SAM2-Adapter: Evaluating & Adapting Segment Anything 2 in Downstream Tasks: Camouflage, Shadow, Medical Image Segmentation, and More

Tianrun Chen *, Ankang Lu, Lanyun Zhu, Chaotao Ding, Chunan Yu, Deyi Ji, Zejian Li, Lingyun Sun, Ying Zang *

Posted Date: 9 August 2024

doi: 10.20944/preprints202408.0622.v1

Keywords: Segment Anything; Adapter; Visual Prompting; Camouflaged; Shadow; Image Segmentation; Polyp Segmentation; Medical Image Segmentation

Article

SAM2-Adapter: Evaluating & Adapting Segment Anything 2 in Downstream Tasks: Camouflage, Shadow, Medical Image Segmentation, and More †

Tianrun Chen [1,2*,‡], Ankang Lu [3,‡], Lanyun Zhu [4,‡], Chaotao Ding [1,‡], Chunan Yu [3], Deyi Ji [5], Zejian Li [6], Lingyun Sun [2], Papa Mao[1] and Ying Zang [3,*]

[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 Information Science and Technology, University of Science and Technology of China, P.R. China
[6] School of Software Technology, Zhejiang University, Hangzhou, Zhejiang, P.R. China
* Correspondence: tianrun.chen@zju.edu.cn (TC.); 02750@zjhu.edu.cn (Y.Z.)

† **This Work is Built Upon the SAM-Adapter - First Available at 14 April, 2023**
TL; DR. SAM2, enhanced with our new adapter, can replace SAM as the backbone for downstream tasks, **establishing new state-of-the-art (SOTA) results!**

‡ Equal Contribution.

Abstract: The advent of large models, also known as foundation models, has significantly transformed the AI research landscape, with models like Segment Anything (SAM) achieving notable success in diverse image segmentation scenarios. Despite its advancements, SAM encountered limitations in handling some complex low-level segmentation tasks like camouflaged object and medical imaging. In response, in 2023, we introduced SAM-Adapter, which demonstrated improved performance on these challenging tasks. Now, with the release of Segment Anything 2 (SAM2)—a successor with enhanced architecture and a larger training corpus—we reassess these challenges. This paper introduces SAM2-Adapter, the first adapter designed to overcome the persistent limitations observed in SAM2 and achieve new state-of-the-art (SOTA) results in specific downstream tasks including medical image segmentation, camouflaged (concealed) object detection, and shadow detection. SAM2-Adapter builds on the SAM-Adapter's strengths, offering enhanced generalizability and composability for diverse applications. We present extensive experimental results demonstrating SAM2-Adapter's effectiveness. We show the potential and encourage the research community to leverage the SAM2 model with our SAM2-Adapter for achieving superior segmentation outcomes. Code, pre-trained models, and data processing protocols are available at http://tianrun-chen.github.io/SAM-Adaptor/

Keywords: Segment Anything; Adapter; Visual Prompting; camouflaged; shadow; image segmentation; polyp segmentation; medical image segmentation

1. Introduction

The AI research landscape has been transformed by foundation models trained on vast data [1–4]. Recently, among the foundation models, Among these, Segment Anything (SAM) [5] stands out as a highly successful image segmentation model with demonstrated success in diverse scenarios. However, in our previously study, we found that SAM's performance was limited in some challenging low-level structural segmentation tasks, such as camouflaged object detection and shadow detection. To address this, in 2023, within two weeks of SAM's release, we proposed the SAM-Adapter [6,7], which aimed to leverage the power of the SAM model to deliver better performance on these challenging downstream tasks. The success of the SAM-Adapter, with its training and evaluation code and checkpoints made publicly available, has already been a valuable resource for many researchers in the community to experiment with and build upon, demonstrating its effectiveness on a variety of downstream tasks.

Now, the research community has pushed the boundaries further with the introduction of an even more capable and versatile successor to SAM, known as Segment Anything 2 (SAM2). Boasting

 doi:10.20944/preprints202408.0622.v1

further enhancements in its network architecture and training on an even larger visual corpus, SAM2 has certainly piqued our interest. This naturally leads us to the questions:

- Do the challenges faced by SAM in downstream tasks persist in SAM2?
- Can we replicate the success of SAM-Adapter and leverage SAM2's more powerful pre-trained encoder and decoder to achieve new state-of-the-art (SOTA) results on these tasks?

In this paper, we answer both questions with a resounding "Yes." Our experiments confirm that the challenges SAM encountered in downstream tasks do persist in SAM2, due to the inherent limitations of foundation models—where training data cannot cover the entire corpus and working scenarios vary [1]. However, we have devised a solution to address this challenge. By introducing the **SAM2-Adapter**, we've created a multi-adapter configuration that leverages SAM2's enhanced components to achieve new SOTA results in tasks including medical image segmentation, camouflaged object detection, and shadow detection.

Just like SAM-Adapter [6,7], **this pioneering work is the first attempt to adapt the large pre-trained segmentation model SAM2 to specific downstream tasks and achieve new SOTA performance**. SAM2-Adapter builds on the strengths of the original SAM-Adapter while introducing significant advancements.

SAM2-Adapter inherits the core advantages of SAM-Adapter, including:

- **Generalizability**: SAM2-Adapter can be directly applied to customized datasets of various tasks, enhancing performance with minimal additional data. This flexibility ensures that the model can adapt to a wide range of applications, from medical imaging to environmental monitoring.
- **Composability**: SAM2-Adapter supports the easy integration of multiple conditions to fine-tune SAM2, improving task-specific outcomes. This composability allows for the combination of different adaptation strategies to meet the specific requirements of diverse downstream tasks.

SAM2-Adapter enhances these benefits by adapting to SAM2's multi-resolution hierarchical Transformer architecture. By employing multiple adapters working in tandem, SAM2-Adapter effectively leverages SAM2's multi-resolution and hierarchical features for more precise and robust segmentation, which maximizes the potential of the already-powerful SAM2. We perform extensive experiments on multiple tasks and datasets, including ISTD for shadow detection [8] and COD10K [9], CHAMELEON [10], CAMO [11] for camouflaged object detection task, and kvasir-SEG [12] for polyp segmentation (medical image segmentation) task. Benefiting from the capability of SAM2 and our SAM-Adapter, our method achieves state-of-the-art (SOTA) performance on both tasks. The contributions of this work can be summarized as follows:

- We are the first to identify and analyze the limitations of the Segment Anything 2 (SAM2) model in specific downstream tasks, continuing our research from SAM.
- Second, we are the first to propose the adaptation approach, **SAM2-Adapter**, to adapt SAM2 to downstream tasks and achieve enhanced performance. This method effectively integrates task-specific knowledge with the general knowledge learned by the large model.
- Third, despite SAM2's backbone being a simple plain model lacking specialized structures tailored for the specific downstream tasks, our extensive experiments demonstrate that SAM2-Adapter achieves SOTA results on challenging segmentation tasks, setting new benchmarks and proving its effectiveness in diverse applications.

