^2 is set to be 0.3 as suggested in [1] to emphasize the precision. In this paper, we report the **maximum F-measure** score, denoted as (F_m), across the binary maps of different thresholds. **Mean Absolute Error** (**M**) measures the approximation degree between the saliency map and ground-truth map at the pixel level. **S-measure** (**S_m**) [4] evaluates the similarities between object-aware (S_o) and region-aware (S_r) structures of the saliency map compared to the ground truth. Mathematically, we have: $S_m = \alpha \cdot S_o + (1-\alpha)\cdot S_r$, where α is set to be 0.5. **E-measure** (**E_m**) evaluates both image-level statistics and local pixel-matching information. Mathematically, we have: $E_m = \frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\phi_{FM}(i,j)$, where $\phi_{FM}(i,j)$ stands for the enhanced-alignment matrix as presented in [5].

2. Cross-Dataset Cross-Task Comparison

We present in Figures 1 & 2 the performance of different methods in max F-measure on COD and SOD tasks. We compare in total **46 methods** (•), where our method offers state-of-the-art results despite their task specialization. Details and more comparisons can be found in Tables 2 & 1.

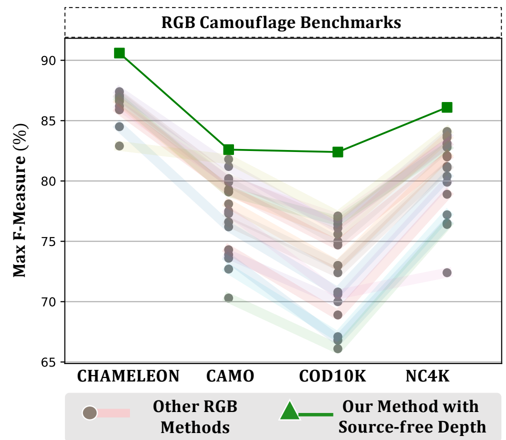

Figure 1. Comparison against 27 RGB SOTA COD models.

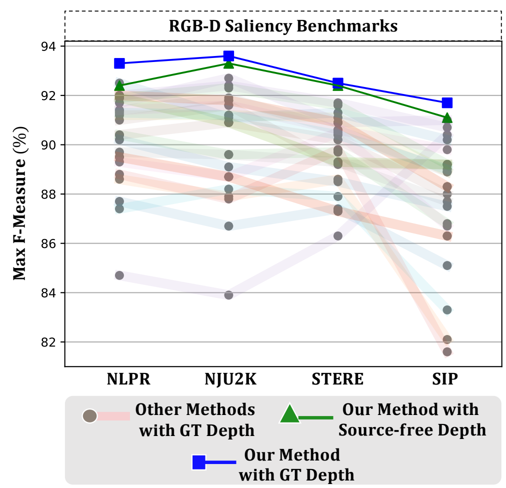

Figure 2. Comparison against 19 RGB-D SOTA SOD models.

3. Detailed Comparisons on SOD

We also provide an exhaustive comparison against SOTA RGB-D SOD models, as shown in Table 1. It is important to notice that our method with source-free depth (✗) already achieves better performance than most competing RGB-D methods with ground truth depth (✓). When using the ground truth depth on a par with other RGB-D methods, our method further improves the performance by a significant margin.

Table 1. Quantitative comparison on RGB-D SOD datasets. ↑ (↓) denotes that the higher (lower) is better. We use the Mean Absolute Error (M), max F-measure (F_m), S-measure (S_m), and max E-measure (E_m) as evaluation metrics. GD stands for GT Depth. **Bold** denotes the best performance. Underline denotes the second-best performance. **Ours** stands for our method with 352×352 resolution, which is on par with other counterparts. **Ours+** stands for our method with 512×512 resolution.

