Edge-Aware Camouflaged Object Detection

Patricia L. Suárez[1] and Angel D. Sappa[1,2(✉)]

[1] ESPOL Polytechnic University, Guayaquil, Ecuador
{plsuarez,asappa}@espol.edu.ec
[2] Computer Vision Center, 08193 Bellaterra, Barcelona, Spain
asappa@cvc.uab.es

Abstract. This paper presents a novel edge-aware transformer-based framework for camouflaged object detection. By integrating a spatial attention module guided by structural cues extracted from edge information, the model is directed toward visually ambiguous regions that commonly hinder segmentation performance. This design enables more effective global context modeling and improves the delineation of camouflaged object boundaries. Extensive experiments on multiple benchmark datasets validate the effectiveness of the proposed approach, demonstrating consistent performance gains over state-of-the-art methods in all key evaluation metrics.

Keywords: Edge · Attention mechanisms · Camouflaged Objects · Pyramid Vision Transformer

1 Introduction

Camouflaged object detection (COD) poses a significant challenge in computer vision due to the minimal appearance differences between foreground objects and their backgrounds. In such settings, traditional detection strategies often struggle to identify target regions, particularly when objects are intentionally embedded within complex visual contexts. Addressing this problem requires models capable of capturing both high-level semantic features and low-level structural cues.

Recent advances in vision transformers have demonstrated strong potential for dense prediction tasks, thanks to their ability to model long-range dependencies and contextual relationships across the image [6]. Additionally, attention mechanisms enhance feature selectivity by enabling the model to prioritize task-relevant regions [26], making these components particularly well-suited for COD, where visual signals are often sparse and ambiguous.

In scenarios where texture boundaries are weak or indistinct, global semantic features alone may not suffice. Structural information, such as edges, plays a critical role in enhancing the model's sensitivity to object contours. Prior work highlights the benefit of integrating such information into the learning process to improve boundary precision [22].

M. Castrillón-Santana et al. (Eds.): CAIP 2025, LNCS 15621, pp. 197–208, 2026.
https://doi.org/10.1007/978-3-032-04968-1_17

This work proposes a transformer-based framework for camouflaged object detection, enhanced with a spatial attention mechanism that integrates structural cues from edge information. Building upon the OAFormer backbone, the method introduces key innovations such as multistage edge injection, edge-guided attention, and an edge-aware loss function, which collectively improve boundary sensitivity and detection accuracy in challenging scenarios.

The method is evaluated on four public COD benchmarks, showing consistent improvements in detection quality. These results demonstrate the effectiveness of combining edge-guided attention and structural supervision within a transformer-based architecture. The main contributions of this paper are:

- A transformer-based architecture is enhanced with a spatial attention mechanism guided by edge-derived structural cues.
- An edge-aware loss function is proposed to improve the consistency between predicted masks and the underlying structural layout of the image.

The remainder of this paper is structured as follows, Sect. 2 reviews relevant literature. Section 3 presents the proposed methodology, including the architecture and fusion mechanism. Section 4 discusses results and evaluations, and Sect. 5 concludes the paper with future directions.

2 Related Work

Camouflaged object detection has been extensively studied, and numerous approaches have been proposed to address the challenges posed by the high similarity between foreground objects and background environments. Unlike standard object detection, COD has required models to capture both global context, to differentiate between subtle environmental variations, and local edge cues, to delineate objects from their surroundings. Early COD methods have been based primarily on convolutional neural networks [5], but recent advances have shown that transformer-based architectures can achieve superior performance by exploiting self-attention to model long-range dependencies across the image (e.g., [30,35]).

On the other hand, to enhance the capability of camouflaged object detection in visually ambiguous scenes, structure-aware methods have been developed to enhance spatial discrimination and improve the localization of subtle visual cues (e.g., [3,13]). To further advance contextual understanding, hierarchical graph interaction transformers have been introduced. These models leverage dynamic token clustering to strengthen the detection process and have demonstrated increased robustness in complex environments [33]. In parallel, adaptive guidance learning has been proposed to tailor the detection pipeline to the specific characteristics of camouflaged targets, offering adaptive supervision that improves segmentation accuracy [4]. Additionally, zero-shot learning techniques have been explored to reduce reliance on large-scale annotated datasets, enhancing the scalability and applicability of camouflaged object detection systems [11].

