SAM Fails to Segment Anything? – SAM-Adaptor: Adapting SAM in Underperformed Scenes

Tianrun Chen[1,2+*] Lanyun Zhu[4+] Chaotao Ding[3+] Runlong Cao[3+]

Shangzhan Zhang[2] Yan Wang [5] Papa Mao[1]

Ying Zang[3*]

14 April, 2023

[+] Equal Contribution [*] Corresponding Author
{tianrun.chen@zju.edu.cn; 02750@zjhu.edu.cn}
[1]KOKONI, Moxin (Huzhou) Tech. Co., LTD, Huzhou, Zhejiang, P.R. China.
[2]College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, P.R. China.
[3]School of Information Engineering, Huzhou University, Huzhou, Zhejiang, P.R. China.
[4] Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore.
[5] School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, P.R. China.

Abstract

The emergence of large models, also known as foundation models, has brought significant advancements to AI research. One such model is Segment Anything (SAM), which is designed for image segmentation tasks. However, as with other foundation models, our experimental findings suggest that SAM may fail or perform poorly in certain segmentation tasks, such as shadow detection and camouflaged object detection. This study first paves the way for applying the large pre-trained image segmentation model SAM to these downstream tasks, even in situations where SAM performs poorly. Rather than fine-tuning the SAM network, we propose **SAM-Adaptor**, which incorporates domain-specific information or visual prompts into the segmentation network by using simple yet effective adaptors. Our extensive experiments show that SAM-Adaptor can significantly elevate the performance of SAM in challenging tasks and we can even achieve state-of-the-art performance. We believe our work opens up opportunities for utilizing SAM in downstream tasks, with potential applications in various fields, including medical image processing, agriculture, remote sensing, and more.

1 Introduction

AI research has witnessed a paradigm shift with models trained on vast amounts of data at scale. These models, or known as foundation models, such as BERT, DALL-E, and GPT-3 have shown promising results in many language or vision tasks[1]. Recently, among the foundation models, Segment Anything (SAM)[2] has a distinct position as a generic image segmentation model trained on the large visual corpus [2]. It has been demonstrated that SAM has successful segmentation capabilities in diverse scenarios, which makes it a groundbreaking step toward image segmentation and related fields of computer vision.

However, as computer vision encompasses a broad spectrum of problems, SAM's incompleteness is evident, which is similar to other foundation models since the training data cannot encompass the entire corpus, and working scenarios are subject to variation [1]. In this study, we first test SAM in some challenging low-level structural segmentation tasks including camouflaged object detection (concealed scenes) and shadow detection, and we find that the SAM model trained on general images cannot perfectly "Segment Anything" in these cases.

As such, a crucial research problem is: *How to harness the capabilities acquired by large models from massive corpora and leverage them to benefit downstream tasks?*

Here, we introduce the **SAM-adaptor**, which serves as a solution to the research problem mentioned above. ***This pioneering work is the first attempt to adapt the large pre-trained image segmentation model SAM to specific downstream tasks with enhanced performance.*** As its name states, **SAM-adaptor** is a very simple yet effective adaptation technique that leverages internal knowledge and external control signal. Specifically, it is a lightweight model that can learn alignment with a relatively small amount of data and serves as an additional network to inject task-specific guidance information from the samples of that task. Information is conveyed to the network using visual prompts [3, 4], which has been demonstrated to be efficient and effective in adapting a frozen large foundation model to many downstream tasks with a minimum number of additional trainable parameters.

Specifically, we show that our method is:

- **Generalizable**: SAM-adaptor can be directly applied to customized datasets of various tasks to enhance performance with the assistance of SAM.
- **Composable**: It is effortless to combine multiple explicit conditions to fine-tune SAM with multi-condition control.

We perform extensive experiments on multiple tasks and datasets, including ISTD for shadow detection [5] and COD10K [6] for camouflaged object detection task. Benefiting from the capability of SAM and our SAM-adaptor, our method achieves state-of-the-art (SOTA) performance on both datasets. The contributions of this work can be summarized as follows:

- First, we pioneer the analysis of the incompleteness of the Segment Anything (SAM) model and propose a research problem of how to utilize the SAM model to serve downstream tasks.
- Second, we are the first to propose the adaptation approach, **SAM-adaptor**, to adapt SAM to downstream tasks and achieve enhanced performance.
- Third, despite SAM's backbone being a simple plain model lacking specialized structures tailored for the two specific downstream tasks, our approach still surpasses existing methods and attains state-of-the-art (SOTA) performance in these tasks.

