Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 2, 2023

Research on Derived Tasks and Realistic Applications of Segment Anything Model: A Literature Review

Juanhua Zhang *

School of Biomedical Engineering, Northeastern University, Shenyang, Liaoning 110016, China
*** Corresponding author Email:** zhangjuanhua_neu@163.com

Abstract: With the rapid development of deep learning technology, unprecedented achievements have been made in the field of computer vision, and the release of the Segment Anything Model (SAM) has shocked the world even more. However, with the continuous evolution of technology, researchers have begun to pay attention to the model in more complex scenarios and problem situations. This review will delve into a series of potential derivative tasks and practical applications of SAM, as well as how to address these tasks through innovative SAM based methods. The author will explore the applications, advantages, and challenges of SAM model in image segmentation, object detection, image restoration, remote sensing, and medical fields, aiming to provide inspiration for researchers in the field of computer vision.

Keywords: Segment Anything; Image Segmentation; Object Detection; Image Inpainting; Remote Sensing Images; Medical Images.

1. Introduction

The Segment Anything Model (SAM) is a fundamental model for image segmentation, proposed by Alexander Kirillov et al. This model draws on the basic models in the NLP field and the idea of constructing large datasets using billions of tokens. SAM can generate high-quality object masks based on input prompts (such as points or boxes), which are used to generate masks for all objects in the image. This model has been trained on a dataset containing 11 million images and 1.1 billion masks, demonstrating excellent zero sample performance and is suitable for various segmentation tasks. On the other hand, it provides a fully automated and suggestive image segmentation model construction method that requires almost no human intervention. Compared to early deep learning methods, SAM no longer requires specialized training data collection, manual annotation, and several hours of training process [1].

In addition, image segmentation, object detection, and image restoration are some attractive application directions, and the combination of SAM and them is intriguing. Meanwhile, the superior performance of SAM is necessary for discussions in remote sensing and medicine.

This review will delve into the potential of SAM in these derived tasks and real-world applications, aiming to provide insights and inspiration for researchers in the field of computer vision. By introducing these innovative methods, the adaptability of SAM models in multiple fields and their potential value in solving specific tasks will be elucidated.

2. Potential Derivative Tasks

2.1. Image Segmentation

2.1.1. Using Adapter to Address Poor Performance of SAM in Specific Scenarios

SAM performs well in general segmentation but has limitations in specific scenarios such as hidden object detection and shadow detection. Therefore, Tianrun Chen et al. introduced the SAM-Adapter, which integrates domain specific knowledge and visual cues into the segmentation network through effective adapters. By combining task-specific insights with the foundational knowledge gained from SAM, the SAM-Adapter significantly improves performance in challenging tasks, which has been comprehensively experimentally validated. It is worth noting that SAM-Adapter has surpassed task specific network models in covert object detection and shadow detection, achieving the latest technological achievements. In addition, this method also improves the performance of medical image segmentation, such as polyp segmentation. This groundbreaking work reveals the adaptability of SAM in different fields, with excellent accuracy. Although evaluated only on a limited dataset, SAM-Adapter demonstrates the potential to become a multifunctional tool suitable for various downstream segmentation tasks, including fields such as medicine and agriculture. This study is of milestone significance in applying large-scale pre trained image models to a wide range of research and industrial applications [2].

2.1.2. One Shot for SAM

Renrui Zhang et al. proposed a personalized segmentation model SAM method for specific visual concepts, called PerSAM. The study achieved customization of SAM through a single image with a reference mask, without the need for manual intervention. PerSAM determines the target concept through prior localization and utilizes three techniques: target guided attention, target semantic guidance, and cascading post-processing to segment other images or videos. In addition, the study also introduced PerSAM-F, an efficient one-time fine-tuning variant. By freezing the entire SAM structure, PerSAM-F only uses two learnable weights to handle multi-scale masks, completing training within 10 seconds, thereby achieving performance improvement. The study validated the effectiveness of the method on the annotated personalized evaluation dataset PerSeg and achieved competitive performance in video object segmentation. In addition, this method can also enhance DreamBooth for personalized and stable diffusion of text to image generation, reducing background interference and improving target appearance learning. On the other hand, this study can also inspire future use of parameter efficient

methods to personalize basic segmentation models [3].

