Personalizing Federated Learning Guided by Site-aggregated Representation for Multi-site One-shot Medical Image Segmentation

Jia Wang
School of Software Technology
Dalian University of Technology
Dalian, China
wangjia@mail.dlut.edu.cn

Yuchen Sun
School of Software Technology
Dalian University of Technology
Dalian, China
syc2004615@gmail.com

Yunan Mei
School of Software Technology
Dalian University of Technology
Dalian, China
dlut_myn@163.com

Zihao Xu
School of Software Technology
Dalian University of Technology
Dalian, China
18167127296@mail.dlut.edu.cn

Xin Fan
School of Software Technology
Dalian University of Technology
Dalian, China
xin.fan@dlut.edu.cn

ABSTRACT

Personalized federated learning for medical image segmentation enables collaborative model training across multiple clinical sites without sharing patient data, by synchronizing a subset of global model parameters while retaining others for local adaptation. However, previous methods adopt paired labeled images to train the model, which is hard to apply in real scenarios due to time-consuming medical experts on manual annotation. Additionally, these methods focus on local parameter learning ignoring inter-site consistencies during local training. To address these challenges, we propose a personalizing **Fed**erated framework guided by **S**ite-aggregated **R**epresentation (**FedSR**) for multi-site one-shot medical image segmentation, which exploits the site-invariant latent information to boost segmentation performance. Specifically, we propose to learn an omniscient encoder by federated learning, which can not only model the data distribution between multi-site datasets but also adapt the multi-task in an efficient way. With the learned robust representation, we further propose to learn site-aggregated representation between multi-site data by mutual information maximization, and then adopt such site-aggregated latent representation to guide the personalized dual-task head decoder. Extensive experiments conducted on two MIS tasks demonstrate that the proposed FedSR outperforms state-of-the-art one-shot MIS methods on segmentation.

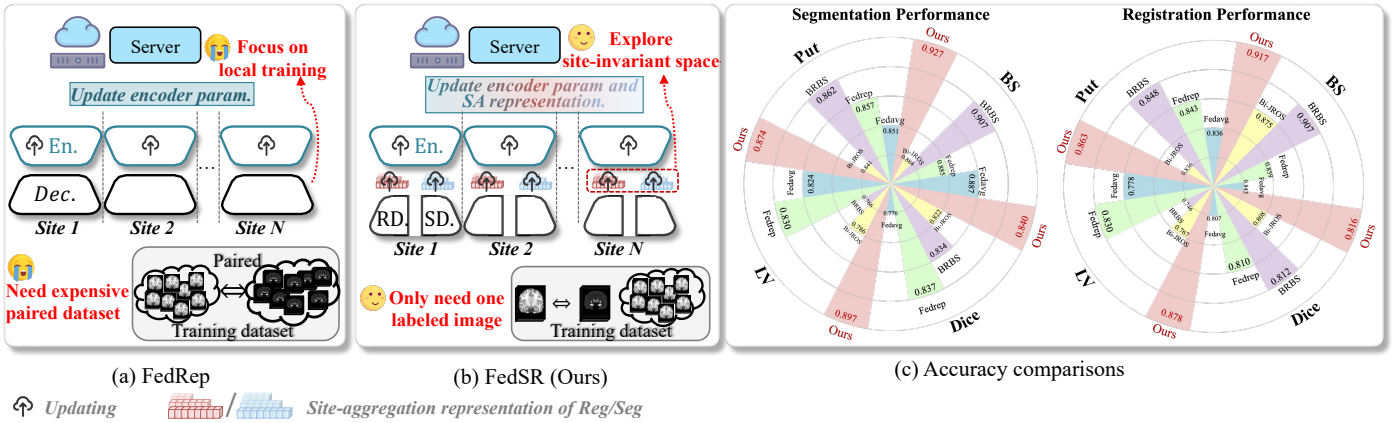
I. INTRODUCTION

As abundant training data does not exit in clinical settings, collaborative learning across multiple medical institutions is a promising solution for maximizing the potential of the data-driven deep learning model in medical image segmentation (MIS). However, data communication across multi-site encounters obstacles to patient privacy protection. Federated learning (FL) [1]–[4] draws much attention for a privacy-preserving solution, which allows the different sites to train a global model without sharing and centralizing data. Specifically, each local medical client trains a model downloaded from the server with its own data, then updates the local model parameters to the global server, and finally the updated global parameters are broadcast back to each client. Although FL has achieved promising performance in MIS, most existing methods adopt a single model to adapt to various data distributions from multi-clients, resulting in poor performance.

Recent personalized FL proposed to alleviate the above problems by reducing the communication part of local parameters and only updating partial parameters to the global server. As shown in Fig.1(b), the recent personalized FL method FedRep [5] splits the model into the encoder part and decoder part, where the former is updated to the server by global averaging and the latter is trained with local data. Despite promising performance, there still exist two main limitations: i) Previous FL-based MIS methods [6]–[8] heavily rely on extensive manual labeling data resulting in impractical. ii) They only consider the site-specific discriminative representation when performing the *local training*.

One-shot MIS has emerged to pursue high segmentation performance by employing one-labeled data to alleviate the challenge. The mainstream one-shot MIS methods adopt joint registration and segmentation (JRS) framework, by exploring voxel-wise correspondence between the atlas and unlabeled images, addressing the predicament of paucity in training samples. The JRS paradigm [9]–[12] combines medical image registration and segmentation, fostering a reciprocal enhancement of both tasks. Specifically, the output of registration serves as an input to the segmentation model, and conversely, the output from segmentation guides and constrains the learning trajectory of the registration model. This symbiosis enhances the segmentation accuracy through improved spatial consistency afforded by registration, while precise segmentation yields additional structured information to direct the registration model toward a more accurate alignment of anatomical features. Despite these methods alleviating the labeling challenge, *they still require training a model from scratch for each site to adapt to the distribution differences in the datasets*, which is time-consuming. In this work, we focus on leveraging the advantages of federated learning and one-shot MIS.

