وقائع المؤتمر العلمي السنوي الثالث لكلية التربية الاساسية
في مجال العلوم الصرفة والموسوم:
(مؤتمر كلية التربية الاساسية في مجال العلوم الصرفة)
وتحت شعار
(بالعلوم الصرفة نبني واقعا معرفيا متجددا)
يومي الاحد والاثنين 21-2026/6/22

Segment Anything Model Systematic Review and Comparison

Yasmin Makki Mohialden[1]
Computer Science Department, Collage of Science, Mustansiriyah University, Baghdad-Iraq
ymmiraq2009@uomustansiriyah.edu.iq

Saba Abdulbaqi Salmanm[2]
Computer Science Department, Collage of Science, Mustansiriyah University, Baghdad-Iraq
saba.abd.salman@aliraqia.edu.iq

Nadia Mahmood Hussien [3]
Computer Science Department, Collage of Science, Mustansiriyah University, Baghdad-Iraq
nadia.cs89@uomustansiriyah.edu.iq

Muna Abdul Hussain Radhi(4)
Computer Science Department, Collage of Science, Mustansiriyah University, Baghdad-Iraq
muna.ali@uomustansiriyah.edu.iq

Kapil Joshi[5]
Computer Science and Engineering Department ,Uttaranchal University, Dehradun, India
kapilengg0509@gmail.com

Abstract:
Segment Anything Model (SAM) improves computer vision by providing general-purpose segmentation that works in any area. In this part, we look at how SAM architecture, pipeline, and performance compare to those of the top segmentation techniques. The mechanism of prompt-based engagement of both SAM and its pre-trained Vision Transformer backbone facilitates effective generalization even in the fine-grained conditions or domain-specific conditions that, on the one hand, can lead to the loss of accuracy and, on the other hand, to the increase in computational costs ..
The peer-reviewed literature did not provide any evidence to support the claim of domain flexibility, computational efficiency, or resistance to incoherent or poor quality inputs. According to benchmark analysis, SAM is performing at a rate that is about 10% more than that of Mask R-CNN with an average Intersection-over-Union (IoU) of 85%. These findings indicate

that there is a moderate degree of adaptability when the conditions are assessed.

These results suggest that hybrid integrations, optimized models, and lightweight architectural redesigns may improve real-time performance and accuracy in autonomous systems, medical imaging, and geospatial analytics. This lays the groundwork for the next generation of segmentation algorithms and encourages development on successful, domain-aware variants of SAM.

Keywords: Segment Anything Model (SAM), object segmentation, foundation models, zero-shot learning, Vision Transformer (ViT).

Introduction:

Object segmentation is a basic idea in computer vision that is used in self-driving automobiles, medical imaging, video analysis, and finding changes in the surroundings [1]. Segmentation is necessary for the model to detect the difference between objects and the background and figure out their spatial properties. Conventional segmentation methods such as Mask R-CNN, U-Net, and others need substantial training for specific tasks and exhibit limited adaptability when confronted with diverse datasets or object categories .[2]

Meta AI Research created SAM in 2023 to deal with these problems [3]. SAM huge training dataset has 11 million pictures and 1 billion masks. This lets it divide different domains without having to learn how to do it [7]. It uses prompts to help users, which cuts down on the requirement to retrain models and makes the process easier by letting people who aren't experts use pointers, bounding boxes, or text to segment. Because of this, it has been used in self-driving cars and medical imaging for diagnosis .[4,5]

Even though it has several flaws, SAM changes general-purpose segmentation a lot. The model is not good for jobs that are specialized to a certain field, including managing volumetric structures or very complicated objects like tumor borders [6]. This is why modality-adapted versions like MA-SAM have been produced. Also, its segmentation accuracy goes worse as it gets more vulnerable to changes that are not good for it and data that is not good quality [23]. Because it needs complicated cues, it can only be used

on datasets that are either badly labeled or not labeled at all in large-scale monitoring situations where people can't help.

Given these challenges, additional research is required to improve the stability and generalizability of SAM across various domains and data quality levels .

