

Supplementary Material for “Semi-Supervised Camouflaged Object Detection from Noisy Data”

Anonymous Author(s)

1 ABLATION STUDY

This section further validates the effectiveness of other proposed components that have not been experimentally validated in our manuscript, on four benchmarks, *i.e.*, CAMO [5], CHAMELEON [8], COD10K [1, 2], and NC4K [6]. The alternatives include: 1) **w/o condition**. Simply aggregating the parameters learned at every epochs, without assessing whether the losses decrease in comparison to the last epochs; 2) **w/o pixel-level uncertainties**. Abandoning the pixel-level uncertainties calculated in Eq. 5 in our manuscript, and directly averaging the predictions by two networks using equal weights, say $\tilde{Y}_m := (O_m^1 + O_m^2)/2$; 3) **w/o Dist(\cdot, \cdot)**. Removing Dist(\cdot, \cdot) in Eq. 2 in our manuscript; 4) **w/o mixed training**. For each epoch, the samples in the labelled set \mathbf{D}_l are first used for training, and after training these labelled samples, the training of other unlabelled samples in \mathbf{D}_u begins; 5) **w/o \mathcal{L}_{iou}** . Removing \mathcal{L}_{iou} during training, and only using \mathcal{L}_{bce} ; 6) **w/o \mathcal{L}_{bce}** . Removing \mathcal{L}_{bce} during training, and only using \mathcal{L}_{iou} .

2 EXPLORATION OF NETWORK STRUCTURE

In this section, we embark on a series of rigorous experiments aimed at investigating the impact of altering the network architecture on accuracy. We first discuss the effectiveness of dilated convolutions in our proposed cross receptive field fusion module (CRFM), the quantitative results of which are shown in Tab. 2. To discuss the impacts of dilation rates in our CRFM, we adjust the dilation rates of five dilated convolutions used in CRFM. As depicted in Tab. 2, “ $D := 1, 1, 1, 1, 1$ ”, “ $D := 2, 2, 2, 2, 2$ ”, “ $D := 3, 3, 3, 3, 3$ ”, “ $D := 4, 4, 4, 4, 4$ ”, and “ $D := 5, 5, 5, 5, 5$ ” represent that the dilation rates of five convolutions are all set to be 1, 2, 3, 4, and 5, respectively. “ $D := 1, 2, 3, 4, 5$ ” denotes that the dilation rates from the 1-st to 5-th convolutions, are set to 1, 2, 3, 4, and 5, respectively.

The quantity of modules exerts a notable influence on performance, necessitating a thorough discussion. Based on this, we conduct the experiments to change the number of our proposed SIM. For example, “stage 1,2,3,4” in Tab.2 indicates that the features produced by the 1-st, 2-nd, 3-rd, and 4-th blocks are processed by SIM, while the features generated by the 5-th block are not processed by SIM.

We also report the results of different image sizes utilized as inputs for ResNet50, represented as “ 678×678 ” and “ 646×646 ” in Tab. 2.

3 MORE RESULTS

In addition to Fig. 3 in our manuscript, we provide more visual comparison with other fully-supervised competitors, as shown in Fig. 1 and Fig. 2. The competitors include: SiNet [2], C2FNet [9], FEDER [3], FSPNet [4], and ZoomNet [7].

REFERENCES

- [1] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. 2022. Concealed object detection. *TPAMI* 44, 10 (2022), 6024–6042.
- [2] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. 2020. Camouflaged object detection. In *CVPR*. 2774–2784.
- [3] Chunming He, Kai Li, Yachao Zhang, Longxiang Tang, Yulun Zhang, Zhenhua Guo, and Xiu Li. 2023. Camouflaged object detection with feature decomposition and edge reconstruction. In *CVPR*. 22046–22055.
- [4] Zhou Huang, Hang Dai, Tian-Zhu Xiang, Shuo Wang, Huai-Xin Chen, Jie Qin, and Huan Xiong. 2023. Feature shrinkage pyramid for camouflaged object detection with transformers. In *CVPR*. 5557–5566.
- [5] Trung-Nghia Le, Tam V. Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. 2019. Anabran network for camouflaged object segmentation. *Comput. Vis. Image Underst.* 184 (2019), 45–56. <https://doi.org/10.1016/J.CVIU.2019.04.006>
- [6] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. 2021. Simultaneously localize, segment and rank the camouflaged objects. In *CVPR*. 11591–11601.
- [7] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan Lu. 2022. Zoom in and out: a mixed-scale triplet network for camouflaged object detection. In *CVPR*. 2150–2160.
- [8] Przemysław Skurowski, Hassan Abdulameer, J Błaszczyk, Tomasz Depta, Adam Kornacki, and P Koziel. 2017. Animal camouflage analysis: chameleon database. <http://kgwisc.aei.polsl.pl/index.php/pl/dataset/63-animal-camouflage-analysis> (2017).
- [9] Yujia Sun, Geng Chen, Tao Zhou, Yi Zhang, and Nian Liu. 2021. Context-aware cross-level fusion network for camouflaged object detection. In *IJCAI*, Zhi-Hua Zhou (Ed.). 1025–1031.