By further building upon the success of the SAM-Adapter, the SAM2-Adapter inherents the advantages of SAM-Adapter and demonstrates the exceptional ability of the SAM2 model to transfer its knowledge to specific data domains, pushing the boundaries of what is possible in downstream segmentation tasks. We encourage the research community to adopt SAM2 as the backbone in conjunction with our SAM2-Adapter, to achieve even better segmentation results in various research fields and industrial applications. We are releasing our code, pre-trained model, and data processing protocols in http://tianrun-chen.github.io/SAM-Adaptor/.

 doi:10.20944/preprints202408.0622.v1

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) [13], encoder-decoder structures [14–19], dilated convolutions [20–25], pyramid structures [22,23,26–29], attention modules [30–34], and transformers [2,35–38]. Recent advancements have improved SAM's performance, such as [39], which introduces a High-Quality output token and trains the model on fine-grained masks. Other efforts have focused on enhancing SAM's efficiency for broader real-world and mobile use, exemplified by [40–42]. The widespread success of SAM has led to its adoption in various fields, including medical imaging [43–46], remote sensing [47,48], motion segmentation [49], and camouflaged object detection [50]. Notably, our previous work SAM-Adapter [6,7] tested camouflaged object detection, polyp segmentation, and shadow segmentation, and provide with the first adapter-based method to integrate the SAM's exceptional capability to these downstream tasks.

Adapters. The concept of Adapters was first introduced in the NLP community [51] as a tool to fine-tune a large pre-trained model for each downstream task with a compact and scalable model. In [52], multi-task learning was explored with a single BERT model shared among a few task-specific parameters. In the computer vision community, [53] suggested fine-tuning the ViT [54] for object detection with minimal modifications. Recently, ViT-Adapter [55] leveraged Adapters to enable a plain ViT to perform various downstream tasks. [56] introduce an Explicit Visual Prompting (EVP) technique that can incorporate explicit visual cues to the Adapter. However, no prior work has tried to apply Adapters to leverage pretrained image segmentation model SAM trained at large image corpus. Here, we mitigate the research gap.

Polyp Segmentation. In recent years, there has been notable progress in polyp segmentation [57] due to deep-learning approaches. These techniques employ deep neural networks to derive more discriminative features from endoscopic polyp images. Nonetheless, the use of bounding-box detectors often leads to inaccurate polyp boundary localization. To resolve this, [58] leveraged fully convolutional networks (FCN) with pre-trained models to identify and segment polyps. [59] introduced a technique utilizing Fully Convolutional Neural Networks (FCNNs) to predict 2D Gaussian shapes. Subsequently, the U-Net [60] architecture, featuring a contracting path for context capture and a symmetric expanding path for precise localization, achieved favorable segmentation results. However, these strategies focus primarily on entire polyp regions, neglecting boundary constraints. Therefore, Psi-Net [61] incorporated both region and boundary constraints for polyp segmentation, yet the interplay between regions and boundaries remained underexplored. [62] introduced PolypSegNet, an enhanced encoder-decoder architecture designed for the automated segmentation of polyps in colonoscopy images. To address the issue of non-equivalent images and pixels, [63] proposed a confidence-aware resampling method for polyp segmentation tasks. Specifically for polyp segmentation, works done by [64] and [6] present promising results using an unprompted SAM and a domain-adapted SAM respectively. Additionally, Polyp-SAM [65] used SAM for the same task. [66] evaluated the zero-shot capabilities of SAM on the organ segmentation task.

Camouflaged Object Detection (COD). Camouflaged object detection, or concealed object detection is a challenging but useful task that identifies objects blend in with their surroundings. COD has wide applications in medicine, agriculture, and art. Initially, researches of camouflage detection relied on low-level features like texture, brightness, and color [67–70] to distinguish foreground from background. It is worth noting that some of these prior knowledge is critical in identifying the objects, and is used to guide the neural network in this paper.

Le et al.[11] first proposed an end-to-end network consisting of a classification and a segmentation branch. Recent advances in deep learning-based methods have shown a superior ability to detect complex camouflaged objects [9,71,72]. In this work, we leverage the advanced neural network backbone (a foundation model – SAM2) with the input of task-specific prior knowledge to achieve the state-of-the-art (SOTA) performance.

 doi:10.20944/preprints202408.0622.v1

Shadow Detection. Shadows can occur when an object surface is not directly exposed to light. They offer hints on light source direction and scene illumination that can aid scene comprehension [73,74]. They can also negatively impact the performance of computer vision tasks [75,76]. Early method use hand-crafted heuristic cues like chromacity, intensity and texture [74,77,78]. Deep learning approaches leverage the knowledge learnt from data and use delicately designed neural network structure to capture the information (e.g. learned attention modules) [79–81]. This work leverage the heuristic priors with large neural network models to achieve the state-of-the-art (SOTA) performance.

3. Method

3.1. Using SAM 2 as the Backbone

The core of our SAM2-Adapter is built upon the powerful image encoder and mask decoder components of the SAM2 model. Specifically, we leverage the MAE pre-trained Hiera image encoder from SAM2, keeping its weights frozen to preserve the rich visual representations it has learned from pretraining on large-scale datasets. Additionally, we utilize the mask decoder module from the original SAM2 model, initializing its weights with the pretrained SAM2 parameters and then fine-tuning it during the training of our adapter. We do not provide any additional prompts as input to the original SAM2 mask decoder.

Similar to the successful approach of the SAM-Adapter [6], we next learn and inject task-specific knowledge F^i into the network via Adapters. We employ the concept of prompting, which utilizes the fact that foundation models like SAM2 have been trained on large-scale datasets. Using appropriate prompts to introduce task-specific knowledge [56] can enhance the model's generalization ability on downstream tasks, especially when annotated data is scarce.

The architecture of the proposed SAM2-Adapter is illustrated in Figure 1. We aim to keep the design of the adapter to be simple and efficient. Therefore, we choose to use an adapter that consists of only two MLPs and an activate function within two MLPs [56]. It is worth noting that the different from SAM[5], the image encoder of SAM2 has four stages with hierarchical resolutions. Therefore, we initialized four different adapter and insert the four adapter in different layers of each stage. In each stage, the weight of the adapter is shared. Specifically, each of the adapter takes the information F^i and obtains the prompt P^i:

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

in which MLP^i_{tune} are linear layers used to generate task-specific prompts for each Adapter. MLP_{up} is an up-projection layer shared across all Adapters 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 [82]. The information F^i can be chosen to be in various forms.