In this paper, we propose a personalizing federating framework guide by site-aggregation representation (FedSR) for one-shot MIS, unifying the personalized feature and site-aggregated representation to boost segmentation performance.



(a) FedRep (b) FedSR (Ours) (c) Accuracy comparisons

Fig. 1. (a) FedRep: A typical federated representation learning framework that updates encoder parameters locally, relying heavily on expensive paired datasets and lacking inter-site collaboration. (b) FedSR (Ours): Our framework aggregates site-invariant representations for registration and segmentation (RD and SD), requiring only one labeled image per site and enabling effective cross-site knowledge sharing. (c) Performance comparison on segmentation and registration tasks on the ABIDE dataset. Our method achieves the best results on each structure as well as the highest overall average performance.

Algorithm 1: The training procedure of FedSR.

Input: Number of communication rounds I, number of sites N, a labeled image $\mathcal{X}_a$; The set of multi-site unlabeled data $\{\mathcal{T}^n\} = \mathcal{X}_{u.n}$
Output: deformation field ψ, predicted label $\hat{\mathcal{Y}}$
Server executes:
Initial global encoder ω_e^g;
Initial global SA representation $f_u \in \{f_s, f_r\}$;
Initial local decoder $\omega_{d.n}$
for $i = 0, 1, ..., I-1$ **do**
  for $n = 0, 1, ..., N-1$ **do**
    Send the global encoder model $\omega_{e,i}^g$ and SA representation f_i^g to $\mathcal{T}_i$
    $\omega_{e.i}^l, f_i^l \leftarrow LocalTraining(i, \omega_e^g, f_i^g)$
  $\omega_{e.i+1}^g \leftarrow \frac{1}{N}\sum_{n=1}^{N} \omega_{e.i}^l$
  $f_{i+1}^g \leftarrow \max \Phi^{MI}(f_i^l, f_i^g)$
return f_{i+1}^g *and* ω_{i+1}^g
LocalTraining(i, ω_e^g, f_i^g)**:**
for *epoch e = 0,1,...,E* **do**
  for *each batch b =* $\{\mathcal{X}_a, \mathcal{X}_{u.n}\}$ **do**
    # Optimize the registration task
    $(\omega_{e.n}^l, \omega_{d.n}) \leftarrow (\omega_{e.n}^l, \omega_{d.n}) - \lambda_1 \nabla \mathcal{L}_{smo} - \lambda_2 \nabla \mathcal{L}_{sim} - \lambda_3 \nabla \mathcal{L}_{dice}$
    $f_{r.(n+1)} \leftarrow \Phi^{MI}(f_r^g, f_{r.n})$
    # Optimize the segmentation task
    $(\omega_{e.n}^l, \omega_{d.n}) \leftarrow (\omega_{e.n}^l, \omega_{d.n}) - \lambda_1 \nabla \mathcal{L}_{smo} - \lambda_2 \nabla \mathcal{L}_{sim} - \lambda_3 \nabla \mathcal{L}_{dice}$
    $f_{s.(n+1)} \leftarrow \Phi^{MI}(f_s^g, f_{s.n})$
return $f_{s.(n+1)}, f_{r.(n+1)}, \omega_{e.(n+1)}$

Our framework consists of a shared encoder and two task-specific decoders, which federates the multi-site encoder to get an omniscient encoder for generating robust features boosting the registration and segmentation task training as shown in Fig.1 (b). Moreover, different from the previous personalized FL methods that focus on modeling local data distribution while ignoring the utilization of site-aggregated (SA) information during local training, we propose a site-aggregated latent representation to preserve common information between multi-site data by mutual information maximization. Such representation will be updated as the client uploads the parameters to the server. With the learned SA latent representations, we propose a site-common information enhancer (SCIE) to further guide dual-task decoder training.

The major contributions of this work can be summarized as follows:

- We propose a novel personalizing federated learning framework guided by site-aggregated representation for one-shot MIS, performing segmentation with one labeled image and achieving privacy protection.
- We propose a site-aggregated information latent representation to preserve multi-site latent information, and further adopt it to guide dual-task training.
- Under the challenging experimental setting of one-shot segmentation, validated through extensive experiments, our methods significantly outperformed the state-of-the-art methods on one-shot MIS (see Figure 1 (c)).

II. Proposed Method

A. The overall process of the FedSR

In one-shot medical image segmentation (MIS) scenario, the training datasets consist of a labeled image pair, denoted by $(\mathcal{X}_a, \mathcal{Y}_a)$, a large amount of multi-site unlabeled data, denoted by $\{\mathcal{T}^n\}_{n=1}^N = \{\mathcal{X}_{u.n}^m\}_{m=1}^M$, where N denotes the number of sites, M denotes the number of the unlabeled datasets in site n. $\mathcal{X} \in \mathbb{R}^{H \times W \times D}$ is the 3D volume of a medical image and $\mathcal{Y} \in \{0, 1, ...C\}^{H \times W \times D}$ is the the per-voxel label map, where C denote the class number. The goal of one-shot MIS is to predict label map $\hat{\mathcal{Y}}$ with only one labeled image.

