Can SAM Segment Anything?

-When SAM Meets Camouflaged Object Detection

Lv Tang[1]   Haoke Xiao[3]   Bo Li[2*]

[1]Independent Researcher, Beijing, China
[2]Youtu Lab, Tencent, Shanghai, China
[3]Xiamen University, Xiamen, China

luckybird1994@gmail.com, hk.xiao.me@gmail.com, libraboli@tencent.com

Abstract

SAM is a segmentation model recently released by Meta AI Research and has been gaining attention quickly due to its impressive performance in generic object segmentation. However, its ability to generalize to specific scenes such as camouflaged scenes is still unknown. Camouflaged object detection (COD) involves identifying objects that are seamlessly integrated into their surroundings and has numerous practical applications in fields such as medicine, art, and agriculture. In this study, we try to ask if SAM can address the COD task and evaluate the performance of SAM on the COD benchmark by employing maximum segmentation evaluation and camouflage location evaluation. We also compare SAM's performance with 22 state-of-the-art COD methods. Our results indicate that while SAM shows promise in generic object segmentation, its performance on the COD task is limited. This presents an opportunity for further research to explore how to build a stronger SAM that may address the COD task. The results of this paper are provided in https://github.com/luckybird1994/SAMCOD.

1. Introduction

Large language models (LLMs), also known as foundation models [1,2], have achieved significant performance on various natural language processing (NLP) tasks. Recently, OpenAI released a chatbot service called ChatGPT, which is a variant of the Generative Pre-trained Transformers (GPT) family. Thanks to its user-friendly interface and exceptional performance, ChatGPT has attracted over 100 million users within just two months.

Unlike NLP, there is still much to explore in the area of foundation models in computer vision (CV). While some models like CLIP [24] and ALIGN [14] have been designed, many problems in CV remain unsolved. Recently, Meta AI Research developed a foundation model called SAM [16] for image segmentation. SAM is a promptable model that is pre-trained on a broad dataset using a task that enables powerful generalization. The remarkable performance of SAM in image segmentation has astonished researchers in CV. Some have even gone as far as to exclaim that "CV does not exist", a reference to a famous line from the science fiction novel Three-Body.

In this technical report, we evaluate the performance of SAM when applied to the task of camouflaged object detection (COD). The goal of COD is to accurately identify objects that are seamlessly blended into their background. This task has significant scientific and practical value in various fields, such as medicine [9,11], agriculture [4,27], and art [5]. To evaluate SAM's performance on COD benchmarks, we follow two rules: 1) Can SAM accurately segment camouflaged objects? and 2) Can SAM accurately locate camouflaged objects? Therefore, we use maximum segmentation evaluation and camouflage location evaluation to assess the performance of SAM. Our comparison results between SAM and 22 state-of-the-art COD methods reveal that SAM has limitations when it comes to COD tasks. We hope that our findings can inspire and give confidence to researchers in the COD field, despite SAM's limitations in addressing the COD task. Furthermore, we believe that this presents an opportunity for further research to explore how to effectively apply SAM to the COD task.

---

*Corresponding author

2. Experiments

2.1. Datasets, Comparison and Metrics

Datasets. To validate the effectiveness of SAM, we evaluate its performance on three benchmark datasets, containing CAMO [17], COD10K [8] and NC4K [20].

Comparison Methods. We compare SAM with 22 state-of-the-art COD methods, including the SINet [8], C2FNet [25], LSR [20], PFNet [22], MGL [29], JCOD [18], TANet [34], BGNet [26], FDCOD [31], SegMaR [15], ZoomNet [23], BSANet [33], SINetV2 [7], FAPNet [32], the extension version of C2FNet [3], DGNet [12], CubeNet [36], ERRNet [13], VST [19], UGTR [28], ICON [35] and TPRNet [30].

Evaluation Metrics. We use six widely used metrics: structure-measure (S_α) [6], mean E-measure (E_ϕ) [10], F-measure (F_β), weighted F-measure (F_β^w), max F-measure (F_β^{max}) [21], and mean absolute error (MAE).

