Federated Client-Tailored Adapter for Medical Image Segmentation

Guyue Hu, *Member, IEEE*, Siyuan Song, Yukun Kang, Zhu Yin, Gangming Zhao, Chenglong Li, *Member, IEEE*, and Jin Tang

***Abstract*—Medical image segmentation in X-ray images is beneficial for computer-aided diagnosis and lesion localization. Existing methods mainly fall into a centralized learning paradigm, which is inapplicable in the practical medical scenario that only has access to distributed data islands. Federated Learning has the potential to offer a distributed solution but struggles with heavy training instability due to client-wise domain heterogeneity (including distribution diversity and class imbalance). In this paper, we propose a novel Federated Client-tailored Adapter (FCA) framework for medical image segmentation, which achieves stable and client-tailored adaptive segmentation without sharing sensitive local data. Specifically, the federated adapter stirs universal knowledge in off-the-shelf medical foundation models to stabilize the federated training process. In addition, we develop two client-tailored federated updating strategies that adaptively decompose the adapter into common and individual components, then globally and independently update the parameter groups associated with common client-invariant and individual client-specific units, respectively. They further stabilize the heterogeneous federated learning process and realize optimal client-tailored instead of sub-optimal global-compromised segmentation models. Extensive experiments on three large-scale datasets demonstrate the effectiveness and superiority of the proposed FCA framework for federated medical segmentation.**

***Index Terms*—Federated learning, parameter-efficient fine-tuning, medical image segmentation.**

I. Introduction

Medical images segmentation plays a critical role in various medical applications, such as epicardial fat segmentation [1], brain tumor segmentation [2], facilitating more accurate diagnoses and reducing the burden on healthcare professionals. Recent progress in deep learning and large foundation models have significantly advanced the field of medical image segmentation. The pioneer U-Net [3] utilizing an encoder-decoder architecture is one of the most famous approaches in medical image segmentation. Subsequently, its variants based on various architectures have been designed to handle medical image segmentation tasks, such as CNN-based variants [3], [4], [5] and transformer-based variants [6], [7]. In recent years, numerous large medical foundation models [8], [9], [10] empowered with powerful capabilities of cross-domain knowledge understanding, logical reasoning, and language generation, have also significantly facilitated the field of medical image segmentation. With the aid of these approaches, medical image segmentation significantly improves the accuracy of computer-aid diagnoses and effectively streamlines clinical workflows.

Despite such huge success, existing medical image segmentation methods mainly fall into a centralized learning paradigm, where medical image data from different sources (clients) are fully delivered to a central server to collectively learn a single optimal segmentation model, as shown in Fig. 1 (a). In practical medical scenarios, we usually only have limited access to distributed “data islands” where sharing local medical data among different clients (*e.g.* hospitals) is forbidden [11] and only insensitive model weights are allowed to be shared since various factors such as strict privacy regulations in hospitals, limited network bandwidth, etc. Thus, the existing centralized approaches are no longer suitable for the distributed medical scenario.

Federated Learning (FL) is one typical decentralized training technique, which collectively learns a global model in a central server from multiple distributed clients without

Received 29 October 2024; revised 17 February 2025, 15 April 2025, and 29 May 2025; accepted 16 June 2025. Date of publication 19 June 2025; date of current version 1 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62376004, in part by the Anhui Provincial Natural Science Foundation under Grant 2408085QF201, in part by the Natural Science Foundation of Anhui Higher Education Institution under Grant 2022AH040014, in part by the Open Project of Anhui Provincial Key Laboratory of Security Artificial Intelligence under Grant SAI2024003, and in part by the Open Project of Anhui Provincial Key Laboratory of Intelligent Detection and Diagnosis for Traffic Infrastructure under Grant KY-2025-03. The associate editor coordinating the review of this article and approving it for publication was Prof. Edgar Weippl. *(Corresponding author: Chenglong Li.)*

Guyue Hu is with the State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui Provincial Key Laboratory of Security Artificial Intelligence, and the School of Artificial Intelligence, Anhui University, Hefei 230601, China, and also with Anhui Provincial Key Laboratory of Intelligent Detection and Diagnosis for Traffic Infrastructure, Anhui Jiaojian Traffic Development and Research Center Company Ltd., Hefei 230051, China (e-mail: guyue.hu@ahu.edu.cn).

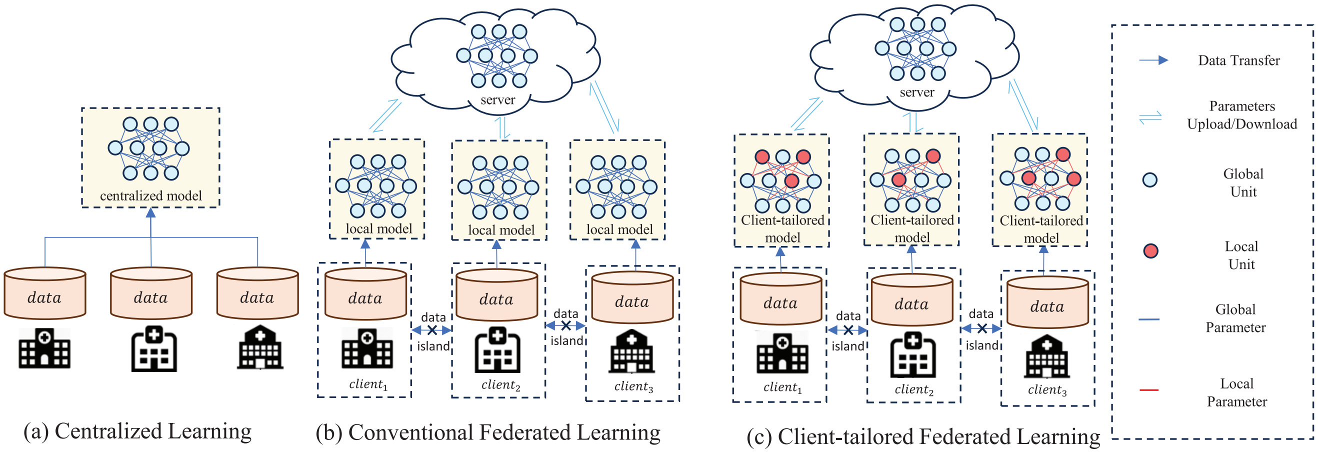
Siyuan Song and Chenglong Li are with the State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui Provincial Key Laboratory of Security Artificial Intelligence, and the School of Artificial Intelligence, Anhui University, Hefei 230601, China (e-mail: ssy136@stu.ahu.edu.cn; lcl1314@foxmail.com).

Yukun Kang and Jin Tang are with the State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, and the School of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: kyk30@stu.ahu.edu.cn; tangjin@ahu.edu.cn).

Zhu Yin is with the School of Internet, Anhui University, Hefei 230601, China (e-mail: yinzhu@ahu.edu.cn).

Gangming Zhao is with the Department of Computer Science, The University of Hong Kong, Hong Kong (e-mail: gmzhao@connect.hku.hk).

Digital Object Identifier 10.1109/TIFS.2025.3581432

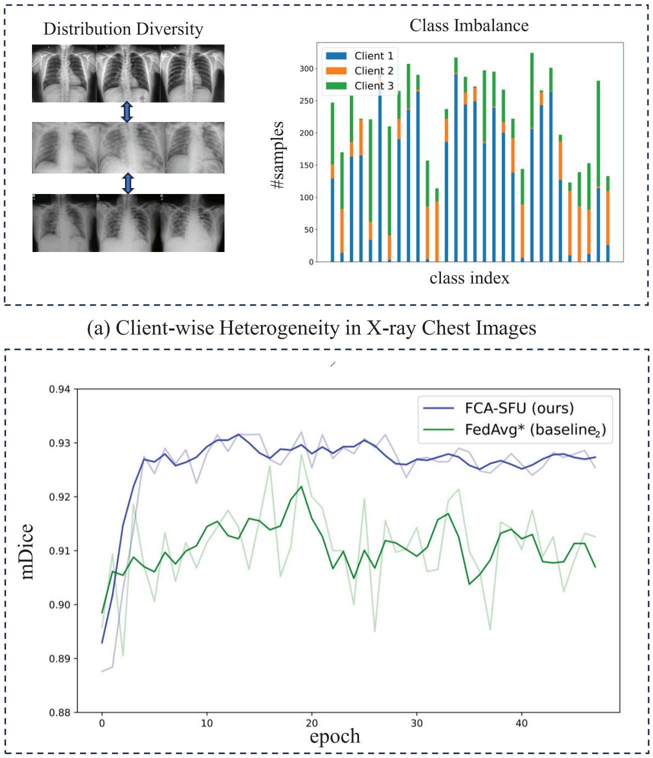


(a) Centralized Learning (b) Conventional Federated Learning (c) Client-tailored Federated Learning

Fig. 1. Conventional learning paradigms for tackling heterogeneous distributed medical data. (a) Centralized learning aggregates all data together to train a single model. (b) Conventional federated learning trains a global-compromised model for all clients without sharing sensitive local data. (c) The proposed client-tailored federated learning trains client-customized models for each client without sharing sensitive local data.

sharing the distributed local data [12], [13], [14]. It was initially introduced in [15], aiming at leveraging distributed training methodologies to accommodate data from various users with disparate data scales. Since then, federated learning has achieved rapid advancements and has been applied to various fields in computer vision, such as image classification [16] and image segmentation [17], [18]. When embracing the federated learning principle, medical image segmentation has the potential to capitalize on distributed data resources while upholding privacy regulation. As shown in Fig. 1 (b), conventional FL first uploads local model parameters in every edge client to the global model in the central server, and then each client downloads aggregating weights from the server, eventually obtaining a global-compromised model for all clients.

