BA-SAM: Scalable Bias-Mode Attention Mask for Segment Anything Model

Supplementary Material

1. Datasets Details

1.1. Object Segmentation Tasks

Salient object segmentation [2] identifies and extracts the most visually prominent objects in an image. *DUTS* [8] is currently the largest dataset for salient object segmentation, consisting of 10,553 training images and 5,019 test images.
Dichotomous image segmentation [7] aims to segment highly accurate objects from natural images. *DIS-TE4* [7] dataset provides 500 highly accurate (the range covers from 2k to 6k) pixel-level masks with complex object shapes.
Camouflaged object detection [4] aims to identify and localize objects that are visually disguised or blended into their surrounding environments. This task poses significant challenges due to camouflage patterns' complex and diverse nature. *COD10K* [4] is a substantial dataset for camouflaged object segmentation, featuring 10,000 high-definition (1080p) images characterized by similar foreground objects and surroundings.
Skin Lesion Segmentation is to segment the skin lesions from medical images, and *ISIC* [3] dataset consists of several sizes of high-resolution images of skin lesions.

1.2. Instance Segmentation

Instance segmentation is a computer vision task that combines object detection and semantic segmentation. It goes beyond traditional object detection by not only identifying objects in an image but also distinguishing between individual object instances, even if they belong to the same class. COCO [6] dataset is a widely used benchmark, for instance segmentation task. It provides a large and diverse collection of images with pixel-level annotations, making it a valuable resource for developing and evaluating instance segmentation algorithms.

2. More Implementation Details

Our BA-SAM can be applied in both zero-shot and fine-tuning scenarios, with the complete implementation code provided in the accompanying zip file and relevant execution commands detailed in the readme file.
Experiments without fine-tuning (i.e., zero-shot). We follow the same inference pipeline of SAM. Under such cases, we fixed hyperparameters λ. The New Scaling Factor, as computed following the formula described in the main paper, remains constant throughout the entire training process. The Bias-Mode Attention Mask utilizes a static slope, defaulting to 0.1. In the ablation experiments, we demonstrated the robustness of our design concerning the choice of slope.

Size	0.1	0.5	1	1.5	2	2.5
MAE	0.0479	0.0482	0.0481	0.0481	0.0482	0.0481

Table 1. Impact of the slope β in BM-AM (finetune setting).

Experiments with fine-tuning. For fair comparisons, we keep most of the hyperparameters in BA-SAM the same as SAM [5] and MobileSAM [9]. No additional data processing operations are employed, and Binary Cross-Entropy loss is used for training. The AdamW optimizer is used for all experiments with an initial learning rate of 1e-4. StepLR, with a step size of 10 and a gamma value of 0.5, is used for the learning rate decay. Unless otherwise stated, all models are initialized with parameters provided by [9] and fine-tuned for 50 epochs on downstream tasks. The New Scaling Factor remains unchanged compared to the scenarios without fine-tuning. The Bias-Mode Attention Mask employs learnable parameters to replace the static slope setting.

Tab. 1 reveals the impact of different slopes of BM-AM on the DIS [7] dataset after fine-tuning.

3. Visualization

In the main manuscript, we presented the segmentation performance of the proposed BA-SAM at different resolutions to demonstrate its effectiveness. Here, additional visualization comparisons with the SAM model are provided in Fig. 2, highlighting a significant improvement in segmentation on these datasets.

References

[1] Lucas Beyer, Pavel Izmailov, Alexander Kolesnikov, Mathilde Caron, Simon Kornblith, Xiaohua Zhai, Matthias Minderer, Michael Tschannen, Ibrahim Alabdulmohsin, and Filip Pavetic. Flexivit: One model for all patch sizes. In *CVPR*, 2023. 1

[2] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. Salient object detection: A survey. *CVM*, 5, 2019. 1

