Federated Medical Image Segmentation under Modality Heterogeneity via Specialized Adapters

Obed Jamir[1(✉)] and Angshuman Paul[1]

Indian Institute of Technology Jodhpur, Rajasthan, India
✉Correspondence: d23cse004@iitj.ac.in

Abstract. Differences in imaging modalities across institutions (e.g., MRI and CT) introduce significant challenges in deploying federated learning in real-world settings for medical image segmentation. To address this challenge, we propose FedMoSA, a robust federated framework for handling severe modality heterogeneity, where clients may be fully modality-disjoint. FedMoSA explicitly decouples modality-specific and modality-agnostic representation learning through the integration of specialized adapters and shared adapters within the image encoder of SAM(Segment Anything Model). To enable stable optimization under modality heterogeneity, we introduce modality-restricted aggregation for specialized adapters and modality-balanced aggregation for shared components. Furthermore, we incorporate a prototype-based alignment mechanism to promote consistent shared adapter representations across clients. We evaluate FedMoSA on publicly available multi-institution, multi-modal liver segmentation datasets under diverse federated configurations, including explicit generalizability testing. FedMoSA achieves up to +8.5 Dice improvement under fully modality-disjoint setting over state-of-the-art federated medical image segmentation methods, with a substantial reduction in inter-client performance variance. The code is available at https://github.com/obedjamir/FedMoSA.

Keywords: Medical Image Segmentation · Modality Heterogeneous · Federated Learning.

1 Introduction

In practical federated learning scenarios for medical image segmentation, clients often have image data from different imaging modalities (e.g., MRI or CT). In the fully modality-disjoint case, each client contains only a single modality. Aggregating modality-specific updates across such clients can introduce negative transfer and cross-modality interference, leading to performance degradation. Recent studies, such as FedMSA [6] and FednnU-Net [11], have adopted FL specifically for medical image segmentation. However, these approaches generally assume homogeneous modality distributions across clients and are not explicitly designed to handle modality heterogeneity.

To address this challenge, we propose FedMoSA (Federated Learning via Modality Specialized Adapters), a federated framework designed specifically for

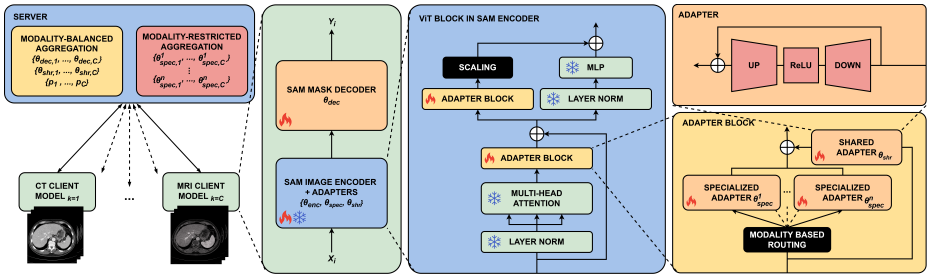


Fig. 1: The proposed FedMoSA with n specialized adapters (one per modality) and a single shared adapter per adapter block. During federated training, only the parameters of the adapter $\theta^m_{spec}, \theta_{shr}$ and the mask decoder θ_{dec} are communicated and aggregated at the server.

modality-heterogeneous medical image segmentation. FedMoSA extends MSA by integrating specialized adapters to capture modality-specific representations and shared adapters that capture modality-agnostic representations. We also introduce a prototype-based [12] alignment mechanism to enforce consistent shared adapter representations across clients. We further propose modality-restricted and modality-balanced aggregation strategies for specialized adapters and shared components, respectively. Together, these design choices enable robust training while mitigating cross-modality interference. Our contributions are:

- A federated framework for robust medical image segmentation under modality heterogeneity.
- Modality-specialized adapters to mitigate cross-modality interference and shared adapters to enable cross-modality interaction.
- A prototype-based alignment mechanism to promote modality-agnostic shared representations across heterogeneous clients.
- Modality-restricted and modality-balanced aggregation strategies to mitigate negative transfer and improve generalization.

2 Method

We consider C clients, each holding a private dataset $\mathcal{D}_k = \{(x_i, y_i)\}$ of images and segmentation masks. Clients may hold one or multiple modalities (e.g., MRI and CT). The most challenging case is the fully modality-disjoint setting where each client has a single modality. The objective is to learn a global segmentation model that performs well across all clients.

