专题：AI赋能通信网络

针对视觉–语言模型的个性化混合专家联邦微调框架

丁美琳[1]，靳晓嘉[2]，李荣盛[2]，赵东明[2]，高菲[1]，赵云凤[1]，仇超[1]，王晓飞[1]
（1. 天津大学智能与计算学部，天津 300350；
2. 中国移动通信集团天津有限公司人工智能产业研究院，天津 300020）

摘　要：在视觉–语言模型（vision-language model，VLM）广泛应用于图文检索、图像标注与视觉问答等任务的背景下，如何在保护跨行业数据隐私与应对计算资源受限的前提下，高效训练大规模模型成为通信运营商的关键挑战。联邦学习（federated learning，FL）通过在不共享原始数据的前提下进行分布式协同训练，缓解了数据孤岛和隐私问题。然而，VLM参数量庞大，训练过程中通信与计算开销高，加之FL环境中数据异质性强，导致全局模型泛化能力受限。为此，提出一种针对VLM的个性化混合专家联邦微调（federated personalized low-rank mixture of experts，FedLRM）框架，结合参数高效微调方法——低秩自适应（low-rank adaptation，LoRA）技术与门控机制，在本地构建融合全局与本地特征的混合专家（mixture of experts，MoE）架构，提升在细粒度数据层级的个性化表现。实验表明，在异构数据场景下，FedLRM相较于对比方法准确度最多提升1.88%，验证了其为个性化VLM的联邦优化提供了有效方案。

关键词：视觉–语言模型；低秩自适应；个性化联邦学习；混合专家
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000–0801.2025197

A personalized mixture-of-experts federated fine-tuning framework for vision-language models

DING Meilin[1], JIN Xiaojia[2], LI Rongsheng[2], ZHAO Dongming[2], GAO Fei[1],
ZHAO Yunfeng[1], QIU Chao[1], WANG Xiaofei[1]
1. College of Intelligence and Computing,Tianjin University, Tianjin 300350, China
2. Artificial Intelligence Industry Research Institute,China Mobile Communications Group Tianjin Co., Ltd., Tianjin 300020, China

Abstract: With the widespread application of vision-language model (VLM) in tasks such as image-text retrieval, image captioning, and visual question answering, efficiently training large-scale models under the constraints of cross-

收稿日期：2025–06–10；**修回日期：**2025–08–11
通信作者：仇超，chao.qiu@tju.edu.cn
基金项目：2024年天津市制造业高质量发展专项资金项目“自主智能算力的通用大模型关键技术研究及产业化应用示范”（No.24ZGZNGX00020）
Foundation Item: 2024 Tianjin Manufacturing High-quality Development Special Project “Research on Key Technologies of General Large Models of Autonomous Intelligent Computing Power and Industrial Application Demonstration” (No.24ZGZNGX00020)

industry data privacy and limited computational resources was recognized as a critical challenge for telecom operators. Federated learning (FL) was employed to address data silos and privacy concerns by enabling distributed collaborative training without sharing raw data. However, the massive number of parameters in VLM was found to lead to high communication and computation costs during training. Additionally, the strong data heterogeneity in FL environments was observed to limit the generalization capability of global models. To address these challenges, federated personalized low-rank mixture of experts (FedLRM), a personalized mixture-of-experts federated fine-tuning framework for VLM was proposed. It combined the parameter-efficient tuning method low-rank adaptation (LoRA) with a gating mechanism to build a local mixture of experts (MoE) architecture that fuses global and local features, enhancing fine-grained personalization. Experiments results show that in heterogeneous data scenarios, FedLRM improves accuracy by up to 1.88% compared to baseline methods, verifying that it provids an effective solution for the federated optimization of personalized vision-language models.

