

MiNet: Weakly-Supervised Camouflaged Object Detection through Mutual Interaction between Region and Edge Cues (Supplementary Materials)

Yuzhen Niu
Fuzhou University
Fuzhou, China
yuzhenniu@gmail.com

Lifen Yang
Fuzhou University
Fuzhou, China
yanglifena@gmail.com

Rui Xu
Fuzhou University
Fuzhou, China
xurui.ryan.chn@gmail.com

Yuezhou Li
Fuzhou University
Fuzhou, China
liyuezhou.cm@gmail.com

Yuzhong Chen
Fuzhou University
Fuzhou, China
yzchen@fzu.edu.cn

In this supplementary material, we provide more analyses on our proposed MiNet. Firstly, we present additional examples from the training dataset in Section 1 to illustrate the scribble annotation. Secondly, we offer a more comprehensive visual comparison with other methods in Section 2. Finally, we conduct more ablation studies in Section 3 to demonstrate the effectiveness and generalization of our MiNet.

1 Illustration of Scribble Annotation

The scribble annotation method offers great flexibility, significantly reducing the time and labor costs required for annotating a large scale dataset. Additionally, it effectively addresses the limitation of pixel-wise ground truth, such as missing the primary structure of an object. The scribble-based camouflaged object detection dataset (S-COD) [2] is proposed recently and employed for training purposes, which comprises 4,040 images, meticulously conducted by re-labeling 1,000 images from the CAMO [3] dataset and 3,040 images from the COD10K [1] dataset using the scribble annotation method. Following the settings of [2], we also use the S-COD dataset for training in our experiments.

Furthermore, to provide a more elucidating depiction of the scribble annotation, we present several examples from S-COD in Fig. 2, which encompasses challenging scenarios such as multiple objects, occluded objects, disruptive coloration object, elongated object, small object, uncertain object, and large object. The binary labeling approach typically uses ‘0’ to represent the background and ‘1’ to represent the foreground. However, in the case of scribble-based ground truth (GT), foreground, background, and unknown pixels are identified as ‘1’, ‘2’, and ‘0’, respectively. In Fig. 2, the third and sixth columns showcase the corresponding scribble GTs of the input images. For illustration in Fig. 2, we use red and blue scribbles to show the foreground and background, and the unknown pixels are showed in their original colors.

2 More Visual Comparisons

2.1 Visual comparisons with other methods

As shown in Fig. 3, we present more qualitative results comparing our method with other state-of-the-art methods (including three fully-supervised camouflaged object detection (COD) methods, i.e.,

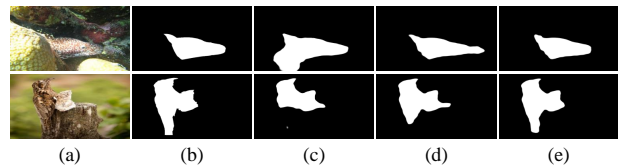


Figure 1: Visual demonstration of different levels of interaction. (a) Input; (b) GT; (c) Only deepest-level interaction; (d) Only shallowest-level interaction; (e) Multi-level interaction.

ZoomNet [6], UGTR [9], and SINet [1], one scribble-based weakly-supervised camouflaged object detection (WSCOD) method, i.e., CRNet [2], and two scribble-based weakly-supervised salient object detection (SOD) methods, i.e., SCWS [10] and SS [11]) on four COD benchmark datasets (i.e., CAMO [3], CHAMELEON [7], COD10K [1], and NC4K [5]). It can be observed that when the surroundings of the camouflaged object are highly cluttered, our method achieves more precise boundary localization of the object. However, some methods are influenced by the cluttered background (2-nd row), leading to inaccuracies in object boundary localization. In scenarios featuring large object, some methods even fail to capture the primary structure of the object (6-th row). For some scenes do not have cluttered backgrounds, since the boundaries of the camouflaged objects seamlessly blend with the environment, they also pose a great challenge for boundary localization (1-st row). In contrast, our method demonstrates superior detection performance in boundary localization for a variety of scenarios.

3 More Ablation Studies

In this section, we conduct more ablation studies on two COD datasets (i.e., CAMO [3] and COD10K [1]) and two SOD datasets (i.e., ECSSD [8] and HKU-IS [4]) to validate the effectiveness and generalization of our proposed MiNet.

