SAM3-Adapter: Efficient Adaptation of Segment Anything 3 for Camouflage Object Segmentation, Shadow Detection, and Medical Image Segmentation

Tianrun Chen[1,2*+] Runlong Cao[3+] Xinda Yu[4+] Lanyun Zhu[5] Chaotao Ding[1]

Deyi Ji [7] Cheng Chen[6] Qi Zhu[7] Chunyan Xu[3] Papa Mao[1]

Ying Zang[4*]

This work is based on the SAM-Adapter, which was originally released on April 14, 2023.
TL; DR. SAM3, enhanced with our proposed adapter, surpasses its predecessor as a backbone for segmentation and establishes new state-of-the-art (SOTA) results across a range of downstream tasks

[+] 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 Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, P.R. China.

[4]School of Information Engineering, Huzhou University, Huzhou, P.R. China.

[5] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

[6] College of Computing and Data Science, Nanyang Technological University, Singapore.

[7] School of Information Science and Technology, University of Science and Technology of China, P.R. China.

Project Page: http://tianrun-chen.github.io/SAM-Adaptor/

Abstract

The rapid rise of large-scale foundation models has reshaped the landscape of image segmentation, with models such as Segment Anything achieving unprecedented versatility across diverse vision tasks. However, previous generations—including SAM and its successor—still struggle with fine-grained, low-level segmentation challenges such as camouflaged object detection, medical image segmentation, cell image segmentation, and shadow detection. To address these limitations, we originally proposed SAM-Adapter in 2023, demonstrating substantial gains on these difficult scenarios. With the emergence of Segment Anything 3 (SAM3)—a more efficient and higher-performing evolution with a redesigned architecture and improved training pipeline—we revisit these long-standing challenges. In this work, we present SAM3-Adapter, the first adapter framework tailored for SAM3 that unlocks its full segmentation capability. SAM3-Adapter not only reduces computational overhead but also consistently surpasses both SAM and SAM2-based solutions, establishing new state-of-the-art results across multiple downstream tasks, including medical imaging, camouflaged (concealed) object segmentation, and shadow detection. Built upon the modular and composable design philosophy of the original SAM-Adapter, SAM3-Adapter provides stronger generalizability, richer task adaptability, and significantly improved segmentation precision. Extensive experiments confirm that integrating SAM3 with our adapter yields superior accuracy, robustness, and efficiency compared to all prior SAM-based adaptations. We hope SAM3-Adapter can serve as a foundation for future research and practical segmentation applications. Code, pre-trained models, and data processing pipelines are available at: http://tianrun-chen.github.io/SAM-Adaptor/.

1. Introduction

The AI research landscape has been revolutionized by foundation models trained on vast datasets [2, 11, 104, 106]. Among these, the Segment Anything (SAM) series [45] has become a cornerstone for image segmentation. Our prior work, SAM-Adapter [7, 8] and SAM2-Adapter [10] were pioneering efforts to bridge the gap between the first SAM's general capabilities and the nuanced demands of downstream tasks, a contribution widely adopted by the community.

Now, the release of Segment Anything 3 (SAM3) marks a new era. With its significantly scaled-up architecture and a more extensive training corpus, SAM3 offers a vastly superior foundation and unprecedented potential for segmentation tasks. This advancement shifts the fundamental research question from "how do we fix model limitations?" to "how do we fully unlock and channel the immense power of this scaled-up model for specialized applications?"

This paper provides a definitive answer. We introduce SAM3-Adapter, a highly efficient and synergistic adaptation method designed specifically to unleash the full capabilities of SAM3. We demonstrate that by pairing the powerful SAM3 backbone with our lightweight adapter, we can achieve new state-of-the-art (SOTA) performance on a diverse set of challenging downstream tasks, including medical image, camouflage, and shadow segmentation. This work is the first to prove that the scaling of SAM3, when properly harnessed, directly translates into breakthrough performance in these specialized domains.

Our SAM3-Adapter is engineered to be both Generalizable and Composable. It can be seamlessly applied to custom datasets with minimal data, and its components can be flexibly combined to meet diverse task requirements. Critically, it is tailored to SAM3's advanced hierarchical architecture, ensuring that every part of the powerful backbone is effectively utilized. This synergy allows SAM3-Adapter to not only achieve superior accuracy but also maintain remarkable parameter efficiency. We conduct extensive experiments on multiple benchmarks, including ISTD [82], COD10K [19], and Kvasir-SEG [32]. The results are unequivocal: the combination of SAM3 and SAM3-Adapter consistently outperforms all previous methods. Our contributions are:

- We are the first to demonstrate and unlock the latent potential of the scaled-up SAM3 model for specialized downstream tasks, showing that its advanced architecture provides a superior foundation for achieving SOTA performance.
- We propose SAM3-Adapter, a novel, parameter-efficient adaptation framework specifically designed to synergize with SAM3, effectively channeling its generalist power into specialist excellence.
- We establish new state-of-the-art results across multiple challenging segmentation benchmarks, proving that a generalist backbone like SAM3, when enhanced by our adapter, can outperform highly specialized models.

We advocate for adopting the SAM3 and SAM3-Adapter combination to push the frontiers of image segmentation in both research and industry. We encourage the research community to adopt SAM3 as the backbone in conjunction with our SAM3-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/.

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) [56], encoder-decoder structures [1, 9, 20, 35, 37, 41, 72], dilated convolutions [5, 6, 28, 55, 88], pyramid structures [5, 6, 24, 34, 83, 90, 100, 102], attention modules [36, 40, 103, 105, 107], and transformers [13, 77, 85, 93, 104]. Recent advancements have improved SAM's performance, such as [43], 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 [87, 89, 91]. The widespread success of SAM has led to its adoption in various fields, including medical imaging [16, 58, 59, 63, 84], remote sensing [4, 38, 71], motion segmentation [22, 80, 81, 86], and camouflaged object detection [78]. Notably, our previous work SAM-Adapter [7, 8] tested camouflaged object detection, polyp segmentation, and shadow segmentation, and provide 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 [27] as a tool to fine-tune a large pre-trained model for each downstream task with a compact and scalable model. In [76], multi-task learning was explored with a single BERT model shared among a few task-specific parameters. In the computer vision community, [33, 39, 50] suggested fine-tuning the ViT [17] for object detection with minimal modifications. Recently, ViT-Adapter [12] leveraged Adapters to enable a plain ViT to perform various downstream tasks. [54] 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 no-