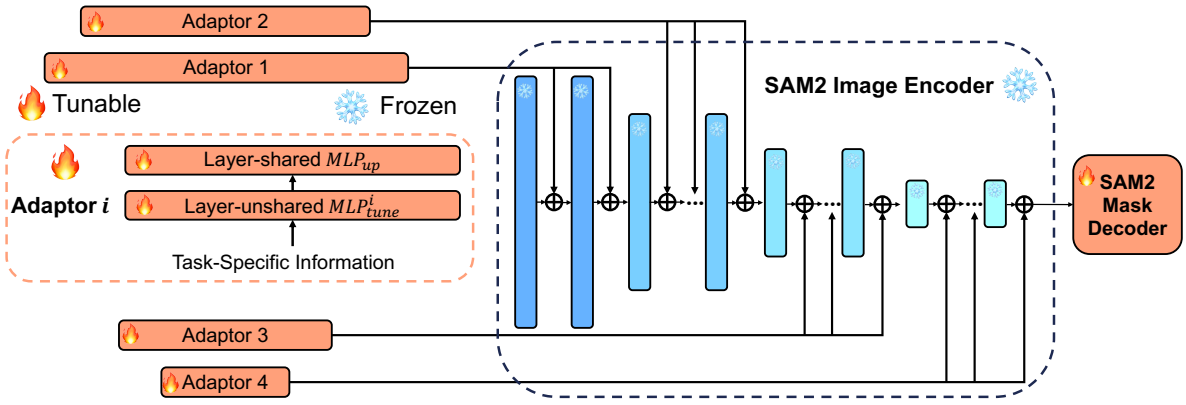

Figure 1. The architecture of the proposed SAM-Adapter.

For more information, please refer to the original SAM-Adapter paper [6].

 doi:10.20944/preprints202408.0622.v1

3.2. 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 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. For more information, please refer to the original SAM-Adapter paper [6].

4. Experiments

4.1. Tasks and Datasets

In our experiments, we selected two challenging low-level structural segmentation tasks and one medical imaging task to evaluate the performance of the SAM2-Adapter: camouflaged object detection and shadow detection, and polyp segmentation.

For the camouflaged object detection task, we utilized three prominent datasets: COD10K [9], CHAMELEON [10], and CAMO [11]. COD10K is the largest dataset for camouflaged object detection, containing 3,040 training and 2,026 testing samples. CHAMELEON includes 76 images collected from the internet for testing. The CAMO dataset consists of 1,250 images, with 1,000 for training and 250 for testing. Following the training protocol in [9], we used the combined dataset of CAMO and the training set of COD10K for model training. For evaluation, we used the test sets of CAMO and COD10K, as well as the entire CHAMELEON dataset. For the shadow detection task, we employed the ISTD dataset [8], which contains 1,330 training images and 540 test images.For polyp segmentation (medical image segmentation), we use the kvasir-SEG dataset [12]. The train-test split followed the settings of the Medico multimedia task at MediaEval 2020: Automatic Polyp Segmentation [83].

For evaluation metrics, we followed the protocol in [56] and used commonly-used metrics such as S-measure (S_m), mean E-measure (E_ϕ), and MAE for the camouflaged object detection task. For the shadow detection task, we used the balance error rate (BER) metric. For the polyp segmentation task, we used mean Dice score (mDice) and mean Intersection-over-Union (mIoU) as the evaluation measures.

For more details, please refer to the original SAM-Adapter paper [6].

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 [56], 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 one linear layer 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 hiera-large version of SAM2. Balanced BCE loss is used for shadow detection. BCE loss and IOU loss are used for camouflaged object detection and polyp segmentation. 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 of camouflaged object segmentation is performed for 20 epochs. Shadow segmentation is trained for 90 epochs. Polyp segmentation is trained for 20 epochs. The experiments are implemented using PyTorch on three NVIDIA Tesla A100 GPUs. For more information, please refer to the original SAM-Adapter paper [6] and our codebase.

 doi:10.20944/preprints202408.0622.v1

4.3. Experiments for Camouflaged Object Detection

We first evaluated SAM on the challenging task of camouflaged object detection, where foreground objects often blend with visually similar background patterns. Our experiments revealed that SAM did not perform well in this task. As shown in Figure 2, SAM failed to detect several concealed objects. This was further confirmed by the quantitative results presented in Table 1, where SAM's performance was significantly lower than existing state-of-the-art methods across all evaluated metrics, while SAM2, on its own, had the lowest performance, which fails to produce any meaningful results.

In contrast, Figure 3 clearly demonstrates that by introducing the SAM2-Adapter, our method significantly elevates the model's performance. 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 methods.

Furthermore, the SAM2-Adapter set a new SOTA performance. Visualized results show that SAM2-Adapter segments more precisely without adding extra false information, further demonstrating the robustness and accuracy of our approach.

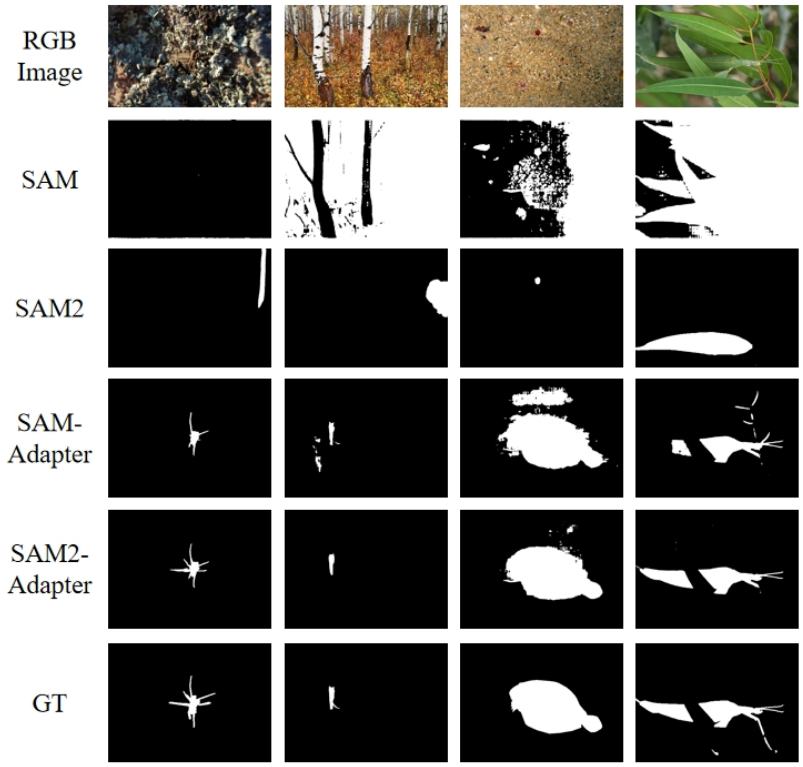

Figure 2. Shadow Detection Visualization As shown in the figure, SAM often fails to detect animals that are visually camouflaged within their natural environments and can sometimes produce irrelevant results. SAM2 also struggles with similar issues and produce non-meaningful outcomes. However, by incorporating SAM-Adapter, our approach significantly improves object segmentation performance. Furthermore, SAM2-Adapter demonstrates even better performance than SAM-Adapter. The samples depicted are from the CHAMELEON dataset.