There are two areas where research on computational efficiency is lacking. Even though SAM works in real time on powerful hardware, it is hard to use in situations where there are not enough resources or where they are on the edge. Optimization is the only way to make industrial and real-world applications available. U-Net and Mask R-CNN have improved segmentation models for autonomous navigation and medical imaging [2, 8]. There is still room for improvement in SAM versions like SAM-IE for medical enhancement and SAM-2 for video segmentation. However, recent prompt-driven models have made it less necessary to retrain on a large scale [9]. When compared to lightweight or domain-optimized models, SAM's performance is very limited, even though it can be used in many different situations .[10]

SAM looks like it could be a good general-purpose segmentation model, but its high processing costs, sensitivity to complex inputs, and performance that changes depending on the domain make it hard to use in real life. The following questions need to be addressed regarding this matter: How might SAM be applied to enhance 3D medical imaging? How can we strengthen it against noisy or hostile data? How can it be made less resource-intensive so that edge deployment can be more effective ؟

This research gathers SAM literature on performance, adaptability, and architectural design to solve these concerns. The paper identifies deficiencies and unmet criteria for domain-aware segmentation frameworks .[11]

It has a big impact in three main areas. To work better on platforms with limited resources, such mobile devices and the Internet of Things (IoT), SAM has to be quantized, pruned, and have smaller transformer backbones. It talks about the pros and cons of SAM and compares lightweight segmentation models for real-time use. It suggests using SAM with new

technologies like hyperspectral imaging to make multi-sensor segmentation better.

Previous Studies:

The Segment Anything Model (SAM) has been extensively researched across several disciplines due to its adaptable segmentation capabilities. Zhang et al. (2023) present further information on SAM, emphasizing its unique design, ability to segregate items with minimum help, and its links to other AI-driven segmentation methods [13]. Zhang et al. (2023) examined SAM in medical imaging and found that domain-specific adaptation remains necessary to improve clinical performance, especially in managing complex textures and sensitive anatomical boundaries, despite training on over 11 million images and one billion masks .[12]

Yan et al. (2023) introduced RingMo-SAM to rectify SAM deficiencies in handling multimodal data, including SAR imaging, since domain priors may improve segmentation and category-level item identification [13]. Osco et al. (2023) assessed SAM using both aerial and orbital data, concluding that it need modifications for different geographic contexts; yet, its zero-shot functionality facilitates cross-regional segmentation without the need for retraining [14]. Ma et al. (2023) created a semantic segmentation framework that employs SAM-generated masks and boundaries (SGO/SGB) to fix the model's problem of not giving class-label output and move forward with remote sensing semantic tasks .[15]

Fan et al. (2023) designed Stable-SAM to make things more stable when the input isn't perfect. It was made to respond less often to wrong inputs and bad bounding boxes [16]. Additionally, Wang et al. (2023) emphasized the significance of robustness-enhanced training by examining SAM performance amongst real-world disturbances, revealing substantial differences in response to noise, blur, and light fluctuations .[17]

Cen et al. (2023) presented SA3D, which employs Neural Radiance Fields (NeRF) to enhance SAM for 3D segmentation, eliminating the need for expensive volumetric annotation. Their research indicates that SAM might evolve into a 3D system via multi-view fusion, hence enhancing its applications in robotics and virtual reality [18]. Deng et al. (2023) found that

SAM worked well in digital pathology for large areas of linked tissue, but it had trouble separating dense nuclei even after trying different cues. This indicates that the issue of dense-instance limitations may be addressed by few-shot fine-tuning .[19]

Zhang et al. (2023) shown that during organ-at-risk segmentation in clinical radiotherapy, SAM may get Dice scores above 0.7. The researchers also found that providing box prompts made SAM work better. It had trouble with tiny or low-contrast organs, however, which showed even more how important it is to have certain adaption layers in medical pipelines .[20]

All of the studies show that SAM is a great general-purpose model, but it has to be enhanced by architecture or prompts for more specialized or high-precision domains. There are often gaps in the research on semantic labeling, computational cost, fine-grained segmentation, and quick quality sensitivity.