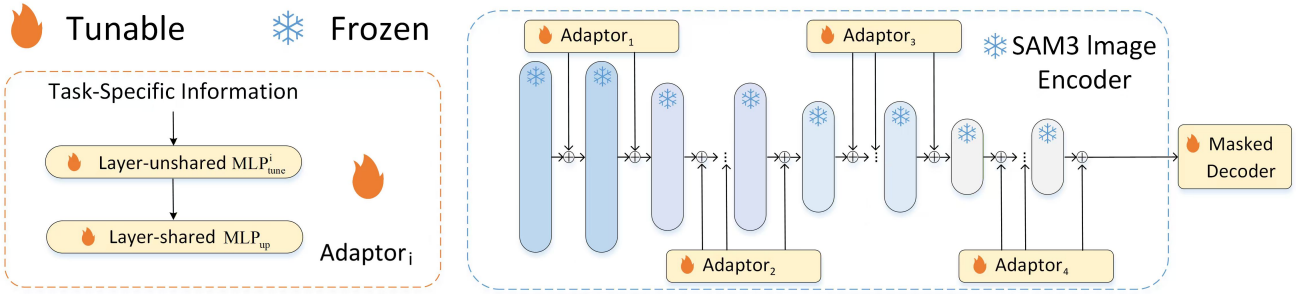

Figure 1. **The architecture of the proposed SAM3-Adapter.**

table progress in polyp segmentation [95] 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, [3] leveraged fully convolutional networks (FCN) with pre-trained models to identify and segment polyps. [69] introduced a technique utilizing Fully Convolutional Neural Networks (FCNNs) to predict 2D Gaussian shapes. Subsequently, the U-Net [44] 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 [66] 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, [25] proposed a confidence-aware resampling method for polyp segmentation tasks. Specifically for polyp segmentation, works done by [94] and [8] present promising results using an unprompted SAM and a domain-adapted SAM respectively. Additionally, Polyp-SAM [51] used SAM for the same task. [73] 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 that blend in with their surroundings. COD has wide applications in medicine, agriculture, and art. Initially, research of camouflage detection relied on low-level features like texture, brightness, and color [23, 26, 68, 74] to distinguish foreground from background. It is worth noting that some of this prior knowledge is critical in identifying the objects, and is used to guide the neural network in this paper.

Le et al.[48] 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 [19, 53, 65]. In this work, we leverage the advanced neural network backbone (a foundation model – SAM2) with the input of task-specific prior knowledge to achieve state-of-the-art (SOTA) performance.

Shadow Detection. Shadows can occur when an object's surface is not directly exposed to light. They offer hints on light source direction and scene illumination that can aid scene comprehension [42, 46]. They can also negatively impact the performance of computer vision tasks [14, 67]. Early methods use hand-crafted heuristic cues like chromaticity, intensity, and texture [30, 46, 97]. Deep learning approaches leverage the knowledge learned from data and use delicately designed neural network structures to capture the information (e.g. learned attention modules) [15, 47, 99]. This work leverages the heuristic priors with large neural network models to achieve the state-of-the-art (SOTA) performance.

3. Method

3.1. Using SAM 3 as the Backbone

The core of our approach is built upon the formidable vision backbone of the SAM3 model. SAM3 represents a significant architectural evolution, featuring a unified backbone shared between a DETR-based detector and a video tracker, designed to process complex visual and concept-based prompts.

In our work, we leverage this powerful, pre-trained vision encoder from SAM3. We keep its weights frozen during training. This strategy is crucial as it preserves the incredibly rich and generalizable visual representations learned from SAM3's extensive training on the massive SA-Co dataset. By doing so, we build upon a superior foundation without incurring the prohibitive costs of re-training the entire model. For the segmentation head, we utilize the mask decoder architecture from the SAM family, initializing it with pre-trained weights and subsequently fine-tuning it alongside our adapter.

While the SAM3 encoder provides a state-of-the-art foundation, unlocking its full potential for specialized domains requires a mechanism to inject task-specific knowledge. Following the successful paradigm of our previous work [8], we introduce lightweight adapters to achieve this.

3.2. SAM3-Adapter for Task-Specific Specialization

The architecture of our SAM3-Adapter is designed for simplicity and efficiency, as illustrated in Figure 1. The SAM3 vision encoder features a hierarchical, multi-stage architecture. To complement this, we introduce a set of adapters, one for each stage of the encoder. The weights of the adapter are shared within each stage to maintain parameter efficiency.

Specifically, each adapter processes task-specific information, F_i, to generate a conditioning prompt, P_i. This process is defined as:

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

where $\mathrm{MLP}tune^i$ is a tunable linear layer that creates a task-specific prompt from the input information F_i. The $\mathrm{MLP}up$ is an up-projection layer, shared across all adapters, that aligns the prompt's dimensions with the transformer features. The resulting prompt P^i is then integrated into the transformer layers of the corresponding stage, effectively guiding the model's focus toward task-relevant features.

3.3. Flexible Task-Specific Inputs

A key strength of our framework is the flexibility of the task-specific information, F_i. This input can be engineered in various forms depending on the downstream application. For instance, it can be derived from dataset-specific statistics (e.g., texture, frequency information) or based on handcrafted rules relevant to the task.

Furthermore, F_i can be a composition of multiple guidance signals, allowing for nuanced control:

$$F_i = \sum_{j=1}^{N} w_j F_j \quad (2)$$

Here, each F_j represents a distinct type of knowledge or feature, and w_j is a learnable weight controlling its influence. This composability enables the model to integrate diverse sources of information, enhancing its adaptability. For a more detailed exploration of this concept, we refer readers to our original SAM-Adapter paper [8].

4. Experiments

4.1. Tasks and Datasets

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

For the camouflaged object detection task, we use three widely used benchmarks: COD10K [19], CHAMELEON [75], and CAMO [48]. COD10K is currently the most comprehensive resource in this area, offering 3,040 samples for training and 2,026 for testing. The CHAMELEON set provides 76 internet-sourced images and is used solely for evaluation. The CAMO dataset comprises 1,250 images, with 1,000 designated for training and 250 for testing. Following the training strategy proposed in [19], our model was trained using CAMO together with the COD10K training split, while performance was assessed on the test splits of CAMO and COD10K, as well as the full CHAMELEON dataset.

For the shadow detection task, we adopted the ISTD dataset [82], which includes 1,330 training and 540 testing samples. For polyp segmentation in medical imaging, we used the Kvasir-SEG dataset [32], following the train–test division specified in the Medico 2020 Automatic Polyp Segmentation challenge [31]. In cell segmentation, we obtain dataset from NeurIPS 2022 Cell Segmentation Challenge [61], which focuses on cell segmentation in various microscopy images. We used 1000 images for training and 101 images for evaluation. The original task was an instance segmentation task. In our experiments, we converted instance segmentation into a semantic segmentation task following the data processing method described in [60].

