Supplemental Material for 'SAM-COD: SAM-guided Unified Framework for Weakly-Supervised Camouflaged Object Detection'

Huafeng Chen[1,3], Pengxu Wei[2,4], Guangqian Guo[1,3], and Shan Gao[1,3⋆]

[1] Northwestern Polytechnical University, [2] Sun Yat-Sen University, [3] National Key Laboratory of Unmanned Aerial Vehicle Technology, [4] Peng Cheng Laboratory, China
{{chf, guogq21}@mail, gaoshan@}.nwpu.edu.cn, weipx3@mail.sysu.edu.cn

Table 1: Quantitative comparison with state-of-the-arts on three popular polyp image segmentation benchmarks. Red and blue represent the first and second best performance, respectively.

Methods	Label	CVC-ColonDB				ETIS				Kvasir			
		MAE ↓	S_m ↑	F_β^w ↑	E_m ↑	MAE ↓	S_m ↑	F_β^w ↑	E_m ↑	MAE ↓	S_m ↑	F_β^w ↑	E_m ↑
U-Net [27]	F	0.061	0.711	0.498	-	0.036	0.682	0.366	-	0.055	0.858	0.794	-
U-Net++ [41]	F	0.064	0.691	0.467	-	0.035	0.681	0.390	-	0.048	0.862	0.808	-
SFA [8]	F	0.094	0.634	0.379	-	0.109	0.557	0.231	-	0.075	0.782	0.670	-
PraNet [7]	F	0.043	0.820	0.699	-	0.031	0.791	0.600	-	0.030	0.815	0.885	-
MSNet [40]	F	0.041	0.836	0.737	-	0.020	0.840	0.678	-	0.028	0.822	0.893	-
SAM [15]	-	0.479	0.427	0.343	0.419	0.429	0.503	0.439	0.512	0.320	0.582	0.545	0.564
SAM-P [15]	P	0.194	0.671	0.587	0.664	0.144	0.715	0.625	0.719	0.108	0.802	0.793	0.811
WSSA [39]	P	0.127	0.713	0.645	0.732	0.123	0.762	0.647	0.733	0.082	0.828	0.822	0.852
SCWS [37]	P	0.082	0.787	0.674	0.758	0.085	0.731	0.646	0.768	0.078	0.831	0.837	0.860
TEL [19]	P	0.089	0.761	0.669	0.743	0.083	0.726	0.639	0.776	0.091	0.804	0.810	0.826
CRNet [13]	P	0.077	0.802	0.691	0.795	0.071	0.766	0.664	0.802	0.071	0.836	0.853	0.877
WSSAM [12]	P	0.043	0.816	-	0.839	0.037	0.797	-	0.849	0.046	0.877	-	0.917
SAM-COD	P	**0.036**	**0.839**	**0.737**	**0.848**	**0.020**	**0.841**	**0.651**	**0.849**	**0.029**	**0.861**	**0.902**	**0.930**
SAM-COD	B	**0.031**	**0.844**	**0.748**	**0.889**	**0.019**	**0.854**	**0.679**	**0.854**	**0.025**	**0.926**	**0.908**	**0.942**

1 Framework Details

1.1 Detailed Structure of the Encoder and Decoder.

As shown in Fig. 1, we use PVT [32] as the encoder, for an input image $I \in \mathbb{R}^{3\times H\times W}$, we put it into the encoder to get the output features $Feat_i$ for the i-th. Then, we get the multi-scale features $(F_{eat1}, F_{eat2}, F_{eat3}, F_{eat4})$ with $(\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32})$ resolution of input images. We downsize the channel dimension of F_{eati} into 64 by using 3×3 convolutional layers. Next, these feature maps are unified

⋆ Corresponding author

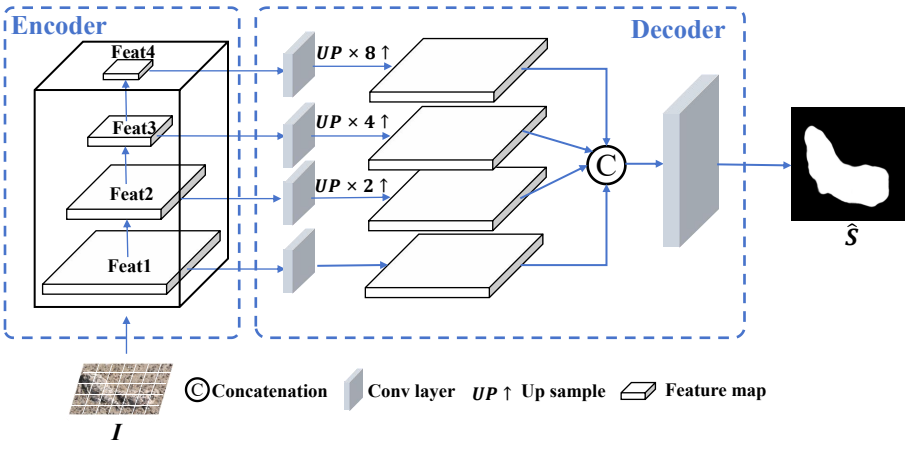

Fig. 1: The architecture of encoder and decoder.