The training process of the proposed personalizing **Fed**erated framework guided by **S**ite-aggregated **R**epresentation (**FedSR**) is shown in Fig. 2, consisting

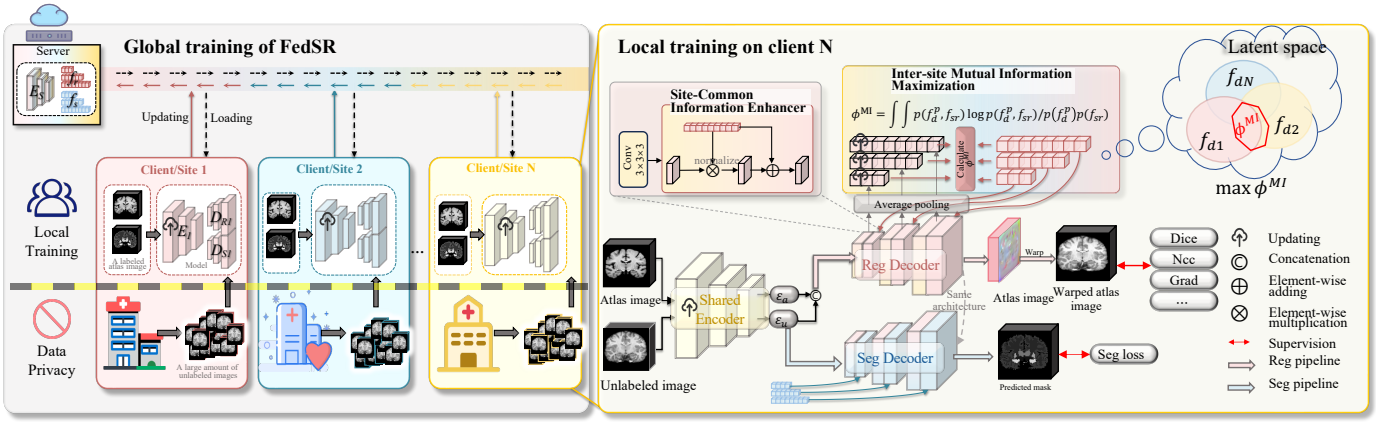


Fig. 2. Overview of the proposed FedSR framework for one-shot MIS.Overall training pipeline of the proposed FedSR. Left: In the global training phase, each client/site uploads selected encoder parameters and site-aggregated features to the server for collaborative model aggregation while preserving data privacy. Right: In local training, inter-site mutual information (MI) maximization is first applied to learn site-invariant representations. Then, a Site-Common Information Enhancer (SCIE) further refines these representations to facilitate effective optimization of both registration and segmentation tasks.

of a shared encoder E, a registration decoder $\mathcal{D}_r$, and a segmentation decoder $\mathcal{D}_s$. We take the entire training process at site n as an example to illustrate our algorithm, with the overall training procedure of FedSR detailed in Algorithm 1. We first randomly pick an unlabeled image $\mathcal{X}_i$ from the training set, and initialize the site-aggregated representation f with a labeled image ($\mathcal{X}_a$, $\mathcal{Y}_a$), Then, we feed $\mathcal{X}_i$ and $\mathcal{X}_a$ into the shared encoder respectively, as

$$\begin{aligned} \varepsilon_a &= E(\omega_e^g; \mathcal{X}_a) \\ \varepsilon_u &= E(\omega_e^g; \mathcal{X}_u^i) \end{aligned} \tag{1}$$

where ω_e^g denotes the parameters of the shared encoder, ε_a and ε_u denote the features of the labeled image and the unlabeled image. Then, concatenating the ε_a and ε_u as input feeds them to $\mathcal{D}_r$ to generate a deformation filed ψ^i, and taking ε_u and f as the input feeds it to $\mathcal{D}_s$ to get a predicted label map $\hat{\mathcal{Y}}$ (The details are presented in Sec. II-C). The above process can be summarized as

$$\begin{aligned} \psi^i &= \mathcal{D}_r([\varepsilon_u, f_a]f) \\ \hat{\mathcal{Y}} &= \mathcal{D}_s(\varepsilon_u|f) \end{aligned} \tag{2}$$

where f defines the prior site-common information which is constantly updated with iteration. Finally, we apply ψ to warp the $\mathcal{Y}_a$ to obtain the Pseudo-label for $\mathcal{X}_i$ as

$$\mathcal{Y}_p = \mathcal{Y}_a \circ \psi \tag{3}$$

Thus, we can get the paired ($\mathcal{X}_u$, $\mathcal{Y}_p$) to train the segmentation model.

After the local training on site i is accomplished, we update the parameter of local encoder $\omega_{e.n}^l$ to the global encoder, and update the local feature representation to the global site-common representation. With such parameter communication and feature communication, the proposed FedSR not only leverages multi-site latent information to guide the segmentation task but also preserves the personal feature without data sharing. In addition, we only use one paired image in the whole training process, alleviating the expensive manual labeling.

B. Omniscient feature extractor

Most previous one-shot MIS paradigms typically require training the whole network from scratch (see Figure 2 (a)) to adapt to a new site-specific dataset resulting time time-consuming, computational complexity, and even overfitting. Although federated learning can learn a global model to adapt to multi-site datasets, our model consists of parameters for two tasks, making it difficult to obtain a highly generalized model.

To this end, we propose to learn an omniscient encoder by federated learning. Specifically, we adopt $\mathcal{T}^n$ as the set of N distributed source domains involved in federated learning. *Each domain only contains images without corresponding labels, which are sampled from a site-specific distribution.* The goal of FedSR is to learn an omniscient global encoder E_e^g using the N distribution source domain, such that it can directly generalize to a completely unseen fine-tune domain and adapt to dual-task head.