For the evaluation protocol, we followed the guidelines in [54]. For camouflaged object detection, we adopted widely used metrics, including S-measure (S_m), mean E-measure (E_ϕ), and MAE. The shadow detection task was assessed using the balanced error rate (BER). For polyp segmentation, model performance was quantified using the mean Dice score (mDice) and mean Intersection-over-Union (mIoU).

For more details, please refer to the original SAM-Adapter paper [8].

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 [54], 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. Balanced BCE loss is used for shadow detection. BCE loss and IOU loss are used for camouflaged object detection and polyp segmentation. The AdamW optimizer is used for all experiments. The initial learning rate is set to 2e-4. Cosine decay

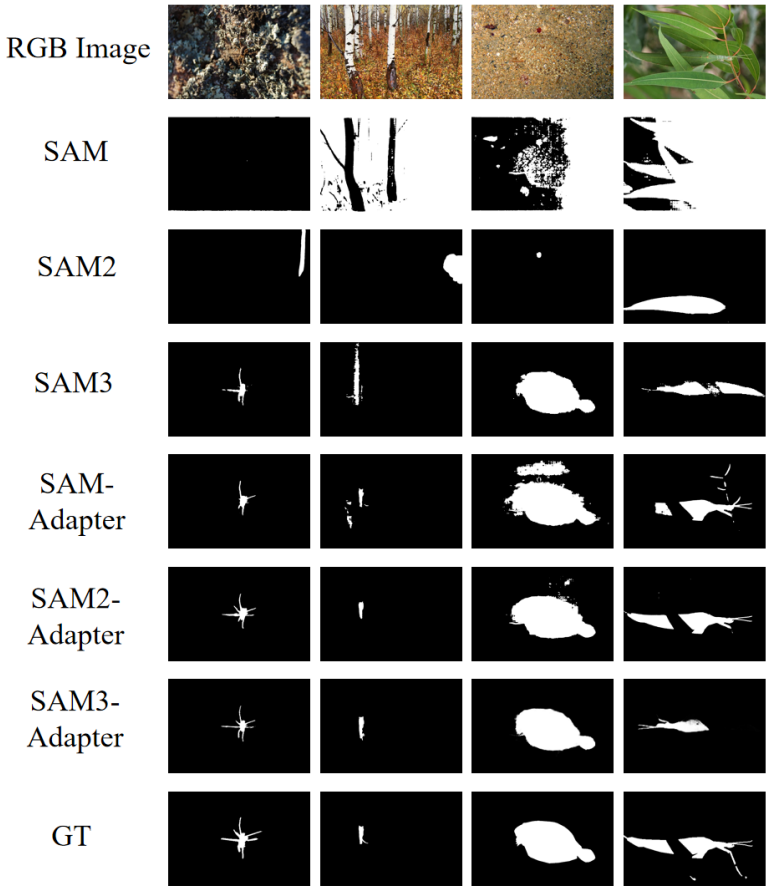

Figure 2. **Segmentation performance visualization on CHAMELEON samples.** The figure highlights the limitations of SAM, SAM2, and SAM3 in handling severe camouflage, where they produce non-meaningful outcomes. Although SAM-Adapter enhances segmentation quality, our SAM3-Adapter delivers the best performance, generating precise masks that closely match the ground truth compared to its predecessors.

Method	CHAMELEON [75]				CAMO [48]				COD10K [19]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	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[18]	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[57]	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 [49]	0.870	0.924	-	0.039	0.792	0.839	-	0.82	0.800	0.872	-	0.041
PFNet [64]	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 [52]	0.888	0.939	0.828	0.032	0.783	0.839	0.702	0.081	0.809	0.889	0.684	0.035
MM-SAM [52]	0.923	0.946	0.853	0.027	0.863	0.901	0.782	0.059	0.896	0.907	0.808	0.023
SENet [52]	0.888	0.932	0.847	0.039	0.918	0.957	0.878	0.019	0.865	0.925	0.780	0.024
SAM [45]	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 [70]	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 [7, 8]	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 [10]	0.915	0.955	0.889	0.018	0.855	0.909	0.810	0.051	0.899	0.950	0.850	0.018
SAM3-Adapter (Ours)	**0.944**	**0.972**	**0.908**	**0.016**	**0.919**	**0.954**	**0.875**	**0.029**	**0.927**	**0.965**	**0.882**	**0.015**

Table 1. Quantitative Segmentation Result Comparison for Camouflaged Object Detection

is applied to the learning rate. The batch size is 2. The training of camouflaged object segmentation is performed for 29 epochs. Shadow segmentation is trained for 29 epochs. Polyp segmentation is trained for 100 epochs. The experiments are implemented using PyTorch on NVIDIA Tesla A800 GPUs. For more information, please refer to the orig-