GD	Public.	Dataset	NLPR [34]				NJUK [19]				STERE [30]				SIP [9]			
		Metric	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$
Performance of RGB-D Models Trained with GT Depth																		
✓	$CVPR_{19}$ [53]	CPFP	.036	.867	.888	.932	.053	.877	.878	.923	.051	.874	.879	.925	.064	.851	.850	.903
✓	$ICCV_{19}$ [35]	DMRA	.031	.879	.899	.947	.051	.886	.886	.927	.047	.886	.886	.938	.085	.821	.806	.875
✓	TIP_{20} [40]	DRLF	.031	.904	.903	.929	.055	.896	.886	.913	.050	.897	.887	.916	.070	.868	.850	.881
✓	$ECCV_{20}$ [10]	BBSNet	.023	.918	.930	.961	.035	.920	.921	.949	.041	.909	.908	.942	.055	.883	.879	.922
✓	$ECCV_{20}$ [31]	HDFNet	.031	.839	.898	.942	.051	.847	.885	.920	.039	.863	.906	.937	.050	.904	.878	.920
✓	$ECCV_{20}$ [55]	DANet	.028	.916	.915	.953	.045	.910	.899	.935	.043	.892	.901	.937	.054	.892	.875	.918
✓	$ECCV_{20}$ [17]	CoNet	.031	.887	.908	.945	.046	.893	.895	.937	.040	.905	.908	**.949**	.063	.867	.858	.913
✓	$ECCV_{20}$ [23]	CMMS	.027	.896	.915	.949	.044	.897	.900	.936	.043	.893	.895	.939	.058	.877	.872	.911
✓	$ECCV_{20}$ [25]	CMWNet	.029	.912	.917	.941	.045	.912	.902	.924	.043	.911	.905	.929	.062	.889	.867	.901
✓	$CVPR_{20}$ [36]	A2dele	.029	.882	.898	.944	.051	.874	.871	.916	.044	.879	.878	.928	.070	.833	.828	.889
✓	$CVPR_{20}$ [11]	JLDCF	.021	.925	.925	.954	.041	.911	.902	.935	.040	.913	.902	.933	.049	.902	.880	.918
✓	TMM_{21} [15]	EBFSPI	.028	.887	.909	.940	.038	.895	.907	.936	.041	.873	.900	.926	.052	.863	.877	.911
✓	TMM_{21} [58]	CCAFNet	.020	.909	.928	.960	.030	.919	.920	.925	.033	.893	.908	.927	.043	.892	.886	.924
✓	TIP_{21} [42]	DSNet	.024	.924	.926	.951	.034	.928	.921	.946	.036	.922	.915	.941	.051	.899	.876	.910
✓	TIP_{21} [52]	BIANet	.032	.888	.900	.930	.056	.878	.867	.898	.048	.898	.895	.918	.091	.816	.802	.847
✓	TIP_{21} [24]	HAINet	.024	.920	.924	.956	.037	.924	.911	.940	.040	.917	.907	.938	.052	.907	.879	.917
✓	$TNNLS_{21}$ [9]	D3Net	.029	.904	.911	.942	.046	.909	.899	.927	.044	.902	.906	.925	.063	.880	.860	.897
✓	$TPAMI_{21}$ [43]	MobileSal	.025	.916	.920	.961	.041	.914	.905	.942	.041	.906	.903	.940	.053	.898	.873	.916
✓	$AAAI_{21}$ [2]	RD3D	.022	.927	.930	.959	.036	.923	.916	.941	.037	.917	.911	.939	.048	.905	.885	.918
✓	MM_{21} [27]	TriTransNet	.020	.909	.928	.960	.030	.919	.920	.925	.033	.893	.908	.927	.043	.892	.886	.924
✓	$CVPR_{21}$ [16]	DCFNet	.021	.891	.920	.957	.035	.902	.905	.924	.039	.885	.903	.927	.051	.875	.873	.920
✓	$ICCV_{21}$ [26]	VST	.023	.918	.930	.961	.035	.920	.921	.949	.041	.909	.908	.942	.055	.883	.879	.922
✓	TIP_{22} [3]	CIRNet	.028	.887	.909	.940	.038	.895	.907	.936	.041	.873	.900	.926	.052	.863	.877	.911
✓	$ECCV_{22}$ [21]	SPSN	.023	.917	.923	.956	.032	.927	.918	.949	.035	.909	.906	.941	.043	.910	.891	.932
✗	**Ours**	**PopNet**	.022	.925	.926	.956	.031	.931	.920	.949	.032	.922	.916	.947	.046	.911	.885	.926
✗	**Ours +**	**PopNet**	.023	.924	.926	.954	.031	.933	.922	.951	.032	.924	.917	.947	.044	.911	.890	.927
✓	**Ours**	**PopNet**	**.019**	.927	.932	.963	.030	**.936**	.924	.952	.033	.924	.917	.947	**.040**	**.923**	**.897**	**.937**
✓	**Ours+**	**PopNet**	.018	**.933**	**.934**	**.964**	**.029**	**.936**	**.925**	**.953**	**.031**	**.925**	**.918**	**.949**	.042	.917	.894	.933

4. Detailed Comparisons on COD

We show in Table 2 more comparisons against the SOTA COD models to complement the manuscript. It can be seen that our method performs favorably against all existing COD models by a large margin, especially with the S-measure and E-measure. The superior performances on these two metrics show that our method can efficiently leverage the depth cues to better preserve the object structure. Our method with higher resolution, denoted as **Ours+**, can further improve the performance, showing that we can efficiently deal with inputs with different resolutions. More detailed comparisons can be found in Section 7 Table 4.