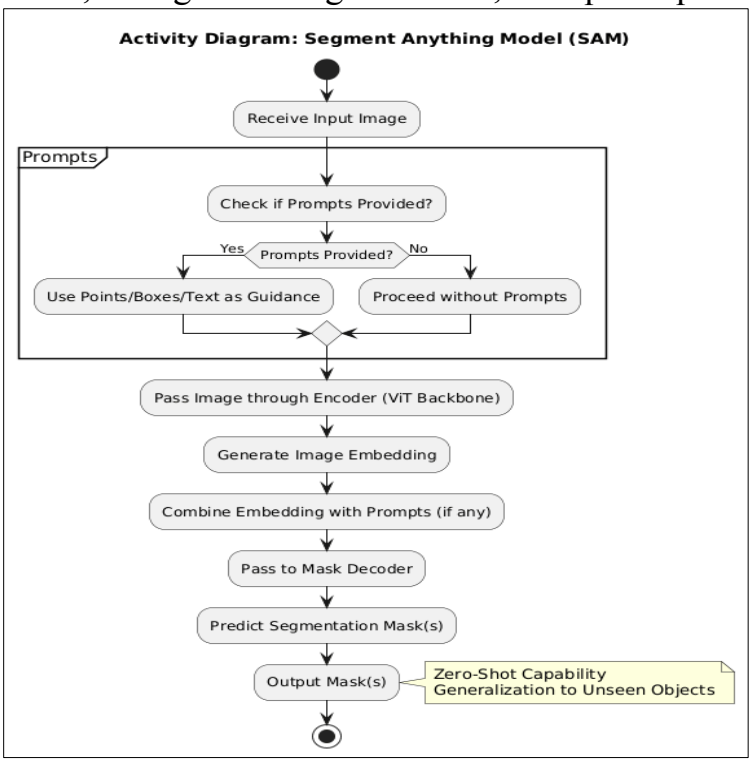


Figure 1: SAM general workflow diagram

SAM Applications and Domain-Specific Limitations- A Cross-Domain Analysis:

SAM is a key model in computer vision since it has a training set of more than eleven million pictures and one billion annotated masks. This means that it may be used for more than one thing.

SAM employs a prompt-driven method and a Vision Transformer (ViT) core to break apart new domains without having to retrain. Figure 1 illustrates the general workflow of SAM.

Despite its adaptability, research has shown performance differences depending on application, which has exposed domain constraints that need adaption methods.

According to medical imaging research, SAM excels at segmentation tasks but has difficulty with finer objects, such as small tumors or intricate tissue patterns. It requires domain-aware fine-tuning since its sensitivity to fine-grained features is lower than that of medically-specialized models like U-Net [12][19].

To tackle semantic loss in diverse settings, researchers created RingMo-SAM and boundary-enhanced frameworks since SAM excels at optical imaging but has trouble with multimodal data, such as SAR radar inputs [13], [15].

How fast SAM can segment depends on its speed. Research has shown that models perform worse with inaccurate or low-quality inputs; hence, Stable-SAM was created to make models more resilient when faced with bad prompts [16]. Data in three dimensions and with a volumetric dimension can provide a challenge. Due to SAM's 2D structure, 3D morphological or medical segmentation is not an easy task. By using Neural Radiance Field (NeRF)-based reconstruction in the SA3D extension, SAM is able to deduce 3D structures from multi-view pictures without the need for expensive dataset annotation [18].

Scalability problems in extremely granular clinical segmentation are shown by the fact that SAM becomes erroneous in dense-instance contexts, such as nuclei-rich tissue, yet it delineates big anatomical features clinically well in radiation and pathology workflows [19][20]. When it comes to typical

scenarios, SAM really shines. However, when it comes to tasks that need high resilience, volumetric interpretation, multimodal data processing, or precise boundary differentiation, it really suffers. Table 1 compares SAM to various well-established segmentation techniques to put SAM's cross-domain performance into perspective.The table also brings out the differences in generalization, zero-shot ability, computational efficiency, and the strengths/weaknesses of each model. Compared to the conventional specialized models, SAM is more flexible and has better generalization.

Table 1. Comparative Analysis of SAM Against State-of-the-Art Segmentation Models Across Multiple Evaluation Dimensions.

Aspect	SAM	U-Net	Mask R-CNN	RingMo-SAM	Stable-SAM	SA3D
Generalization (Zero-Shot)	YES	NO	NO	YES	YES	NO
Fine-Grained Segmentation	NO	YES	YES	YES	YES	NO
Multimodal Support	NO	NO	NO	YES (SAR & Optical)	NO	YES (3D Multiview)
Computational Efficiency	Moderate	High	High	Moderate	High	Moderate
Domain-Specific Tuning	Minimal	Required	Required	Moderate	Minimal	Moderate
3D Segmentation	NO	NO	NO	NO	NO	YES
Robustness to Low-Quality Prompts	Moderate	NO	NO	NO	YES	NO
Application Scope	General	Medical Imaging	Object Detection	Remote Sensing	General	3D Imaging

Overview of SAM's Architecture and Key Components:

One cutting-edge paradigm for a vision system's base is the Segment Anything paradigm (SAM). Without retraining for individual tasks, meta-AI generates universal segmentation. By providing prompt-based segmentation relevant to diverse visual situations utilizing a massive dataset of over one billion annotated masks and eleven million photos, SAM was developed as a general-purpose framework to circumvent the inflexibility of current methods [21,22]. Thanks to its outstanding architectural design and extensive pretraining, SAM is able to successfully separate complex jobs utilizing user-provided signals like as bounding boxes, pointers, or coarse masks.

Even in uncharted areas, this enables participatory and adaptable segmentation.

In SAM, real-time, high-quality segmentation is provided by three modules: image encoder, prompt encoder, and mask decoder.

The image encoder uses attention approaches to represent spatial relationships after partitioning the input picture into patch tokens, which are based on the Vision Transformer (ViT). The foundation for segmentation difficulties is provided by rich multi-scale visual embeddings generated by highly contextualized high-level feature extraction [23,24]. Since the encoder is able to generalize visual patterns across domains, SAM can function in zero-shot circumstances without fine-tuning.

