基于不确定性协作图和双层聚合机制的个性化联邦医学图像分割

杜晓刚[1,2],魏 征[1,2],雷 涛[1,2*],刘统飞[1,2],王营博[1,2]

(1. 陕西科技大学电子信息与人工智能学院,陕西西安 710021;2. 陕西科技大学陕西省人工智能联合实验室,陕西西安 710021)

摘 要: 个性化联邦学习作为一种分布式机器学习范式,能够在不泄露客户端原始数据的前提下,实现多客户端模型的协同训练,已成为医学影像智能处理与分析领域的研究热点。然而,现有的个性化联邦学习方法主要通过全局协同或聚类分组协同来建模客户端关系,其整体协同粒度粗且灵活性不足。近年来,基于协作图的个性化联邦学习方法通过图结构建模客户端之间的协作关系,能够实现较细粒度的动态协同,有效缓解了全局协同与聚类协同的固有缺陷。但是,其仅以数据量和模型相似度来更新客户端协作图,未考虑医学图像分割任务中固有的高不确定性,导致其易受高不确定性客户端的影响而使得分割精度下降。为了解决该问题,提出了一种基于不确定性协作图和双层聚合机制的个性化联邦医学图像分割方法。该方法的核心优势主要包括两个方面:一是设计了不确定性惩罚项并将其引入服务器端目标函数中来优化协作图更新过程,生成适配医学图像分割任务的不确定性协作图,通过动态调整各个客户端之间的协作权重并避免高噪声参数混入导致知识污染,有效保障了协同训练的稳定性。二是提出了基于不确定性协作图的双层聚合机制。第一层聚合实现基于协作图的客户端局部协同,挖掘相似客户端之间的有效知识;第二层聚合通过融合局部协同结果与全局模型,来平衡全局模型通用性与本地客户端的个性化需求,实现了高质量知识的有效传递,提升了客户端本地模型的分割性能。为了全面验证所提方法的有效性与鲁棒性,在四个公开的息肉分割数据集上开展了大量的实验。实验结果表明:与其他先进的医学图像分割方法相比,提出的方法在多个客户端测试数据上取得了更优异的分割性能,为临床医疗场景下的个性化联邦医学图像分割提供了一种新的技术方案。

关键词: 个性化联邦学习;医学图像分割;协作图;不确定性;双层聚合;证据理论

基金项目: 国家自然科学基金(No.62271296, No.62201334);西安市中青年科技创新领军人才项目(No.25ZQRC00019);陕西省创新能力支持计划(No.2025RS-CXTD-012);陕西省教育厅青年创新团队科研计划(No.23JP022,No.23JP014,No.25JP023)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2026)03-1078-16

电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250856

Personalized Federated Medical Image Segmentation Based on Uncertain Collaboration Graph and Dual-Layer Aggregation Mechanism

DU Xiaogang[1,2], WEI Zheng[1,2], LEI Tao[1,2*], LIU Tongfei[1,2], WANG Yingbo[1,2]

(1. *School of Electronic Information and Artificial Intelligence*, *Shaanxi University of Science and Technology*, *Xi'an*, *Shaanxi* 710021, *China*;
2. *Shaanxi Joint Laboratory of Artificial Intelligence*, *Shaanxi University of Science and Technology*, *Xi'an*, *Shaanxi* 710021, *China*)

Abstract: As a distributed machine learning paradigm, personalized federated learning can realize the collaborative training of multi-client models without leaking the original data of the client, and has become a research hotspot in the field of medical image intelligent processing and analysis. However, the existing personalized federated learning methods mainly model client relationships through global collaboration or clustering group collaboration, and their overall collaboration granularity is coarse and lack of flexibility. In recent years, personalized federated learning methods based on collaboration graph model the collaboration relationship between clients by graph structure, which can achieve fine-grained dynamic collaboration and effectively alleviate the inherent defects of global collaboration and clustering collaboration. However, it only uses the amount of data and model similarity to update the client collaboration graph, and does not consider the inherent high uncertainty in the medical image segmentation task, which makes it vulnerable to high uncertainty clients and reduces the segmentation accuracy. In order to solve this problem, we propose a personalized federated medical image segmentation method based on uncertain collaboration graph and dual-layer aggregation mechanism in this paper. The core advantages of this method mainly include two aspects. Firstly, an uncertainty penalty term is designed and introduced into the server-side

收稿日期:2025-09-28;录用日期:2026-02-25;责任编辑:覃怀银

*通信作者:雷涛

objective function to optimize the updating process of the collaboration graph, and generate an uncertain collaboration graph suitable for the medical image segmentation task. By dynamically adjusting the collaboration weights between each client and avoiding knowledge pollution caused by high noise parameters, the stability of collaborative training is effectively guaranteed. Secondly, a dual-layer aggregation mechanism based on uncertain collaboration graph is proposed. The first layer of aggregation realizes the local collaboration of clients based on collaboration graph, and mines the effective knowledge between similar clients. The second layer of aggregation balances the generality of the global model and the personalized requirements of the local client by fusing the local collaborative results and the global model, realizes the effective transfer of high-quality knowledge, and improves the segmentation performance of the client-side local model. In order to fully verify the effectiveness and robustness of the proposed method, a large number of experiments are carried out on four public polyp segmentation datasets. The experimental results show that compared with other advanced medical image segmentation methods, the proposed method achieves better segmentation performance on multiple client test data, which provides a new technical solution for personalized federal medical image segmentation in clinical medical scenarios.

Keywords: personalized federated learning; medical image segmentation; collaboration graph; uncertainty; dual-layer aggregation; evidence theory

Foundation Item(s): National Natural Science Foundation of China (No.62271296, No.62201334); Young Science and Technology Innovation Leading Talents Program of Xi'an City (No. 25ZQRC00019); Innovation Capability Support Plan Project in Shaanxi Province (No. 2025RS-CXTD-012); Scientific Research Program Funded by Shaanxi Provincial Education Department (No.23JP022, No.23JP014, No.25JP023)

0 引言

医学图像分割是从CT、MRI等医学图像中提取器官、病灶等感兴趣区域轮廓的过程,在病情诊断、放射治疗、图像引导手术等临床实践中具有重要的应用价值。

随着深度学习技术的迅速发展,在医学图像分割领域涌现出U-Net[1]、U-Net++[2]、TransUNet[3]、Swin-UNet[4]等很多代表性的集中式医学图像分割方法。然而,集中式医学图像分割方法面临以下来自临床实践的挑战[5-6]:首先,医学图像数据天然分散于各级医疗机构,使得集中式学习难以充分利用这些分散数据进行模型训练;其次,由于医学图像数据包含患者的隐私信息,而集中式分割方法需要将数据汇聚到单一中心,导致数据传输与存储过程中面临很大的隐私泄露风险。

联邦学习[7]作为一种新兴的分布式机器学习方法,支持多个站点利用私有数据训练本地模型,然后通过在服务器上聚合所有本地模型参数得到全局模型。整个过程无需传输患者数据,同时,在保障数据隐私安全的前提下能够充分利用各客户端的数据,从而有效提升模型的分割性能[8]。但是,由于不同机构在成像设备、扫描参数、图像分辨率及标注规范上有所不同,使得各客户端的数据分布具有较大差异,导致全局模型在各客户端上的泛化能力欠佳[9-11]。为了解决该问题,个性化联邦学习技术应运而生[12-13]。

个性化联邦学习是在联邦学习框架的基础上,允许参与的客户端在保护数据隐私的前提下,利用全局模型的知识训练出更适合本地数据分布和任务需求的局部个性化模型[14]。然而,为了提升客户端的局部性能,一些现有的个性化联邦学习方法过于聚焦本地数据适配[15-17],在联邦学习时没有进行多中心关系建模并利用多中心关系进行深度协作,导致共性知识挖掘不充分、个性需求适配不到位,最终造成全局模型性能提升受限且各中心分割精度参差不齐,跨中心泛化能力与鲁棒性下降。为了解决该问题,一些学者在医学图像分类任务上通过利用图结构来挖掘客户端之间的潜在协作关系[18-19]。但是,在医学图像分割领域,基于协作图的个性化联邦学习方法还未被探索。同时,医学图像分割任务的高不确定性容易导致协作图更新时存在客户端关联性误判,使得协作权重频繁调整而削弱关系稳定性,最终导致协作图无法传递有效知识,反而因放大不确定性而背离协同优化目标。

针对以上问题,本文提出了一种基于不确定性协作图和双层聚合机制的个性化联邦医学图像分割方法,简称GraphFedSeg。本文的主要贡献包括以下三点:

(1)本文提出的GraphFedSeg首次将协作图引入医学图像分割任务中来建模客户端之间的协作关系,为挖掘个性化联邦医学图像分割方法中的客户端协作关系提供了一种新思路。实验结果表明GraphFedSeg的分割性能优于最先进的个性化联邦学习方法。

(2)设计了一种基于不确定性预测的客户端协作图构建方法。通过量化客户端模型预测分割的不确定性,在协作图更新过程中创新性地引入不确定性惩罚项,生成不确定性协作图(Uncertainty-aware Collabo-

ration Graph，UCG），并动态调整客户端之间的协作权重，从而更精准地反映客户端间的协作关系，为个性化联邦医学图像分割方法实现客户端之间的有效协作奠定了可靠基础。

（3）提出了一种基于不确定性协作图的双层聚合机制（Dual-layer Aggregation Mechanism，DAM）。该机制采用数据基础聚合和邻域增强聚合的双层结构。数据基础聚合保留了基于数据量权重的全局一致性，确保数据规模大的客户端主导基础模型的稳定性，邻域增强聚合通过不确定性协作图强化相似客户端的邻域协作以适应数据异质性，从而在全局一致性与局部个性化间取得了良好的平衡，提升了模型在不同客户端上的分割性能。

1 相关工作

1.1 联邦医学图像分割

联邦学习能够在不共享原始数据的前提下实现多客户端的协同分布式训练，并能得到在所有客户端上都表现良好的全局模型。FedAvg[7]作为联邦学习的基础算法，它的实现前提是客户端数据满足独立同分布。然而，在医学图像分割场景中，客户端成像设备差异[20-23]、数据标注类别差异[24-26]、标注准确性差异[27-28]等问题会导致各客户端的数据通常是非独立同分布的，具有较为显著的数据异质性。