[3] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *ISBI*, 2018. 1, 2

[4] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, 2020. 1, 2

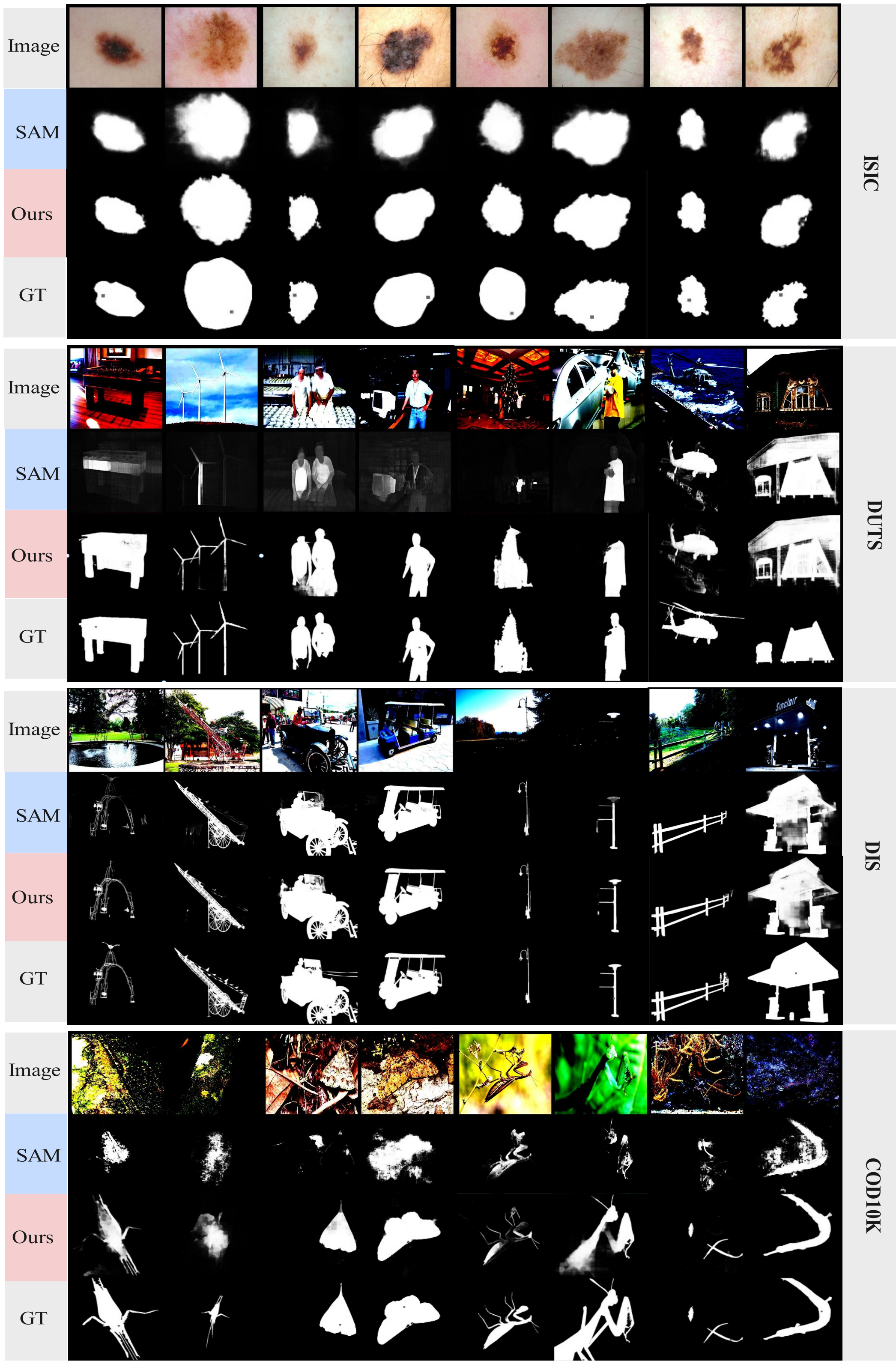

Figure 2. Visualization results of our BA-SAM on four object segmentation tasks compared with SAM, *i.e.,* skin lesion segmentation, salient object segmentation, complex object segmentation, and camouflaged object detection, corresponding to four datasets, respectively: ISIC [3], DUTS [8], DIS-TE4 [7], and COD10K [4].

[5] 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. *ICCV*, 2023. 1

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1

[7] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool. Highly accurate dichotomous image segmentation. In *ECCV*, 2022. 1, 2

[8] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017. 1, 2

[9] Chaoning Zhang, Dongshen Han, Yu Qiao, Jung Uk Kim, Sung-Ho Bae, Seungkyu Lee, and Choong Seon Hong. Faster segment anything: Towards lightweight sam for mobile applications. *arXiv preprint arXiv:2306.14289*, 2023. 1

[10] Kaidong Zhang and Dong Liu. Customized segment anything model for medical image segmentation. In *arXiv preprint arXiv:2304.13785*, 2023. 1