2.1.3. An Innovative Method for Audio Visual Joint Localization and Segmentation

Shentong Mo and Yapeng Tian proposed "AV SAM: Segment Anything Model Meets Audio Visual Localization and Segmentation" and explored a new method for applying deep learning model SAM in audio visual tasks. They focus on the correlation between audio and vision, especially the correlation between audio signals and visual objects, and propose a solution for audio visual joint learning. The AV-SAM framework was introduced in the study, extending the functionality of the SAM model to the audio-visual field. Its core idea is to achieve the localization and segmentation of sound objects by integrating audio and visual information. Researchers use pre trained visual and audio features to aggregate cross modal representations through pixel level fusion methods. These features are used to prompt the encoder and mask decoder, generate audio visual segmentation masks, and achieve joint localization and segmentation. The effectiveness of AV-SAM was validated through experiments on FlickrSoundNet and AVSBench datasets. The results show that AV-SAM performs well in sound object localization and segmentation tasks, and is also competitive in audio visual tasks compared to SAM. They also explored the challenges of audio-visual localization and segmentation tasks, emphasizing the misalignment between audio signals and objects in the video. To address this challenge, AV-SAM adopts pixel level audio visual fusion to learn visual features of audio alignment from videos to better guide segmentation tasks. AV-SAM provides an innovative method for audio visual joint tasks, achieving satisfactory results [4].

2.1.4. Semantic Segmentation in SAM

In the study by authors such as Shehbaz Tariq, the impact of image segmentation on the training of semantic encoding models was explored. A practical image segmentation method is proposed by using the Segment Anything Model (SAM) as the basic model. This method does not require expert data annotation and only requires a simple prompt to segment the region of interest. The research results indicate that this method can improve image restoration performance in noisy communication channels. Although this study mainly focuses on image restoration, namely the impact of segmentation in semantic encoding models, future research can further explore its role in task completion ability. On the other hand, basic models such as BERT and GPT have extensive generalization capabilities on multimodal data through self-supervised learning. On this basis, this study explores the application of SAM as a suggestive image segmentation model that can perform zero sample segmentation tasks without explicit training. By fully utilizing the segmentation ability of SAM, this study proposes a practical method for image transmission, which adopts a lightweight neural network architecture in source signal and channel encoding and decoding, significantly reducing communication overhead while preserving higher image quality. The author believes that this method has potential value in practical applications as it not only eliminates the need for segmentation model training, but also applies to various semantic encoding architectures [5].

2.2. Target Detection

2.2.1. Rotating Object Detection

Qingyun Li implemented the application of SAM by using MMRotate to generate rotation bounding boxes. This method was compared with 'H2RBox v2: Boosting HBox supervised Oriented Object Detection via Symmetric Learning'. With the strong zero sample capability demonstrated by SAM, researchers provided well-trained horizontal FCOS detectors as prompts to SAM to generate corresponding masks, and ultimately obtained rotated RBox by performing minimum bounding rectangle operations on the predicted masks. Thanks to its powerful zero sample capability, ViT-B based SAM-RBox achieved an accuracy of 63.94%. However, due to time-consuming post-processing, only 1.7 FPS was achieved during the inference process [6].

2.2.2. Abnormal Target Detection

Yunkang Cao et al. introduced the "Segment Any Anomaly+(SAA+)" innovative framework to achieve zero sample anomaly target detection using the Segment Anything Model (SAM). Traditional anomaly segmentation models require domain fine-tuning and are limited to generalization between different anomaly patterns. Therefore, this study combines basic models such as SAM to apply multimodal knowledge for anomaly localization for the first time, solving the generalization problem. To adapt to anomaly segmentation, mixed prompt normalization is introduced. The SAA+ framework has two steps. Use a hint-based object detection model to find abnormal areas, and then refine the segmentation mask based on the hint-based segmentation model. Different from traditional prompts, combining text prompts, domain knowledge, and image context. This fusion strategy has a significant effect. Domain knowledge provides target exception descriptions, constructs specific prompts, aligns basic models and dataset content. Based on image context prompts, identify, and calibrate anomaly segmentation, and accurately model the degree of anomalies. Experiments have confirmed that SAA+ reaches a new level in multiple datasets. SAM's zero sample generalization ability, mixed hint normalization, SAA+ without additional training, accurately detects various texture anomalies, and is an efficient anomaly detection scheme [7].

