Original Article

DOI: 10.5582/bst.2023.01128

The ability of Segmenting Anything Model (SAM) to segment ultrasound images

Fang Chen[1,2,*], Lingyu Chen[1,2], Haojie Han[1,2], Sainan Zhang[1,2], Daoqiang Zhang[1,2], Hongen Liao[3]

[1] Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China;

[2] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China;

[3] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China.

SUMMARY Accurate ultrasound (US) image segmentation is important for disease screening, diagnosis, and prognosis assessment. However, US images typically have shadow artifacts and ambiguous boundaries that affect US segmentation. Recently, Segmenting Anything Model (SAM) from Meta AI has demonstrated remarkable potential in a wide range of applications. The purpose of this paper was to conduct an initial evaluation of the ability for SAM to segment US images, particularly in the event of shadow artifacts and ambiguous boundaries. We evaluated SAM's performance on three US datasets of different tissues, including multi-structure cardiac tissue, thyroid nodules, and the fetal head. Results indicated that SAM generally performs well with US images with clear tissue structures, but it has limited performance in the event of shadow artifacts and ambiguous boundaries. Thus, creating an improved SAM that considers the characteristics of US images is significant for automatic and accurate US segmentation.

Keywords Segment Anything Model, ultrasound images, shadow artifacts

1. Introduction

Automatic segmentation of ultrasound (US) images can help disease screening, diagnosis, and assessment of prognosis. However, accurate US segmentation is a challenge due to the following difficulties. First, US images often suffer from a low signal-to-noise ratio (SNR) (*1*) and inhomogeneous intensity distribution (*2*). Second, shadows are a common occurrence due to inadequate contact between the US probe and the body surface or the presence of anatomical structures that interfere with the scanned tissue interfaces (*3*). These shadow regions, with their low intensity or dark pixels, are often integral to anatomical areas and lesions (*4*). As shown in Figure 1, shadow artifacts and ambiguous lesion boundaries are often observed in US images, posing significant challenges to accurate US segmentation.

Recently, the Segment Anything Model (SAM) (*5*) from Meta AI has been proposed as a promotable foundational model for natural image segmentation with minimal human intervention. SAM is a deep learning model (transformer-based) that has been trained on a huge number of images and masks - more than 1 billion masks in 11 million images. SAM is driven by various segmentation prompts (*e.g.*, points, boxes, masks) to achieve zero-shot image segmentation. Due to its promising performance in several computer vision benchmarks, SAM has garnered a great deal of attention for use in medical image segmentation tasks (*6-9*). Specifically, Deng *et al*. (*6*) conducted experiments with SAM for tumor, non-tumor tissue, and cell nuclei segmentation, and empirical results indicated that SAM is amenable to the tasks of segmenting large connected objects. He *et al*. (*7*) evaluated more than 12 medical image segmentation datasets that used 5 imaging modalities (2D X-ray, histology, endoscopy, *etc*.) and that include different organs such as the brain, chest, lungs, and skin (*8*) in an attempt to validate the out-of-the-box zero-shot capabilities of SAM with an abdominal CT organ segmentation dataset, and they examined multiple scenarios, such as marking multiple points or boxes as prompts to obtain segmentation accuracy. Hu *et al*. (*9*) concluded that the more prompts were made, the more precise segmentation results were obtained by analyzing liver tumor segmentation for contrast-enhanced computed tomography volumes.

Although the aforementioned studies investigated

SAM's performance on medical images, they lack the comprehensive and in-depth assessment of SAM's performance on US images with shadow artifacts and inhomogeneous intensity distribution. The current study evaluated SAM's performance on three US datasets of different tissues in order to perform a comprehensive analysis of SAM's performance on US images. The hope is that this study can provide the community with some insights into the future development of an improved SAM for US image segmentation.

2. Materials and Methods

2.1. Overview

Figure 2 depicts the testing pipeline for SAM as applied to various US images in this study. SAM has three main components: an image encoder, a prompt encoder, and a mask decoder. The image encoder uses the Vision Transformer (ViT) (*10*) as its backbone and is pre-trained using the masked strategy from the masked autoencoder (MAE) (*11*). Its role is to provide the embedding of the input tensor so that it can be combined with the embedding of manual prompts in subsequent steps. The prompt encoder handles various types of sparse (multiple points, boxes, or texts) and dense (masks) prompts *via* distinct branches comprising a basic convolutional neural network. Ultimately, the mask decoder uses all embedding to determine the segmentation labels.