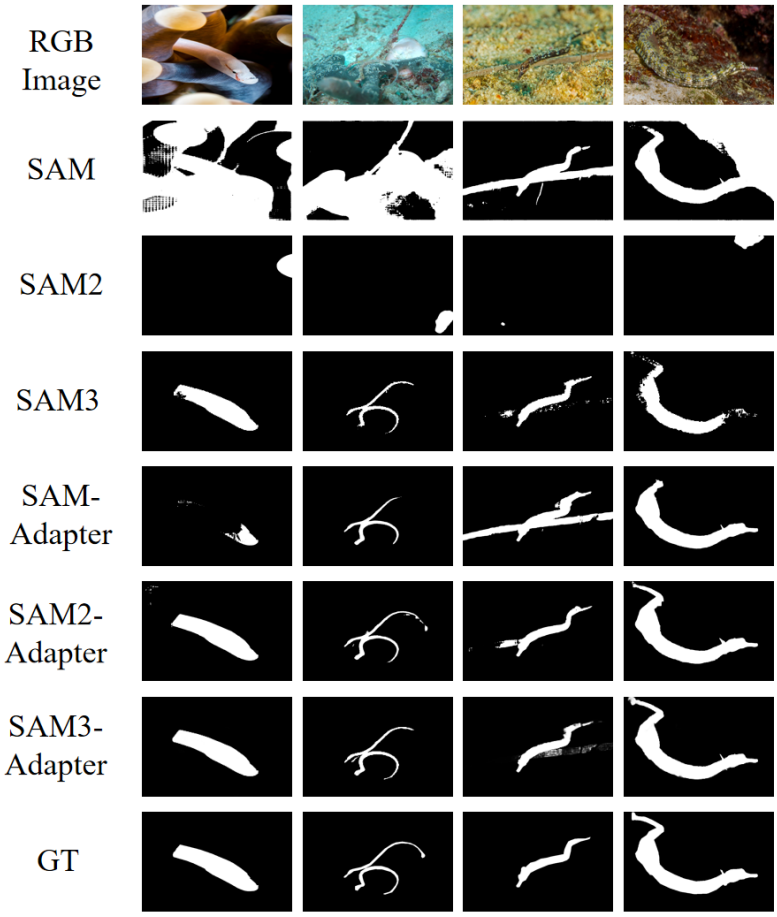

Figure 3. **Camouflaged image segmentation on the COD-10K dataset.** Examples from the COD-10K dataset illustrating animals that are strongly camouflaged within their natural backgrounds. The original SAM frequently fails to accurately localize these targets and can produce fragmented or semantically incoherent segmentations; SAM2 and SAM3 exhibits similar limitations, occasionally producing no mask or incorrect outputs. With the integration of SAM3-Adapter, segmentation reliability on these challenging instances is substantially improved, achieving clear gains over earlier SAM2-Adapter variants.

inal SAM-Adapter paper [8] and our codebase. Note that our codebase also supports Ascend 910b NPU.

4.3. Experiments for Camouflaged Object Detection

We first evaluated the performance on the challenging task of camouflaged object detection, where objects are intentionally blended into their surroundings. Our initial analysis revealed that the powerful Segment Anything 3 (SAM3) model, on its own, already possesses a foundational capability to discern these concealed objects. Unlike its predecessors (SAM and SAM2) which often failed to produce meaningful results, SAM3 can typically locate the camouflaged targets. However, as shown in the baseline results in Figure 2, its segmentation masks often lack precision, with blurry boundaries and incomplete coverage. This is quantitatively reflected in Table 1, where SAM3's standalone performance, while promising, does not yet match the state-of-the-art.

This is where SAM3-Adapter demonstrates its transformative impact. As vividly illustrated in Figures 2, 3, and 4, the introduction of our adapter dramatically enhances SAM3's native ability. The segmentation results are not just improved; they are refined to a new level of accuracy. Our method produces masks with sharper, more precise contours that adhere tightly to the true object boundaries, effectively separating the camouflaged object from its visually similar background.

The quantitative results confirm this visual evidence. With the enhancement from SAM3-Adapter, our method not only surpasses the standalone SAM3 but also establishes a new state-of-the-art (SOTA) across all evaluated metrics. This significant performance leap, achieved by refining an already strong baseline, underscores the effectiveness of our adapter in unlocking and focusing the full poten-

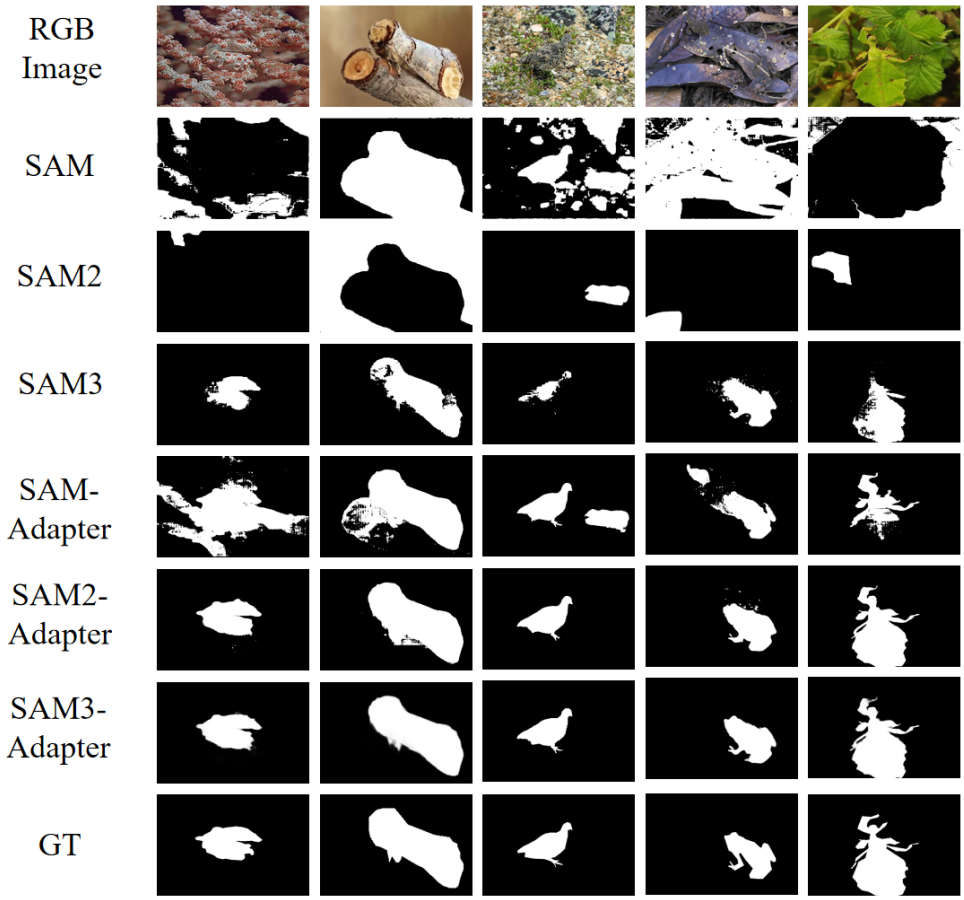

Figure 4. **Camouflaged examples from the CAMO dataset.** The original SAM, SAM2 struggle to perceive animals that are visually concealed within their natural surroundings. SAM3, however, have already gained the capability of distinguish the camouflaged object as we can observe. Integrating SAM3-Adapter further improves the model's ability to segment the concealed targets.

tial of the SAM3 backbone for high-fidelity segmentation.