Table 1. Comparison Results between SAM and 22 state-of-the-art COD methods.

| Methods	Pub.	CAMO-Test (250 images)						COD10K-Test (2026 images)						NC4K (4121 images)					
		S_α	E_ϕ	F_β^w	F_β	F_β^{max}	MAE	S_α	E_ϕ	F_β^w	F_β	F_β^{max}	MAE	S_α	E_ϕ	F_β^w	F_β	F_β^{max}	MAE
CNN-Based Models																			
SINet	CVPR'20	0.751	0.771	0.606	0.675	0.706	0.100	0.771	0.806	0.551	0.634	0.676	0.051	0.808	0.871	0.723	0.769	0.775	0.058
C2FNet	IJCAI'21	0.796	0.854	0.719	0.762	0.771	0.080	0.812	0.890	0.686	0.723	0.743	0.036	0.838	0.897	0.762	0.795	0.810	0.049
LSR	CVPR'21	0.787	0.838	0.696	0.744	0.753	0.080	0.804	0.880	0.673	0.715	0.732	0.037	0.840	0.895	0.766	0.804	0.815	0.048
PFNet	CVPR'21	0.782	0.841	0.695	0.746	0.758	0.085	0.800	0.877	0.660	0.701	0.725	0.040	0.829	0.887	0.745	0.784	0.799	0.053
MGL	CVPR'21	0.775	0.812	0.673	0.726	0.740	0.088	0.814	0.851	0.666	0.710	0.738	0.035	0.833	0.867	0.739	0.782	0.800	0.053
JCOD	CVPR'21	0.800	0.859	0.728	0.772	0.779	0.073	0.809	0.884	0.684	0.721	0.738	0.035	0.842	0.898	0.771	0.806	0.816	0.047
TANet	AAAI'21	0.781	0.847	0.678	-	-	0.087	0.793	0.848	0.635	-	-	0.043	-	-	-	-	-	-
BGNet	IJCAI'22	0.813	0.870	0.749	0.789	0.799	0.073	0.831	0.901	0.722	0.753	0.774	0.033	0.851	0.907	0.788	0.820	0.833	0.044
FDCOD	CVPR'22	0.828	0.883	0.748	0.781	0.804	0.068	0.832	0.907	0.706	0.733	0.776	0.033	0.834	0.893	0.750	0.784	0.804	0.051
SegMaR	CVPR'22	0.815	0.874	0.753	0.795	0.803	0.071	0.833	0.899	0.724	0.757	0.774	0.034	0.841	0.896	0.781	0.821	0.826	0.046
ZoomNet	CVPR'22	0.820	0.877	0.752	0.794	0.805	0.066	0.838	0.888	0.729	0.766	0.780	0.029	0.853	0.896	0.784	0.818	0.828	0.043