2.3. Image Inpainting

Tao Yu et al. innovatively explored the field of maskless image restoration for the first time based on the Segment Anything Model (SAM), proposing a new paradigm of "Inpaint Anything (IA)". The core concept of IA integrates the advantages of multiple models to build efficient and easy-to-use repair solutions. IA revolves around three main characteristics, firstly "Remove Anything". When a user selects a target, IA removes the content and soothes the context to fill the "hole"; Then there is "Fill Anything", where users need to provide text prompts to remove objects, and IA fills them with AIGC models such as Stable Diffusion; Finally, there is "Replace Anything", which preserves the selected objects and generates a new background to replace the original one. IA integrates Segmentation Anything's visual basic model, leading restorer, and AIGC model, supporting "click delete, prompt fill" and adapting to various images, including 2K resolution. The author showcases the potential of large-scale AI models and reveals the prospects of "composable AI". In the future, the research team plans to further develop IA by incorporating functions such as image matting and editing, which will be widely applied in practical scenarios [8].

3. Practical Applications in Reality

3.1. Remote Sensing

To fully utilize the excellent characteristics of SAM, innovative methods have been used to segment remote sensing images using SAM. For example, Keyan Chen et al. proposed a method based on prompt learning, named RSPrompt, to address the universality and zero sample capability of SAM in image segmentation, as well as the remaining issues in remote sensing image segmentation tasks. This method can automatically generate appropriate prompts suitable for SAM input, enabling SAM to generate segmentation results with semantic differentiation ability in remote sensing images [9]. In addition, Di Wang et al. explored how to quickly generate large-scale remote sensing image segmentation datasets (SAMRS) through SAM using existing remote sensing object detection datasets. In fact, SAMRS not only has a large scale, but also provides object category, location, and instance information for semantic segmentation, instance segmentation, and object detection, providing strong resource support for research in the field of remote sensing image segmentation [10]. These two studies jointly demonstrate the innovative application of SAM in remote sensing image segmentation and how to overcome the difficulty of remote sensing data annotation, providing new ideas and resources for remote sensing image segmentation research.

3.2. Medical Field

To explore the effectiveness of SAM in medical images, Sheng He et al. tested the accuracy of SAM in different medical image segmentation tasks and explored influencing factors. They tested SAM on 12 datasets. Measure accuracy using Dice coefficient and compare with 5 medical image segmentation algorithms. However, in terms of results, the Dice coefficient of SAM in 12 datasets was significantly lower than the other 5 algorithms, with a difference between 0.1 and 0.7 [11]. Additionally, Maciej A. Mazurowski et al. found through research that SAM exhibits impressive zero shot segmentation performance on certain medical image datasets but has moderate to poor performance for other datasets [12]. However, Peilun Shi et al. believe that SAM has limitations in fields such as medical imaging, and its zero-sample segmentation performance is inconsistent in different medical fields. For certain structured targets such as blood vessels, zero sample segmentation fails. Nevertheless, by fine-tuning a small amount of data, the segmentation quality can be significantly improved, demonstrating the enormous potential and feasibility of fine-tuning SAM in precise medical image segmentation [13]

To address this issue, Jun Ma et al. proposed MedSAM, which is based on SAM and used in medicine. With a dataset constructed from over one million images, MedSAM not only outperforms existing segmentation basic models in terms of performance, but also exhibits comparable or even superior performance in many professional models. In addition, MedSAM can accurately extract key biomarkers for quantifying tumor burden, thus possessing the potential to accelerate the development of diagnostic tools and personalized treatment. By achieving accurate and efficient segmentation in multiple tasks through MedSAM, it is expected to promote the further development of diagnostic tools and personalized treatment plans [14]. Furthermore, Junde Wu et al. proposed the Med SAM Adapter, which integrates medical domain knowledge into segmentation models through simple adaptation techniques. Although it is one of the rare practices of natural language processing technology Adapters in the field of vision, it performs excellently in medical image segmentation. The Medical SAM Adapter (MSA) has demonstrated excellent performance in 19 tasks, covering CT, MRI, ultrasound, fundus, and dermatoscopy images. Compared to fully tuned MedSAM in terms of nnUNet, TransUNet, UNetr, MedSegDiff, etc., the performance gap is significant [15].