However, distributed medical "data island" typically exhibits heavy client-wise heterogeneity including class imbalance and distribution diversity (see Fig. 2 (a)). Class imbalance in medical data is common for abundant reasons such as different hospitals specializing in different diseases, different anatomical regions having different probabilities being examined in medical imaging equipment, etc. The distribution diversity is usually induced by diverse data collection conditions (such as equipment, personnel, and environmental factors). Our initial experiments indicate that directly applying conventional FL methods to medical image segmentation suffers from considerable instability and slow convergence (see Fig. 2 (b)) since these client-wise heterogeneities. Besides, the obtained global-compromised model from conventional FL also deviates too far from the individual optimality of each client due to such heterogeneity, thus largely suppressing the advantage of utilizing distributed data for learning facilitation. Although there are a few pioneer works attempt to customize models for different clients, they usually decouple at coarse-grained levels [14], [19] and allocate parameter groups [20], [21] statically and binarily, lacking enough adaptability in tackling heterogeneous distributed medical image segmentation.



(a) Client-wise Heterogeneity in X-ray Chest Images

(b) Stability Comparison

Fig. 2. (a) Client-wise heterogeneity in X-ray chest images consists of common class imbalance (long-tail distribution) and various distribution diversity. (b) In heterogeneous distributed scenarios, conventional federated learning (*e.g.* FedAvg* [15]) suffers from considerable instability and slow convergence while our federated client-tailored adapter (FCA-SFU) effectively alleviates this issue. The transparent lines represent the original experimental results, while the solid lines represent smoothed results that facilitate visualization.

To move beyond such limitations, we propose a novel Federated Client-tailored Adapter (FCA) framework to achieve stable distributed medical segmentation without sharing sensitive local data. Specifically, we first construct parameter-efficient federated adapters to distill the client-invariant universal knowledge in off-the-shelf large medical foundation models to stabilize heterogeneous distributed medical image segmentation. In addition, we dynamically decompose the fine-grained adapter parameters into common and individual units through binary or probabilistic decomposition, as shown in Fig. 1 (c). The client-invariant components undergo a global federated updating while the client-specific individual components are updated client-independently. The decomposed federated updating strategy achieves two advantages: further stabilizing the heterogeneous federated learning process and realizing optimal client-tailored segmentation model for each client rather than sub-optimal global-compromised segmentation model for all clients.

In summary, the main contributions of this paper could be summarized as follows:

- We identify the training instability issue in conventional federated learning induced by client-wise heterogeneity (including class imbalance and distribution diversity) in medical image segmentation and alleviate it by seeking optimal client-tailored models rather than a sub-optimal global-compromised model.
- We propose a Federated Client-tailored Adapter for medical image segmentation, achieving stable and customized federated segmentation without sharing sensitive local data. It tightly couples parameter-efficient adapters with FL, thus effectively distilling the universal (common) knowledge in off-the-shelf medical foundation models to stabilize heterogeneous federated learning.
- The innovative Client-tailored Federated Updating strategies adaptively decompose the adapter units into common and individual components, empowering a novel client-tailored and parameter-efficient updating and further stabilizing the heterogeneous federated training process.
- The proposed FCA achieves state-of-the-art performance on three large-scale datasets, demonstrating its effectiveness and superiority in tackling heterogeneous distributed medical image segmentation.

II. Related Works

A. Medical Image Segmentation

Medical image segmentation is a crucial technique in healthcare that involves assigning each pixel in medical images with the corresponding class label. Existing approaches for medical image segmentation primarily fall into three paradigms: CNN-based, transformer-based, and hybridized methods. Specifically, the CNN-based methods, represented by the well-known U-Net and its variants (such as U-Net++ and nnUNet), [3], [4], [5] typically employ a U-shaped encoder-decoder architecture with skip connections to preserve detailed anatomical information. These methods have demonstrated excellent performance on small-scale datasets, significantly advancing the field of medical image segmentation. Transformer-based segmentation approaches, represented by TransUnet [22], leverage the robust long-range information acquisition capabilities of Vision Transformers (ViT) [23], further improving representation capability and driving rapid advancements in medical image segmentation. Besides, VM-UNet [24] integrates Vision-Mamba with the classical U-Net, further introducing long-distance dependencies while maintaining linear computational complexity simultaneously.

In past a few years, supervised pre-training and fine-tuning paradigm was the mainstream methods for various computer vision tasks [25], [26], [27], [28]. Beyond traditional supervised learning paradigm, segmentation approaches that utilize foundation models and Parameter-Efficient Fine-Tuning (PEFT) techniques have also significantly reshaped and facilitated the image segmentation field. For instance, the pioneer Segment Anything Model (SAM) [10] achieved a notable breakthrough in image segmentation by introducing a novel prompt-driven approach. The MedSAM [29] achieves high-precision segmentation across various data modalities and segmentation targets. The nnSAM [30] combines the powerful capability of representation learning from SAM with the adaptive configuration ability from classical nnUNet [4], realizing effective dataset-specific representation learning for medical image segmentation. Despite such huge success, existing segmentation methods mainly fall into a centralized learning paradigm [31], [32] and do not perform well in distributed medical scenarios.

B. Federated Learning

Benefiting from the promising ability to leverage distributed data while preserving privacy, federated learning has attracted considerable attention in recent years. FedAvg [15], one of the pioneering works in Federated Learning, offers the most fundamental framework to this paradigm. It employs a simple weighted averaging strategy to update a global model in the central server. However, recent studies have highlighted the phenomenon of client drift induced by client-wise heterogeneity [18], [33], resulting in inconsistency issues regarding optimal models for each client. Therefore, some improvements have been developed to alleviate the client-wise non-iid challenge. The FedProx [34], FedDC [33], FedSM [35], and FedSeg [18] reform the federated aggregation strategy or loss functions to seek more matchable global or local models. FedNH [36] and Fed-CBS [37] alleviate client-wise class imbalance by a client sampling strategy to generate grouped class-balanced datasets and utilizing the uniformity and semantics of class prototypes, receptively. Scaffold [38], and FedNP [39] alleviate non-iid data issues by deliberately handling distribution diversity in distributed datasets. FedST [40] and FedA3I [41] enhance the contribution of high-quality local models during aggregation, aiming to achieve superior aggregated models. FSAR [42] divides the topology connection in the graph neural network into global and local parts, realizing adaptive federated action recognition. FedNH [36] improves the client drift phenomenon during aggregation by enhancing the generalization of local models. Recently, there are some pioneer works attempt to introduce large foundation models into classical federated learning [43], [44], [45] but without client-tailored consideration. In our work, we develop a novel client-tailored federated updating strategy that adaptive and fine-grained decomposes global client-invariant and local client-specific units, realizing optimal client-customized models for each client rather than a sub-optimal global-compromised model for all clients.

III. Method

A. Preliminary

1) Conventional Federated Learning: We first introduce a vanilla baseline under the umbrella of conventional federated learning (FL) to demonstrate some preliminary knowledge, illustrated in Fig. 1 (b). Consider a distributed medical scenario with N edge clients and one central server, where each client holds some private data that cannot be shared among clients and the central server. During training, all clients collaboratively update a shared segmentation model in the central server. For each global updating round, every client trains its latest local model received from the server for N_e epochs based on its private data. Then, each local client sends the updated parameters to the central server for aggregating and updating the global segmentation model. The updated global model is subsequently distributed to each local client for parameter replacement and waiting for the next updating round. As a result, conventional FL usually leads to a *global-compromised* segmentation model for *all clients* rather than *client-tailored* segmentation models for *each client*, and each client utilizes the same model during the reference period. Moreover, conventional FL also encounters considerable instability and slow convergence issues induced by client-wise heterogeneity (class imbalance and distribution diversity) in distributed medical data (see Fig. 2 (a) for details). In this paper, we proposed a Federated Client-tailored Adapter (FCA) for medical image segmentation to address these issues.

2) Medical Foundation Models: In the last few years, large medical foundation models (MFMs) have significantly advanced various tasks in the medical image processing field. For example, one of the most famous MFMs is SAM-Med2D [10], which fine-tunes conventional segmentation model SAM [46] via Parameter-Efficient Fine-Tuning (PEFT) techniques of adapter learning and prompt learning [47]. Most MFMs for medical image segmentation (such as MedSAM [29] and Med-SA [48]) usually follow a similar pipeline that includes a transformer-based image encoder, a transformer-based mask decoder, a prompt generator, and a prompt encoder. The encoder layers are implemented with a Vision Transformer (ViT) that extracts image features through multiple stacked transformer layers. A prompt encoder encodes the information of prompt hints generated from text, points, or boxes. Finally, the mask decoder integrates the representations from the image encoder and the prompt encoder to generate corresponding segmentation masks. For simplicity, we take the famous SAM-Med2D [10] as an example to demonstrate our FCA framework in the method section and validate its generalization ability in the experimental section (Section IV). The off-the-shelf MFMs inherently contain abundant general medical knowledge because of their huge model parameters and massive training data. Therefore, it is potential to distill the client-invariant universal knowledge from off-the-shelf MFMs as one effective measure to stabilize the non-centralized medical image segmentation.