BSANet	AAAI'22	0.794	0.851	0.717	0.763	0.770	0.079	0.817	0.891	0.699	0.738	0.753	0.034	0.841	0.897	0.771	0.808	0.817	0.048
SINetV2	PAMI'22	0.822	0.882	0.743	0.782	0.801	0.070	0.815	0.887	0.680	0.718	0.752	0.037	0.847	0.903	0.770	0.805	0.823	0.048
FAPNet	TIP'22	0.815	0.865	0.734	0.776	0.792	0.076	0.822	0.887	0.694	0.731	0.758	0.036	0.851	0.899	0.775	0.810	0.825	0.046
C2FNet-ext	TCSVT'22	0.799	0.859	0.730	0.770	0.779	0.077	0.811	0.887	0.691	0.725	0.742	0.036	0.840	0.896	0.770	0.802	0.814	0.048
DGNet	MIR'22	0.839	0.900	0.768	0.806	0.822	0.057	0.822	0.896	0.693	0.728	0.759	0.033	0.854	0.909	0.783	0.813	0.830	0.043
CubeNet	PR'22	0.788	0.838	0.682	0.732	0.750	0.085	0.795	0.865	0.643	0.692	0.715	0.041	-	-	-	-	-	-
ERRNet	PR'22	0.779	0.842	0.679	0.729	0.742	0.085	0.786	0.867	0.630	0.675	0.702	0.043	0.827	0.887	0.737	0.778	0.794	0.054
Transformer-Based Models																			
VST	ICCV'21	0.807	0.848	0.713	0.758	0.777	0.081	0.820	0.879	0.698	0.738	0.754	0.037	0.845	0.893	0.767	0.804	0.817	0.048
UGTR	ICCV'21	0.785	0.822	0.685	0.737	0.753	0.086	0.818	0.852	0.667	0.712	0.742	0.035	0.839	0.874	0.746	0.787	0.807	0.052
ICON	PAMI'22	0.840	0.894	0.769	0.796	0.824	0.058	0.818	0.904	0.688	0.717	0.756	0.033	0.847	0.911	0.784	0.697	0.817	0.045
TPRNet	TVCJ'22	0.807	0.861	0.725	0.772	0.785	0.074	0.817	0.887	0.683	0.724	0.748	0.036	0.846	0.898	0.768	0.805	0.820	0.048
SAM	arXiv'23	0.684	0.687	0.606	0.680	0.681	0.132	0.783	0.798	0.701	0.756	0.758	0.050	0.767	0.776	0.696	0.752	0.754	0.078
		19%	24%	21%	16%	17%	56%	7%	12%	4%	1%	3%	42%	10%	15%	12%	4%	9%	45%
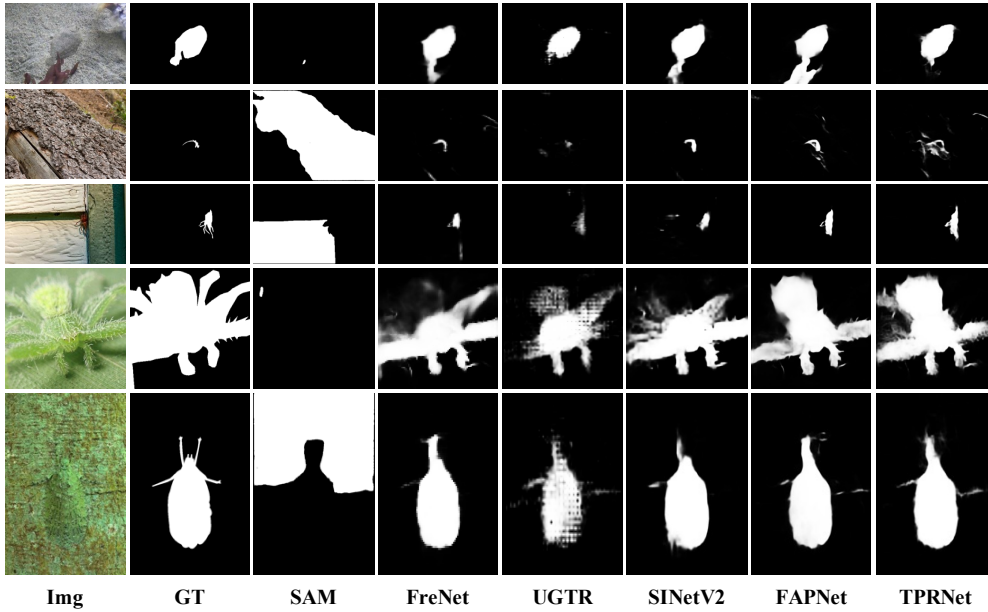
Figure 1. Some bad cases of SAM.

