SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

January 2025, Vol. 68, Iss. 1, 112104:1–112104:15
https://doi.org/10.1007/s11432-024-4233-9

COMPrompter: reconceptualized segment anything model with multiprompt network for camouflaged object detection

Xiaoqin ZHANG[1], Zhenni YU[1], Li ZHAO[1], Deng-Ping FAN[2,3] & Guobao XIAO[4*]

[1]*Zhejiang Province Key Laboratory of Intelligent Informatics for Safety and Emergency, Wenzhou University, Wenzhou 325035, China;*
[2]*Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Shenzhen 518045, China;*
[3]*College of Computer Science, Nankai University, Tianjin 300071, China;*
[4]*School of Computer Science and Technology, Tongji University, Shanghai 201804, China*

Received 28 January 2024/Revised 25 July 2024/Accepted 3 September 2024/Published online 17 December 2024

Abstract We rethink the segment anything model (SAM) and propose a novel multiprompt network called COMPrompter for camouflaged object detection (COD). SAM has zero-shot generalization ability beyond other models and can provide an ideal framework for COD. Our network aims to enhance the single prompt strategy in SAM to a multiprompt strategy. To achieve this, we propose an edge gradient extraction module, which generates a mask containing gradient information regarding the boundaries of camouflaged objects. This gradient mask is then used as a novel boundary prompt, enhancing the segmentation process. Thereafter, we design a box-boundary mutual guidance module, which fosters more precise and comprehensive feature extraction via mutual guidance between a boundary prompt and a box prompt. This collaboration enhances the model's ability to accurately detect camouflaged objects. Moreover, we employ the discrete wavelet transform to extract high-frequency features from image embeddings. The high-frequency features serve as a supplementary component to the multiprompt system. Finally, our COMPrompter guides the network to achieve enhanced segmentation results, thereby advancing the development of SAM in terms of COD. Experimental results across COD benchmarks demonstrate that COMPrompter achieves a cutting-edge performance, surpassing the current leading model by an average positive metric of 2.2% in COD10K. In the specific application of COD, the experimental results in polyp segmentation show that our model is superior to top-tier methods as well. The code will be made available at https://github.com/guobaoxiao/COMPrompter.

Keywords segment anything model, camouflaged object detection, boundary, prompt

Citation Zhang X Q, Yu Z N, Zhao L, et al. COMPrompter: reconceptualized segment anything model with multiprompt network for camouflaged object detection. Sci China Inf Sci, 2025, 68(1): 112104, https://doi.org/10.1007/s11432-024-4233-9

1 Introduction

Camouflaged object detection (COD) [1] has been extensively studied as a subset of image segmentation tasks. It finds various applications in medical image segmentation [2], nature conservation and wildlife research [3], and search and rescue missions.

In the field of COD, diverse methods are focused on essential information sources, including context (e.g., MSCAF-Net [4], C2FNet [5]), edge (e.g., JCSOD [6], R-MGL [7], TINet [8]), and gradient (e.g., DGNet [9]). Other methods employ a range of effective strategies such as amplification (e.g., ZoomNet [10], ZoomNeXt [11]), humans attention (e.g., SegMaR [12]), predation (e.g., SINetV2 [13], LSR [14], PFNet [15], PraNet [2], SINet [16]), and uncertainty (e.g., UGTR [17], UCNet [18]). In the recent studies, segment anything model (SAM)-Adapter [19] and MedSAM [20] leveraged SAM [21] to perform COD. SAM excels in segmentation across various scenarios, including camouflage, because of its robust zero-shot generalization ability. SAM accomplishes COD with preliminary segmentation results at minimal computational cost. This capability enables researchers to develop more customized approaches for the unique characteristics of camouflaged targets. However, despite these strides, the existing methods often ignore the constraints associated with a single prompt. More critically, these methods fail to explore alternative prompt types other than those provided by SAM. In the context of COD, a noticeable disparity persists between SAM-based methods and the current state-of-the-art (SOTA) methods.

* Corresponding author (email: x-gb@163.com)

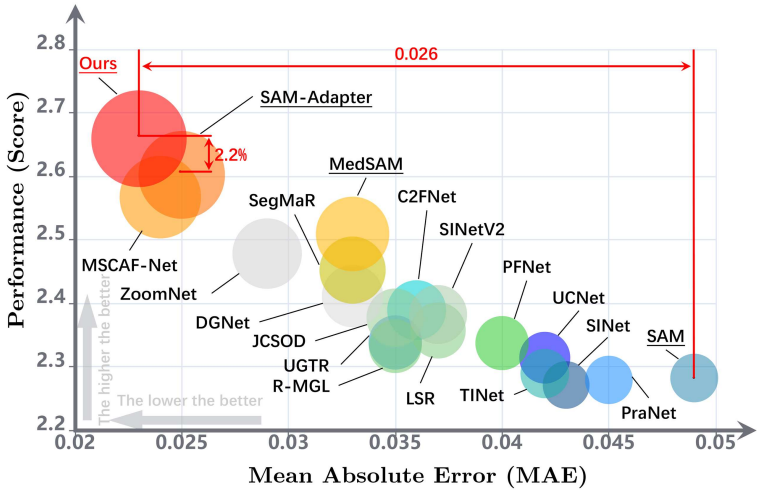

Figure 1 (Color online) Scatter plot representing the performance of competitors and our model on COD10K-Test. F_β^ω, S_α, and E_ϕ are positive-oriented, while M is negative-oriented. The order of magnitude of M and the other indices are different. For a more effective comparison, we take M as the X-axis, and the sum of the other three indicators as the Y-axis. The underline represents the segment anything model (SAM)-based method. Score $= F_\beta^\omega + S_\alpha + E_\phi$.

This study proposes COMPrompter, a multiprompt network for COD. COMPrompter leverages the strengths of SAM and expands the utility thereof to the COD domain. Deviating from the direct use of SAM, COMPrompter introduces a multiprompt strategy. This strategy integrates both the original box prompt of SAM and the boundary prompt. In the boundary prompt, we particularly emphasize edges and gradients because of their extensive exploration. Identification of edges is more straightforward than recognition of an entire camouflage target, and gradients offer a fresh perspective on segmentation. Still, both approaches suffer their challenges: designing an edge module considerably increases computational overhead, and the targeted nature of internal gradients for various objects also poses limitations. Similar challenges exist in the specific application domain of COD, such as polyp segmentation.

Interestingly, we do not perform edge prediction in response to these challenges. Instead, we integrate edge information with gradients into the network via prompts, resulting in increased accuracy. Our proposed boundary prompt prevents the abovementioned issues. In addition, in conjunction with a box prompt, it provides a more accurate prior for COD. Specifically, the boundary prompt is derived from our proposed edge gradient extraction module (EGEM). EGEM employs dilation and canny operations on ground truth (GT) and image, respectively. Thereafter, EGEM obtains edge masks containing gradient instead of the entire camouflage target. Acquiring the gradient-enhanced boundary operation is straightforward yet innovative because prior research has not emphasized the gradient at the edge. To effectively guide segmentation using both the box and boundary prompts, we introduce the box-boundary mutual guidance module (BBMG). BBMG strengthens the connection between the dense box embedding and dense boundary embedding via adapted pointwise convolution of depthwise separable convolution [22] and residual connection. In addition, inspired by He et al. [23], we incorporate the discrete wavelet transform (DWT) to obtain high-frequency signals. These signals represent details or rapidly changing parts of the image.

Finally, through a judicious combination of EGEM, BBMG, and introduced DWT, we present COMPrompter for COD, a SAM variant tailored for COD. In COMPrompter, we adjust the SAM structure to accommodate our multiprompt strategy. Explicitly, we integrate a prompt encoder into SAM to address the proposed boundary prompt. Experimental results on four benchmark datasets substantiate the superiority of our method to all other SOTA methods, as shown in Figure 1. Figure 2 shows the complete network structure.

Our contributions are summarized as follows:

• We propose a multiprompt network called COMPrompter for COD, a structural variant of SAM. Precisely, we propose a multiprompt strategy, in which the original box prompt of SAM and the newly designed boundary prompt are used as the user prompt of COMPrompter.