5. Details of our Poping Network

Our proposed PopNet follows the conventional encoder-decoder design with skip connection by addition. The encoder can be any classical backbone. We have tested with ResNet-18 [14] for its lightweight architecture, as well as Res2Net-50 [12] for its great performance. Our decoder is composed of 5 layers. Each layer contains Conv2D, BN, ReLU, and upsampling. While the encoder is with ResNet-18 (**R_18**), the dimensions for decoding convolutions are 512, 256, 128, 64, 64, and 1. The model size becomes 48.7 MB with around 12.7M additional learning parameters. The additional computational cost over our baseline [57] reduces the FPS from 14 to 10, while the performance gain is significant as shown in Figure 2 of the main manuscript. This trade-off is very encouraging. While the encoder is with Res2Net-50 (**R2_50**), the dimensions for decoding convolutions are 2048, 1024, 512, 256, 64, 1. The model size becomes 185.5 MB with around 48.5M additional learning parameters. By replacing **R_18** with **R2_50**, we can further boost our performance as shown in Table 3.

6. Qualitative Comparison

We provide in Figure 3 more qualitative comparisons on challenging scenarios. It can be seen that our network can better preserve the object structure and deal with occlusion. This can be attributed to our depth poping network which brings the object to the background surface, making it easier to be popped out and segmented from the camouflaged scene.

7. Towards Higher Resolution

We provide more experimental results on different resolutions in Table 4. We retrain all the methods in an end-to-end manner with only RGB inputs, without requiring the GT depth. It can be seen that our network can more effectively handle high-resolution images and the results are very promising compared to the SOTA counterparts. While with the same resolution, our method outperforms the others with large margins. Our method also achieves the best trade-off between performance and computational cost. The

Table 2. Quantitative comparison on benchmark COD datasets. ↑ (↓) denotes that the higher (lower) is better. We use the Mean Absolute Error (M), max F-measure (F_m), S-measure (S_m), and max E-measure (E_m) as evaluation metrics. **Bold** denotes the best performance. Underline denotes the second-best performance. **Ours** stands for our method with 352×352 resolution, which is on par with other counterparts. **Ours+** stands for our method with 512×512 resolution.

Pseudo	Public.	Dataset	CAMO [20]				CHAMELEON [37]				COD10K [7]				NC4K [28]			
		Metric	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$
Performance of RGB COD Models																		
✓	$CVPR_{19}$ [44]	CPD	.113	.727	.726	.754	-	-	-	-	.057	.671	.748	.799	.074	.772	.788	.829
✓	$ICCV_{19}$ [54]	EGNet	.109	.736	.732	.801	-	-	-	-	.060	.668	.736	.809	.075	.765	.777	.840
✓	$ICCV_{19}$ [45]	SCRN	.089	.781	.778	.816	-	-	-	-	.046	.730	.789	.840	.059	.821	.830	.869
✓	$AAAI_{20}$ [41]	F3Net	.108	.703	.711	.752	-	-	-	-	.051	.661	.738	.806	.069	.764	.780	.833
✓	$CVPR_{20}$ [56]	ITSD	.101	.743	.748	.797	-	-	-	-	.059	.689	.766	.833	.063	.789	.810	.861
✓	$CVPR_{20}$ [33]	MINetR	.095	.739	.737	.770	-	-	-	-	.043	.700	.766	.831	.060	.799	.804	.851
✓	$CVPR_{20}$ [49]	UCNet	.093	.766	.739	.789	-	-	-	-	.041	.724	.776	.860	.055	.811	.811	.874
✗	$CVPR_{20}$ [7]	SINet	.099	.762	.751	.790	.044	.845	.868	.908	.051	.708	.771	.832	.058	.804	.808	.873
✓	$ECCV_{20}$ [13]	CSNet	.091	.775	.770	.813	-	-	-	-	.047	.706	.775	.838	.087	.724	.750	.777
✓	$MICCAI_{20}$ [8]	PraNet	.094	.773	.769	.827	-	-	-	-	.045	.730	.789	.862	.058	.812	.822	.877
✗	$CVPR_{21}$ [28]	SLSR	.080	.791	.787	.843	.030	.866	.889	.938	.037	.756	.804	.854	.048	.836	.839	.898
✗	$CVPR_{21}$ [47]	MGL-R	.088	.791	.775	.820	.031	.868	.893	.932	.035	.767	.813	.874	.053	.828	.832	.876
✗	$CVPR_{21}$ [29]	PFNet	.085	.793	.782	.845	.033	.859	.882	.927	.040	.747	.800	.880	.053	.820	.829	.891
✗	$CVPR_{21}$ [22]	UJSC	**.072**	.812	.800	.861	.030	.874	.891	.948	.035	.761	.808	.886	.047	.838	.841	.900
✗	$IJCAI_{21}$ [38]	C2FNet	.079	.802	.796	.856	.032	.871	.888	.936	.036	.764	.813	.894	.049	.831	.838	.898
✗	$ICCV_{21}$ [46]	UGTR	.086	.800	.783	.829	.031	.862	.887	.926	.036	.769	.816	.873	.052	.831	.839	.884
✗	$CVPR_{22}$ [18]	SegMAR	.080	.799	.794	.857	.032	.871	.887	.935	.039	.750	.799	.876	.050	.828	.836	.893
✗	$CVPR_{22}$ [32]	ZoomNet	.074	.818	.801	.858	.033	.829	.859	.915	.034	.771	.808	.872	.045	.841	.843	.893
Performance of RGB-D Models Retrained with Source-free Depth																		
✓	$ECCV_{20}$ [10]	BBSNet	.088	.783	.779	.821	.040	.859	.876	.917	.044	.735	.763	.840	.059	.818	.825	.869
✓	MM_{21} [48]	CDINet	.100	.638	.732	.766	.036	.787	.879	.903	.044	.610	.778	.821	.067	.697	.793	.830
✓	$CVPR_{21}$ [16]	DCF	.089	.724	.749	.834	.037	.821	.850	.923	.040	.685	.766	.864	.061	.765	.791	.878
✓	TIP_{21} [24]	HAINet	.084	.782	.760	.829	.028	.876	.876	.942	.049	.735	.781	.865	.057	.809	.804	.872
✓	$ICCV_{21}$ [50]	CMINet	.087	.798	.782	.827	.032	.881	.891	.930	.039	.768	.811	.868	.053	.832	.839	.888
✓	$ICCV_{21}$ [57]	SPNet	.083	.807	.783	.831	.033	.872	.888	.930	.037	.776	.808	.869	.054	.828	.825	.874
✓	TIP_{22} [39]	DCMF	.115	.737	.728	.757	.059	.807	.830	.853	.063	.679	.748	.776	.077	.782	.794	.820
✓	$ECCV_{22}$ [21]	SPSN	.084	.782	.773	.829	.032	.866	.887	.932	.042	.727	.789	.854	.059	.803	.813	.867
✓	**Ours**	**PopNet**	.073	.821	.806	**.869**	.022	.893	.910	.962	.031	.789	.827	.897	.043	.852	.852	.908
✓	**Ours+**	**PopNet**	.076	**.826**	**.808**	.863	**.020**	**.907**	**.917**	**.968**	**.027**	**.824**	**.851**	**.914**	**.042**	**.861**	**.861**	**.913**