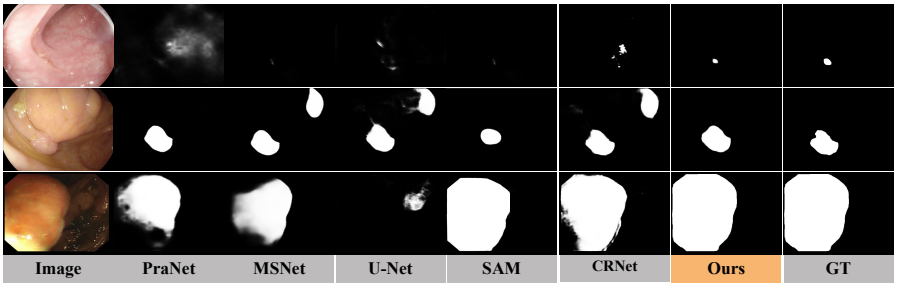

Fig. 2: Visualized results for polyp image segmentation.

into the same size by an up-sampling operation, and combined through the concatenation. Finally the output map $\hat{S} \in \mathbb{R}^{1\times W\times H}$ is obtained by the 3×3 convolution layer.

1.2 Details about Step 1 of Training.

The original labels (where boxes are transformed into pixel-level supervision through SAM) are directly used as the supervision information, and Partial Cross-Entropy is employed as the loss function for training, following the settings outlined in Section 4.1.

2 Comparison with using ResNet50 Backbone.

We provide additional comparisons using ResNet50 as the backbone, as shown in the Tab. 2. Our model outperforms significantly CRNet and WS-SAM when using the ResNet50 backbone even with weaker supervision (point).

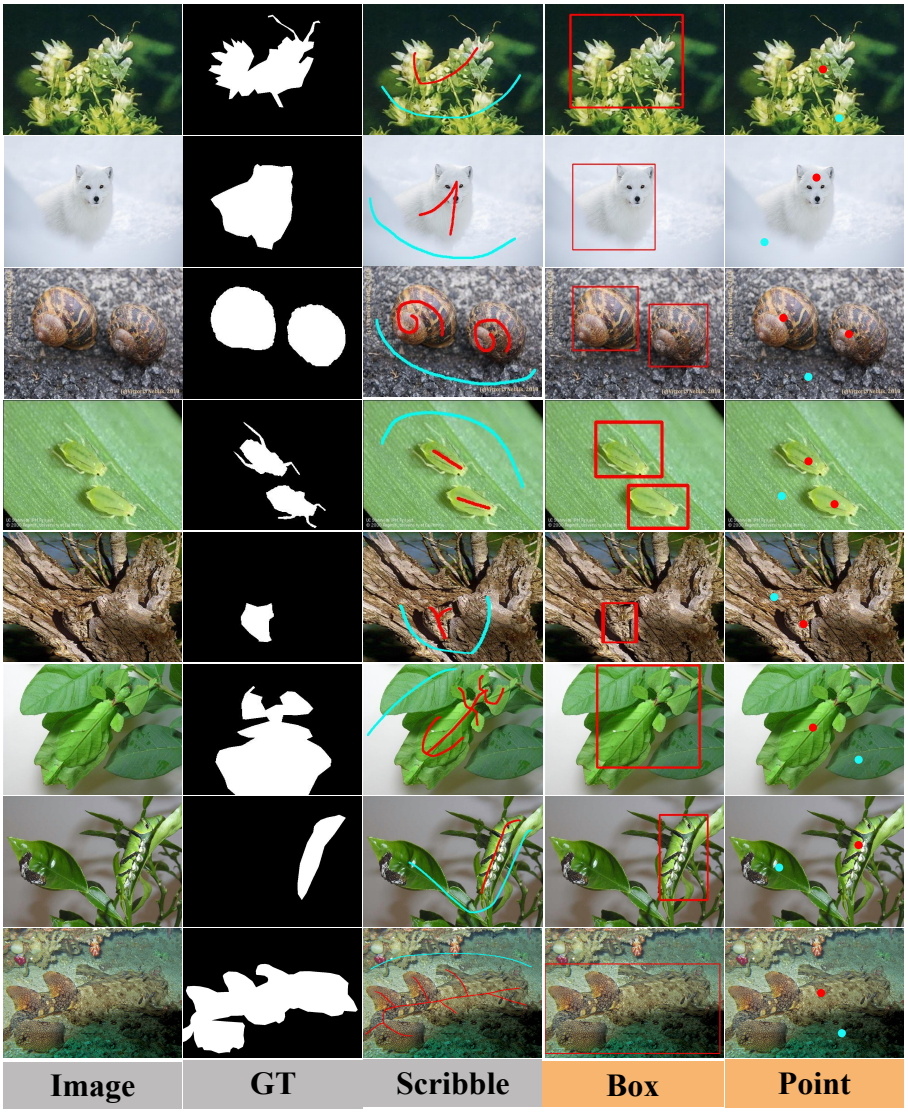

Fig. 3: Examples of BCOD and PCOD datasets. It includes many categories of animals in challenging scenarios.

Table 2: Quantitative comparison with state-of-the-arts on three datasets using ResNet50 as a backbone. "S", "P", "B" denote scribble, point, and box labels, respectively. "–" is not available. Red and blue represent the first and second best performance, respectively.

