Correspondence

DOI: 10.5582/bst.2023.01119

The potential of 'Segment Anything' (SAM) for universal intelligent ultrasound image guidance

Guochen Ning[1,*], **Hanyin Liang**[1], **Zhongliang Jiang**[2], **Hui Zhang**[1], **Hongen Liao**[1,*]

[1] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China;
[2] Computer Aided Medical Procedures, Technical University of Munich, Germany.

SUMMARY Ultrasound image guidance is a method often used to help provide care, and it relies on accurate perception of information, and particularly tissue recognition, to guide medical procedures. It is widely used in various scenarios that are often complex. Recent breakthroughs in large models, such as ChatGPT for natural language processing and Segment Anything Model (SAM) for image segmentation, have revolutionized interaction with information. These large models exhibit a revolutionized understanding of basic information, holding promise for medicine, including the potential for universal autonomous ultrasound image guidance. The current study evaluated the performance of SAM on commonly used ultrasound images and it discusses SAM's potential contribution to an intelligent image-guided framework, with a specific focus on autonomous and universal ultrasound image guidance. Results indicate that SAM performs well in ultrasound image segmentation and has the potential to enable universal intelligent ultrasound image guidance.

Keywords medical artificial intelligence, intelligent medical image guidance, large models, ultrasound image

Ultrasound image guidance is a method often used to help provide care, and it relies on the accurate perception of information from ultrasound images to guide medical procedures. It has become an indispensable tool in many clinical settings, ranging from diagnostics and interventions to minimally invasive therapies (*1*). Moreover, ultrasound images can be used as guidance to direct surgery and can also be acquired by a robot to further develop an autonomous image-guided process (*2*). One critical aspect of ultrasound image guidance is image segmentation, which involves identifying and delineating specific structures or regions of interest within ultrasound images. Image segmentation plays a crucial role in ultrasound-guided procedures as it enables physicians to accurately locate and target the tissues or structures of interest, monitor their changes in real-time, and guide the intervention or therapy accordingly (*3*). Moreover, the segmented structures can serve as the basis for planning and guiding robotic ultrasound systems, enabling autonomic positioning of an ultrasound probe for optimal imaging of the regions of interest (*4*). Therefore, accurate segmentation is essential for precise targeting and localization of structures, and it is currently achieved with AI models that have different structures. However, the reliance on specific AI models for tissue segmentation limits autonomous ultrasound image-guided procedures, like high-intensity focused ultrasound (HIFU) therapy, in clinical settings (*5*). It may require frequent updates and retraining of the AI models to adapt to different tissues, which is time-consuming and may not always be feasible in clinical settings (*6*). In addition, the availability of labeled ultrasound datasets for training may be limited, and especially those for rare diseases or specific procedures, further adding to the challenges of developing accurate and robust tissue segmentation models for intelligent ultrasound image guidance (*6*).

Large models, also called foundation models, have brought about a revolution in many fields, driving breakthroughs in diverse applications (*7*). One such model, ChatGPT, has garnered considerable recognition for its ability to generate human-like text responses and engage in interactive conversations (*8*). Numerous experts have described the potential for large models to be used in healthcare and their impact on traditional healthcare models (*9*). Inspired by the success of ChatGPT, the Segment Anything Model (SAM) is a state-of-the-art foundation model that enhances the generalization capability of image segmentation to new heights (*10*). Although several studies are attempting to improve segmentation generalizability by introducing new structures, large models are still only trained for a single anatomy. In contrast, SAM leverages advanced machine learning algorithms to automatically identify and delineate objects, structures, or regions of interest within