 doi:10.20944/preprints202408.0622.v1

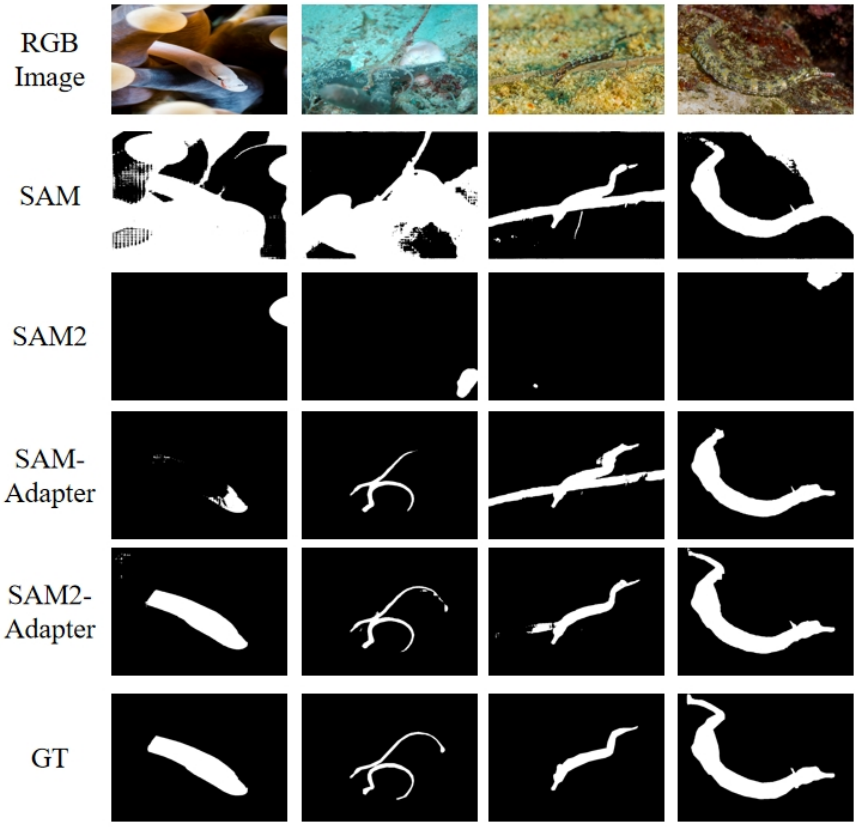

Figure 3. Visualization for Camouflaged Image Segmentation in COD-10K dataset As shown in the figure, SAM struggles to detect animals that are visually camouflaged within their natural environments and can sometimes produce results that lack meaningful segmentation. SAM2 also faces similar challenges, often resulting in no output or false results. However, by incorporating SAM2-Adapter, our method significantly improves object segmentation performance, surpassing SAM-Adapter. For other dataset, please refer to *More Results* section.

Table 1. Quantitative Segmentation Result Comparison for Camouflaged Object Detection

Method	CHAMELEON [10]				CAMO [11]				COD10K [9]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$
SINet[84]	0.869	0.891	0.740	0.440	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051
RankNet[85]	0.846	0.913	0.767	0.045	0.712	0.791	0.583	0.104	0.767	0.861	0.611	0.045
JCOD [86]	0.870	0.924	-	0.039	0.792	0.839	-	0.82	0.800	0.872	-	0.041
PFNet [87]	0.882	0.942	0.810	0.330	0.782	0.852	0.695	0.085	0.800	0.868	0.660	0.040
FBNet [88]	0.888	0.939	0.828	0.032	0.783	0.839	0.702	0.081	0.809	0.889	0.684	0.035
SAM [5]	0.727	0.734	0.639	0.081	0.684	0.687	0.606	0.132	0.783	0.798	0.701	0.050
SAM2 [89]	0.359	0.375	0.115	0.357	0.350	0.411	0.079	0.311	0.429	0.505	0.115	0.218
SAM-Adapter [6,7]	0.896	0.919	0.824	0.033	0.847	0.873	0.765	0.070	0.883	0.918	0.801	0.025
SAM2-Adapter (Ours)	**0.915**	**0.955**	**0.889**	**0.018**	**0.855**	**0.909**	**0.810**	**0.051**	**0.899**	**0.950**	**0.850**	**0.018**

4.4. Experiments for Shadow Detection

We also evaluated SAM on shadow detection. However, as depicted in Figure 4, SAM struggled to differentiate between the shadow and the background, with parts missing or mistakenly added.

Similarly, SAM2 also struggled with the "shadow" concept without proper prompting, failing to produce meaningful results. In our study, we compared various methods for shadow detection and found that SAM's performance was significantly poorer than existing methods. However, by integrating the SAM-Adapter, we achieved a substantial improvement in performance. The SAM-Adapter enhanced the detection of shadow regions, making them more clearly identifiable. Furthermore, SAM2-Adapter worked just as effectively as SAM-Adapter, delivering comparable results. Our findings were validated through quantitative analysis, and Table 2 demonstrates the significant performance boost provided by the SAM-Adapter and matched by the SAM2-Adapter for shadow detection.

 doi:10.20944/preprints202408.0622.v1

Table 2. Result for Shadow Detection

Method	BER ↓
Stacked CNN [90]	8.60
BDRAR [91]	2.69
DSC [92]	3.42
DSD [93]	2.17
FDRNet [94]	1.55
SAM [5]	40.51
SAM2 [89]	50.81
SAM-Adapter	**1.43**
SAM2-Adapter (Ours)	**1.43**