Methods	Label	CAMO (250)				COD10K (2026)				NC4K (4121)			
		MAE	S_m	E_m	F_β^w	MAE	S_m	E_m	F_β^w	MAE	S_m	E_m	F_β^w
CRNet [11]	S	.092	.735	.815	.641	.049	.733	.832	.576	.063	.775	.855	.688
SAM-S [15]	S	.105	.731	.774	-	.046	.772	.828	-	.071	.763	.832	-
WS-SAM [12]	S	.092	.759	.818	-	.038	.803	.878	-	.052	.829	.886	-
SAM-P [15]	P	.123	.677	.693	-	.069	.765	.796	-	.082	.776	.786	-
WS-SAM [12]	P	.102	.718	.757	-	.039	.790	.856	-	.057	.813	.859	-
SAM-COD	P	.080	.768	.832	.687	.034	.797	.867	.685	.054	.824	.883	.753
SAM-COD	B	**.073**	**.804**	**.879**	**.733**	**.032**	**.801**	**.883**	**.698**	**.048**	**.837**	**.897**	**.769**
SAM-COD	S	**.070**	**.818**	**.884**	**.751**	**.032**	**.803**	**.885**	**.699**	**.049**	**.836**	**.896**	**.766**

3 Experiment on Polyp Segmentation.

Common concealed scenarios include camouflaged object detection, medical image segmentation (polyp segmentation), and transparent object detection. Therefore, we attempt to apply the SAM-COD model to polyp segmentation and transparent object detection.

3.1 Datasets.

We evaluate the proposed model on three benchmark datasets: CVC-ColonDB [29], ETIS [28], Kvasir [14]. We adopt the same training set as the previous polyp segmentation method [7], that is, 900 samples from the Kvasir and 550 samples from the CVC-ClinicDB are used for training. The remaining images and the other three datasets are used for testing.

3.2 Evaluation Metrics.

We adopt four evaluation metrics: Mean Absolute Error (MAE), S-measure (S_m) [4], E-measure (E_m) [5], and weighted F-measure (F_β^w) [22].

3.3 Implementation Details.

We implement our method with PyTorch and conduct experiments on one GeForce RTX4090 GPU and use ViT-H version of SAM. We chose PVT-B4 as the encoder. We use the stochastic gradient descent optimizer with a momentum of 0.9, a weight decay of 5e-4, and triangle learning rate schedule with maximum learning rate 1e-3. The batch size is 8, and the training epoch is 60. We adopt

Table 3: Quantitative comparison with state-of-the-arts on CAMO dataset. "F", "U", "S", "P", "Mix" denote fully-supervised label, unsupervised, scribble, point, and mixed random selection of one of three weakly-supervised labels, respectively. "–" is not available. Red and blue represent the first and second best performance, respectively.

Methods	Label	CAMO (250)								
		MAE ↓	S_m ↑	E^m ↑	F_β^w ↑	E_m^a ↑	E_m^x ↑	F_β^a ↑	F_β^m ↑	F_β^x ↑
SINet [CVPR20] [6]	F	.092	.745	.804	.644	.825	.829	.712	.702	.708
MGL-R [CVPR21] [38]	F	.088	.775	.812	.673	.848	.842	.738	.726	.740
PFNet [CVPR21] [23]	F	.085	.782	.841	.695	.855	.855	.751	.746	.758
UGTR [ICCV21] [36]	F	.086	.784	.822	.684	.861	.854	.749	.738	.754
UJSC [CVPR21] [17]	F	.073	.800	.859	.728	.865	.873	.779	.772	.779
ZoomNet [CVPR22] [24]	F	**.066**	**.820**	**.892**	.752	**.883**	**.892**	**.792**	**.794**	**.805**
FEDER [CVPR23] [11]	F	**.069**	.807	.873	**.785**	**.877**	**.873**	**.786**	**.781**	**.789**
SAM-Adapter [ICCV23] [2]	F	.070	**.847**	**.873**	**.765**	-	-	-	-	-
CRNet [AAAI23] [13]	S	.092	.735	.815	.641	.829	.830	.709	.701	.707
SAM [ICCV23] [15]	-	.132	.684	.687	.606	.742	.742	.705	.705	.710
SAM-S [ICCV23] [15]	S	.105	.731	.774	-	-	-	.709	-	-
WS-SAM [NIPS23] [12]	S	.092	.759	.818	-	-	-	.742	-	-
SAM-P [ICCV23] [15]	P	.123	.677	.693	-	-	-	.649	-	-
WS-SAM [NIPS23] [12]	P	.102	.718	.757	-	-	-	.703	-	-
SAM-COD	S	**.060**	.836	**.903**	.779	.897	.897	**.801**	**.804**	.821
SAM-COD	B	.062	**.837**	.901	**.786**	**.901**	**.907**	**.805**	**.809**	**.834**
SAM-COD	P	.066	.820	.885	.760	.888	.889	.784	.787	.804
SAM-COD	Mix	**.058**	**.839**	**.907**	**.784**	**.905**	**.920**	.798	.803	**.840**

the offline distillation, where SAM is pre-computed, and forward computation is performed only once. So, it only takes around 3h in training. During training and inference, input images are resized to 512×512.

3.4 Comparisons with state-of-the-art

We compare our SAM-COD with U-Net [27], U-Net++ [41], SFA [8], PraNet [7], MSNet [40], WSSA [39], SCWS [37], TEL [19], CRNet [13], SAM [15], WS-SAM [12] and SAM-P [15] which fine-tune the mask decoder of SAM with point supervision. To be fair, the predictions of these competitors are directly provided by their respective authors or computed by their released codes.