2.2. Maximum Segmentation Evaluation

In general, when evaluating the performance of a COD method on a certain benchmark, the performance of each predicted map is calculated first, and then they are averaged. This process for calculating the F_β score can be written as follows:

$$F_\beta = \frac{1}{N} \sum_{i=1}^{i=N} F_\beta^i,$$ (1)

where N is the total number of a dataset.

SAM generates multiple binary maps for each image, but not all of them contain the camouflaged object. To evaluate SAM's upper segmentation performance, we propose the maximum segmentation evaluation. For a given image, we calculate the mean F-measure for all predicted maps and select the one with the highest F-measure value as SAM's predicted map. Consequently, we express SAM's F_β on a particular dataset as:

$$HF^i = max\{F_1^i, F_2^i, ..., F_m^i\},$$
$$F_\beta = \frac{1}{N} \sum_{i=1}^{i=N} HF^i,$$ (2)

where m is the total number of predicted binary maps for one image, and F_m^i is the mean F-measure of one binary map. Then, we can calculate the remaining metrics for the SAM. As shown in Table. 1 and Fig. 1, the performance of SAM is still worse than state-of-the-art COD methods.

2.3. General Location Evaluation

Table 2. The ratios of predicted binary maps that exceed a certain F_β threshold

Thresholds	0.2	0.3	0.4	0.5	0.7
CAMO	2.55%	1.83%	1.39%	1.15%	0.82%
COD10K	2.32%	1.80%	1.48%	1.26%	0.93%
NC4K	2.94%	2.22%	1.82%	1.53%	1.13%

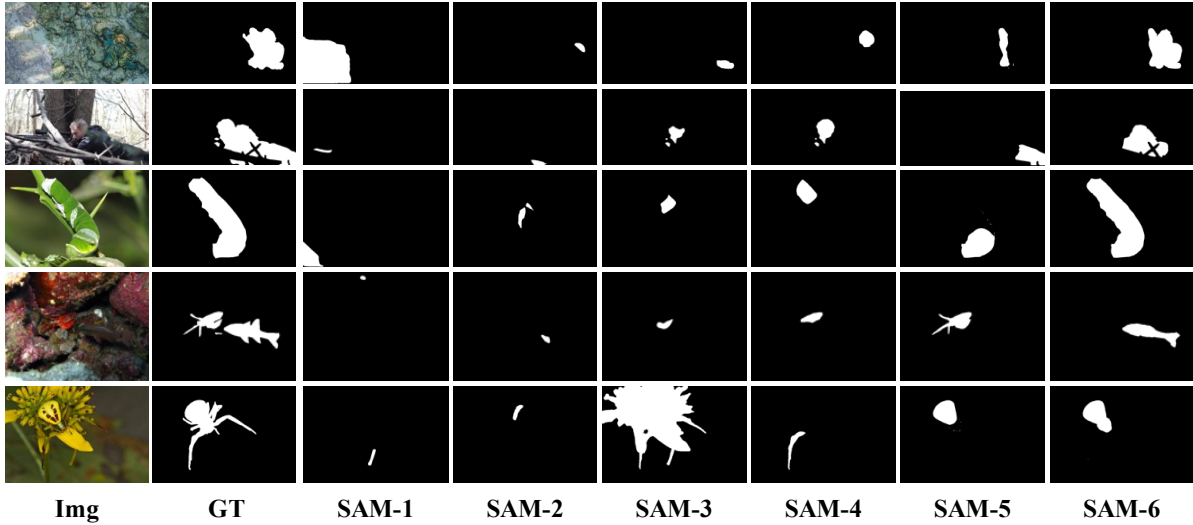| Img | GT | SAM-1 | SAM-2 | SAM-3 | SAM-4 | SAM-5 | SAM-6 |

Figure 2. The location ability of the SAM.

SAM generates multiple binary maps for each image, and we have designed the general location evaluation to assess how accurately SAM can detect the camouflaged object. Specifically, for a given image, we calculate the F_β for all predicted maps and then calculate the ratios of maps that exceed a certain F_β threshold. As shown in Table. 2 and Fig. 2, the location performance of SAM is needed to be further improved.