针对因客户端成像设备差异导致数据异质性的问题，学者通过数据增强[20]、改进参数聚合方式[21]、动态调整聚合[22]、抑制过拟合客户端[23]等手段提出了许多联邦医学图像分割方法。具体地，FedDG[20]提出连续频率空间插值方法，通过频率域混合不同客户端的数据特征，模拟跨域数据分布，提升全局模型在其他客户端上的泛化能力。FedBN[21]通过客户端独立计算和更新批归一化层参数，并在服务器端仅聚合非归一化层参数，最终在多个异构数据集上提升了模型的分割准确率。FedEvi[22]通过分解总体不确定性来动态评估客户端泛化能力与数据可靠性，进而调整聚合权重，改善了模型的分割性能。FedCLAM[23]通过客户端自适应动量与阻尼模块动态调整聚合权重，同时以前景强度匹配损失引导模型学习通用解剖特征并规避设备特异性，有效提升了多中心场景下的模型分割精度与鲁棒性。

针对客户端之间的标注类别差异导致模型训练出现语义混淆与偏差的问题，学者通过改善特征提取与增强策略提出了多种针对性的解决方案[24-26]。例如：UFPS[24]通过统一标签学习生成去噪伪标签、优化高质量本地模型聚合权重并动态调整损失权重，然后借助稀疏统一锐度感知最小化和强数据增强来优化全局方向、通过梯度掩码加速并补充全局特征。FedMENU[25]通过多编码网络将多器官特征学习分解为独立子任务，客户端仅微调所标注器官对应的子编码器及共享解码器以规避未标注器官干扰，同时通过辅助通用解码器正则化训练，增强器官特征的跨领域不变性与辨识度。FUNAvg[26]提出了一种联邦不确定性加权平均方法，通过联合学习所有标注结构，并利用贝叶斯技术挖掘未标注结构的信息。这些方法在联邦学习框架下都实现了对本地未知多器官的精准分割。

针对不同客户端的标注准确性差异导致模型对同类目标的像素级语义理解出现偏差的问题，学者从模型聚合策略优化和噪声处理方面提出了相应的解决方法[27-28]。Wu 等人[27]提出标注质量感知的联邦学习聚合策略，通过轮廓演化模型来建模噪声，结合质量感知聚合策略来动态调整客户端权重。Xiang 等人[28]提出动态权重调整与噪声校正框架，通过量化客户端标注质量来区分可靠与低质量数据源，提升完整标注客户端权重并校正不完整标注噪声，结合数据量与标注质量来动态平衡客户端贡献。这些方法有效改善了因标注准确性差异引起的分割性能下降的问题。

然而，以上这些方法均专注于训练全局模型，没有充分利用每个客户端的数据特点，忽略了每个客户端的本地需求。因此，在改善全局模型的分割性能时，考虑各客户端的数据特点和个性需求，研究个性化联邦医学图像分割方法显得尤为重要。

1.2 个性化联邦医学图像分割

个性化联邦学习将优化目标从全局统一模型转向全局知识共享和本地个性化适配，让每个客户端既能利用全局协作的共性知识提升基础性能，又能根据本地数据特性与任务需求来训练模型，有效实现全局协作与本地个性的平衡。

针对因数据异构性导致客户端数据分布漂移的问题，学者在个性化联邦医学图像分割方法中提出了不同的技术手段，如局部校准[29]、自注意力网络[30]、内外个性化协同设计[31]、正样本局部特征增强[32]等。具体地，Wang 等人[29]提出了基于局部校准的个性化联邦学习方法 FedLC。FedLC 通过挖掘特征层和预测层的跨站点不一致性，实现了局部模型的精确校准，平衡了全局知识共享与局部个性化需求。FedDP[30]利用自注意力网络的长程依赖建模能力来处理数据异质性，通过特征层面的长程依赖个性化和预测层面的不一致性引导校准，协同提升了联邦学习框架下的医学图像分割性能。Jiang 等人[31]通过内部个性化和外部个性化的协同设计，同时提升了参与训练的本地

客户端的分割性能与未参与训练的外部客户端的泛化能力。pFLFE[32]通过利用正样本的局部特征增强模块提升特征分离度,并利用这些特征来学习分割掩码,从而实现高性能医学图像分割。以上方法虽然有效抑制了客户端的数据分布漂移,但这些方法都专注于训练适应本地数据分布的个性化模型,忽略了学习客户端之间的协作关系。

为了解决该问题,集群联邦学习(Clustered Federated Learning,CFL)[33]根据数据特征或模型更新相似性将客户端划分为独立集群,每个集群训练专用模型以适应局部数据特性,同时结合差分隐私与安全聚合技术确保隐私安全。pFedHN[34]通过超网络生成客户端个性化模型参数,结合嵌入向量传输和动态聚合实现了高效的个性化学习。为了进一步挖掘客户端之间的潜在关系,有些学者利用图结构来建模客户端之间的关系。例如:Ye 等人[18]通过推断协作图建模客户端之间的协作关系,并通过平衡本地任务损失与模型相似度正则化来优化个性化模型,避免本地数据过拟合或盲目协作;Zhou 等人[19]通过在服务器端部署图注意力网络(graph attention networks)来捕获客户端之间的潜在关系,图网络的节点为参与联邦学习的各个客户端,节点间的信息交换主要通过将节点特征投影到高维空间,并计算客户端对的原始关联,再通过多注意力头融合得到最终权重分配矩阵来实现。然而,这些方法没有考虑医学图像分割任务中对边界区域、模糊病灶的预测不确定性较高的问题。

为了解决以上问题,本文提出了 GraphFedSeg。与以上方法不同的是,GraphFedSeg 在协作图更新过程中创新性引入基于证据理论的不确定性惩罚项,生成不确定性协作图并动态调整协作权重以过滤高噪声客户端干扰;另外,在 GraphFedSeg 中,通过设计双层聚合机制,在保留全局模型稳定性的同时,利用不确定性协作图注入适配性强的局部分割领域知识,最终实现全局分割性能与本地个性化需求的协同提升。

2　方法

2.1　GraphFedSeg 的整体结构

本文提出的 GraphFedSeg 的主要结构如图 1 所示,整体架构分为客户端和服务器端。在客户端,每个客户端在本地数据集上进行训练,然后上传模型参数至服务器端完成模型聚合。具体地,在 GraphFedSeg 中,客户端需要进行本地模型训练和不确定性计算,服务器端需要进行不确定性协作图生成和基于不确定性协作图的双层聚合。

不确定性协作图是 GraphFedSeg 技术核心。首先,通过证据理论量化客户端预测不确定性;其次,利用数据量占比、模型相似度与不确定性等因素来构造用于优化协作图的目标函数;最后,生成能准确捕捉高价值协作关系(高数据量、高模型相似度且低不确定性客户端间强关联)的协作图邻接矩阵,并同时削弱高噪声或异常客户端的干扰,为后续聚合提供可靠协作依据。

基于不确定性协作图的双层聚合机制采用数据基础聚合与邻域增强聚合相结合的层次化设计,在保留数据量主导的全局一致性基础上,充分利用不确定性协作图捕捉到的局部相似性和有效协作关系,提升模型在客户端的分割性能。

2.2　不确定性协作图

基于基础协作图的个性化联邦学习方法仅应用于图像分类任务,且基础协作图仅依模型相似度与数据量更新。医学图像分割的高不确定性会导致基础协作图对医学图像分割任务中的不可靠协作存在无差别容忍,易让噪声扩散。为了解决该问题,本文提出了一种基于不确定性预测的协作图构建方法。通过引入基于证据理论的不确定性惩罚项来构建差异度量和动态调整协作权重,既延续核心协作逻辑,又让协作图能分辨不确定性,从而实现精准协作与噪声过滤,更加有效地适配医学图像分割场景。

2.2.1　服务器端优化

在个性化联邦学习系统中,定义 K 个客户端 $\{c_1, c_2, c_3, \cdots, c_k\}$,客户端 c_i 的本地数据集为 D_i,数据规模为 $n_i=|D_i|$,θ_i 表示客户端 c_i 的本地分割模型的全部可训练参数集合。本文使用协作图 $\boldsymbol{G}=(V, \boldsymbol{W})$ 来刻画各客户端间的协作潜力。其中,V 为协作图的节点集,即 $V=\{c_1, c_2, c_3, \cdots, c_k\}$,$V$ 中的每个节点对应一个客户端;$\boldsymbol{W}\in\mathbb{R}^{k\times k}$ 为协作图 $\boldsymbol{G}$ 的邻接矩阵,矩阵元素 $W_{i,j}$ 表示客户端 c_i 与 c_j 的协作强度,反映 c_i 从 c_j 获取知识的权重。基于协作图的个性化联邦学习以最小化全局损失和最大化模型协作效益为优化目标[18],在服务器端构造协作图的目标函数如式(1)所示:

$$\min_{\{W_{i,j}\}_j}\sum_j\left(W_{i,j}-p_j\right)^2-\alpha\sum_j W_{i,j}\cos\left(\theta_i,\theta_j\right)$$
$$\text{s.t.}\ \sum_{j=1}^{k}W_{i,j}=1,\forall i;\qquad W_{i,j}\geqslant 0,\forall i,j \tag{1}$$

其中:p_j 表示数据集的数据量占比,即 $p_j=n_j\Big/\sum_{i=1}^{k}n_i$,衡量客户端 c_j 的数据贡献度;α 是一个超参数,$\alpha=K\times 0.08$;模型相似性通过余弦相似度来计算,即 $\cos\left(\theta_i,\theta_j\right)=\theta_i\cdot\theta_j\big/\|\theta_i\|\cdot\|\theta_j\|$,用来度量客户端模型的知识一致性;矩阵 $\boldsymbol{W}$ 的每行元素和为 1,即