4. Conclusion

The application of SAM model in multiple fields has demonstrated its potential and value in image segmentation, object detection, image restoration, as well as remote sensing and medical fields. In the field of image segmentation, the SAM model is not only widely used for different tasks, such as medical image segmentation, but also achieves significant performance improvement by combining domain specific knowledge and visual cues. In the direction of object detection, SAM combined with innovative technologies has further expanded the research on zero sample rotation object detection and abnormal object detection, providing more powerful solutions for these fields. In the field of image restoration, the SAM model provides the foundation for the proposal of the "Input Anything" method and opens a new paradigm for image restoration tasks.

Moreover, the application of SAM in the field of remote sensing not only improves the effectiveness of remote sensing image segmentation by automatically generating prompts and constructing datasets suitable for SAM, but also provides more support for research in this field. In the medical field, the application of SAM has shown complex effects, although it performs well in some cases and may perform mediocrely in other data. However, through methods such as MedSAM and MedSAM Adapter, the application of SAM in medical image segmentation has been improved, making positive contributions to the development of medical diagnosis and personalized treatment.

In summary, the widespread application of SAM models in multiple fields and their performance advantages in different tasks highlight their important role in promoting progress in the field of image processing and analysis. With further research and innovation, we believe that the SAM model will continue to bring more possibilities and opportunities to various fields.

References

[1] Kirillov, A., Mintun, E., Ravi, N., et al. (2023). Segment Anything. arXiv. https://doi.org/10.48550/arXiv.2304.02643.

[2] Chen, T., Zhu, L., Ding, C., et al. (2023). SAM Fails to Segment Anything? -- SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More. arXiv. https://doi. org/10. 48550/ arXiv.2304.09148.

[3] Zhang, R., Jiang, Z., Guo, Z., et al. (2023). Personalize Segment Anything Model with One Shot. arXiv. https:// doi.org/ 10.48550/arXiv.2305.03048.

[4] Mo, S., & Tian, Y. (2023). AV-SAM: Segment Anything Model Meets Audio-Visual Localization and Segmentation. arXiv. https://doi.org/10.48550/arXiv.2305.01836.

[5] Tariq, S., Arfeto, B. E., Zhang, C., et al. (2023). Segment Anything Meets Semantic Communication. arXiv. https:// doi.org/ 10.48550/arXiv.2306.02094.

[6] GitHub - Li-Qingyun/sam-mmrotate: SAM (Segment Anything Model) for generating rotated bounding boxes with MMRotate, which is a comparison method of H2RBox-v2. (n.d.-f). Retrieved 11 August 2023, from https://github.com/Li-Qingyun/sam-mmrotate.

[7] Cao, Y., Xu, X., Sun, C., et al. (2023). Segment Any Anomaly without Training via Hybrid Prompt Regularization. arXiv. https://doi.org/10.48550/arXiv.2305.10724.

[8] Yu, T., Feng, R., Feng, R., et al. (2023). Inpaint Anything: Segment Anything Meets Image Inpainting. arXiv. https:// doi.org/ 10.48550/arXiv.2304.06790.

[9] Chen, K., Liu, C., Chen, H., et al. (2023). RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation based on Visual Foundation Model. arXiv. https://doi.org/ 10. 48550/arXiv.2306.16269.

[10] Wang, D., Zhang, J., Du, B., et al. (2023). Scaling-up Remote Sensing Segmentation Dataset with Segment Anything Model. arXiv. https://doi.org/10.48550/arXiv.2305.02034.

[11] He, S., Bao, R., Li, J., et al. (2023). Computer-Vision Benchmark Segment-Anything Model (SAM) in Medical Images: Accuracy in 12 Datasets. arXiv. https://doi.org/ 10. 48550/ arXiv.2304.09324.

[12] Mazurowski, M. A., Dong, H., Gu, H., et al. (2023). Segment anything model for medical image analysis: An experimental study. Medical Image Analysis, 102918. https://doi.org/ 10. 1016 / j.media.2023.102918.

[13] Shi, P., Qiu, J., Abaxi, S. M. D., et al. (2023). Generalist Vision Foundation Models for Medical Imaging: A Case Study of Segment Anything Model on Zero-Shot Medical Segmentation. Diagnostics, 13(11), 1947. https://doi.org/ 10.3390/ diagnostics 13111947.

[14] Ma, J., He, Y., Li, F., et al. (2023). Segment Anything in Medical Images. arXiv. https://doi.org/ 10.48550/ arXiv. 2304. 12306.

[15] Wu, J., Zhang, Y., Fu, R., et al. (2023). Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. arXiv. https://doi.org/ 10.48550/ arXiv. 2304. 12620.