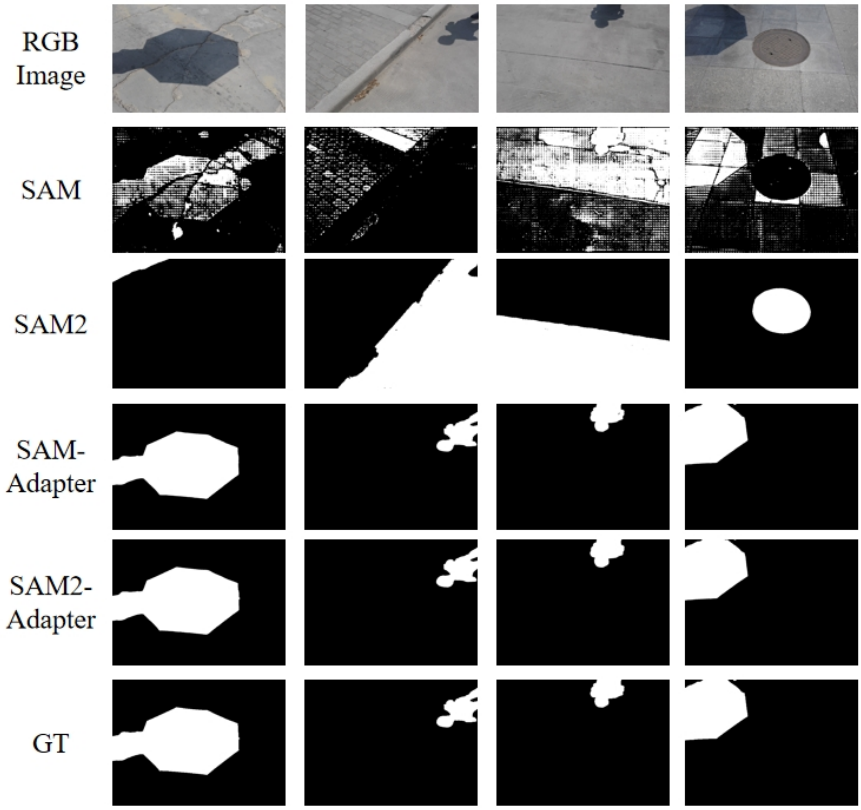

Figure 4. Shadow Detection Visualized. Both SAM and SAM2 have no understanding about the "shadow" concept without proper prompting. They produce meaningless results. SAM-Adapter and SAM2-Adapter perform equally well in shadow detection tasks.

4.5. Experiments for Polyp Segmentation

We illustrate the application of SAM2-Adapter in the context of medical image segmentation, specifically focusing on polyp segmentation. Polyps, which have the potential to become malignant, are identified during colonoscopy and removed through polypectomy. Accurate and swift detection and removal of polyps are crucial in preventing colorectal cancer, a leading cause of cancer-related deaths globally.

While numerous deep learning approaches have been developed for polyp identification, and the pre-trained SAM model shows promise in identifying some polyps, its performance can be significantly improved with our SAM-Adapter approach. However, without proper prompting, the SAM2 model fails to produce meaningful results. Our SAM2-Adapter addresses this issue and outperforms the original SAM-Adapter. The results of our study, presented in Table 3 and the visualization results in

 doi:10.20944/preprints202408.0622.v1

Figure 6, underscore the effectiveness of SAM2-Adapter in improving the accuracy and reliability of polyp detection.

Table 3. Quantitative Result for Polyp Segmentation

Method	mDice ↑	mIoU ↑
UNet [14]	0.821	0.756
UNet++ [95]	0.824	0.753
SFA [96]	0.725	0.619
SAM [5]	0.778	0.707
SAM2 [89]	0.200	0.029
SAM-Adapter	0.850	0.776
SAM2-Adapter (Ours)	**0.873**	**0.806**

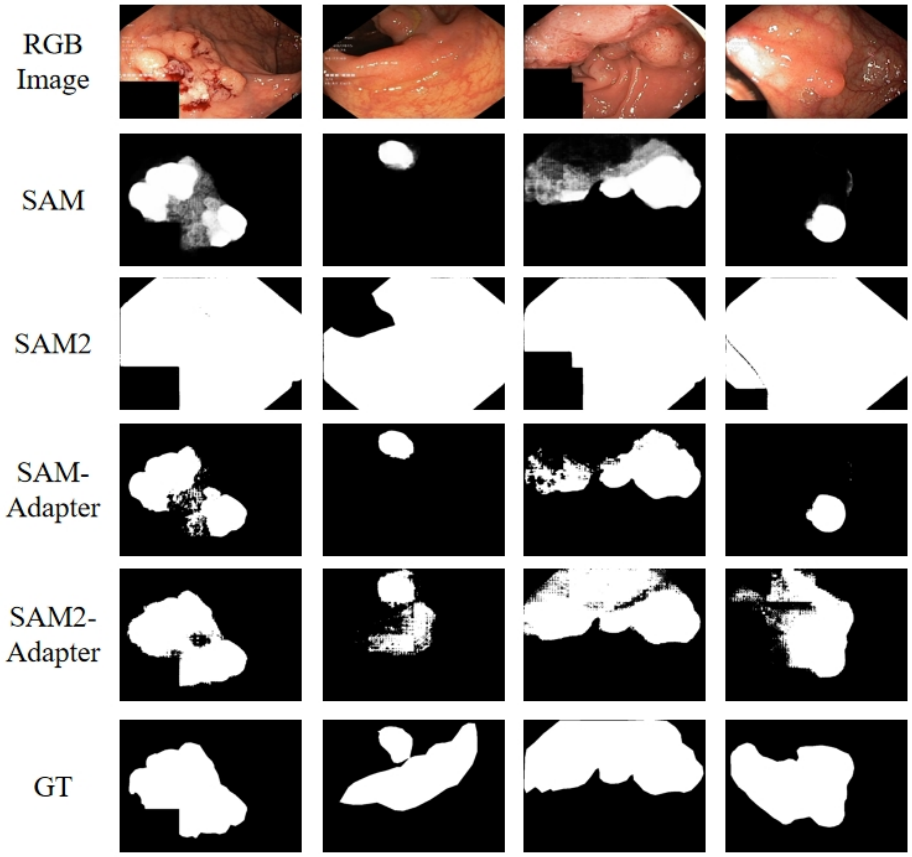

Figure 5. Visualization of Polyp Segmentation Results. As illustrated in the figure, although SAM can identify some polyp structures in the image, the result is not accurate. Without proper prompting, SAM 2 failed to deliver meaningful polyp segmentation results. By using SAM2-Adapter, our approach significantly outperform SAM-Adapter with more accurate (and complete) segmentation results.

5. Conclusion and Future Work

In this paper, we introduced SAM2-Adapter, a novel adaptation method designed to leverage the advanced capabilities of the Segment Anything 2 (SAM2) model for specific downstream segmentation tasks. Building on the success of the original SAM-Adapter, SAM2-Adapter utilizes a multi-adapter configuration that is specifically tailored to SAM2's multi-resolution hierarchical Transformer architecture. This approach effectively addresses the limitations encountered with SAM, enabling the achievement of new state-of-the-art (SOTA) performance in challenging segmentation tasks such as camouflaged object detection, shadow detection, and polyp segmentation.

 doi:10.20944/preprints202408.0622.v1

Our experiments demonstrate that SAM2-Adapter not only retains the beneficial features of its predecessor, including generalizability and composability but also enhances these capabilities by integrating seamlessly with SAM2's advanced architecture. This integration allows SAM2-Adapter to outperform previous methods and set new benchmarks across various datasets and tasks.

The continued presence of challenges from SAM in SAM2 highlights the inherent complexities of applying foundation models to diverse real-world scenarios. Nevertheless, SAM2-Adapter effectively addresses these issues, showcasing its potential as a robust tool for high-quality segmentation in a range of applications.