• We propose two efficient designs in COMPrompter: EGEM and BBMG. EGEM obtains the gradient mask of the boundary from the image and GT. A box prompt and boundary prompt guide and

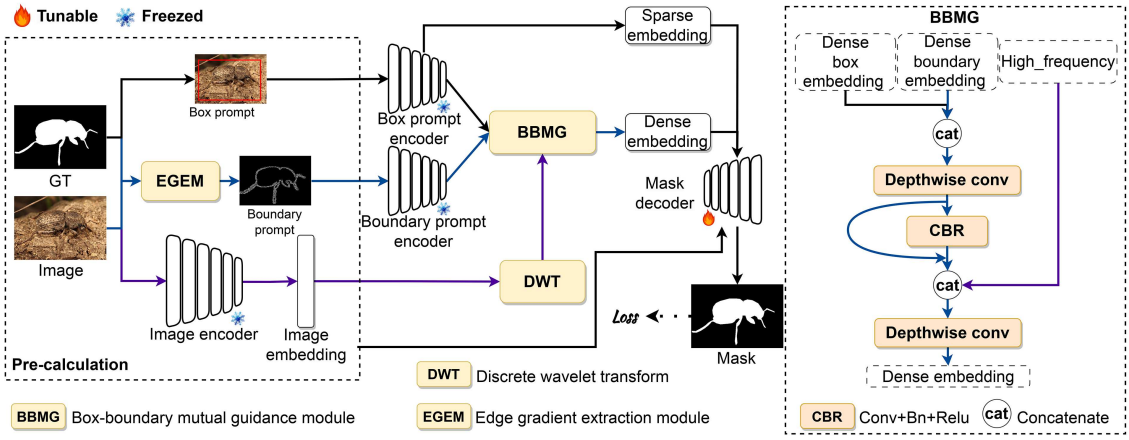

Figure 2 (Color online) Pipeline of our COMPrompter framework (left) and details of the box-boundary mutual guidance module (BBMG) (right). Regarding the modules in SAM, the parameters in the module with a snowflake are fixed, while whose in the module with a spark can be optimized via training. Purple arrows represent image processing, while blue ones represent the processing of boundary prompts. The dashed arrow represents loss calculation. To decrease the amount of computation, we have calculated in advance the part in the left dashed box.

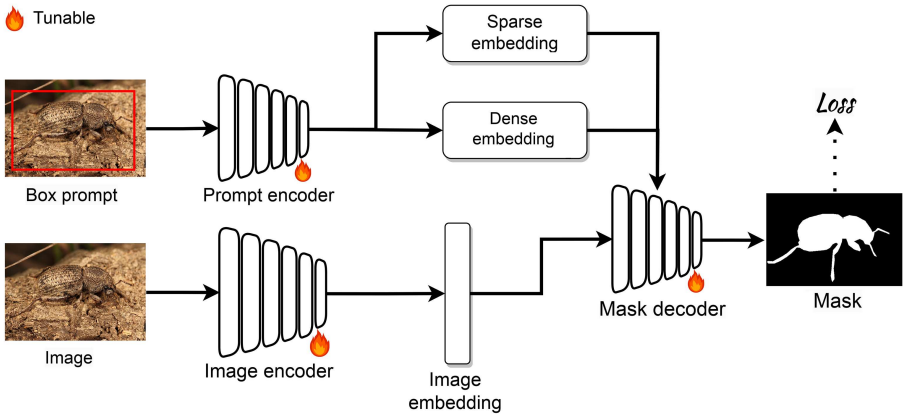

Figure 3 (Color online) Overview of SAM [21] with box prompt.

complement each other via BBMG for accurate prompts.

- We confirm the performance of COMPrompter on COD benchmark datasets. COMPrompter is observed to outperform the existing SOTA methods. We also conduct extensive experiments in polyp segmentation, and conclude that COMPrompter achieves a cutting-edge performance in this domain.

2 Related work

2.1 Segment anything model

As a new foundation model, SAM uses a massive dataset for training and has a remarkable zero-shot ability. As shown in Figure 3, SAM comprises three modules: image encoder, prompt encoder, and mask decoder. The image encoder is large-scale pretrained via masked auto-encoder modeling and has satisfactory feature extraction ability. The prompt encoder encodes the user prompt to obtain sparse embedding and dense embedding. Important for the mask decoder is the design of the self-attention and cross-attention mechanisms. However, SAM often splits objects with the same semantic information into multiple masks [24] because SAM lacks the guidance of semantic information. In a specific field, SAM is unable to accurately segment fine structures because of a lack of professional knowledge or lack of a strong prior [20]. Ma et al. [20] employed a box prompt as a manual prompt, fine-tuned mask decoder, and achieved a satisfactory improvement in segmentation of fine structures. Chen et al. [19] used an adapter to adapt SAM to COD and achieved satisfactory, if not remarkable, results. On account of

the inherent complexity of COD tasks, there remains significant room for improving the application of SAM in the COD domain. We propose a multiprompt strategy that leverages both a box prompt and a novel boundary prompt. The boundary prompt, a concept we propose, captures critical edge gradient information regarding the target object. The multiprompt strategy enhances the precision of prompts, thereby adapting to the difficulty of COD segmentation.

2.2 Camouflaged object detection

Object detection [25,26] is a crucial task in computer vision. It aims to identify specific objects present in images and determine their locations. Its application in videos involves object tracking [27]. The subtask of detection of camouflaged targets is called COD [28,29]. Simultaneous detection of camouflaged objects with similar properties across a set of images is called collaborative COD [30]. Several COD approaches integrate base models. Huang et al. [31] focused on locality modeling and feature aggregation to mitigate the limitations of transformers. Luo et al. [32] employed a diffusion model to generate salient objects in camouflaged scenes for training on multipattern images. Some COD methods give attention to the visual features (color, texture, brightness, etc.) of an object. Compared with these approaches, the strategy of giving attention to boundaries in COD is being increasingly accepted on a wide basis. Zhu et al. [33] designed a boundary guider module to accurately highlight the boundaries of hidden objects. Zhai et al. [7] designed specified modules to enhance the visualization of edges. Ji et al. [34] obtained an initial edge prior via selective edge aggregation. Sun et al. [35] employed excavation and integration of boundary-related edge semantics to increase the efficacy of COD. Lyu et al. [36] decoupled uncertainty reasoning and boundary estimation into two branches: uncertainty and boundary-guided features. These branches were then effectively aggregated to provide accurate segmentation information. Sun et al. [37] proposed EAMNet, comprising an edge detection branch and a segmentation branch. The edge detection branch provided enhanced foreground representations, thereby facilitating the edge detection process. Dong et al. [38], grounded in the unified query-based paradigm, proposed UQFormer, which employed queries to derive boundary cues. Meanwhile, object gradient generation was also applied in COD as an auxiliary task. Ji et al. [9] mined texture information by learning object-level gradients. The application of object-level gradients for obtaining texture information is more deterministic than boundary modeling. It eliminates potential noise due to modeling.

However, a single boundary provides limited information. When providing the gradient of an entire target, the features learned by the network are messy because of the various types of camouflaged targets. Therefore, this study proposes a novel boundary mask with gradient information of the object-background junction. Compared with gradient information of an entire object, gradient information of object edges is easier to learn.

3 Methodology

In this section, we present the details of multiprompt network (COMPrompter). First, we describe the overall architecture of COMPrompter, as shown in Figure 2. Thereafter, we explain the core of this study, i.e., boundary prompt. Last, we discuss the necessity of introducing DWT.

3.1 Overall architecture

The feature extraction part of COMPrompter can be categorized into two parts: the prompt encoder part and the image encoder part. Inspired by [20], we freeze the image encoder and prompt encoder and fine-tune the mask decoder. In addition, the image encoder is precomputed and stored as an npz file. The results of the image encoder can be directly read during real training. In this manner, the calculation amount and training threshold of the SAM large model are considerably decreased. We design the boundary prompt branch, which contains EGEM, BBMG, and a parameter frozen prompt encoder. Among these, EGEM is designed to be computed in advance. The resulting image embedding of the image encoder is saved as an npz file, so that it can be directly read later. The boundary prompt combined with the original box prompt is used as the user prompt of COMPrompter. During inference, we use the GT to generate boundaries and boxes as user prompts to simulate scenarios of user interaction. We apply DWT to the image embedding to complement the frequency details of boundary features.

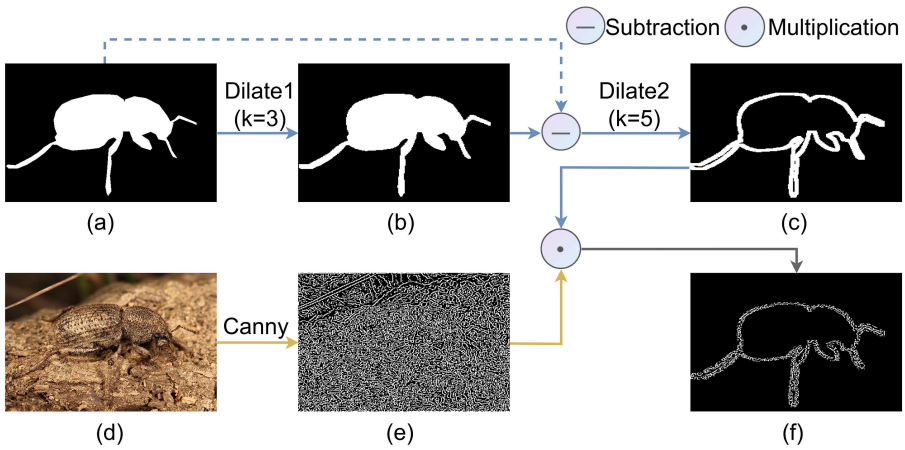

Figure 4 (Color online) EGEM details. The top half of the figure represents the process of object edge extraction. The bottom half represents the process of extracting the gradient of the whole image. Finally the two images are multiplied to obtain the edge map containing the gradient. (a) Ground truth; (b) dilated ground truth; (c) boundary; (d) image; (e) image gradient; (f) boundary gradient.