Quantitative Evaluation. As shown in Tab. 1, our method significantly outperforms the second-ranked weakly supervised CRNet method and similarly surpasses many fully supervised methods. SAM and SAM-P do not perform well on this task, further substantiating their weakness in this challenging segmentation task. This again verifies our benefit in handling challenging Concealed Object Segmentation tasks.

Table 4: Quantitative comparison with state-of-the-arts on COD10K dataset. "F", "U", "S", "P", "Mix" denote fully-supervised label, unsupervised, scribble, point, and mixed random selection of one of three weakly-supervised labels, respectively. "–" is not available. Red and blue represent the first and second best performance, respectively.

Methods	Label	COD10K (2026)								
		MAE ↓	S_m ↑	E^m ↑	F_β^w ↑	E_m^a ↑	E_m^x ↑	F_β^a ↑	F_β^m ↑	F_β^x ↑
SINet [CVPR20] [6]	F	.043	.776	.864	.631	.867	.874	.667	.679	.691
MGL-R [CVPR21] [38]	F	.035	.814	.851	.666	.865	.890	.681	.711	.738
PFNet [CVPR21] [23]	F	.040	.800	.877	.660	.868	.890	.676	.701	.725
UGTR [ICCV21] [36]	F	.036	.817	.852	.666	.850	.891	.671	.712	.742
UJSC [CVPR21] [17]	F	.035	.809	.884	.684	.882	.891	.705	.721	.738
ZoomNet [CVPR22] [24]	F	**.029**	**.838**	**.911**	.729	**.893**	**.911**	**.741**	**.766**	**.780**
FEDER [CVPR23] [11]	F	.032	.823	.900	**.740**	**.901**	**.905**	**.740**	**.751**	**.768**
SAM-Adapter [ICCV23] [2]	F	**.025**	**.883**	**.918**	**.801**	-	-	-	-	-
CRNet [AAAI23] [13]	S	.049	.733	.832	.576	.845	.845	.637	.633	.636
SAM [ICCV23] [15]	-	.050	.783	.798	.701	.800	.800	.758	.756	.758
SAM-S [15]	S	.046	.772	.828	-	-	-	.695	-	-
WS-SAM [12]	S	.038	.803	.878	-	-	-	.719	-	-
SAM-P [15]	P	.069	.765	.796	-	-	-	.694	-	-
WS-SAM [12]	P	.039	.790	.856	-	-	-	.698	-	-
SAM-COD	S	**.029**	**.833**	**.904**	**.728**	**.900**	**.926**	**.739**	**.763**	**.803**
SAM-COD	B	**.028**	**.842**	**.914**	**.745**	**.901**	**.928**	**.740**	**.763**	**.803**
SAM-COD	P	.031	.831	.901	.725	.887	.920	.720	.743	.794
SAM-COD	Mix	.031	.833	.903	.725	.888	.921	.719	.744	.793

Table 5: Quantitative comparison with state-of-the-arts on NC4K dataset. "F", "U", "S", "P", "Mix" denote fully-supervised label, unsupervised, scribble, point, and mixed random selection of one of three weakly-supervised labels, respectively. "–" is not available. Red and blue represent the first and second best performance, respectively.

Methods	Label	NC4K (4121)								
		MAE ↓	S_m ↑	E^m ↑	F_β^w ↑	E_m^a ↑	E_m^x ↑	F_β^a ↑	F_β^m ↑	F_β^x ↑
SINet [CVPR20] [6]	F	.058	.808	.871	.723	.883	.883	.768	.769	.775
MGL-R [CVPR21] [38]	F	.052	.833	.867	.740	.890	.893	.778	.782	.800
PFNet [CVPR21] [23]	F	.053	.829	.887	.745	.894	.898	.779	.784	.799
UGTR [ICCV21] [36]	F	.052	.839	.874	.747	.889	.899	.779	.787	.807
UJSC [CVPR21] [17]	F	.047	.842	.898	.771	.903	.907	.803	.806	.816
ZoomNet [CVPR22] [24]	F	**.043**	**.853**	**.896**	**.784**	**.907**	**.912**	**.814**	**.818**	**.828**
FEDER [CVPR23] [11]	F	**.045**	**.846**	**.905**	**.817**	**.913**	**.915**	**.822**	**.824**	**.833**
CRNet [AAAI23] [13]	S	.063	.775	.855	.688	.885	.887	.682	.680	.682
SAM [ICCV23] [15]	-	.078	.767	.776	.696	.778	.778	.754	.752	.754
SAM-S [15]	S	.071	.763	.832	-	-	-	.747	-	-
WS-SAM [12]	S	.052	.829	.886	-	-	-	.802	-	-
SAM-P [15]	P	.082	.776	.786	-	-	-	.728	-	-
WS-SAM [12]	P	.057	.813	.859	-	-	-	.801	-	-
SAM-COD	S	**.039**	**.859**	.912	.795	.912	.917	.803	.813	.848
SAM-COD	B	**.037**	**.867**	**.923**	**.813**	**.920**	**.931**	**.819**	**.828**	**.855**
SAM-COD	P	.041	.858	**.918**	**.802**	**.915**	.925	.720	.743	.794
SAM-COD	Mix	.039	.862	.912	.798	.912	**.928**	**.803**	**.813**	**.848**

Table 6: Quantitative comparison with state-of-the-arts on four popular SOD benchmarks. "F", "S", "P" denote fully-, scribble-, and point-supervised methods, respectively. Red and blue represent the first and second best performance, respectively.