We encourage researchers and engineers to adopt SAM2 as the backbone for their segmentation tasks, coupled with SAM2-Adapter, to realize improved performance and advance the field of image segmentation. Our work not only extends the capabilities of SAM2 but also paves the way for future innovations in adapting large pre-trained models for specialized applications.

6. More Results

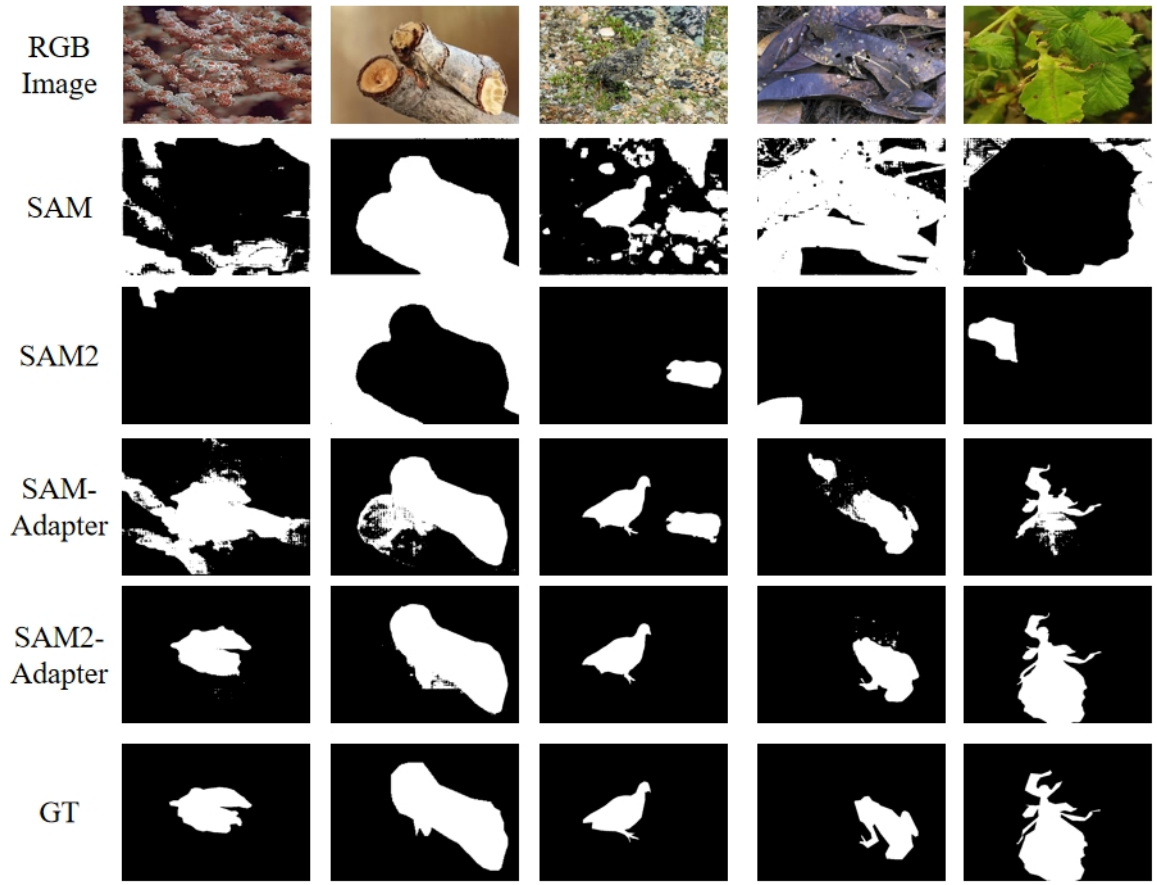

Figure 6. Camouflaged Segmentation of CAMO dataset. The SAM and SAM 2 failed to perceive those animals that are visually 'hidden'/concealed in their natural surroundings. By using SAM-Adapter, our approach can significantly elevate the performance of object segmentation with SAM.

References

1. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; others. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* **2021**.
2. Zhu, L.; Chen, T.; Ji, D.; Ye, J.; Liu, J. LLaFS: When Large Language Models Meet Few-Shot Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3065–3075.

 doi:10.20944/preprints202408.0622.v1

3. Zhu, L.; Ji, D.; Chen, T.; Xu, P.; Ye, J.; Liu, J. Ibd: Alleviating hallucinations in large vision-language models via image-biased decoding. *arXiv preprint arXiv:2402.18476* **2024**.
4. Chen, T.; Yu, C.; Li, J.; Zhang, J.; Zhu, L.; Ji, D.; Zhang, Y.; Zang, Y.; Li, Z.; Sun, L. Reasoning3D–Grounding and Reasoning in 3D: Fine-Grained Zero-Shot Open-Vocabulary 3D Reasoning Part Segmentation via Large Vision-Language Models. *arXiv preprint arXiv:2405.19326* **2024**.
5. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; others. Segment anything. *arXiv preprint arXiv:2304.02643* **2023**.
6. Chen, T.; Zhu, L.; Ding, C.; Cao, R.; Wang, Y.; Li, Z.; Sun, L.; Mao, P.; Zang, Y. SAM Fails to Segment Anything? – SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More, 2023, [arXiv:cs.CV/2304.09148].
7. Chen, T.; Zhu, L.; Deng, C.; Cao, R.; Wang, Y.; Zhang, S.; Li, Z.; Sun, L.; Zang, Y.; Mao, P. Sam-adapter: Adapting segment anything in underperformed scenes. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3367–3375.
8. Wang, J.; Li, X.; Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1788–1797.
9. Fan, D.P.; Ji, G.P.; Sun, G.; Cheng, M.M.; Shen, J.; Shao, L. Camouflaged object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2777–2787.
10. Skurowski, P.; Abdulameer, H.; Błaszczyk, J.; Depta, T.; Kornacki, A.; Kozieł, P. Animal camouflage analysis: Chameleon database. *Unpublished manuscript* **2018**, *2*, 7.
11. Le, T.N.; Nguyen, T.V.; Nie, Z.; Tran, M.T.; Sugimoto, A. Anabranch network for camouflaged object segmentation. *Computer vision and image understanding* **2019**, *184*, 45–56.
12. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Halvorsen, P.; de Lange, T.; Johansen, D.; Johansen, H.D. Kvasir-seg: A segmented polyp dataset. MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26. Springer, 2020, pp. 451–462.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
15. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 325–341.
16. Fan, M.; Lai, S.; Huang, J.; Wei, X.; Chai, Z.; Luo, J.; Wei, X. Rethinking BiSeNet For Real-time Semantic Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9716–9725.
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *39*, 2481–2495.
18. Li, X.; You, A.; Zhu, Z.; Zhao, H.; Yang, M.; Yang, K.; Tan, S.; Tong, Y. Semantic Flow for Fast and Accurate Scene Parsing. European Conference on Computer Vision. Springer, 2020, pp. 775–793.
19. Chen, T.; Ding, C.; Zhu, L.; Xu, T.; Ji, D.; Zang, Y.; Li, Z. xLSTM-UNet can be an Effective 2D\& 3D Medical Image Segmentation Backbone with Vision-LSTM (ViL) better than its Mamba Counterpart. *arXiv preprint arXiv:2407.01530* **2024**.
20. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062* **2014**.
21. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *40*, 834–848.
22. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* **2017**.
23. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