Table 1: More ablation studies on four datasets. The best results are highlighted in Bold, while the second best results are marked in *Italic*. All the results are trained through selecting 40% from the training dataset as the labelled samples, and the remaining data is unlabelled.

Method	CAMO				CHAMELEON				COD10K				NC4K			
	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑
w/o condition	.061	.830	.798	.890	.026	.901	.884	.953	.027	.854	.794	.921	.043	.871	.809	.912
w/o pixel-level uncertainties	.060	.831	.796	.892	.025	.904	.881	.959	.026	.857	.796	.923	.041	.870	.812	.915
w/o Dist(\cdot, \cdot)	.059	.835	.803	.910	.023	.912	.887	.962	.027	.856	.791	.919	.040	.870	.815	.920
w/o mixed training	.063	.826	.798	.895	.029	.892	.874	.950	.032	.844	.781	.903	.046	.862	.794	.905
w/o \mathcal{L}_{iou}	.060	.819	.791	.889	.023	.910	.890	.958	.027	.853	.792	.918	.040	.870	.815	.916
w/o \mathcal{L}_{bce}	.061	.832	.791	.891	.027	.901	.883	.955	.028	.851	.796	.914	.042	.871	.811	.913
Ours	.057	.839	.805	.914	.020	.917	.899	.968	.025	.861	.797	.935	.039	.873	.821	.924

Table 2: Quantitative results of altering the network structure. The best results are highlighted in Bold, while the second best results are marked in *Italic*. All the results are trained through selecting 40% from the training dataset as the labelled samples, and the remaining data is unlabelled.

Method	CAMO				CHAMELEON				COD10K				NC4K			
	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑	MAE↓	S_m ↑	F_β^w ↑	E_m ↑
$D := 1, 1, 1, 1, 1$.064	.817	.788	.886	.029	.882	.873	.947	.032	.840	.780	.901	.047	.868	.802	.899
$D := 2, 2, 2, 2, 2$.061	.830	.794	.890	.026	.903	.882	.957	.027	.855	.797	.919	.041	.870	.808	.904
$D := 3, 3, 3, 3, 3$.062	.826	.790	.892	.030	.892	.872	.950	.031	.843	.781	.906	.042	.866	.794	.903
$D := 4, 4, 4, 4, 4$.063	.825	.792	.890	.031	.890	.871	.948	.030	.844	.779	.902	.043	.861	.792	.907
$D := 5, 5, 5, 5, 5$.061	.827	.794	.893	.028	.893	.876	.955	.029	.847	.785	.905	.045	.863	.795	.908
$D := 1, 2, 3, 4, 5$.062	.824	.797	.891	.030	.891	.873	.952	.030	.846	.782	.904	.044	.865	.793	.907
Stage 1,2,3,4	.064	.821	.790	.887	.028	.886	.875	.950	.031	.842	.790	.905	.045	.862	.805	.903
Stage 2,3,4,5	.061	.829	.794	.893	.024	.902	.884	.957	.027	.859	.798	.925	.042	.873	.814	.913
Stage 1,3,4,5	.062	.822	.798	.903	.025	.905	.886	.954	.030	.850	.793	.901	.044	.870	.809	.907
Stage 1,2,4,5	.061	.827	.796	.906	.026	.909	.883	.957	.031	.847	.789	.909	.041	.865	.812	.916
Stage 1,2,3,5	.060	.825	.801	.911	.023	.908	.889	.962	.027	.855	.793	.914	.042	.867	.818	.914
678×678	.059	.831	.800	.907	.025	.904	.883	.959	.030	.851	.785	.907	.040	.869	.815	.917
646×646	.061	.828	.799	.898	.028	.895	.877	.953	.031	.847	.783	.905	.045	.864	.797	.908
Ours	.057	.839	.805	.914	.020	.917	.899	.968	.025	.861	.797	.935	.039	.873	.821	.924