Table 3. Quantitative results with different encoders for our Poping Network. **R_18** and **R2_50** stand for ResNet-18 [14] and Res2Net-50 [12] as the popping network backbone, respectively.

Pseudo	Public.	Dataset	CAMO [20]				CHAMELEON [37]				COD10K [7]				NC4K [28]			
		Metric	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$
✓	**Ours+ (R_18)**	**PopNet**	.078	**.829**	.802	.854	.024	.896	.908	.952	.028	.814	.844	.910	**.042**	.860	.857	.908
✓	**Ours+ (R2_50)**	**PopNet**	**.076**	.826	**.808**	**.863**	**.020**	**.907**	**.917**	**.968**	**.027**	**.824**	**.851**	**.914**	**.042**	**.861**	**.861**	**.913**

full performance of our method at higher resolution, denoted as **Ours+**, can be found in Tables 2 & 1. We also provide in Table 5 the ablation study on the losses at higher resolution. Our losses perform similarly across resolutions, validating their effectiveness. The semantic loss $\mathcal{L}_{sem}$ is not ablated because it is (a) always essential and (b) not the contribution of our paper.

8. Sensitivity Analysis on Hyperparameters

We conduct experiments on the hyperparameters finetuning. In our PopNet, we have 1 hyperparameter σ controlling the slope of the sigmoid function for separation. In addition, we have another 5 hyperparameters controlling the proportion of each loss. The sensitivity analysis of different hyperparameters can be found in Tables 6 & 7. Our method performs well with a large variation of hyperparameters.

Table 4. End-to-end comparison with different resolutions on SOD and COD benchmarks. Our method with source-free depth generalizes significantly better compared to SOTA COD models.