Key words: vision-language model, low-rank adaptation, personalized federated learning, mixture-of-experts

0 引言

视觉–语言模型（vision-language model，VLM）的出现打破了视觉与语言之间的壁垒[1]。通过联合学习图像的视觉特征与文本的语义信息，VLM能够更准确地捕捉图像中的细节并与文本描述进行匹配，从而在图文检索、图像标注、视觉问答等跨模态任务中展现出强大的性能。然而，这类大规模预训练模型在实际部署中仍面临一系列挑战，特别是在需要适配特定任务或场景时，通常依赖集中式数据收集与训练流程，受到隐私保护需求、数据孤岛现象及分布式计算资源受限等因素的制约[2]。为应对这些挑战，研究者开始探索分布式学习框架与VLM的结合，以去中心化的方式优化模型。联邦学习（federated learning，FL）[3]作为一种代表性分布式机器学习范式，通过仅共享模型参数而非原始数据，实现多个数据孤岛间的协同训练，有效缓解了隐私泄露与合规性风险，为分布式优化大规模VLM提供了可行的解决方案。

然而，VLM本身结构庞大、计算密集，FL在训练此类模型时面临两大主要挑战：一是通信成本高，频繁的参数同步使得网络带宽成为瓶颈；二是计算资源消耗大，部分边缘设备难以承载大模型的完整训练。此外，联邦环境中的数据通常是非独立同分布（non-independent and identically distributed，non-IID）的，不同客户端所拥有的数据具有显著差异，这使得统一的全局模型在多个客户端上的泛化能力受限。这一数据异质性挑战成为制约FL性能提升的关键瓶颈。

参数高效微调（parameter efficient fine-tuning，PEFT）技术[4]的提出则进一步提升了VLM在特定任务或数据集上的适应性，其针对特定需求进行少量参数调整以优化模型的参数配置，从而在保持泛化能力的同时，显著提升在特定任务上的准确性和效率。最近，一种仅由微调低秩矩阵组成的方法，低秩自适应（low-rank adaptation，LoRA）因其具有轻量化的特性而被广泛应用[5]。

同时，为提升FL在异构环境下的个性化能力，个性化FL（personalized federated learning，PFL）成为当前研究热点[6]。典型方法包括多任务学习[7-8]、动态客户端约束策略，如FedProx[9]等。最近，混合专家（mixture-of-experts，MoE）模型[10]也被引入PFL中[11-12]，通过门控机制将客户端的数据动态分配给更匹配的“专家”模型，从而更好地适应数据异质性带来的挑战。

在数联网技术飞速发展的当下，通信运营商纷纷聚焦跨行业数据融通赋能，带动产业发展，体现数据价值。但是，相关技术缺少跨行业的数据交互保护基础手段，对数据外部交互的可行性

和合规性信心和手段不足，严重阻碍了跨行业的数据融合应用；尤其是缺少行业化安全知识融通识别能力，数据专业背景知识构成复杂；在多维数据联合分析背景下，缺少行业化的知识语料实现对复杂行业数据的精准识别分析能力，严重阻碍数据融合分析效能和实现能力；缺少深度语义理解和认知分析手段，对大模型等数据智能化技术应用能力不足，难以实现基于智能化技术的自动数据分析能力调整和优化能力。

为了解决行业痛点问题，本文的目标是支持企业打造通信与汽车、政务、金融、文旅数据融合后的语言大模型系统，在业务、服务领域实现知识融合、案例融合，将VLM融入产业知识问答场景，并利用其提取的视觉和文本语义信息，实现个性化精准推荐。

在此基础上，本文提出了一种面向VLM[13]的个性化混合专家联邦微调（federated personalized low-rank mixture-of-experts，FedLRM）框架，旨在解决VLM在分布式训练环境下的3个关键挑战：数据隐私保护、通信计算效率优化以及non-IID数据分布下的模型个性化问题。首先，本框架基于FL框架构建分布式协同训练架构，在严格保护原始数据隐私的前提下完成知识融合，同时引入LoRA技术对VLM进行PEFT，降低训练参数计算量和通信参数量，缓解带宽压力与计算负载；其次，本框架创新性地设计了基于MoE技术的个性化增强机制，其中每个客户端的模型由本地门控网络、个性化的LoRA特征提取器（即本地专家），以及全局可共享的LoRA特征提取器（即全局专家）用于提取广义特征，从而形成本地MoE。在本地训练期间，门控网络能够针对每个数据样本动态生成自适应权重，并通过加权混合表示实现全局特征和个性化特征的融合，在细粒度数据级别增强了本地模型个性化，同时保证模型泛化能力。实验结果表明，本文提出的框架通过LoRA进行PEFT将传输参数量压缩至全量微调的16.4%。同时在异构数据设置下相对于基线方案有效提升其精度性能，展现出更好的适应能力。