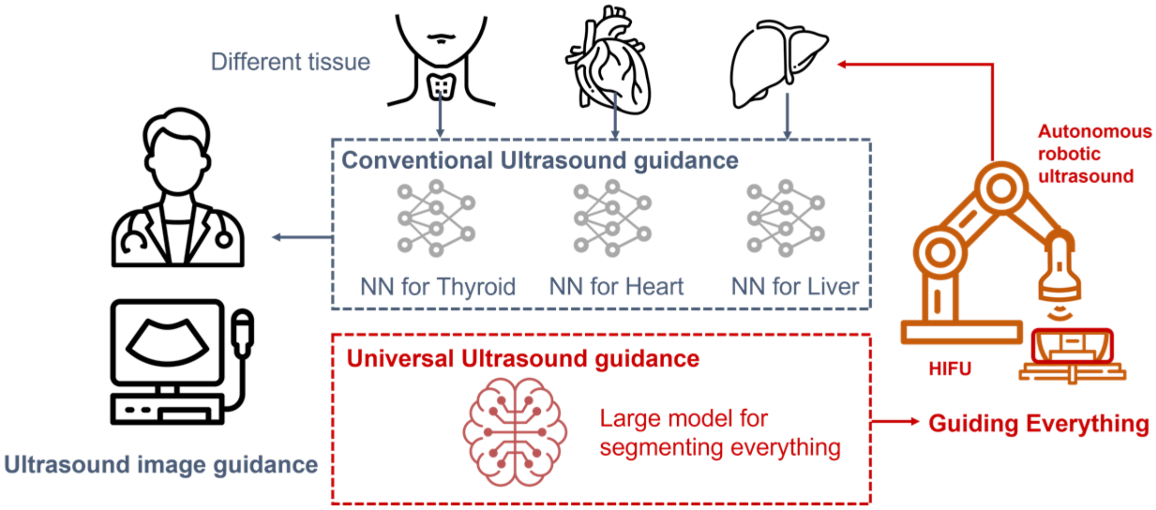

Figure 1. Can "Segment Anything" help with fully autonomous ultrasound image guidance?

images, regardless of their multimodality or diversity. With its unprecedented capabilities, SAM opens up new possibilities for visual analysis, understanding, and manipulation in different fields. The versatility of SAM in segmenting different organs enhances the potential of intelligent medical image guidance in various scenarios. Unlike traditional models that require the training of specific segmentation models for each organ, SAM's ability to segment different organs with a high level of accuracy allows for a more generalized approach (*11*). The current study has evaluated the performance of SAM on ultrasound image segmentation and assessed if the results can serve as a catalyst for a universal image-guided system, as shown in Figure 1.

The evaluation was performed in two parts, focusing on SAM's performance on ultrasound image data and the reliability of the segmentation results for guidance. To assess SAM's performance, three models of interaction were selected, namely Everything mode, Click mode, and Box mode, with each representing a level of interaction. In Everything mode, SAM automatically segments the image within its range of detection. In Click mode, SAM utilizes human clicks on the image as prompts for segmentation, with these clicks randomly generated within the region of the ultrasound image segmentation label. Lastly, in Box mode, SAM segments are based on boxes drawn by a human on the image as prompts, where the enclosing boxes are generated from the segmentation labels of the ultrasound image. In addition, Click mode and Box mode were combined in Click-Box mode to include more interactive information.

The quantitative evaluation metric used in this study is the DICE coefficient between the labels and the segmentation results obtained from SAM (*12*). DICE measures the similarity between two samples and is a reproducibility validation metric most frequently used in medical image segmentation. Three public ultrasound image datasets, denoted as Carotid Artery (*13*), Thyroid (*14*), and Heart (*15*), were selected for evaluation. These datasets consist of continuous ultrasound images, which are more representative of the real-time acquisition of ultrasound images during guidance. In the Click model, the accuracy of segmentation was 0.877 ± 0.115 for the Carotid artery, 0.427 ± 0.320 for the Thyroid, and 0.326 ± 0.153 for the Heart. In the Box model, the accuracy of segmentation was 0.908 ± 0.041 for the Carotid artery, 0.829 ± 0.126 for the Thyroid, and 0.841 ± 0.098 for the Heart. In the Click-Box model, the accuracy of segmentation was 0.909 ± 0.041 for the Carotid artery, 0.829 ± 0.121 for the Thyroid, and 0.867 ± 0.051 for the Heart. Figure 2 shows some ultrasound images obtained using different segmentation models. Results generated by the Everything model are difficult to evaluate quantitatively due to the inclusion of multiple objectives. The results of quantitative segmentation reveal that SAM can achieve a high level of accuracy in carotid artery segmentation when provided with correct prompts, such as Click mode and Box mode. Click-Box mode performs better as the prompts are more informative. In contrast, Everything mode segments all potential structures, including blood vessels, and may have difficulty depicting the major structure of interest. This suggests that ultrasound image segmentation requires more informative prompts for SAM compared to natural image segmentation in order to accurately reveal the target structure.