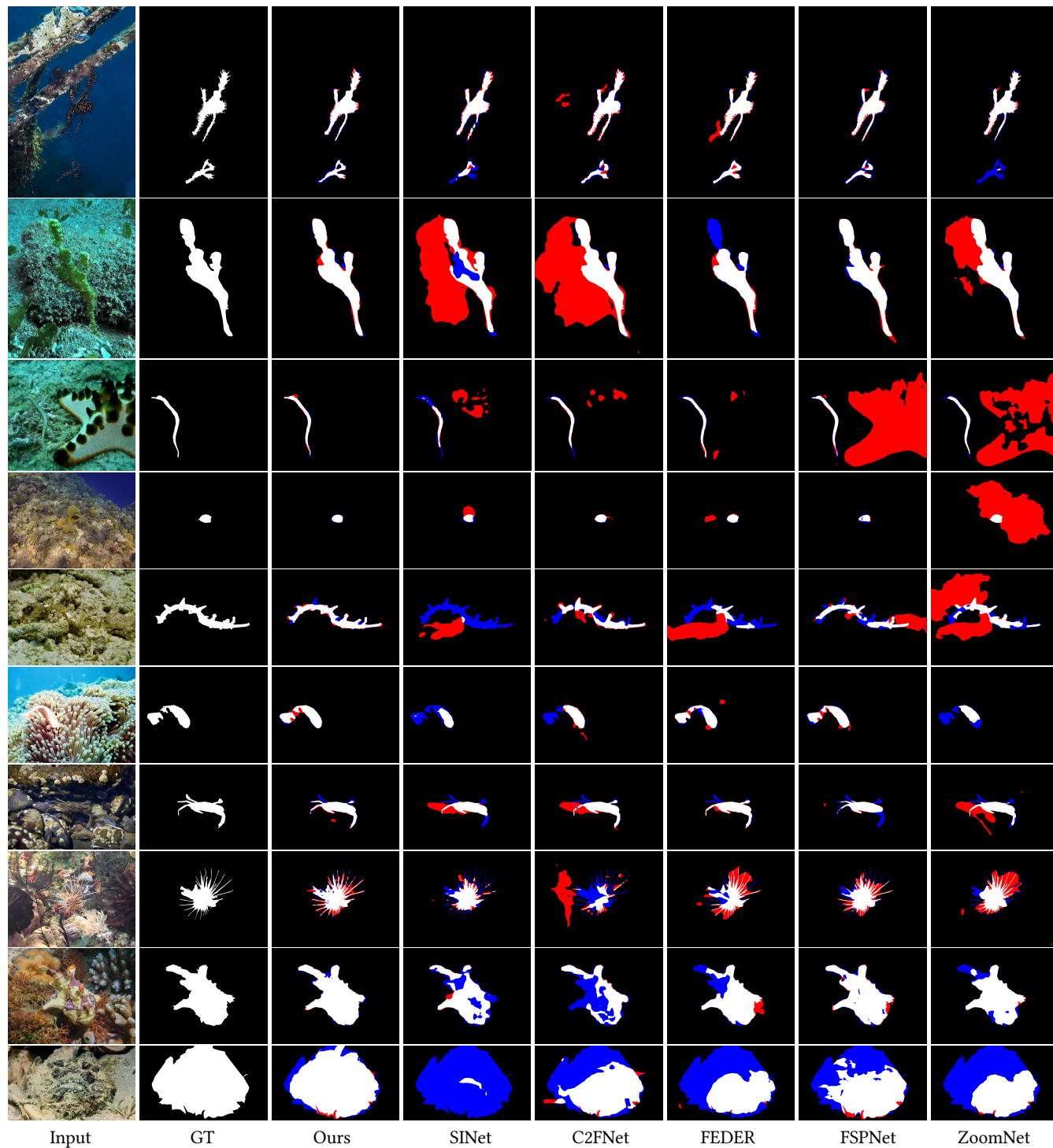


Figure 1: Visual comparison with other fully-supervised COD methods. The red and blue regions represent false-positive and false-negative predictions, respectively.