The field has additionally evolved to tackle complex scenarios involving multi-object camouflage and overlapping instances. Multi-object detection pipelines have been proposed that group relevant tokens and refine them using edge-based representations [27]. Collaborative models such as CoCOD [14] have been designed to utilize multi-image contexts for improved inference in challenging environments. These efforts have addressed issues related to boundary ambiguity, occlusion, and limited data availability. Furthermore, minimal-supervision and zero-shot approaches have also been explored, such as ZeroScope [12], which has demonstrated effective detection in data-scarce conditions by leveraging generalized feature representations.

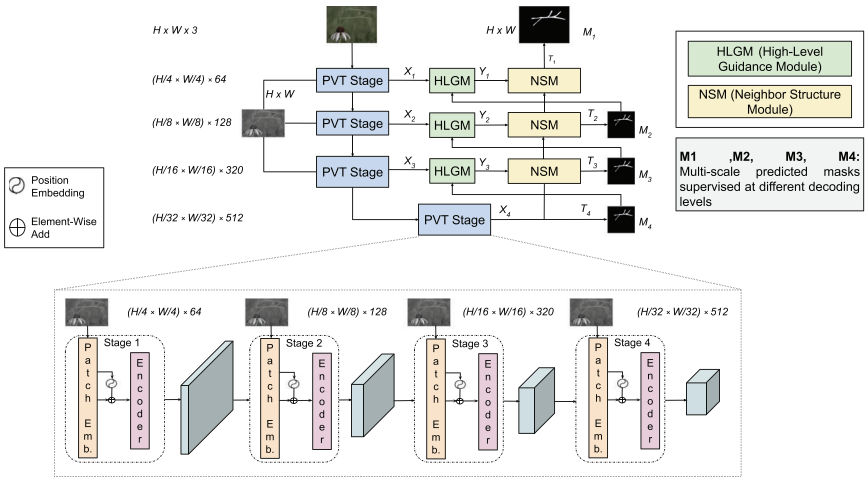

Fig. 1. Overview of the proposed architecture. A four-stage PVT backbone extracts multi-scale features, refined by HLGM and NSM modules to produce predictions at different levels.

Beyond architectural advances, attention mechanisms have been widely adopted in image segmentation to enable models to concentrate on the most relevant features [2], while edge detection has served to improve the localization of object boundaries [22].

3 Proposed Approach

An edge-aware transformer design is introduced to enhance the model's ability to capture spatial structure. This component builds upon the baseline architecture by integrating structural edge information to guide attention in visually ambiguous regions. This design improves the model's ability to capture fine-grained variations between camouflaged objects and their surroundings, thereby enabling more accurate and precise segmentation.

3.1 Edge-Based Attention Mechanism

This paper proposes to integrate edge detection into the model to supply structural information about object edges. The Canny edge detector [1] has been applied to the RGB input images to extract edge maps, which are subsequently normalized and used to implement the attention map. This configuration has allowed the network to better localize object contours that are otherwise difficult to distinguish in camouflaged environments. The attention map A has been computed through a convolutional operation over the edge map E followed by a sigmoid activation:

$$A = \sigma\left(\mathrm{Conv}(E)\right), \tag{1}$$

where $\sigma(\cdot)$ denotes the sigmoid function. The feature map X from the RGB image has been modulated as follows:

$$X_{\mathrm{att}} = X \odot (1 + A), \tag{2}$$

where $\odot$ denotes element-wise multiplication. The term X_{att} represents the spatially enhanced feature map where structural information from the edge map A reinforces the original feature X.

3.2 Model Architecture

The proposed approach extends the OAFormer framework [32], originally developed for camouflaged object detection, by incorporating an edge-guided attention mechanism into its architecture (see Fig. 1). In this design, the PVT backbone is composed of four hierarchical stages, each consisting of a patch embedding layer and a transformer encoder. At each stage, edge-salience maps are injected before the encoder to enrich spatial encoding. The outputs are progressively refined through HLGM and NSM modules. This guidance is integrated into the transformer encoder alongside the RGB input, allowing the model to capture both contextual semantics and spatial patterns. The resulting features are subsequently refined through convolutional layers to produce the final segmentation output.