2.1 Model Architecture

Our framework extends MSA [14], which adapts SAM [4] via parameter-efficient adapters. We adopt SAM ViT-B, freezing the image encoder and inserting adapter

modules into the self-attention and MLP branches of each transformer block; the prompt encoder is fixed, and the mask decoder is jointly trained. To address modality heterogeneity, each adapter block contains a modality-specialized adapter and a shared adapter. Input tokens are simultaneously routed to the modality-specific adapter θ^m_{spec} and the shared adapter θ_{shr}, capturing modality-specific and modality-agnostic representations, respectively. All adapters follow a bottleneck structure (down-projection, non-linearity, up-projection).

2.2 Client Training

At communication round t, each client receives the aggregated model parameters $\{\theta_{shr}, \theta^m_{spec}, \theta_{dec}\}$ and global prototype $\mathbf{p}_{global}$. Local training minimizes a composite objective. The first component of this objective is the segmentation loss

$$\mathcal{L}_{seg}(x_i, y_i) = \lambda_{bce}\mathcal{L}_{bce}(x_i, y_i) + \lambda_{dice}\mathcal{L}_{dice}(x_i, y_i), \tag{1}$$

where λ_{bce} and λ_{dice} are the weighting coefficients. To encourage consistent shared adapter representations across heterogeneous clients, we define a prototype alignment loss

$$\mathcal{L}_{proto}(x_i) = \|f_{shr}(x_i) - \mathbf{p}_{global}\|_2^2, \tag{2}$$

where $f_{shr}(x_i) \in \mathbb{R}^d$ denotes the shared adapter representation of input sample x_i with $d = 768$, and $\mathbf{p}_{global}$ is the global prototype aggregated from local shared adapter representations. This loss aligns shared features across clients, promoting modality-agnostic representation learning. The overall objective for client k is

$$\mathcal{L}_k = \frac{1}{|\mathcal{D}_k|} \sum_{(x_i, y_i) \in \mathcal{D}_k} \left(\mathcal{L}_{seg}(x_i, y_i) + \lambda_{proto}\mathcal{L}_{proto}(x_i)\right), \tag{3}$$

where λ_{proto} is the weighting coefficient. The prototype loss is disabled at $t = 0$ as the global prototype $\mathbf{p}_{global}$ is not informative before the first round of aggregation. Each client maintains a local shared prototype $\mathbf{p}_k = \mathbf{E}_{x \sim \mathcal{D}_k}[f_{shr}(x)]$, where $\mathbf{E}[\cdot]$ denotes expectation approximated via a running mean accumulated over mini-batches throughout each training round. $\mathbf{p}_k$ is transmitted to the server along with the model parameters at the end of each communication round.

2.3 Server Aggregation

Modality-restricted aggregation. Specialized adapters encode modality-specific information and are aggregated (only across clients sharing modality m) as

$$\theta^m_{spec} \leftarrow \sum_{k=1}^{C} w^m_k \, \theta^m_{spec,k}, \tag{4}$$

where $\theta^m_{spec,k}$ denotes the specialized adapter parameters at client k and C is the total number of participating clients. The weight is $w^m_k \frac{|\mathcal{D}^m_k|}{\sum_{j=1}^{C} |\mathcal{D}^m_j|}$, where

$\mathcal{D}_k^m \subseteq \mathcal{D}_k$ denotes samples of modality m at client k. Clients without modality m contribute $|\mathcal{D}_k^m| = 0$ and are thus excluded from the aggregation.

Modality-balanced aggregation. The shared adapter, mask decoder, and local prototypes are aggregated across all clients using weight

$$\alpha_k = \frac{|\mathcal{M}_k|\,|\mathcal{D}_k|}{\sum_{j=1}^{C} |\mathcal{M}_j|\,|\mathcal{D}_j|}, \tag{5}$$

where $\mathcal{M}_k$ denotes the set of imaging modalities available at client k, $|\mathcal{M}_k|$ is the number of modalities, and $|\mathcal{D}_k|$ is the size of the local dataset at client k. The shared components are aggregated as

$$\theta_{shr} = \sum_{k=1}^{C} \alpha_k \theta_{shr,k}, \quad \theta_{dec} = \sum_{k=1}^{C} \alpha_k \theta_{dec,k}, \quad \mathbf{p}_{global} = \sum_{k=1}^{C} \alpha_k \mathbf{p}_k, \tag{6}$$

where $\theta_{shr,k}$, $\theta_{dec,k}$, and $\mathbf{p}_k$ denote the shared adapter parameters, decoder parameters, and local prototype at client k, respectively.