To the best of our knowledge, this work pioneers to demonstrate the exceptional ability of SAM to transfer to other specific data domains with remarkable accuracy. While we only tested it on a few datasets, we expect SAM-adaptor can serve as an effective and adaptable tool for various downstream segmentation tasks in different fields, including medical and agriculture. This study will usher in a new era of utilizing large pre-trained image models in diverse research fields and industrial applications.

2 Related Work

Semantic Segmentation. In recent years, semantic segmentation has made significant progress, primarily due to the remarkable advancements in deep-learning-based methods such as fully convolutional networks (FCN) [7], encoder-decoder structures [8, 9, 10, 11, 12], dilated convolutions [13, 14, 15, 16, 17], pyramid structures [18, 15, 19, 16, 20], attention modules [21, 22, 23], and transformers [24, 25, 26, 27]. Building upon previous research, Segment Anything (SAM) [2] introduces a large ViT-based model trained on a large visual corpus. This work aims to leverage the SAM to solve specific downstream image segmentation tasks.

Adaptors. The concept of adaptors was first introduced in the NLP community [28] as a tool to fine-tune a large pre-trained model for each downstream task with a compact and scalable model. In [29], multi-task learning was explored with a single BERT model shared among a few task-specific parameters. In the computer vision community, [30] suggested fine-tuning the ViT [31] for object detection with minimal modifications. Recently, ViT-adaptor [32] leveraged adaptors to enable a

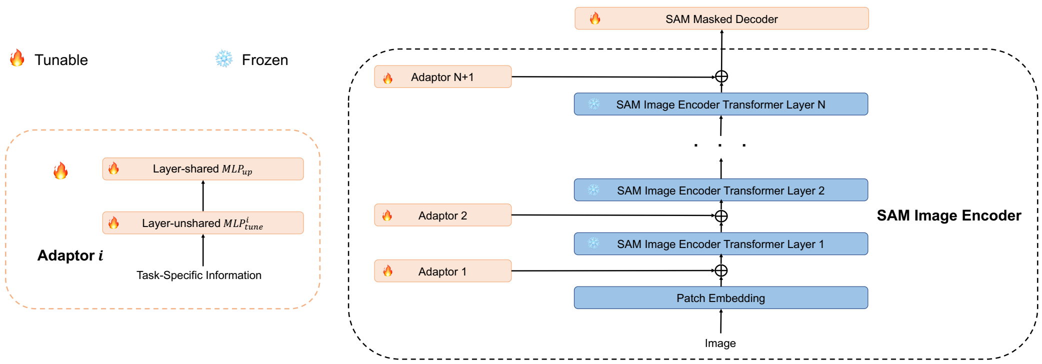

Figure 1: **The architecture of the proposed SAM-Adaptor.**

plain ViT to perform various downstream tasks. [4] introduce an Explicit Visual Prompting (EVP) technique that can incorporate explicit visual cues to the adaptor. However, no prior work has tried to apply adaptors to leverage pretrained image segmentation model SAM trained at large image corpus. Here, we mitigate the research gap.

3 Method

3.1 Using SAM as the Backbone

As previously illustrated, the goal of the SAM-adaptor is to leverage the knowledge learned from the SAM. Therefore, we use SAM as the backbone of the segmentation network. The image encoder of SAM is a ViT-H/16 model with 14x14 windowed attention and four equally-spaced global attention blocks. We keep the weight of pretrained image encoder frozen. We also leverage the mask decoder of the SAM, which consists of a modified transformer decoder block followed by a dynamic mask prediction head. We use the pretrained SAM's weight to initialize the weight of the mask decoder of our approach and tune the mask decoder during training. We input no prompts into the original mask decoder of SAM.

3.2 Adaptors

Next, the task-specific knowledge F^i is learned and injected into the network via adaptors. We employ the concept of prompting, which utilizes the fact that foundation models have been trained on large-scale datasets. This can enhance the model's generalization ability on downstream tasks, especially when annotated data is scarce.