3.3 Loss Function

The original OAFormer loss [32] is adopted and extended to improve structural accuracy. It combines three components: weighted binary cross-entropy $\mathcal{L}^{\omega}_{\mathrm{bce}}$, weighted IoU $\mathcal{L}^{\omega}_{\mathrm{iou}}$ [29], and uncertainty-aware loss $\mathcal{L}_{\mathrm{ual}}$ [8]. These are computed at four decoder stages M_i, with outputs resized to the input resolution. The OAFormer base loss is defined as:

$$\mathcal{L}_{\mathrm{OAFormer}} = \sum_{i=1}^{4} \frac{1}{2^{i-1}} \left(\mathcal{L}^{\omega}_{\mathrm{bce}}(M_i, G) + \mathcal{L}^{\omega}_{\mathrm{iou}}(M_i, G) + \mathcal{L}_{\mathrm{ual}}(M_i)\right), \tag{3}$$

where G denotes the ground-truth mask, and M_i represents the predicted mask produced at the i-th decoding stage. To enhance contour alignment, an additional edge loss is introduced:

$$\mathcal{L}_{\text{edge}} = \frac{1}{N} \sum_{i,j} |\nabla PM_{ij} - E_{ij}|, \tag{4}$$

where PM is the final predicted mask, ∇PM its gradient, E edge map, and N the number of pixels.

The final training objective is formulated as:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{OAFormer}} + \lambda_{\text{edge}} \cdot \mathcal{L}_{\text{edge}}, \tag{5}$$

the weight $\lambda_{\text{edge}} = 0.3$ has been empirically selected to balance the impact of each component, enhancing edge quality without disrupting the original training dynamics of OAFormer.

4 Experimental Results

This section presents the experimental results obtained with the proposed architecture. The model's object detection performance is evaluated and compared with state-of-the-art approaches.

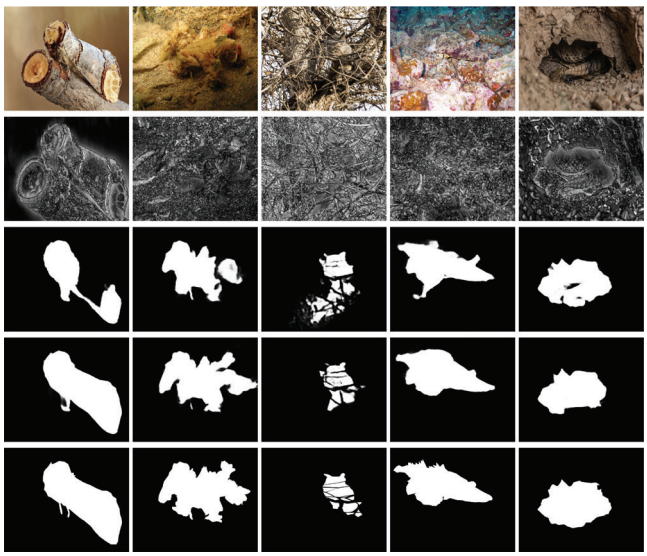

Fig. 2. Results from COD10K dataset: (*1st. row*) Input RGB images; (*2nd. row*) Edge salience images; (*3rd. row*) Camouflaged masks detected by [32]; (*4th. row*) Camouflaged masks from the proposed approach; (*5th. row*) Ground truth masks.

4.1 Datasets

The model has been trained and evaluated using four benchmark datasets for camouflaged object detection: COD10K [8], CAMO [10], CHAMELEON [21], and NC4K [16]. Images from COD10K and CAMO datasets have been jointly considered for training and validation, while CHAMELEON and NC4K were used just for evaluating the generalization capability of the proposed framework. COD10K contains 10,000 high-resolution images, of which 6,066 have been used for training and validation; and 2,026 images have been used for testing. CAMO comprises 1,250 images, with 1,000 used for training and validation; and 250 images have been used for testing. In other words, the model has been trained and validated with 7,066 images and tested with 2,276 images. On the other hand, CHAMELEON includes 76 challenging images and NC4K consists of 4,121 images, collected from various sources. The model was trained and evaluated using two NVIDIA A100 GPUs, each with 24 GB of VRAM.