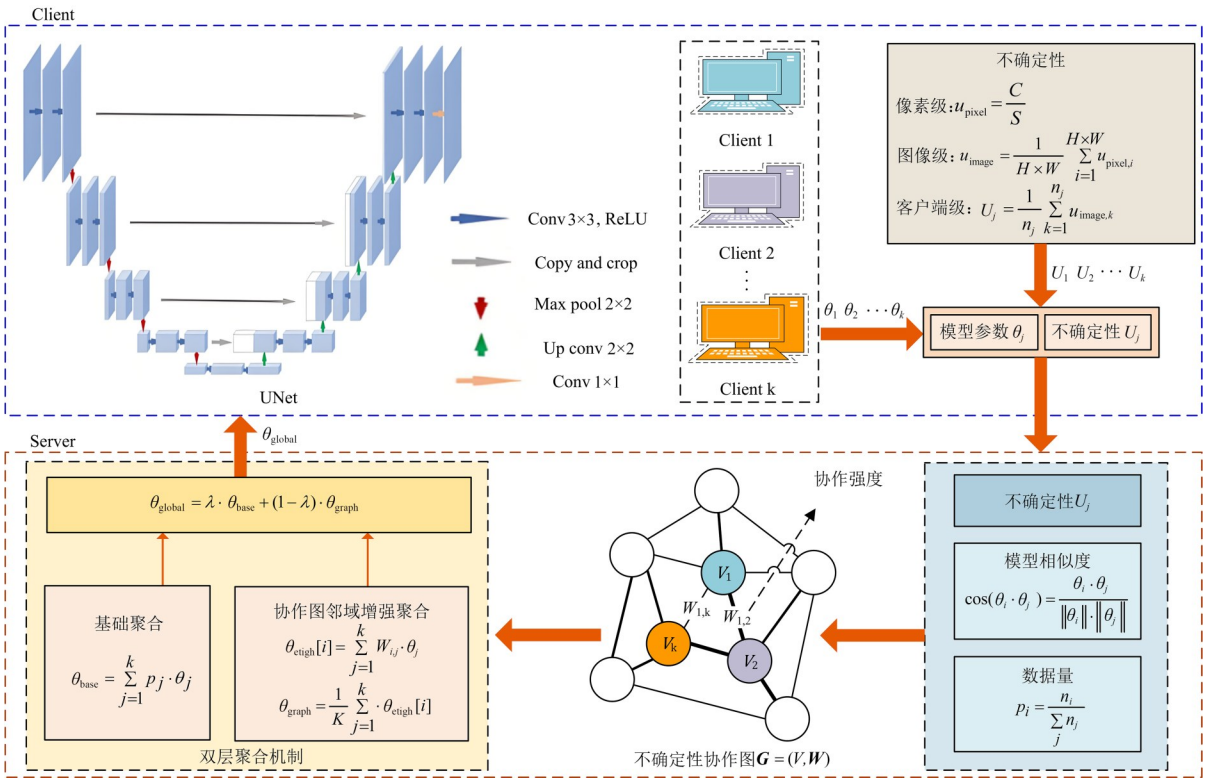


图1 GraphFedSeg的整体框架

Figure 1 The overall framework of GraphFedSeg

$\sum_{j=1}^{k}W_{i,j}=1$，确保c_i对其他客户端的协作权重构成概率分布；$W_{i,j}\geqslant 0$保证协作强度为正向贡献，从而保证知识传递时不引入反向干扰。

式(1)所示的服务器端的目标函数仅考虑了客户端的数据量占比和模型相似度，未纳入客户端的本地分割模型的预测不确定性。因此，容易给高不确定性的客户端分配过高的权重，且难以抵御异常客户端干扰，引入不可靠协作关系。因此，为了提升协作图质量，实现更可靠的权重分配，本文计算客户端的预测不确定性，并将其作为不确定性惩罚项引入服务器端的目标函数中。因此，修正服务器端的目标函数如式(2)所示：

$$\min_{\{W_{i,j}\}_j}\sum_j\left(W_{i,j}-p_j\right)^2-\alpha\sum_j W_{i,j}\cos\left(\theta_i,\theta_j\right)+\gamma\sum_j W_{i,j}U_j,$$
$$\text{s.t.}\sum_{j=1}^{k}W_{i,j}=1,\ \forall i;\ W_{i,j}\geqslant 0,\forall i,j \tag{2}$$

其中，$\gamma\sum_j W_{i,j}U_j$为不确定性惩罚项，γ为惩罚系数且$\gamma\geqslant 0$。将不确定性惩罚项融入协作图优化目标后，服务器端的优化函数扩展为同时平衡数据集大小、模型余弦相似度与不确定性惩罚的形式。

模型预测的不确定性通过证据理论进行量化，反映模型对像素级分类的犹豫程度。客户端级不确定性U_j可根据来自像素、图像、客户端的三级平均过程计算，具体步骤如下：

首先，计算每个像素的不确定性。针对图像中单个像素，模型输出的Logit向量经指数变换得到证据值e_c，并计算信念度$\alpha_c=e_c+1$。然后将所有类别的信念度α_c求和得总信念度S，并计算像素不确定性$u_{\text{pixel}}=C/S$，其中C是类别数。总信念度S越大，该像素的不确定性越小。

其次，计算单张图像的不确定性。对一张图像的所有像素，求其u_{pixel}的平均值，得到该图像的不确定性u_{image}如式(3)所示：

$$u_{\text{image}}=\frac{1}{M\times N}\sum_{i=1}^{M\times N}u_{\text{pixel},i} \tag{3}$$

其中，$M\times N$为图像尺寸。

最后，计算客户端的不确定性如式(4)所示，对客户端c_j本地所有图像求其u_{image}的平均值：

$$U_j=\frac{1}{n_j}\sum_{k=1}^{n_j}u_{\text{image},k} \tag{4}$$

其中，n_j为客户端c_j的图像数量。

不确定性惩罚项的核心作用是通过成本调控来过滤不可靠协作。式(2)将$W_{i,j}$作为优化目标，利用U_j筛选可靠协作源，把U_j作为客户端整体可靠性指

标来调节协作权重 $W_{i,j}$。具体地,当客户端 c_j 的预测不确定性 U_j 较高时,惩罚项会提高其参与协作的成本,从而在优化过程中降低 $W_{i,j}$,过滤来自客户端的噪声数据或异常干扰;反之,对于 U_j 较低的客户端,惩罚项的影响可忽略不计,此时协作权重主要由数据量占比与模型相似度共同决定。因此,利用不确定性惩罚项既能过滤高不确定性客户端携带的噪声对全局模型优化的负面影响,又能充分利用优质客户端的协作价值,使其深度参与模型聚合。

2.2.2　客户端优化

引入不确定性惩罚项后,其调控逻辑完全由服务器端负责。服务器通过惩罚项优化协作图权重 $W_{i,j}$ 筛选出可靠的协作关系,而客户端的目标函数无需直接加入不确定性相关计算。客户端仍聚焦于两个核心任务:一是最小化本地任务损失,确保模型能适配本地医学图像数据;二是满足本地模型与聚合模型的相似度约束,维持与全局协作的一致性。对于客户端 c_i,优化目标如式(5)所示:

$$\min_{\theta_i}\left(F_i\left(\sum_j W_{i,j}\theta_j\right)-\frac{\lambda_1}{2}\sum_j W_{i,j}\cos\left(\theta_i,\theta_j\right)\right) \tag{5}$$

其中,$F_i\left(\sum_j W_{i,j}\theta_j\right)$ 是客户端 c_i 基于服务器生成的聚合模型计算的本地经验损失,确保模型对本地数据的拟合能力。此时的 $W_{i,j}$ 是已通过式(2)优化得到的可靠协作权重矩阵。$-\frac{\lambda_1}{2}\sum_j W_{i,j}\cos\left(\theta_i,\theta_j\right)$ 是正则化项,最大化本地模型 θ_i 与其他客户端模型 θ_j 的相似度(加权由协作图 $W_{i,j}$ 决定),避免模型过度拟合本地数据。值得注意的是,此处的 θ_i 是上一轮服务器聚合后的分发给每个客户端的 θ_{global};θ_j 是其他客户端上一个联邦回合在本地数据集上训练得到的个性化原始参数。

虽然客户端的目标函数没有直接进行不确定性计算,但客户端模型优化会间接受到不确定性惩罚项的影响。因为客户端用于对齐的聚合模型更多吸收了低不确定性、高可靠性客户端的参数,减少了高不确定性客户端噪声的干扰。客户端在与这样的聚合模型对齐时,间接避免了不可靠协作的负面影响,有效实现了客户端轻量化与协作可靠性的平衡。

2.3　基于协作图的双层聚合机制

本文设计双层聚合机制,通过数据基础聚合保证全局稳定性,并通过基于不确定性协作图的邻域增强聚合保证本地个性化。双层聚合机制的计算方式如式(6)所示:

$$\theta_{\text{global}}=\lambda\cdot\theta_{\text{base}}+(1-\lambda)\cdot\theta_{\text{graph}} \tag{6}$$

其中:θ_{base} 是利用数据基础聚合得到的全局模型参数;θ_{graph} 是利用基于不确定性协作图的邻域增强聚合得到的全局模型参数;λ 为融合系数,取值范围为[0,1]。双层聚合机制保留了数据基础聚合来保障全局模型的稳定性,同时又添加了基于不确定性协作图的邻域增强聚合来注入局部个性化知识,从而平衡全局共性与局部适配,提升医学图像分割精度与可靠性,尤其是在常见组织器官通用分割与特殊病灶个性化适配之间取得了良好的平衡。

2.3.1　数据基础聚合

数据基础聚合是实现模型全局一致性的重要保障。数据基础聚合以客户端的数据量占比为核心权重,通过加权求和计算全局模型如式(7)所示:

$$\theta_{\text{base}}=\sum_{j=1}^{K}p_j\cdot\theta_j \tag{7}$$

其中:p_j 为客户端 c_j 的数据量占比,$p_j=n_j\Big/\sum_{j=1}^{K}n_j$,$n_j$ 为客户端 c_j 的样本量;θ_j 为客户端 c_j 在本次联邦回合经本地训练后的模型参数。

在医学图像分割场景中,数据规模与质量关联临床价值,如三甲医院的数据集样本多样、标注可靠,其本地模型甚至已学习到通用组织器官的特征。以数据量占比 p_j 为权重的数据基础聚合,能让样本覆盖全面、标注可靠的客户端主导全局模型,使 θ_{base} 具备临床通用性,既对常见组织器官保持稳定的分割精度,避免全局模型产生偏见,又作为邻域增强聚合的基准,与后续邻域增强的局部个性化信息形成互补。

2.3.2　邻域增强聚合

基于不确定性协作图的邻域增强聚合是双层聚合机制的核心。其利用不确定性协作图的邻接矩阵 $\boldsymbol{W}$,实现局部信息的精准聚合,主要包括以下两步。