3.2 Boundary prompt

We design a new boundary prompt branch with main modules comprising EGEM and BBMG. We first obtain the boundary mask with gradient by image, GT, edge detection, and other operations. EGEM takes the boundary mask with gradient as input to the second prompt encoder to obtain the boundary embedding. Boundary prompt compensates for the original box prompt's inability to provide precise boundary information. Specifically, the original box embedding and boundary embedding are fused by BBMG to achieve mutual guidance. Finally a key embedding is obtained as the input of the mask decoder.

3.2.1 *Edge gradient extraction module*

The boundary mask indicates the masks containing the gradient between the object and the background. Because the target is disguised, direct identification of the target is in COD. However, the difficulty of identifying the boundary is relatively easy, so boundary detection is added to COD as an auxiliary task [7]. Gradient information is also widely used as another important information source. Ji et al. [9] mined texture information by learning object-level gradients. However, compared with the gradient of the overall object, the gradient of the boundary undergoes a more rapid change. In addition, these gradients are more representative. Hence, we propose a boundary mask with a gradient. As shown in Figure 4, EGEM employs operations such as dilate and canny to extract boundary masks with gradients from images and GT. The mask is the input of the prompt encoder. Specifically, we first perform the dilation operation of GT with kernel = 3, and the image obtained is subtracted from the GT. In contrast to Premachandran et al. [39], wherein a pixel-wide boundary was employed, we use a dilation operation with a 3×3 kernel to obtain a thicker boundary. This choice is motivated by the necessity for not only the boundary but also its gradient. An excessively narrow boundary might lack sufficient information. Conversely, an excessively wide boundary might extend beyond the desired range and introduce noise. At this point, we obtain the edge image of the camouflaged target in the image. For learning more boundary information, this study performs the dilation operation of GT with kernel = 5 on the obtained edge image to expand the edge of the image. In the prediction, the dilation operation of GT with kernel = 5 also models the case wherein the user prompt may be inaccurate. The resulting boundaries are broader compared to the actual ones. This decreases the difficulty of obtaining boundaries at inference. The expanded edge image is multiplied with the canny image to obtain the final edge image with gradient information, the boundary mask. The boundary mask with gradient BG is generated as follows:

$$\mathrm{BG} = \gamma\big(\mu(\mathrm{GT}) - \mathrm{GT}\big) * C(I), \tag{1}$$

where $\gamma(\cdot)$ and $\mu(\cdot)$ denote dilation operations with kernel = 5 and kernel = 3, respectively, $*$ denotes multiplication, $C(\cdot)$ denotes canny operation, and I denotes original image corresponding to GT.

In addition, we adopt the strategy of precalculation of EGEM like that in the image encoder because of EGEM's independence from subsequent modules, decreasing the amount of calculation during training. Thereafter, we store the calculated images and features in npz files rather than common image formats.

3.2.2 *Box-boundary mutual guidance*

We propose BBMG with the intention of making a box prompt and boundary prompt guide each other and fuse each other. The box prompt is a sparse prompt, while the boundary prompt is a dense prompt. The box prompt indicates the location of the target in the form of four points. The boundary prompt employs a mask to segment the boundary of the target to compensate for the lack of boundary details in the box prompt. Dense box embedding and dense boundary embedding are the results of the box and boundary prompts after the prompt encoder, respectively. As shown in the right-hand side of Figure 2, we use the dense box embedding and dense boundary embedding to perform the join operation on the channel dimension. Thereafter, we apply a residual operation and pass through basic units of convolution, batch normalization, and ReLU (CBR module), defined as CBR(·). The optimized box-boundary embedding (OBB) is generated as follows:

$$\mathrm{EM} = \mathrm{DC}\big(\mathrm{cat}(E_{\mathrm{box}}, E_{\mathrm{boundary}})\big), \tag{2}$$

$$\mathrm{OBB} = \mathrm{cat}\big(\mathrm{CBR}(\mathrm{EM}), EM\big), \tag{3}$$

where E_{box} represents dense box embedding and E_{boundary} dense boundary embedding. cat(·) denotes the join operation on channel dimension. DC(·) represents an adapted pointwise convolution of depthwise separable convolutions [22].

3.3 Discrete wavelet transform

In image processing, DWT can capture features with various frequencies. Among these, the high-frequency features represent edges and subtle changes in the image. Wang et al. [40] extracted the high-frequency features of an image. Liu et al. [41] used a high-frequency component as a prompt to adapt to various downstream tasks.

Inspired by He et al. [23], we apply DWT to the image embedding. DWT focuses on diagonal high-frequency regions in the image. Specifically, it captures rapid changes in signal in the diagonal direction. DWT obtains the diagonal high frequency via the diagonal difference of the original signal. The diagonal high-frequency information (HF) is defined as follows:

$$\mathrm{HF} = x_1 - x_2 - x_3 + x_4, \tag{4}$$

where x_1 and x_2 represent the horizontal and vertical components of the low-frequency signal, respectively, and x_3 and x_4 represent the horizontal and vertical components of the high-frequency signal, respectively. The high frequency extracted via DWT is added to BBMG. The OBB is connected with HF as per dimension. The adapted pointwise convolution of depthwise separable convolutions [22] is then performed. The optimized dense embedding (ODE) is generated as follows:

$$\mathrm{ODE} = \mathrm{DC}\big(\mathrm{cat}(\mathrm{HF}, \mathrm{OBB})\big), \tag{5}$$

where cat(·) denotes the join operation by channel dimension, and DC(·) denotes an adapted pointwise convolution of depthwise separable convolutions [22].

4 Experiments

4.1 Datatset

We conducted experiments on four widely recognized datasets, namely CAMO [42], CHAMELEON [43], COD10K [13], and NC4K [14], to examine the effect of COMPrompter in the task of COD. CAMO comprised 1250 images, randomly split into a training dataset of 1000 images and a test dataset of 250 images. CHAMELEON had 76 images for COD. COD10K had 5066 images, of which 3040 were of a training dataset and 2026 of a test dataset. NC4K was fairly large, with 4121 images. It was used as a test dataset for experiments to examine the generalization ability of COMPrompter. Following Fan et al. [13], we adopted 3040 images of COD10K and 1000 images of CAMO as the training dataset. The remaining

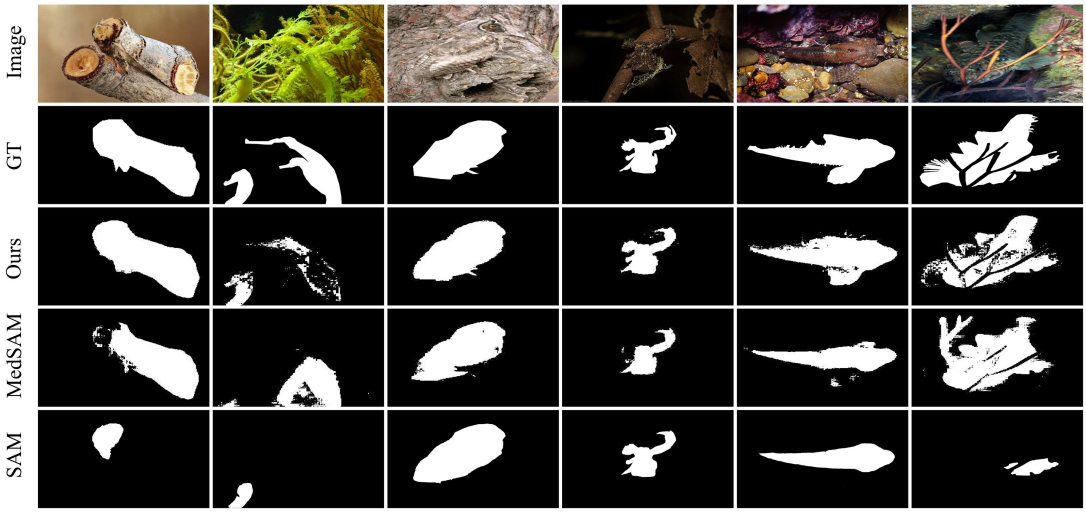

Figure 5 (Color online) Comparison of our COMPrompter and other methods, including MedSAM [20] and SAM [21], in terms of COD. Columns 1–3 are for the CAMO dataset, and Columns 4–6 are for the COD10K dataset.

images of COD10K and CAMO, the entire NC4K dataset, and the entire CHAMELEON dataset were used as the test dataset. In addition, we tested COMPrompter on a more specific application, polyp segmentation, for a more in-depth evaluation. Following Fan et al. [2], we used five public benchmarks datasets, ETIS-Larib [44], CVC-ClinicDB [45], CVC-ColonDB [46], CVC-300 [47], and Kvasir-SEG [48].