The standard federated learning paradigm involves communication between a central server and multiple local clients. In each federated round t, each client k receives identical global model weight ω_e^g from the central server and updates the encoder using their local dataset $\mathcal{T}^n$ for t epochs. The central server then gathers the updated local parameters ω_{en} from all clients and aggregates them to update the global model. This iterative process continues until the global model achieves convergence. In this study, we focus on the widely utilized Federated Averaging algorithm (FedAvg) [13], which aggregates local encoder parameters with weights proportional to the size of each local dataset to update the global encoder, *i.e.*, $\omega_e^g = \frac{1}{N}\sum_{i=1}^{I} \omega_{e.i}^i$.

C. SA representation guided dual-task head

Site-Common Information Enhancer. Given the omniscient feature extractor $E_g(\cdot)$ determined by sec.II-B, we further learn the local dual-task head guided by site-aggregated representation. With the robust multi-scale feature representation $f \in \{f_1, f_2, f_3\}$, $f_1 \in \mathbb{R}^{H/4\times W/4\times D/4}$ is fed into

one $3 \times 3 \times 3$ convolutional kernel generating features f_{d1}. Then, globally pooled f_{d1} (denotes as f_{d1}^{p}) and f_{sr1} compute mutual information and update the global site aggregation representation f_1 by *maximizing inter-site mutual information* to preserve latent information between multiple sites. f_{d1} is further fed into a site-common information enhancer (SCIE) block to generate features with abundant medical information from multi-site institutions guided by SA representation.

Inter-site mutual information maximization. As is well-known, multi-site data appear to have independent data distribution, but there is a certain connection between them. Mathematically, there is a representation pair $\{f_{d1}, f_{d1}\}$ that comes from the segmentation head at site 1 and site 2, respectively. We assume that their prior distribution f^1 are known as

$$\begin{aligned} P(f_{d1}) &= P(f_{d1}|f_1) \\ P(f_{d1} &= P(f_{d1}|f_1) \end{aligned} \tag{4}$$

Then, we can get their joint distribution as

$$P(f_{d1}, f_{d1}) = P(f_{d1}, f_{d1}|f_1)P(f_1) \tag{5}$$

Through the above analysis, we can input the mutual information into the multi-site model learning to filter out task-independent site-specific random noise, and preserve site-invariant information. We calculate the mutual information of f_1 and f_{d1} as

$$\Phi^{MI} = \int\int p(f_{d1}, f_1) \log p(f_{d1}, f_1)/p(f_{d1}, f_1) \tag{6}$$

Then, we calculate $\max \Phi^{MI}$ to update the f. With such a design, we can keep the site-invariant representation during the current site training. The updated site-aggregated representation on local training f and the last iteration f^{i-1} are weighted to average as

$$\hat{f}^i = (1-\alpha)f^{i-1} + \alpha f^i \tag{7}$$

where α denotes the smoothing coefficient parameter. Finally, the f^i updates to the server.

With the learned site-aggregated representation f, we further propose the SCIE block to enhance the site-invariant information to each local training as shown in Fig.2. It is worth noting that the network architecture of the registration header is the same as the segmentation head, except that the inputs are different.

D. Loss Function

Loss for registration. To ensure the smoothness of the deformation field, We first adopt a smoothness loss function $\mathcal{L}_{smooth}$ to constrain the deformation field as:

$$\mathcal{L}_{smooth}(\psi) = \sum_{i \in \psi} ||\nabla \psi(i)||^2. \tag{8}$$

where the i is the position of the voxels in ψ, the $\nabla\psi(i)$ is the gradient of i^{th} position on the deformation field ψ. Then, we simultaneously calculate the similarity between the warped image and the target image y pair as

$$\mathcal{L}_{sim} = l_{smi}(\mathcal{X}_a \circ \psi, \mathcal{X}_u) \tag{9}$$

where $\circ$ denotes the warp operation.

Considering the challenge of anatomical structures of images that can assist in registration, we further adopt segmentation results from the segmentation decoder to boost the optimization. Specifically, we quantify the warped y_a and the predicted segmentation maps $\hat{y}_u^t$ generated from the segmentation decoder using the Dice score [14] as

$$\mathcal{L}_{dice} = l_{dice}(\mathcal{Y}_a \circ \phi, \hat{\mathcal{Y}}_u^m) \tag{10}$$

The total loss function for the registration task can be given by:

$$\begin{aligned} \mathcal{L}_{reg} =& \lambda_1 \mathcal{L}_{smooth} + \lambda_2 \mathcal{L}_{sim} + \lambda_3 \mathcal{L}_{dice} \\ &+ \lambda_4 \max \Phi^{MI} \end{aligned} \tag{11}$$

where λ_1, λ_2, λ_3, and λ_4 are the hyper-parameters to balance the trade-off of three components.

Loss for segmentation. We adopt the same loss to train the segmentation task for both general and medical branches, denoting as $\mathcal{L}_{seg}^g$ and $\mathcal{L}_{seg}^m$, respectively. Specifically, the pseudo labels $y_p^i = y_a \circ \phi$ are generated by the registration task to supervise the segmentation task as

$$\mathcal{L}_{seg}^g = l_{dice}(\mathcal{Y}, \hat{\mathcal{Y}}_u^g) + \lambda \max \Phi^{MI} \tag{12}$$

III. Experiments

Our proposed method is validated on brain MRI image registration and segmentation tasks. Section 3.1 introduces the datasets we used in our experiment. Section 3.2 provides a detailed description of comparative experiments, demonstrating that our method outperforms all state-of-the-art approaches. The ablation experiments in Section 3.3 demonstrate the effectiveness of the mutual information loss and attention mechanism in our method and verify the stability of the proposed FedSR framework.

A. Experiments configurations

1) Data Preparation: Our method leverages OASIS, ABIDE, ADNI, and PPMI as source domains. Subsequently, it is fine-tuned on the OASIS and HCPChild datasets, enabling our FedSR to adapt effectively across multiple sites. The training and testing data partition for these five datasets is detailed in Table II, and the publicly available atlas from [17] is the single labeled template image in training.