 doi:10.20944/preprints202408.0622.v1

24. Liu, Z.; Zhu, L. Label-guided attention distillation for lane segmentation. *Neurocomputing* **2021**, *438*, 312–322.
25. Zang, Y.; Fu, C.; Cao, R.; Zhu, D.; Zhang, M.; Hu, W.; Zhu, L.; Chen, T. Resmatch: Referring expression segmentation in a semi-supervised manner. *arXiv preprint arXiv:2402.05589* **2024**.
26. Zhu, L.; Ji, D.; Zhu, S.; Gan, W.; Wu, W.; Yan, J. Learning Statistical Texture for Semantic Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12537–12546.
27. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
28. Zhu, L.; Chen, T.; Yin, J.; See, S.; Liu, J. Continual Semantic Segmentation with Automatic Memory Sample Selection. *arXiv preprint arXiv:2304.05015* **2023**.
29. Fu, X.; Zhang, S.; Chen, T.; Lu, Y.; Zhu, L.; Zhou, X.; Geiger, A.; Liao, Y. Panoptic nerf: 3d-to-2d label transfer for panoptic urban scene segmentation. *arXiv preprint arXiv:2203.15224* **2022**.
30. Zhang, F.; Chen, Y.; Li, Z.; Hong, Z.; Liu, J.; Ma, F.; Han, J.; Ding, E. Acfnet: Attentional class feature network for semantic segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6798–6807.
31. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
32. Zhu, Z.; Xu, M.; Bai, S.; Huang, T.; Bai, X. Asymmetric non-local neural networks for semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 593–602.
33. Zhu, L.; Chen, T.; Yin, J.; See, S.; Liu, J. Addressing Background Context Bias in Few-Shot Segmentation through Iterative Modulation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3370–3379.
34. Zhu, L.; Chen, T.; Yin, J.; See, S.; Liu, J. Learning gabor texture features for fine-grained recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1621–1631.
35. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; others. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 6881–6890.
36. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems* **2021**, *34*, 12077–12090.
37. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 7262–7272.
38. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1290–1299.
39. Ke, L.; Ye, M.; Danelljan, M.; Liu, Y.; Tai, Y.W.; Tang, C.K.; Yu, F. Segment Anything in High Quality, 2023, [arXiv:cs.CV/2306.01567].
40. Xiong, Y.; Varadarajan, B.; Wu, L.; Xiang, X.; Xiao, F.; Zhu, C.; Dai, X.; Wang, D.; Sun, F.; Iandola, F.; Krishnamoorthi, R.; Chandra, V. EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything, 2023, [arXiv:cs.CV/2312.00863].
41. Zhang, C.; Han, D.; Qiao, Y.; Kim, J.U.; Bae, S.H.; Lee, S.; Hong, C.S. Faster Segment Anything: Towards Lightweight SAM for Mobile Applications, 2023, [arXiv:cs.CV/2306.14289].
42. Zhao, X.; Ding, W.; An, Y.; Du, Y.; Yu, T.; Li, M.; Tang, M.; Wang, J. Fast Segment Anything, 2023, [arXiv:cs.CV/2306.12156].
43. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nature Communications* **2024**, *15*. doi:10.1038/s41467-024-44824-z.
44. Deng, R.; Cui, C.; Liu, Q.; Yao, T.; Remedios, L.W.; Bao, S.; Landman, B.A.; Wheless, L.E.; Coburn, L.A.; Wilson, K.T.; Wang, Y.; Zhao, S.; Fogo, A.B.; Yang, H.; Tang, Y.; Huo, Y. Segment Anything Model (SAM) for Digital Pathology: Assess Zero-shot Segmentation on Whole Slide Imaging, 2023, [arXiv:eess.IV/2304.04155].
45. Mazurowski, M.A.; Dong, H.; Gu, H.; Yang, J.; Konz, N.; Zhang, Y. Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis* **2023**, *89*, 102918. doi:10.1016/j.media.2023.102918.
46. Wu, J.; Ji, W.; Liu, Y.; Fu, H.; Xu, M.; Xu, Y.; Jin, Y. Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation, 2023, [arXiv:cs.CV/2304.12620].