4.4. Experiments for Shadow Detection

We extended our evaluation to the task of shadow detection, a challenge that requires discerning subtle, low-contrast regions from the background. Our analysis shows that the Segment Anything 3 (SAM3) model, on its own, demonstrates a clear, foundational understanding of the "shadow" concept. Unlike previous models that often failed entirely, SAM3 is generally able to identify the presence and approximate location of shadows in an image. However, as illustrated by the baseline results in Figure 5, these initial predictions often suffer from inaccuracies, such as incomplete segmentation, bleeding into non-shadow areas, and poorly defined edges. This provides the perfect opportunity for SAM3-Adapter to showcase its value. By integrating our lightweight adapter, we transform SAM3's foundational capability into expert-level performance. The visual results in Figure 5 are striking: where the standalone SAM3 produced ambiguous or noisy masks, our SAM3-Adapter method yields clean, precise shadow segmentations with sharp, well-defined contours. The adapter effectively teaches the model to respect the subtle boundaries of the shadow, eliminating the previously observed missing parts and erroneous additions. The quantitative data presented in Table 2 rigorously supports these visual improvements. The integration of SAM3-Adapter brings a significant performance boost, elevating the results far beyond the SAM3 baseline and setting a new state-of-the-art (SOTA) for shadow detection. This success further validates our core hypothesis: the most effective path to advancing segmentation is not just using a powerful backbone like SAM3, but synergistically enhancing it with intelligent, task-specific adapters.

4.5. Experiments for Polyp Segmentation

We then evaluated our method in the critical domain of medical image segmentation, specifically focusing on polyp

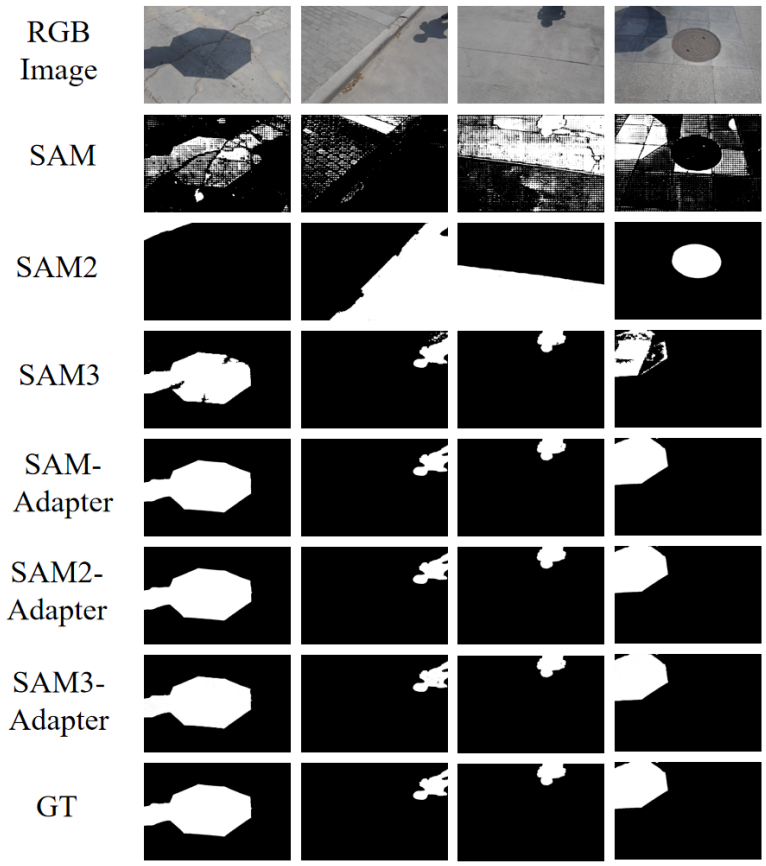

Figure 5. **Visualization of Shadow Detection results.** SAM and SAM2 fails to identify shadows. Standalone SAM3 demonstrates a foundational ability to identify shadows, but struggles with precise boundaries. Our SAM3-Adapter unlocks SAM3's full potential, transforming its initial perception into state-of-the-art segmentation masks with sharp, accurate contours.

Method	BER ↓
Stacked CNN [79]	8.60
BDRAR [98]	2.69
DSC [29]	3.42
DSD [92]	2.17
FDRNet [101]	1.55
SAM [45]	40.51
SAM2 [70]	50.81
SAM-Adapter	1.43
SAM2-Adapter	1.43
SAM3-Adapter (Ours)	**1.14**

Table 2. Result for Shadow Detection

segmentation. Accurate and reliable segmentation of polyps is paramount for the early detection and prevention of colorectal cancer, a leading cause of cancer-related deaths globally.

Our analysis begins by assessing the standalone capabilities of the Segment Anything 3 (SAM3) model. We found that SAM3, owing to its powerful architecture and vast pretraining, possesses a strong foundational ability to identify and locate polyps within colonoscopy images. This marks a significant improvement over prior generalist models. However, for a task where clinical precision is essential, SAM3's baseline performance reveals limitations: the resulting segmentation masks, while correctly positioned, often suffer from incomplete coverage and fuzzy boundaries, failing to capture the entire polyp structure accurately (as shown in the baseline examples in Figure 6).

This is precisely the gap that SAM3-Adapter is designed to fill. By integrating our efficient adapter, we elevate SAM3's performance from foundational to state-of-the-art. The visual results, presented in Figure 6, are compelling. Our method transforms SAM3's initial, coarse predictions into highly accurate, holistic segmentation masks that precisely delineate the full contour of the polyp tissue. The

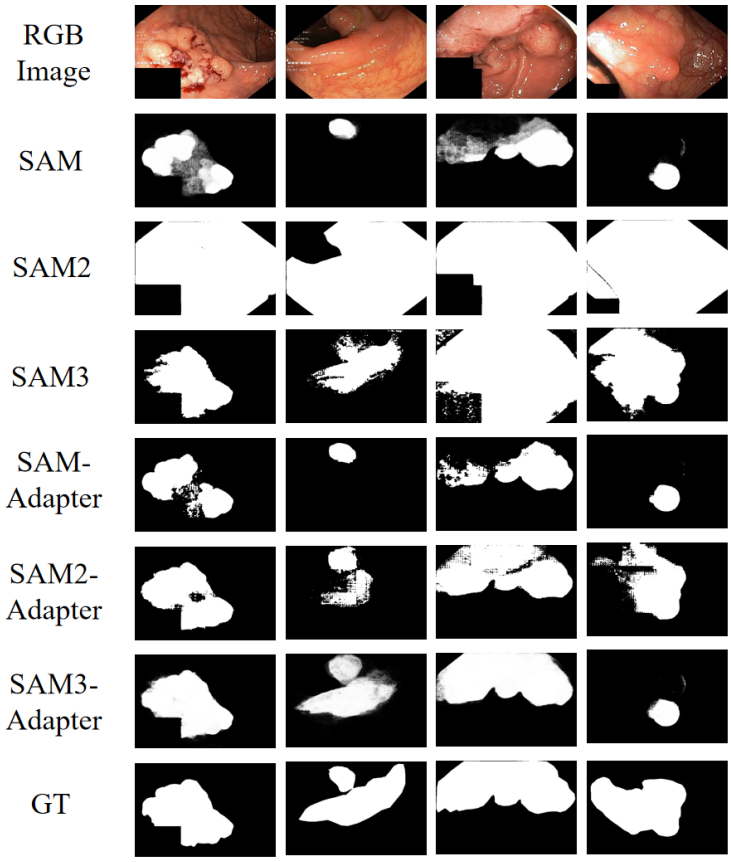

Figure 6. **Qualitative results for Polyp Segmentation.** The figure illustrates that SAM struggles to accurately delineate polyp boundaries, and SAM2 produce non-meaningful outputs. While the powerful SAM3 model can successfully locate polyp tissues, its standalone segmentation often results in incomplete masks with poorly defined boundaries. Our SAM3-Adapter dramatically enhances this foundational capability, guiding the model to produce highly accurate and holistic segmentations. The resulting masks precisely delineate the entire polyp structure, significantly outperforming all baseline models [9, 60]