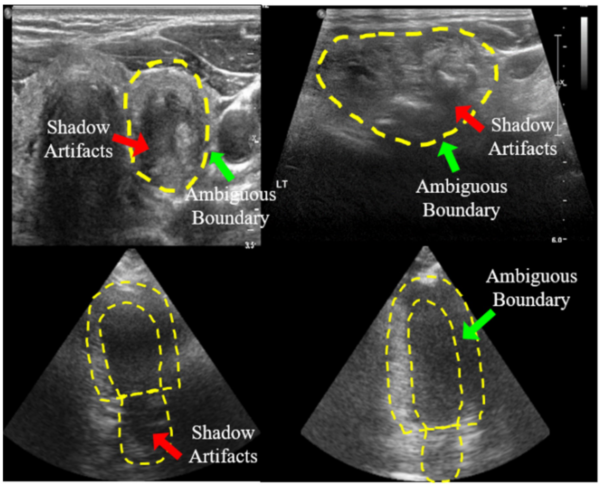

Figure 1. Difficultly of segmenting US images that contain shadow artifacts or ambiguous boundaries.

During the testing phase, SAM was comprehensively compared to related deep segmentation models using three different US datasets pertaining to various tissues, including multi-structure cardiac tissue, thyroid nodules, and the fetal head. Moreover, the testing involving US images was divided into two sets, the first consisting of images with shadow artifacts and the second consisting of clean images without any obvious shadow artifacts. This division enabled evaluation of SAM's effectiveness on US images with shadow artifacts. Moreover, four different methods of prompt selection were attempted and a regular grid of foreground points was used as prompts to generate US image segmentation results.

Positive and negative prompt points exist, indicating foreground or background points, respectively. To ensure experimental and model reproducibility, randomness, and accuracy, prompt points were chosen using the following three methods: (*i*) SAM-MPP: foreground points from the GT mask were randomly selected to serve as positive prompt points, with a range of 1-10 points; (*ii*) SAM-MPN: a background point was randomly marked as a negative point and multiple positive points were marked; (*iii*) SAM-CP: the central point of the image was identified and whether it is a positive or negative prompt

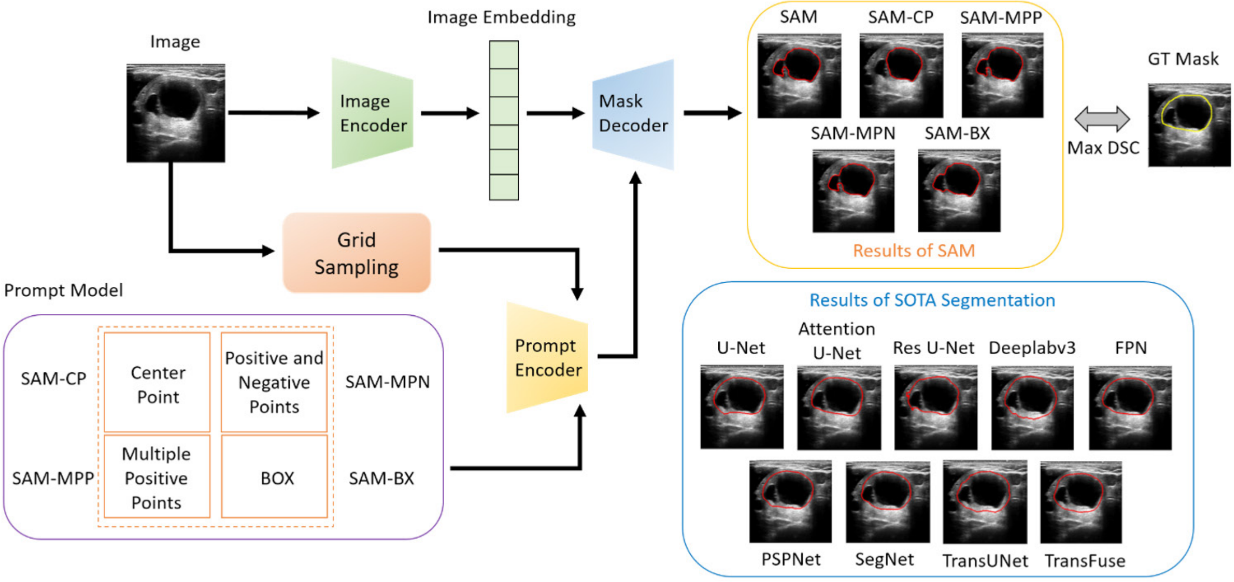

Figure 2. Testing pipeline for SAM in this study.

was determined using the GT mask. Additionally, (*iv*) SAM-BX: the bounding box of the GT mask was directly acquired without further steps. To evaluate SAM's segmentation performance, regular grid sampling was used to predict multiple masks per image and the highest quality mask was selected as the final segmentation result after comparison.