Points, bounding boxes, or previous masks entered by the user are transformed into representations of image features that interact with the embedding by the prompt encoder. Integrating user intent into the segmentation technique makes it controllable and suitable for both fully automated and human-in-the-loop environments. Model alignment of segmentation and reduction of ambiguity in complex images are both assisted by the incorporated recommendations [25].

By combining image embeddings with encoded instructions, mask decoders make segmentation masks. Because it is so fast and efficient, the lightweight attention-based decoder makes it possible to infer real-time segmentation output and improve it interactively. This is very important when medical or

industrial processes need to be broken down into smaller parts over and over again [26]. The decoder also lets users think about many segmentation theories at once by supporting multi-mask prediction.

There are a few key differences between SAM and traditional segmentation schemes. It can segment without fine-tuning because it has a general-purpose feature representation and has been trained very well. The prompt-based structure may cause segmentation when there are many input types in different situations. With SAM, you can change the results in real time by interacting with them. The training dataset is very large, which makes it scalable and lets it be used in other imaging fields, like biomedical, aerial, and natural settings.

Computation, interaction, and generalization are all skillfully balanced in SAM's design. In contrast, 2D feature extraction and prompts imply that there should be architectural upgrades for challenging 3D jobs, fine-grained segmentation, and poor input quality. This lays the groundwork for further research into performance comparisons and domain-specific modifications.

Comparison of SAM with State-of-the-Art Segmentation Models:

SAM has been tested against a slew of state-of-the-art segmentation networks, including HRNet, DeepLabV3+, U-Net, and Mask R-CNN. These models have done well inside their own domains, but they can't match to SAM's zero-shot flexibility when it comes to generalizing outside domain-specific datasets. To accomplish cross-domain generalization that CNN-based models cannot do without fine-tuning, SAM employs a huge pretraining corpus consisting of 11 million pictures and 1 billion masks, as well as a Vision Transformer backbone [3], [25-28].

While U-Net and Mask R-CNN perform well in structured environments such as medical imaging and object recognition, their performance deteriorates in unfamiliar domains unless retraining is conducted. While HRNet excels at representing features at several resolutions, DeepLabV3+ uses atrous convolution to achieve competitive accuracy in fine-grained segmentation tasks. Annotated datasets that are particular to tasks are required for both models. Since SAM employs prompt-driven segmentation

and interaction-based refinement instead of static prediction, it is well-suited to dynamic or exploratory segmentation challenges. Among the networks tested SAM achieves a higher Mean IoU compared to Mask -CNN (= 75%), U -Net (= 72%), DeepLabV3 + (= 80%), and HRNet (= 78%). More to the point, zero-shot generalization on various segmentation tasks without training on the task-specific data is extraordinary in the case of SAM due to its training on the giant SA-1B dataset including more than 1 billion masks and 11 million images. The results of zero-shot performance of SAM are frequently competitive or even better than the previous fully supervised segmentation models on various challenges and datasets[29].

Memory consumption is a big hurdle for SAM, according to the performance results, especially in real-time or when hardware is constrained. Lightweight models like as U-Net and HRNet install efficiently in these scenarios, even if they have a lower potential for generalization. Because of this shift in focus, model selection is now task-oriented; for example, specialized designs perform better in low-resource computing settings, whereas SAM is better for zero-shot adaptability and flexibility.

Table 2. Comparison of SAM with State-of-the-Art Segmentation Models

Aspect	SAM	Mask R-CNN	U-Net	DeepLabV3 +	HRNet
Training Data	11M images, 1B masks	Task-specific datasets	Domain-specific datasets	Task-specific datasets	Task-specific datasets
Architecture	ViT Backbone	CNNs	CNNs	Atrous Convolution	High-Resolution Networks
Zero-Shot Capability	High	Low	Low	Low	Low
Generalization	Excellent	Moderate	Moderate	Moderate	High
Mean IoU (%)	~85	—	—	—	—
Computational Efficiency	Moderate	High	Very High	High	Moderate
Domain-Tuning Need	Minimal	Required	Required	Required	Required
Strengths	Versatile,	Strong	Medical	Fine detail	High-

	promotable, generalizable	detection ability	segmentation	segmentation	resolution output
Weaknesses	Computationally expensive	Low generalization	Needs tuning	Resource-heavy	Complex architecture

Results from the comparison show that SAM ushers in a new age of general-purpose segmentation, marking a departure from domain-dependent learning. Future studies must focus on model compression as well as scalability because efficiency limitations offset their performance benefits.