4.2 Pre-proccesing

To ensure consistency in preprocessing across all datasets, each image and its corresponding ground-truth mask has been resized to a fixed resolution of 416 $\times$ 416 pixels before being input into the model. To enhance robustness and generalization during training, several data augmentation techniques have been applied. These have included random horizontal flipping, random cropping, and resizing, while color jittering has been introduced in selected experiments to simulate lighting variations.

4.3 Results and Comparisons

Results from the proposed approach have been compared with respect to several state-of-the-art models using four widely adopted metrics: Mean Absolute Error (MAE, $\downarrow$), Enhanced-alignment Measure (E_m, $\uparrow$), Structure Measure (S_m, $\uparrow$), and F-measure (F_β^w, $\uparrow$). In the results table, "$\uparrow$"/"$\downarrow$" means that larger/smaller is better. As shown in Table 1, the approach demonstrates superior performance on the CAMO and COD10K datasets, which are used for both training and validation. In addition, Table 2 presents results on the CHAMELEON and NC4K datasets, which were used exclusively for evaluating the generalization capability of the proposed approach. It can be appreciated that the model also achieves consistent improvements across all metrics—in all the metrics it obtains the best result.

To contextualize the evaluated models, Table 3 summarizes the publication venue, year, and number of parameters for each approach considered in the comparison. This information provides a reference for understanding differences in model complexity and development timelines.

Complementing the quantitative results, qualitative comparisons are illustrated in Figs. 2, 3, 4 and 5. The masks generated by the proposed model exhibit

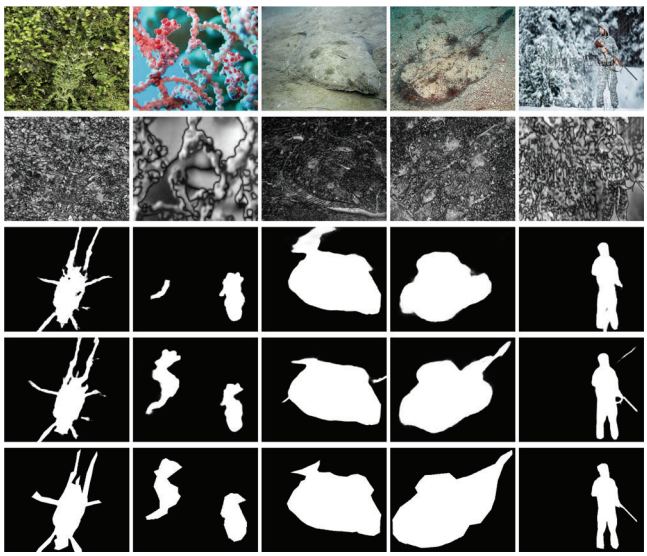

Fig. 3. Results from CAMO dataset: (*1st. row*) Input RGB images; (*2nd. row*) Edge salience images; (*3rd. row*) Camouflaged mask detected by [32]; (*4th. row*) Camouflaged mask from the proposed approach; (*5th. row*) Ground truth masks.

Fig. 4. Results from CHAMELEON dataset: (*1st. row*) Input RGB images; (*2nd. row*) Edge salience images; (*3rd. row*) Camouflaged masks detected by [32]; (*4th. row*) Camouflaged masks from proposed approach; (*5th. row*) Ground truth masks.