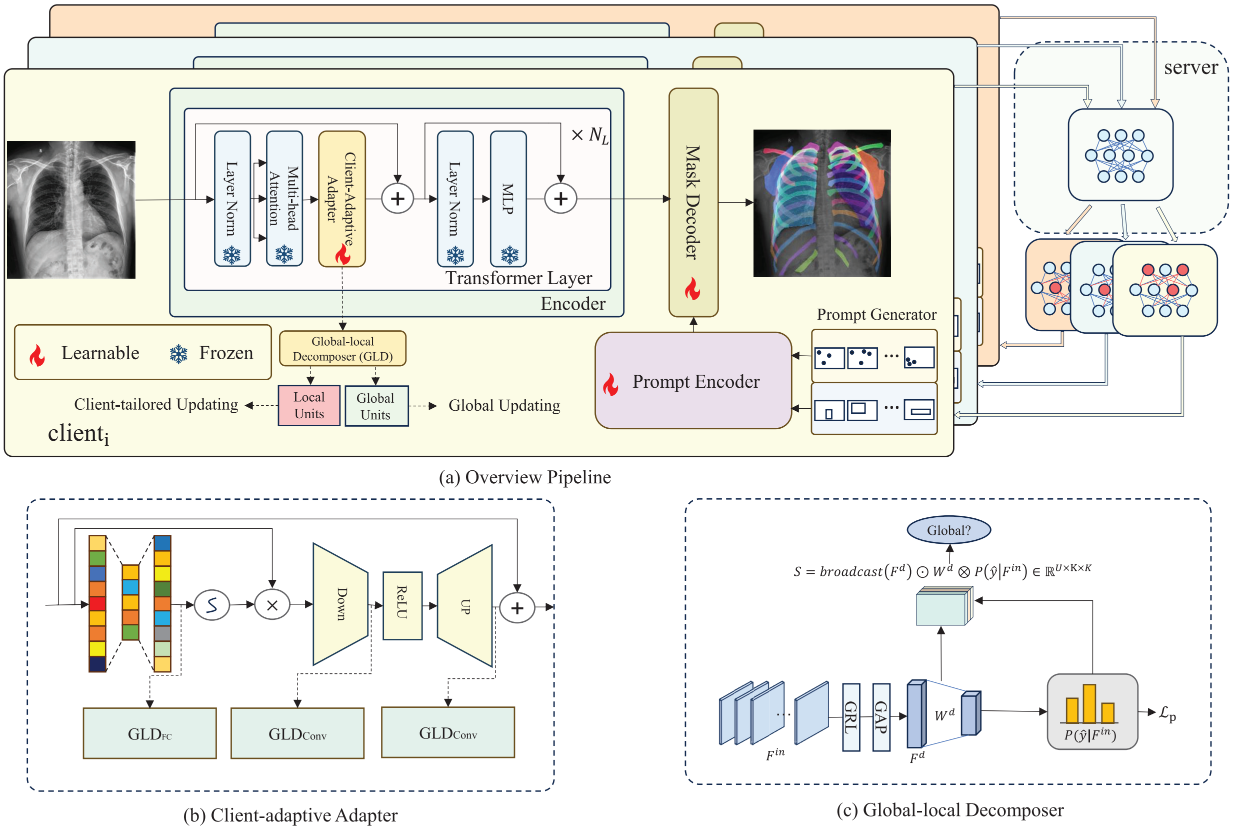
B. Overview

The overview pipeline of our Federated Client-tailored Adapter (FCA) for tackling heterogeneous medical image segmentation is shown in Fig. 3 (a). The basic segmentation model in each local client consists of a large medical foundation model (specified as the SAM-Med2D in Fig. 3 (a)) and inserted adapter layers in each transformer layer of the MFMs encoder. Most layers of the MFMs encoder are frozen, which contain rich prior knowledge (including both client-specific and client-invariant knowledge) inherited from off-the-shelf MFMs. Only the lightweight adapter layers, prompt encoder, and mask decoder with a few learnable parameters will go through parameter-efficient fine-tuning. As a result, each local client can efficiently distill client-specific dark knowledge in the MFMs to facilitate its client-tailored knowledge learning process. They also stir client-invariant general knowledge to benefit and stabilize the sequential client-invariant federated updating process.

Then, the central server and all edge clients fed with the above basic segmentation models conduct federated learning with heterogeneous distributed medical "data islands". If directly applying conventional federated updating strategy (Fig. 1 (b)), it could only obtain a *global-compromised* segmentation model and also encounters considerable instability and slow convergence. Therefore, we develop a Global-local decompose mechanism (GLD) to adaptively decompose each adaptor into client-invariant global units and client-specific local units. Thereafter, the proposed client-tailored federated updating strategies are explored to binary or smoothly to update the adapters in the local clients and central server. As a result, our FCA framework could obtain more optimal *client-tailored* segmentation models for *each client* rather than a sub-optimal *global-compromised* segmentation model for *all clients*. Moreover, our client-tailored federated updating strategies also alleviate training instability and slow convergence issues during heterogeneous federated learning.

C. Adaptive Adaptor Decomposition

As shown in Fig. 3 (b), a typical adapter (borrowed from [10]) for large MFMs usually consists of a series of fully connected (FC) and convolutional (Conv) layers. We append an auxiliary Global-local Decomposer (GLD) branch to each Conv and FC layer to adaptively decompose the adaptor into global and local units, which are denoted as GLD_{Conv} and GLD_{FC} in Fig. 3 (b). The Global-local Decomposer aims to decompose the units (channel for Conv layers or neuron for FC layers) in each adapter layer into common client-invariant and individual client-specific components. The parameters group associated with global and local units will be updated globally and locally thereafter. Taking the GLD_{Conv} as an example, we describe its implementation details in Fig. 3 (c). The GLD_{Conv} comprises a lightweight client discriminator that distinguishes which client (domain) the input representation comes from. Specifically, the side input representation $\boldsymbol{F}^{in}(i) \in \mathbb{R}^{H\times W\times C}$ corresponding sample i first goes through a global average pooling (GAP) layer along the spatial dimensions to obtain a channel-wised presentation $\boldsymbol{F}^{d}(i) \in \mathbb{R}^{C}$, where C is the channel



(a) Overview Pipeline

(b) Client-adaptive Adapter

(c) Global-local Decomposer

Fig. 3. (a) Overview of the proposed federated client-tailored adapter (FCA) framework. (b) Detailed structure of the client-adaptive adapter. (c) The mechanism of the Global-local decomposer (GLD).

number of the feature map. Then, we construct a pretext task called client discrimination to dynamically distinguish the representation source and determine the individual unit contribution during each global communication round. The client discriminator takes in representation from different clients and will be trained to distinguish the representation source. Given an input representation $\boldsymbol{F}^{in}(i)$ of sample i from one specific client, the client discriminator yields a classification probability $P(\hat{\boldsymbol{y}}_k(i)|\boldsymbol{F}^{in}(i))$ voting it sources from the k-th client. The train loss for this pretext task is defined as follows,

$$\mathcal{L}_p = \frac{1}{N_I}\sum_{i=1}^{N_I}\sum_{j=1}^{K}\mathbb{1}_{[j=k]}\log\left(\boldsymbol{W}^d(GAP(\boldsymbol{F}_k^{in}(i)))\right) \qquad (1)$$

where N_I and K are respectively the numbers of samples and clients, and $\mathbb{1}$ denotes the indicator function. To ensure the auxiliary branch does not affect the original training process of the main branch in Fig. 3 (b), we elaborately insert a Gradient Reversal Layer (GRL) before the client discriminator to truncate the gradients from the above discrimination loss.

Taking one convolutional layer in the MFMs adapter as an example, we treat each channel as a unit and let $U = C$. Intuitively, the channels contributing more to client prediction contain more client-specific information while those contributing less may contain more client-invariant global information. Thus, we first quantify the contribution score $\hat{\boldsymbol{S}}_{u,k}(i)$ of each channel (unit) u utilized for determining an input feature $\boldsymbol{F}^{in}(i)$ corresponding to sample i is sourced from client k via its weighted activation and then weight it by predicted client probability $P(\hat{\boldsymbol{y}}_j^i|\boldsymbol{F}^{in}(i))$. The whole process of computing weighted the contribution score could be formulated as $\boldsymbol{S}_{u,k,j}(i) = \boldsymbol{F}_u^d(i)\cdot\boldsymbol{W}_{u,k}^d\cdot P(\hat{\boldsymbol{y}}_j^i|\boldsymbol{F}^{in}(i))$. For brevity, we rewrite the process in matrix form, *i.e.*

$$\hat{\boldsymbol{S}}(i) = broadcast(\boldsymbol{F}^d(i)) \odot \boldsymbol{W^d} \qquad (2)$$

where $\hat{\boldsymbol{S}}(i) \in \mathbb{R}^{U\times K}$, the function *broadcast* denotes element duplicating for shape matching, $\boldsymbol{W}^d \in \mathbb{R}^{U\times K}$ is the classifier weight of the GLD (its various elements can represent the contribution level of each unit judged as a certain category), and $\odot$ is element-wised multiplication. $\hat{\boldsymbol{S}}(i)$ represents the contribution of each unit to determining whether the entire feature belongs to a certain category. After weighing the previous contribution score of each unit with the predicted domain probability $P(\hat{\boldsymbol{y}}^i|\boldsymbol{F}^{in}(i)) \in \mathbb{R}^K$ corresponding to the input feature, we obtain the weighted contribution score matrix $\boldsymbol{S}(i)$ corresponding sample i,

$$\boldsymbol{S}(i) = \hat{\boldsymbol{S}}(i) \otimes P(\hat{\boldsymbol{y}}^i|\boldsymbol{F}^{in}(i)) \qquad (3)$$

where the weighted contribution score $\boldsymbol{S}(i) \in \mathbb{R}^{U\times K\times K}$, and $\otimes$ represents the outer product. Finally, we conduct a sample-wise average to obtain the final matrix of contribution score, *i.e.* $\boldsymbol{S} = \sum_{i=1}^{N_I}\boldsymbol{S}(i)$, where N_I is the number of samples. To maintain brevity, we will exclude the sample index i for all related matrices from here unless explicitly stated.