SOD Benchmarks										
Model	Size	Flops (G)	NJUK [19]				SIP [9]			
			$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$
SegMAR [18]	352^2	67.3	.036	.921	.909	.941	.052	.893	.872	.914
ZoomNet [32]	352^2	167.8	.037	.926	.914	.940	.054	.891	.868	.909
Ours	352^2	228.8	**.031**	**.931**	**.920**	**.949**	**.046**	**.911**	**.885**	**.926**
SegMAR [18]	512^2	142.4	.035	.927	.914	.943	.050	.899	.878	.917
ZoomNet [32]	512^2	353.4	.036	.926	.915	.942	.052	.895	.873	.910
Ours+	512^2	484.0	**.031**	**.933**	**.922**	**.951**	**.044**	**.911**	**.890**	**.927**
COD Benchmarks										
Model	Size	Flops (G)	CHAMELEON [37]				COD10K [7]			
			$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$M\downarrow$	$F_m\uparrow$	$S_m\uparrow$	$E_m\uparrow$
SegMAR [18]	352^2	67.3	.032	.871	.887	.935	.039	.750	.799	.876
ZoomNet [32]	352^2	167.8	.033	.829	.859	.915	.034	.771	.808	.872
Ours	352^2	228.8	**.022**	**.893**	**.910**	**.962**	**.031**	**.789**	**.827**	**.897**
SegMAR [18]	512^2	142.4	.035	.861	.875	.923	.042	.741	.792	.856
ZoomNet [32]	512^2	353.4	.025	.888	.895	.942	.029	.816	.838	.895
Ours+	512^2	484.0	**.020**	**.906**	**.917**	**.969**	**.027**	**.824**	**.851**	**.914**

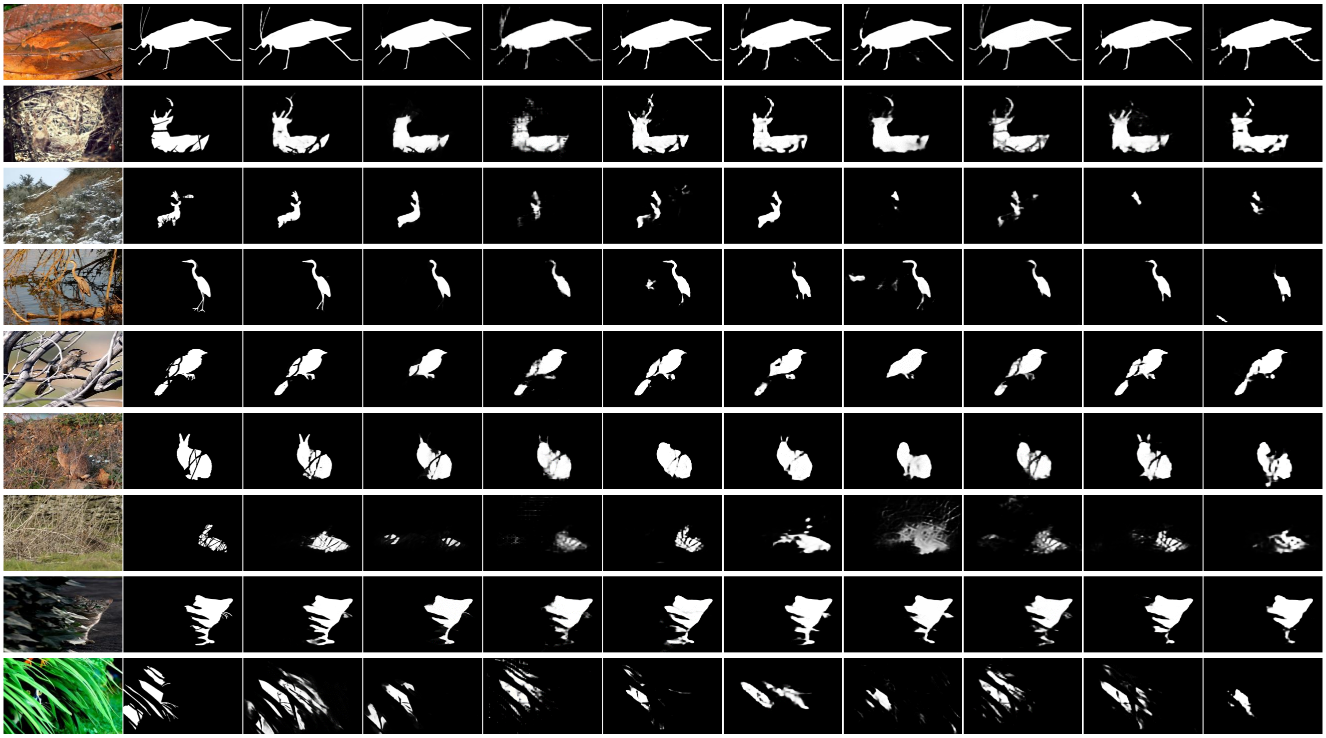

Figure 3. **Qualitative comparison**. Our method can better preserve the object structure. Please, zoom in for more details.

Table 5. Ablation study on the proposed losses on 512×512.

