This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

FedFMS: Exploring Federated Foundation Models for Medical Image Segmentation

Yuxi Liu[1], Guibo Luo[1](✉), and Yuesheng Zhu[1](✉)

School of Electronic and Computer Engineering, Peking University, Shenzhen, China
✉Correspondences: {luogb,zhuys}@pku.edu.cn

Abstract. Medical image segmentation is crucial for clinical diagnosis. The Segmentation Anything Model (SAM) serves as a powerful foundation model for visual segmentation and can be adapted for medical image segmentation. However, medical imaging data typically contain privacy-sensitive information, making it challenging to train foundation models with centralized storage and sharing. To date, there are few foundation models tailored for medical image deployment within the federated learning framework, and the segmentation performance, as well as the efficiency of communication and training, remain unexplored. In response to these issues, we developed Federated Foundation models for Medical image Segmentation (FedFMS), which includes the Federated SAM (FedSAM) and a communication and training-efficient Federated SAM with Medical SAM Adapter (FedMSA). Comprehensive experiments on diverse datasets are conducted to investigate the performance disparities between centralized training and federated learning across various configurations of FedFMS. The experiments revealed that FedFMS could achieve performance comparable to models trained via centralized training methods while maintaining privacy. Furthermore, FedMSA demonstrated the potential to enhance communication and training efficiency. Our model implementation codes are available at https://github.com/LIU-YUXI/FedFMS.

Keywords: Medical image segmentation · Federated learning · Foundation model.

1 Introduction

Medical image segmentation aims to identify and separate structures or regions in medical images [25, 16], which is crucial for clinical care. Very recently, the Segmentation Anything Model (SAM) [7] has garnered widespread attention as a powerful foundation model for visual segmentation. Many works fine-tuning SAM in the medical images have achieved advanced results, such as the Medical SAM Adapter (MSA) [26], 3DSAM-adapter [5] and [27].

However, medical imaging data typically contain privacy-sensitive information, making it difficult to centralized storage and sharing [11, 22]. Moreover, training large models often involves data that is distributed across various geographic locations or institutes. Transmitting large volumes of data can lead to increased communication costs and delays in transmission. Federated learning [8] offers a solution by enabling model training on distributed datasets without the need to centralize data in one location [13]. Additionally, distributed training of large models allows for the distribution of computational requirements [19].

To date, deploying foundation models for medical images within the federated learning framework is rare. There are two main issues that remain unexplored: First, can foundation models trained based on federated learning harness the powerful capabilities of foundation models, and maintain performance comparable to those trained based on centralized training when facing Non-Independent and Identically Distributed (Non-IID) datasets? Second, the federated learning of foundation models requires significant communication resources and training costs, is there a more efficient method for its federated learning training?

To address the above issues, we have collected a large number of real multi-center medical datasets and developed Federated Foundation models for Medical image Segmentation (**FedFMS**) to investigate both its performance of segmentation and training efficiency. FedFMS includes two federated foundation models, the Federated SAM (FedSAM) and a communication and training-efficient Federated SAM with Medical SAM Adapter (FedMSA). For FedSAM, we fine-tune all parameters of the pre-trained SAM on each client. For FedMSA, we efficiently fine-tune the parameters of the adapters and decoder of the pre-trained MSA on each client. Then, we aggregate the parameters on the global server using the FedAvg [14] algorithm. To our knowledge, this study is the first comprehensive investigation into the application of federated foundation models within the medical domain. Our contributions can be summarized as follows:

(1) **Dataset Collection**. We have collected various multi-institutional datasets to serve as benchmarks for evaluating the performance of federated foundation models in medical image segmentation. This offers comprehensive and reliable evaluation data for federated medical segmentation.

(2) **Model Development**. We have developed a federated learning framework named FedSAM based on the foundation model SAM, which enables distributed training of medical images and demonstrates stability and effectiveness. To further explore more efficient methods, we have also built the FedMSA framework. These models could serve as baselines and be beneficial for further promoting the federated foundation models for medical image segmentation.

(3) **Experimental Analysis**. We have conducted an in-depth investigation into the performance disparities between centralized training and federated learning across different configurations of FedFMS with various datasets. Our investigation will provide an insight overview of the feasibility and effectiveness of a federated large-scale model for medical images in real-world clinical settings.