The architecture of the proposed SAM-Adaptor is illustrated in Figure 1. We aim to keep the design of the adaptor to be simple and efficient. Therefore, we choose to use an adaptor that consists of only two MLPs and an activate function within two MLPs [4]. Specifically, the adaptor takes the information F^i and obtains the prompt P^i:

$$P^i = \mathrm{MLP}_{up}\left(\mathrm{GELU}\left(\mathrm{MLP}^i_{tune}\left(F_i\right)\right)\right) \tag{1}$$

in which MLP^i_{tune} are linear layers used to generate task-specific prompts for each adaptor. MLP_{up} is an up-projection layer shared across all adaptors that adjusts the dimensions of transformer features. P^i refers to the output prompt that is attached to each transformer layer of SAM model. GELU is the GELU activation function [33]. The information F^i can be chosen to be in various forms.

3.3 Input Task-Specific Information

It is worth noting that the information F^i can be in various forms depending on the task and flexibly designed. For example, it can be extracted from the given samples of the specific dataset of the task in some form, such as texture or frequency information, or some hand-crafted rules. Moreover, the

	SINet[34]	RankNet[35]	JCOD[36]	PFNet[37]	FBNet[38]	SAM[2]	SAM-Adaptor (Ours)
$S_\alpha \uparrow$	0.771	0.767	0.800	0.800	0.809	0.781	**0.921**
$E_\phi \uparrow$	0.806	0.861	0.872	0.868	0.889	0.800	**0.930**
MAE $\downarrow$	0.551	0.045	0.041	0.040	0.035	0.054	**0.023**

Table 1: Quantitative Result for Camouflaged Object Detection

F^i can be in a composition form consisting multiple guidance information:

$$F_i = \sum_{1}^{N} w_j F_j \tag{2}$$

in which F^j can be one specific type of knowledge/features and w^j is an adjustable weight to control the composed strength.

4 Experiments

4.1 Tasks and Datasets

We select two challenging low-level structural segmentation task for SAM – camouflaged object detection and shadow detection. For camouflaged object detection, we choose COD10K dataset [6] in our experiment, which is the largest dataset for camouflaged object detection containing 3,040 training and 2,026 testing samples. ISTD dataset [5] contains 1,330 training images and 540 test images. For evaluation metrics, we follow the protocol in [4] and use commonly-used S-measure (S_m), mean E-measure (E_ϕ), and MAE for evaluation of camouflaged object detection, and use balance error rate (BER) for shadow detection.

4.2 Implementation Details

In the experiment, we choose two types of visual knowledge, patch embedding F_{pe} and high-frequency components F_{hfc}, following the same setting in [4], which has been demonstrated effective in various of vision tasks. w^j is set to 1. Therefore, the F_i is derived by $F_i = F_{hfc} + F_{pe}$.

The MLP^i_{tune} has 32 linear layers and MLP^i_{up} is one linear layer that maps the output from GELU activation to the number of inputs of the transformer layer. We use ViT-H version of SAM. Balanced BCE loss is used for shadow detection. BCE loss and IOU loss are used for camouflaged object detection. AdamW optimizer is used for all the experiments. The initial learning rate is set to 2e-4. Cosine decay is applied to the learning rate. The training is performed for 20 epochs. The experiments are implemented using PyTorch on four NVIDIA Tesla A100 GPUs.

4.3 Experimental Result for Camouflaged Object Detection

We first evaluate SAM in camouflaged object detection task, which is a very challenging task as foreground objects are often with visual similar patterns to the background. Our experiments revealed that SAM did not perform well in this task. As shown in Figure 2, SAM failed to detect some concealed objects. This can be further confirmed by the quantitative results presented in Table Y. In fact, SAM's performance was significantly lower than the existing state-of-the-art methods in all metrics evaluated.

In Figure 2, it can be found clearly that by introducing the SAM-adaptor, our method significantly elevates the performance of the model (+17.9% in S_α). Our approach successfully identifies concealed objects, as evidenced by clear visual results. Quantitative results also show that our method outperforms the existing state-of-the-art method.

4.4 Experimental Result for Shadow Detection

We also evaluated SAM on the task of shadow detection. However, as depicted in Figure 3, SAM struggled to differentiate between the shadow and the background information.