adapter effectively channels SAM3's powerful features to focus on the subtle yet critical details required for medical-grade accuracy.

The quantitative results in Table 3 provide definitive proof of this superiority. The combination of SAM3 and SAM3-Adapter not only dramatically outperforms the standalone SAM3 but also establishes a new state-of-the-art (SOTA), surpassing all previous methods. This achievement underscores the immense value of our approach: by unlocking and refining the power of a scaled-up foundation model, we provide a more accurate and reliable tool for a crucial clinical application.

4.6. Experiments for Cell Segmentation

To further test the generalizability of our approach, we applied it to the highly challenging task of cell segmentation. This domain demands extreme precision to distinguish individual cells in dense, often overlapping clusters. As evidenced by both qualitative results and quantitative metrics in Table **??**, our method achieved a staggering improvement over all previous state-of-the-art methods. The performance leap was more substantial here than in any other downstream task that we evaluated, showcasing the immense potential of our approach for biomedical research and diagnostics.

5. Conclusion and Future Work

In this work, we presented SAM3-Adapter, a parameter-efficient tuning method that elevates Segment Anything 3 (SAM3) to a new level of performance for specialized segmentation. Through comprehensive experiments, we have demonstrated that SAM3-Adapter sets a new state-of-the-art (SOTA) in difficult segmentation tasks such as medical, camouflage, and shadow detection. Crucially, it achieves these results with higher computational efficiency.

Our extensive experiments validate this hypothesis. By synergizing our lightweight adapter with the powerful

Method	mDice ↑	mIoU ↑
UNet [72]	0.821	0.756
UNet++ [96]	0.824	0.753
SFA [21]	0.725	0.619
SAM [45]	0.778	0.707
SAM2 [70]	0.200	0.029
SAM-Adapter	0.850	0.776
SAM2-Adapter	0.873	0.806
SAM3-Adapter (Ours)	**0.906**	**0.842**

Table 3. Quantitative Result for Polyp Segmentation

Table 4. F1 Score Comparison on Cells in Microscopy

Methods	F1 ↑
nnU-Net	0.5383
SegResNet	0.5411
UNETR	0.4357
SwinUNETR	0.3967
U-Mamba_Bot	0.5389
U-Mamba_Enc	0.5607
xLSTM-UNet_bot	0.5818
xLSTM-UNet _enc	0.6036
SAM3-Adapter (Ours)	**0.7525**

SAM3 backbone, we have established new state-of-the-art (SOTA) benchmarks in challenging domains like medical, camouflage, and shadow segmentation. This combination not only surpasses previous SAM2-based methods in accuracy but also does so with remarkable parameter efficiency, showcasing the tangible benefits of SAM3's advanced design when properly adapted.

The success of SAM3-Adapter provides a clear demonstration: scaling up foundation models, when paired with intelligent and efficient adaptation techniques, directly translates to significant gains in specialized, real-world applications. Our method acts as a catalyst, enabling the powerful general representations learned by SAM3 to be effectively channeled into high-fidelity, domain-specific segmentation.

We advocate for the adoption of SAM3, amplified by our SAM3-Adapter, as the new frontier for high-performance image segmentation. We are releasing our code, models, and protocols to empower the community to build upon this powerful synergy and drive further progress in adapting large-scale vision models. Code, pre-trained models, and data processing protocols are available at http://tianrun-chen.github.io/SAM-Adaptor/

References

[1] 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. 2

[2] 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

[3] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. 3

[4] Keyan Chen, Chenyang Liu, Hao Chen, Haotian Zhang, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. Rsprompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model, 2023. 2

[5] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 2

[6] 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. 2

[7] Tianrun Chen, Lanyun Zhu, Chaotao Deng, Runlong Cao, Yan Wang, Shangzhan Zhang, Zejian Li, Lingyun Sun, Ying Zang, and Papa Mao. Sam-adapter: Adapting segment anything in underperformed scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3367–3375, 2023. 2, 5

[8] Tianrun Chen, Lanyun Zhu, Chaotao Ding, Runlong Cao, Yan Wang, Zejian Li, Lingyun Sun, Papa Mao, and Ying Zang. Sam fails to segment anything? – sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, medical image segmentation, and more, 2023. 2, 3, 4, 5, 6

[9] Tianrun Chen, Chaotao Ding, Lanyun Zhu, Tao Xu, Deyi Ji, Ying Zang, and Zejian Li. xlstm-unet can be an effec-

tive 2d\& 3d medical image segmentation backbone with vision-lstm (vil) better than its mamba counterpart. *arXiv preprint arXiv:2407.01530*, 2024. 2, 9

[10] Tianrun Chen, Ankang Lu, Lanyun Zhu, Chaotao Ding, Chunan Yu, Deyi Ji, Zejian Li, Lingyun Sun, Papa Mao, and Ying Zang. Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. *arXiv preprint arXiv:2408.04579*, 2024. 2, 5

[11] Tianrun Chen, Chunan Yu, Jing Li, Jianqi Zhang, Lanyun Zhu, Deyi Ji, Yong Zhang, Ying Zang, Zejian Li, and Lingyun Sun. 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. 2

[12] 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. 2

[13] 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. 2

[14] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 25(10):1337–1342, 2003. 3

[15] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10680–10687, 2020. 3

[16] Ruining Deng, Can Cui, Quan Liu, Tianyuan Yao, Lucas W. Remedios, Shunxing Bao, Bennett A. Landman, Lee E. Wheless, Lori A. Coburn, Keith T. Wilson, Yaohong Wang, Shilin Zhao, Agnes B. Fogo, Haichun Yang, Yucheng Tang, and Yuankai Huo. Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging, 2023. 2