2 Method

2.1 Preliminary Methods

SAM Architecture SAM is a large data-driven image segmentation model. It constructs a dataset named SA-1B, consisting of 11 million images and 100

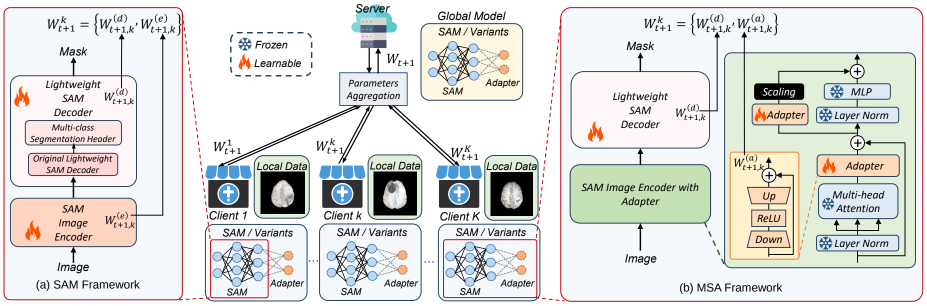

Fig. 1. The illustration of FedSAM and FedMSA. FedSAM is a federated learning framework with (a) SAM. FedMSA is a federated learning framework with (b) MSA.

million masks, to drive its training. The image encoder utilizes a standard Vision Transformer (ViT) [2, 23] pre-trained by Masked Autoencoders. In our study, we set SAM using the ViT-B/16 and ViT-L/16 variants. ViT-B/16 represents the base-scale version of ViT, implemented with 768 convolutional kernels. ViT-L represents the large-scale version, implemented with 1024 convolutional kernels. The output of the image encoder is a 16× downsampled embedding of the input image. The mask decoder is a lightweight modified Transformer decoder block that includes bidirectional cross-attention and a dynamic mask prediction head.

MSA Architecture MSA efficiently fine-tunes the SAM architecture for medical images to enhance its medical capabilities. Fine-tuning allows the model to retain the knowledge gained from extensive data while strengthening its abilities in new domains. In the encoder, MSA freezes the pre-trained SAM parameters and inserts two adapter modules at each ViT block. The adapter is a bottleneck model that sequentially uses down-projection, ReLU activation, and up-projection. Both projections are implemented using simple MLP layers. MSA's decoder is the same as SAM's. MSA fine-tunes all parameters of the decoder.

2.2 Federated Foundation Models

Figure 1 illustrates the FedFMS framework, comprising multiple clients for local training and a server for parameter aggregation, all utilizing the same foundational model (*e.g.*, SAM or MSA). The federated learning of SAM and its more efficient variant MSA corresponds to FedSAM and FedMSA, respectively.

Client-side Model Training Each client possesses a fixed local dataset and sufficient computational resources to perform mini-batch updates. The number of clients is K. Each client adopts the same BCE loss and the same model (SAM or MSA), which is initialized with pre-trained SAM parameters before training.
FedSAM To achieve simultaneous segmentation of multiple classes, we omit the input prompts and prompt encoder, perform a multi-class segmentation header

by adopting a two-dimensional convolution with a 1×1 kernel after the original SAM decoder, mapping the output mask to $H \times W \times c$, where c is the number of segmentation classes, H is the height and W is the width of the predictive mask. Our SAM is shown in Figure 1 (a). The local SAM is initialized by global parameters $W_t = \{W_t^{(d)}, W_t^{(e)}\}$, where $W_t^{(e)}$ is the parameters of the encoder and $W_t^{(d)}$ is the parameters of the decoder. In the k-th client, the updated parameters is $W_{t+1}^k = \{W_{t+1,k}^{(d)}, W_{t+1,k}^{(e)}\}$, where $W_{t+1,k}^{(e)}$ is the updated parameters of the encoder and $W_{t+1,k}^{(d)}$ is the updated parameters of the decoder.
FedMSA For MSA, we fine-tune the parameters of adapters in the encoder (denoted as $W_t^{(a)}$), and all parameters in the decoder. Our MSA's decoder adopts the same multi-class segmentation decoder as our SAM. The features obtained by fine-tuning the encoder propagate to the top layers of the decoder, so all parameters of the decoder need to be fine-tuned. The parameter of the SAM decoder is lightweight, resulting in a low fine-tuning cost. The structure of MSA is shown in Figure 1 (b). We use MSA to build a more efficient federated learning framework for three reasons. (1) MSA performs well in fine-tuning tasks for medical image segmentation. (2) Only training adapters and decoder requires less computational cost compared to training the entire SAM. (3) During global parameter aggregation, only the parameters of the adapters and decoder need to be transmitted and calculated. During the federated t-th round, the local model is initialized by fetching global model parameters $W_t = \{W_t^{(d)}, W_t^{(a)}\}$ from the server. In the k-th client, the updated parameters is $W_{t+1}^k = \{W_{t+1,k}^{(d)}, W_{t+1,k}^{(a)}\}$, where $W_{t+1,k}^{(a)}$ is the updated parameters of the adapters.