ABIDE: A publicly accessible dataset consisting of 1,112 R-fMRI data sets with corresponding structural MRI and phenotypic information from 1,112 participants. The participants include 539 individuals with Autism Spectrum Disorders (ASDs) and 573 age-matched typical controls (TCs), with ages ranging from 7 to 64.

PPMI: An international public dataset for Parkinson's disease progression, consisting of data from 400 recently

TABLE I

QUANTITATIVE COMPARISON OF SEGMENTATION AND REGISTRATION TASKS ACROSS THREE SITE-SPECIFIC DATASETS: ABIDE, PPMI, AND OASIS. OUR METHOD CONSISTENTLY ACHIEVES COMPETITIVE PERFORMANCE ACROSS ALL DATASETS AND BOTH TASKS. THE BEST RESULTS ARE IN **RED BOLD**, WHILE THE SECOND BEST ONES ARE LIGHT RED UNDERLINED

Task	Methods	ABIDE		PPMI		OASIS	
		Dice(%)↑	Hd95↓	Dice(%)↑	Hd95↓	Dice(%)↑	Hd95↓
Segmentation	DeepAtlas [15]	74.40±2.68	3.60±0.88	75.39±2.48	2.81±1.01	75.44±2.65	3.57±0.92
	TBIOneShot [11]	79.65±0.06	2.18±0.26	79.64±7.78	3.28±0.43	80.81±0.90	**2.95±0.52**
	BRBS [16]	83.43±1.25	1.68±0.16	82.42±1.48	2.54±0.33	80.30±1.10	1.68±1.44
	Bi-JROS [12]	82.21±1.22	1.78±0.14	81.47±1.24	2.64±0.29	81.43±0.60	2.74±0.31
	FedAvg [13]	77.64±0.86	1.90±0.15	76.58±1.07	2.98±0.31	76.40±0.66	2.12±0.11
	FedRep [5]	83.70±0.95	1.70±0.12	82.86±1.28	2.68±0.28	82.54±0.82	1.88±0.11
	Ours	**84.01±0.77**	**1.64±0.12**	**82.91±1.03**	**2.51±0.22**	**83.70±0.70**	1.60±0.15
Registration		Dice(%)↑	Ncc↑	Dice(%)↑	Ncc↑	Dice(%)↑	Ncc↑
	DeepAtlas [15]	73.41±2.88	0.201±0.099	69.17±3.01	0.248±0.003	70.56±2.01	0.310±0.007
	TBIOneShot [11]	71.60±2.04	0.285±0.007	71.32±3.19	0.279±0.006	70.41±0.90	0.304±0.003
	BRBS [16]	81.20±1.41	0.301±0.004	80.56±1.44	0.317±0.004	79.50±1.00	0.342±0.006
	Bi-JROS [12]	80.76±1.27	0.370±0.009	79.92±1.26	0.369±0.005	80.00±0.60	0.362±0.004
	FedAvg [13]	80.69±1.18	0.381±0.009	80.22±1.21	**0.382±0.005**	79.56±0.85	**0.352±0.004**
	FedRep [5]	81.03±1.18	0.379±0.009	**80.69±1.17**	0.379±0.005	80.03±0.86	0.371±0.004
	Ours	**81.58±1.08**	**0.381±0.082**	80.61±0.90	0.384±0.041	**80.80±0.90**	0.374±0.003

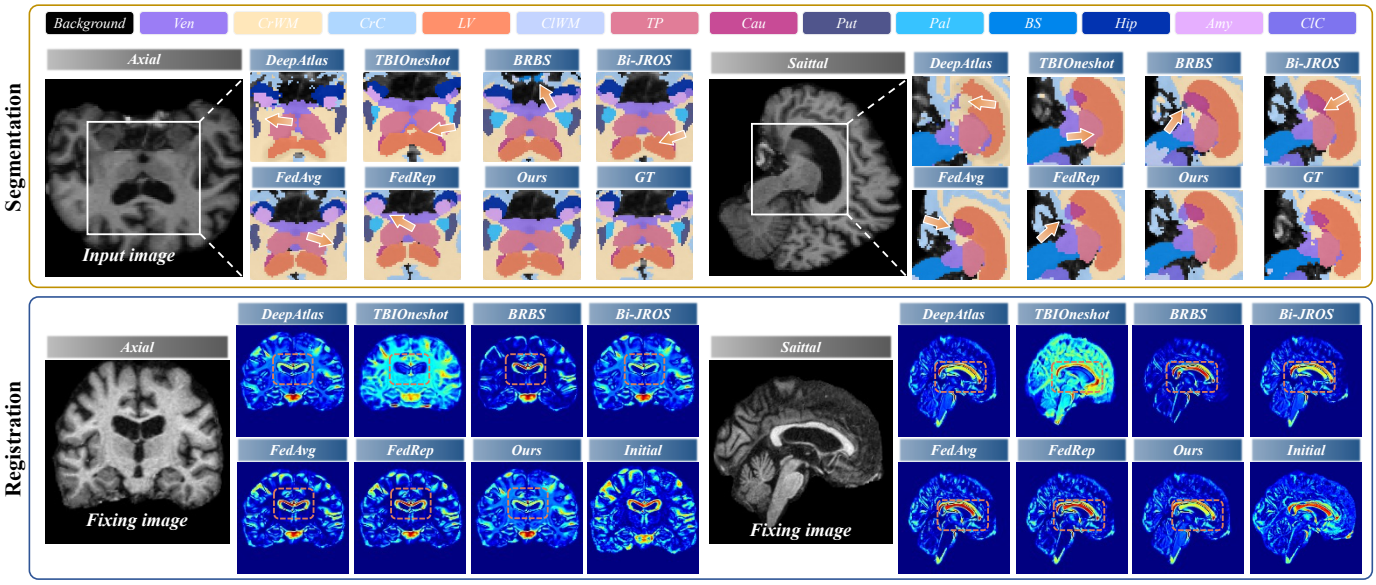


Fig. 3. Qualitative visualization results of segmentation (**top**) and registration (**bottom**) tasks across multiple methods in Axial and Sagittal views, respectively. The orange arrows indicate segmentation errors. In the top part, the orange arrows highlight segmentation errors. In the bottom part, we show the error maps between the registered atlas image and the target, where brighter regions indicate larger registration errors.

diagnosed PD patients and 200 healthy subjects, gathered longitudinally at twenty-one clinical sites.