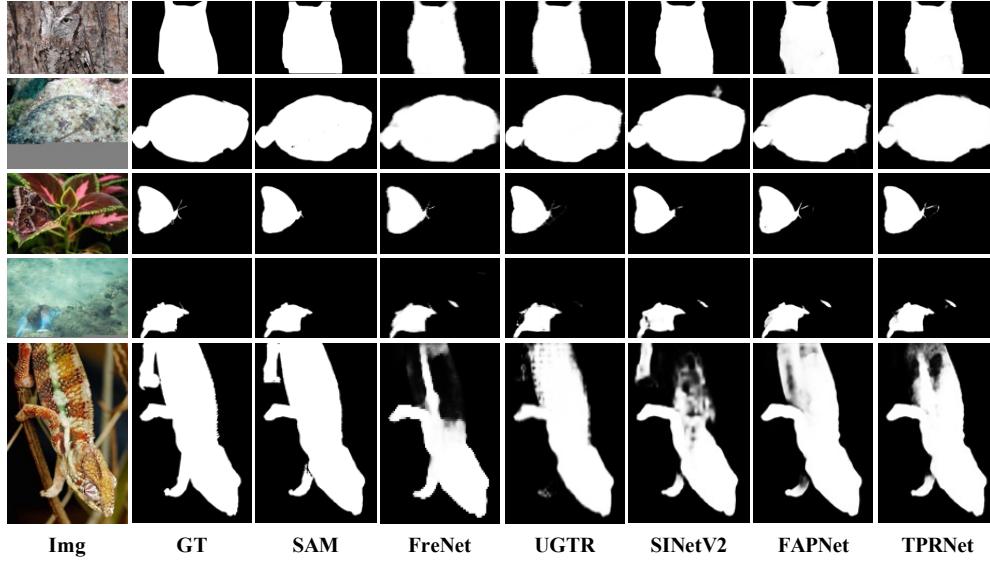| Img | GT | SAM | FreNet | UGTR | SINetV2 | FAPNet | TPRNet |

Figure 3. Some good cases of SAM.

3. Conclusion

This paper presents a preliminary evaluation of SAM's performance when applied to the COD task. While we acknowledge the advancement of large CV foundation models, our experiments show that there is still room for improvement in SAM's performance on this specific task. Although SAM's performance lags behind some state-of-the-art COD methods, Table. 1 and Fig. 3 show that SAM can achieve a noteworthy performance compared to some methods published in 2021. We hope that our evaluation and analysis can provide insights for future research, such as designing a stronger foundation model or improving the network architecture of SAM.

References

[1] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ B. Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri S. Chatterji, Annie S. Chen, Kathleen Creel, Jared Quincy Davis, Dorottya Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah D. Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark S. Krass, Ranjay Krishna, Rohith Kuditipudi, and et al. On the opportunities and risks of foundation models. *CoRR*, abs/2108.07258, 2021. 1

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Adv. Neural Inform. Process. Syst.*, 2020. 1

[3] Geng Chen, Si-Jie Liu, Yu-Jia Sun, Ge-Peng Ji, Ya-Feng Wu, and Tao Zhou. Camouflaged object detection via context-aware cross-level fusion. *IEEE Trans. Circuits Syst. Video Technol.*, 32(10):6981–6993, 2022. 2

[4] Xi Cheng, Youhua Zhang, Yiqiong Chen, Yun-Zhi Wu, and Yi Yue. Pest identification via deep residual learning in complex background. *Comput. Electron. Agric.*, 141:351–356, 2017. 1

[5] Hung-Kuo Chu, Wei-Hsin Hsu, Niloy J. Mitra, Daniel Cohen-Or, Tien-Tsin Wong, and Tong-Yee Lee. Camouflage images. *ACM Trans. Graph.*, 29(4):51:1–51:8, 2010. 1

[6] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 4558–4567. IEEE Computer Society, 2017. 2

[7] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(10):6024–6042, 2022. 2

[8] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *Conf. Comput. Vis. Pattern Recog.*, pages 2774–2784. IEEE, 2020. 2

[9] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans. Medical Imaging*, 39(8):2626–2637, 2020. 1

[10] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. *SCIENTIA SINICA Informationis*, 2021. 2

[11] 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 *Med. Image. Comput. Comput. Assist. Interv.*, pages 263–273. Springer, 2020. 1

[12] Ge-Peng Ji, Deng-Ping Fan, Yu-Cheng Chou, Dengxin Dai, Alexander Liniger, and Luc Van Gool. Deep gradient learning for efficient camouflaged object detection. *Int. J. Autom. Comput.*, 20(1):92–108, 2023. 2

[13] Ge-Peng Ji, Lei Zhu, Mingchen Zhuge, and Keren Fu. Fast camouflaged object detection via edge-based reversible re-calibration network. *Pattern Recognit.*, 123:108414, 2022. 2