Server-side Model Aggregation The server distributes a global model and receives synchronized updates from all clients at each federated round. We use FedAvg as the aggregation method. The aggregation is formalized as $W_{t+1} \leftarrow \frac{1}{\sum_{k=1}^K N_k^{(local)}} \sum_{k=1}^K (N_k^{(local)} \cdot W_{t+1}^k)$, where $N_k^{(local)}$ is the amount of data in client k. Though more complex algorithms could also be considered, FedAvg has shown good performance due to the strong generalization capabilities of SAM.

2.3 Dataset Preparation

We collected and constructed four Non-IID federated learning datasets with different modalities and types from public datasets. The example cases and sample numbers of each data are presented in Figure 2. Following the methodology of MSA, all images are preprocessed to a shape of $1024 \times 1024 \times 3$ before input, and the size of the output mask is 256×256.

- **Prostate Cancer**. Extracted from public prostate cancer MRI imaging datasets from various medical institutions [12, 10] and NCI-ISBI 2013.
- **Brain Tumor**. Derived from FeTS2022 [18], which is a collection of multi-institutional clinical acquisition mp-MRI scans of gliomas. The segmentation target we use is the GD-enhancing tumor (ET - label 4) on T1ce images.

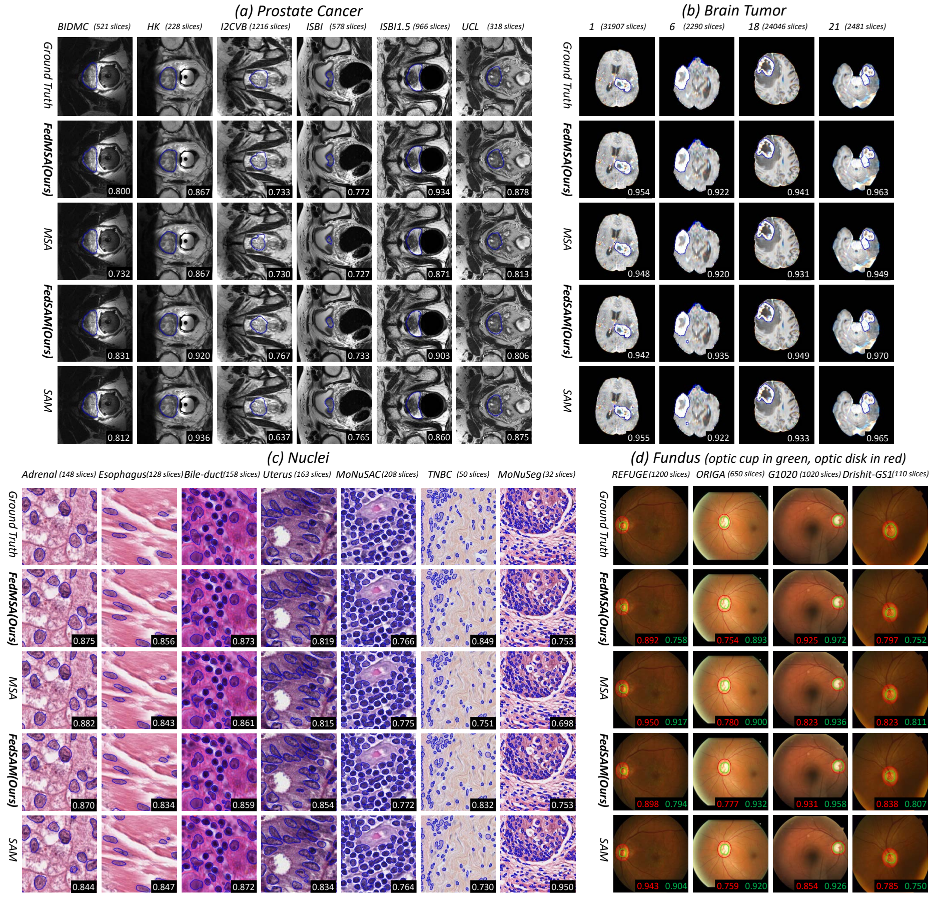

Fig. 2. The example of images and Ground Truth from different clients across four datasets (a-d) and the comparison in the results of FedMSA, MSA, FedSAM and SAM. The bottom right corner of each image indicates the Dice for it or the whole nii.

- **Nuclei**. It is a nuclei segmentation dataset from [4, 3, 24, 15, 9]. Cells from different tissues in the PanNuke dataset are distributed across different clients.
- **Fundus**. Gathered from four distinct fundus photography images datasets [21, 17, 1, 28] for optic cup (OC) and optic disc (OD) segmentation tasks.

Brain Tumor and Prostate Cancer images are in 3D nii format, while SAM can only handle 2D images. Therefore, we slice them along the depth direction, converting a 3D image with depth d into d slices of 2D images. Since the maximum pixel value in nii format is much larger than 255, we calculate the 1st percentile of pixel values for each nii file as the upper bound of high-intensity pixels and perform linear normalization. Other RGB images are also linearly normalized to the range (0,1) with 255 as the maximum value.

3 Experiments

As the first study to investigate image segmentation foundation models with federated learning, we conducted experiments to investigate three questions:
(1) FedFMS performance: Can SAM maintain its capabilities when trained under federated learning, comparable to SAM under centralized training?
(2) Model Efficiency: Is FedMSA a more efficient and cost-saving method in federated learning training? How does its performance compare to FedSAM?