2.4 Federated Optimization Procedure

At each round, clients update specialized adapters, shared adapters, and the mask decoder using (3), and transmit updated parameters and local prototypes to the server. The server performs modality-restricted aggregation for specialized adapters and modality-balanced aggregation for shared components and prototypes. The server then broadcasts the updated specialized adapters $\{\theta_{spec}^m\}_m$, shared adapters θ_{shr}, decoder θ_{dec}, and the aggregated global prototype $\mathbf{p}_{global}$ to all clients for the next communication round.

3 Experiments and Results

Datasets and Experimental Setup. We construct a multi-institution, modality-heterogeneous liver segmentation benchmark using two MRI datasets (Duke Liver Dataset [7], CHAOS MRI [2]) and four CT datasets (LiTS17 [1], CHAOS CT [2], 3D-IRCADb-01 [13], and CT-ORG [9]). These datasets provide abdominal MRI and contrast-enhanced CT volumes with expert-annotated liver masks; for datasets with multi-organ annotations, only liver masks are retained.

Data are partitioned across five clients under predefined protocols, with each patient assigned to exactly one client and split patient-level into train/val/test (80:10:10). During each training epoch, 25% of slices per 3D volume are randomly sampled; inference is on individual 2D slices. Performance is measured by IoU and Dice on both local and global test sets. Local metrics reflect performance on each client's own test data, while global metrics reflect performance on a pooled multi-modal test set with equal, non-overlapping samples from every client. We consider three federated configurations: (i) **Protocol 1**: all five clients contain both modalities; (ii) **Protocol 2**: two clients contain both MRI and CT data,

Table 1: Client-wise IoU (%) and Dice (%) on local test data, and performance (mean ± std) on the global multi-modal (CT and MRI) test set under three modality-heterogeneous federated protocols (k: client index).

Method	Local Test Data										Global Test Data	
	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$		Mean ± Std	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Protocol 1 (Modality-Complete)												
nnU-Net-local	85.64	91.84	80.11	87.72	88.44	93.55	82.84	90.19	75.64	84.06	82.76 ± 3.88	89.95 ± 2.56
MSA-local	86.42	92.56	80.71	88.87	86.39	92.62	80.94	89.31	82.36	90.15	83.50 ± 3.10	90.72 ± 1.88
MoSA-local	83.99	91.17	82.89	90.27	86.98	92.96	81.71	89.81	85.11	91.86	83.59 ± 2.75	90.82 ± 1.68
FednnU-Net (AFA)	90.11	94.53	89.21	94.05	92.59	96.10	90.33	94.63	89.01	94.03	**90.77 ± 1.56**	**94.99 ± 0.90**
FednnU-Net (FFE)	88.86	93.63	87.38	92.81	91.32	95.38	88.46	93.49	87.66	93.20	89.08 ± 2.27	93.94 ± 1.41
FedProx (MSA)	85.19	91.88	86.42	92.67	86.68	92.80	88.26	93.67	83.11	90.67	86.73 ± 3.53	92.73 ± 2.12
FedMSA	88.60	93.80	86.25	92.42	89.89	94.63	89.09	94.12	87.98	93.52	88.49 ± 1.82	93.77 ± 1.07
FedMoSA	89.91	94.57	88.69	93.87	91.01	95.25	89.82	94.58	88.80	94.02	89.76 ± 1.39	94.53 ± 0.80
Protocol 2 (Partially Modality-Disjoint)												
nnU-Net-local	88.28	93.59	85.91	92.11	72.38	83.73	86.88	92.85	90.43	94.95	70.12 ± 17.57	76.36 ± 16.24
MSA-local	86.92	92.91	85.13	91.80	82.24	90.14	85.65	92.13	87.07	93.06	69.43 ± 16.24	76.68 ± 14.79
MoSA-local	86.72	92.80	87.31	93.11	84.52	91.53	85.99	92.37	89.40	94.38	72.96 ± 16.61	81.32 ± 14.03
FednnU-Net (AFA)	77.40	86.27	73.51	82.36	28.38	43.51	78.57	87.63	84.95	91.81	66.94 ± 17.85	77.56 ± 15.01
FednnU-Net (FFE)	88.88	93.85	86.66	92.54	80.52	89.13	91.35	95.45	91.16	95.35	88.56 ± 2.21	93.74 ± 1.37
FedProx (MSA)	88.97	94.08	86.98	92.94	78.54	87.85	89.48	94.42	90.42	94.95	87.55 ± 0.91	93.29 ± 0.55
FedMSA	88.85	94.04	86.87	92.78	84.28	91.40	89.43	94.38	90.05	94.75	88.37 ± 1.27	93.75 ± 0.74
FedMoSA	89.74	94.55	88.48	93.81	87.27	93.15	91.23	95.40	90.59	95.05	**89.74 ± 0.89**	**94.55 ± 0.52**
Protocol 3 (Fully Modality-Disjoint)												
nnU-Net-local	78.07	87.50	78.13	87.58	70.03	81.83	88.97	94.10	88.39	93.73	48.49 ± 15.78	58.09 ± 12.60
MSA-local	82.21	90.15	83.52	90.96	83.09	90.67	85.41	92.01	82.94	90.39	51.27 ± 10.82	62.24 ± 7.18
MoSA-local	82.31	90.21	85.39	92.06	82.23	90.11	82.84	90.39	79.81	88.28	48.63 ± 17.85	59.71 ± 16.87
FednnU-Net (AFA)	9.19	15.94	28.10	43.24	34.32	50.43	69.22	81.36	70.43	82.20	49.15 ± 14.04	61.07 ± 10.98
FednnU-Net (FFE)	46.69	62.60	42.73	59.01	37.56	53.53	85.30	91.89	86.41	92.53	75.00 ± 8.30	83.60 ± 6.62
FedProx (MSA)	32.46	48.25	35.58	51.59	41.22	57.80	87.00	93.00	86.74	92.76	69.23 ± 8.71	79.49 ± 7.09
FedMSA	46.34	62.80	45.19	61.76	40.45	56.74	81.32	89.52	79.98	88.43	72.24 ± 7.50	83.01 ± 5.57
FedMoSA	82.16	90.12	84.63	91.59	83.62	90.98	86.14	92.41	85.29	91.77	**84.85 ± 3.90**	**91.54 ± 2.48**