As for the implementation of a Global-local Decomposer for FC layers (GLD_{FC}), we could treat it as a special convolution layer with a kernel of 1×1. Then, the GLD_{FC} could be implemented similar as the GLD_{Conv}. Eventually, the final contribution score $\boldsymbol{S}$ will be delivered to the central server and then broadcast to every client for global and local unit discrimination and client-tailored federated updating.

D. Client-Tailored Federated Updating

In this section, we split the adapter units into common client-invariant and individual client-specific components according to the weighted contribution score $\boldsymbol{S}$. Then, we conduct global updating for common client-invariant units and independent updating for individual client-specific units. Specifically, we first establish a model copy in the central server including all components of each local client in Fig. 3 (a) except for the insertion of the client-tailored adapters. Then, we devise two client-tailored federated updating strategies to conduct client-tailored federated learning, including binary federated updating (BFU) and smooth federated updating (SFU). The SFU is conceptually built upon BFU and could be treated as a generalized version of BFU. Our experiments indicate that both strategies are very effective in alleviating training instability and slow convergence issues during heterogeneous federated learning, among which the generalized strategy SFU achieves better performance. The detailed algorithm is shown in Algorithm 1. We will introduce the implementation details below.

1) Binary Federated Updating Strategy: In each round, the Binary Federated Updating Strategy (BFU) binary distinguishes the adapter units in each local model into client-specific local parts and client-invariant global parts (Fig. 4 (a)). Then, the parameter group corresponding to client-invariant global parts will communicate with the central server for joint global updating while the parameter group corresponding to client-specific local parts only be updated locally. The weighted contribution score $\boldsymbol{S}$ contains the contribution of each unit for client discrimination and serves as a good indicator to determine whether the units should be treated locally or globally. Intuitively, the adapter units that contribute more to client prediction likely contain more client-specific information. In contrast, the units that contribute less contain more client-invariant global information. To determine whether the parameter group in a unit should be treated globally or locally, we examine the uniformity of its weighted contribution scores for client discrimination. If the weighted contribution scores are similar across all clients, it suggests that the parameter group in this unit is client-specific and thus can be considered a global parameter. In contrast, if there are significant score differences across clients, it indicates that the parameter group in the unit may be domain-specific and should be treated locally. Therefore, we quantify the client-wise uniformity of the weighted contribution score by computing the normalized entropy of score distribution, *i.e.*

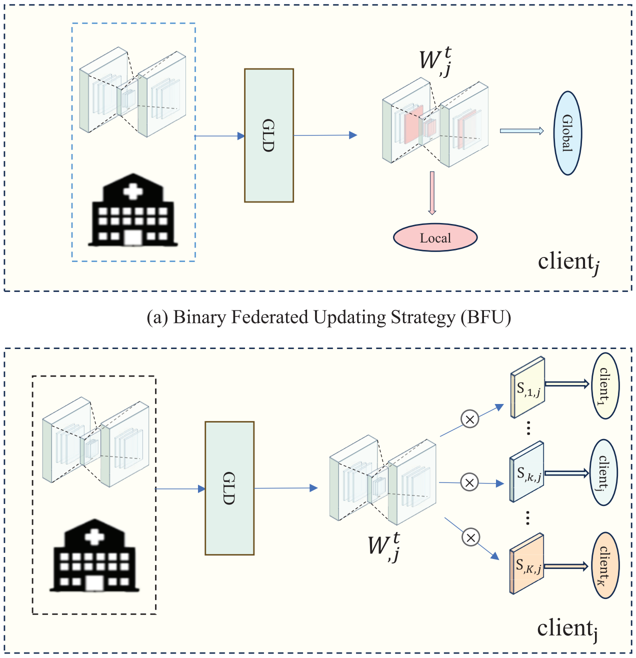
$$\boldsymbol{D}_{u,k} = \frac{Entropy(\boldsymbol{S}_{u,k})}{Entropy(\mathcal{U}(1,K))} = \frac{\sum_{j=1}^{K} \boldsymbol{S}_{u,k,j} \times \log_2 \boldsymbol{S}_{u,k,j}}{\log_2 K} \tag{4}$$

where $\boldsymbol{D}_{u,k}$ represents the diversity degree of score distribution for the unit u of the k-th client, and the $\mathcal{U}$ represents the uniform distribution. The $\log_2 K$ is the theoretical upper bound for the entropy of client-wise score distribution when the distribution is a standard uniform distribution (*i.e.* $Entropy(\mathcal{U}(1,K)) = \log_2 K$). As a result, the larger the diversity degree of a unit, its distribution is closer to a uniform distribution, and its parameter group is more likely to be global. Therefore, we could conveniently distinguish whether each adapter unit is global (client-invariant) or local (client-specific) through binarizing $\boldsymbol{D}_{u,k}$ with a threshold δ (see Fig. 4 (a) for details). If $\boldsymbol{D}_{u,k}$ for a unit u of client k is greater than δ, the unit is treated as a global unit. Otherwise, if $\boldsymbol{D}_{u,k}$ is less than δ, the unit is treated as a local unit. The detailed formula is as follows:

$$M_{u,k} = \begin{cases} 1, & D_{u,k} > \delta \quad (Global\ Unit) \\ 0, & D_{u,k} < \delta \quad (Local\ Unit) \end{cases} \tag{5}$$

Algorithm 1 Client-Tailored Federated Updating (FCA)

```
Input: The number of input samples $N_I$, the number of clients
    $K$, the learnable parameters of encoders in each client $\boldsymbol{W}$,
    the unit number in a adapter layer $U$
Output: Client-tailored adaptor for every client $W^{T+1}(\theta, \hat{\theta})$
    for $t = 1$ to $T$ do
      // Update local models and get contribution score $\boldsymbol{S}$
        for $k = 1$ to $K$ do
            $\boldsymbol{W}_k^{t+1}(\theta, \hat{\theta}) \leftarrow \boldsymbol{W}_k^t(\theta, \hat{\theta}) - \eta \bigtriangledown l(\boldsymbol{W}_k^t)(\theta, \hat{\theta})$
            // $\theta$, $\hat{\theta}$ denotes global and local updating parameters
            $\boldsymbol{S}(i) \leftarrow broadcast(\boldsymbol{F}^d(i)) \odot \boldsymbol{W^d} \otimes P(\hat{\boldsymbol{y}}^i | \boldsymbol{F}^{in}(i))$
            // Eq. 2 & Eq. 3
            $\boldsymbol{S} \leftarrow \frac{1}{N_I} \sum_{i=1}^{N_I} S(i)$
        end for
        // Binary Federated Updating Strategy (BFU)
        if apply BFU Strategy then
            for $u, k = (1, 1)$ to $(U, K)$ do
                $\boldsymbol{D}_{u,k} \leftarrow \frac{\sum_{j=1}^{K} \boldsymbol{S}_{u,k,j} \times \log_2 \boldsymbol{S}_{u,k,j}}{\log_2 K}$  // Eq. 4
                $\boldsymbol{M}_{u,k} \leftarrow \begin{cases} 1 & \boldsymbol{D}_{u,k} > \delta \\ 0 & \boldsymbol{D}_{u,k} < \delta \end{cases}$  // Eq. 5
            end for
            // Integration($\boldsymbol{W}_1^t(\theta), \boldsymbol{W}_2^t(\theta), \cdots, \boldsymbol{W}_k^t(\theta)$)
            for $u, k = (1, 1)$ to $(U, K)$ do
                if $\boldsymbol{M}_{u,k} == 1$ then
                    $\boldsymbol{W}_u^{t+1} \leftarrow \frac{\sum_{k=1}^{K} \boldsymbol{M}_{u,k} \times \boldsymbol{W}_{u,k}^t}{\sum_{k=1}^{K} \boldsymbol{M}_{u,k}}$  // Eq. 6
                end if
            end for
        end if
        // Smooth Federated Updating Strategy (SFU)
        if apply SFU Strategy then
            // Integration($\boldsymbol{W}_1^t(\theta, \hat{\theta}), \boldsymbol{W}_2^t(\theta, \hat{\theta}), \cdots, \boldsymbol{W}_k^t(\theta, \hat{\theta})$)
            for $u, j, = (1, 1)$ to $U, K$ do
                $\boldsymbol{W}_{u,j}^{t+1} = \frac{\sum_{k=1}^{K} \boldsymbol{S}_{u,k,j} \times \boldsymbol{W}_{u,k}^t}{\sum_{k=1}^{K} \boldsymbol{S}_{u,k,j}}$  // Eq. 7
            end for
        end if
    end for
```



(a) Binary Federated Updating Strategy (BFU)

(b) Smooth Federated Updating Strategy (SFU)

Fig. 4. (a) The binary federated updating (BFU) strategy binary distinguishes the adapter units in each client into local client-specific and global client-invariant units and respectively updates them locally and globally. (b) The smooth federated updating (SFU) strategy probabilistically distributes each unit to all clients and thus each unit probabilistically participates in the federated parameter updating of all clients.

where the mask $\boldsymbol{M}_{u,k}$ denotes whether an adapter unit u of the k-th client is global or not.