 doi:10.20944/preprints202408.0622.v1

47. Chen, K.; Liu, C.; Chen, H.; Zhang, H.; Li, W.; Zou, Z.; Shi, Z. RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation based on Visual Foundation Model, 2023, [arXiv:cs.CV/2306.16269].
48. Ren, S.; Luzi, F.; Lahrichi, S.; Kassaw, K.; Collins, L.M.; Bradbury, K.; Malof, J.M. Segment anything, from space?, 2023, [arXiv:cs.CV/2304.13000].
49. Xie, J.; Yang, C.; Xie, W.; Zisserman, A. Moving Object Segmentation: All You Need Is SAM (and Flow), 2024, [arXiv:cs.CV/2404.12389].
50. Tang, L.; Xiao, H.; Li, B. Can SAM Segment Anything? When SAM Meets Camouflaged Object Detection, 2023, [arXiv:cs.CV/2304.04709].
51. Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; Gelly, S. Parameter-efficient transfer learning for NLP. International Conference on Machine Learning. PMLR, 2019, pp. 2790–2799.
52. Stickland, A.C.; Murray, I. Bert and pals: Projected attention layers for efficient adaptation in multi-task learning. International Conference on Machine Learning. PMLR, 2019, pp. 5986–5995.
53. Li, Y.; Mao, H.; Girshick, R.; He, K. Exploring plain vision transformer backbones for object detection. Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX. Springer, 2022, pp. 280–296.
54. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; others. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* **2020**.
55. Chen, Z.; Duan, Y.; Wang, W.; He, J.; Lu, T.; Dai, J.; Qiao, Y. Vision transformer adapter for dense predictions. *arXiv preprint arXiv:2205.08534* **2022**.
56. Liu, W.; Shen, X.; Pun, C.M.; Cun, X. Explicit Visual Prompting for Low-Level Structure Segmentations. *arXiv preprint arXiv:2303.10883* **2023**.
57. Zhou, Y.; Wang, H.; Huo, S.; Wang, B. Full-attention based Neural Architecture Search using Context Auto-regression, 2021, [arXiv:cs.CV/2111.07139].
58. Canny, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* **1986**, pp. 679–698.
59. Qadir, H.A.; Shin, Y.; Solhusvik, J.; Bergsland, J.; Aabakken, L.; Balasingham, I. Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction. *Medical Image Analysis* **2021**, *68*, 101897.
60. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, 2017, [arXiv:cs.LG/1412.6980].
61. Murugesan, B.; Sarveswaran, K.; Shankaranarayana, S.M.; Ram, K.; Sivaprakasam, M. Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation, 2019, [arXiv:cs.CV/1902.04099].
62. Mahmud, T.; Paul, B.; Fattah, S.A. PolypSegNet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images. *Computers in biology and medicine* **2021**, *128*, 104119.
63. Guo, X.; Chen, Z.; Liu, J.; Yuan, Y. Non-equivalent images and pixels: Confidence-aware resampling with meta-learning mixup for polyp segmentation. *Medical image analysis* **2022**, *78*, 102394.
64. Zhou, T.; Zhang, Y.; Zhou, Y.; Wu, Y.; Gong, C. Can SAM Segment Polyps?, 2023, [arXiv:cs.CV/2304.07583].
65. Li, Y.; Hu, M.; Yang, X. Polyp-SAM: Transfer SAM for Polyp Segmentation, 2023, [arXiv:eess.IV/2305.00293].
66. Roy, S.; Wald, T.; Koehler, G.; Rokuss, M.R.; Disch, N.; Holzschuh, J.; Zimmerer, D.; Maier-Hein, K.H. SAM.MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model, 2023, [arXiv:eess.IV/2304.05396].
67. Feng, X.; Guoying, C.; Wei, S. Camouflage texture evaluation using saliency map. Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, 2013, pp. 93–96.
68. Pike, T.W. Quantifying camouflage and conspicuousness using visual salience. *Methods in Ecology and Evolution* **2018**, *9*, 1883–1895.
69. Hou, J.Y.Y.H.W.; Li, J. Detection of the mobile object with camouflage color under dynamic background based on optical flow. *Procedia Engineering* **2011**, *15*, 2201–2205.
70. Sengottuvelan, P.; Wahi, A.; Shanmugam, A. Performance of decamouflaging through exploratory image analysis. 2008 First International Conference on Emerging Trends in Engineering and Technology. IEEE, 2008, pp. 6–10.
71. Mei, H.; Ji, G.P.; Wei, Z.; Yang, X.; Wei, X.; Fan, D.P. Camouflaged object segmentation with distraction mining. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8772–8781.

 doi:10.20944/preprints202408.0622.v1

72. Lin, J.; Tan, X.; Xu, K.; Ma, L.; Lau, R.W. Frequency-aware camouflaged object detection. *ACM Transactions on Multimedia Computing, Communications and Applications* **2023**, *19*, 1–16.
73. Karsch, K.; Hedau, V.; Forsyth, D.; Hoiem, D. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (TOG)* **2011**, *30*, 1–12.
74. Lalonde, J.F.; Efros, A.A.; Narasimhan, S.G. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision* **2012**, *98*, 123–145.
75. Nadimi, S.; Bhanu, B. Physical models for moving shadow and object detection in video. *IEEE transactions on pattern analysis and machine intelligence* **2004**, *26*, 1079–1087.
76. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence* **2003**, *25*, 1337–1342.
77. Huang, X.; Hua, G.; Tumblin, J.; Williams, L. What characterizes a shadow boundary under the sun and sky? 2011 international conference on computer vision. IEEE, 2011, pp. 898–905.
78. Zhu, J.; Samuel, K.G.; Masood, S.Z.; Tappen, M.F. Learning to recognize shadows in monochromatic natural images. 2010 IEEE Computer Society conference on computer vision and pattern recognition. IEEE, 2010, pp. 223–230.
79. Le, H.; Vicente, T.F.Y.; Nguyen, V.; Hoai, M.; Samaras, D. A+ D Net: Training a shadow detector with adversarial shadow attenuation. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 662–678.
80. Cun, X.; Pun, C.M.; Shi, C. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, Vol. 34, pp. 10680–10687.
81. Zhu, L.; Deng, Z.; Hu, X.; Fu, C.W.; Xu, X.; Qin, J.; Heng, P.A. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 121–136.
82. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415* **2016**.
83. Jha, D.; Hicks, S.A.; Emanuelsen, K.; Johansen, H.; Johansen, D.; de Lange, T.; Riegler, M.A.; Halvorsen, P. Medico multimedia task at mediaeval 2020: Automatic polyp segmentation. *arXiv preprint arXiv:2012.15244* **2020**.
84. Fan, D.P.; Ji, G.P.; Sun, G.; Cheng, M.M.; Shen, J.; Shao, L. Camouflaged object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2777–2787.
85. Lv, Y.; Zhang, J.; Dai, Y.; Li, A.; Liu, B.; Barnes, N.; Fan, D.P. Simultaneously localize, segment and rank the camouflaged objects. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 11591–11601.
86. Li, A.; Zhang, J.; Lv, Y.; Liu, B.; Zhang, T.; Dai, Y. Uncertainty-aware joint salient object and camouflaged object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10071–10081.
87. Mei, H.; Ji, G.P.; Wei, Z.; Yang, X.; Wei, X.; Fan, D.P. Camouflaged object segmentation with distraction mining. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8772–8781.
88. Lin, J.; Tan, X.; Xu, K.; Ma, L.; Lau, R.W. Frequency-aware camouflaged object detection. *ACM Transactions on Multimedia Computing, Communications and Applications* **2023**, *19*, 1–16.
89. Ravi, N.; Gabeur, V.; Hu, Y.T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; Mintun, E.; Pan, J.; Alwala, K.V.; Carion, N.; Wu, C.Y.; Girshick, R.; Dollár, P.; Feichtenhofer, C. SAM 2: Segment Anything in Images and Videos, 2024, [arXiv:cs.CV/2408.00714].
90. Vicente, T.F.Y.; Hou, L.; Yu, C.P.; Hoai, M.; Samaras, D. Large-scale training of shadow detectors with noisily-annotated shadow examples. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. Springer, 2016, pp. 816–832.
91. Zhu, L.; Deng, Z.; Hu, X.; Fu, C.W.; Xu, X.; Qin, J.; Heng, P.A. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 121–136.
92. Hu, X.; Zhu, L.; Fu, C.W.; Qin, J.; Heng, P.A. Direction-aware spatial context features for shadow detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7454–7462.

 doi:10.20944/preprints202408.0622.v1

93. Zheng, Q.; Qiao, X.; Cao, Y.; Lau, R.W. Distraction-aware shadow detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5167–5176.
94. Zhu, L.; Xu, K.; Ke, Z.; Lau, R.W. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4702–4711.
95. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*; Springer, 2018; pp. 3–11.
96. Fang, Y.; Chen, C.; Yuan, Y.; Tong, K.y. Selective feature aggregation network with area-boundary constraints for polyp segmentation. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Springer, 2019, pp. 302–310.