2.2. US datasets

SAM's performance in the US segmentation task was evaluated using three classic publicly available datasets. These datasets contain different scenarios of cardiac ultrasound, thyroid, and fetal head. The first dataset, CAMUS (*12*), is a large, fully annotated 2D echocardiographic assessment dataset collected from 500 patients and it includes manual annotations of the left ventricular endocardium (LV(Endo)), myocardium (LV(Epi)), and left atrium (LA) by experts. The second dataset, TN-SCUI (*13*), from the MICCAI 2020 Challenge, presents a challenging US segmentation task due to the various shapes of thyroid nodules, missing areas of lesions, ambiguous boundaries, and artifacts due to the way US is imaged. This dataset contains 3644 thyroid nodules from 3,644 patients that were manually annotated by experienced radiologists. The third dataset, HC18 (*14*), is a fetal US dataset consisting of 1,334 images used to measure the fetal head circumference.

2.3. Evaluation metrics

This study used four evaluation metrics to assess the performance of the different segmentation methods: the Dice Similarity Coefficient (DSC), Jaccard index (JI), Hausdorff Distance (HD), and Average Surface Distance (ASD).

Dice Similarity Coefficient (DSC, %) (*15*): This measures the similarity between the prediction and ground-truth sets, with a value range of [0,1]. A higher value indicates better model performance and it is often used to calculate the similarity of closed regions.

Jaccard index (JI, %) (*16*): This measures the ratio between the intersection and union of a category prediction and the ground-truth using fuzzy set theory.

Average Surface Distance (ASD, pixel) (*17*): This measures the average surface distance from all points of the prediction to the ground-truth, which assesses the surface variation between the segmentation and the GT.

Hausdorff Distance (HD, pixel) (*18*): This calculates the distance between the two sets of the prediction and ground-truth, with smaller values indicating higher similarity between the two sets. It is more sensitive to boundaries than DSC.

2.4. Methods of comparison

This study compared segmentation by SAM to several classic segmentation methods including U-Net, Attention U-Net, ResU-Net, DeepLabV3, PSPNet, SegNet, FPN, TransUnet, and TransFuse.

U-Net (*19*): U-Net is a U-shaped structure that uses skip connections to capture contextual information.

Attention U-Net (*20*): To address the problem that many redundant underlying features are extracted due to U-Net skip connections, Attention U-Net adds an attention module in skip connections to effectively suppress activations in irrelevant regions, thereby reducing the number of redundant features.

ResU-Net (*21*): ResUNet is a deep learning model based on residual connectivity for image segmentation tasks. It combines the advantages of ResNet and U-Net to better solve the problems of gradient disappearance and missing semantic information.

DeepLabV3 (*22*): DeeplabV3 provides the ability to arbitrarily control the resolution of features extracted by the encoder, with the encoder section having a large number of hole convolutions to balance accuracy and time consumption without loss of information.

PSPNet (*23*): PSPNet is based on FCN with a global mean pooling operation and feature fusion to obtain more contextual information. The features have a pyramid structure, so it is also called pyramid pooling.

SegNet (*24*): The decoder structure of SegNet performs non-linear upsampling using the pooling index computed in the maximum pooling step of the corresponding encoder. This reduces the number of parameters and operations compared to deconvolution and eliminates the need to learn upsampling.

FPN (*25*): The FPN algorithm uses both the high resolution of the lower-layer features and the semantic information of the higher-layer features to achieve prediction by fusing these different feature layers. In addition, the prediction is performed on each fused feature layer separately.

TransUnet (*26*): This is the first time that transformer was used as a promising alternative for medical image segmentation, and it has the merits of both transformers and U-Net.

TransFuse (*27*): TransFuse fuses a transformer and CNN to achieve long-range dependency modeling and reduce computational redundancy.

2.5. Implementation details

This study divided three publicly available datasets into training and testing sets in a 4:1 ratio. The training and testing datasets for each dataset are shown in Table 1.