首先是个性化邻域聚合。对每个客户端 c_i,其个性化邻域参数由不确定性协作图中值得信任的相似客户端的参数进行加权计算如式(8)所示:

$$\theta_{\text{neigh}}[i]=\sum_{j=1}^{K}W_{i,j}\cdot\theta_j \tag{8}$$

其中,$W_{i,j}$ 为不确定性协作图的邻接矩阵元素,在协作图优化阶段,服务器的目标函数为数据量占比损失、模型相似度收益、不确定性惩罚的和。在式(2)的不确定性惩罚项 $\gamma\sum_j W_{i,j}U_j$ 中,惩罚项的数值会随 U_j 增大而升高,导致总优化目标函数值增大。为了让服务器端目标函数最小化则要降低 $W_{i,j}$,其对 $\theta_{\text{neigh}}[i]$ 的贡献即可忽略;反之,低不确定性、高模型相似度的客户端 c_j 将以更高权重参与模型聚合。

其次是全局邻域平均。将所有客户端的个性化邻域参数取平均值,得到邻域增强聚合的最终输出如式(9)所示:

$$\theta_{\text{graph}}=\frac{1}{K}\sum_{i=1}^{K}\theta_{\text{neigh}}[i] \tag{9}$$

其中，$\theta_{\text{neigh}}[i]$是式(8)生成的各客户端个性化邻域参数，其作用是将所有客户端的可靠局部知识进行平均融合。相较于传统图聚合的无差别邻域融合，邻域增强聚合通过不确定性协作图的可靠性筛选实现了精准的局部知识传递，使注入的局部信息更精准、更安全，从而有效提升了模型对特殊病灶和复杂客户端的泛化能力。

2.4 算法流程

本文提出的GraphFedSeg的主要步骤包括：首先，初始化客户端本地模型和服务器全局模型和协作图，客户端基于本地数据进行训练并计算不确定性，将模型参数以及不确定性上传至服务器；其次，服务器根据客户端数据量占比、模型相似度以及不确定性更新不确定性协作图；最后，采用数据基础聚合和邻域增强聚合的层次化设计，完成模型聚合并发送给客户端，客户端基于新的聚合模型进行新一轮迭代。GraphFedSeg的具体流程如算法1所示。

算法1 GraphFedSeg

输入：客户端数量K，客户端数据、全局迭代轮数T，本地训练轮数E，学习率η等

输出：本地个性化模型θ_i

Process：

初始化全局模型θ_{global}、协作图邻接矩阵$\boldsymbol{W}$、本地数据集D_i、本地模型$\theta_i=\theta_{\text{global}}$

for t = 1 to T： // 全局迭代

// 客户端本地训练与基于证据理论的不确定性计算

for $i\in\{1, 2, \cdots, K\}$ do in parallel：

for e = 1 to E： // 本地训练E轮

从D_i中采样批次数据利用Adam优化器进行本地更新θ_i

//基于证据理论计算本地模型预测不确定性U_i

for $I\in D_i$： // 遍历每张图像

for $p\in I$： // 遍历图像中每个像素

计算模型输出的logit向量：$z=\theta_i(p)$

计算证据值$e_c=\exp(z_c)$和信念度$\alpha_c=e_c+1$

计算总信念度：$S=\sum_{c=1}^{C}\alpha_c$

计算像素级不确定性：$u_{\text{pixel}}=\frac{C}{S}$

计算图像级不确定性：$u_{\text{image}}=\frac{1}{M\times N}\sum_{i=1}^{M\times N}u_{\text{pixel},i}$

计算客户端级不确定性：$\sum U+=u_{\text{image}}$；$U_i=\sum U/|D_i|$

Sent (θ_i, U_i, n_i) to Server

// 服务器更新不确定性协作图$\boldsymbol{W}$

计算数据量占比$p_i=n_i/\text{sum}(n_i)$

计算模型余弦相似度矩阵$\cos_{ij}=\cos(\theta_i,\theta_j)$

优化协作图邻接矩阵$\boldsymbol{w}_i=\arg\min_x x^Tx+(-2p-\alpha s_i+\gamma U_i)^Tx$ //由式(2)得出

// 服务器执行双层聚合机制

进行数据量基础聚合$\theta_{\text{base}}=\sum_{j=1}^{K}p_j\cdot\theta_j$

进行个性化邻域聚合$\theta_{\text{neigh}}[i]=\sum_{j=1}^{K}W_{i,j}\cdot\theta_j$

进行全局邻域平均 $\theta_{\text{graph}}=\frac{1}{K}\sum_{i=1}^{K}\theta_{\text{neigh}}[i]$

进行两层结果融合$\theta_{\text{global}}=\lambda\cdot\theta_{\text{base}}+(1-\lambda)\cdot\theta_{\text{graph}}$

对每个客户端同步新一轮全局模型$\theta_i=\theta_{\text{global}}$

3 实验结果与分析

3.1 数据集

本文实验所用数据集为公开的内窥镜息肉数据集。数据来自四个中心，分别是Kvasir[35]、ETIS[36]、CVC-ColonDB[37]、CVC-ClinicDB[38]。为了确保实验的准确性和模型训练的有效性，每个中心的数据都被视为独立的本地客户端进行处理。其中，四个客户端的样本数量分别为1 000、196、380、612，每个客户端的本地数据都按照7∶2∶1划分为训练集、测试集和验证集，所有图像的尺寸都调整为384 × 384。以50%概率对训练集进行随机水平或垂直翻转，增加训练数据的多样性。

3.2 评估指标

为了全面评估模型的分割性能，本文实验采用医学影像分割任务中两个广泛使用的量化评估指标：Dice系数和HD95。二者可分别从区域完整性和边界精细度两方面客观地反映模型的分割效果。

Dice系数能有效评估目标区域的分割准确性，计算方式如式(10)所示：

$$\text{Dice}=2\times\frac{|A\cap B|}{|A|+|B|} \tag{10}$$

其中，A表示模型预测的分割结果，B表示真实标签。Dice系数取值范围为[0, 1]，值越接近1表示分割结果与真实标签越相似。

豪斯多夫距离(Hausdorff Distance，HD)定义为两个点集中最远点对的距离的最小值，用来量化两个点集之间的相似度。从A到B的单向豪斯多夫距离如式(11)所示：

$$h(A,B)=\max_{a\in A}\left(\min_{b\in B}\|a-b\|\right) \tag{11}$$

其中，$\|a-b\|$表示点a与点b之间的欧氏距离。

HD95是豪斯多夫距离的95%分位数，即忽略距离最大的5%极端值后的值，能更稳健地反映整体边界误差。

3.3 实验环境与参数设置

本文实验的硬件环境配置为:Intel(R) Xeon(R) Gold 6326 CPU,50 GB 内存,NVIDIA Tesla A30 GPU,24 GB 显存。实现 GraphFedSeg 所使用的深度学习框架为 PyTorch。GraphFedSeg 采用的分割网络为 2D UNet,使用 Adam 优化器训练局部模型,学习率设置为 5×10^{-4},Batch 大小设置为 4,惩罚系数 $\gamma=0.4$,融合系数 $\lambda=0.2$。训练 GraphFedSeg 的联邦回合数 $T=250$,本地训练 Epochs 设置为 2。

3.4 消融实验与分析

为充分验证 GraphFedSeg 中各模块的有效性,本文在息肉分割数据集上设计了消融实验。首先,选择 FedAvg 作为本文消融实验的 Baseline。其次,在此基础上,按照模块功能的递进关系与逻辑关联依次添加协作图、不确定性协作图(UCG)、双层聚合机制(DAM),来分别验证每个新增模块对模型性能的影响。同时,为确保实验公平性,严格保持除所添加模块外的其他实验条件(如网络参数设置、训练迭代次数、数据输入格式等)一致。最后,通过对不同模块组合的实验结果进行对比与分析,明确各模块的独立作用并探究模块间的协同效应。在息肉分割数据集上的消融实验结果如表 1 所示。

表 1　在息肉分割数据集上的消融实验结果

Table 1　Ablation experimental results on the polyp segmentation dataset

Baseline	协作图	UCG	DAM	Dice/%(↑)					HD95(↓)				
				C1	C2	C3	C4	Avg	C1	C2	C3	C4	Avg
√	×	×	×	85.02	75.82	82.19	88.92	82.99	35.01	43.22	26.15	22.51	31.72
√	√	×	×	88.31	79.33	82.16	89.01	84.70	32.45	35.94	32.45	24.56	31.35
√	√	√	×	87.72	**84.65**	82.41	89.02	85.95	35.19	**16.49**	32.39	23.62	26.92
√	√	√	√	**89.10**	84.47	**84.35**	**90.47**	**87.10**	**31.20**	22.38	**25.24**	**20.93**	**24.94**

注:最优结果用粗体表示,C1 到 C4 分别为 4 个客户端,下表同。

在表 1 中,Baseline 的平均 Dice 得分为 82.99%,平均 HD95 为 31.72,能够完成基本的息肉分割任务。但是,在面对复杂数据(如客户端 C2)时的分割效果欠佳,无论是区域覆盖准确性还是边界定位精度都有较大的提升空间。

在 Baseline 的基础上引入协作图之后,平均 Dice 得分提升至 84.70%,相较于 Baseline 提高了 1.71%,平均 HD95 降至 31.35。这表明协作图模块通过构建客户端之间的协作关系,能够有效整合多客户端资源,提升分割性能,尤其对客户端 C1 和 C2 的改善比较明显。但同时也可以看到,该模块对不同客户端的提升效果存在差异,且整体提升幅度比较有限。

在“Baseline+协作图”的基础上引入不确定性惩罚后,各客户端的性能得到了更为显著的提升。引入不确定性惩罚后的平均 Dice 升至 85.95%,相较于“Baseline+协作图”方法提升了 1.25%,平均 HD95 降至 26.92,下降了 4.43。其主要原因是不确定性惩罚项可有效区分客户端信息可靠性,通过优化协作图信息传递来提升分割性能,尤其显著提升了客户端 C2 的表现,大幅缩小了客户端性能差距。这充分说明了不确定性惩罚项的有效性。