Another important aspect to consider in evaluating the usability of segmentation results for autonomic guidance is continuity, which determines the stability and flexibility of guidance in a continuous process (*16-18*). To assess this, the stability of the segmentation results was evaluated based on sequential images, in specific terms, the standard deviation of the segmentation results within an image sequence. For each dataset, five sets of images were selected for evaluation. Each contains 10 consecutive images. The standard deviation of the

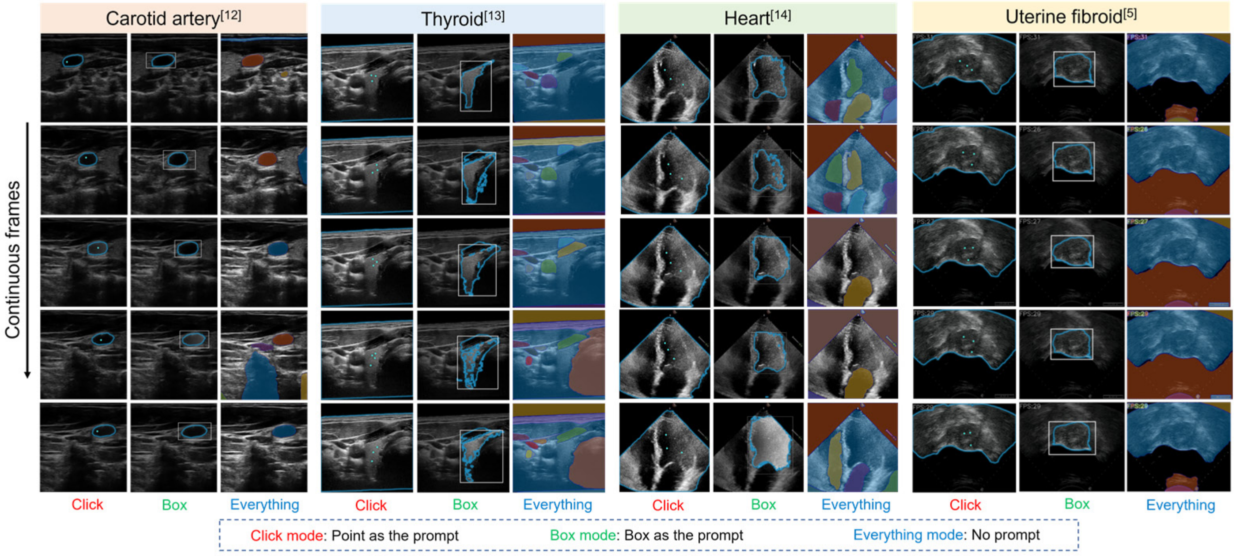

Figure 2. Qualitative results of SAM segmentation of ultrasound images. For simple and obvious structures, SAM yields accurate segmentation results and good sequence stability, even in less interactive Click mode. For less clear tissues such as the thyroid, however, SAM still fails to segment the tissue accurately in Box mode. And due to the lack of prompt input, Everything mode can automatically segment easily discerned tissues in ultrasound images but the results obtained are not those desired. Pictures were generated from: *https://segment-anything.com*