3.1 Efficacy of different levels of interaction

In order to investigate the efficacy of multi-level interaction within the region-boundary refinement net, we devise two variant models utilizing single-level interaction: one that interacts edge map M_e and region feature F_4 solely at the deepest level and another that



Figure 2: Illustration of the scribble annotation across various scenarios in the S-COD dataset [2]. Each subfigure displays, from left to right, the Image, the Pixel-wise Ground Truth, and the Scribble-based Ground Truth.

interacts edge map M_e and region feature F_1 exclusively at the shallowest level.

Fig. 1 presents the visual comparisons of above three models. It is evident that the interaction at the shallowest level provides better guidance for learning object boundaries (Fig. 1(d)), compared to the interaction on the deepest-level (Fig. 1(c)). Furthermore, by facilitating the iterative refinement of object boundaries in a multi-level manner with edge map M_e , it becomes more feasible to enhance the precision in localizing the object boundaries by combining the semantic and detail information (Fig. 1(e)).

3.2 Impact of different combined probability maps

In the region-aware guidance module (RGM), we adopt a probability-based spatial suppressing (PSS) operation to help suppress non-object noise. In this operation, diverse probability maps P_i (calculated based on s_i) are aggregated to compute the combined probability map P . In Table 1, we present several comparison results to investigate the impact of different s_i sets on the combined probability maps P . It can be seen that the number of elements in the s_i set has varying impacts on the performance of the model. As the number of elements in the s_i set increases, the model's performance gradually improves and reaches the best result at $s_i \in \{\frac{1}{2}N, \dots, \frac{4}{5}N\}$. If we

Table 1: Impact of different s_i sets on combined probability maps P , where $N = HW$.

No.	Set of s_i	CAMO				COD10K			
		$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$
①	$s_i \in \{\frac{1}{2}N\}$	0.094	0.749	0.828	0.658	0.052	0.745	0.835	0.585
②	$s_i \in \{\frac{1}{2}N, \frac{2}{3}N\}$	0.092	0.747	0.834	0.659	0.050	0.745	0.838	0.589
③	$s_i \in \{\frac{1}{2}N, \frac{2}{3}N, \frac{3}{4}N\}$	0.091	0.749	0.834	0.664	0.050	0.748	0.840	0.593
④	$s_i \in \{\frac{1}{2}N, \dots, \frac{4}{5}N\}$	0.091	0.750	0.840	0.669	0.049	0.749	0.840	0.596
⑤	$s_i \in \{\frac{1}{2}N, \dots, \frac{5}{6}N\}$	0.092	0.749	0.837	0.665	0.051	0.746	0.837	0.588

keep increasing the number of elements in the s_i set, the model may retain too much uncertainty in the region feature, potentially preserving a significant amount of non-object noise, thereby reducing the discriminability of the edge map M_e and gradually decreasing the model's performance.

3.3 Impact of each loss function

The choice of loss functions is critical for WSCOD task. We analyze the impact of individual loss functions, and the results are presented in Table 2. Specifically, the performance differences between Table 2 ① and ⑤ highlights the performance improvements resulting from the incorporation of discriminative edge cues. Table 2 ② and ③ demonstrate the indispensability of region loss (\mathcal{L}_{reg})