Methods	Label	ECSSD			DUT-O			HKU-IS			DUTS-TE		
		MAE ↓	S_m ↑	F_β^{max} ↑	MAE ↓	S_m ↑	F_β^{max} ↑	MAE ↓	S_m ↑	F_β^{max} ↑	MAE ↓	S_m ↑	F_β^{max} ↑
RAS [1]	F	0.056	0.893	0.921	0.062	0.814	0.786	0.045	0.887	0.913	0.059	0.839	0.831
R^3Net [3]	F	0.056	0.903	0.925	0.071	0.818	0.788	0.048	0.892	0.910	0.066	0.836	0.824
DGR [31]	F	0.041	0.903	0.922	0.062	0.806	0.774	0.036	0.892	0.910	0.050	0.842	0.828
PiNet [20]	F	0.046	0.917	0.935	0.065	0.832	0.803	0.043	0.904	0.919	0.051	0.869	0.860
MLMS [33]	F	0.045	0.911	0.928	0.064	0.809	0.774	0.039	0.907	0.921	0.049	0.862	0.852
AFNet [9]	F	0.042	0.913	0.935	0.057	0.826	0.797	0.036	0.905	0.923	0.046	0.867	0.863
BASNet [26]	F	0.037	0.916	0.943	0.057	0.836	0.805	0.032	0.909	0.928	0.048	0.866	0.859
MFNet [25]	S	0.084	0.834	0.879	0.087	0.741	0.706	0.059	0.846	0.876	0.076	0.774	0.770
SCSOD [37]	S	0.049	0.881	0.914	0.060	0.811	0.782	0.038	0.882	0.908	0.049	0.853	0.858
PSOD [10]	P	0.036	0.913	0.935	0.064	0.824	0.808	0.033	0.901	0.923	0.045	0.853	0.858
SAM-COD	P	**0.033**	**0.925**	**0.947**	0.051	0.844	0.826	0.024	0.946	0.941	0.034	0.892	0.898
SAM-COD	S	0.034	0.921	0.944	**0.050**	**0.846**	**0.829**	**0.024**	**0.947**	**0.944**	**0.033**	**0.898**	**0.901**
SAM-COD	B	**0.031**	**0.929**	**0.952**	**0.051**	**0.844**	**0.828**	**0.023**	**0.952**	**0.949**	**0.033**	**0.899**	**0.903**

Qualitative Evaluation. Fig. 2 illustrates visual comparison with other approaches. It can be seen that the proposed method has good detection performance for small, medium, and large scale polyps (see the 1st - 3rd rows).

4 Experiment Details

4.1 Experiments on SOD.

In order to show good generalization and further verify the rationality of the structural design, we evaluate the proposed model on the SOD task.
Datasets. Our experiment on SOD is based on the existing four SOD datasets, ECSSD [34], DUT-O [35], HKU-IS [18], and DUTS-test [30]. We only use the training set of DUTS for training. During the test phase, we use the remaining data for inference.
Implementation Details. We use the stochastic gradient descent optimizer with a momentum of 0.9, a weight decay of $5e-4$, and triangle learning rate schedule with maximum learning rate $1e-3$. The batch size is 8, and the training epoch is 60. During training and inference, input images are resized to 512×512.

4.2 Experiments on COD.

PCOD Dataset Our experiments are conducted on three COD benchmarks, CAMO [16], COD10K [6], and NC4K [21]. To evaluate the performance of our approach under point supervision, we relabel 4040 images (3040 from COD10K, 1000 from CAMO) and propose the Point-supervised Dataset for training and the remaining is for testing. Three annotators participate in the annotation task. To mitigate personal bias, we randomly choose one annotation from the three for each image. For every image, we annotate one foreground point for each camouflaged object and one background point for the background. More examples of our dataset are shown in Fig. 3 (We exaggerate the size of the labeled position in visualization).
BCOD Dataset. To evaluate the performance of our approach under bounding box supervision, we relabel 4040 images (3040 from COD10K, 1000 from CAMO) and propose the Box-supervised Dataset for training and the remaining is for testing. Three annotators participate in the annotation task, we select the best one of the three annotations as the final annotation of each image. For each image, annotate a bounding box for each camouflaged object. More examples of our dataset are shown in Fig. 3

5 More Results and Analysis.

5.1 Qualitative Comparison

Due to the space limitations of the manuscript, we add more visual comparisons to this supplementary material for further demonstration of the performance of