$\mathcal{L}_{dep}$	$\mathcal{L}_{loc}$	$\mathcal{L}_{wtv}$	$\mathcal{L}_{sep}$	Size	COD10K [7] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$	NC4K [28] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$
-	-	-	-	512^2	.034	.787	.821	.895	.051	.839	.840	.892
✓	-	-	-	512^2	.031	.801	.832	.901	.046	.847	.848	.901
-	✓	-	-	512^2	.031	.807	.837	.904	.046	.849	.841	.897
-	-	✓	-	512^2	.032	.798	.832	.898	.045	.852	.850	.905
-	-	-	✓	512^2	.030	.811	.842	.906	.045	.851	.852	.907
✓	✓	-	-	512^2	.031	.810	.841	.905	.046	.851	.849	.898
✓	-	✓	-	512^2	.029	.821	.841	.904	.045	.854	.851	.902
✓	-	✓	✓	512^2	.028	.821	.846	.910	.044	.857	.856	.908
✓	✓	-	✓	512^2	.029	.818	.845	.909	.044	.855	.851	.904
✓	✓	✓	✓	512^2	**.027**	**.824**	**.851**	**.914**	**.042**	**.861**	**.861**	**.913**

Table 6. Sensitivity analysis on hyperparameters. The ablation studies are realized at higher resolution, *i.e.*, 512×512.

$\mathcal{L}_{dep}$	$\mathcal{L}_{loc}$	$\mathcal{L}_{wtv}$	$\mathcal{L}_{sep}$	$\mathcal{L}_{sem}$	COD10K [7] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$	NC4K [28] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$
1	1	1	1	0.5	.030	.805	.834	.901	.046	.839	.844	.885
1	0.5	1	1	1	.031	.802	.831	.899	.046	.831	.845	.888
1	1	0.5	1	1	.030	.808	.836	.903	.046	.838	.841	.887
0.5	1	1	1	1	.030	.802	.837	.901	.044	.847	.851	.903
0.5	0.5	0.5	1	1	.031	.799	.833	.898	.045	.850	.854	.905
1	1	1	0.5	1	.029	.818	.844	.907	.044	.853	.855	.904
0.5	0.5	0.5	0.5	1	.030	.815	.844	.909	.045	.857	.856	.907
0.1	0.1	0.1	0.1	1	.029	.822	.849	.911	.043	**.862**	.859	.911
1	1	0.1	0.1	1	**.027**	**.824**	**.851**	**.914**	**.042**	.861	**.861**	**.913**

Table 7. Sensitivity analysis on the $\mathcal{L}_{sep}$. The ablation studies are realized at higher resolution, *i.e.*, 512×512.

σ	CHAMELEON [37] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$	COD10K [7] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$	NC4K [28] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$
0.1	.035	.836	.853	.903	.034	.742	.797	.851	.048	.815	.827	.869
1	.031	.858	.875	.931	.035	.741	.799	.857	.046	.828	.836	.884
5	.026	.898	.904	.946	.030	.808	.838	.902	.046	.853	.850	.900
10	**.020**	**.906**	**.917**	**.969**	**.027**	**.824**	**.851**	**.914**	**.042**	**.861**	**.861**	**.913**
20	.026	.887	.900	.949	.034	.775	.849	.892	.047	.845	.846	.901
50	.034	.845	.862	.912	.035	.736	.794	.848	.048	.819	.828	.869

Table 8. Comparison with other models with the same Res2Net-50 backbone. Our network performs favorably with a large margin over the counterparts.

Model	Public.	CHAMELEON [37] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$	COD10K [7] $M \downarrow$	$F_m \uparrow$	$S_m \uparrow$	$E_m \uparrow$
SINetv2	*TPAMI*$_{21}$ [6]	.029	.873	.888	.944	.036	.769	.815	.888
C2FNet	*IJCAI*$_{21}$ [38]	.032	.871	.888	.936	.036	.764	.813	.894
SPNet	*ICCV*$_{21}$ [57]	.033	.872	.888	.930	.037	.776	.808	.869
PreyNet	*MM*$_{22}$ [51]	.028	.880	.895	.955	.034	.775	.813	.884
PopNet	**Ours**	.022	.893	.910	.962	.031	.789	.827	.897
PopNet	**Ours+**	**.020**	**.906**	**.917**	**.969**	**.027**	**.824**	**.851**	**.914**

9. Evaluation with the Same Backbone

Our RGB-D baseline is built upon [57] with Res2Net-50 backbone (**R2_50**). Under the consideration of a fair comparison, we provide in Table 8 the quantitative comparison of all the methods using the same Res2Net-50 backbone (**R2_50**). It can be seen that our method performs favorably against all the counterparts.