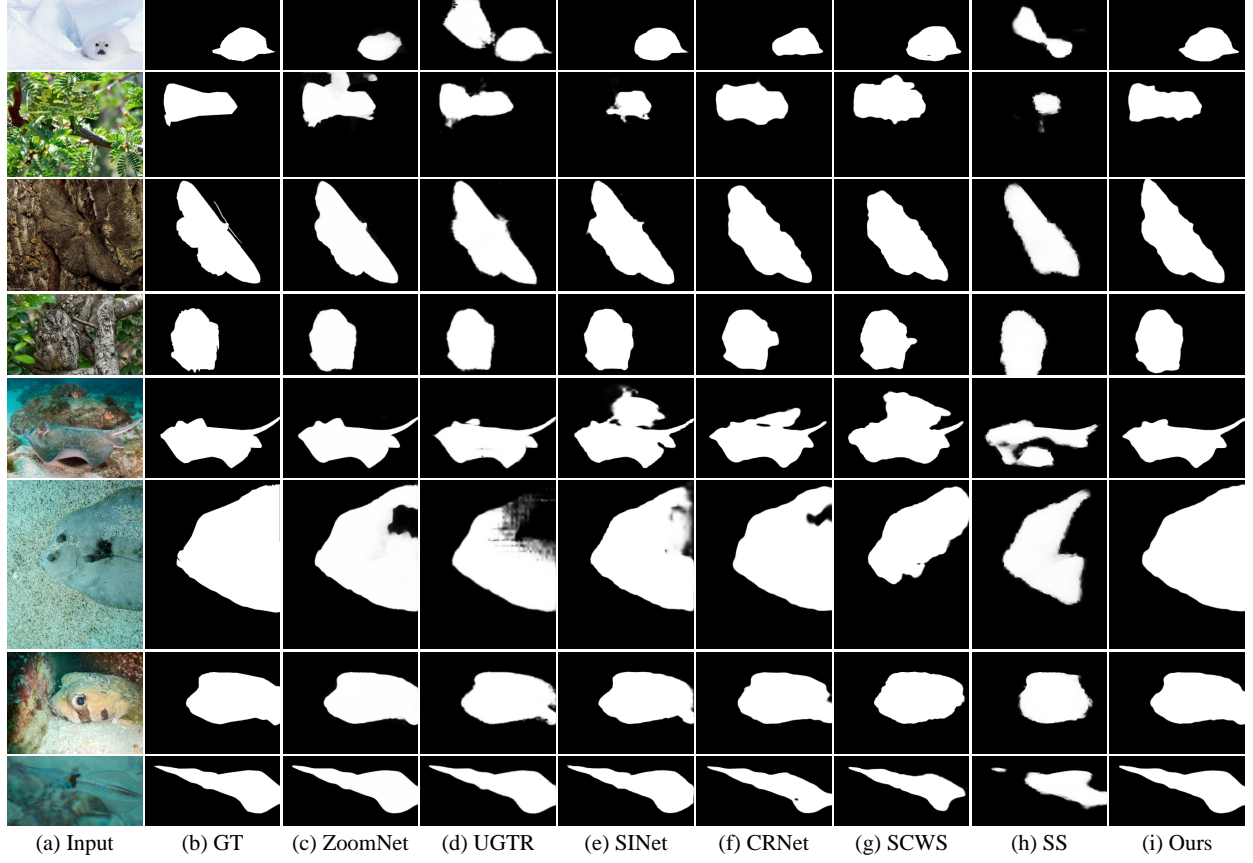


Figure 3: Qualitative comparison of the proposed MiNet with three fully-supervised COD methods, one scribble-based weakly-supervised COD method, and another two scribble-based weakly-supervised SOD methods. The visual results from top to bottom are for CAMO [3], CHAMELEON [7], COD10K [1], and NC4K [5] datasets, with two examples given for each dataset.

and boundary localization loss (\mathcal{L}_{bl}). Furthermore, Table 2 ④ illustrates that employing deep supervision strategy for the model’s intermediate prediction results also contributes to enhancing the model’s performance.

3.4 Performance of MiNet in SOD

The salient object detection (SOD) task and COD task share a certain degree of commonality, both being binary classification tasks. However, camouflaged objects are typically concealed within complex environments, greatly increasing the challenge of detection. In this subsection, we further validate the generalization of our proposed MiNet on scribble-based weakly-supervised salient object detection task. The quantitative results on ECSSD [8] and HKU-IS [4] datasets are reported in Table 3, revealing that our proposed MiNet outperforms others in weakly-supervised salient object detection task. Additionally, we present a qualitative comparison of our proposed MiNet with two state-of-the-art scribble-based weakly-supervised SOD methods (i.e., SCWS [10] and SS [11]) in Fig. 4. It’s evident that our proposed MiNet effectively delineates the boundaries of salient objects.

Table 2: Ablation study on loss function.

No.	Settings	CAMO				COD10K			
		$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$
①	w/o \mathcal{L}_{edge}	0.099	0.738	0.824	0.645	0.055	0.736	0.828	0.571
②	w/o \mathcal{L}_{reg}	0.286	0.426	0.268	0.118	0.228	0.454	0.284	0.068
③	w/o \mathcal{L}_{bl}	0.194	0.617	0.677	0.444	0.229	0.510	0.507	0.254
④	w/o \mathcal{L}_{aux}	0.096	0.747	0.829	0.655	0.054	0.744	0.830	0.580
⑤	Ours	0.091	0.750	0.840	0.669	0.049	0.749	0.840	0.596

Table 3: Quantitative results on two SOD datasets.