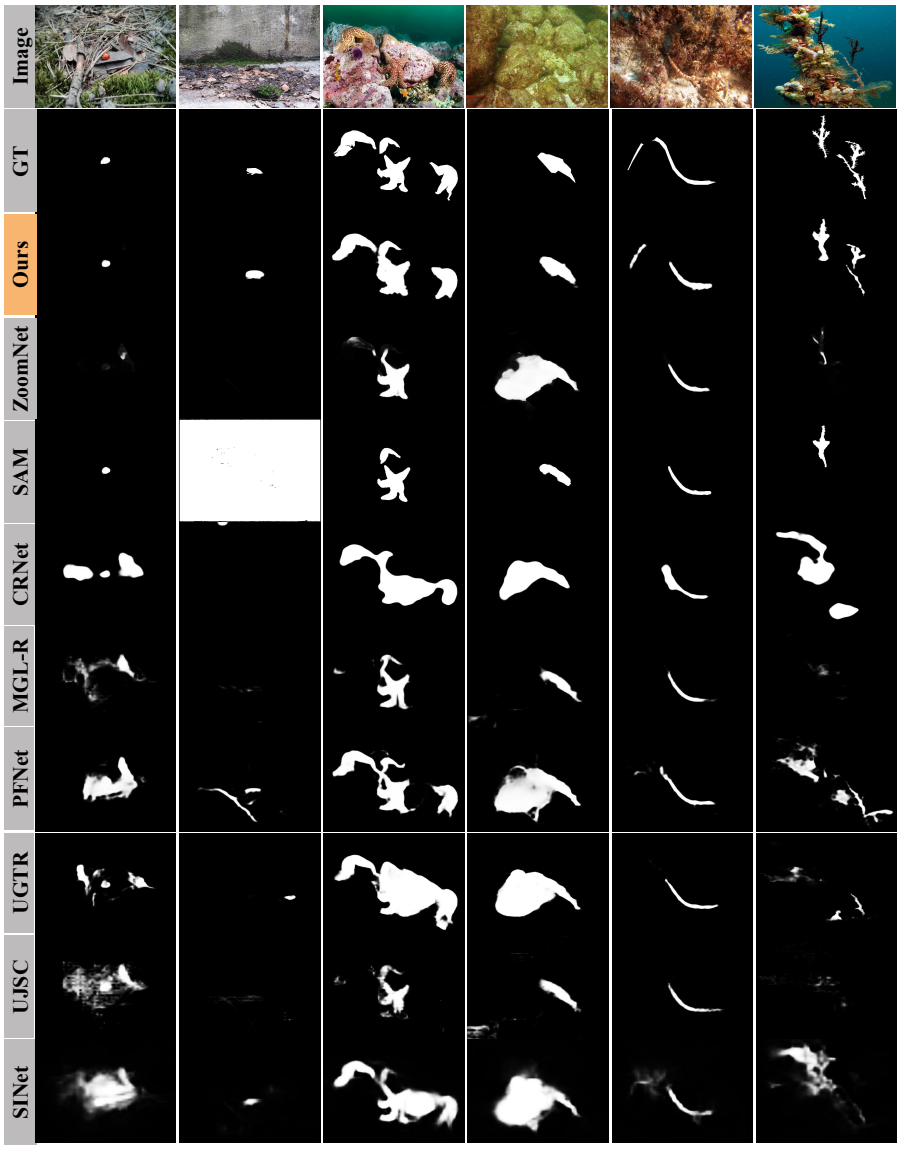

Fig. 4: Visual comparson with other competitors in detecting **small** camouflaged objects. Please zoom in for details.

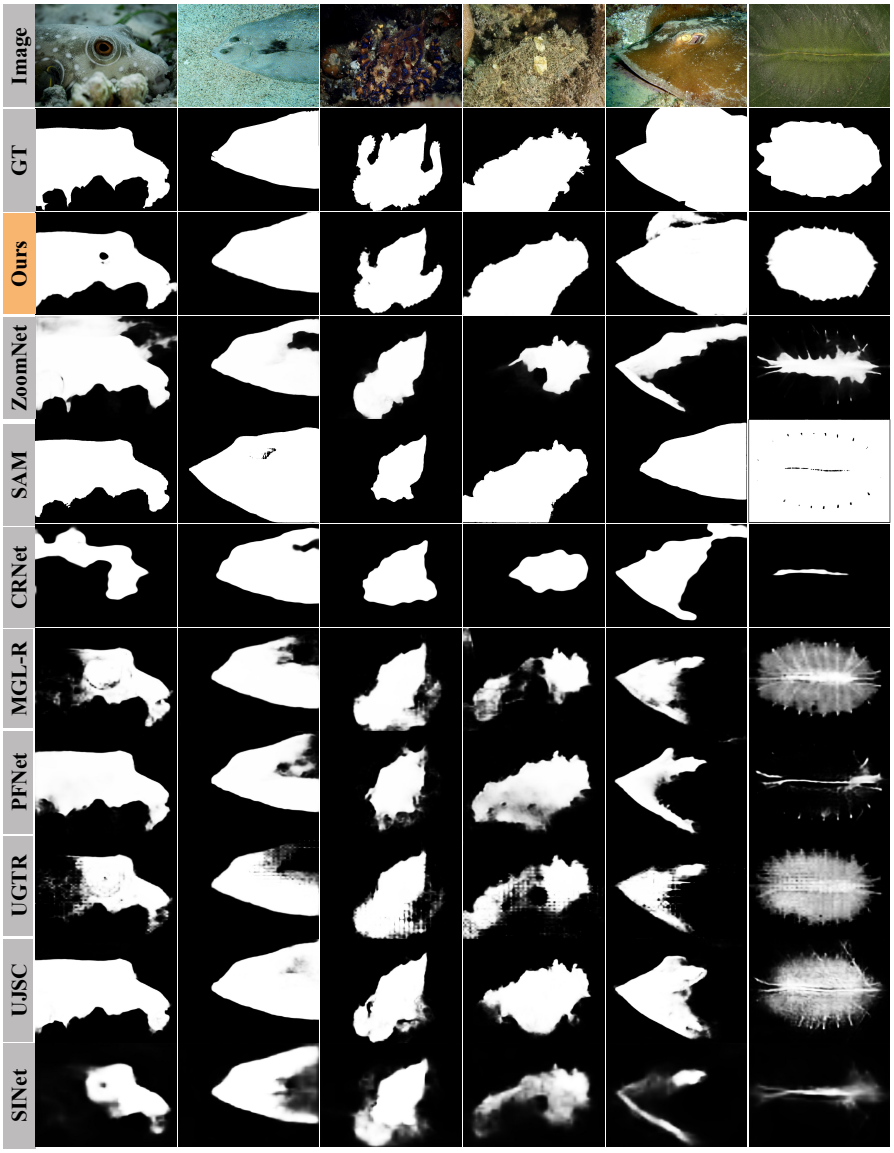

Fig. 5: Visual comparison with other competitors in detecting **big** camouflaged objects. Please zoom in for details.