As shown in Algorithm 1, our binary federated updating (BFU) strategy applies different updating processes for the local and global units during federated updating. As for global units ($\boldsymbol{M}_{u,k} = 1$) in a particular client, their parameter group will be updated as conventional FL. Specifically, the parameters in a local client will be uploaded to the central server and averaged with corresponding parameters from other clients. Then, the average value will be sent back to the local client for updating. The updating process for global units could be formulated as follows:

$$\boldsymbol{W}_u^{t+1} = \frac{\sum_k^K \boldsymbol{M}_{u,k} \times \boldsymbol{W}_{u,k}^t}{\sum_k^K \boldsymbol{M}_{u,k}} \tag{6}$$

where t represents the current updating round, and $t+1$ represents next round. In contrast, if the unit is identified as a local unit ($\boldsymbol{M}_{u,k} = 0$), its parameter group is only updated locally and does not undergo any global updating process. Eventually, our BFS advances the conventional federated learning approaches by enabling binary parameter categorization (*i.e.* client-invariant global parameters and client-specific local parameters), then realizing client-tailored federated learning for heterogeneous medical image segmentation. The detailed updating process of the proposed binary federated updating (BFU) strategy is shown in Algorithm 1.

2) Smooth Federated Updating Strategy: Although the above BFU strategy already realizes client-tailored federated learning through binary parameter categorization, it is somewhat rigid and not good enough for complex scenarios where parameters simultaneously encapsulate complicated information from multiple client domains. To solve such limitation, we further proposed a Smooth Federated Updating (SFU) strategy. It is conceptually built upon the above BFU and could be treated as a more generalized version of BFU because it utilizes a smoother parameter association substituting the binary parameter categorization. Specifically, unlike the binary strategy that assigns each unit exclusively to be local or global, our smooth federated updating strategy considers each unit probabilistically (rather than exclusively) belonging to a client. As shown in Fig. 4 (b), we exploit the weighted contribution score of each unit on different clients to weight the parameter group of each unit on each client, generating a unit-wise weighted model for each client. Then, each client exploits the weighted models of all clients to update itself. Specifically, when updating the parameter group $\boldsymbol{W}_{u,j}^t$ of unit u on client j at updating round t, we use the probabilistic contribution score for every client to weight the parameter updating from every client k, *i.e.*

$$\boldsymbol{W}_{u,j}^{t+1} = \frac{\sum_k^K \boldsymbol{S}_{u,k,j} \times \boldsymbol{W}_{u,k}^t}{\sum_k^K \boldsymbol{S}_{u,k,j}}. \tag{7}$$

As a result, our SFU incorporates the weighted contribution score regarding every client as smoothly adjusting weights for federated aggregating, which enables a more adaptive and dynamic updating of model parameters and realizes smooth client-tailored segmentation for each client. It allows the parameter update in a local client contributing to the global aggregating in proportion to its relevance to other clients.

In summary, our BFU strategy offers a straightforward parameter categorization pipeline for heterogeneous federated learning. The SFU further leverages probabilistic weighting based on client-specific scores, providing a more nuanced and flexible mechanism for parameter management during federated updating, significantly enhancing the performance of heterogeneous federated segmentation.

IV. Experiments

A. Datasets

1) CXRS-HG Dataset: The CXRS-HG dataset is a heterogeneous distributed dataset constructed from the Chest X-ray Segmentation (CXRS) dataset [53]. Original CXRS is an in-house dataset comprising 1250 chest X-ray images of 30 different anatomical structures (including 24 ribs, 2 clavicles, 2 scapulae, and 2 lungs) for medical image segmentation, and each image is annotated with 30 anatomical segmentation masks. We reform the centralized CXRS dataset into a distributed dataset CXRS-HG containing client-wise heterogeneity. Following the classical protocol for federated learning in [54], we first distribute the image samples to distributed clients following a standard Dirichlet distribution to simulate the heterogeneity of class imbalance in a practical medical scenario (as shown in Fig. 2 (a). Then, the distributed X-ray images in different local clients will undergo different image transformation strategies (original images for client$_1$,

3×3 mean blur filtering for client$_2$, half down-sampling and then restoring resolution for client$_3$) to simulate the heterogeneity induced by various environmental factors (such as surroundings and imaging devices). As a result, the distributed CXRS-HG contains abundant client-wise heterogeneity including class imbalance and distribution diversity which is highly similar to non-centralized medical segmentation scenarios. Following [40], [55], we employ the mean Dice Similarity Coefficient (mDice) as the metric to evaluate segmentation performance on the CXRS-HG dataset, which is the class-wise mean of the Dice Similarity Coefficient for each class.

2) HLS Dataset: To examine the proposed FCA framework on real-world distributed medical dataset, we exploit several public datasets released by different institutes for lung segmentation in X-ray images, forming a heterogeneous lung segmentation dataset (referred to as HLS). The HLS dataset consists of four independent subsets, including COVID-19 x-ray dataset [56], covid-chestxray dataset [57], QaTa-COV19 dataset [58], and COVID-19 Chest X-ray Segmentation dataset [59]. Specifically, the COVID-19 X-ray dataset [56] contains 6500 images of chest X-rays with pixel-level polygonal lung segmentation masks, among which 517 cases are from COVID-19 patients. The COVID-19 covid-chestxray dataset [57] contains 542 chest X-ray and CT images from COVID-19 patients or other viral and bacterial pneumonia (MERS, SARS, and ARDS) patients. The QaTa-COV19 dataset [58] consists of 9258 COVID-19 chest X-ray images collected by Qatar University and Tampere University. The COVID-19 Chest X-ray Segmentation [59] consists of a collection of a total of 100 Chest X-ray images from the Novel Coronavirus (COVID-19) cases. We assign each subset of HLS to different clients forming multiple "data islands" containing real heterogeneous environmental factors (such as races, surroundings, and imaging devices). The standard mDice serves as the metric to evaluate the segmentation performance on this dataset.

3) AMD-SD-HG: The AMD-SD [60] dataset contains 3049 B-scan images from 138 patients with segmentation categories of subretinal effusion, subretinal effusion, elliptical continuity, subretinal hyperreflective material, and pigment epithelium detachment. Similar to the above CXRS-HG dataset, we transform it into a heterogeneous distributed dataset AMD-SD-HG with the same transformation strategies to simulate client-wise heterogeneity in practical medical scenarios. We use the standard mDice metric to evaluate the segmentation performance on this dataset.

B. Implementation Details

The threshold δ for score binarization in our BFU strategy (Eq. 5) is empirically set as 0.25. We employ the SAM-Med2D [10] as an example of MFMs in Fig. 3 (a) to examine our FCA framework if not specially mentioned. The layer number N_L in the image encoder (Fig. 3 (a)) of SAM-Med2D is set as 12. As for the training details, the standard Binary Cross-Entropy (BCE) loss [3] is applied for the segmentation head, and the Cross-Entropy loss is applied to optimize the client discriminator. During federated learning, the global model in the central server is updated in total for $T = 60$ rounds. During each global round, the local clients are individually trained for N_e local epochs ($N_e = 5$ for the CXRS-HG dataset and $N_e = 3$ for the HLS dataset), then the updated local parameters will be uploaded to the central server for sequential integration, which is similar to [15]. For fair comparisons, this setup is applied to the proposed FCA and all comparisons. All the models are trained via the Adam optimizer with a learning rate of 0.001 and a weight decay of 0.0001 on 4 NVIDIA RTX 3090 graphics cards.

C. Comparison With State-of-the-Art Methods

To validate the effectiveness and superiority of the proposed federated client-tailored adapter (FCA) framework,we conduct comparison experiments on three large-scale heterogeneous distributed datasets for medical segmentation, including the CXRS-HG, HLS and AMD-SD-HG datasets. We report the mDice on three local clients and the average mDice across all clients for comprehensive performance comparison.