Table 1. Number of training and testing images for TN-SCUI, CAMUS, and HC18

Dataset	TN-SCUI	CAMUS	HC18
Training Dataset Size	2,916	1,440	800
Testing Dataset Size	728	360	199

Due to the availability of training weights for SAM, this study is consistent with related studies on SAM (*6,28-29*) to evaluate SAM's performance on ultrasound images using the testing sets. Inspired by studies that compared SAM to related segmentation methods (*30-31*), segmentation models for comparison were first trained on the training set, and then corresponding segmentation metrics were obtained using the testing set.

3. Results

3.1. Segmentation results for US images overall

Quantitative comparison: Tables 2, 3, and 4 show the quantitative results of different segmentation methods in terms of the four evaluation metrics using different US datasets. In this experiment, the US images for testing were selected based on the same data partitioning of the public dataset. Results indicated that:

(1) Interestingly, SAM and SAM-CP performed poorly on the TN-SCUI dataset in terms of segmentation performance, possibly due to the unique image characteristics of the US dataset since it contains shadow artifacts and missing or unclear boundaries, hampering the model's ability to differentiate between foreground and background. However, SAM-MPP and SAM-MPN had substantially improved segmentation performance by incorporating manually labeled point prompts, achieving comparable or even better results than ResU-Net. Moreover, SAM-BX performed well on all four evaluation metrics, surpassing popular fully supervised segmentation models such as Deeplabv3, FPN, and SegNet, indicating that SAM is more effective at segmenting large connected areas. SAM-BX also benefited from the effective prompt of the bounding box, allowing the model to focus only on the box region and achieve a higher accuracy. However, providing box prompts may be time-consuming in real clinical scenarios, so the focus of this study is primarily on the segmentation results of the baseline SAM.

(2) Experiments on three different anatomical structures were conducted using CAMUS and Table 3 shows a comparative analysis of the results obtained. Results indicate that the best DSC obtained by SAM was 0.617 for the LV (Endo), 0.380 for the LV (Epi),

Table 2. Comparison to seven state-of-the-art fully-supervised methods using the TN-SCUI dataset

Items	DSC	JI	ASD	HD
U-Net	0.846	0.764	16.177	4.423
Attention U-Net	0.827	0.745	15.433	4.644
ResU-Net	0.696	0.572	44.865	10.661
Deeplabv3	0.861	0.778	13.256	4.028
FPN	0.869	0.792	12.049	4.029
PSPNet	0.858	0.780	23.516	6.183
SegNet	0.867	0.791	11.186	3.483
TransUNet	0.787	0.683	14.746	4.623
TransFuse	0.802	0.703	19.237	5.725
SAM	0.195	0.118	106.075	31.253
SAM-CP	0.345	0.257	76.281	28.056
SAM-MPP	0.716	0.603	28.419	9.222
SAM-MPN	0.721	0.605	28.536	9.029
SAM-BX	0.889	0.805	9.155	2.873

Table 4. Comparison to seven state-of-the-art fully-supervised methods using the HC18 dataset

Items	DSC	JI	ASD	HD
U-Net	0.979	0.958	3.984	1.411
Attention U-Net	0.978	0.957	4.310	1.504
ResU-Net	0.964	0.933	21.059	3.645
Deeplabv3	0.980	0.960	3.517	1.338
FPN	0.979	0.959	3.320	1.327
PSPNet	0.979	0.959	3.230	1.326
SegNet	0.980	0.960	3.108	1.304
TransUNet	0.966	0.936	6.059	2.105
TransFuse	0.974	0.950	3.654	1.532
SAM	0.539	0.380	58.622	17.771
SAM-CP	0.820	0.709	23.516	8.473
SAM-MPP	0.856	0.760	19.742	6.751
SAM-MPN	0.860	0.764	21.130	7.197
SAM-BX	0.951	0.908	8.330	2.601

Table 3. Comparison to seven state-of-the-art fully-supervised methods using the CAMUS dataset