(3) Pre-training Impact: Can pre-training on large dataset enrich the prior knowledge of our federated foundation models, thereby surpassing conventional ones?

3.1 Experimental Settings

We adopt two commonly used metrics, Dice (the Dice coefficient) and IOU (Intersection over Union), to quantitatively evaluate the segmentation results. We treat the dataset of each client as an unseen test set, while the data of each remaining client is divided into training and validation sets at a ratio of 9:1.

In the federated learning process, all clients use the same hyper-parameter settings, and the local model is trained using Adam optimizer with a batch size of 6. The momentum parameters for Adam are set to 0.9 and 0.999, respectively. The pre-trained model utilized is provided by SAM publicly. We conducted a total of 100 federated training rounds, with each local epoch set to 1. The framework is implemented using PyTorch and trained on NVIDIA A800.

3.2 Results

Overall Comparison To explore the impact of federated learning on foundation model, we compare FedMSA with MSA and compare FedSAM with SAM. We use FedU-Net and FednnU-Net as baselines. U-Net [20] is a commonly used and effective convolutional network for biomedical image segmentation. nnU-Net [6] is a more robust method compared to U-Net. We use the pre-trained ViT-B architecture of SAM as default. FedMSA-L and SAM-L are extension experiments using ViT-L. Results are presented in Table 1 and Figure 2.
FedFMS performance Both federated foundation models (FedMSA, FedSAM) and non-federated foundation models (MSA, SAM) achieve promising results across various tasks. FedSAM, which fine-tunes all parameters, outperforms FedMSA in Prostate, Nuclei, and Fundus segmentation. In Brain Tumor dataset, FedMSA outperforms FedSAM. FedMSA-L and MSA-L with larger parameter count also perform similarly, and both outperform models based on Vit-B. The results demonstrate the potential of SAM in federated learning for medical image segmentation. This suggests the feasibility of further extending advanced federated learning algorithms to foundation models for the medical image domain.

The performance of FedSAM, FedMSA and FedMSA-L is significantly higher than FedU-Net and FednnU-Net. The preprocessing of nnU-Net under federated learning is limited in effectiveness, and its training on Non-IID datasets is also unstable. Using pre-trained SAM is beneficial for medical image segmentation as it can mitigate unseen domain issues, attributed to its abundant background knowledge. The varying data quantities and distributions across clients result in inconsistent convergence directions among different clients. This inconsistency further leads to suboptimal performance of the globally aggregated model on the server. The foundation model demonstrates higher robustness and stability, which can alleviate the above issues. Therefore, through fine-tuning SAM, FedMSA, FedSAM and FedMSA-L can achieve advanced performance.