4.2 Experimental setup

Implementation details. COMPrompter was implemented using PyTorch, employing the Adam optimizer with a learning rate of $1e^{-5}$. The model underwent 300 epochs to achieve optimal performance. The process was completed in approximately 4.2 h on an NVIDIA 3080TI GPU with a batch size of 32. We scaled all the input images to 1024×1024 via bilinear interpolation, scaling them either up or down. In addition, we truncated and normalized the input image data. This ensured the pixel values were in the appropriate range while maintaining the relative distribution relationship of the data.

Evaluation metrics. We adopted four metrics from COD10K [13] that are widely used and recognized in the field of COD: structure measure (S_α), weighted F-measure (F_β^ω), mean enhanced-alignment measure (E_ϕ), and mean absolute error (M). In the polyp segmentation experiment, we selected the mean dice similarity coefficient (mDice) and mean intersection over union (mIoU). The structure measure quantifies the structural similarity between predicted results and actual segmented regions. The weighted F-measure combines precision and recall, and weights them. The enhanced-alignment measure evaluates prediction results by comparing the alignment relationship between the predicted value and the actual value. The mean absolute error is a quantification of the mean absolute error between the predicted value and the true value.

4.3 Comparisons with cutting-edge methods

We now compare COMPrompter with SAM [21] and other existing COD algorithms, such as UCNet [18], SINet [16], PraNet [2], C2FNet [5], TINet [8], UGTR [17], PFNet [15], R-MGL [7], LSR [14], JCSOD [6], SINetV2 [13], ZoomNet [10], SegMaR [12], DGNet [9], MSCAF-Net [4], SAM-Adapter [19], and MedSAM [20]. This qualitative results are shown in Figures 5–7. To assess the practical usability of COMPrompter, we provided metric data based on the boundaries generated by the pretrained model. These results are shown in the Ours* column of Table 1. The procedure of the boundary generation is shown in Figure 8. First, we obtained the edges using UEDG [36]. Next, we achieved a clearer mask via binarization with a flexible threshold value. This threshold value comprised the pixel value with the highest percentage (the background pixel value, computed from the histogram) plus an offset value of 15. Thereafter, we set the pixels outside the bounding box to zero as per the box prompt and multiplied the result with the gradient map. Finally, we obtained the generated boundary with a gradient. Although MedSAM is used in the field of medical image processing, this method can be used to improve SAM for universal applications. Upon applying the algorithm to the COD task, the metric indexes were considerably improved. Hence,

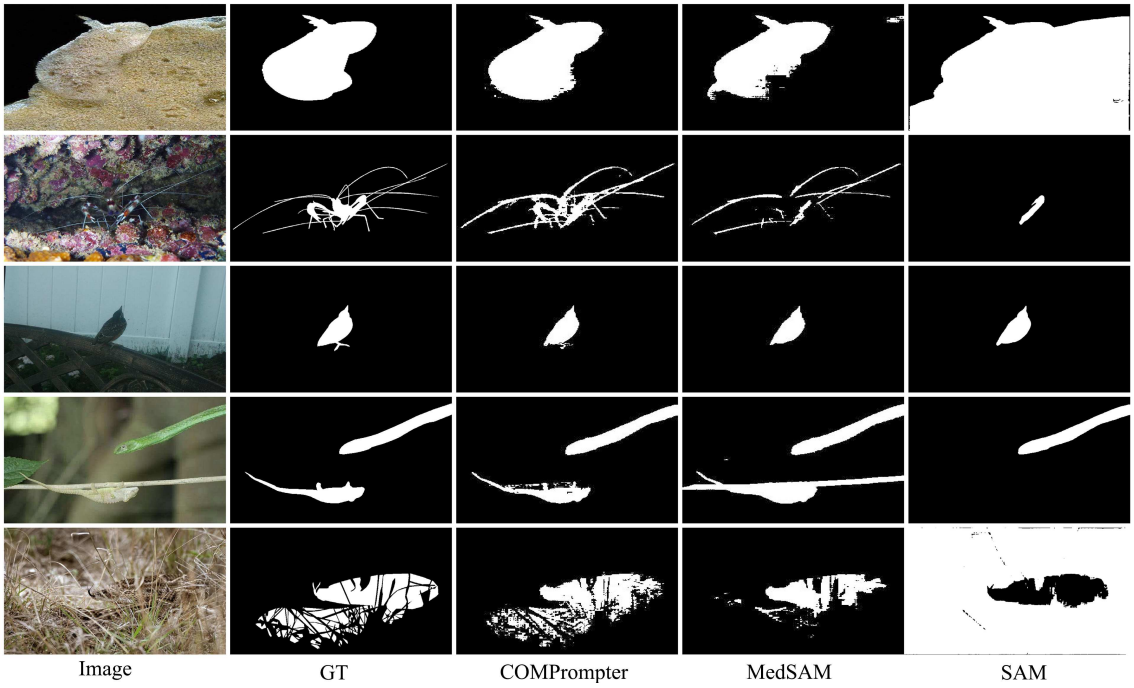

Figure 6 (Color online) Comparison of our COMPrompter with other methods, including MedSAM [20] and SAM [21], in terms of COD. The selected images are from NC4K and contain various shapes, categories, and camouflage methods.

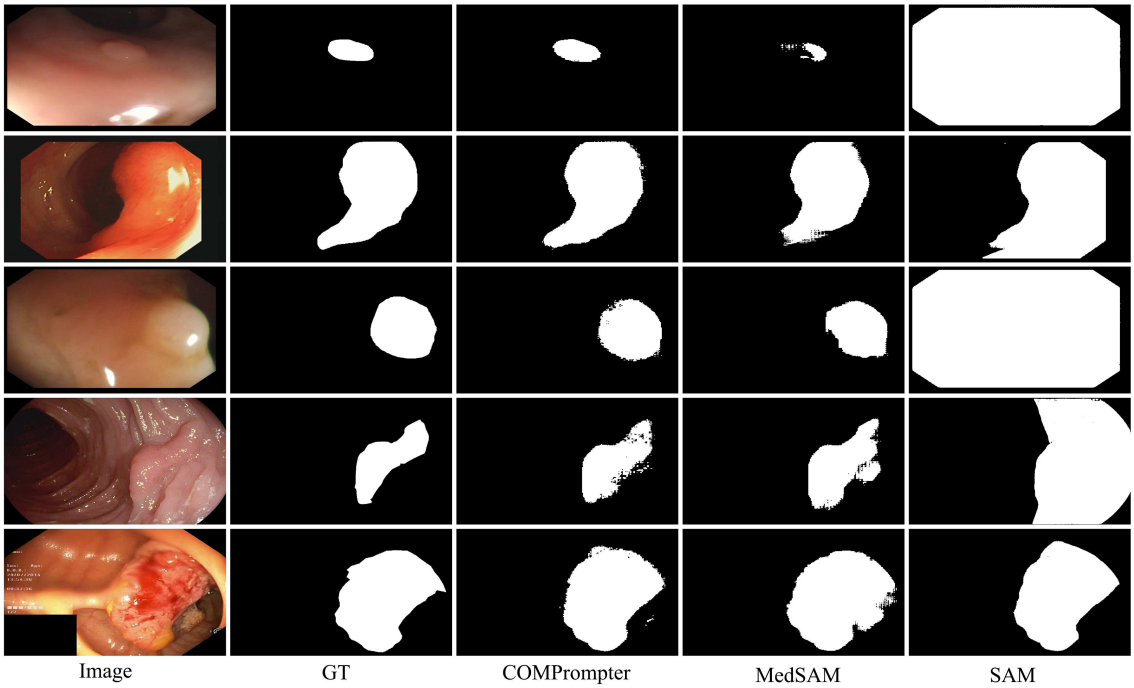

Figure 7 (Color online) Comparison of COMPrompter with other methods, including MedSAM [20] and SAM [21], in polyp datasets. We have provided different examples for a comprehensive comparison.

we listed MedSAM as one of the comparison algorithms. For MedSAM, we performed retraining and validation on the basis of the official code. We also compared COMPrompter with the existing methods vis-à-vis polyp segmentation, such as U-Net [49], UNet++ [50], ResUNet++ [51], SFA [52], PraNet [2], EU-Net [53], SANet [54], LDNet [55], FAPNet [56], SAM [21], and MedSAM [20]. As shown in Tables 1 and 2, our proposed COMPrompter achieved SOTA performance in the COD and polyp segmentation domains. We later present a detailed qualitative and quantitative analysis of the results in both these

Table 1 Quantitative results on four different datasets: CAMO, CHAMELEON, COD10K, and NC4K. The scores in bold represent the best results, while the underlined scores indicate the second and third best results. ↑ indicates that the higher the score the better and ↓ indicates that the lower the score the better.