Items	LV (Endo)				LV (Epi)				LA			
	DSC	JI	ASD	HD	DSC	JI	ASD	HD	DSC	JI	ASD	HD
U-Net	0.929	0.875	63.511	21.203	0.866	0.769	64.539	11.791	0.888	0.818	135.457	28.424
Attention U-Net	0.930	0.877	63.795	21.246	0.863	0.766	65.042	11.851	0.884	0.814	135.666	28.557
ResU-Net	0.920	0.856	68.042	22.057	0.834	0.720	67.437	13.167	0.858	0.771	135.818	29.149
Deeplabv3	0.936	0.883	64.414	21.291	0.873	0.778	65.703	11.841	0.901	0.828	135.437	28.454
FPN	0.934	0.882	65.458	21.506	0.873	0.780	66.679	11.937	0.905	0.835	136.691	28.515
PSPNet	0.937	0.886	64.289	21.165	0.875	0.783	65.515	11.709	0.899	0.932	135.567	28.321
SegNet	0.934	0.880	64.319	21.476	0.868	0.771	65.552	11.923	0.891	0.822	136.105	28.553
TransUNet	0.914	0.851	63.276	21.450	0.831	0.720	64.811	11.835	0.879	0.797	135.278	28.670
TransFuse	0.915	0.850	65.394	21.991	0.832	0.720	66.766	12.255	0.876	0.799	136.011	28.856
SAM	0.241	0.140	89.938	35.827	0.233	0.133	90.724	34.634	0.140	0.076	152.234	61.710
SAM-CP	0.555	0.413	49.529	12.554	0.256	0.148	53.860	17.010	0.161	0.089	128.767	37.364
SAM-MPP	0.595	0.452	49.035	11.639	0.280	0.165	77.102	28.331	0.428	0.319	75.077	22.800
SAM-MPN	0.617	0.470	46.164	11.182	0.290	0.172	72.311	26.111	0.477	0.356	57.552	17.158
SAM-BX	0.600	0.440	26.224	8.093	0.380	0.238	31.484	9.699	0.867	0.770	8.813	2.907

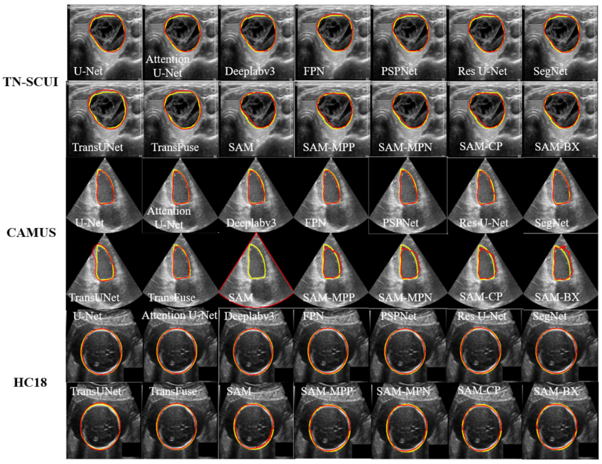

Figure 3. Good examples of segmentation on three US datasets.

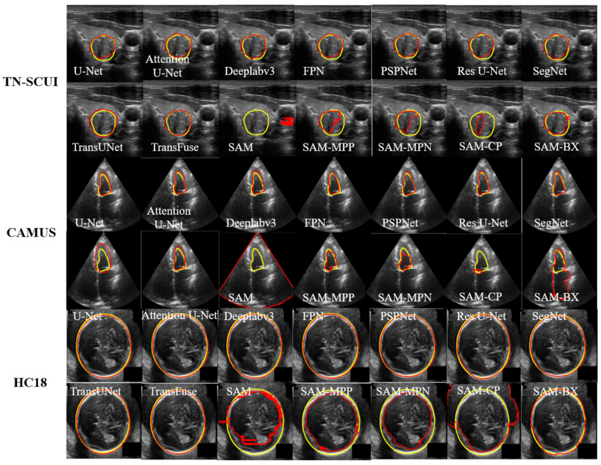

Figure 4. Bad examples of segmentation on three US datasets.

and 0.867 for the LA, which are significantly lower than the best DSC results obtained using representative deep networks (0.937, 0.875, and 0.905). This difference can be attributed to several reasons. First, the heart has complex and multiple structures that differ greatly from the shapes of thyroid nodules and the fetal head. Second, CAMUS images contain a large number of shadow artifacts and ambiguous boundaries, hampering SAM's ability to accurately distinguish between foreground and background regions. Lastly, the segmentation performance of the LV (Epi) was substantially lower compared to the LV (Endo) and LA, which may be due to interference from the LV (Endo) structure surrounding the LV (Epi).

(3) The results for HC18 are shown in Table 4, where all four types of SAM-based models, except for SAM, achieved DSC scores exceeding 0.8, and the accuracy of SAM-BX reached 0.95. This excellent performance can be attributed to the fact that the target anatomical regions in the HC18 dataset have large connected regions, which allows the additional prompts provided by the models to help achieve accurate segmentation. However, there is a noticeable gap between the ASD and HD obtained by SAM and classic fully supervised segmentation models, indicating that SAM has a weakness in fine segmentation tasks.