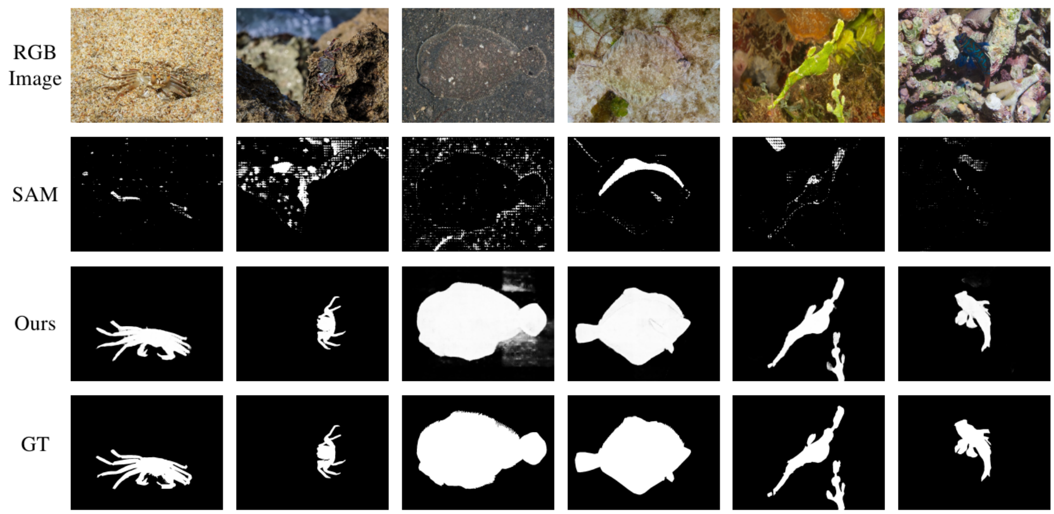

Figure 2: **The Visualization Results of Camouflaged Image Segmentation.** As illustrated in the figure, the SAM failed to perceive those animals that are visually 'hidden'/concealed in their natural surroundings. By using SAM-adaptor, our approach can significantly elevate the performance of object segmentation with SAM.

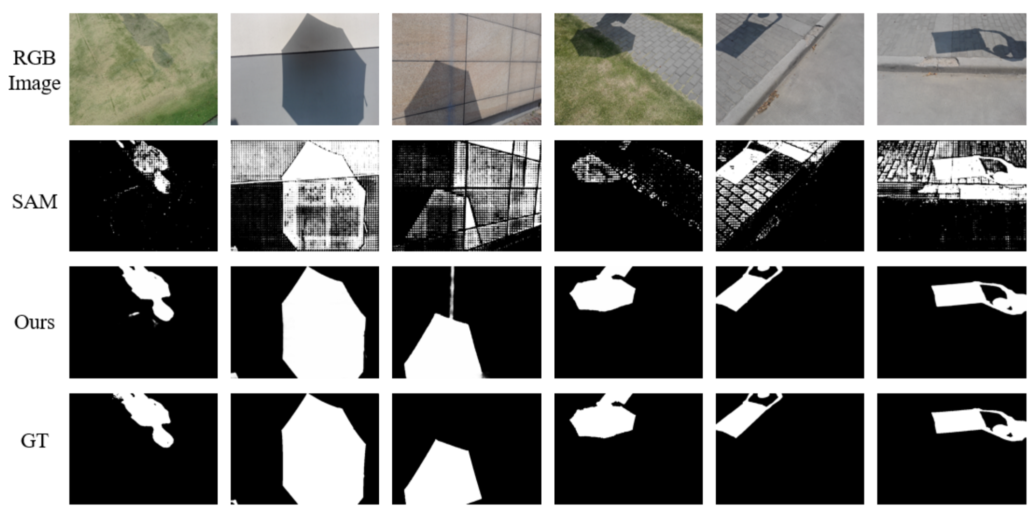

Figure 3: **The Visualization Results of Shadow Detection.** As illustrated in the figure, the SAM failed to distinguish the shadow and the background object. By using SAM-adaptor, our approach can significantly elevate the performance of object segmentation with SAM.

Our evaluation metrics for shadow detection were significantly poorer compared to existing methods. However, by integrating the **SAM-Adapter**, we were able to significantly improve the performance. The shadow region can be clearly identified by our approach. The quantitative results also verify the performance boost, as demonstrated in Table. 2.