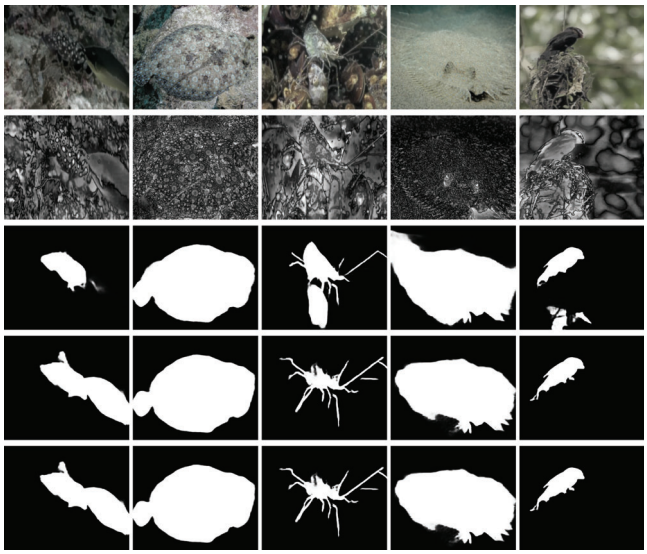

Fig. 5. Results from NC4K dataset: (*1st. row*) Input RGB images; (*2nd. row*) Edge salience images; (*3rd. row*) Camouflaged masks detected by [32]; (*4th. row*) Camouflaged masks from the proposed approach; (*5th. row*) Ground truth masks.

Table 1. Comparison of the proposed method with state-of-the-art methods on CAMO and COD10K datasets. "↑"/"↓" means that larger/smaller is better (best values in bold black; second-best values in bold blue).

Method	CAMO [10]				COD10K [8]			
	$S_m \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$MAE \downarrow$	$S_m \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$MAE \downarrow$
EGNet [36]	0.732	0.7996	0.6036	0.1095	0.7365	0.8097	0.5174	0.0607
PraNet [9]	0.7692	0.8245	0.6625	0.0942	0.7894	0.8606	0.6294	0.0451
F2Net [28]	0.7113	0.7407	0.5636	0.1087	0.7386	0.7951	0.5438	0.0513
MINet [19]	0.748	0.7911	0.637	0.0903	0.7697	0.832	0.6085	0.0417
SINet [8]	0.7454	0.8035	0.6443	0.0915	0.7794	0.8642	0.631	0.0426
C2FNet [24]	0.7961	0.8537	0.7187	0.0799	0.813	0.8902	0.6862	0.036
PFNet [17]	0.7823	0.841	0.6952	0.0849	0.7998	0.8772	0.6599	0.0396
MGL [34]	0.7755	0.816	0.6728	0.0884	0.8139	0.8513	0.66566	0.0353
UGTR [31]	0.7839	0.8215	0.6836	0.0863	0.8171	0.8519	0.6656	0.0356
SINet [7]	0.8201	0.8817	0.7426	0.0705	0.8151	0.887	0.6796	0.0368
BSANet [20]	0.7943	0.8511	0.7174	0.0786	0.8176	0.8905	0.699	0.0342
OCENet [15]	0.8019	0.8518	0.7234	0.0804	0.8272	0.8935	0.7071	0.0327
BGNet [25]	0.8116	0.8698	0.7485	0.0734	0.8307	0.9007	0.7219	0.0326
ZoomNet [18]	0.8197	0.877	0.7522	0.0659	0.8384	0.8876	0.7288	0.0289
OAFormer [32]	0.8659	0.9238	0.8263	0.048	0.8599	0.9274	0.7732	0.0245
FSENet [23]	**0.885**	**0.942**	**0.85**	**0.04**	**0.873**	**0.928**	**0.8**	**0.021**
Our Approach	**0.8912**	**0.9574**	**0.8699**	**0.042**	**0.8674**	**0.9324**	**0.7798**	**0.0221**

†FSENet omits CHAMELEON dataset, limiting fair comparison and generalization assessment.

Table 2. Comparison of the proposed method with state-of-the-art methods on CHAMELEON and NC4K datasets. "↑"/"↓" means that larger/smaller is better (best values in bold black; second-best values in bold blue).