Table 1. The comparison of FedSAM, SAM, FedMSA, MSA, FedU-Net, and ablation variants (denoted as *italics*) on different medical image datasets. The p-values between FedSAM and SAM, as well as between FedMSA and MSA, are both greater than 0.5.

Dataset	Prostate Cancer													
Client	BIDMC		HK		I2CVB		ISBI		ISBI1.5		UCL		Average	
Model	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU
FedU-Net	0.498	0.498	0.684	0.645	0.034	0.023	0.649	0.590	0.671	0.640	0.563	0.542	0.516	0.490
FednnU-Net	0.457	0.375	0.602	0.534	0.448	0.372	0.747	0.681	0.612	0.544	0.671	0.591	0.590	0.516
FedSAM	0.810	0.774	0.841	0.808	**0.798**	**0.768**	0.837	0.792	0.785	0.754	0.844	0.809	0.819	0.784
SAM	0.793	0.758	0.837	0.799	0.731	0.697	0.812	0.759	0.786	0.754	0.848	0.807	0.801	0.762
FedSAM (-PT)	0.688	0.493	0.542	0.490	0.583	0.567	0.459	0.393	0.378	0.352	0.474	0.449	0.521	0.457
SAM (-PT)	0.515	0.500	0.679	0.619	0.511	0.486	0.577	0.496	0.517	0.470	0.520	0.471	0.553	0.507
FedMSA	0.769	0.737	0.809	0.777	0.754	0.720	0.821	0.771	0.795	0.765	0.859	0.820	0.801	0.765
MSA	0.749	0.711	0.813	0.775	0.748	0.716	0.803	0.758	0.782	0.752	0.841	0.801	0.789	0.752
FedMSA (-PT)	0.499	0.499	0.527	0.495	0.582	0.564	0.565	0.481	0.525	0.504	0.419	0.400	0.520	0.491
MSA (-PT)	0.546	0.522	0.422	0.389	0.474	0.444	0.495	0.423	0.476	0.445	0.482	0.427	0.482	0.442
FedMSA-L	0.806	0.777	**0.869**	**0.838**	0.772	0.743	**0.838**	**0.793**	**0.821**	**0.793**	0.867	**0.834**	**0.829**	**0.796**
MSA-L	**0.810**	**0.779**	0.845	0.814	0.764	0.742	0.836	0.792	0.811	0.782	**0.869**	0.834	0.823	0.791

Dataset	Brain Tumor										Fundus			
											REFUGE			
Client	1		6		18		21		Average		OD		OC	
Model	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU
FedU-Net	0.860	0.822	0.851	0.806	0.861	0.824	0.857	0.817	0.857	0.817	0.848	0.743	0.840	0.733
FednnU-Net	0.772	0.711	0.804	0.740	0.781	0.721	0.785	0.727	0.786	0.725	0.843	0.733	0.796	0.671
FedSAM	0.869	0.831	0.880	0.836	0.879	0.839	0.860	0.822	0.872	0.832	0.869	0.772	**0.873**	0.781
SAM	0.867	0.830	0.877	0.833	0.876	0.838	0.849	0.809	0.867	0.828	0.859	0.758	0.855	0.758
FedSAM (-PT)	0.809	0.767	0.853	0.807	0.830	0.790	0.832	0.791	0.831	0.789	0.857	0.756	0.842	0.736
SAM (-PT)	0.827	0.785	0.851	0.803	0.838	0.798	0.818	0.774	0.834	0.790	0.833	0.721	0.821	0.710
FedMSA	0.877	0.838	0.876	0.832	0.884	0.847	0.862	0.823	0.875	0.835	0.860	0.760	0.858	0.763
MSA	0.876	0.837	0.871	0.828	0.883	0.845	0.853	0.814	0.871	0.831	**0.881**	0.792	0.866	0.773
FedMSA (-PT)	0.808	0.767	0.856	0.811	0.811	0.770	0.826	0.784	0.825	0.783	0.846	0.742	0.842	0.736
MSA (-PT)	0.837	0.797	0.849	0.804	0.844	0.805	0.830	0.790	0.840	0.799	0.845	0.739	0.829	0.719
FedMSA-L	**0.887**	**0.850**	**0.883**	**0.841**	**0.895**	**0.859**	0.876	0.840	**0.885**	**0.847**	0.869	0.772	0.869	**0.867**
MSA-L	0.886	0.850	0.875	0.832	0.890	0.854	**0.877**	**0.840**	0.882	0.844	0.879	**0.794**	0.817	0.739