OASIS: A publicly available dataset consisting of a longitudinal collection of 150 subjects aged 60 to 96, all scanned on the same machine using identical sequences. Each subject underwent two or more scans, spaced at least 1 year apart, resulting in 373 imaging sessions. Subjects were classified based on the Clinical Dementia Rating (CDR) as nondemented or with very mild to mild Alzheimer's disease.

HCPChild: A publicly available dataset, divided into 60 percent and 40 percent for training and testing, respectively. This includes brain images and outcome volumes from the state-of-the-art Siemens 3T Connector imaging system at MGH, and data from naive imaging systems.

TABLE II
TRAINING AND TESTING SPLIT NUMBERS OF ABIDE, ADNI, PPMI, OASIS, AND HCPCHILD DATASETS.

Datasets	ABIDE [19]	PPMI [20]	OASIS [21]	HCP [22]
Total	164	145	111	12
Train	147	102	86	7
Test	17	43	25	5

We performed standard preprocessing steps, including motion correction, NU intensity correction, normalization, and affine normalization by FreeSurfer and FSL tools [18]. Subsequently, all scans were cropped and resized to 128×128×128 with a 1 mm isotropic resolution. For evaluation, anatomical segmentation was conducted on all test MRI scans using FreeSurfer to extract 13 anatomical structures, resulting in a total of 14 labels. Specifically, the 13 anatomical structures include: 3rd/4th Ventricle (Ven), Cerebral White Matter (CrWM), Cerebral Cortex (CrC), Lateral Ventricle (LV), Cerebellum White Matter (CIWM), Thalamus Proper (TP), Caudate (Cau), Putamen (Put), Pallidum (Pal), Brain Stem (BS), Hippocampus (Hip), Amygdala (Amy), and Cerebellum Cortex (ClC). These structures cover both cortical and subcortical regions and are representative of diverse anatomical variability across individuals and sites.

2) Evaluation Metrics: In the evaluation of segmentation tasks, we employ Dice similarity coefficient (Dice), Jaccard index, Average Volume Difference (AVD), and Average Symmetric Surface Distance (ASD) as primary metrics. The Dice similarity coefficient is a widely utilized metric for spatial overlap assessment, primarily used to calculate the similarity between two binary images, making it particularly suitable for evaluating the alignment of anatomical structures. The Jaccard index computes the ratio of the intersection to the union of the predicted and actual regions, with higher values indicating greater similarity. The AVD metric measures the volumetric difference between the predicted and actual regions, calculated as the absolute difference between the predicted and actual volumes, averaged over all voxels; lower AVD values indicate closer agreement between predicted and actual volumes. The ASD metric calculates the average distance between the predicted and actual boundaries, with lower values signifying higher precision in boundary segmentation.

3) Implementation Details: Our network architecture comprises a shared encoder and two task-specific decoders. The training process is divided into two phases: (a) The encoder training phase, in which we employ the federated learning FedAvg algorithm to train an encoder that performs robustly across multi-site data; (b) The fine-tuning phase, where the decoder parameters are fixed, and separate loss function searches are conducted for the registration and segmentation decoders to further optimize performance. Notably, we achieve optimal alignment for both registration and segmentation tasks through bi-level learning.

Our framework is implemented in PyTorch and runs on a Tesla V100 with 32GB of RAM. During training, the Adam [23] optimizer is utilized with a learning rate set to 1×10^{-4}. Additionally, the trade-off factors λ_{sim} = 5, λ_{smo} = 7.5, and λ_{ce} = 1 are set. In the multi-site training process, encoder parameters are shared across sites and updated to the global server model, while decoders remain fixed. This approach ensures consistency and generalizability across multi-tasks and multi-site data.

TABLE III
QUANTITATIVE COMPARISON AMONG VARIOUS METHODS FOR REGISTRATION AND SEGMENTATION TASKS ON HCPCHILD DATASET. THE TOP-RANKED METHOD IS HIGHLIGHTED IN **RED BOLDED** FORM.

Methods	Segmentation	Registration	
	Dice (%)	Dice (%)	Ncc
UResNet [25]	81.9 ± 0.4	79.6 ± 0.8	0.362 ± 0.004
BRBS [26]	81.7 ± 0.9	80.3 ± 1.0	0.351 ± 0.009
Bi-JROS [27]	81.7 ± 0.8	80.6 ± 0.9	0.367 ± 0.009
Ours	**83.7 ± 0.7**	**80.8 ± 0.9**	**0.370 ± 0.006**

B. Comparison Experiments

Comparison settings. To comprehensively evaluate the effectiveness of our proposed FedSR framework, we conduct extensive comparisons against a variety of representative methods for medical image segmentation and registration under the one-shot setting, as shown in Table I. Specifically, the compared baselines include: (1) classic deep learning-based methods, such as DeepAtlas [15], TBIOneShot [24], BRBS [10], and Bi-JROS [12], which serve as strong fully centralized baselines; (2) representative federated learning frameworks, including FedAvg [13] and FedRep [5], which are adapted to the one-shot scenario by distributing the labeled atlas and unlabeled samples across sites. For fairness, all methods are trained using only one labeled image per site (the atlas), while all other samples are considered unlabeled and used in an unsupervised or weakly-supervised manner.