在添加双层聚合机制后,C2 的 Dice 下降、HD95 上升,可能是由于双层聚合机制对其数据的不确定性估计偏差、跨客户端信息适配性不足、参数设置不匹配其数据特性,导致聚合引入的干扰超过有效信息所致。但是,综合来看,引入双层聚合机制后的模型的平均 Dice 达到 87.10%,相较于“Baseline+协作图+不确定性惩罚项”提升了 1.15%,平均 HD95 降至 24.94,下降了 1.98。这表明双层聚合机制通过优化模型聚合方式,能够更充分地挖掘和利用多客户端的协同信息,进一步提升了模型的分割性能。另外,该模块与不确定性协作图形成了良好的协同效应,使得模型在各客户端的表现更加均衡且优异,充分验证了本文提出的模块组合协同的有效性。

3.5 对比实验与分析

为了验证 GraphFedSeg 的有效性,本文在息肉分割数据集上将 GraphFedSeg 与基线方法 FedAvg[7] 以及 12 种先进联邦学习方法进行了比较。12 种先进方法可以大致分为三类:第一类主要是针对客户端间数据分布不均的挑战,包括 FedDG[20]、FedProx[39];第二类方法主要是聚焦于模型参数聚合和知识传递效果,包括 FedRep[40]、FedBABU[41]、FedGKD[42]、FedCLAM[23]、FedEvi[22];第三类主要注重模型的个性化,包括 Ditto[43]、FedDP[30]、FedLC[29]、IOP-FL[31]、FedLPPA[44]。

为确保对比实验的公平性,所有方法采用相同的实验设置、相同尺寸的输入图像、相同的数据增强策略。评估指标均采用 Dice 系数和 HD95,所有方法均使用 U-Net 作为客户端本地网络。GraphFedSeg、基线方法 FedAvg 及 12 种先进方法在息肉分割数据集上的分割结果如表 2 所示。

由表 2 可以看出:在客户端 C1、C3 和 C4 上,GraphFedSeg 均取得了最高的 Dice 分数,分别为

表2 GraphFedSeg与其他对比方法在息肉数据集上的分割结果

Table 2 Segmentation results of GraphFedSeg and other comparison methods on the polyp dataset

Methods	Dice/%(↑)					HD95(↓)				
	C1	C2	C3	C4	Avg	C1	C2	C3	C4	Avg
FedAvg(AISTATS 2017)[7]	85.02	75.82	82.19	88.92	82.99	35.01	43.22	26.15	22.51	31.72
FedProx(ICML 2020)[39]	85.04	73.31	84.26	90.24	83.21	34.10	51.04	25.29	21.81	33.06
Ditto(ICML 2021)[43]	79.85	85.65	79.56	88.62	83.42	42.40	22.42	35.97	22.18	30.74
FedRep(ICML 2021)[40]	78.75	88.83	83.23	86.91	84.43	43.24	17.50	26.83	22.81	27.60
FedDG(CVPR 2021)[20]	85.44	80.36	83.77	88.66	84.56	34.59	30.44	28.04	20.96	28.51
FedBABU(ICLR 2022)[41]	85.70	83.21	83.65	90.16	85.68	38.62	20.39	29.32	21.96	27.57
FedLC(ECCV 2022)[29]	86.07	82.26	84.12	90.44	85.72	41.23	23.78	30.97	21.20	29.30
IOP-FL(IEEE TMI 2023)[31]	86.08	82.35	81.42	88.12	84.49	39.16	22.08	32.09	22.28	28.90
FedGKD(IEEE TC 2024)[42]	83.07	83.34	80.24	90.05	84.18	42.41	23.81	30.50	21.32	29.51
FedDP(IEEE TMI 2024)[30]	84.15	**89.21**	83.95	88.46	86.44	41.73	**15.04**	26.88	22.09	26.44
FedEvi(MICCAI 2024)[22]	86.78	85.84	84.11	89.31	86.51	37.18	22.69	25.62	22.58	27.02
FedCLAM(MICCAI 2025)[23]	84.49	86.69	83.21	90.25	86.16	40.20	19.60	27.56	22.60	27.49
FedLPPA(IEEE TMI 2025)[44]	87.21	86.35	82.30	88.95	86.20	35.22	25.60	28.25	22.18	27.79
GraphFedSeg	**89.10**	84.47	**84.35**	**90.47**	**87.10**	**31.20**	22.38	**25.24**	**20.93**	**24.94**

89.10%、84.35%、90.47%，FedLPPA在客户端C1上为87.21%(排第二)。在客户端C2上FedDP的Dice分数以89.21%领先，GraphFedSeg为84.47%。GraphFedSeg的平均Dice得分以87.10%位居榜首，较FedAvg提升4.11%。从HD95指标看，在客户端C1、C3、C4上，GraphFedSeg分别以31.20、25.24、20.93为最低，在客户端C2上，FedDP以15.04为最佳。GraphFedSeg的平均HD95以24.94为最低，较FedAvg降低6.78。综上，在平均Dice和HD95上，GraphFedSeg均超越了所有对比方法，取得了最好的分割性能。

GraphFedSeg与其他对比方法的可视化分割结果如图2所示。由图2可以看出：GraphFedSeg在不同客户端的各类息肉分割中均展现出优势。具体地，对于客户端C1中特征明显的息肉样本，其分割结果在区域覆盖完整性和边界准确性上均优于其他方法，比如FedAvg对息肉边缘的分割较模糊，而GraphFedSeg能更精准地勾勒出轮廓。针对客户端C2上的边界模糊的息肉样本，尽管定量指标提升不显著，但与FedProx等方法相比，GraphFedSeg能更好地捕捉息肉的大致范围，减少因边界模糊导致的过分割或欠分割。针对客户端C3上的形态复杂的息肉样本，FedGKD等方法容易出现分割不连续的情况，FedLPPA和FedCLAM等方法分割不完整，GraphFedSeg的分割结果更接近真实标签。对于客户端C4中特征突出的息肉样本，GraphFedSeg与FedGKD、FedRep等方法的分割结果都较为准确，且在息肉与周围组织的细微边界处表现优异，而FedLPPA和FedEvi等方法则表现欠佳。由此可见，GraphFedSeg在不同类型息肉样本的区域覆盖和边界定位上均取得了更好的表现。

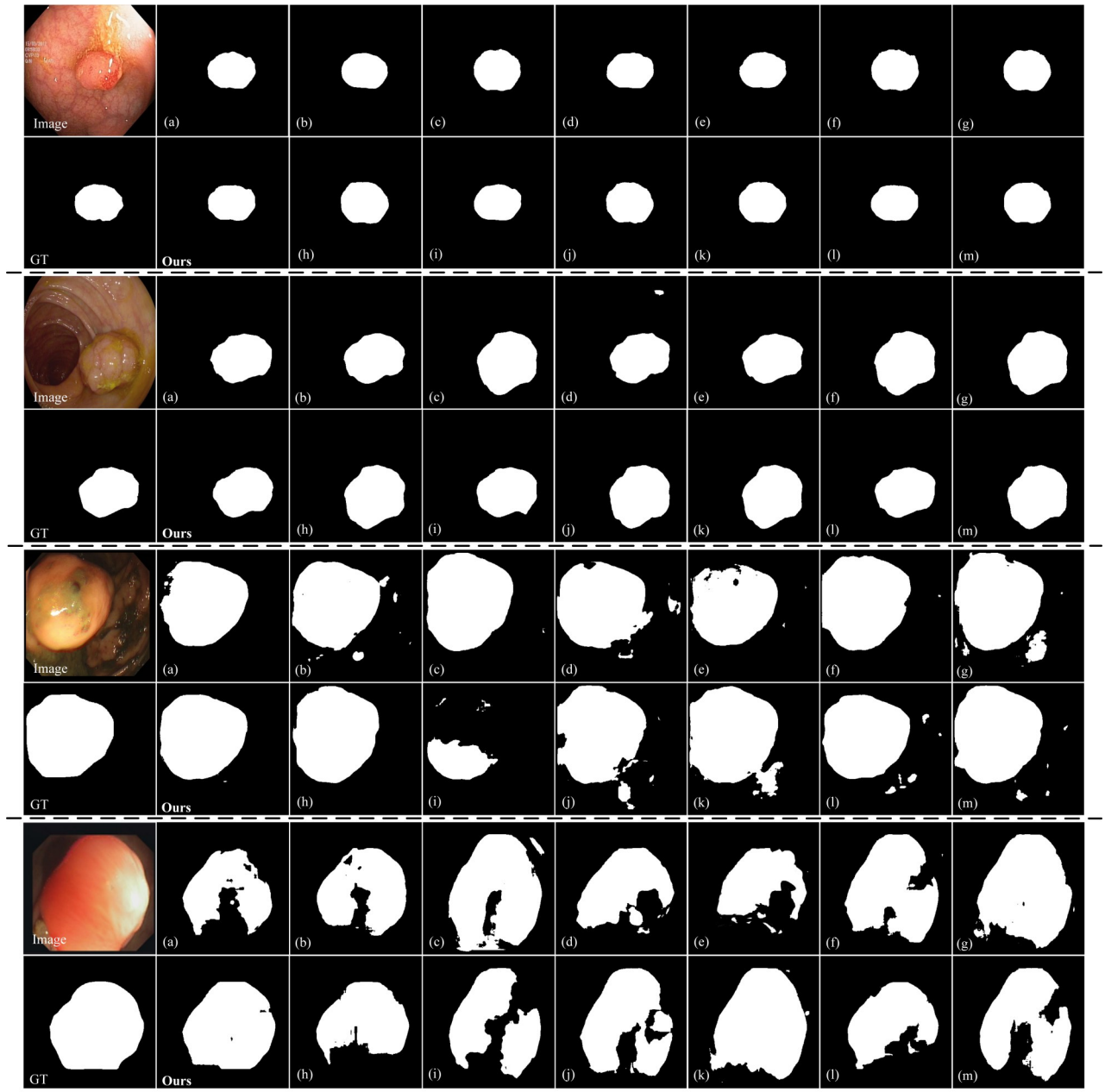
3.6 超参数讨论

惩罚系数γ和融合系数λ直接决定了不确定性协作图和双层聚合机制的运行效果，对最终的分割性能至关重要。为了探究其影响，本文开展超参数分析实验。针对融合系数和惩罚系数，在固定其他实验参数的前提下，以0.1为步长在设定范围内进行等间隔采样，同时记录对应的平均Dice系数与HD95值，进而分析超参数对模型分割性能的影响。