References

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *IEEE CVPR*, 2009. 1

[2] Qian Chen, Ze Liu, Yi Zhang, Keren Fu, Qijun Zhao, and Hongwei Du. Rgb-d salient object detection via 3d convolutional neural networks. In *AAAI*, 2021. 2

[3] Runmin Cong, Qinwei Lin, Chen Zhang, Chongyi Li, Xiaochun Cao, Qingming Huang, and Yao Zhao. Cir-net: Cross-modality interaction and refinement for rgb-d salient object detection. *IEEE TIP*, 2022. 2

[4] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *IEEE ICCV*, 2017. 1

[5] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment measure for binary foreground map evaluation. In *IJCAI*, 2018. 1

[6] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE TPAMI*, 44(10):6024–6042, 2022. 4

[7] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *IEEE CVPR*, 2020. 3, 4

[8] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *MICCAI*, 2020. 3

[9] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. Rethinking RGB-D salient object detection: Models, datasets, and large-scale benchmarks. *IEEE TNNLS*, 32(5):2075–2089, 2021. 2, 3

[10] Deng-Ping Fan, Yingjie Zhai, Ali Borji, Jufeng Yang, and Ling Shao. BBS-Net: RGB-D salient object detection with a bifurcated backbone strategy network. In *ECCV*, 2020. 2, 3

[11] Keren Fu, Deng-Ping Fan, Ge-Peng Ji, and Qijun Zhao. JL-DCF: Joint learning and densely-cooperative fusion framework for RGB-D salient object detection. In *IEEE CVPR*, 2020. 2

[12] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE TPAMI*, 43(2):652–662, 2021. 2, 3

[13] Shang-Hua Gao, Yong-Qiang Tan, Ming-Ming Cheng, Chengze Lu, Yunpeng Chen, and Shuicheng Yan. Highly efficient salient object detection with 100k parameters. In *ECCV*, 2020. 3

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE CVPR*, 2016. 2, 3

[15] Nianchang Huang, Yang Yang, Dingwen Zhang, Qiang Zhang, and Jungong Han. Employing bilinear fusion and saliency prior information for rgb-d salient object detection. *TMM*, 24:1651–1664, 2021. 2

[16] Wei Ji, Jingjing Li, Shuang Yu, Miao Zhang, Yongri Piao, Shunyu Yao, Qi Bi, Kai Ma, Yefeng Zheng, Huchuan Lu, et al. Calibrated RGB-D salient object detection. In *IEEE CVPR*, 2021. 2, 3

[17] Wei Ji, Jingjing Li, Miao Zhang, Yongri Piao, and Huchuan Lu. Accurate RGB-D salient object detection via collaborative learning. In *ECCV*, 2020. 2

[18] Qi Jia, Shuilian Yao, Yu Liu, Xin Fan, Risheng Liu, and Zhongxuan Luo. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *IEEE CVPR*, 2022. 3

[19] Ran Ju, Ling Ge, Wenjing Geng, Tongwei Ren, and Gangshan Wu. Depth saliency based on anisotropic center-surround difference. In *IEEE ICIP*, 2014. 2, 3

[20] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *CVIU*, 184:45–56, 2019. 3

[21] Minhyeok Lee, Chaewon Park, Suhwan Cho, and Sangyoun Lee. Spsn: Superpixel prototype sampling network for rgb-d salient object detection. In *ECCV*, 2022. 2, 3

[22] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *IEEE CVPR*, 2021. 3

[23] Chongyi Li, Runmin Cong, Yongri Piao, Qianqian Xu, and Chen Change Loy. RGB-D salient object detection with cross-modality modulation and selection. In *ECCV*, 2020. 2

[24] Gongyang Li, Zhi Liu, Minyu Chen, Zhen Bai, Weisi Lin, and Haibin Ling. Hierarchical alternate interaction network for rgb-d salient object detection. *IEEE TIP*, 30:3528–3542, 2021. 2, 3

[25] Gongyang Li, Zhi Liu, Linwei Ye, Yang Wang, and Haibin Ling. Cross-modal weighting network for RGB-D salient object detection. In *ECCV*, 2020. 2

[26] Nian Liu, Ni Zhang, Kaiyuan Wan, Ling Shao, and Junwei Han. Visual saliency transformer. In *IEEE ICCV*, 2021. 2

[27] Zhengyi Liu, Wang Yuan, Zhengzheng Tu, Yun Xiao, and Bin Tang. TriTransNet: RGB-D salient object detection with a triplet transformer embedding network. *ACM MM*, 2021. 2

[28] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *IEEE CVPR*, 2021. 3, 4

[29] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *IEEE CVPR*, 2021. 3

[30] Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu. Leveraging stereopsis for saliency analysis. In *IEEE CVPR*, 2012. 2

[31] Youwei Pang, Lihe Zhang, Xiaoqi Zhao, and Huchuan Lu. Hierarchical dynamic filtering network for RGB-D salient object detection. In *ECCV*, 2020. 2

[32] 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 *IEEE CVPR*, 2022. 3

[33] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *IEEE CVPR*, 2020. 3

[34] Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji. RGBD salient object detection: a benchmark and algorithms. In *ECCV*, 2014. 2

[35] Yongri Piao, Wei Ji, Jingjing Li, Miao Zhang, and Huchuan Lu. Depth-induced multi-scale recurrent attention network for saliency detection. In *IEEE ICCV*, 2019. 2