Comparison results. Table I summarizes the quantitative results of FedSR versus the current state-of-the-art one-shot medical image models on three public brain datasets—OASIS, PPMI, and ABIDE. Across almost all sites, FedSR attains the highest Dice score and the lowest 95th-percentile Hausdorff distance (Hd95), indicating more accurate boundary delineation. FedSR likewise delivers the top Dice score and the best (highest) normalized cross-correlation (NCC), reflecting superior anatomical alignment. Benefiting from the multi-site federated training, our model has learned more generalized features that account for the variations in data distribution across different sites. This approach ensures that our model is not only highly effective but also robust across diverse data environments, enhancing its applicability in real-world clinical settings. Notably, since the above three datasets consist exclusively of adult brain MRIs, we further conduct experiments on the HCPChild dataset to evaluate the generalizability of our method in pediatric populations. As shown in Table III, FedSR still achieves competitive or superior performance, confirming the robustness of our approach in dealing with significant anatomical and distributional shifts.

TABLE IV
THE ABLATION STUDY DEMONSTRATES THE CONTRIBUTION OF OUR INNOVATIONS. THE TOP-RANKED METHOD IS HIGHLIGHTED IN **RED BOLDED** FORM.

Baseline	FL	MIL	SCIE	S-Dice(%)	R-Dice(%)
✓	✗	✗	✗	80.0±0.8	78.1±1.0
✓	✓	✗	✗	81.7±0.9	79.8±0.9
✓	✓	✓	✗	82.1±0.8	80.1±0.9
✓	✓	✓	✓	**83.7±0.7**	**80.8±0.9**

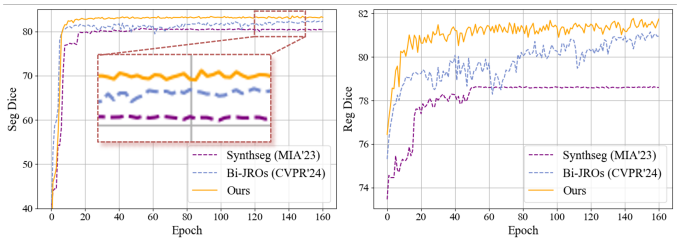


Fig. 4. Training process of segmentation (left) and registration (right) Dice scores using different encoder backbones: Synthseg, Bi-JROS, and our proposed method. Our approasch consistently achieves superior performance acrooss the training process.

We further show the segmentation and registration results of six representative methods compared to our approach on three medical datasets—ABIDE,PPMI and OASIS,showcasing the results from three different perspectives.The corresponding visualizations are summarized in Figure 3. In the segmentation task, our method achieves the highest overlap with the Ground Truth (GT) in large brain white matter structures and small lateral ventricle tissues. Current state-of-the-art methods perform segmentation by training a shared encoder; however, their adaptation to unseen data is limited due to learning from data specific to a single site. Our proposed multi-site learning framework with a shared encoder demonstrates satisfactory adaptability, enhancing the segmentation performance of our approach. In the segmentation and registration tasks, we achieve the best alignment of regions compared to other methods. Previous methods have resulted in some significant misalignments due to updating the parameters of the decoders through only a single loss function. Considering this, we introduce loss search to adjust the weights in the loss function, aiming to find the optimal loss function, which further improves alignment accuracy.

The experimental results reveal that, thanks to the parallel processing capability and model generalizability of the federated learning framework, our method outperforms others in terms of processing speed. In summary, we can draw the following conclusions: *Firstly*, most one-shot MIS methods outperform standalone registration or segmentation approaches, highlighting the importance of implementing joint registration and segmentation. *Secondly*, the proposed bilevel optimization framework, which utilizes a loss search strategy, is crucial for effectively optimizing the parameters of task-specific decoders, thus improving the performance of both registration and segmentation tasks. *Thirdly*, the use of a pre-trained, universal plug-and-play encoder, as part of our federated learning setup, enables efficient multi-site training while enhancing the model's generalization ability across diverse data distributions. Although the federated setup results in slightly longer training times, our method outperforms others in terms of registration accuracy, segmentation performance, and computational efficiency.

C. Ablation Experiments

Effectiveness of federated learning.

We first validated the effectiveness of our Federated Learning (FL) framework. To assess its necessity, we conducted a variant where data from all sites were centrally merged to pre-train the encoder, which was then fixed during decoder training. As shown in Table IV, this non-federated setting yielded an S-Dice of 80.0% and R-Dice of 78.1%, serving as our baseline. In contrast, FedSR, by continuously updating the encoder via inter-site parameter sharing, significantly improved performance, boosting S-Dice to 83.7% and R-Dice to 80.8%. To further assess the encoder quality, we compared it with two widely used alternatives: SynthSeg [28], a generative model capable of segmenting brain MRIs of varying contrast and resolution without retraining, and Bi-JROS [29], which uses an encoder pre-trained on 295 MR scans from multi-site datasets. The training curves are visualized in Figure 4, showing that our method consistently outperforms the others across the training process. Therefore, federated learning is pivotal for learning a powerful and generalizable encoder from decentralized data, ensuring robustness, stability, and cross-site transferability.