To summarize, the segmentation performance of basic SAM needs to be improved according to all three US datasets. This could be the result of the unique image characteristics of US, such as shadow artifacts and ambiguous or missing boundaries, posing significant challenges for SAM in identifying foreground and background regions accurately. In addition, various prompt methods were used, and results indicated that using the bounding box of the GT mask is the most effective solution. However, this approach is also very stringent and restricts the analysis and clinical application of medical images. A series of SAM models displayed better segmentation performance in areas with large connectivity and regularity but struggled with complex anatomical structures.

Qualitative comparison: Figure 3 shows some good examples of segmentation achieved by SAM, while Figure 4 displays examples where SAM failed to segment the target areas accurately. Figure 3 indicates that SAM performs as well as or better than other methods in instances where the tissue or anatomical structures are relatively distinct. However, SAM struggles with segmenting complex structures such as thyroid nodules or cardiac structures, which may be obscured by shadow artifacts or have ambiguous or missing boundaries. As shown in Figure 4, this results in low accuracy for SAM-based models, with a significant performance gap compared to popular deep models.

3.2. Segmentation results for US images with and without shadow artifacts

Tables 2-4 indicate that SAM consistently produces the worst segmentation results compared to other popular segmentation models using the TN-SCUI and CAMUS datasets. To investigate whether the presence of shadow artifacts in US images of the TN-SCUI dataset affected SAM's performance, the dataset was divided into two groups: US samples with shadow artifacts and clear US images without obvious shadow artifacts. With the guidance of a sonographer with five years of experience, segmentation results for US images with and without shadow artifacts were compared using different methods.

Quantitative comparison: As shown in Table 5, SAM methods differ most dramatically between the TN-SCUI with and without shadows. Especially with SAM, SAM-MPP, SAM-MPN, and SAM-CP, the DSC difference was 22.1% (SAM-MPN), along with maximum differences in the ASD of 18.882 and the HD of 4.919 (SAM-MPN). In addition, SAM-BX had an 8% DSC difference. The analysis of complete datasets identified the bounding box as the most

Table 5. Comparison to seven state-of-the-art fully-supervised methods using the SHADOW TN-SCUI dataset

Items	w/ Shadow Artifacts				w/o Shadow Artifacts			
	DSC	JI	ASD	HD	DSC	JI	ASD	HD
U-Net	0.828	0.734	17.663	4.920	0.865	0.796	14.605	3.897
Attention U-Net	0.805	0.713	16.736	5.305	0.851	0.780	14.086	3.962
ResU-Net	0.676	0.545	45.193	10.896	0.717	0.600	44.520	10.415
Deeplabv3	0.830	0.735	15.322	4.899	0.894	0.825	11.084	3.113
FPN	0.840	0.751	14.181	5.061	0.901	0.835	9.805	2.944
PSPNet	0.830	0.740	27.350	7.978	0.889	0.823	19.499	4.301
SegNet	0.841	0.754	12.304	4.047	0.895	0.829	10.015	2.892
TransUNet	0.773	0.663	16.580	5.587	0.832	0.745	14.101	4.182
TransFuse	0.787	0.679	19.408	5.818	0.817	0.728	19.059	5.627
SAM	0.178	0.108	107.762	33.683	0.216	0.129	104.455	28.844
SAM-CP	0.281	0.196	83.101	30.453	0.412	0.322	69.102	25.534
SAM-MPP	0.633	0.502	36.132	12.103	0.820	0.720	20.527	6.285
SAM-MPN	0.616	0.482	36.832	12.195	0.837	0.740	17.950	5.254
SAM-BX	0.811	0.705	15.923	8.714	0.889	0.828	10.243	5.989

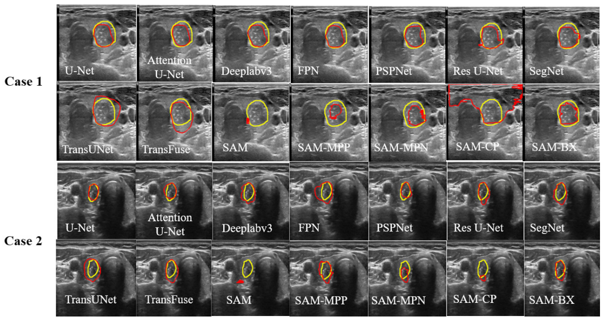

Figure 5. Failure to segment US images of the thyroid with shadow artifacts.

powerful prompt for improving SAM's segmentation performance. However, the significant difference in results between images with and without shadows persisted despite using a given GT mask's bounding box, indicating that the US dataset contains a significant number of shadow artifacts that can have a considerable impact on SAM's segmentation results.