Dataset	Fundus (Continued Table)															
Client	ORIGA				G1020				Drishit-GS1				Average			
	OD		OC		OD		OC		OD		OC		OD		OC	
Model	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU
FedU-Net	0.794	0.667	0.830	0.722	0.550	0.427	0.449	0.346	0.677	0.529	0.779	0.650	0.717	0.592	0.725	0.613
FednnU-Net	0.771	0.636	0.804	0.686	0.541	0.404	0.406	0.297	0.679	0.536	0.725	0.607	0.709	0.577	0.683	0.565
FedSAM	0.816	0.698	0.846	0.748	0.717	0.602	0.556	0.456	0.715	0.571	0.777	0.643	0.779	0.661	0.763	0.657
SAM	0.820	0.703	0.831	0.729	0.621	0.512	0.546	0.444	0.676	0.526	0.767	0.632	0.744	0.625	0.750	0.641
FedSAM (-PT)	0.765	0.628	0.811	0.697	0.420	0.314	0.347	0.266	0.537	0.395	0.640	0.490	0.645	0.523	0.660	0.547
SAM (-PT)	0.777	0.645	0.778	0.777	0.482	0.364	0.432	0.335	0.509	0.371	0.718	0.579	0.650	0.525	0.687	0.600
FedMSA	0.822	0.705	**0.850**	**0.753**	0.733	0.614	0.560	0.463	0.692	0.545	0.769	0.632	0.777	0.656	0.759	0.653
MSA	0.834	0.723	0.844	0.746	0.693	0.573	0.551	0.449	0.716	**0.576**	**0.798**	**0.672**	0.781	0.666	0.765	0.660
FedMSA (-PT)	0.789	0.661	0.817	0.704	0.475	0.362	0.465	0.374	0.557	0.408	0.683	0.536	0.667	0.543	0.702	0.587
MSA (-PT)	0.804	0.683	0.811	0.704	0.573	0.451	0.489	0.393	0.481	0.347	0.683	0.536	0.676	0.555	0.703	0.588
FedMSA-L	0.836	0.725	0.845	0.748	0.732	0.630	**0.587**	**0.494**	0.704	0.559	0.762	0.625	0.785	0.672	**0.766**	**0.684**
MSA-L	**0.855**	**0.754**	0.846	0.750	**0.734**	**0.632**	0.582	0.485	**0.717**	0.573	0.776	0.643	**0.796**	**0.688**	0.755	0.654

Dataset	Nuclei															
Client	Adrenal		Esophagus		Bile-duct		Uterus		MoNuSAC		TNBC		MoNuSeg		Average	
Model	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU	Dice	IOU
FedU-Net	0.742	0.618	0.781	0.655	0.717	0.598	0.777	0.645	0.559	0.404	0.703	0.548	0.678	0.519	0.708	0.569
FednnU-Net	0.798	0.684	0.807	0.690	0.754	0.642	0.803	0.681	0.588	0.435	0.747	0.606	0.713	0.563	0.744	0.614
FedSAM	0.810	0.698	0.807	0.693	0.765	0.659	**0.832**	**0.717**	0.623	0.472	**0.776**	**0.643**	0.746	0.602	0.765	0.640
SAM	0.819	0.709	0.777	0.655	0.768	0.665	0.824	0.706	0.606	0.457	0.709	0.564	0.677	0.518	0.740	0.611
FedSAM (-PT)	0.666	0.533	0.736	0.597	0.690	0.561	0.742	0.598	0.572	0.415	0.636	0.485	0.629	0.468	0.667	0.523
SAM (-PT)	0.686	0.548	0.711	0.571	0.669	0.539	0.713	0.564	0.587	0.434	0.667	0.513	0.643	0.479	0.668	0.521
FedMSA	0.806	0.694	0.802	0.688	0.763	0.655	0.805	0.682	0.640	0.490	0.730	0.600	0.741	0.598	0.755	0.630
MSA	0.813	0.702	0.805	0.693	0.769	0.664	0.824	0.707	0.629	0.477	0.630	0.486	0.665	0.506	0.733	0.605
FedMSA (-PT)	0.692	0.558	0.728	0.589	0.684	0.556	0.739	0.596	0.562	0.406	0.628	0.477	0.642	0.482	0.668	0.523
MSA (-PT)	0.649	0.507	0.662	0.523	0.696	0.552	0.662	0.527	0.559	0.406	0.622	0.458	0.658	0.517	0.644	0.499
FedMSA-L	0.811	0.701	0.805	0.699	0.767	0.662	0.824	0.706	**0.644**	**0.494**	0.769	0.635	**0.761**	**0.622**	**0.769**	**0.646**
MSA-L	**0.820**	**0.711**	**0.814**	**0.703**	**0.802**	**0.703**	0.823	0.704	0.630	0.480	0.683	0.542	0.700	0.547	0.753	0.627