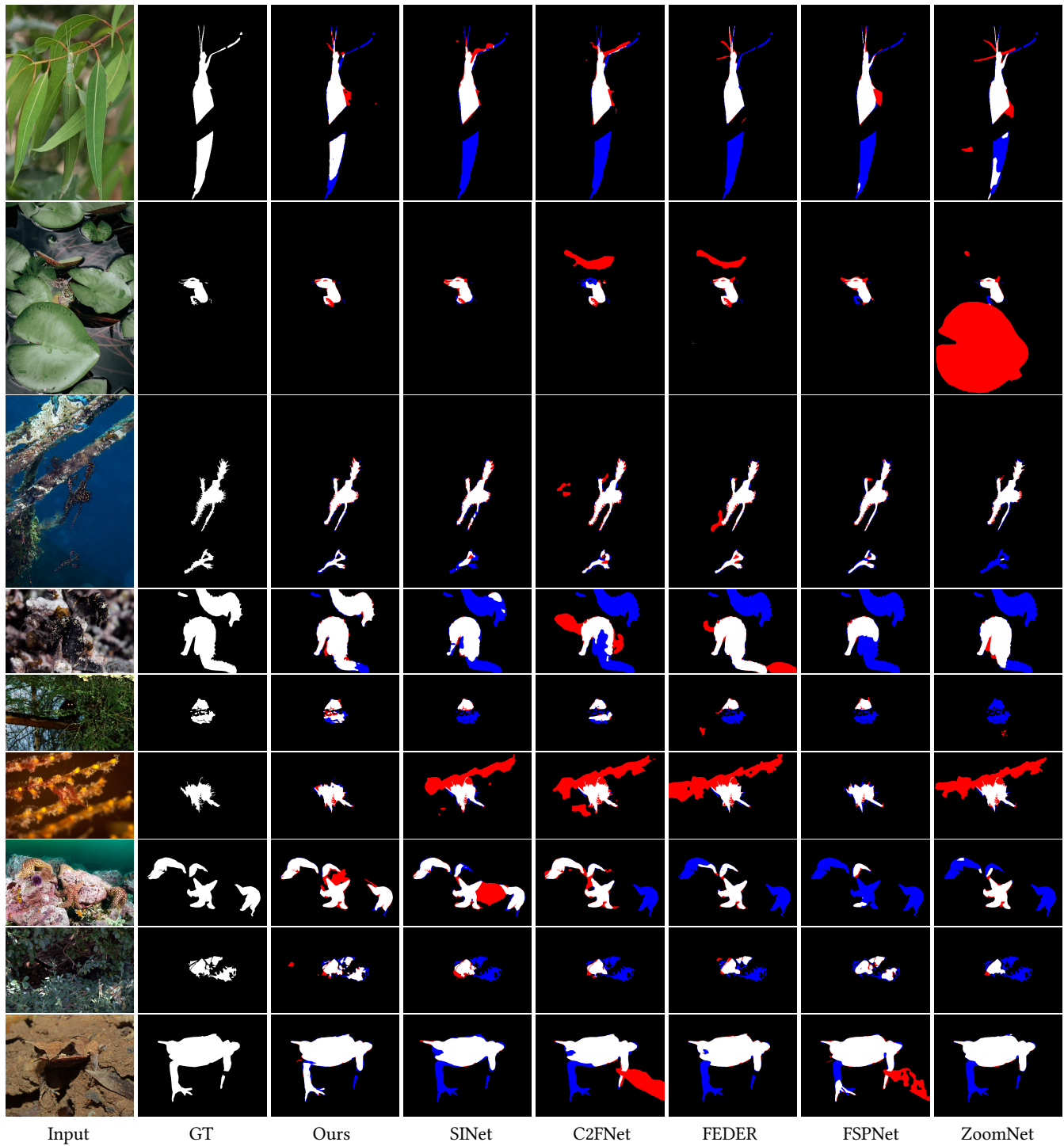


Figure 2: Visual comparison with other fully-supervised COD methods. The red and blue regions represent false-positive and false-negative predictions, respectively.