[36] Yongri Piao, Zhengkun Rong, Miao Zhang, Weisong Ren, and Huchuan Lu. A2dele: Adaptive and attentive depth distiller for efficient RGB-D salient object detection. In *IEEE CVPR*, 2020. 2

[37] Przemysław Skurowski, Hassan Abdulameer, J Błaszczyk, Tomasz Depta, Adam Kornacki, and P Kozieł. Animal camouflage analysis: Chameleon database. *Unpublished manuscript*, 2(6):7, 2018. 3, 4

[38] Yujia Sun, Geng Chen, Tao Zhou, Yi Zhang, and Nian Liu. Context-aware cross-level fusion network for camouflaged object detection. In *IJCAI*, 2021. 3, 4

[39] Fengyun Wang, Jinshan Pan, Shoukun Xu, and Jinhui Tang. Learning discriminative cross-modality features for RGB-D saliency detection. *IEEE TIP*, 31:1285–1297, 2022. 3

[40] Xuehao Wang, Shuai Li, Chenglizhao Chen, Yuming Fang, Aimin Hao, and Hong Qin. Data-level recombination and lightweight fusion scheme for RGB-D salient object detection. *IEEE TIP*, 30:458–471, 2020. 2

[41] Jun Wei, Shuhui Wang, and Qingming Huang. F^3net: Fusion, feedback and focus for salient object detection. In *AAAI*, 2020. 3

[42] Hongfa Wen, Chenggang Yan, Xiaofei Zhou, Runmin Cong, Yaoqi Sun, Bolun Zheng, Jiyong Zhang, Yongjun Bao, and Guiguang Ding. Dynamic selective network for rgb-d salient object detection. *TIP*, 30:9179–9192, 2021. 2

[43] Yu-Huan Wu, Yun Liu, Jun Xu, Jia-Wang Bian, Yu-Chao Gu, and Ming-Ming Cheng. Mobilesal: Extremely efficient rgb-d salient object detection. *IEEE TPAMI*, 44(12):10261–10269, 2022. 2

[44] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE CVPR*, 2019. 3

[45] Zhe Wu, Li Su, and Qingming Huang. Stacked cross refinement network for edge-aware salient object detection. In *IEEE ICCV*, 2019. 3

[46] Fan Yang, Qiang Zhai, Xin Li, Rui Huang, Ao Luo, Hong Cheng, and Deng-Ping Fan. Uncertainty-guided transformer reasoning for camouflaged object detection. In *IEEE ICCV*, 2021. 3

[47] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *IEEE CVPR*, 2021. 3

[48] Chen Zhang, Runmin Cong, Qinwei Lin, Lin Ma, Feng Li, Yao Zhao, and Sam Kwong. Cross-modality discrepant interaction network for RGB-D salient object detection. In *ACM MM*, 2021. 3

[49] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Sadat Saleh, Tong Zhang, and Nick Barnes. Uc-net: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In *IEEE CVPR*, 2020. 3

[50] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Xin Yu, Yiran Zhong, Nick Barnes, and Ling Shao. RGB-D saliency detection via cascaded mutual information minimization. In *IEEE ICCV*, 2021. 3

[51] Miao Zhang, Shuang Xu, Yongri Piao, Dongxiang Shi, Shusen Lin, and Huchuan Lu. Preynet: Preying on camouflaged objects. In *ACM MM*, 2022. 4

[52] Zhao Zhang, Zheng Lin, Jun Xu, Wen-Da Jin, Shao-Ping Lu, and Deng-Ping Fan. Bilateral attention network for rgb-d salient object detection. *IEEE TIP*, 30:1949–1961, 2021. 2

[53] Jia-Xing Zhao, Yang Cao, Deng-Ping Fan, Ming-Ming Cheng, Xuan-Yi Li, and Le Zhang. Contrast prior and fluid pyramid integration for rgbd salient object detection. In *IEEE CVPR*, 2019. 2

[54] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. EGNet: Edge Guidance Network for salient object detection. In *IEEE ICCV*, 2019. 3

[55] Xiaoqi Zhao, Lihe Zhang, Youwei Pang, Huchuan Lu, and Lei Zhang. A single stream network for robust and real-time RGB-D salient object detection. In *ECCV*, 2020. 2

[56] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. Interactive two-stream decoder for accurate and fast saliency detection. In *IEEE CVPR*, 2020. 3

[57] Tao Zhou, Huazhu Fu, Geng Chen, Yi Zhou, Deng-Ping Fan, and Ling Shao. Specificity-preserving RGB-D saliency detection. In *IEEE ICCV*, 2021. 2, 3, 4

[58] Wujie Zhou, Yun Zhu, Jingsheng Lei, Jian Wan, and Lu Yu. Ccafnet: Crossflow and cross-scale adaptive fusion network for detecting salient objects in rgb-d images. *TMM*, 24:2192–2204, 2021. 2