Dataset	Metric	UCNet 2020 [18]	SINet 2020 [16]	PraNet 2020 [2]	C2FNet 2021 [5]	TINet 2021 [8]	UGTR 2021 [17]	PFNet 2021 [15]	R-MGL 2021 [7]	LSR 2021 [14]	JCSOD 2021 [6]
CAMO	$F_\beta^\omega \uparrow$	0.700	0.644	0.663	0.719	0.678	0.684	0.695	0.673	0.696	0.728
	$S_\alpha \uparrow$	0.739	0.745	0.769	0.796	0.781	0.784	0.782	0.775	0.787	0.800
	$E_\phi \uparrow$	0.787	0.829	0.837	0.864	0.848	0.851	0.842	0.847	0.854	0.873
	$M \downarrow$	0.095	0.092	0.094	0.080	0.087	0.086	0.085	0.088	0.080	0.073
CHAM-ELEON	$F_\beta^\omega \uparrow$	0.836	0.806	0.763	0.828	0.783	0.794	0.810	0.813	0.839	0.848
	$S_\alpha \uparrow$	0.880	0.872	0.860	0.888	0.874	0.888	0.882	0.893	0.893	0.894
	$E_\phi \uparrow$	0.930	0.946	0.907	0.935	0.916	0.940	0.931	0.923	0.938	0.943
	$M \downarrow$	0.036	0.034	0.044	0.032	0.038	0.031	0.033	0.030	0.033	0.030
COD10K	$F_\beta^\omega \uparrow$	0.681	0.631	0.629	0.686	0.635	0.667	0.660	0.666	0.673	0.684
	$S_\alpha \uparrow$	0.776	0.776	0.789	0.813	0.793	0.818	0.800	0.814	0.804	0.809
	$E_\phi \uparrow$	0.857	0.864	0.861	0.890	0.861	0.853	0.877	0.852	0.880	0.884
	$M \downarrow$	0.042	0.043	0.045	0.036	0.042	0.035	0.040	0.035	0.037	0.035
NC4K	$F_\beta^\omega \uparrow$	0.777	0.723	0.724	0.762	0.734	0.747	0.745	0.731	0.766	0.771
	$S_\alpha \uparrow$	0.813	0.808	0.822	0.838	0.829	0.839	0.829	0.833	0.840	0.842
	$E_\phi \uparrow$	0.872	0.871	0.876	0.897	0.879	0.874	0.888	0.867	0.895	0.898
	$M \downarrow$	0.055	0.058	0.059	0.049	0.055	0.052	0.053	0.052	0.048	0.047

Dataset	Metric	SINetV2 2022 [13]	ZoomNet 2022 [10]	SegMaR 2022 [12]	DGNet 2023 [9]	MSCAF-Net 2023 [4]	SAM 2023 [21]	SAM-Adapter 2023 [19]	MedSAM 2023 [20]	Ours* –	Ours –
CAMO	$F_\beta^\omega \uparrow$	0.743	0.752	0.742	0.769	0.828	0.606	0.765	0.779	0.819	**0.858**
	$S_\alpha \uparrow$	0.820	0.820	0.815	0.839	0.873	0.684	0.847	0.820	0.853	**0.882**
	$E_\phi \uparrow$	0.882	0.892	0.872	0.901	0.929	0.687	0.873	0.904	0.919	**0.942**
	$M \downarrow$	0.070	0.066	0.071	0.057	0.046	0.132	0.070	0.065	0.054	**0.044**
CHAM-ELEON	$F_\beta^\omega \uparrow$	0.816	0.845	0.860	0.816	**0.865**	0.639	0.824	0.813	0.830	0.857
	$S_\alpha \uparrow$	0.888	0.902	0.906	0.890	**0.912**	0.727	0.896	0.868	0.884	0.906
	$E_\phi \uparrow$	0.942	**0.958**	0.954	0.934	**0.958**	0.734	0.919	0.936	0.946	0.955
	$M \downarrow$	0.030	0.023	0.025	0.029	**0.022**	0.081	0.033	0.036	0.030	0.026
COD10K	$F_\beta^\omega \uparrow$	0.680	0.729	0.724	0.693	0.775	0.701	0.801	0.751	0.779	**0.821**
	$S_\alpha \uparrow$	0.815	0.838	0.833	0.822	0.865	0.783	0.883	0.841	0.861	**0.889**
	$E_\phi \uparrow$	0.887	0.911	0.895	0.896	0.927	0.798	0.918	0.917	0.933	**0.949**
	$M \downarrow$	0.037	0.029	0.033	0.033	0.024	0.049	0.025	0.033	0.026	**0.023**
NC4K	$F_\beta^\omega \uparrow$	0.770	0.784	0.781	0.784	0.839	0.696	–	0.821	0.840	**0.876**
	$S_\alpha \uparrow$	0.847	0.853	0.841	0.857	0.887	0.767	–	0.866	0.880	**0.907**
	$E_\phi \uparrow$	0.903	0.912	0.905	0.911	0.935	0.776	–	0.929	0.935	**0.955**
	$M \downarrow$	0.048	0.043	0.046	0.042	0.032	0.078	–	0.041	0.036	**0.030**

domains.

Quantitative result. COMPrompter introduces detailed prompts and fine-tuning techniques particularly designed for COD tasks. In comparison with SAM, COMPrompter resulted in considerable advancement in the evaluation metrics. Compared with SAM-Adapter, which is another SAM-based COD method, COMPrompter achieved distinction with several advantages. On the COD10K dataset, COMPrompter achieved enhancements, including a 2% rise in F_β^ω, 0.6% increase in S_α, 3.1% boost in E_ϕ, and 0.2% improvement in M versus SAM-Adapter. COMPrompter achieved an average improvement of 4.9%, 1.7%, 4.5%, and 1.2% across F_β^ω, S_α, E_ϕ, and M, respectively, on three datasets in comparison with SAM-Adapter. COMPrompter demonstrates its superiority to non-SAM methods as well. On the CAMO dataset, COMPrompter outperformed MSCAFNet with a 3% boost in F_β^ω, 0.9% enhancements in S_α, 1.3% progress in E_ϕ, and 0.2% increase in M. From a holistic dataset viewpoint, COMPrompter exhibited an average enhancement of 2.6%, 1.2%, 1.3%, and 0.03% across F_β^ω, S_α, E_ϕ, and M, respectively, compared with MSCAFNet on the four evaluated datasets. When compared with MSCAFNet on the CHAMELEON dataset, COMPrompter demonstrated certain shortcomings. Overall, these results highlighted the effectiveness of COMPrompter in COD tasks. The experimental results of polyp seg-

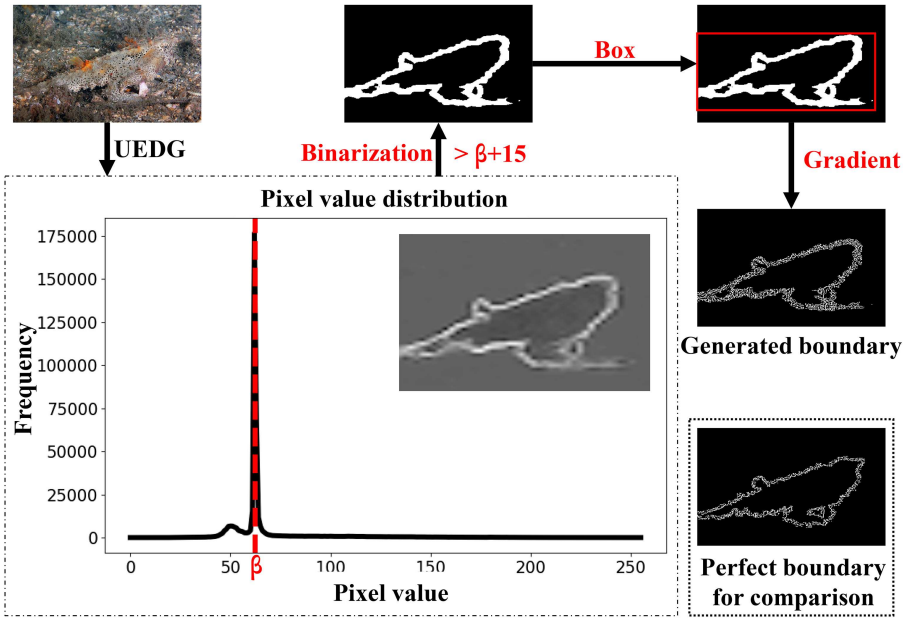

Figure 8 (Color online) Process of the generation of boundary with gradient. β represents the pixel value with the largest proportion, which is unfixed. The number 15 represents the offset value, which is fixed. The key step is in red.

mentation are presented in Table 2. Compared with MedSAM, COMPrompter had an average gain of 3.2% in mDice and 4.5% in mIoU. For COMPrompter with the generated boundary, the accuracy ranked within the top three.