Method	BER ↓
Stacked CNN [39]	8.60
BDRAR [40]	2.69
DSC [41]	3.42
DSD [42]	2.17
SAM [2]	40.51
SAM-Adaptor (Ours)	**1.88**

Table 2: Quantitative Result for Shadow Detection

5 Conclusion

In this work, we first extend the Segment Anything (SAM) model and apply it to some downstream tasks. Our experiments reveal that, like other foundational models, SAM is not effective in some vision tasks, for example, dealing with concealed objects. Therefore, we propose the SAM-adaptor, which utilizes SAM as the backbone and injects customized information into the network through simple yet effective adaptors to enhance performance in specific tasks. We evaluate our approach in camouflaged object detection and shadow detection tasks and demonstrate that the SAM-adaptor not only significantly improves SAM's performance but also achieves state-of-the-art (SOTA) results. We anticipate that this work will pave the way for applying SAM in downstream tasks and will have significant impacts in various image segmentation and computer vision fields.

6 Future Work

This study showcases the effectiveness and versatility of using adaptors and large foundation models. Moving forward, we plan to extend the SAM-Adaptor to tackle even more challenging image segmentation tasks and broaden its application to other fields. We also anticipate the development of more specialized designs tailored to specific tasks.

References

[1] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[2] 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.

[3] Senqiao Yang, Jiarui Wu, Jiaming Liu, Xiaoqi Li, Qizhe Zhang, Mingjie Pan, and Shanghang Zhang. Exploring sparse visual prompt for cross-domain semantic segmentation. *arXiv preprint arXiv:2303.09792*, 2023.

[4] Weihuang Liu, Xi Shen, Chi-Man Pun, and Xiaodong Cun. Explicit visual prompting for low-level structure segmentations. *arXiv preprint arXiv:2303.10883*, 2023.

[5] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1788–1797, 2018.

[6] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2777–2787, 2020.

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[9] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 325–341, 2018.

[10] Mingyuan Fan, Shenqi Lai, Junshi Huang, Xiaoming Wei, Zhenhua Chai, Junfeng Luo, and Xiaolin Wei. Rethinking bisenet for real-time semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9716–9725, 2021.

[11] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[12] Xiangtai Li, Ansheng You, Zhen Zhu, Houlong Zhao, Maoke Yang, Kuiyuan Yang, Shaohua Tan, and Yunhai Tong. Semantic flow for fast and accurate scene parsing. In *European Conference on Computer Vision*, pages 775–793. Springer, 2020.

[13] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[15] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[16] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[17] Zhikang Liu and Lanyun Zhu. Label-guided attention distillation for lane segmentation. *Neurocomputing*, 438:312–322, 2021.

[18] Lanyun Zhu, Deyi Ji, Shiping Zhu, Weihao Gan, Wei Wu, and Junjie Yan. Learning statistical texture for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12537–12546, 2021.

[19] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[20] Lanyun Zhu, Tianrun Chen, Jianxiong Yin, Simon See, and Jun Liu. Continual semantic segmentation with automatic memory sample selection. *arXiv preprint arXiv:2304.05015*, 2023.

[21] Fan Zhang, Yanqin Chen, Zhihang Li, Zhibin Hong, Jingtuo Liu, Feifei Ma, Junyu Han, and Errui Ding. Acfnet: Attentional class feature network for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6798–6807, 2019.

[22] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019.

[23] Zhen Zhu, Mengde Xu, Song Bai, Tengteng Huang, and Xiang Bai. Asymmetric non-local neural networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 593–602, 2019.

[24] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890, 2021.

[25] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021.

[26] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7262–7272, 2021.

[27] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1290–1299, 2022.

[28] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR, 2019.

[29] Asa Cooper Stickland and Iain Murray. Bert and pals: Projected attention layers for efficient adaptation in multi-task learning. In *International Conference on Machine Learning*, pages 5986–5995. PMLR, 2019.

[30] Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring plain vision transformer backbones for object detection. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX*, pages 280–296. Springer, 2022.

[31] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[32] Zhe Chen, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, and Yu Qiao. Vision transformer adapter for dense predictions. *arXiv preprint arXiv:2205.08534*, 2022.

[33] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[34] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2777–2787, 2020.

[35] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11591–11601, 2021.

[36] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10071–10081, 2021.

[37] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8772–8781, 2021.

[38] Jiaying Lin, Xin Tan, Ke Xu, Lizhuang Ma, and Rynson WH Lau. Frequency-aware camouflaged object detection. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2):1–16, 2023.

[39] Tomás F Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pages 816–832. Springer, 2016.

[40] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 121–136, 2018.

[41] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7454–7462, 2018.

[42] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. Distraction-aware shadow detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5167–5176, 2019.