Method	CHAMELEON [21]				NC4K [16]			
	$S_m \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$MAE \downarrow$	$S_m \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$MAE \downarrow$
EGNet [36]	0.7975	0.8599	0.6486	0.0648	0.7771	0.8408	0.6386	0.0751
PraNet [9]	0.86	0.9073	0.7633	0.0437	0.8222	0.8761	0.7245	0.0588
F^3Net [28]	0.848	0.8943	0.7436	0.0467	0.78	0.8244	0.6556	0.0695
MINet [19]	0.8548	0.9139	0.7713	0.0358	0.8122	0.8618	0.7195	0.0555
SINet [8]	0.872	0.9363	0.8056	0.0341	0.808	0.8173	0.7227	0.0576
C^2FNet [24]	0.8881	0.9353	0.8284	0.0316	0.8383	0.8974	0.7624	0.049
PFNet [17]	0.8819	0.9306	0.8099	0.0325	0.829	0.8874	0.7453	0.0527
MGL [34]	0.8932	0.9171	0.8123	0.0305	0.8326	0.8666	0.7392	0.0526
UGTR [31]	0.8857	0.9097	0.7939	0.0314	0.8394	0.8744	0.7463	0.0519
SINet-v2 [7]	0.8882	0.9417	0.816	0.0297	0.8472	0.9027	0.7698	0.0476
BSANet [20]	0.8954	0.9458	0.841	0.0272	0.8414	0.8968	0.7708	0.0479
OCENet [15]	0.8972	0.9402	0.8333	0.0269	0.8533	0.9025	0.7846	0.045
BGNet [25]	0.9012	0.943	0.8505	0.0268	0.851	0.9067	0.7884	0.0444
ZoomNet [18]	0.9017	0.9427	0.8451	0.0229	0.8528	0.8957	0.7844	0.0434
OAFormer [32]	**0.9035**	**0.9608**	**0.8583**	**0.0227**	0.8833	0.9343	0.8372	0.0325
FSENet† [23]	-	-	-	-	**0.892**	**0.941**	**0.853**	**0.03**
Our Approach	**0.9067**	**0.9714**	**0.8620**	**0.0223**	**0.8986**	**0.9476**	**0.8418**	**0.0297**

†FSENet is not evaluated on CHAMELEON dataset, limiting generalization assessment.

Table 3. Model size and venue details by approach.

Approach	Venue'Year	Parameters (M)
EGNet [36]	ICCV'2019	56.02
PraNet [9]	MICCAI'2020	32.55
F^3Net [28]	AAAI'2020	25.54
MINet [19]	CVPR'2020	162.38
SINet [8]	CVPR'2020	48.95
C^2FNet [24]	IJCAI'2021	28.41
PFNet [17]	CVPR'2021	46.5
MGL [34]	ICCV'2021	63.6
UGTR [31]	ICCV'2021	48.87
SINet-v2 [7]	TPAMI'2022	26.98
BSANet [20]	AAAI'2022	32.58
OCENet [15]	WACV'2022	58.17
BGNet [25]	IJCAI'2022	79.85
ZoomNet [18]	CVPR'2022	32.38
OAFormer [32]	ICME'2023	39.86
Our Approach		39.86

sharper contours and more complete object coverage than those of other evaluated approaches. These visual results confirm that the integration of edge-guided attention enables the model to distinguish camouflaged objects more effectively from their background, even in scenarios with minimal visual contrast.

5 Conclusions

This work presents a transformer-based camouflaged object detection framework enhanced with edge-guided attention, extending the OAFormer backbone to explicitly model structural cues. The integration of edge features and an edge-aware loss function enables more accurate contour localization and improves mask consistency. Extensive experiments across four benchmark datasets confirm consistent improvements over state-of-the-art methods in both quantitative and qualitative evaluations. Future research will explore the incorporation of additional domain modalities, such as depth or thermal imagery, to further improve segmentation under challenging camouflage scenarios.

Acknowledgements. This work was supported in part by the ESPOL project CIDIS-003-2024-T, in part by Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/5011000 11033 and by "ERDF A way of making Europe", in part by the Air Force Office of Scientific Research Under Award FA9550-24-1-0206. The second author acknowledges the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021 SGR01499.