Qualitative results. With a view to more intuitively showing the segmentation effect of our proposed COMPrompter on COD datasets and polyp datasets, we compared the original image and GT with the predictions generated by COMPrompter, MedSAM, and SAM. This comparison is shown in Figures 5–7. Figures 5 and 6 show the learning ability and generalization ability of COMPrompter, respectively. Because there is a lack of certain semantic information when SAM is directly applied to COD, only a part of the target was segmented. This is illustrated in the first and last columns in Figure 5. Because of providing a strong prior, the segmentation effect of MedSAM was observed to be greatly improved. In view of the fact that the bounding box only provides an approximate location, there still exists a certain level of semantic ambiguity. This resulted in occasional segmentation errors, as shown in the second image of Figure 5 for MedSAM. In addition, the processing of details of the edges and occlusions of the target was not particularly ideal. In particular, in the fifth line in Figure 6, we can see that MedSAM is affected by the weed and does not segment out the target hidden behind it. However, our proposed COMPrompter can segment the entire object and consider the details. One can see that the foot of the bird in the third column in Figure 5 is segmented. In terms of foreground occlusion, COMPrompter can clearly distinguish between target and occlusion, and finely segment them out, as shown in the sixth column of Figure 5 and the fifth line of Figure 6.

4.4 Ablation study

We confirmed the effectiveness of the box prompt, boundary prompt, and DWT modules via ablation experiments. We added the boundary prompt, box prompt, and DWT modules to SAM in turn, using the same experimental setup as for training. In particular, we designed five models to confirm the effectiveness of the three modules. The results on COD10K and NC4K are presented in Table 3. In general, each module played a positive role in boosting the experimental results. Finally, our proposed COMPrompter achieved SOTA performance.

As shown in Figure 7, the direct application of SAM to polyp segmentation resulted in a large number of incorrectly segmented regions. Although MedSAM could roughly segment the polyp, the edge was not clear. COMPrompter added a boundary prompt on the basis of the box prompt and more clearly segmented polyps.

Efficiency analysis. To comprehensively describe our model, we compared its input size, parameters, and inference speed with those of the COD-related and polyp-related models. Table 4 shows that the

Table 2 Quantitative results on five different datasets: CVC-ClinicDB, Kvasir, CVC-300, CVC-ColonDB, and ETIS-LaribPolypDB. The scores in bold are the best ones. mD represents mean dice similarity coefficient (mDice) and mI denotes mean intersection over union (mIoU). ↑ indicates that the higher the score the better, and ↓ indicates that the lower the score the better.

	Kvasir		CVC-ClinicDB		CVC-ColonDB		CVC-300		ETIS-LaribPolypDB	
	$mD\uparrow$	$mI\uparrow$	$mD\uparrow$	$mI\uparrow$	$mD\uparrow$	$mI\uparrow$	$mD\uparrow$	$mI\uparrow$	$mD\uparrow$	$mI\uparrow$
U-Net [49]	0.818	0.746	0.823	0.755	0.512	0.444	0.710	0.627	0.398	0.335
UNet++ [50]	0.821	0.743	0.794	0.729	0.483	0.410	0.707	0.624	0.401	0.344
ResU-Net++ [51]	0.813	0.793	0.796	0.796	–	–	–	–	–	–
SFA [52]	0.723	0.611	0.700	0.607	0.469	0.347	0.467	0.329	0.297	0.217
PraNet [2]	0.898	0.840	0.899	0.849	0.709	0.640	0.871	0.797	0.628	0.567
EU-Net [53]	0.908	0.854	0.902	0.846	0.756	0.681	0.837	0.765	0.687	0.609
SANet [54]	0.904	0.847	0.916	0.859	0.753	0.670	0.888	0.815	0.750	0.654
LDNet [55]	0.902	0.847	0.909	0.856	0.752	0.678	0.850	0.781	0.605	0.542
FAPNet [56]	0.902	0.849	0.925	0.877	0.731	0.658	0.893	0.826	0.717	0.643
SAM [21]	0.799	0.720	0.580	0.518	0.488	0.423	0.669	0.614	0.538	0.488
Med-SAM [20]	0.909	0.857	0.916	0.858	0.877	0.798	0.914	0.848	0.855	0.783
Ours	**0.935**	**0.892**	**0.931**	**0.883**	**0.917**	**0.856**	**0.938**	**0.889**	**0.910**	**0.849**

Table 3 Ablation study for each module of the proposed COMPrompter on COD datasets and the comparative experiments with different frequencies obtained using the DWT module. SAM (M1): this setting is the mode for segmenting all objects in SAM. We evaluated the mask with the best segmentation quality in this mode as the final result of segmentation. SAM + Box (M2): providing box prompt guidance based on M1. SAM + Boundary (M3): providing boundary prompt guidance on the basis of M1. SAM + Box + Boundary (M4): providing boundary prompt and box prompt on the basis of M1. SAM + Box + Boundary + DWT (M5): adding DWT module based on M4. LL denotes the low-frequency part of image. HH denotes the high-frequency part in the diagonal direction. LH denotes the combination of the high-frequency part in the horizontal direction and the low-frequency part in the vertical direction. HL denotes the converse.

Ablation study								
Model	COD10K				NC4K			
	F_β^ω	S_α	E_ϕ	M	F_β^ω	S_α	E_ϕ	M
M1	0.701	0.783	0.798	0.049	0.696	0.767	0.776	0.078
M2	0.752	0.841	0.918	0.034	0.819	0.865	0.928	0.042
M3	0.813	0.887	0.948	0.024	0.867	0.903	0.953	0.032
M4	0.813	0.884	0.947	0.024	0.873	0.905	0.955	0.030
M5	0.821	0.889	0.949	0.023	0.876	0.907	0.955	0.030
Frequencies comparative experiments								
Setting	COD10K				NC4K			
	F_β^ω	S_α	E_ϕ	M	F_β^ω	S_α	E_ϕ	M
LL	0.814	0.885	0.946	0.024	0.871	0.904	0.954	0.031
LH	0.812	0.883	0.946	0.025	0.869	0.902	0.953	0.032
HL	0.818	0.887	0.949	0.024	0.873	0.905	0.955	0.031
HH	0.821	0.889	0.949	0.023	0.876	0.907	0.955	0.030

SAM-based model has larger parameters and a longer inference time. Compared with SAM, COMPrompter greatly decreased the number of parameters and increased the inference speed by four times. Most importantly, the segmentation ability of COMPrompter greatly exceeded that of the existing models, irrespective of whether they were based on SAM or not. The inference times in Table 4 were obtained via testing in one NVIDIA RTX 3080TI GPU. Except for SAM, one can refer to the model performance metrics on the official website.

Effectiveness of box prompt. The effectiveness of the box prompt was confirmed by comparing two pairs of models: from M1 to M2 and from M3 to M4. From M1 to M2, we can see that the box prompt greatly enhanced the model performance. The enhancement achieved on COD10K was 5.1% in F_β^ω, while on NC4K, the enhancement was 12.3% in F_β^ω. From M3 to M4, other than the guidance already with the boundary prompt, we also saw a performance increase of 0.6% in F_β^ω owing to the introduction of the box prompt in NC4K.

Effectiveness of boundary prompt. The effectiveness of the boundary prompt was confirmed by comparing two pairs of models: from M1 to M3 and from M2 to M4. Looking at the four metrics, the

Table 4 Comparison of network complexity.