while the remaining three are modality-specific (one MRI-only and two CT-only); and (iii) **Protocol 3**: clients are fully modality-disjoint, comprising three MRI-only and two CT-only clients.

Implementation Details. We compare FedMoSA with FedMSA, two FednnU-Net variants (FFE and AFA), and local baselines (MoSA-local, MSA-local, nnU-Net-local). MoSA-local trains the FedMoSA architecture per-client without federation. FedMSA updates and aggregates only adapter and decoder parameters. FedProx (MSA) adds a proximal term [5] to FedMSA's local objective, which constrains each client's update toward the global model. Since FedProx addresses client heterogeneity, we include this baseline to examine whether a simpler approach can handle modality heterogeneity. FFE derives a unified configuration from aggregated client fingerprints; AFA aggregates only structurally identical components. All SAM-based adapters use a bottleneck design with reduction factor 16 (dimension 48). FedMoSA is trained using the composite loss in (3) with $\lambda_{bce} = \lambda_{dice} = 0.5$ and $\lambda_{proto} = 0.01$. All methods are trained for 25 communication rounds, with each client performing two training epochs per round under identical protocols and data splits to ensure fairness. We use SGD [10] optimizer for FednnU-Net, while Adam [3] is used for FedMSA and FedMoSA. Learning rates are selected from $\{10^{-2}, 10^{-3}, 10^{-4}\}$ based on validation performance,