[17] 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. 2

[18] 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. 5

[19] 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. 2, 3, 4, 5

[20] 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. 2

[21] Yuqi Fang, Cheng Chen, Yixuan Yuan, and Kai-yu Tong. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 302–310. Springer, 2019. 10

[22] Weitao Feng, Deyi Ji, Yiru Wang, Shuorong Chang, Hansheng Ren, and Weihao Gan. Challenges on large scale surveillance video analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 69–76, 2018. 2

[23] Xue Feng, Cui Guoying, and Song Wei. Camouflage texture evaluation using saliency map. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, pages 93–96, 2013. 3

[24] Xiao Fu, Shangzhan Zhang, Tianrun Chen, Yichong Lu, Lanyun Zhu, Xiaowei Zhou, Andreas Geiger, and Yiyi Liao. Panoptic nerf: 3d-to-2d label transfer for panoptic urban scene segmentation. *arXiv preprint arXiv:2203.15224*, 2022. 2

[25] Xiaoqing Guo, Zhen Chen, Jun Liu, and Yixuan Yuan. Non-equivalent images and pixels: Confidence-aware resampling with meta-learning mixup for polyp segmentation. *Medical image analysis*, 78:102394, 2022. 3

[26] Jianqin Yin Yanbin Han Wendi Hou and Jinping Li. Detection of the mobile object with camouflage color under dynamic background based on optical flow. *Procedia Engineering*, 15:2201–2205, 2011. 3

[27] 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. 2

[28] Hanzhe Hu, Deyi Ji, Weihao Gan, Shuai Bai, Wei Wu, and Junjie Yan. Class-wise dynamic graph convolution for semantic segmentation. In *European Conference on Computer Vision*, pages 1–17. Springer, 2020. 2

[29] 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. 8

[30] Xiang Huang, Gang Hua, Jack Tumblin, and Lance Williams. What characterizes a shadow boundary under the sun and sky? In *2011 international conference on computer vision*, pages 898–905. IEEE, 2011. 3

[31] Debesh Jha, Steven A Hicks, Krister Emanuelsen, Håvard Johansen, Dag Johansen, Thomas de Lange, Michael A Riegler, and Pål Halvorsen. Medico multimedia task at mediaeval 2020: Automatic polyp segmentation. *arXiv preprint arXiv:2012.15244*, 2020. 4

[32] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D

Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, pages 451–462. Springer, 2020. 2, 4

[33] Deyi Ji, Hongtao Lu, and Tongzhen Zhang. End to end multi-scale convolutional neural network for crowd counting. In *Eleventh international conference on machine vision*, pages 761–766. SPIE, 2019. 2

[34] Deyi Ji, Haoran Wang, Hanzhe Hu, Weihao Gan, Wei Wu, and Junjie Yan. Context-aware graph convolution network for target re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1646–1654, 2021. 2

[35] Deyi Ji, Haoran Wang, Mingyuan Tao, Jianqiang Huang, Xian-Sheng Hua, and Hongtao Lu. Structural and statistical texture knowledge distillation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16876–16885, 2022. 2

[36] Deyi Ji, Feng Zhao, and Hongtao Lu. Guided patch-grouping wavelet transformer with spatial congruence for ultra-high resolution segmentation. *International Joint Conference on Artificial Intelligence*, pages 920–928, 2023. 2

[37] Deyi Ji, Feng Zhao, Hongtao Lu, Mingyuan Tao, and Jieping Ye. Ultra-high resolution segmentation with ultra-rich context: A novel benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23621–23630, 2023. 2

[38] Deyi Ji, Siqi Gao, Mingyuan Tao, Hongtao Lu, and Feng Zhao. Changenet: Multi-temporal asymmetric change detection dataset. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2725–2729. IEEE, 2024. 2

[39] Deyi Ji, Siqi Gao, Lanyun Zhu, Qi Zhu, Yiru Zhao, Peng Xu, Hongtao Lu, Feng Zhao, and Jieping Ye. View-centric multi-object tracking with homographic matching in moving uav. *arXiv preprint arXiv:2403.10830*, 2024. 2

[40] Deyi Ji, Wenwei Jin, Hongtao Lu, and Feng Zhao. Pptformer: Pseudo multi-perspective transformer for uav segmentation. *International Joint Conference on Artificial Intelligence*, 2024. 2

[41] Deyi Ji, Feng Zhao, Lanyun Zhu, Wenwei Jin, Hongtao Lu, and Jieping Ye. Discrete latent perspective learning for segmentation and detection. In *Forty-first International Conference on Machine Learning*, 2024. 2

[42] Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (TOG)*, 30(6): 1–12, 2011. 3

[43] Lei Ke, Mingqiao Ye, Martin Danelljan, Yifan Liu, Yu-Wing Tai, Chi-Keung Tang, and Fisher Yu. Segment anything in high quality, 2023. 2

[44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 3

[45] 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. 2, 5, 8, 10

[46] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98:123–145, 2012. 3

[47] Hieu Le, Tomas F Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+ d net: Training a shadow detector with adversarial shadow attenuation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 662–678, 2018. 3

[48] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Computer vision and image understanding*, 184:45–56, 2019. 3, 4, 5

[49] 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. 5

[50] 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. 2

[51] Yuheng Li, Mingzhe Hu, and Xiaofeng Yang. Polyp-sam: Transfer sam for polyp segmentation, 2023. 3

[52] 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. 5

[53] 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. 3

[54] Weihuang Liu, Xi Shen, Chi-Man Pun, and Xiaodong Cun. Explicit visual prompting for low-level structure segmentations. *arXiv preprint arXiv:2303.10883*, 2023. 2, 4

[55] Zhikang Liu and Lanyun Zhu. Label-guided attention distillation for lane segmentation. *Neurocomputing*, 438:312–322, 2021. 2

[56] 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. 2

[57] 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. 5

[58] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1), 2024. 2

[59] Jun Ma, Sumin Kim, Feifei Li, Mohammed Baharoon, Reza Asakereh, Hongwei Lyu, and Bo Wang. Segment anything in medical images and videos: Benchmark and deployment. *arXiv preprint arXiv:2408.03322*, 2024. 2

[60] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. 4, 9

[61] Jun Ma, Ronald Xie, Shamini Ayyadhury, Cheng Ge, Anubha Gupta, Ritu Gupta, Song Gu, Yao Zhang, Gihun Lee, Joonkee Kim, et al. The multimodality cell segmentation challenge: toward universal solutions. *Nature methods*, 21(6):1103–1113, 2024. 4