Table 2. Model efficiency analysis on FedMSA and FedSAM.

Learnable Parameter		Training Time		GPU Memory Usage		FLOPs		Predicting Time	
FedMSA	FedSAM	FedMSA	FedSAM	FedMSA	FedSAM	FedMSA	FedSAM	FedMSA	FedSAM
14.7 B	93.7 B	739.9 min	911.4 min	52,274 MiB	58,478 MiB	5.7 T	13.4 T	0.127 s	0.118 s

Further Discussion The test results of FedMSA and MSA are generally similar across various datasets, and the performance of FedSAM and SAM is also similar. The federated learning paradigm leads to slight differences in their performance. The discrepancies are slightly larger in their tests on clients TNBC and MoNuSeg for Nuclei segmentation. The Nuclei dataset has the smallest dataset among all the datasets. Moreover, there are large differences among different types of cells. These factors lead to inconsistent convergence directions in federated learning.

Model Efficiency Analysis We calculate the learnable parameter count, training time in GPU (the average training time in Fundus dataset), GPU memory usage in each client, and the estimated FLOPs (Floating Point Operations) for both forward and backward propagation for FedMSA and FedSAM, as shown in Table 2. The number of parameters (denoted as n) to be trained and updated determines the model's training speed and communication cost. The amount of parameters to be communicated in each round of federated learning is $2n$. The results show that FedMSA freezes a large number of parameters in the encoder, resulting in a significantly reduced parameter count and FLOPs compared to FedSAM, and consequently reducing communication and training costs. We calculate the average time required to predict each 2D image, as shown in "Predicting time" in Table 2. During the prediction process, FedMSA takes a bit longer because it has more parameters from adapters compared to FedSAM.

Pre-training Impact SAM is pre-trained on a large-scale natural dataset. To investigate the effectiveness of this pretraining for medical image segmentation and its impact on federated learning, we conducted an ablation study. For FedSAM, FedMSA, SAM and MSA, we constructed variants *FedSAM (-PT)*, *FedMSA (-PT)*, *SAM (-PT)* and *MSA (-PT)* without using pre-trained parameters, and the experimental results are illustrated in Table 1. The results show that not using pre-trained parameters from SAM leads to a drastic decrease in performance. In some cases, *FedSAM (-PT)* performs worse than *SAM (-PT)*, which is due to the convergence of no pre-trained SAM under federated learning is not stable. *FedSAM (-PT)* and *FedMSA (-PT)* sometimes perform worse than the lightweight model FedU-Net, for example in experiments Adrenal, Esophagus, Bile-duct, Uterus, TNBC, MoNuSeg on Nuclei dataset and HK, ISBI, ISBI1.5, UCL on Prostate Cancer dataset. This indicates that the pretraining knowledge of SAM is crucial for its effectiveness under the federated learning paradigm, which enables our federated foundation models to far surpass the traditional federated learning models (*e.g.*, FednnU-Net). The code for our implementation of FednnU-Net is available at https://github.com/LMIAPC/FednnU-Net.