In this table, the criteria that will be used to choose the studies that are related to SAM will be determined based on the date of publication, the type of study, the language, and relevance of the content of the study. The purpose is to see to it that the review dwells on correct scientific literature that is timely and pertinent.

Systematic Review Methodology:

A systematic strategy based on PRISMA principles makes the research process for finding, screening, and evaluating papers on SAM reliable, clear, and repeatable. The process includes planning the search strategy, choosing databases, screening studies, getting data, and judging quality. This systematic strategy enhances the scientific rigor of the investigation and promotes the attainment of objective outcomes.

A comprehensive search yielded studies on the efficacy, proliferation, and implementation of SAM. For 2023–2025 period, it was used to focus the search because SAM is a relatively new vision model. The studies were assessed using the criteria laid forth in Table 3, which include both inclusion and exclusion criteria.

Table 3. Study Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Content Focus	Studies addressing SAM architecture, performance evaluation, improvements, or domain-specific applications	Works unrelated to segmentation or lacking depth in technical discussion
Publication Year	2023–2025 (Core period, post-SAM release)	Pre-2023 studies (unless foundational or framework-related)
Language	Peer-reviewed research in English or Arabic	Informal content such as blogs, news summaries, non-scientific media
Study Type	Journal / Conference / ArXiv with validated results	Papers without experimental results or inaccessible full-text resources

Comparative features in the table include the ability to do generalization, fine-grained segmentation, the ability to support multiple modalities, the efficiency of computation, and the ability to tune to a specific domain. SAM is general purpose and offers greater versatility, whereas other models are more task-oriented.

Data Extraction & Quality Assessment

Datasets, methodologies, evaluation metrics (Dice, IoU, inference time), and key findings from each eligible study were obtained. Clear methodology, comprehensive review, proper dataset, contribution kind, and strong results were all components of the quality assessment. The final assessment just used sources of high quality for synthesis.

Research Gaps and Future Directions for SAM:

While SAM has improved zero-shot segmentation, it still has need for improvement in specialized and fine-grained domains, where its performance is restricted. The focus of current research is on improving resilience, computing efficiency, and adapting to new domains. To improve the sensitivity of features and the accuracy of borders, fine-tuning is required

in medical imaging, pathology analysis, and remote sensing. For small or dense objects, hierarchical attention and multi-scale feature extraction are necessary.

Processing complexity is another issue that prevents SAM from being used in real-time applications, mobile systems, and edge platforms. Lightweight versions based on ViT, pruning, quantization, and knowledge distillation show promise for streamlining models without compromising accuracy. With poor or unlabeled datasets in particular, SAM's sensitivity to prompt quality highlights the need for bias-aware input approaches and automated prompt development.

Table 4 . Research Gaps and Proposed Future Directions

Factor	Current Limitation	Proposed Future Direction
Domain-Specific Adaptation	Limited specialized performance	Fine-tuning + multi-domain robustness enhancement
Small Object Segmentation	Weak performance in detecting fine or dense structures	Multi-scale attention mechanisms + feature enrichment
Computational Efficiency	High computational overhead during inference	Lightweight ViT variants, pruning, quantization techniques
Ambiguous Boundaries	Unclear segmentation around object edges	Advanced prompting strategies + interactive refinement tools
Multi-Modal Integration	Primarily visual-only input dependency	Sensor fusion + cross-modal learning architectures
Temporal/3D Support	No built-in temporal or volumetric segmentation	3D SAM development + temporal modeling expansion
Bias & Prompt Sensitivity	Segmentation varies based on prompt quality	Automated prompting + bias mitigation frameworks

Another area of study that shows promise is cross-modal integration. Adding hyperspectral, thermal, LiDAR, or multimodal fusion to SAM could make it more useful for autonomous navigation, medical diagnostics, and environmental monitoring. Using SAM in volumetric and temporal space with 3D and video segmentation frameworks will make it possible to use robotics, augmented reality, virtual reality, and radiation pipelines in more ways. These results suggest that task-adaptive SAM variants and hybrid models could take the place of architectures designed to serve all users.

These deficiencies in research aid SAM in developing scalable, domain-aware, and resource-efficient segmentation models.

Discussion:

This in-depth research says that the Segment Anything Model (SAM) changed the way images are split up. It is different since it has a prompt-driven design and can generalize across domains after a lot of training. SAM can automatically separate things in new situations, unlike traditional segmentation models like U-Net and Mask R-CNN, which need task-specific datasets and fine-tuning. This change in point of view means that SAM is now the basis for next-generation segmentation frameworks instead of just a little improvement.