3.6.1 惩罚系数

在协作图的优化目标函数中，为了平衡噪声过滤与有效知识保留，使协作图更聚焦低噪声、高价值客户端的协作，本文添加了不确定性惩罚项，其中惩罚系数γ用于调控惩罚强度。由于过高或过低的惩罚强度对于协作图的优化都会产生不利影响，本文实验的γ取值范围为$[0, 1.4]$，客户端的平均Dice以及HD95随γ的变化趋势如图3所示。

由图3可以看出：首先，在低惩罚区间$[0, 0.2]$的惩罚项作用极弱，高不确定性客户端(预测不可靠、含噪声数据)的干扰未被有效过滤，协作图可靠性差。此时聚合的知识混杂噪声，导致分割区域一致性差且边界误差大。其次，在惩罚区间$[0.2, 0.5]$的惩罚强度适中，既有效压制高不确定性客户端的噪声干扰，又保留了低不确定性客户端的有效知识，在$\gamma = 0.4$时协作图可靠性达到最优。此时聚合知识的质量最高，分割区域一致性与边界精度同步最优，达到了噪声过滤与知识保留的有效平衡。最后，在0.5以后的惩罚强度过大，过度抑制甚至错误过滤不确定性略高



注:由上至下的四个样本分别来自客户端C1、C2、C3、C4; (a)FedLPPA; (b)FedAvg; (c)FedBABU; (d)FedDG; (e)FedEvi; (f)FedLC; (g)FedRep; (h) Ditto; (i)FedCLAM; (j)FedDP; (k)FedGKD; (l) FedProx; (m)IOP-FL。

图2　GraphFedSeg与其他对比方法的可视化结果

Figure 2　Visualization of GraphFedSeg compared to other methods

但仍有价值的客户端(如包含边界模糊息肉或样本数量少但临床关键的数据集),导致协作图丢失部分有效协作关系,聚合知识的完整性不足,此时分割区域一致性下降,边界精度同步恶化。后续[0.8, 1.2]区间内的波动则是由于不确定性计算的近似性,以及实验随机因素(数据采样、优化器的梯度波动)引发的误差振荡,但整体性能趋势仍为持续恶化。

3.6.2　融合系数

在基于不确定性协作图的双层聚合机制中,融合系数是平衡数据基础聚合与邻域增强聚合的核心参数,其取值直接决定两种聚合策略的权重配比,进而影响模型的分割性能。λ取值范围为[0, 1],客户端的平均Dice以及HD95随λ的变化趋势如图4所示。

观察图4中的整体趋势可以看出:随着λ的不断

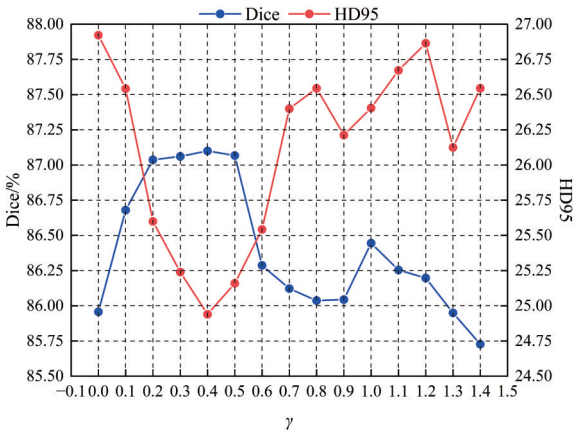


图3 客户端的平均Dice以及HD95随γ的变化趋势图

Figure 3 Plot of average Dice and HD95 versus γ for the client

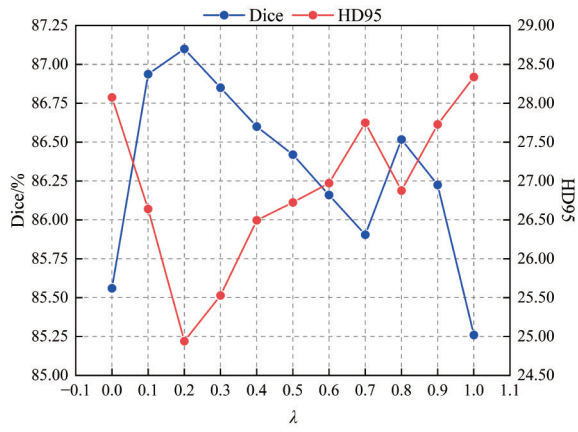


图4 客户端的平均Dice以及HD95随λ的变化趋势图

Figure 4 Plot of average Dice and HD95 versus λ for the client

增大,客户端的平均分割性能呈现先提升后下降的趋势。其主要原因是客户端的平均分割性能由融合系数λ对全局一致性与局部个性化的平衡效果来决定。具体而言,当λ取值在0.2附近时,能最佳平衡全局共性传递与局部个性化需求,平均Dice达到峰值、HD95降至最低;而当λ偏离这一区间尤其是向1趋近时,模型会过度强调全局一致性,进而大幅削弱对各客户端特殊分割需求的适配能力,最终使得平均Dice系数逐渐下降、HD95逐渐上升,分割的区域一致性与边界精度均持续恶化。在该过程中由于实验的随机因素带来了小幅波动,但是并不会改变全局一致性和局部个性化平衡主导分割性能的核心规律。

3.7 不确定性计算方法对比

为了验证不同确定性计算方法的性能与开销情况,将基于证据理论的不确定性计算方法和基于蒙特卡罗Dropout的不确定性计算方法进行了对比实验,主要从分割性能和时空开销两方面进行比较。为了保证对比实验的公平性,仅将GraphFedSeg中的不确定性计算替换为基于蒙特卡罗Dropout的不确定性计算方法,其他实验设置保持不变。

3.7.1 性能对比

其中,对蒙特卡罗Dropout方法进行了10次采样,不同客户端的分割性能以及四个客户端的平均性能如表3所示。

表3 不同不确定性方法性能对比

Table 3 Performance comparison of different uncertainty methods

方法	Dice/%(↑)					HD95(↓)				
	C1	C2	C3	C4	Avg	C1	C2	C3	C4	Avg
蒙特卡罗Dropout	84.99	**87.92**	**84.90**	89.33	86.78	39.42	**16.51**	**25.23**	23.77	26.23
证据理论	**89.10**	84.47	84.35	**90.47**	**87.10**	**31.20**	22.38	25.24	**20.93**	**24.94**

由表3可以看出:在训练结束后,证据理论的平均Dice较蒙特卡罗Dropout高0.32%,平均HD95低1.29,在区域完整性和边界精细度上均实现提升。尤其在数据量较大、临床价值更高的客户端C1和C4上,Dice分别领先4.11%和1.14%,更能适配多中心协作中核心医疗机构的分割需求。

3.7.2 开销对比

为了进一步验证基于证据理论的不确定性计算方法在医疗场景中的优势,补充了其与蒙特卡罗Dropout(5次、10次采样)在时间开销和显存占用两方面的对比实验。其中,完整训练步骤时间为一个Batch的数据训练时间,GPU显存是指特定计算过程中GPU显存的最大占用量。具体结果如表4所示。

根据表4的结果,从以下四个方面进行分析。

(1)基础推理效率:证据理论的前向传播时间(4.80 ms)仅为蒙特卡罗Dropout(5次采样)的18.5%、蒙特卡罗Dropout(10次采样)的9.6%,单次推理速度远超蒙特卡罗Dropout。这是因为证据理论无需多次随机失活采样,仅通过单次前向传播输出的Logit向量即可计算证据值与信念度,计算流程更简洁。

(2)不确定性计算效率:在“前向传播+不确定性计算”的总时间方面,证据理论(14.22 ms)较蒙特卡罗Dropout5次采样(32.83 ms)降低56.7%,较10次采样(57.51 ms)降低75.3%。联邦学习中每轮通信需要所有客户端同步完成不确定性计算并上传结果,证据理论的低时间开销可避免因部分客户端计算延迟导致的协作卡顿,提升全局训练效率。

(3)完整训练效率:证据理论的完整训练步骤时

间(41.84 ms)较蒙特卡罗 Dropout(5 次采样)节省35.2%,较 10 次采样节省 51.8%。在 250 轮联邦回合的训练中,证据理论可累计节省训练时间约 3.4 h(相较于 10 次采样),大幅缩短了模型训练周期。

(4)显存占用:证据理论的GPU显存占用(150.39 MB)与蒙特卡罗 Dropout(151.39~152.64 MB)基本持平且略低。这表明证据理论在实现“像素、图像、客户端”三级不确定性量化的同时,未额外增加显存负担;蒙特卡罗 Dropout 的显存占用随采样次数增加略有上升(10 次采样较 5 次增加 1.25 MB),而证据理论的显存占用更稳定,无采样次数相关的额外开销。

表 4　不同不确定性方法开销对比

Table 4　Overhead comparison of different uncertainty methods

方法	仅前向传播时间/ms	前向传播+不确定性计算时间/ms	完整训练步骤时间/ms	GPU 显存/MB
证据理论	4.80	14.22	41.84	150.39
蒙特卡罗 Dropout(5 次采样)	25.90	32.83	64.55	151.39
蒙特卡罗 Dropout(10 次采样)	49.95	57.51	86.76	152.64

综上,证据理论实现了更高性能和更低开销的协同优化,在时间开销和显存占用上均展现出优势,既能满足联邦医学图像分割的多中心实时协作需求,同时保持更优的分割性能。

3.8　客户端权重变化分析

为直观分析客户端权重的动态变化规律,本文进行了相关实验。总训练轮次为 250 轮,首先得到每一轮四个客户端的聚合权重,并对每 10 轮的结果进行平均,从而得到四个客户端的聚合权重随训练轮次变化的折线图,具体如图 5 所示。

由图 5 可以看出:客户端 C2 的聚合权重始终处于最低区间,随训练轮次增加有微小提升,全程无明显波动。这是因为客户端 C2 的数据集包含更多复杂病灶样本,不确定性惩罚项持续抑制其全局贡献占比,避免高噪声知识干扰全局聚合。客户端 C1 聚合权重始终最高,波动幅度较小。主要是因为客户端 C1 的数据集标注质量高、病灶特征典型,不确定性量化结果最低,协作图优先为其分配高权重,强化可靠知识的协同传递。客户端 C3、C4 的聚合权重稳定在中间范围,与各自的不确定性程度精准匹配,避免了传统个性化联邦学习无差别平均聚合的弊端。随着训练轮次推进,所有客户端的聚合权重波动逐渐减小并趋于稳定。这表明协作图已通过多轮迭代,精准捕捉到各客户端的不确定性特征与数据价值,权重分配从初始的近似均匀分布,逐步收敛至稳定,验证了协作图动态优化机制的鲁棒性。