4 Conclusion

In this study, we propose a solution to deploy the foundation model SAM and its efficient variant MSA within the federated learning framework, referred to as FedSAM and FedMSA respectively. We collected various multi-institutional federated datasets for our experiment. By leveraging rich pre-training knowledge, FedSAM and FedMSA demonstrate excellent performance in addressing the inherent training issues of federated learning, achieving comparable results to the foundation models in centralized training. Additionally, we conducted an efficiency analysis between FedMSA and FedSAM. Our study is the first to introduce foundation models for federated learning in the medical image domain. It will encourage the integration of more foundation models into privacy-preserving federated learning frameworks, which holds profound practical significance.

Acknowledgments. This work is supported by Shenzhen Science and Technology Program (No.JCYJ20230807120800001), and 2023 Shenzhen sustainable supporting funds for colleges and universities (No.20231121165240001). The authors sincerely appreciate the computing environment supported by the China Unicom Shenzhen Intelligent Computing Center.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bajwa, M.N., Singh, G.A.P., Neumeier, W., Malik, M.I., Dengel, A., Ahmed, S.: G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–7. IEEE (2020)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
3. Gamper, J., Koohbanani, N.A., Benet, K., Khuram, A., Rajpoot, N.: Pannuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In: European Congress on Digital Pathology. pp. 11–19. Springer (2019)
4. Gamper, J., Koohbanani, N.A., Graham, S., Jahanifar, M., Benet, K., Khurram, S.A., Azam, A., Hewitt, K., Rajpoot, N.: PanNuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778 (2020)
5. Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.A., Dou, Q.: 3DSAM-adapter: Holistic Adaptation of SAM from 2D to 3D for Promptable Medical Image Segmentation. arXiv preprint arXiv:2306.13465 (2023)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
7. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)

8. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 (2016)
9. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE transactions on medical imaging **36**(7), 1550–1560 (2017)
10. Lemaître, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M., Meriaudeau, F.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. Computers in biology and medicine **60**, 8–31 (2015)
11. Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M.J., et al.: Privacy-preserving federated brain tumour segmentation. In: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10. pp. 133–141. Springer (2019)
12. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the PROMISE12 challenge. Medical image analysis **18**(2), 359–373 (2014)
13. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1013–1023 (2021)
14. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
15. Naylor, P., Laé, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE transactions on medical imaging **38**(2), 448–459 (2018)
16. Norouzi, A., Rahim, M.S.M., Altameem, A., Saba, T., Rad, A.E., Rehman, A., Uddin, M.: Medical image segmentation methods, algorithms, and applications. IETE Technical Review **31**(3), 199–213 (2014)
17. Orlando, J.I., Fu, H., Breda, J.B., Van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis **59**, 101570 (2020)
18. Pati, S., Baid, U., Zenk, M., Edwards, B., Sheller, M., Reina, G.A., Foley, P., Gruzdev, A., Martin, J., Albarqouni, S., et al.: The federated tumor segmentation (fets) challenge. arXiv preprint arXiv:2105.05874 (2021)
19. Rauniyar, A., Hagos, D.H., Jha, D., Håkegård, J.E., Bagci, U., Rawat, D.B., Vlassov, V.: Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions. IEEE Internet of Things Journal (2023)
20. 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. pp. 234–241. Springer (2015)
21. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers **2**(1), 1004 (2015)

22. Sohan, M.F., Basalamah, A.: A Systematic Review on Federated Learning in Medical Image Analysis. IEEE Access (2023)
23. Steiner, A., Kolesnikov, A., Zhai, X., Wightman, R., Uszkoreit, J., Beyer, L.: How to train your vit? data, augmentation, and regularization in vision transformers. arXiv preprint arXiv:2106.10270 (2021)
24. Verma, R., Kumar, N., Patil, A., Kurian, N.C., Rane, S., Graham, S., Vu, Q.D., Zwager, M., Raza, S.E.A., Rajpoot, N., et al.: MoNuSAC2020: A multi-organ nuclei segmentation and classification challenge. IEEE Transactions on Medical Imaging **40**(12), 3413–3423 (2021)
25. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET Image Processing **16**(5), 1243–1267 (2022)
26. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
27. Xie, W., Willems, N., Patil, S., Li, Y., Kumar, M.: SAM Fewshot Finetuning for Anatomical Segmentation in Medical Images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3253–3261 (2024)
28. Zhang, Z., Yin, F.S., Liu, J., Wong, W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual international conference of the IEEE engineering in medicine and biology. pp. 3065–3068. IEEE (2010)