[14] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 4904–4916. PMLR, 2021. 1

[15] Qi Jia, Shuilian Yao, Yu Liu, Xin Fan, Risheng Liu, and Zhongxuan Luo. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *CVPR*, pages 4703–4712. IEEE, 2022. 2

[16] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. 1

[17] Trung-Nghia Le, Tam V. Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Comput. Vis. Image Underst.*, 184:45–56, 2019. 2

[18] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *CVPR*, pages 10071–10081. Computer Vision Foundation / IEEE, 2021. 2

[19] Nian Liu, Ni Zhang, Kaiyuan Wan, Ling Shao, and Junwei Han. Visual saliency transformer. In *ICCV*, pages 4702–4712. IEEE, 2021. 2

[20] 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 *Conf. Comput. Vis. Pattern Recog.*, pages 11591–11601. IEEE, 2021. 2

[21] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 248–255. IEEE Computer Society, 2014. 2

[22] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *Conf. Comput. Vis. Pattern Recog.*, pages 8772–8781. IEEE, 2021. 2

[23] 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 *Conf. Comput. Vis. Pattern Recog.*, pages 2150–2160. IEEE, 2022. 2

[24] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021. 1

[25] Yujia Sun, Geng Chen, Tao Zhou, Yi Zhang, and Nian Liu. Context-aware cross-level fusion network for camouflaged object detection. In *Int. Joint Conf. Artif. Intell*, pages 1025–1031. ijcai.org, 2021. 2

[26] Yujia Sun, Shuo Wang, Chenglizhao Chen, and Tian-Zhu Xiang. Boundary-guided camouflaged object detection. In *Int. Joint Conf. Artif. Intell*, pages 1335–1341. ijcai.org, 2022. 2

[27] Xiaoping Wu, Chi Zhan, Yu-Kun Lai, Ming-Ming Cheng, and Jufeng Yang. IP102: A large-scale benchmark dataset for insect pest recognition. In *Conf. Comput. Vis. Pattern Recog.*, pages 8787–8796. IEEE, 2019. 1

[28] 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 *Int. Conf. Comput. Vis.*, pages 4126–4135. IEEE, 2021. 2

[29] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *Conf. Comput. Vis. Pattern Recog.*, pages 12997–13007. IEEE, 2021. 2

[30] Qiao Zhang, Yanliang Ge, Cong Zhang, and Hongbo Bi. Tprnet: camouflaged object detection via transformer-induced progressive refinement network. *The Visual Computer*, pages 1–15, 2022. 2

[31] Yijie Zhong, Bo Li, Lv Tang, Senyun Kuang, Shuang Wu, and Shouhong Ding. Detecting camouflaged object in frequency domain. In *Conf. Comput. Vis. Pattern Recog.*, pages 4494–4503. IEEE, 2022. 2

[32] Tao Zhou, Yi Zhou, Chen Gong, Jian Yang, and Yu Zhang. Feature aggregation and propagation network for camouflaged object detection. *IEEE Trans. Image Process.*, 31:7036–7047, 2022. 2

[33] Hongwei Zhu, Peng Li, Haoran Xie, Xuefeng Yan, Dong Liang, Dapeng Chen, Mingqiang Wei, and Jing Qin. I can find you! boundary-guided separated attention network for camouflaged object detection. In *AAAI*, pages 3608–3616. AAAI Press, 2022. 2

[34] Jinchao Zhu, Xiaoyu Zhang, Shuo Zhang, and Junnan Liu. Inferring camouflaged objects by texture-aware interactive guidance network. In *AAAI*, pages 3599–3607. AAAI Press, 2021. 2

[35] Mingchen Zhuge, Deng-Ping Fan, Nian Liu, Dingwen Zhang, Dong Xu, and Ling Shao. Salient object detection via integrity learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(3):3738–3752, 2023. 2

[36] Mingchen Zhuge, Xiankai Lu, Yiyou Guo, Zhihua Cai, and Shuhan Chen. Cubenet: X-shape connection for camouflaged object detection. *Pattern Recognit.*, 127:108644, 2022. 2