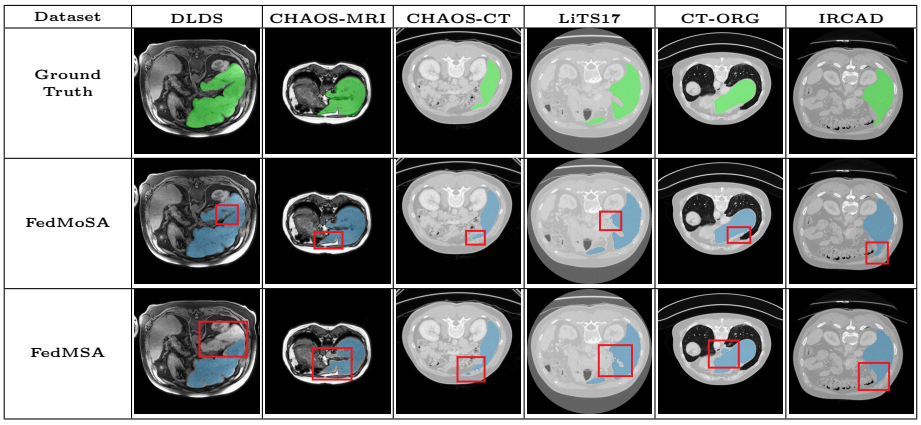
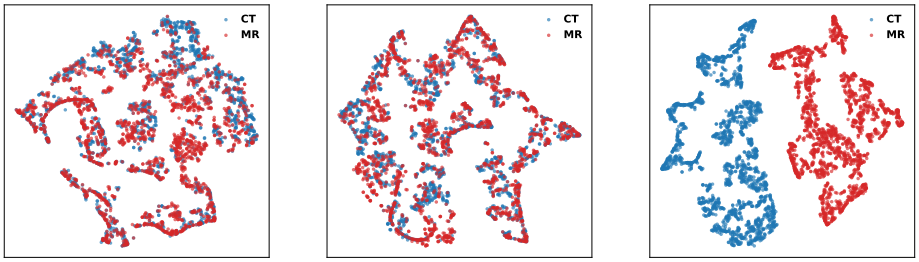


Fig. 2: Segmentation results for FedMoSA and FedMSA on the global multi-modal (CT and MRI) test set under Protocol 3 (Row 1: Ground truth (GT), Row 2: Predictions by the proposed FedMoSA, Row 3: Predictions by FedMSA, Red boxes: representative regions with segmentation inaccuracies.

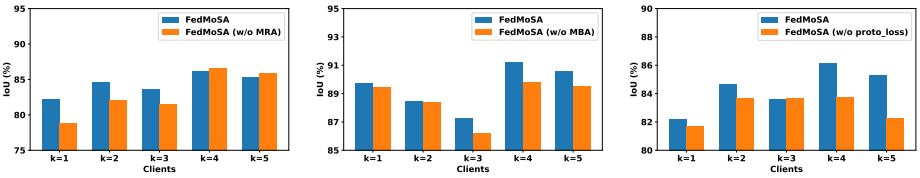
with the final learning rate set to 10^{-4} for SAM-based methods (FedMoSA and FedMSA) and 10^{-2} for FednnU-Net, using a batch size of 8. All experiments are conducted with a fixed random seed of 32.

Comparative and Representation Analysis. In **Protocol 3** (fully modality-disjoint), FedMSA suffers notable performance degradation (72.24% IoU, 83.01% Dice), while FedMoSA maintains stable performance (84.85% IoU, 91.54% Dice), demonstrating strong robustness under severe modality heterogeneity (see Table 1). FednnU-Net (AFA) collapses in this setting, as its fingerprint-driven client-specific model initialization introduces additional architectural heterogeneity across clients. This, combined with fully modality-disjoint data, severely limits effective federated aggregation. Notably, compared to local training (59.71% Dice), FedMoSA achieves a substantial improvement of +31.83 Dice on the global test data. Under **Protocol 2**, FedMoSA attains the best global performance (89.74% IoU, 94.55% Dice), with consistent gains across both multi-modal and modality-specific clients. Finally, under **Protocol 1**, where heterogeneity is minimal, FednnU-Net (AFA) achieves the highest global performance (90.77% IoU, 94.99% Dice), while FedMoSA remains competitive (89.76% IoU, 94.53% Dice) and consistently improves over FedMSA. Qualitative results in Fig. 2 further show that FedMoSA produces anatomically accurate segmentations across both MRI and CT datasets. Notably, FedProx (MSA) fails to consistently improve performance across all clients under **Protocol 2** and **Protocol 3**, and degrades performance in the modality-complete setting (**Protocol 1**). This shows that proximal regularization alone cannot resolve modality-induced client drift. Paired Wilcoxon signed-rank test on per-sample Dice scores computed on the global test set confirms that FedMoSA significantly outperforms the strongest baseline FedMSA under Protocol 3 ($p < 10^{-6}$).