We first compare our FCA framework with other state-of-the-art FL methods, including the FedSeg [18], FedProx [34], HarmoFL [49], IOP-FL [17], FedA3I [41], FedCross [51], IOP-FL [17], PerFedAvg [50], and MAP [52]. For fair comparison, we further re-implemented several enhanced variants by enhancing them with the same MFMs (*i.e.* SAM-Med2D) backbone, then training with the same protocols as the proposed method. As shown in Table I, the re-implemented MFMs enhanced variants perform significantly better than those without MFMs since stirring universal prior knowledge in MFMs helps to improve and stabilize heterogeneous federated learning (see Fig. 1 (b)). In addition, our FCA respectively outperforms the second-best method [17] equipped with the same MFMs by large margins of 3.51%, 1.65% on the CXRS-HG, HLS datasets mainly attributed to the following reasons: (1) Unlike non-tailored FL methods (*i.e.*, FedAvg, FedProx), our FCA allows each client to undergo customized federated updating, thus enabling the optimal tailored model for each client. (2) Although the IOP-FL [17], FedCross [51], MAP [52], and PerFedAvg [50] also attempt to customize federated models for local clients, they decouple at coarse-grained model or module levels [14], [19] and allocate parameter groups [20], [21] statically, lacking enough adaptability. While our FCA framework achieves better performance via dynamic and fine-grained parameter decompose. (3) Our FCA framework equipped with the smooth federated updating strategy (FCA-SFU) performs better than that equipped with the binary federated updating strategy (FCA-BFU). It is because our SFU probabilistically distributes the adapter units to multiple clients, thus achieving more fine-grained and adaptive federated updating.

To further validate the generalization across different scales and modalities, we compared our FCA framework with several representative methods on the large-scale AMD-SD-HG dataset, including two baselines (FedAvg, FedAvg* [15]), a personalized approach (FedCross* [51]), a decoupling approach (MAP* [52]), and the existing state-of-the-art approach (IOP-FL* [17]). As shown in Table II, our FCA framework consistently outperforms these approaches with large margins. Moreover, the standard deviation (STD) of

TABLE I

Comparing With the State-of-the-Art and Baseline Methods Regarding mDice on the CXRS-HG and HLS Datasets

Methods	CXRS-HG				HLS			
	Average	Client$_1$	Client$_2$	Client$_3$	Average	Client$_1$	Client$_2$	Client$_3$
FedAvg [15] (baseline$_1$)	36.94	35.82	37.71	37.29	58.41	58.39	58.80	58.04
FedProx [34]	37.74	37.40	37.97	37.85	57.20	56.81	57.47	57.32
HarmoFL [49]	42.85	42.73	42.97	42.85	51.13	50.27	51.52	51.60
FedSeg [18]	40.35	38.42	41.56	41.07	81.72	79.13	82.69	83.34
FedA3I [41]	41.99	53.75	34.02	38.22	81.58	80.98	83.30	80.48
FedAvg* [15] (baseline$_2$)	60.64	60.58	60.69	60.67	90.79	88.76	90.87	92.75
FedProx* [34]	61.06	61.27	61.48	60.43	91.06	88.80	91.38	92.84
PerFedAvg* [50]	61.35	61.11	61.39	61.56	91.37	89.95	91.28	92.90
FedCross* [51]	61.44	61.09	61.75	61.29	91.03	89.21	91.01	92.87
MAP* [52]	61.54	61.22	61.39	62.01	91.49	90.18	91.54	92.75
IOP-FL* [17]	62.00	61.21	61.86	61.69	91.68	90.34	91.96	92.69
FCA-BFU (**ours**)	**63.41**	**63.31**	**63.20**	**63.15**	**92.12**	**90.60**	**92.26**	**93.50**
FCA-SFU (**ours**)	**64.15**	**64.08**	**64.47**	**63.90**	**92.44**	**90.82**	**92.52**	**93.98**

* Our implementation of MFMs enhanced variants of conventional FL methods.

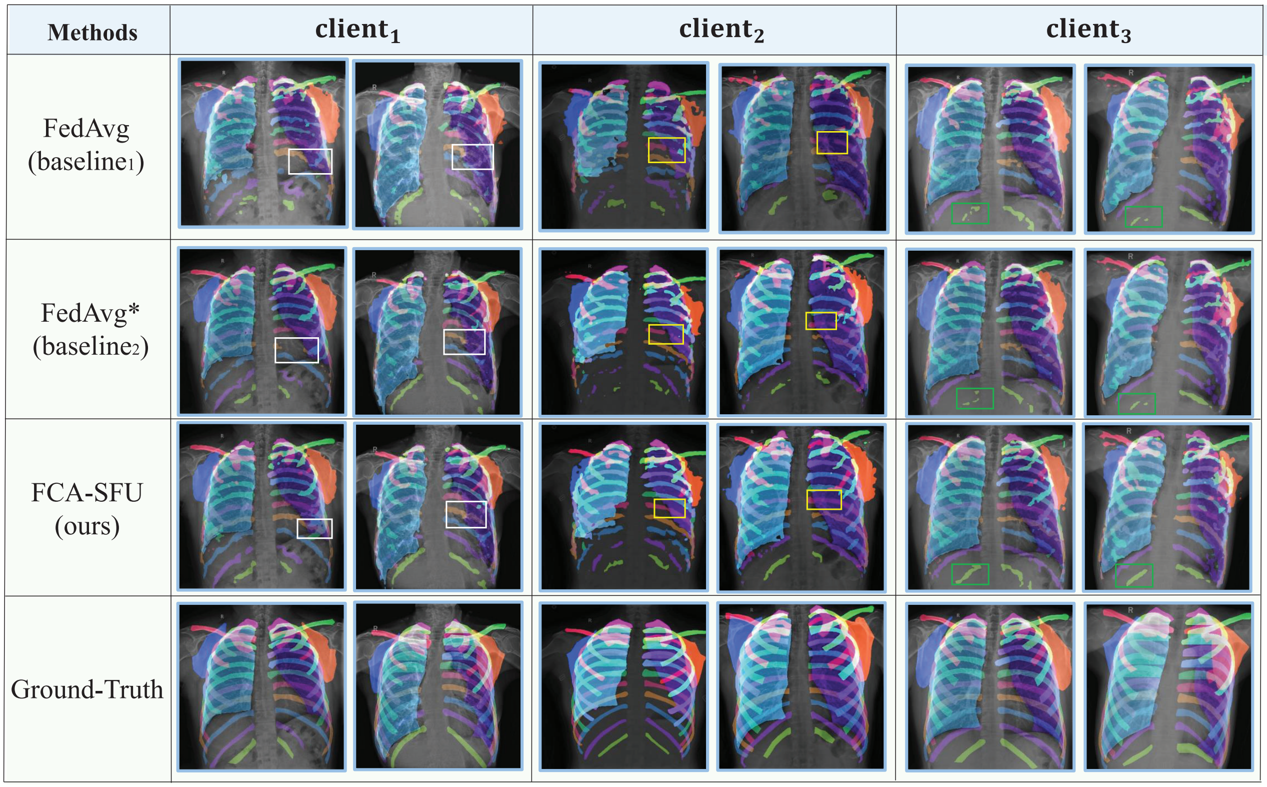


Fig. 5. Visualization comparison of heterogeneous federated segmentation results on the CXRS-HG dataset.

performance across clients also indirectly reflects the personalize (client-tailored) degree in distributed clients. As shown in Table II, our FCA method achieve high diversity, which effectively validates the effectiveness of our client-tailored updating strategies. We note that the high STD values obtained from FedAvg [15] and FedAvg* [15] methods should owe to bad global compromise rather than the client-tailored updating since they are global-compromised not client-tailored federated learning methods.

In addition, we qualitatively compare the segmentation masks obtained from our FCA-SFU, the FedAvg* (the MFMs enhanced variant of the FedAvg [15]), and the original FedAvg [15] for effectiveness validation. As shown in Fig. 5 and Fig. 6, our FCA-SFU achieves more precision segmentation than the FedAvg* variant on every local client. Besides, the difficulty of segmenting the same anatomical structure differs across clients (*e.g.* the boxed areas in Fig. 5) since the complicated client-wise heterogeneity. Conventional FL methods like the FedAvg* variant only obtain a global-compromised model for all clients that leads to a sub-optimal compromised segmentation. In contrast, our FCA-SFU customizes a local model for each client that achieves an optimal client-tailored segmentation for each client. In Fig. 6, some clients (such as *client*$_2$) require simultaneously segmenting out the shadow area that overlapped by the mediastinum and right lung (see corresponding ground-truth annotations) while other clients