No.	Methods	ECSSD			HKU-IS		
		$M \downarrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$F_\beta \uparrow$
①	SS [11]	0.0610	0.9077	0.8650	0.0470	0.9232	0.8576
②	SCWS [10]	0.0489	0.9079	0.8995	0.0375	0.9376	0.8962
③	Ours	0.0439	0.9364	0.8996	0.0349	0.9473	0.8924

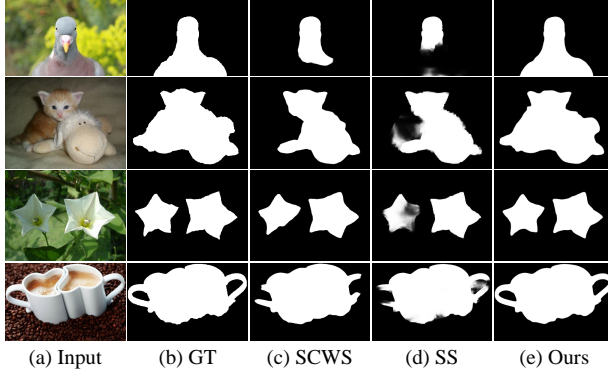


Figure 4: Qualitative comparison of the proposed MiNet with two scribble-based weakly-supervised SOD methods.

Table 4: Quantitative results on different degrees of motion blur.

No.	Settings	CAMO				COD10K			
		$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$
①	severe blur	0.100	0.733	0.818	0.637	0.052	0.747	0.835	0.593
②	moderate blur	0.096	0.742	0.831	0.653	0.050	0.750	0.837	0.597
③	w/o blur	0.091	0.750	0.840	0.669	0.049	0.749	0.840	0.596

Table 5: Ablation study on label noise.

No.	Settings	CAMO				COD10K			
		$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$
①	10% labels shifting	0.096	0.738	0.821	0.649	0.045	0.819	0.903	0.736
①	Ours	0.091	0.750	0.840	0.669	0.049	0.749	0.840	0.596

3.5 Noise Sensitivity of Weakly-supervised Learning

We also conduct some experiments to explore the noise sensitivity of weakly-supervised learning. First, we employ different degrees of motion blur to achieve blurred object boundaries. As shown in

Table 4, the model shows stable performance with mild blur levels. However, when severe blur is applied, there is a slight decline in performance. For instance, F_β^ω decrease by 1% on COD10K. This highlights that our method maintains reliability against moderate levels of boundary blur. In addition, we introduce noise to the scribble labels by shifting labels horizontally and vertically for 10% training images. As shown in Table 5, label noise has a slight impact on the model's performance. On COD10K, the S_α and F_β^ω decrease by 1% and 2%, respectively.

References

- [1] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. 2020. Camouflaged object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2777–2787.
- [2] Ruozhen He, Qihua Dong, Jiaying Lin, and Rynson WH Lau. 2023. Weakly-supervised camouflaged object detection with scribble annotations. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 781–789.
- [3] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. 2019. Anabran network for camouflaged object segmentation. *Computer Vision and Image Understanding* 184 (2019), 45–56.
- [4] Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. 2014. The secrets of salient object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 280–287.
- [5] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 11591–11601.
- [6] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2160–2170.
- [7] Przemyslaw Skurowski, Hassan Abdulameer, J Blaszczuk, Tomasz Depta, Adam Kornacki, and P Kozielec. 2018. Animal camouflage analysis: Chameleon database. *Unpublished Manuscript* 2, 6 (2018), 7.
- [8] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. 2013. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1155–1162.
- [9] Fan Yang, Qiang Zhai, Xin Li, Rui Huang, Ao Luo, Hong Cheng, and Deng-Ping Fan. 2021. Uncertainty-guided transformer reasoning for camouflaged object detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 4146–4155.
- [10] Siyue Yu, Bingfeng Zhang, Jimin Xiao, and Eng Gee Lim. 2021. Structure-consistent weakly supervised salient object detection with local saliency coherence. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3234–3242.
- [11] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai. 2020. Weakly-supervised salient object detection via scribble annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 12546–12555.