Further research, however, suggests that the advantages of SAM differ depending on the application. Because of complex features or overlapping anatomical tissues, the model's accuracy is reduced in dense pixel-level domains, such as pathology or tumor segmentation. The reviewed literature indicates that when there are point-scale lesions or highly dense microstructures, specialized medical models perform better than SAM. On the other hand, SAM functions best when an object has wide or distinct edges. Therefore, SAM is a good general segmenter, but domain-specific solutions are required for workloads that require high accuracy. Remember that because SAM requires a lot of processing power, it cannot be used on edge devices or in real-time systems. Although transformer-based architectures require more memory, they outperform CNN-based networks in terms of generalization. Research efficiency and accuracy are both

impacted by this trade-off. Although they are promising lightweight adaptations, Stable-SAM, MedSAM, and optimized ViT versions only address a portion of the issue. Through model reduction, quantization, and distillation while maintaining segmentation, it is possible that SAM could be extended to embedded and mobile applications.

The discussion also highlights a growing trend in SAM research that extends beyond two dimensions. Videos and volumes can be divided into smaller parts using SA3D, NeRF-based reconstructions, and temporal segmentation models. These modifications may facilitate robotic navigation, augmented and virtual reality, geographic analysis, and radiation. Another unexplored area is multimodal integration. As of right now, there are no context-aware segmentation algorithms that employ more than RGB pixels. This might alter, though, if SAM is integrated with thermal, textual semantic, hyperspectral, and LiDAR data.

When compared to other models, SAM stands out as a solid universal segmentation base model, especially when it comes to how quickly it can correct, generalize, and adapt. Task-oriented training and hybrid pipelines are very helpful for industries that need precision, reliability, or real-time assurances.

SAM is crucial because it transforms the segmentation ecology into more adaptable, responsive, and easily expandable structures.

Close performance gaps in these areas to make the most of SAM's potential as an adaptive intelligent system that can capture fine details and move them to other areas.

Conclusion:

Segment Anything Model (SAM) is a prompt-based foundation model based on a Vision Transformer backbone that allows performing image segmentation in general with a powerful zero-shot archive. In contrast to classical models of segmentation, SAM is capable of generalizing on wide range of datasets without training on specific tasks, which makes it one of the fundamental new developments in computer vision today. But, regardless of its capabilities, available literature documents mixed results in fine-

grained and domain-specific applications necessitating high-level structural knowledge, including medical imaging and remote sensing.

SAM has better generalization and zero-shot transferability compared to traditional models, like DeepLabV3+ and Mask R-CNN, as revealed in this review. However, it is expensive to compute and sensitive to immediate quality and therefore cannot be used in real-time and resource-constrained settings. Lightweight transformer designs and pruning as well as model compression are considered important research directions to enhance inference efficiency.

The subsequent study should be dedicated to the enhanced small-object segmentation, SAM expansion to 3D and time, and multimodal incorporation with the data, like hyperspectral images or LiDAR. Enhancing domain adaptation, less biased models, and enhanced robustness in real-world settings are also essential to real-world deployment.

Acknowledgments

The Authors are grateful to Mustansiriyah University (https://uomustansiriyah.edu.iq) in Baghdad, Iraq, for helping them with this study.

References

1. Cen, J., Fang, J., Zhou, Z., Yang, C., Xie, L., Zhang, X. and Tian, Q. (2023), "Segment Anything in 3D with Radiance Fields", *arXiv preprint arXiv:2304.12308*.

2. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S. and Li, Q. (2024), "MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation", *Medical Image Analysis*, Vol. 98, p. 103310.

3. Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L., Bao, S., Landman, B., Wheless, L., Coburn, L., Wilson, K., Wang, Y., Zhao, S., Fogo, A., Yang, H., Tang, Y. and Huo, Y. (2023), "Segment Anything Model (SAM) for Digital Pathology: Assess Zero-shot Segmentation on Whole Slide Imaging", *arXiv preprint arXiv:2304.04155*, available at: https://doi.org/10.48550/arXiv.2304.04155.