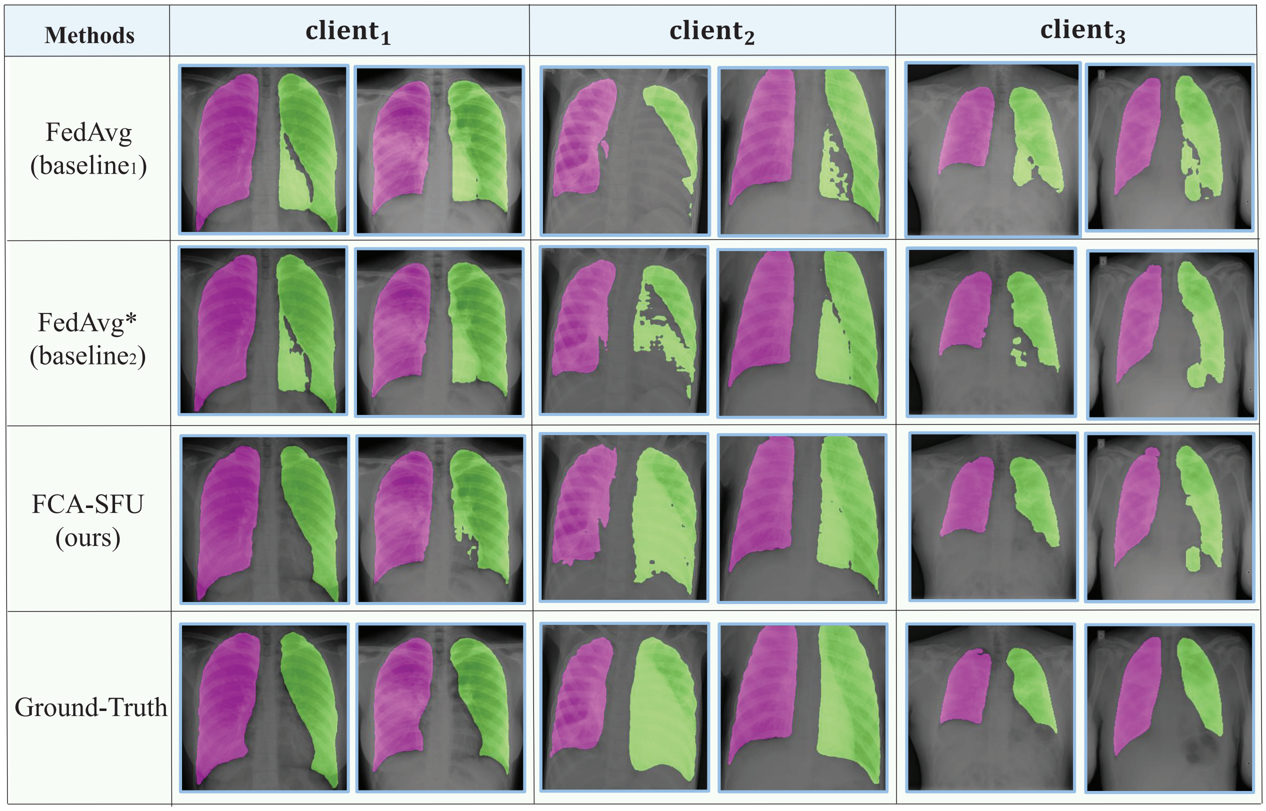


Fig. 6. Visualization comparison of heterogeneous federated segmentation results on the HLS dataset.

TABLE II
COMPARING WITH THE STATE-OF-THE-ART AND BASELINE METHODS REGARDING MDICE ON THE AMD-SD-HG DATASET

Methods	Client_1	Client_2	Client_3	Average	STD
FedAvg [15] (baseline_1)	44.25	46.03	45.1	45.12	0.7269
FedAvg* [15] (baseline_2)	59.55	61.42	60.31	60.42	0.7679
FedCross* [51]	60.56	61.01	60.56	60.71	0.2121
MAP* [52]	60.96	61.65	60.69	61.1	0.4042
IOP-FL* [17]	61.24	61.88	61.13	61.41	0.3307
FCA-BFU (**ours**)	**62.15**	**63.12**	**62.41**	**62.56**	**0.4100**
FCA-SFU (**ours**)	**62.5**	**64.47**	**63.24**	**63.40**	**0.8125**

* Our implementation of MFMs enhanced variants.

(*client*$_1$, *client*$_3$) do not need to segment this area. Since the FedAvg* variant only has a global-compromised segmentation model, the *client*$_1$ and *client*$_3$ mis-segmented out this shadow area. In contrast, our FCA-SFU was not affected by this annotation heterogeneity demonstrating the superiority and effectiveness of our client-tailored adapter framework.

D. Ablation Studies

To examine the effectiveness of each component, we compare our FCA with a series of baselines and variants, the main results already contained in Table I. Specifically, the FCA-BFU and FCA-SFU respectively are the proposed Federated Client-tailored Adapter (FCA) equipped with our binary (BFU) and smooth (SFU) federated updating strategies. FedAvg* [15] (baseline_2) is a baseline constructed with the same network (*i.e.* SAM-Med2D) as our FCA except it utilizes a conventional federated updating strategy FedAvg [15]. The FedAvg [15] (baseline_1) is the vanilla FedAvg baseline without assistance from off-the-shelf MFMs. The results of baseline_1 and baseline_2 indicate that the prior knowledge contained in off-the-shelf MFMs indeed benefits a lot for improving segmentation performance (Table I) and stabilizing the heterogeneous federated updating (see Fig. 1 (b)). Eventually, our client-tailored adapter FCA-SFU further improves the individual performance of each client and sets a new start-of-art result on both CXRS-HG and HLS datasets.

TABLE III
EXPERIMENTAL VALIDATION OF ALLEVIATING VARIOUS CLIENT-WISE HETEROGENEITY ON THE CXRS-HG DATASET

Heterogeneity	FedAvg* [15] (baseline_2)	FCA-SFU (ours)
IID	69.70	**70.89**
Class Imbalance	61.78	**64.27**
Distribution Diversity	62.91	**64.05**
Imbalance+Diversity	60.64	**64.15**

E. Experimental Validation of Alleviating Heterogeneity

In this section, we further examine the capability of our FCA for alleviating various client-wise heterogeneity and providing client-tailored segmentation. Since the CXRS-HG dataset contains two kinds of heterogeneity including class imbalance and distribution diversity, we split them alone and construct corresponding dataset variants to train two variant models. As shown in Table III, our FCA-SFU significantly boosts

TABLE IV

EXPERIMENTAL VALIDATION OF GENERALIZATION ABILITY FOR DIFFERENT MFMS ON THE CXRS-HG DATASET

Medical Foundation Models (MFMs)	**Client$_k$**	**Average**
FedAvg [15] without MFMs (baseline$_1$)	35.82	36.94
FedAvg* [15] with H-SAM [61]	59.13	58.36
FCA-SFU with H-SAM [61] (**ours**)	62.51	62.15
FedAvg* [15] with Med-SA [48]	47.10	45.92
FCA-SFU with Med-SA [48] (**ours**)	53.91	51.53
FedAvg* [15] with SAM-Med2D [10] (baseline$_2$)	60.69	60.64
FCA-SFU with SAM-Med2D [10] (**ours**)	64.47	64.15

* Our implementation of corresponding MFMs enhanced variants.

TABLE V

COMPARISON AND ANALYSIS OF COMPUTATIONAL EFFICIENCY ON THE CXRS-HG DATASET

Methods	#Params (M)	#Time (h)
FedAvg* (baseline$_2$)	180.50	36.14
FedCross*	180.50	108.39
IOP-FL*	180.50	39.61
FCA-SFU* (**ours**)	180.59	36.64

* Our implementation of SAM-Med2D enhanced variants.

performance (mDice) from 61.78 to 64.27 and from 62.91 to 64.05 regarding heterogeneity types of class imbalance and distribution diversity. When dealing with the more complicated heterogeneity mixture (Imbalance+Diversity), the performance drop from our FCA-SFU is negligible while the drop from FedAvg* [15] is prominent. This difference further validates that our FCA-SFU is good at learning client-tailored models for each client and effectively alleviating various heterogeneity existing in distributed medical image segmentation.

F. Experimental Validation of MFMs Generalization

In this section, we validate that our federated client-tailored adapter framework is generalizable to various MFMs consisting of transformers. As shown in Table IV, we examine our FCA-SFU on three medical foundation models including the SAM-Med2D [10], Med-SA [8] and H-SAM [61]. We observe that our FCA-SFU shows consistent improvements over the conventional FL baseline method FedAvg* [15], indicating the generalization ability of the proposed FCA framework for various MFMs.

G. Comparison and Analysis of Computational Efficiency

In this section, we compare and analyze the computational efficiency by measuring the model parameters (#Params) and convergence time (#Time), respectively. For fair comparisons, all the methods are implemented and enhanced by the same MFMs (SAM-Med2D [10]). Besides, since all compared methods nearly involve the same parameter transmission and collection process, the model parameters (#Params) for each local model copy could also be treated as an indirect metric to evaluate the communication overhead during federated learning. As shown in Table V, our FCA-SFU achieves significant performance improvement with a slight increase in model parameters and communication overhead. Moreover, the convergence speed of our FCA-SFU is much faster than other state-of-the-art methods.

V. CONCLUSION

This paper introduces a generalizable framework Federated Adaptive Adapter (FCA) for heterogeneous medical image segmentation. We identify the training instability issue induced by client-wise heterogeneity in the conventional federated learning paradigm and propose two measures to alleviate it. One measure distills the universal knowledge in off-the-shelf medical foundation models to stabilize heterogeneous federated learning via the parameter-efficient adapter. Another measure decomposes the adapter units in each client into client-specific and client-invariant parts and develops different federated updating strategies for them. This measure further stabilizes the heterogeneous federated training process and realizes client-tailor federated learning at the same time. Eventually, the FCA achieves state-of-the-art performance on three datasets for heterogeneous medical image segmentation.

REFERENCES

[1] L. Liu, J. M. Wolterink, C. Brune, and R. N. J. Veldhuis, "Anatomy-aided deep learning for medical image segmentation: A review," *Phys. Med. Biol.*, vol. 66, no. 11, Jun. 2021, Art. no. 11TR01.

[2] J. Sun, Y. Peng, Y. Guo, and D. Li, "Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN," *Neurocomputing*, vol. 423, pp. 34–45, Jan. 2021.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Jan. 2015, pp. 234–241.

[4] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[5] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "A nested U-Net architecture for medical image segmentation," in *Proc. MICCAI*, 2018, pp. 3–11.