The Effectiveness of inter-site mutual information maximization. We further investigated the impact of the inter-site Mutual Information Loss (MIL), designed to enhance feature-level consistency by maximizing shared representations between sites. As shown in Table IV, removing MIL led to a drop in S-Dice from 81.7% to 80.0% and R-Dice from 79.8% to 78.1%, indicating its essential role in improving cross-site alignment. Without MIL, the model struggled to extract site-common feature distributions, amplifying the effects of local heterogeneity and impairing generalization. By encouraging mutual feature sharing, MIL facilitated better inter-site invariant representation learning, effectively addressing distribution shifts and maintaining stable performance across domains.

The Effectiveness of SCIE block. Finally, we evaluated the impact of the proposed Site-Common Information Enhancer (SCIE) block, designed leverage site-invariant features shared across domains. To assess its effect, we removed SCIE and compared performance with and without shared guidance. As shown in Table IV, disabling SCIE led to a drop from 83.7% to 82.1% in S-Dice and from 80.8% to 80.1% in R-Dice, confirming its effectiveness. These results highlight that incorporating consistent inter-site features is vital for generalization. By integrating site-common representations into both tasks, SCIE helps the model focus on reliable structural cues, yielding more stable and accurate predictions in complex multi-site scenarios.

IV. Conclusion

In this paper, we propose a personalizing federated guided by site-aggregated representation for multi-site one-shot medical image segmentation, exploiting the site-invariant latent information to boost segmentation performance. Specifically, we propose to learn an omniscient encoder by federated learning, which can not only model the data distribution between multi-site datasets but also adapt the multi-task in an efficient way. With the learned robust representation, we further propose to learn consistency features between multi-site data by mutual information maximization, and then adopt such latent site-common representation to guide the personalized decoder modeling. Extensive experiments conducted on two MIS tasks demonstrate that the proposed FedSR outperforms state-of-the-art one-shot MIS methods and FL methods.

References

[1] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and trends® in machine learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[2] J. Chen, B. Ma, H. Cui, and Y. Xia, "Think twice before selection: Federated evidential active learning for medical image analysis with domain shifts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 439–11 449.

[3] M. Poggi and F. Tosi, "Federated online adaptation for deep stereo," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 20 165–20 175.

[4] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10 713–10 722.

[5] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Exploiting shared representations for personalized federated learning," in *International conference on machine learning*. PMLR, 2021, pp. 2089–2099.

[6] J. Dong, D. Zhang, Y. Cong, W. Cong, H. Ding, and D. Dai, "Federated incremental semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 3934–3943.

[7] J. Miao, Z. Yang, L. Fan, and Y. Yang, "Fedseg: Class-heterogeneous federated learning for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 8042–8052.

[8] X. Xu, H. H. Deng, J. Gateno, and P. Yan, "Federated multi-organ segmentation with inconsistent labels," *IEEE transactions on medical imaging*, vol. 42, no. 10, pp. 2948–2960, 2023.

[9] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, "Joint learning of motion estimation and segmentation for cardiac mr image sequences," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 472–480.

[10] Y. He, R. Ge, X. Qi, Y. Chen, J. Wu, J.-L. Coatrieux, G. Yang, and S. Li, "Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 2588–2601, 2022.

[11] X. Zhao, Z. Shen, D. Chen, S. Wang, Z. Zhuang, Q. Wang, and L. Zhang, "One-shot traumatic brain segmentation with adversarial training and uncertainty rectification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 120–129.

[12] X. Fan, X. Wang, J. Gao, J. Wang, Z. Luo, and R. Liu, "Bi-level learning of task-specific decoders for joint registration and one-shot medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 726–11 735.

[13] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[14] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[15] Z. Xu and M. Niethammer, "Deepatlas: Joint semi-supervised learning of image registration and segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, pp. 420–429.

[16] D.-K. Ngo, M.-T. Tran, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "Multi-task learning for small brain tumor segmentation from mri," *Applied Sciences*, vol. 10, no. 21, p. 7790, 2020.

[17] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[18] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in fsl," *Neuroimage*, vol. 45, no. 1, pp. S173–S186, 2009.

[19] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.

[20] M. MacKay, P. Vicol, J. Lorraine, D. Duvenaud, and R. Grosse, "Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions," *arXiv preprint arXiv:1903.03088*, 2019.

[21] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[22] J. S. Elam, M. F. Glasser, M. P. Harms, S. N. Sotiropoulos, J. L. Andersson, G. C. Burgess, S. W. Curtiss, R. Oostenveld, L. J. Larson-Prior, J.-M. Schoffelen *et al.*, "The human connectome project: a retrospective," *NeuroImage*, vol. 244, p. 118543, 2021.

[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[24] X. Zhao, Z. Shen, D. Chen, S. Wang, Z. Zhuang, Q. Wang, and L. Zhang, "One-shot traumatic brain segmentation with adversarial training and uncertainty rectification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer Nature Switzerland, 2023, pp. 120–129.

[25] T. Estienne, M. Vakalopoulou, S. Christodoulidis, E. Battistela, M. Lerousseau, A. Carre, G. Klausner, R. Sun, C. Robert, S. Mougiakakou *et al.*, "U-resnet: Ultimate coupling of registration and segmentation with deep nets," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*. Springer, 2019, pp. 310–319.

[26] Y. He, R. Ge, X. Qi, Y. Chen, J. Wu, J.-L. Coatrieux, G. Yang, and S. Li, "Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 2588–2601, 2022.

[27] X. Fan, X. Wang, J. Gao, J. Wang, Z. Luo, and R. Liu, "Bi-level learning of task-specific decoders for joint registration and one-shot medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 726–11 735.

[28] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, J. E. Iglesias *et al.*, "Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining," *Medical image analysis*, vol. 86, p. 102789, 2023.

[29] X. Fan, X. Wang, J. Wang, Z. Luo, and R. Liu, "Bi-level learning of task-specific decoders for joint registration and one-shot medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, p. online.