three SAM segmentation models on the Carotid artery sequence was 0.137, 0.066, and 0.065. The standard deviation of the three SAM segmentation models on the Thyroid sequence was 0.273, 0.153, and 0.150. The standard deviation of the three SAM segmentation models on the Heart sequence was 0.112, 0.117, and 0.074. Data on the Heart had a lower standard deviation than data on carotid artery segmentation, which is inconsistent with segmentation accuracy. These results indicate that while the overall accuracy of the segmentation results may be satisfactory in carotid artery images, there are still fluctuations in the stability of the continuous sequential images. These instabilities can lead to shifts and errors in autonomic guidance during imaging and can accumulate during clinical guidance. This is presumably because SAM does not currently take into account the continuity between input images, as it is designed to segment a single image without considering the contextual relationship between similar frames in video data.

The breakthrough achieved with SAM has brought possibilities for downstream tasks based on image segmentation, including fully autonomous ultrasound image-guided medical procedures. Preliminary results have revealed that SAM can achieve a good segmentation accuracy for ultrasound images with sufficient prompts, although it requires an additional interaction process compared to a specially trained model. However, the instability in continuous image segmentation, due to the lack of contextual information, introduces errors in long-term guidance. Overall, SAM has great potential as a large model for image segmentation in general-purpose tasks, but there are some areas that can be improved. First, the prompts need to be enhanced. Like ChatGPT, SAM relies on well-developed prompts to yield stable and accurate outputs, which requires a certain level of operator experience (*19-22*). This also means that fully autonomous ultrasound image guidance for different tissues still requires a lot of work. Second, the performance on medical data, such as ultrasound images, needs to be improved. Medical images have distinct imaging principles and characteristics compared to natural images, which may limit the applicability of SAM's segmentation model to ultrasound images. Therefore, developing a general segmentation model specifically for medical images is worth investigating (*23*). Moreover, continuous data, such as continuous ultrasound images or video, containing temporal information can provide effective cues for segmentation. Thus, exploring the use of continuous data in conjunction with SAM for image segmentation can be beneficial.

In conclusion, the emergence of SAM and other large models has opened up numerous possibilities for image-based downstream tasks, including an intelligent or autonomous ultrasound image guidance system. With further iteration of large models and refinement of data, the current authors believe that fully intelligent medical image-guiding systems and autonomous image-guided therapy will become a reality in the future.