Fig. 6: Visual comparison with other competitors in detecting **obscured** camouflaged objects. Please zoom in for details.

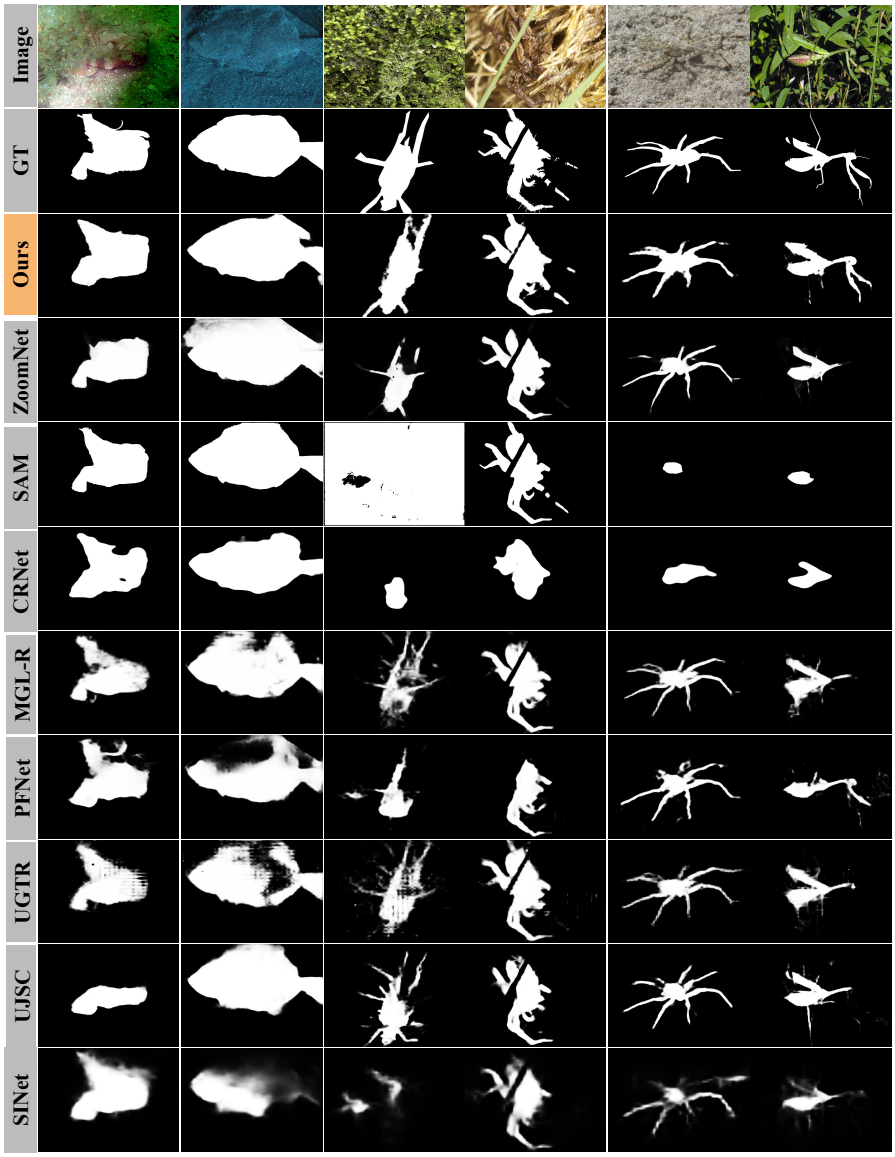

Fig. 7: Visual comparison with other competitors in detecting camouflaged objects with **indistinguishable boundaries** . Please zoom in for details.

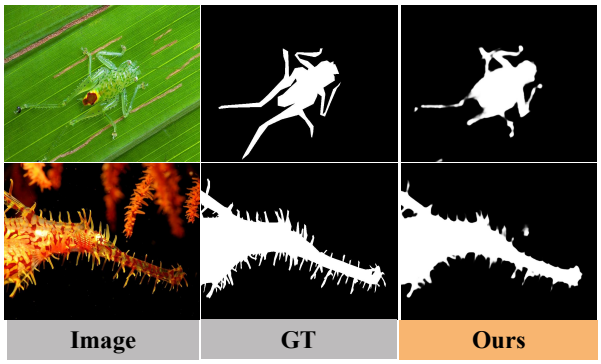

Fig. 8: Two failure cases of our method.

our model. Fig 4, 5, and 6 show examples containing small, large, and obscured objects, respectively. As can be seen from these visual comparisons, our model is more robust to a wide range of challenging scenarios, showing superior visual performance for more accurate and complete predictions.