References

1. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
2. Chen, L., Wang, M., Zhang, K.: Attention-based models for image segmentation: a survey. IEEE Trans. Neural Networks Learn. Syst. **32**(6), 1234–1245 (2021)
3. Chen, M.: Non-local attention mechanisms for COD. CVPR (2023)
4. Chen, Z., Zhang, X., Xiang, T.Z., Tai, Y.: Adaptive guidance learning for camouflaged object detection. preprint arXiv:2401.12345 (2024)
5. Cui, L.: Contrastive learning for camouflaged object detection. TPAMI (2023)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2021)
7. Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. IEEE Trans. Pattern Anal. Mach. Intell. **44**(10), 6024–6042 (2021)
8. 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)
9. Fan, D.P., et al.: Pranet: parallel reverse attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 263–273. Springer (2020)
10. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A.: Anabranch network for camouflaged object segmentation. Comput. Vis. Image Underst. **184**, 45–56 (2019)
11. Lei, C., Fan, J., Li, X., Xiang, T., Li, A., Zhu, C., Zhang, L.: Towards real zero-shot camouflaged object segmentation without camouflaged annotations. arXiv preprint arXiv:2402.12345 (2024)
12. Lei, X.: Zeroscope: Minimal supervision techniques for cod tasks. CVPR (2024)

13. Li, X.: Edge-resnet: Boundary-aware networks for camouflaged object detection. ECCV (2022)
14. Liu, F.e.a.: Collaborative camouflaged object detection: A large-scale dataset and benchmark. Preprint arXiv:2205.11333 (2022)
15. Liu, J., Zhang, J., Barnes, N.: Modeling aleatoric uncertainty for camouflaged object detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1445–1454 (2022)
16. Lv, Y., et al.: 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)
17. 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)
18. 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)
19. Pang, Y., Zhao, X., Zhang, L., Lu, H.: Multi-scale interactive network for salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9413–9422 (2020)
20. 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)
21. Skurowski, P., Abdulameer, H., Błaszczyk, J., Depta, T., Kornacki, A., Kozieł, P.: Animal camouflage analysis: chameleon database. https://www.polsl.pl/rau6/chameleon-database-animal-camouflage-analysis/ (2018)
22. Smith, J., Doe, J.: Edge detection techniques: a comprehensive review. J. Comput. Vision **45**(3), 567–589 (2021)
23. Sun, Y., Xu, C., Yang, J., Xuan, H., Luo, L.: Frequency-spatial entanglement learning for camouflaged object detection. In: European Conference on Computer Vision, pp. 343–360. Springer (2024)
24. Sun, Y., Chen, G., Zhou, T., Zhang, Y., Liu, N.: Context-aware cross-level fusion network for camouflaged object detection. arxiv 2021. preprint arXiv:2105.12555 (2021)
25. Sun, Y., Wang, S., Chen, C., Xiang, T.Z.: Boundary-guided camouflaged object detection. arXiv preprint arXiv:2207.00794 (2022)
26. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. In: Advances in Neural Information Processing System, vol. 30, pp. 5998–6008 (2017)
27. Wang, H.: Multi-scale transformer networks for camouflaged object detection. ICCV (2021)
28. Wei, J., Wang, S.: F^3net: Fusion, feedback and focus for salient object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12321–12328 (2020). https://doi.org/10.1609/aaai.v34i07.6916
29. Wei, J., Wang, S., Huang, Q.: F^3net: fusion, feedback and focus for salient object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12321–12328 (2020)
30. Wu, J.: Lightweight models for camouflaged object detection. NeurIPS (2022)
31. Yang, F., et al.: Uncertainty-guided transformer reasoning for camouflaged object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4146–4155 (2021)

32. Yang, X., Zhu, H., Mao, G., Xing, S.: Oaformer: occlusion aware transformer for camouflaged object detection. In: 2023 IEEE International Conference on Multimedia and Expo (ICME), pp. 1421–1426 (2023)
33. Yao, S., Sun, H., Xiang, T.Z., Wang, X., Cao, X.: Hierarchical graph interaction transformer with dynamic token clustering for camouflaged object detection. IEEE Trans. Image Process. **33**, 1234–1245 (2024)
34. 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)
35. Zhang, J.: Transformer-based object detection. CVPR (2020)
36. Zhao, J.X., Liu, J.J., Fan, D.P., Cao, Y., Yang, J., Cheng, M.M.: EGNET: edge guidance network for salient object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8779–8788 (2019)