(a) MSA (b) Shared Adapter (c) Specialized Adapters

Fig. 3: t-SNE of attention adapter representations (first ViT block, SAM encoder): (a) MSA shows moderate cross-modality separation; (b) shared adapter shows substantial overlap, indicating modality-agnostic features; (c) specialized adapters form well-separated clusters, reflecting modality-specific features.



(a) MRA (b) MBA (c) Prototype Alignment

Fig. 4: Client-wise IoU in FedMoSA without (a) modality-restricted aggregation (MRA) under Protocol 3, (b) modality-balanced aggregation (MBA) under Protocol 2. (c) prototype regularization under Protocol 3.

t-SNE projections of attention adapter outputs (Protocol 2, Fig. 3) show specialized adapters forming well-separated MRI/CT clusters while shared adapters exhibit substantial cross-modality overlap. This is confirmed quantitatively: specialized adapters achieve high inter-modality distance (3.60) with low intra-modality variance (0.0027), versus MSA's moderate separation (0.1967, variance 0.0010), indicating limited modality-aware adaptation. Shared adapters also show negligible inter-client variance (9.8×10^{-7}) and low intra-client variance (6.2×10^{-4}), compared to MSA's higher inter-client (1.5×10^{-5}) and intra-client variance (9.2×10^{-4}), confirming that shared adapters learn compact, client-invariant representations. Together, these results validate the decoupling of modality-specific and modality-agnostic adaptation.

3.1 Ablation Studies.

Impact of Modality-Restricted Aggregation. Under **Protocol 3**, replacing modality-restricted aggregation with standard FedAvg degrades performance, particularly for MRI-only clients ($k \in \{1, 2, 3\}$) (Fig. 4a), confirming that modality-consistent aggregation mitigates cross-modality interference.

Table 2: Communication cost scaling of FedMoSA using SAM ViT-B for full model size and per-round communication cost (including mask decoder, shared adapters, and modality-specific specialized adapters).

# Modalities	Full Model	Mask Decoder	Shared Adapters	Specialized Adapters	Total / Round
2	378.05 MB	15.48 MB	6.82 MB	13.64 MB	35.96 MB
3	384.87 MB	15.48 MB	6.82 MB	20.46 MB	42.78 MB
4	391.70 MB	15.48 MB	6.82 MB	27.28 MB	49.60 MB

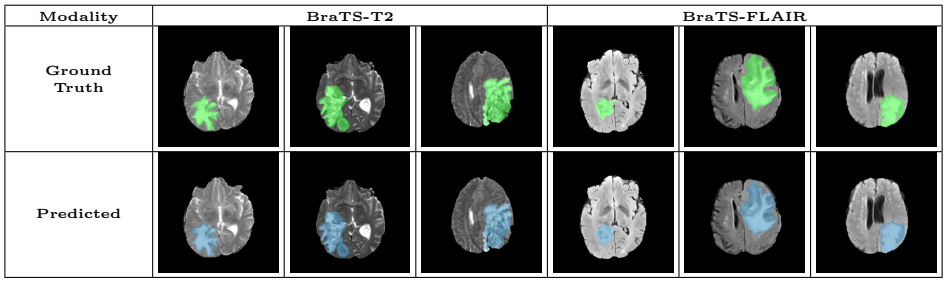


Fig. 5: Out-of-distribution segmentation results of FedMoSA on sample images from the BraTS dataset (row 2).

Significance of Modality-Balanced Aggregation. In **Protocol 2**, replacing modality-balanced aggregation with standard FedAvg for shared components consistently degrades performance (Fig. 4b), confirming that prioritizing multimodal clients promotes more generalizable shared representations. Under **Protocol 3**, modality-balanced aggregation naturally reduces to standard FedAvg.

Significance of Prototype Alignment. We remove the prototype regularization term under **Protocol 3**. As shown in Fig. 4c, performance consistently degrades, demonstrating that aligning shared adapter representations across clients enhances generalization under strong heterogeneity.

Communication Cost Scaling with the Number of Modalities. Table 2 reports the full model size and per-round communication cost of FedMoSA using SAM ViT-B. The communication cost scales linearly with the number of modalities due to specialized adapters, while the mask decoder and shared adapters remain constant. With four modalities, FedMoSA communicates 49.6 MB per round versus a 391.7 MB model. This results in 87.30% reduction in per-round communication compared to the full model FL.