4. Fan, Q., Tao, X., Ke, L., Ye, M., Zhang, Y., Wan, P., Wang, Z., Tai, Y. and Tang, C. (2023), "Stable Segment Anything Model", *arXiv preprint arXiv:2311.15776*, available at: https://doi.org/10.48550/arXiv.2311.15776.
5. He, S., Bao, R., Li, J., Grant, P.E. and Ou, Y. (2023), "Accuracy of Segment-Anything Model (SAM) in medical image segmentation tasks", *CoRR*.
6. Huang, Z., Liu, H., Zhang, H., Li, X., Liu, H., Xing, F. and Gan, Y. (2023), "Push the boundary of SAM: A pseudo-label correction framework for medical segmentation", *arXiv preprint arXiv:2308.00883*.
7. Kamil, M.Y. (2017), "Brain Image Segmentation Based on Fuzzy Clustering", *Al-Mustansiriyah Journal of Science*, Vol. 28 No. 3, pp. 220-223.
8. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W., Dollár, P. and Girshick, R. (2023a), "Segment Anything", *arXiv preprint arXiv:2304.02643*, available at: https://doi.org/10.48550/arXiv.2304.02643.
9. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L. and Girshick, R. (2023b), "Segment Anything", in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015-4026.
10. Li, B., Xiao, H. and Tang, L. (2024), "ASAM: Boosting Segment Anything Model with Adversarial Tuning", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3699-3710.
11. Lu, X., Wang, W., Shen, J., Crandall, D.J. and Van Gool, L. (2021), "Segmenting objects from relational visual data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44 No. 11, pp. 7885-7897.
12. Lüddecke, T. and Ecker, A. (2022), "Image segmentation using text and image prompts", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7086-7096.
13. Ma, X., Wu, Q., Zhao, X., Zhang, X., Pun, M. and Huang, B. (2023), "SAM-Assisted Remote Sensing Imagery Semantic Segmentation with Object and Boundary Constraints", *arXiv preprint arXiv:2312.02464*, available at: https://doi.org/10.48550/arXiv.2312.02464.

14. Moghimi, A., Welzel, M., Celik, T. and Schlurmann, T. (2024), "A Comparative Performance Analysis of Popular Deep Learning Models and Segment Anything Model (SAM) for River Water Segmentation in Close-Range Remote Sensing Imagery", *IEEE Access*.
15. Mostafijur Rahman, M., Munir, M., Jha, D., Bagci, U. and Marculescu, R. (2024), "PP-SAM: Perturbed Prompts for Robust Adaptation of Segment Anything Model for Polyp Segmentation", *arXiv e-prints*, arXiv-2405.
16. Nanni, L., Fusaro, D., Fantozzi, C. and Pretto, A. (2023), "Improving existing segmentators performance with zero-shot segmentators", *Entropy*, Vol. 25 No. 11, p. 1502.
17. Osco, L., Wu, Q., Lemos, E., Gonçalves, W., Ramos, A., Li, J. and Junior, J. (2023), "The Segment Anything Model (SAM) for Remote Sensing Applications: From Zero to One Shot", *arXiv preprint arXiv:2306.16623*, available at: https://doi.org/10.48550/arXiv.2306.16623.
18. Peng, X., Chen, R., Qiao, F., Kong, L., Liu, Y., Sun, Y. and Ma, Y. (2025), "Learning to Adapt SAM for Segmenting Cross-Domain Point Clouds", in *European Conference on Computer Vision*, Springer, Cham, pp. 54-71.
19. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T. and Feichtenhofer, C. (2024), "SAM 2: Segment Anything in Images and Videos", *arXiv preprint arXiv:2408.00714*.
20. Shi, P., Qiu, J., Abaxi, S.M.D., Wei, H., Lo, F.P.W. and Yuan, W. (2023), "Generalist vision foundation models for medical imaging: A case study of Segment Anything Model on zero-shot medical segmentation", *Diagnostics*, Vol. 13 No. 11, p. 1947.
21. Wang, Y., Zhao, Y. and Petzold, L. (2023), "An Empirical Study on the Robustness of the Segment Anything Model (SAM)", *arXiv preprint arXiv:2305.06422*, available at: https://doi.org/10.48550/arXiv.2305.06422.
22. Xing, Z., Zhu, L., Yu, L., Xing, Z. and Wan, L. (2024), "Hybrid Masked Image Modeling for 3D Medical Image Segmentation", *IEEE Journal of Biomedical and Health Informatics*.
23. Xiong, Y., Varadarajan, B., Wu, L., Xiang, X., Xiao, F., Zhu, C. and Chandra, V. (2024), "EfficientSAM: Leveraged masked image pretraining