Further analysis of other representative deep segmentation models revealed that their DSC difference did not exceed 6% (FPN) when tested on TN-SCUI with or without shadows. In addition, the differences in ASD and HD did not exceed 7.851 and 3.677, respectively (PSPNet). These findings suggest that the segmentation performance of deep models was relatively stable. Moreover, deep segmentation models had ASD and HD metrics that were generally better than those of SAM, indicating that additional refinement is necessary to enable SAM to better complete segmentation tasks for shadow artifacts and ambiguous boundaries present in US images.

Qualitative comparison: The segmentation results of different models using US images of the thyroid containing shadow artifacts are shown in Figure 5. Areas with large shadow artifacts are evident, so the segmentation results of SAM are not ideal. In areas with missing or ambiguous boundaries, segmentation results with SAMs were rough, and the corresponding ASD and HD were not as good.

4. Discussion and Conclusion

As a foundational model for image segmentation, SAM has shown great potential for natural images. This study completed an initial evaluation of SAM's ability to perform medical US image segmentation using three US image datasets of different organs, including multi-structure cardiac tissue, thyroid nodules, breast nodules, and the fetal head. This study particularly examined shadow artifacts in US images as a factor affecting SAM's accuracy. While we acknowledge the advancement of large foundational models for CV, the current experiments demonstrate that there is still room for improvement in SAM's performance on this specific task of medical US image segmentation.

Results of medical US image segmentation are compared in Tables 2-4, which indicate that everything mode is not suitable for most medical US datasets. In other words, SAM is not as accurate as dataset specific deep-learning algorithms (*19-27*) for medical US segmentation tasks. Therefore, applying the trained model from SAM directly to a medical US segmentation task will not result in satisfactory performance. In the future, more medical US images need to be used to fine-tune SAM to create a highly accurate benchmark mode.

Given that US imaging is a special imaging modality that commonly contains shadow artifacts, SAM's performance was compared on US images with and without shadow artifacts. Thus, a strength of this study is that it fully explored shadow artifacts as a factor that affect SAM's accuracy in US images. As shown by Table 5 and Figure 2, SAM had significant performance degradation with shadow artifacts. Hence, future research should investigate how to improve

SAM's reliability for US image segmentation with shadow artifacts. A possible solution (not explicitly addressed in the current work) is adding a shadow learning mechanism to SAM. Previous studies have proven that generating and injecting simulated shadows into US images and teaching them is helpful for US segmentation tasks (*32,33*). Another possible solution is using US images with shadow artifacts to finetune SAM. In summary, the current study has shown that additional work is needed to improve SAM's performance on this specific US segmentation task. The hope is that this study can provide the community with some insights into the future development of a improved SAM for US image segmentation.

Funding: This work was supported by the National Nature Science Foundation of China (62271246, U20A20389, 82027807, U22A2051), the National Key Research and Development Program of China (2022YFC2405200), the Beijing Hospital Authority's Special Funding Support for Development of Clinical Medicine (XMLX202141), and the China Postdoctoral Science Foundation (2020M671484).