3.9　协作强度变化分析

为清晰呈现不同训练阶段的协作强度分布,从第 0 轮开始,对协作矩阵每 25 轮进行一次抽样,绘制协作强度热力图如图 6 所示。其中,颜色越深代表协作强度越高,颜色越浅代表协作强度越低。

由图 6 可以看出:初始阶段[第 0 轮:图 6(a)],所有客户端间协作权重呈均匀分布(均为 0.333 3),尚未引入不确定性与模型相似度信息,客户端间呈无差

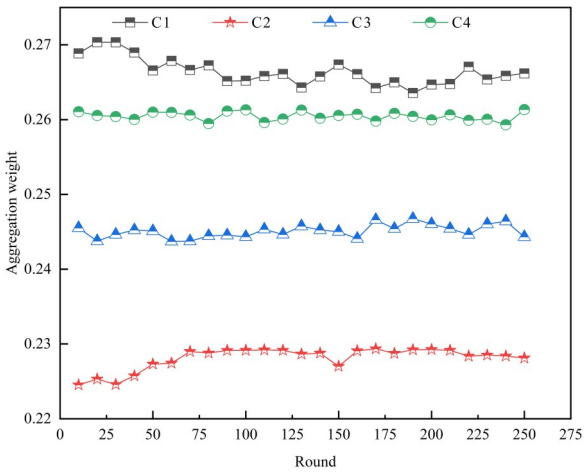
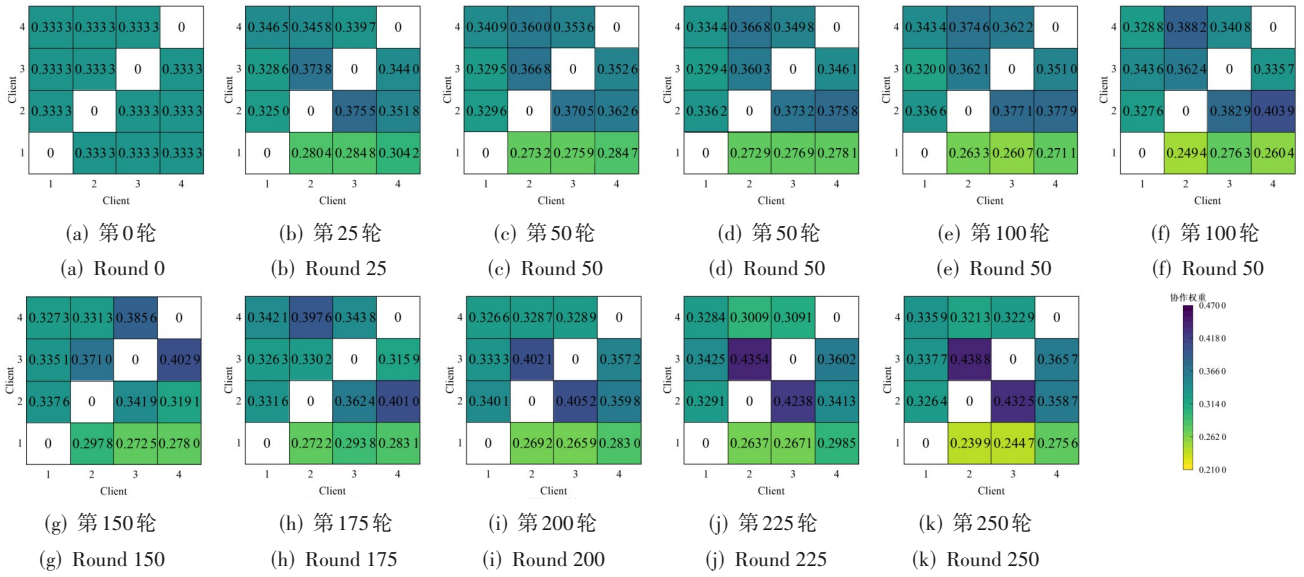


图 5　客户端聚合权重动态变化情况

Figure 5　Client-side aggregated weight dynamics

异协作。早期阶段[第 25~50 轮:图 6(b)~图 6(c)],协作权重开始分化,高不确定性的客户端 C2 向高相似度的客户端 C3 集中,低不确定性的客户端 C1 与其他客户端的协作权重开始降低。中期阶段[第 75~150 轮:图 6(d)~图 6(g)],高不确定性的客户端 C2 与 C3 的协作权重维持高位,显著高于 C2 与其他客户端的协作权重;低不确定性的客户端 C1、C4 与同类客户端的协作强度的变化比较小,形成初步的可靠协作集群。后期阶段[第 175~250 轮:图 6(h)~图 6(k)],客户端 C2 与 C3 的协作权重达峰值 0.432 5。这一过程直观反映了高不确定性客户端 C2 的协作权重向高相似度的客户端 C3 靠近,协作图从均匀分布逐步演化为精准匹配不确定性和模型相似度的稳定结构,清晰反映了 GraphFedSeg 的核心协作逻辑。

由于客户端 C2 与 C3 的数据分布较为相似,模型相似度高,其高相似度的正向作用抵消了 C2 高不确定性的负向影响,协作图保留并强化二者之间的双向协作,成为协作强度最高的组合。客户端 C3 可捕捉 C2 的有效信息,高相似度避免了知识传递偏差,体现



(a) 第0轮 (b) 第25轮 (c) 第50轮 (d) 第50轮 (e) 第100轮 (f) 第100轮

(a) Round 0 (b) Round 25 (c) Round 50 (d) Round 50 (e) Round 50 (f) Round 50

(g) 第150轮 (h) 第175轮 (i) 第200轮 (j) 第225轮 (k) 第250轮

(g) Round 150 (h) Round 175 (i) Round 200 (j) Round 225 (k) Round 250

图6 多阶段协作强度热力图

Figure 6 Heat map of multi-stage collaboration strength

了非单一不确定性决定协作的优化逻辑，避免完全否定高不确定性客户端的价值。而客户端C2与C1、C4因模型相似度较低，正向作用不足，无法抵消高不确定性，协作被自然抑制。此外，C1、C3、C4之间维持较强的协作关系，保障了全局模型的可靠知识积累。

3.10 隐私预算对分割性能的影响

为了验证不同隐私保护强度（隐私预算ε）对GraphFedSeg的分割性能的影响，选取多组不同隐私预算ε值覆盖低、中、高隐私保护场景，观测四个客户端的平均Dice随训练轮次的变化。

本文保持GraphFedSeg的核心模块以及其他实验参数（训练轮数、学习率、模型结构等）不变，仅在客户端本地训练阶段引入差分隐私，固定松弛项$\delta = 1\times10^{-5}$（符合医疗数据隐私合规要求），确保性能差异仅由差分隐私的噪声引入导致。以四个客户端的平均Dice为观测对象，与原方法的性能评价保持一致。设置250轮联邦回合，每10轮进行一次采样，计算四个客户端的平均分割性能，得到客户端平均分割性能随训练轮次变化的折线图，具体结果如图7所示。

由图7可以看出：在低隐私预算（$\varepsilon=1$、5）情况下，客户端的平均分割性能在训练前期短暂小幅上升后出现停滞，整体平均Dice处于20%~30%的低水平。其主要原因是在低隐私预算ε下添加的噪声量极大，客户端本地模型参数严重失真，使模型预测输出偏离真实值，影响不确定性计算。服务器误将噪声引发的预测波动判定为数据本身的高不确定性，进而错误降低有效客户端的协作权重，使得不确定性协作图（UCG）丧失筛选可靠协作的核心能力。同时，噪声持续干扰梯度更新方向，模型无法有效收敛，最终性能

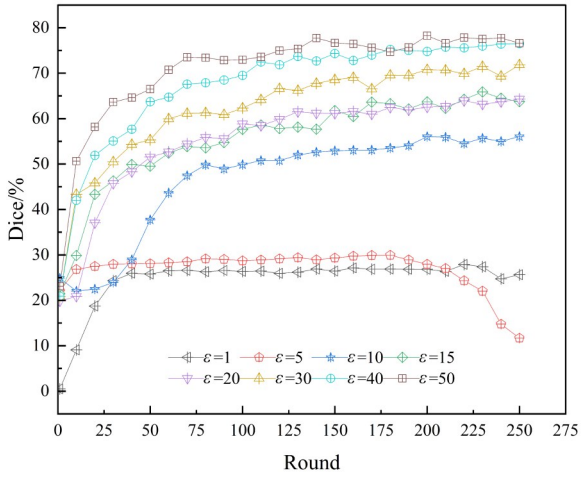


图7 不同ε取值下客户端性能变化

Figure 7 Client-side performance variation for different ε values

持续低迷。

在中隐私预算（$\varepsilon=10$、15、20）情况下，客户端的平均分割性能前期快速上升，约100轮后逐渐趋于稳定，但最终平均Dice维持在50%~70%的中等水平，显著低于高隐私预算场景。此区间的噪声干扰使客户端本地参数的失真程度有限，不确定性计算的伪不确定性占比降低，UCG虽受一定干扰，但仍能部分识别真实高可靠性客户端。双层聚合机制可整合部分有效知识，因此模型能收敛，但噪声仍会削弱知识传递的效率，导致性能无法达到高隐私预算的水平。

在高隐私预算（$\varepsilon=30$、40、50）情况下，性能在50~100轮内快速上升，随后稳定收敛，最终平均Dice接近80%的高水平（仅低于未引入差分隐私时的7%）。其主要原因是在高隐私预算下添加的噪声量极小，客

户端本地模型参数轻微失真,不确定性计算接近数据本身的真实不确定性;UCG能够正常基于真实不确定性调整协作权重,有效筛选可靠客户端并聚合优质知识,充分发挥了GraphFedSeg的两个核心模块(协作图和双层聚合)的优势,因此模型能快速收敛并维持较高分割性能。

综上,实验结果清晰呈现了隐私预算与模型平均分割性能的关联:低隐私预算时,噪声干扰严重,模型难以有效收敛,平均Dice长期维持在低水平;中隐私预算时,噪声干扰部分可控,模型可收敛但性能受限,平均Dice稳定在中等水平;高隐私预算时,噪声干扰轻微,GraphFedSeg的不确定性协作图与双层聚合机制可正常发挥作用,模型快速收敛并维持接近高的分割性能。整体而言,实验验证了隐私保护强度与分割性能负相关,也表明低隐私预算下的噪声会显著干扰GraphFedSeg的有效性。另外,隐私保护强度的提升会伴随模型可用性的下降与通信开销的增加,三者需根据实际医疗场景的隐私需求和性能要求进行灵活权衡。后续工作将结合临床应用场景进一步研究隐私保护、模型可用性和通信开销三者的有效权衡策略。

4　结论

本文提出了一个个性化联邦医学图像分割方法GraphFedSeg。首先,提出了基于不确定性预测的协作图构建方法,并首次将其应用在医学图像分割任务中。通过证据理论量化客户端预测不确定性,在协作图更新中加入不确定性惩罚以动态调整协作强度,有效过滤高噪声干扰,解决了基础协作图可靠性不足的问题,提升了客户端协作关系的精准度与稳定性。进一步地,设计了基于不确定性协作图的双层聚合机制,利用数据量基础聚合和基于不确定性协作图的邻域增强聚合,兼顾了全局一致性和局部个性化,实现了个性化联邦学习中全局模型与局部个性化需求的平衡。在息肉分割数据集上进行了大量实验,证明了GraphFedSeg的有效性。

未来工作将进一步研究双层聚合机制的自适应调节策略,以实现全局与局部平衡的智能优化。

参考文献

[1] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241.