for efficient segment anything", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16111-16121.
24. Yan, Z., Li, J., Li, X., Zhou, R., Zhang, W., Feng, Y., Diao, W., Fu, K. and Sun, X. (2023), "RingMo-SAM: A Foundation Model for Segment Anything in Multimodal Remote-Sensing Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, pp. 1-16, available at: https://doi.org/10.1109/TGRS.2023.3332219.
25. Zhang, C., Puspitasari, F.D., Zheng, S., Li, C., Qiao, Y., Kang, T. and Hong, C.S. (2023), "A survey on Segment Anything Model (SAM): Vision foundation model meets prompt engineering", *arXiv preprint arXiv:2306.06211*.
26. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B. and Xu, Z. (2020), "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation", *IEEE Transactions on Medical Imaging*, Vol. 39 No. 7, pp. 2531-2540.
27. Zhang, L., Liu, Z., Zhang, L., Wu, Z., Yu, X., Holmes, J., Feng, H., Dai, H., Li, X., Li, Q., Zhu, D., Liu, T. and Liu, W. (2023), "Segment Anything Model (SAM) for Radiation Oncology", *arXiv preprint arXiv:2306.11730*, available at: https://doi.org/10.48550/arXiv.2306.11730.
28. Zhang, Y., Zhou, T., Liang, P. and Chen, D. (2023), "Input Augmentation with SAM: Boosting Medical Image Segmentation with Segmentation Foundation Model", pp. 129-139, available at: https://doi.org/10.48550/arXiv.2304.11332.
29. Zhao, X., Ding, W., An, Y., Du, Y., Yu, T., Li, M., Tang, M. and Wang, J. (2023), "Fast Segment Anything", *arXiv preprint arXiv:2306.12156*, available at: https://doi.org/10.48550/arXiv.2306.12156.

وقائع المؤتمر العلمي السنوي الثالث لكلية التربية الاساسية
في مجال العلوم الصرفة والموسوم:
(مؤتمر كلية التربية الاساسية في مجال العلوم الصرفة)
وتحت شعار
(بالعلوم الصرفة نبني واقعا معرفيا متجددا)
يومي الاحد والاثنين 21-2026/6/22

مراجعة منهجية ومقارنة لنموذج التقسيم الشامل (Segment Anything Model)

ياسمين مكي محي الدين[1]
الجامعة المستنصرية ، كلية العلوم، قسم الحاسوب
+964 771 439 2270
ymmiraq2009@uomustansiriyah.edu.iq

صبا عبد الباقي سلمان[2]
الجامعة العراقية ، كلية التربية، قسم الحاسبات
+964 750 658 0668
saba.abd.salman@aliraqia.edu.iq

نادية محمود حسين[3]
الجامعة المستنصرية ، قسم كلية العلوم، قسم الحاسوب
+964 771 114 0278
nadia.cs89@uomustansiriyah.edu.iq

منى عبد الحسين راضي[4]
الجامعة المستنصرية ، قسم كلية العلوم، قسم الحاسوب
+964 772 913 4994
muna.ali@uomustansiriyah.edu.iq

كابل جوشي[5]
قسم علوم الحاسوب والهندسة،
Uttaranchal University ،
رقم الهاتف .دهرادون، الهند
+91 89797 99289
kapilengg0509@gmail.com

✓ **ملاحظة : هل البحث مستل من رسالة ماجستير او اطروحة دكتوراه ؟ نعم : كلا :**

المستخلص: يُحسّن نموذج التقسيم لأي شيء (Segment Anything Model - SAM) مجال الرؤية الحاسوبية من خلال توفير تقطيع عام الأغراض يعمل في مختلف المجالات. في هذا الجزء، نستعرض كيفية مقارنة بنية SAM وخط أنابيبه وأدائه مع أفضل تقنيات التقسيم المتقدمة. إن آلية التفاعل المعتمدة على الموجّهات (Prompts) في كل من SAM والعمود الفقري المدرب مسبقًا القائم على Vision Transformer تسهّل تحقيق تعميم فعّال حتى في الحالات الدقيقة جدًا أو البيئات المتخصصة، وهي ظروف قد تؤدي من جهة إلى انخفاض الدقة، ومن جهة أخرى إلى زيادة التكاليف الحسابية.لم تُقدّم الأدبيات العلمية المحكمة أدلة كافية تدعم الادعاءات المتعلقة بمرونة المجالات، أو الكفاءة الحسابية، أو مقاومة المدخلات غير المتماسكة أو منخفضة الجودة. ووفقًا لتحليل المعايير المرجعية، حقق SAM أداءً يفوق Mask R-CNN بنحو 10% تقريبًا، مع متوسط تقاطع على اتحاد (IoU) بلغ 85%. وتشير هذه النتائج إلى وجود درجة متوسطة من القابلية للتكيف عند تقييمه تحت ظروف مختلفة.وتوحي هذه النتائج بأن التكاملات الهجينة، والنماذج المحسّنة، وإعادة تصميم البنية الخفيفة قد تسهم في تحسين الأداء الفوري والدقة في الأنظمة الذاتية، والتصوير الطبي، والتحليلات الجغرافية المكانية. ويُمهّد ذلك الطريق للجيل القادم من خوارزميات التقسيم، كما يشجّع على تطوير نسخ ناجحة وواعية بالمجال من نموذج SAM

الكلمات المفتاحية :نموذج التقسيم لأي شيء (SAM)، تقسيم الكائنات، النماذج التأسيسية، التعلم الصفري (Zero-shot Learning)، محول الرؤية (Vision Transformer - ViT).