[62] Tanvir Mahmud, Bishmoy Paul, and Shaikh Anowarul Fattah. Polypsegnet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images. *Computers in biology and medicine*, 128:104119, 2021. 3

[63] Maciej A. Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis*, 89:102918, 2023. 2

[64] 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. 5

[65] 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. 3

[66] Balamurali Murugesan, Kaushik Sarveswaran, Sharath M Shankaranarayana, Keerthi Ram, and Mohanasankar Sivaprakasam. Psi-net: Shape and boundary aware joint multi-task deep network for medical image segmentation, 2019. 3

[67] Sohail Nadimi and Bir Bhanu. Physical models for moving shadow and object detection in video. *IEEE transactions on pattern analysis and machine intelligence*, 26(8):1079–1087, 2004. 3

[68] Thomas W Pike. Quantifying camouflage and conspicuousness using visual salience. *Methods in Ecology and Evolution*, 9(8):1883–1895, 2018. 3

[69] Hemin Ali Qadir, Younghak Shin, Johannes Solhusvik, Jacob Bergsland, Lars Aabakken, and Ilangko Balasingham. Toward real-time polyp detection using fully cnns for 2d gaussian shapes prediction. *Medical Image Analysis*, 68: 101897, 2021. 3

[70] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos, 2024. 5, 8, 10

[71] Simiao Ren, Francesco Luzi, Saad Lahrichi, Kaleb Kassaw, Leslie M. Collins, Kyle Bradbury, and Jordan M. Malof. Segment anything, from space?, 2023. 2

[72] 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. 2, 10

[73] Saikat Roy, Tassilo Wald, Gregor Koehler, Maximilian R. Rokuss, Nico Disch, Julius Holzschuh, David Zimmerer, and Klaus H. Maier-Hein. Sam.md: Zero-shot medical image segmentation capabilities of the segment anything model, 2023. 3

[74] P Sengottuvelan, Amitabh Wahi, and A Shanmugam. Performance of decamouflaging through exploratory image analysis. In *2008 First International Conference on Emerging Trends in Engineering and Technology*, pages 6–10. IEEE, 2008. 3

[75] Przemysław Skurowski, Hassan Abdulameer, J Błaszczyk, Tomasz Depta, Adam Kornacki, and P Kozieł. Animal camouflage analysis: Chameleon database. *Unpublished manuscript*, 2(6):7, 2018. 4, 5

[76] 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. 2

[77] 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. 2

[78] Lv Tang, Haoke Xiao, and Bo Li. Can sam segment anything? when sam meets camouflaged object detection, 2023. 2

[79] 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. 8

[80] Haoran Wang, Licheng Jiao, Fang Liu, Lingling Li, Xu Liu, Deyi Ji, and Weihao Gan. Ipgn: Interactiveness proposal graph network for human-object interaction detection. *IEEE Transactions on Image Processing*, 30:6583–6593, 2021. 2

[81] Haoran Wang, Licheng Jiao, Fang Liu, Lingling Li, Xu Liu, Deyi Ji, and Weihao Gan. Learning social spatio-temporal relation graph in the wild and a video benchmark. *IEEE Transactions on Neural Networks and Learning Systems*, 34(6):2951–2964, 2021. 2

[82] 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. 2, 4

[83] Yan Wang, Jian Cheng, Yixin Chen, Shuai Shao, Lanyun Zhu, Zhenzhou Wu, Tao Liu, and Haogang Zhu. Fvp: Fourier visual prompting for source-free unsupervised domain adaptation of medical image segmentation. *IEEE Transactions on Medical Imaging*, 2023. 2

[84] Junde Wu, Wei Ji, Yuanpei Liu, Huazhu Fu, Min Xu, Yanwu Xu, and Yueming Jin. Medical sam adapter: Adapting segment anything model for medical image segmentation, 2023. 2

[85] 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. 2

[86] Junyu Xie, Charig Yang, Weidi Xie, and Andrew Zisserman. Moving object segmentation: All you need is sam (and flow), 2024. 2

[87] Yunyang Xiong, Bala Varadarajan, Lemeng Wu, Xiaoyu Xiang, Fanyi Xiao, Chenchen Zhu, Xiaoliang Dai, Dilin Wang, Fei Sun, Forrest Iandola, Raghuraman Krishnamoorthi, and Vikas Chandra. Efficientsam: Leveraged masked image pretraining for efficient segment anything, 2023. 2

[88] Ying Zang, Chenglong Fu, Runlong Cao, Didi Zhu, Min Zhang, Wenjun Hu, Lanyun Zhu, and Tianrun Chen. Resmatch: Referring expression segmentation in a semi-supervised manner. *arXiv preprint arXiv:2402.05589*, 2024. 2

[89] 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, 2023. 2

[90] 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. 2

[91] Xu Zhao, Wenchao Ding, Yongqi An, Yinglong Du, Tao Yu, Min Li, Ming Tang, and Jinqiao Wang. Fast segment anything, 2023. 2

[92] 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. 8

[93] 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. 2

[94] Tao Zhou, Yizhe Zhang, Yi Zhou, Ye Wu, and Chen Gong. Can sam segment polyps?, 2023. 3

[95] Yuan Zhou, Haiyang Wang, Shuwei Huo, and Boyu Wang. Full-attention based neural architecture search using context auto-regression, 2021. 3

[96] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018. 10

[97] Jiejie Zhu, Kegan GG Samuel, Syed Z Masood, and Marshall F Tappen. Learning to recognize shadows in monochromatic natural images. In *2010 IEEE Computer Society conference on computer vision and pattern recognition*, pages 223–230. IEEE, 2010. 3

[98] 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. 8

[99] 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. 3

[100] 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. 2

[101] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4702–4711, 2021. 8

[102] 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. 2

[103] Lanyun Zhu, Tianrun Chen, Jianxiong Yin, Simon See, and Jun Liu. Learning gabor texture features for fine-grained recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1621–1631, 2023. 2

[104] Lanyun Zhu, Tianrun Chen, Deyi Ji, Jieping Ye, and Jun Liu. Llafs: When large language models meet few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3065–3075, 2024. 2

[105] Lanyun Zhu, Tianrun Chen, Jianxiong Yin, Simon See, and Jun Liu. Addressing background context bias in few-shot segmentation through iterative modulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3370–3379, 2024. 2

[106] Lanyun Zhu, Deyi Ji, Tianrun Chen, Peng Xu, Jieping Ye, and Jun Liu. Ibd: Alleviating hallucinations in large vision-language models via image-biased decoding. *arXiv preprint arXiv:2402.18476*, 2024. 2

[107] 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. 2