Funding: This work has been supported by the National Key Research and Development Program of China (2022YFC2405200), the National Natural Science Foundation of China (82027807, U22A2051), the Beijing Municipal Natural Science Foundation (7212202), the Institute for Intelligent Healthcare, Tsinghua University (2022ZLB001), and the Tsinghua-Foshan Special Fund for Innovation (2021THFS0104).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Noble J A, and Djamal B. Ultrasound image segmentation: A survey. IEEE Trans Med Imaging. 2006; 25: 987-1010.
2. Salcudean SE, Moradi H, Black DG, Navab N. Robot-assisted medical imaging: A review. Proceedings of the IEEE. 2022; 110:951-967.
3. Ning G, Zhang X, and Liao H. Autonomic robotic ultrasound imaging system based on reinforcement learning. IEEE Trans Biomed Engineer 2021; 68:2787-2797.
4. Roshan MC, Pranata A, Isaksson M. Robotic ultrasonography for autonomous non-invasive diagnosis – A systematic literature review. IEEE Trans Med Robotics Bionics. 2022; 4:863-874.
5. Ning G, Zhang X, Zhang Q, Wang Z, Liao H. Real-time and multimodality image-guided intelligent HIFU therapy for uterine fibroid. Theranostics. 2020; 10:4676-4693.
6. Wang L, Guo D, Wang G, Zhang S. Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning. IEEE Trans Med Imaging. 2021; 40:2795-2807.
7. Brown T, Mann B, Ryder N, *et al*. Language models are few-shot learners. Advances in neural information processing systems. 2020; 33:1877-1901.
8. Haupt CE, Marks M. AI-Generated Medical Advice-GPT and Beyond. JAMA. 2023; 329:1349-1350.
9. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ and Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature. 2023; 616:259-265.
10. Kirillov A, Mintun E, Ravi N, Mao HZ, Rolland C, Gustafson L, Xiao TT, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. Segment anything. arXiv preprint arXiv:2304.02643, 2023. https://doi.org/10.48550/arXiv.2304.02643
11. Ji G, Fan D, Xu P, Cheng M, Zhou B, Gool LV, SAM struggles in concealed scenes –Empirical study on "Segment Anything". arXiv:2304.06022, 2023. https://doi.org/10.48550/arXiv.2304.06022
12. Wang G, Zhai S, Lasio G, Zhang B, Yi B, Chen S, Macvittie TJ, Metaxas D, Zhou J, Zhang S. Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung CT scans with multi-scale guided dense attention. IEEE Trans Med Imaging. 2022; 41:531-542.
13. Wunderling T, Golla B, Poudel P, Arens C, Friebe M and Hansen C, Comparison of thyroid segmentation techniques for 3D ultrasound. Proceedings of SPIE Medical Imaging, Orlando, USA, 2017.
14. Ríha K, Mašek J, Burget R, Beneš R, Závodná E. Novel method for localization of common carotid artery transverse section in ultrasound images using modified Viola-Jones detector. Ultrasound Med Biol. 2013; 39:1887-902.
15. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin PM, Grenier T, Lartizien C, Dhooge J, Lovstakken L, Bernard O. Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography. IEEE Trans Med Imaging. 2019; 38:2198-2210.
16. Ouyang D, He B, Ghorbani A, Lungren MP, Ashley EA, Liang DH, Zou JY. Echonet-dynamic: A large new cardiac motion video data resource for medical machine learning, NeurIPS ML4H Workshop: Vancouver, BC, Canada. 2019. *https://echonet.github.io/dynamic/NeuroIPS_2019_ML4H%20Workshop_Paper.pdf* (accessed May 20, 2023).
17. Ning G, Liang H, Zhang X, Liao H. Inverse-reinforcement-learning-based robotic ultrasound active compliance control in uncertain environments. IEEE Trans Indust Electron. 2023:1-10. doi:10.1109/tie.2023.3250767
18. Jiang Z, Li Z, Grimm M, Zhou MC, Esposito M, Wein W, Stechele W, Wendler T, Navab N. Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback. IEEE Trans Indust Electron. 2022; 69:7064-7075.
19. Shen Y, Tan X, Li D, Lu W, Zhuang Y. HuggingGPT: Solving AI tasks with ChatGPT and its friends in Hugging Face. arXiv:230317580. https://doi.org/10.48550/arXiv.2303.17580
20. Deng R, Cui C, Liu Q, *et al*. Segment anything model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv: 2304.04155, 2023. https://doi.org/10.48550/arXiv.2304.04155
21. Yuan B, Jiang ZL, Clarenbach R, Ghotbi R, Karlas A, Navab N. MI-SegNet: Mutual information-based US segmentation for unseen domain generalization. arXiv:2303.12649, 2023. https://doi.org/10.48550/arXiv.2303.12649
22. Jiang Z, Gao Y, Xie L, Navab N. Towards autonomous atlas-based ultrasound acquisitions in presence of articulated motion. IEEE Robot Automation Letters 7.3 2022: 7423-7430.
23. Ma J, Wang B, Segment anything in medical images, arXiv:2304.12306, 2023. https://doi.org/10.48550/arXiv.2304.12306

Received May 23, 2023; Revised June 10, 2023; Accepted June 15, 2023.

**Address correspondence to:*
Guochen Ning, Hongen Liao Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China.
E-mail: liao@tsinghua.edu.cn