Method	Input size	Param (M)	Speed (fps)	Method (COD)	Input size	Param (M)	Speed (fps)
ResUNet++	256 × 256	4.06	1	PFNet	416 × 416	46.50	46
PraNet	352 × 352	32.55	42	R-MGL	473 × 473	67.64	14
EU-Net	256 × 256	31.43	11	LSR	352 × 352	57.90	83
SANet	352 × 352	23.90	67	JCSOD	352 × 352	121.63	53
LDNet	256 × 256	33.38	20	SINetV2	352 × 352	26.98	50
FAPNet	352 × 352	29.52	38	DGNet	352 × 352	21.02	40
PraNet	352 × 352	32.55	42	SAM	1024 × 1024	615	2
C2FNet	352 × 352	28.41	43	MedSAM	1024 × 1024	93.73	8
Ours	1024 × 1024	94.86	8	Ours	1024 × 1024	94.86	8

Table 5 Ablation study results for the dilate parameters of EGEM on COD10K and NC4K. Among the settings, the combinations of 1 × 1 and 1 × 1, and 1 × 1 and 3 × 3 have not been adopted because employing a kernel size of 1 is insufficient to capture the respective boundaries.

Setting	Dilate1	Dilate2	COD10K				NC4K			
			F_β^ω	S_α	E_ϕ	M	F_β^ω	S_α	E_ϕ	M
D1	3 × 3	3 × 3	0.824	0.891	0.951	0.023	0.875	0.906	0.953	0.030
D2 (office)	3 × 3	5 × 5	0.821	0.889	0.949	0.023	0.876	0.907	0.955	0.030
D3	5 × 5	5 × 5	0.820	0.888	0.949	0.023	0.875	0.907	0.954	0.030
D4	5 × 5	7 × 7	0.817	0.886	0.947	0.024	0.874	0.906	0.954	0.030
D5	7 × 7	7 × 7	0.814	0.884	0.948	0.024	0.873	0.905	0.953	0.030

average augmentation of boundary prompt on two datasets was 14.2% (F_β^ω), 12.0% (S_α), 16.4% (E_ϕ), and 3.6% (M) from M1 to M3. Compared with M2, M4 achieved an average increase of 5.8% (F_β^ω), 4.2% (S_α), 2.8% (E_ϕ), and 1.1% (M). These enhancements directly represent the strong effectiveness of the boundary prompt.

In the boundary prompt, EGEM obtains a gradient-containing boundary via appropriate dilation operations. The rationale behind the setting of the dilation parameters is presented in Table 5, which shows a general trend of gradual decrease from D1 to D5. However, considering potential biases in boundary acquisition during inference, the dilation parameter setting of D1 seemed overly precise (see Figure 9). Therefore, we adopted the parameters from D2.

To vividly illustrate the advantages of the multiprompt strategy, we showed the feature maps in the mask decoder of M2, M3, and COMPrompter, as depicted in Figure 10. Solely relying on M2 (SAM + Box prompt) yielded feature maps that roughly captured the target but suffered from edge blurriness (first line). In addition, there was insufficient attention to finer details (second line) and incomplete focus (third line). Feature maps generated by M3 (SAM + Boundary) exhibited higher edge attention (second line) but came with a broader activation range (first line). Meanwhile, COMPrompter with the multiprompt strategy demonstrated superior edge attention and appropriate activation ranges and even achieved complementary activation ranges in some instances (third line).

Effectiveness of DWT. The effectiveness of DWT was confirmed by comparing a pair of models: M4 and M5. Although the improvement brought by DWT is not as notable as that brought by boundary and box prompts, it is still observable. In addition, we conducted comparative experiments on which part of the frequencies in DWT were adopted. The experimental results are presented in Table 3. We have presented the results using a line chart, as shown in Figure 11, which shows that HH obtained the best score across all four metrics. Coincidentally, the curves of LL and HH in Figure 11(d) exactly coincided.

Effectiveness of offset value. An offset value was used for the binarization of the generated boundary. The effectiveness of the fixed offset value of 15 was demonstrated via a comparison of paired groups. For ease of comparison, we separately calculated the average positive and average negative metrics. As presented in Table 6, the set with an offset of 15 achieved the highest accuracy, with performance decreasing on either side. Therefore, we selected 15 as the optimal offset value.

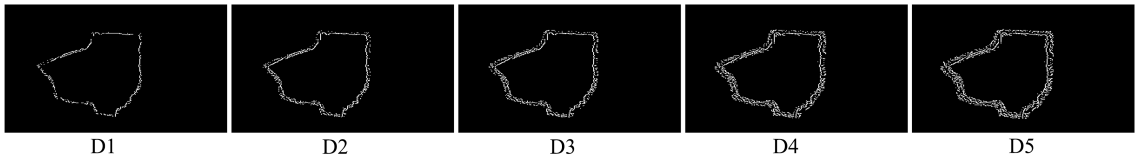

Figure 9 Comparison of boundary containing gradients obtained for various dilate parameter settings. One can see Table 5 for the dilate parameter settings for D1–D5.

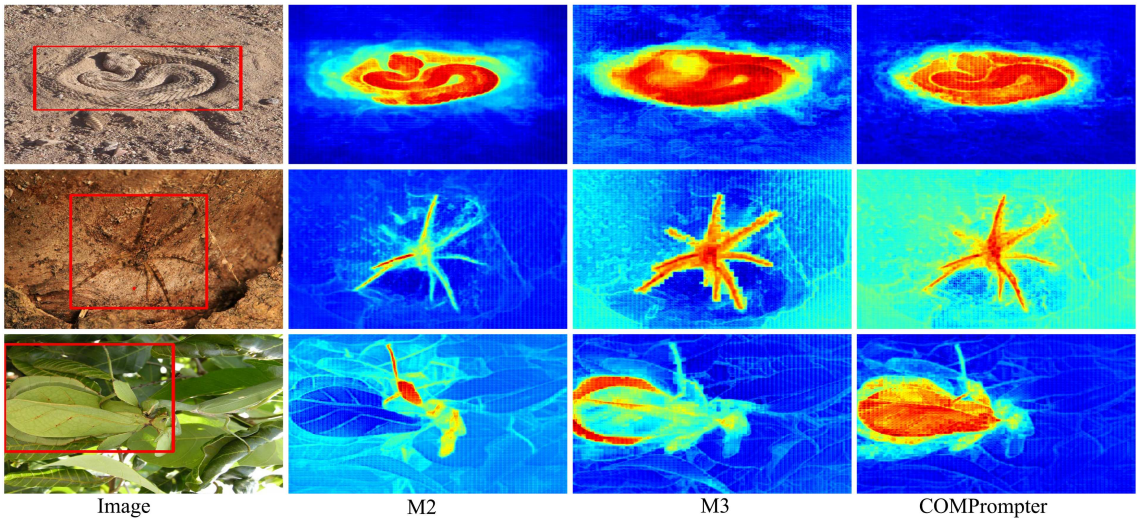

Figure 10 (Color online) Feature visualizations for the box prompt condition, the boundary prompt condition, and the full COMPrompter. Settings for SAM + Box (M2) are used within the box group, while settings for SAM + Boundary (M3) are employed within the boundary group. One can zoom in to see more details.

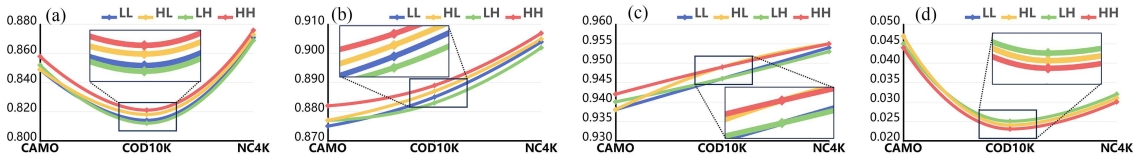

Figure 11 (Color online) Line graph with various frequencies obtained using the DWT module. The performance comparison of LL, LH, HL, and HH is in terms of the following four evaluation metrics: (a) weighted F-measure, (b) structure measure, (c) mean enhanced-alignment measure, and (d) mean absolute error.

Table 6 Ablation study results for the offset value of the gradient-containing generated boundary. Up denotes positive metric, and Down negative metric. The underline indicates the best results.