[2] Zhou Zongwei, Siddiquee M R, Tajbakhsh N, et al. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.

[3] Chen J, Mei J, Li X, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers[J]. Medical Image Analysis, 2024, 97: 103280.

[4] Hatamizadeh A, Nath V, Tang Yucheng, et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images[C]//7th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Heidelberg: Springer, 2021: 272-284.

[5] 杨子瑶, 雷涛, 杜晓刚, 等. 基于可疑像素相互修正的半监督医学图像分割[J]. 电子学报, 2025, 53(5): 1607-1621.
Yang Ziyao, Lei Tao, Du Xiaogang, et al. Semi-supervised medical image segmentation based on suspicious pixel mutual correction[J]. Acta Electronica Sinica, 2025, 53(5): 1607-1621. (in Chinese)

[6] 雷涛, 张峻铭, 杜晓刚, 等. 基于混洗特征编码与门控解码的医学图像分割网络[J]. 电子学报, 2024, 52(12): 4142-4152.
Lei Tao, Zhang Junming, Du Xiaogang, et al. Medical image segmentation network based on shuffled feature encoding and gated decoding[J]. Acta Electronica Sinica, 2024, 52(12): 4142-4152. (in Chinese)

[7] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.

[8] Sheller M J, Edwards B, Reina G A, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data[J]. Scientific Reports, 2020, 10(1): 12598.

[9] Jiang Meirui, Roth H R, Li Wenqi, et al. Fair federated medical image segmentation via client contribution estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 16302-16311.

[10] Ma Benteng, Zhang Jing, Xia Yong, et al. VNAS: Variational neural architecture search[J]. International Journal of Computer Vision, 2024, 132(9): 3689-3713.

[11] Zhang Ling, Wang Xiaosong, Yang Dong, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation[J]. IEEE Transactions on Medical Imaging, 2020, 39(7): 2531-2540.

[12] Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: A model-agnostic

meta-learning approach[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 300.

[13] Yu Tao, Bagdasaryan E, Shmatikov V. Salvaging federated learning by local adaptation[PP/OL]. V3. arXiv (2022-03-03)[2025-09-06]. https://arxiv.org/abs/2002.04758.

[14] Tan A Z, Yu Han, Cui Lizhen, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.

[15] Makhija D, Ghosh J, Ho N. A Bayesian approach for personalized federated learning in heterogeneous settings[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2024: 3252.

[16] Deng Yuyang, Kamani M M, Mahdavi M. Adaptive personalized federated learning[C/OL]//9th International Conference on Learning Representations. 2021: 1-47. https://openreview.net/forum id=g0a-XYjpQ7r.

[17] Arivazhagan M, Aggarwal V, Singh A K, et al. Federated Learning with Personalization Layers[PP/OL]. V1.arXiv (2019-12-02)[2025-09-06]. https://doi.org/10.48550/arXiv.1912.00818.

[18] Ye Rui, Ni Zhenyang, Wu Fangzhao, et al. Personalized federated learning with inferred collaboration graphs[C]//Proceedings of the 40th International Conference on Machine Learning. PMLR, 2023: 39801-39817.

[19] Zhou Ziran, Gao Guanyu, Wu Xiaohu, et al. Personalized federated learning via learning dynamic graphs[PP/OL]. V1.arXiv (2025-03-07)[2025-09-06]. https://arxiv.org/abs/2503.05474.

[20] Liu Quande, Chen Cheng, Qin Jing, et al. FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 1013-1023.

[21] Li Xiaoxiao, Jiang Meirui, Zhang Xiaofei, et al. FedBN: Federated learning on non-IID features via local batch normalization[C/OL]//9th International Conference on Learning Representations. 2021: 1-27. https://openreview.net/forum id=6YEQUn0QICG.

[22] Chen Jiayi, Ma Benteng, Cui Hengfei, et al. FedEvi: Improving federated medical image segmentation via evidential weight aggregation[C]//27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2024: 361-372.

[23] Siomos V, Passerat-Palmbach J, Tarroni G. FedCLAM: Client adaptive momentum with foreground intensity matching for federated medical image segmentation[C]//28th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2025: 247-257.

[24] Jiang Le, Ma Liyan, Zeng Tieyong, et al. UFPS: A unified framework for partially annotated federated segmentation in heterogeneous data distribution[J]. Patterns, 2024, 5(2): 100917.

[25] Xu Xuanang, Deng H H, Gateno J, et al. Federated multi-organ segmentation with inconsistent labels[J]. IEEE Transactions on Medical Imaging, 2023, 42(10): 2948-2960.

[26] Tölle M, Navarro F, Eble S, et al. FUNAvg: Federated uncertainty weighted averaging for datasets with diverse labels[C]//27th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2024: 405-415.

[27] Wu Nannan, Sun Zhaobin, Yan Zengqiang, et al. FedA^{3}I: Annotation quality-aware aggregation for federated medical image segmentation against heterogeneous annotation noise[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(14): 15943-15951.

[28] Xiang Yangyang, Wu Nannan, Yu Li, et al. FedIA: Federated medical image segmentation with heterogeneous annotation completeness[C]//27th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2024: 373-382.

[29] Wang Jiacheng, Jin Yueming, Wang Liansheng. Personalizing federated medical image segmentation via local calibration[C]//17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 456-472.

[30] Wang Jiacheng, Jin Yueming, Stoyanov D, et al. FedDP: Dual personalization in federated medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2024, 43(1): 297-308.

[31] Jiang Meirui, Yang Hongzheng, Cheng Chen, et al. IOP-FL: Inside-outside personalization for federated medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2023, 42(7): 2106-2117.

[32] Xie Luyuan, Lin Manqing, Liu Siyuan, et al. pFLFE: Cross-silo personalized federated learning via feature enhancement on medical image segmentation[C]//27th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2024: 599-610.

[33] Sattler F, Müller K R, Samek W. Clustered federated

learning: Model-agnostic distributed multitask optimization under privacy constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(8): 3710-3722.

[34] Shamsian A, Navon A, Fetaya E, et al. Personalized federated learning using hypernetworks[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 9489-9502.

[35] Bernal J, Sánchez J, Vilariño F. Towards automatic polyp detection with a polyp appearance model[J]. Pattern Recognition, 2012, 45(9): 3166-3182.

[36] Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.

[37] Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9(2): 283-293.

[38] Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-SEG: A segmented polyp dataset[C]//26th International Conference on Multimedia Modeling. Heidelberg: Springer, 2020: 451-462.

[39] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks[C]//Proceedings of the Third Conference on Machine Learning and Systems. mlsys.org, 2020: 429-450.

[40] Collins L, Hassani H, Mokhtari A, et al. Exploiting shared representations for personalized federated learning[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 2089-2099.

[41] Oh J, Kim S, Yun S Y. FedBABU: Toward enhanced representation for federated image classification[C/OL]//10th International Conference on Learning Representations. 2022: 1-29. https://openreview.net/forum id=HuaYQfggn5u.

[42] Yao Dezhong, Pan Wanning, Dai Yutong, et al. FedGKD: Toward heterogeneous federated learning via global knowledge distillation[J]. IEEE Transactions on Computers, 2024, 73(1): 3-17.

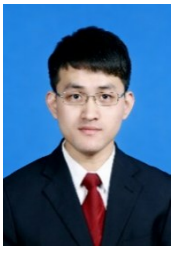
[43] Li Tian, Hu Shengyuan, Beirami A, et al. Ditto: Fair and robust federated learning through personalization[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 6357-6368.

[44] Lin Li, Liu Yixiang, Wu Jiewei, et al. FedLPPA: Learning personalized prompt and aggregation for federated weakly-supervised medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2025, 44(3): 1127-1139.

作者简介

杜晓刚　男,1985年出生于陕西省宝鸡市。现为陕西科技大学电子信息与人工智能学院副教授。主要研究方向为机器学习、计算机视觉、医学图像处理。

E-mail: duxiaogang@sust.edu.cn



刘统飞　男,1995年8月出生于陕西省榆林市。现为陕西科技大学电子信息与人工智能学院讲师、硕士生导师。主要研究方向为深度学习与医学、遥感图像智能解译。

E-mail: liutongfei@sust.edu.cn

魏　征　男,2001年4月出生于河南省许昌市。现为陕西科技大学电子信息与人工智能学院硕士研究生。主要研究方向为联邦学习、医学图像分割。

E-mail: 231612074@sust.edu.cn

王营博　男,1990年3月出生于陕西省宝鸡市。2021年6月毕业于北京理工大学,现为陕西科技大学电子信息与人工智能学院讲师。主要研究方向为图像复原及增强。

E-mail: wangyingbo@sust.edu.cn

雷　涛　男,1981年11月出生于陕西省渭南市。2011年在西北工业大学获得博士学位,现为陕西科技大学电子信息与人工智能学院教授、博士生导师。主要研究方向为图像处理、模式识别和计算机视觉等。中国电子学会会员编号:E190184479M。

E-mail: leitao@sust.edu.cn