Out-of-Distribution Generalization. We fine-tune FedMoSA (trained for liver segmentation) on the BraTS [8] dataset (T2 and FLAIR Brain MRI modalities). We perform five federated fine-tuning rounds across four clients under fully modality-disjoint setting. FedMoSA achieves an average IoU of 80.08 ± 4.60 and Dice of 88.76 ± 2.97 under this minimal adaptation setting (see Fig. 5).

4 Conclusion

We presented FedMoSA, a modality-heterogeneous federated learning framework for medical image segmentation under modality-heterogeneous settings. By integrating specialized and shared adapters within the SAM architecture and introducing modality-aware aggregation strategies, FedMoSA mitigates cross-modality interference while preserving cross-modality interaction. Extensive experiments across diverse federated protocols demonstrate that FedMoSA remains robust under strong modality heterogeneity, particularly in fully modality-disjoint setting where conventional aggregation strategies degrade. Ablation studies further confirm the importance of prototype alignment, modality-restricted, and modality-balanced aggregation mechanisms. Future work will extend FedMoSA to a broader range of anatomical structures and explore dynamic routing strategies to remove explicit modality-based routing and improve adaptability to unseen or partially observed modalities.

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

References

1. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., Lohöfer, F., Holch, J.W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M.M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J.C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E.B., Dou, Q., Fu, C.W., Georgescu, B., i Nieto, X.G., Gruen, F., Han, X., Heng, P.A., Hesser, J., Moltz, J.H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K.C., Khened, M., Kim, I., Kim, J.H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S.C.H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., van Ginneken, B., Greenspan, H., Joskowicz, L., Menze, B.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023). https://doi.org/https://doi.org/10.1016/j.media.2022.102680, `https://www.sciencedirect.com/science/article/pii/S1361841522003085`
2. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: Chaos challenge - combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021). https://doi.org/https://doi.org/10.1016/j.media.2020.101950, `https://www.sciencedirect.com/science/article/pii/S1361841520303145`

3. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017), https://arxiv.org/abs/1412.6980
4. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3992–4003 (2023). https://doi.org/10.1109/ICCV51070.2023.00371
5. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. In: Dhillon, I., Papailiopoulos, D., Sze, V. (eds.) Proceedings of Machine Learning and Systems. vol. 2, pp. 429–450 (2020)
6. Liu, Y., Luo, G., Zhu, Y.: Fedfms: Exploring federated foundation models for medical image segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 283–293. Springer Nature Switzerland, Cham (2024)
7. Macdonald, J.A., Zhu, Z., Konkel, B., Mazurowski, M.A., Wiggins, W.F., Bashir, M.R.: Duke liver dataset: A publicly available liver mri dataset with liver segmentation masks and series labels. Radiology: Artificial Intelligence **5**(5), e220275 (2023). https://doi.org/10.1148/ryai.220275
8. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2015)
9. Rister, B., Shivakumar, K., Nobashi, T., Rubin, D.L.: Ct-org: A dataset of ct volumes with multiple organ segmentations. The Cancer Imaging Archive (2019). https://doi.org/10.7937/TCIA.2019.TT7F4V7O, version 1
10. Robbins, H., Monro, S.: A stochastic approximation method. The Annals of Mathematical Statistics **22**(3), 400–407 (1951). https://doi.org/10.1214/aoms/1177729586, https://www.jstor.org/stable/2236626
11. Skorupko, G., Avgoustidis, F., Martín-Isla, C., Garrucho, L., Kessler, D.A., Pujadas, E.R., Díaz, O., Bobowicz, M., Gwoździewicz, K., Bargalló, X., Jarušević̆ius, P., Osuala, R., Kushibar, K., Lekadir, K.: Federated nnu-net for privacy-preserving medical image segmentation. Scientific Reports **15**(1), 38312 (2025). https://doi.org/10.1038/s41598-025-22239-0, https://doi.org/10.1038/s41598-025-22239-0
12. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
13. Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J.B., Moreau, J., Osswald, A.B., Bouhadjar, M., Marescaux, J.: 3d image reconstruction for comparison of algorithm database (3d-ircadb-01). Tech. rep., IRCAD, Strasbourg, France (2010), https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/, dataset reference for 3D-IRCADb-01 liver segmentation
14. Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. Medical Image Analysis **102**, 103547 (2025). https://doi.org/https://doi.org/10.1016/j.media.2025.103547, https://www.sciencedirect.com/science/article/pii/S1361841525000945