[6] Y. Xie, J. Zhang, C. Shen, and Y. Xia, "CoTr: Efficiently bridging CNN and transformer for 3D medical image segmentation," in *Proc. MICCAI*, Jan. 2021, pp. 171–180.

[7] O. Dalmaz, M. Yurt, and T. Çukur, "ResViT: Residual vision transformers for multimodal medical image synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 10, pp. 2598–2614, Oct. 2022.

[8] J. Wu et al., "Medical SAM adapter: Adapting segment anything model for medical image segmentation," 2023, *arXiv:2304.12620*.

[9] V. I. Butoi, J. J. G. Ortiz, T. Ma, M. R. Sabuncu, J. V. Guttag, and A. V. Dalca, "UniverSeg: Universal medical image segmentation," in *Proc. ICCV*, Oct. 2023, pp. 21438–21451.

[10] J. Cheng et al., "SAM-Med2D," 2023, *arXiv:2308.16184*.

[11] Y. Liu, Z. Ma, X. Liu, S. Ma, and K. Ren, "Privacy-preserving object detection for medical images with faster R-CNN," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 69–84, 2022.

[12] X. Liu, H. Li, G. Xu, Z. Chen, X. Huang, and R. Lu, "Privacy-enhanced federated learning against poisoning adversaries," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4574–4588, 2021.

[13] L. Qiu, J. Cheng, H. Gao, W. Xiong, and H. Ren, "Federated semi-supervised learning for medical image segmentation via pseudo-label denoising," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 10, pp. 4672–4683, Oct. 2023.

[14] X. Yang, B. Xiong, Y. Huang, and C. Xu, "Cross-modal federated human activity recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5345–5361, Aug. 2024.

[15] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[16] A. Jiménez-Sánchez, M. Tardy, M. Á. G. Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," *Comput. Methods Programs Biomed.*, vol. 229, Dec. 2022, Art. no. 107318.

[17] M. Jiang, H. Yang, C. Cheng, and Q. Dou, "IOP-FL: Inside–outside personalization for federated medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 7, pp. 2106–2117, Jul. 2023.

[18] J. Miao, Z. Yang, L. Fan, and Y. Yang, "FedSeg: Class-heterogeneous federated learning for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8042–8052.

[19] Y. Shen, Y. Zhou, and L. Yu, "CD2-pFed: Cyclic distillation-guided channel decoupling for model personalization in federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10031–10040.

[20] L. Xie et al., "PFLFE: Cross-silo personalized federated learning via feature enhancement on medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jan. 2024, pp. 599–610.

[21] X. Chen et al., "Towards optimal customized architecture for heterogeneous federated learning with contrastive cloud-edge model decoupling," *IEEE Trans. Comput.*, vol. 74, no. 4, pp. 1123–1137, Apr. 2024.

[22] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[23] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[24] J. Ruan, J. Li, and S. Xiang, "VM-UNet: Vision mamba UNet for medical image segmentation," 2024, *arXiv:2402.02491*.

[25] G. Hu, B. He, and H. Zhang, "Compositional prompting video-language models to understand procedure in instructional videos," *Mach. Intell. Res.*, vol. 20, no. 2, pp. 249–262, Apr. 2023.

[26] H. Zhu, Y. Chen, G. Hu, and S. Yu, "Contrastive learning via local activity," *Electronics*, vol. 12, no. 1, p. 147, Dec. 2022.

[27] G. Hu, B. Cui, and S. Yu, "Joint learning in the spatio-temporal and frequency domains for skeleton-based action recognition," *IEEE Trans. Multimedia*, vol. 22, no. 9, pp. 2207–2220, Sep. 2020.

[28] H. Zhu, Y. Chen, G. Hu, and S. Yu, "Information-density masking strategy for masked image modeling," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2023, pp. 1619–1624.

[29] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Commun.*, vol. 15, no. 1, p. 654, Jan. 2024.

[30] Y. Li, B. Jing, Z. Li, J. Wang, and Y. Zhang, "NnSAM: Plug-and-play segment anything model improves nnUNet performance," 2023, *arXiv:2309.16967*.

[31] J. Gao, L. Zhao, and X. Li, "NWPU-MOC: A benchmark for fine-grained multicategory object counting in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5606614.

[32] Z. Hu, J. Gao, Y. Yuan, and X. Li, "Contrastive tokens and label activation for remote sensing weakly supervised semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5620211.

[33] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, and C.-Z. Xu, "FedDC: Federated learning with non-IID data via local drift decoupling and correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10112–10121.

[34] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 429–450.

[35] A. Xu et al., "Closing the generalization gap of cross-silo federated medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20866–20875.

[36] Y. Dai, Z. Chen, J. Li, S. Heinecke, L. Sun, and R. Xu, "Tackling data heterogeneity in federated learning with class prototypes," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, 2023, pp. 7314–7322.

[37] J. Zhang et al., "Fed-CBS: A heterogeneity-aware client sampling mechanism for federated learning via class-imbalance reduction," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2022, pp. 41354–41381.

[38] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, Jul. 2020, pp. 5132–5143.

[39] X. Wu, H. Huang, Y. Ding, H. Wang, Y. Wang, and Q. Xu, "FedNP: Towards non-IID federated learning via federated neural propagation," in *Proc. AAAI*, vol. 37, Jun. 2023, pp. 10399–10407.

[40] B. Ma et al., "FedST: Federated style transfer learning for non-IID image segmentation," in *Proc. AAAI*, vol. 38, Mar. 2024, pp. 4053–4061.

[41] N. Wu, Z. Sun, Z. Yan, and L. Yu, "FedA3I: Annotation quality-aware aggregation for federated medical image segmentation against heterogeneous annotation noise," in *Proc. AAAI*, vol. 38, Mar. 2024, pp. 15943–15951.

[42] J. Guo, H. Liu, S. Sun, T. Guo, M. Zhang, and C. Si, "FSAR: Federated skeleton-based action recognition with adaptive topology structure and knowledge distillation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 10400–10410.

[43] M. Li and G. Yang, "Where to begin? From random to foundation model instructed initialization in federated learning for medical image segmentation," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2024, pp. 1–5.

[44] Y.-x. Liu, G. Luo, and Y. Zhu, "FedFMS: Exploring federated foundation models for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Jan. 2024, pp. 283–293.

[45] H. Chen, Y. Zhang, D. Krompass, J. Gu, and V. Tresp, "FedDAT: An approach for foundation model finetuning in multi-modal heterogeneous federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 11285–11293.

[46] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 4015–4026.

[47] V. Lialin, V. Deshpande, X. Yao, and A. Rumshisky, "Scaling down to scale up: A guide to parameter-efficient fine-tuning," 2023, *arXiv:2303.15647*.

[48] Y. Zhang, Z. Shen, and R. Jiao, "Segment anything model for medical image segmentation: Current applications and future directions," *Comput. Biol. Med.*, vol. 171, Feb. 2024, Art. no. 108238.

[49] M. Jiang, Z. Wang, and Q. Dou, "HarmoFL: Harmonizing local and global drifts in federated learning on heterogeneous medical images," in *Proc. 36th AAAI Conf. Artif. Intell.*, 2022, pp. 1087–1095.

[50] J. Zhou et al., "Personalized and privacy-preserving federated heterogeneous medical image analysis with PPPML-HMI," *Comput. Biol. Med.*, vol. 169, Dec. 2023, Art. no. 107861.

[51] X. Xu et al., "Federated cross learning for medical image segmentation," in *Proc. Med. Imag. Deep Learn.*, Jan. 2022, pp. 1441–1452.

[52] X. Li et al., "MAP: Model aggregation and personalization in federated learning with incomplete classes," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 6560–6573, Apr. 2024.

[53] G. Hu, Y. Kang, G. Zhao, Z. Jin, C. Li, and J. Tang, "Dynamic strip convolution and adaptive morphology perception plugin for medical anatomy segmentation," *IEEE Trans. Med. Imag.*, vol. 44, no. 6, pp. 2541–2552, Jun. 2025.

[54] X. Mu et al., "FedProc: Prototypical contrastive federated learning on non-IID data," *Future Gener. Comput. Syst.*, vol. 143, pp. 93–104, Jun. 2023.

[55] J. Wang, Y. Jin, D. Stoyanov, and L. Wang, "FedDP: Dual personalization in federated medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 43, no. 1, pp. 297–308, Jan. 2024.

[56] S. Edwardsson. (2020). *COVID-19-Xray-Dataset*. [Online]. Available: https://github.com/v7labs/covid-19-xray-dataset

[57] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*.

[58] A. Degerli, S. Kiranyaz, M. E. H. Chowdhury, and M. Gabbouj, "Osegnet: Operational segmentation network for COVID-19 detection using chest X-ray images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2306–2310.

[59] A. N. M. Sakib. (2020). *COVID-19-Chest-Xray-Segmentations-Dataset*. [Online]. Available: https://github.com/generalblockchain/covid-19-chest-xray-segmentations-dataset

[60] Y. Hu et al., "AMD-SD: An optical coherence tomography image dataset for wet AMD lesions segmentation," *Scientific Data*, vol. 11, no. 1, p. 1014, Sep. 2024.

[61] Z. Cheng et al., "Unleashing the potential of SAM for medical adaptation via hierarchical decoding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 3511–3522.