Offset	CAMO				CHAMELEON				COD10K				NC4K				Average	
	$F_\beta^\omega\uparrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$M\downarrow$	$F_\beta^\omega\uparrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$M\downarrow$	$F_\beta^\omega\uparrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$M\downarrow$	$F_\beta^\omega\uparrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$M\downarrow$	Up	Down
+5	0.811	0.847	0.918	0.057	0.814	0.875	0.936	0.034	0.770	0.858	0.931	0.028	0.835	0.879	0.935	0.037	0.8674	0.0390
+10	0.816	0.850	0.919	0.054	0.824	0.880	0.942	0.031	0.776	0.860	0.932	0.027	0.840	0.881	0.936	0.037	0.8713	0.0373
+15	0.819	0.853	0.919	0.054	0.830	0.884	0.946	0.030	0.779	0.861	0.933	0.026	0.840	0.880	0.935	0.036	0.8733	0.0365
+20	0.817	0.849	0.917	0.055	0.831	0.882	0.945	0.031	0.782	0.861	0.934	0.026	0.840	0.879	0.934	0.037	0.8726	0.0373
+25	0.816	0.850	0.918	0.055	0.827	0.880	0.941	0.033	0.783	0.861	0.934	0.026	0.841	0.880	0.934	0.037	0.8721	0.0378

5 Conclusion

We proposed COMPrompter, a novel network designed to advance the development of SAM in COD. It uses a multiprompt strategy incorporating a box prompt and boundary prompt for accurate priors. The SOTA performance of COMPrompter was demonstrated on multiple datasets. In addition, we highlighted the challenges suffered by COMPrompter in accurately and completely segmenting multiple objects within a single box. A single box can emphasize nontarget areas between multiple targets, resulting in segmentation errors. A possible solution is to use a box prompt with multiple subboxes to accurately cover all targets. We noted that the potential boundary prompt, might provide a novel perspective for COD and related fields. We expect that COMPrompter can contribute to the advancement

of SAM application to COD.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. U2033210, U24A20242, 62101387, 62475241) and Zhejiang Provincial Natural Science Foundation (Grant No. LDT23F02024F02).

References

1 Fan D P, Ji G P, Xu P, et al. Advances in deep concealed scene understanding. Vis Intell, 2023, 1: 16
2 Fan D-P, Ji G-P, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2020
3 Lidbetter T. Search and rescue in the face of uncertain threats. Eur J Oper Res, 2020, 285: 1153–1160
4 Liu Y, Li H, Cheng J, et al. MSCAF-Net: a general framework for camouflaged object detection via learning multi-scale context-aware features. IEEE Trans Circ Syst Video Technol, 2023, 33: 4934–4947
5 Sun Y, Chen G, Zhou T, et al. Context-aware cross-level fusion network for camouflaged object detection. 2021. ArXiv:210512555
6 Li A, Zhang J, Lv Y, et al. Uncertainty-aware joint salient object and camouflaged object detection. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2021
7 Zhai Q, Li X, Yang F, et al. Mutual graph learning for camouflaged object detection. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2021
8 Zhu J, Zhang X, Zhang S, et al. Inferring camouflaged objects by texture-aware interactive guidance network. In: Proceedings of AAAI, 2021
9 Ji G P, Fan D P, Chou Y C, et al. Deep gradient learning for efficient camouflaged object detection. Mach Intell Res, 2023, 20: 92–108
10 Pang Y, Zhao X, Xiang T-Z, et al. Zoom in and out: a mixed-scale triplet network for camouflaged object detection. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2022
11 Pang Y, Zhao X, Xiang T Z, et al. ZoomNeXt: a unified collaborative pyramid network for camouflaged object detection. IEEE Trans Pattern Anal Mach Intell, 2024, 46: 9205–9220
12 Jia Q, Yao S, Liu Y, et al. Segment, magnify and reiterate: detecting camouflaged objects the hard way. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2022
13 Fan D P, Ji G P, Cheng M M, et al. Concealed object detection. IEEE Trans Pattern Anal Mach Intell, 2021, 44: 6024–6042
14 Lv Y, Zhang J, Dai Y, et al. Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2021
15 Mei H, Ji G-P, Wei Z, et al. Camouflaged object segmentation with distraction mining. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2021
16 Fan D-P, Ji G-P, Sun G, et al. Camouflaged object detection. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2020
17 Yang F, Zhai Q, Li X, et al. Uncertainty-guided transformer reasoning for camouflaged object detection. In: Proceedings of International Conference on Computer Vision, 2021
18 Zhang J, Fan D-P, Dai Y, et al. UC-Net: uncertainty inspired RGB-D saliency detection via conditional variational autoencoders. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2020
19 Chen T, Zhu L, Ding C, et al. SAM fails to segment anything?–SAM-Adapter: adapting SAM in underperformed scenes: camouflage, shadow, and more. 2023. ArXiv:230409148
20 Ma J, Wang B. Segment anything in medical images. 2023. ArXiv:230412306
21 Kirillov A, Mintun E, Ravi N, et al. Segment anything. In: Proceedings of International Conference on Computer Vision, 2023
22 Sifre L. Rigid-motion scattering for image classification. Dissertation for Ph.D. Degree. Paris: Ecole Polytechnique, 2014
23 He C, Li K, Zhang Y, et al. Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2023
24 Ji G-P, Fan D-P, Xu P, et al. SAM struggles in concealed scenes–empirical study on "segment anything". 2023. ArXiv:230406022
25 Zhu S, Jing W, Kang P, et al. Data augmentation and few-shot change detection in forest remote sensing. IEEE J Sel Top Appl Earth Observations Remote Sens, 2023, 16: 5919–5934
26 Peng C, Zhu M, Ren H, et al. Small object detection method based on weighted feature fusion and CSMA attention module. Electronics, 2022, 11: 2546
27 Jiang Y, Yin G, Jing W, et al. Box-spoof attack against single object tracking. Appl Intell, 2024, 54: 1585–1601
28 Lv Y, Zhang J, Dai Y, et al. Toward deeper understanding of camouflaged object detection. IEEE Trans Circ Syst Video Technol, 2023, 33: 3462–3476
29 Li H, Feng C M, Xu Y, et al. Zero-shot camouflaged object detection. IEEE Trans Image Process, 2023, 32: 5126–5137
30 Zhang C, Bi H, Xiang T Z, et al. Collaborative camouflaged object detection: a large-scale dataset and benchmark. IEEE Trans Neural Netw Learn Syst, 2024, 35: 18470–18484
31 Huang Z, Dai H, Xiang T-Z, et al. Feature shrinkage pyramid for camouflaged object detection with transformers. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2023
32 Luo X-J, Wang S, Wu Z W, et al. CamDiff: camouflage image augmentation via diffusion. CAAI Artif Intell Res, 2023, 2: 55–64
33 Zhu H, Li P, Xie H, et al. I can find you! Boundary-guided separated attention network for camouflaged object detection. In: Proceedings of AAAI, 2022
34 Ji G P, Zhu L, Zhuge M, et al. Fast camouflaged object detection via edge-based reversible re-calibration network. Pattern Recogn, 2022, 123: 108414
35 Sun Y, Wang S, Chen C, et al. Boundary-guided camouflaged object detection. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence Main Track, 2022. 1335–1341
36 Lyu Y, Zhang H, Li Y, et al. UEDG: uncertainty-edge dual guided camouflage object detection. IEEE Trans Multimedia, 2024, 26: 4050–4060
37 Sun D, Jiang S, Qi L. Edge-aware mirror network for camouflaged object detection. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2023
38 Dong B, Pei J, Gao R, et al. A unified query-based paradigm for camouflaged instance segmentation. In: Proceedings of ACM International Conference on Multimedia, 2023
39 Premachandran V, Bonev B, Lian X, et al. Pascal boundaries: a semantic boundary dataset with a deep semantic boundary detector. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV), 2017
40 Wang J, Wu Z, Chen J, et al. Objectformer for image manipulation detection and localization. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2022
41 Liu W, Shen X, Pun C-M, et al. Explicit visual prompting for low-level structure segmentations. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2023

42 Le T N, Nguyen T V, Nie Z, et al. Anabranch network for camouflaged object segmentation. Comput Vision Image Underst, 2019, 184: 45–56
43 Skurowski P, Abdulameer H, Błaszczyk J, et al. Animal camouflage analysis: chameleon database. 2018. https://www.polsl.pl/rau6/chameleon-database-animal-camouflage-analysis/
44 Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int J CARS, 2014, 9: 283–293
45 Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computized Med Imag Graph, 2015, 43: 99–111
46 Tajbakhsh N, Gurudu S R, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imag, 2016, 35: 630–644
47 Vázquez D, Bernal J, Sánchez F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images. J Healthc Eng, 2017, 2017: 4037190
48 Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-SEG: a segmented polyp dataset. In: Proceedings of MultiMedia Modeling, 2020
49 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2015
50 Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: a nested U-Net architecture for medical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2018
51 Jha D, Smedsrud P H, Riegler M A, et al. ResUNet++: an advanced architecture for medical image segmentation. In: Proceedings of IEEE International Symposium on Multimedia, 2019
52 Fang Y, Chen C, Yuan Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2019
53 Patel K, Bur A M, Wang G. Enhanced U-Net: a feature enhancement network for polyp segmentation. In: Proceedings of the 18th Conference on Robots and Vision (CRV), 2021
54 Wei J, Hu Y, Zhang R, et al. Shallow attention network for polyp segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2021
55 Zhang R, Lai P, Wan X, et al. Lesion-aware dynamic kernel for polyp segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 2022
56 Zhou T, Zhou Y, Gong C, et al. Feature aggregation and propagation network for camouflaged object detection. IEEE TIP, 2022, 31: 7036–7047