5.2 Quantitative Comparison.

As shown in Tab. 3 and Tab. 5, we further list more comprehensive evaluation results on three COD datasets. It can be seen that our model achieves the best detection performance overall. It is worth noting that our proposed weakly-supervised model SAM-COD outperforms state-of-the-art fully supervised models ZoomNet [24] and FEDER [11]. We also find that our method has the largest improvement in CAMO (significantly surpassing the best fully-supervised COD methods), which is the most challenging one among all of the three COD datasets (worst metric value). This shows that our method is indeed better at discovering complex camouflage objects.

5.3 Results on SOD Datasets

We compare the proposed model with existing methods. All the results are listed in Tab. 6. Our model outperforms all the competitors. Especially, it significantly outperforms the state-of-the-art weakly supervised method PSOD [10] by a substantial margin. It shows that the proposed model can deal with the more general binary segmentation task.

6 Failure Cases

Our method is effective, but also has limitations. Because SAM sometimes fails to fully cover extremely detailed boundaries under particularly limited supervision conditions (especially point supervision). The trained model may perform

below expectations in scenes with particularly complex boundary details. Fig. 8 illustrates two failed cases where our method predicts approximate parts but fails to detect details at the edges, such as spikes and slender legs. In the future, we will pay more attention to local details, especially intricate boundaries, to segment more precisely in the weak supervision.

References

1. Chen, S., Tan, X., Wang, B., Hu, X.: Reverse attention for salient object detection. In: Proceedings of the European conference on computer vision (ECCV). pp. 234–250 (2018)
2. Chen, T., Zhu, L., Ding, C., Cao, R., Zhang, S., Wang, Y., Li, Z., Sun, L., Mao, P., Zang, Y.: Sam fails to segment anything?–sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, and more. arXiv preprint arXiv:2304.09148 (2023)
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: Proceedings of the 27th international joint conference on artificial intelligence. pp. 684–690. AAAI Press Menlo Park, CA, USA (2018)
4. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE international conference on computer vision. pp. 4548–4557 (2017)
5. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv:1805.10421 (2018)
6. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2777–2787 (2020)
7. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 263–273. Springer (2020)
8. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. pp. 302–310. Springer (2019)
9. Feng, M., Lu, H., Ding, E.: Attentive feedback network for boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1623–1632 (2019)
10. Gao, S., Zhang, W., Wang, Y., Guo, Q., Zhang, C., He, Y., Zhang, W.: Weakly-supervised salient object detection using point supervision. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 670–678 (2022)
11. He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22046–22055 (2023)
12. He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., Li, X.: Weakly-supervised concealed object segmentation with sam-based pseudo labeling and

multi-scale feature grouping. Advances in Neural Information Processing Systems **36** (2024)
13. He, R., Dong, Q., Lin, J., Lau, R.W.: Weakly-supervised camouflaged object detection with scribble annotations. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 781–789 (2023)
14. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26. pp. 451–462. Springer (2020)
15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
16. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A.: Anabranch network for camouflaged object segmentation. Computer vision and image understanding **184**, 45–56 (2019)
17. Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., Dai, Y.: Uncertainty-aware joint salient object and camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10071–10081 (2021)
18. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5455–5463 (2015)
19. Liang, Z., Wang, T., Zhang, X., Sun, J., Shen, J.: Tree energy loss: Towards sparsely annotated semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16907–16916 (2022)
20. Liu, N., Han, J., Yang, M.H.: Picanet: Learning pixel-wise contextual attention for saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3089–3098 (2018)
21. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.P.: Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11591–11601 (2021)
22. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 248–255 (2014)
23. Mei, H., Ji, G.P., Wei, Z., Yang, X., Wei, X., Fan, D.P.: Camouflaged object segmentation with distraction mining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8772–8781 (2021)
24. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: 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. pp. 2160–2170 (2022)
25. Piao, Y., Wang, J., Zhang, M., Lu, H.: Mfnet: Multi-filter directive network for weakly supervised salient object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4136–4145 (2021)
26. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7479–7489 (2019)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)

28. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International journal of computer assisted radiology and surgery **9**, 283–293 (2014)
29. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE transactions on medical imaging **35**(2), 630–644 (2015)
30. 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 the IEEE conference on computer vision and pattern recognition. pp. 136–145 (2017)
31. Wang, T., Zhang, L., Wang, S., Lu, H., Yang, G., Ruan, X., Borji, A.: Detect globally, refine locally: A novel approach to saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3127–3135 (2018)
32. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 568–578 (2021)
33. Wu, R., Feng, M., Guan, W., Wang, D., Lu, H., Ding, E.: A mutual learning method for salient object detection with intertwined multi-supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8150–8159 (2019)
34. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1155–1162 (2013)
35. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3166–3173 (2013)
36. Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., Fan, D.P.: Uncertainty-guided transformer reasoning for camouflaged object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4146–4155 (2021)
37. Yu, S., Zhang, B., Xiao, J., Lim, E.G.: Structure-consistent weakly supervised salient object detection with local saliency coherence. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3234–3242 (2021)
38. Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., Fan, D.P.: Mutual graph learning for camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12997–13007 (2021)
39. Zhang, J., Yu, X., Li, A., Song, P., Liu, B., Dai, Y.: Weakly-supervised salient object detection via scribble annotations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12546–12555 (2020)
40. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 120–130. Springer (2021)
41. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. pp. 3–11. Springer (2018)