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

References

1. Honarvar F, Sheikhzadeh H, Moles M, Sinclair AN. Improving the time-resolution and signal-to-noise ratio of ultrasonic NDE signals. Ultrasonics. 2004; 41:755-763.
2. Xiao G, Brady M, Noble JA, Zhang Y. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. IEEE Trans Med Imaging. 2002; 21:48-57.
3. Singla R, Ringstrom C, Hu R, Lessoway V, Reid J, Nguan C, Rohling R. Speckle and shadows: ultrasound-specific physics-based data augmentation applied to kidney segmentation. In: Medical Imaging with Deep Learning. 2022;1-10.
4. Noll M, Puhl J, Wesarg S. Achieving fluid detection by exploiting shadow detection methods. In: Imaging for Patient-Customized Simulations and Systems for Point-of-Care Ultrasound: International Workshops, BIVPCS 2017 and POCUS 2017; 121-128.
5. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y. Segment anything. arXiv:230402643. 2023;1-30. https://doi.org/10.48550/arXiv.2304.02643
6. Deng R, Cui C, Liu Q, Yao T, Remedios LW, Bao S, Landman BA, Wheless LE, Coburn LA, Wilson KT. Segment Anything Model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv:230404155. 2023; 1-6. https://arxiv.org/pdf/2304.04155.pdf
7. He S, Bao R, Li J, Grant PE, Ou Y. Accuracy of Segment-Anything Model (SAM) in medical image segmentation tasks. arXiv:230409324. 2023;1-8. https://arxiv.org/pdf/2304.09324.pdf
8. Roy S, Wald T, Koehler G, Rokuss MR, Disch N, Holzschuh J, Zimmerer D, Maier-Hein KH. SAM. MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model. arXiv:230405396. 2023;1-4. arXiv:230405396
9. Hu C, Li X. When SAM meets medical images: An investigation of Segment Anything Model (SAM) on multi-phase liver tumor segmentation. arXiv:230408506. 2023;1-5. https://doi.org/10.48550/arXiv.2304.08506
10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:201011929. 2020;1-22. https://doi.org/10.48550/arXiv.2010.11929
11. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In: Proceedings IEEE/CVF Conference Comp Vision Pattern Recog. 2022; 16000-16009.
12. Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin P-M, Grenier T. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging. 2019; 38:2198-2210.
13. Gireesha H, Nanda S. Thyroid nodule segmentation and classification in ultrasound images. Internatl J Engineer Res Tech. 2014; 21-31.
14. van den Heuvel TL, de Bruijn D, de Korte CL, Ginneken Bv. Automated measurement of fetal head circumference using 2D ultrasound images. PloS One. 2018; 13:e0200412.
15. Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, Szeskin A, Jacobs C, Mamani GEH, Chartrand G. The liver tumor segmentation benchmark (LiTS). Med Image Anal. 2023; 84:102680;1-24.
16. Crum WR, Camara O, Hill DL. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging. 2006; 25:1451-1461.
17. Dubuisson M-P, Jain AK. A modified Hausdorff distance for object matching. In: Proceedings 12th Internatl Conference Pattern Recog. IEEE, 1994; 566-568.
18. Zhou D, Fang J, Song X, Guan C, Yin J, Dai Y, Yang R. IOU loss for 2d/3d object detection. In: 2019 International Conference on 3D Vision (3DV). IEEE, 2019; 85-94.
19. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015; 234-241.
20. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B. Attention u-net: Learning where to look for the pancreas. arXiv:180403999. 2018;1-10. https://doi.org/10.48550/arXiv.1804.03999
21. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. IEEE GeosciI Remote S. 2018; 15:749-753.
22. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv:170605587. 2017;1-14. https://doi.org/10.48550/arXiv.1706.05587
23. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: Proceedings IEEE Conference Comp Vision Pattern Recog. 2017; 2881-2890.
24. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017; 39:2481-2495.

25. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings IEEE Conference Comp Vision Pattern Recog. 2017; 2117-2125.
26. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation. arXiv:210204306. 2021;1-13. https://arxiv.org/pdf/2102.04306.pdf
27. Zhang Y, Liu H, Hu Q. Transfuse: Fusing transformers and CNNs for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer, 2021; 14-24.
28. Jie L, Zhang H. When SAM meets shadow detection. arXiv:230511513. 2023;1-6. https://doi.org/10.48550/arXiv.2305.11513
29. Zhou T, Zhang Y, Zhou Y, Wu Y, Gong C. Can sam segment polyps? arXiv:230407583. 2023;1-5. https://doi.org/10.48550/arXiv.2304.07583
30. Wang A, Islam M, Xu M, Zhang Y, Ren H. SAM meets robotic surgery: An empirical study in robustness perspective. arXiv:230414674. 2023;1-5. https://doi.org/10.48550/arXiv.2304.14674
31. Williams D, MacFarlane F, Britten A. Leaf Only SAM: A Segment Anything pipeline for zero-shot automated leaf segmentation. arXiv preprint arXiv:230509418. 2023;1-9.
32. Meng Q, Sinclair M, Zimmer V, Hou B, Rajchl M, Toussaint N, Oktay O, Schlemper J, Gomez A, Housden J. Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans Med Imaging. 2019; 38:2755-2767.

33 Yasutomi S, Arakaki T, Matsuoka R, Sakai A, Komatsu R, Shozu K, Dozen A, Machino H, Asada K, Kaneko S. Shadow estimation for ultrasound images using auto-encoding structures and synthetic shadows. Applied Sciences. 2021; 11:1127-1147.

Received June 2, 2023; Revised June 15, 2023; Accepted June 20, 2023.

**Address correspondence to:*
Fang Chen, Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing211016, Jiangsu, China.
E-mail: chenfang@nuaa.edu.cn

Released online in J-STAGE as advance publication June 22, 2023.