1　问题建模和现有研究

1.1　问题建模

本研究的目标是在对比语言–图像预训练（contrastive language-image pre-training，CLIP）架构的基础上实现一个FedLRM。假设联邦微调框架由N个边缘设备和1个云服务器组成，分别表示为$X=\{1,2,\cdots,n,\cdots,N\}$和$C$。边缘设备集对应的数据集表示为$\{D_1,D_2,\cdots,D_N\}$，且数据分布不同，即$P(D_i)\neq P(D_j)$。该框架旨在解决全局模型无法适应具有non-IID数据的设备而导致的性能下降问题。因此，本文提出的框架通过允许训练多个个性化模型以服务于不同设备来修改优化目标，具体优化目标如下：

$$\min_{\Theta}\sum_{n=1}^{N}F_n(\theta_n;D_n) \tag{1}$$

其中，$\Theta=\{\theta_1,\theta_2,\cdots,\theta_N\}$是一组个性化模型，$\theta_n\in\mathbf{R}^d$表示第$n$个设备的个性化模型参数，$D_n$表示第$n$个设备对应的数据集，函数$F_n$基于$\Theta$同时最小化多个本地损失函数。

1.2　CLIP架构

给定一个预训练的CLIP主干网络，它包括一个视觉编码器$\theta_{\mathrm{v}}(\cdot)$和文本编码器$\theta_{\mathrm{t}}(\cdot)$。当处理基于VLM的分类任务时，给定一组$K$个候选类，创建对于类别的文本描述，即所谓的提示，其中每个提示对应一个类，如令表示“a photo of a [class name]”的标记化版本，$k=1,2,\cdots,K$，令$t_k=\theta_{\mathrm{t}}(c_k)$表示$c_k$相应的规范化文本嵌入表示，其中$\theta_{\mathrm{t}}$为文本编码器的参数。类似地，对于边缘设备$C_n$的每个样本图像$x_{n,i},i=1,2,\cdots,m_n$，使用视觉编码器$\theta_{\mathrm{v}}$:$f_{n,i}=\theta_{\mathrm{v}}(x_{n,i})$，将其投影到相同维数的归一化嵌入空间上，以获得图像嵌入表示。将每个文本嵌入t_k与测试图像$x_{n,i}$的视觉嵌入$f_{n,i}$相匹配，进而通

过测量它们的余弦相似度，获得预测得分：

$$l_{n,i,k}=\cos(f_{n,i},t_k) \tag{2}$$

相应地，产生了一个概率预测，即在给定测试输入x_i的情况下，类别k的后验softmax概率表示为：

$$p_{n,i,k}=\frac{\exp(l_{n,i,k}/\tau)}{\sum_{j=1}^{K}\exp(l_{n,i,j}/\tau)} \tag{3}$$

其中，τ为softmax温度参数。

在少样本设置中，为了进一步调整这些模型，假设对于每个目标类别有M/K个标记样本，即所谓的支持集。M表示支持样本的总数，M/K（每个类的标记样本数量）通常很小（小于16）。令$y_{n,i}$表示有标签的支持图像$x_{n,i}$的独热编码标签，即如果图像$x_{n,i}$属于类别k则$y_{n,i,k}$为1；否则$y_{n,i,k}$为0。然后，训练过程中需要优化最小化交叉熵（cross-entropy，CE）损失：

$$L_n=-\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K}y_{n,i,k}\log p_{n,i,k} \tag{4}$$

可以通过微调输入提示$c_k, k=1,2,\cdots,K$；或通过更新一组附加参数，如适配器。其中CLIP与每种方法的结合都带有独特的温度参数τ，该温度参数τ与预训练期间的可学习参数一起进行优化。

2 框架设计

FedLRM框架采用模块化设计，通过两个核心技术创新实现VLM的高效联邦微调：首先，采用FL架构实现隐私保护的VLM分布式训练，同时引入LoRA技术，通过优化低秩矩阵替代全参数更新，显著降低通信和计算开销；其次，设计FedLRM，包含全局共享和本地个性化的LoRA专家模块，通过动态门控网络实现样本级特征融合。该框架的设计在保证数据隐私的同时，实现了参数高效、知识共享和个性化增强的闭环优化。

2.1 基于LoRA的高效微调

针对CLIP在FL环境中所带来的高计算与通信开销问题，本文在边缘设备本地更新阶段引入参数高效的微调方法——LoRA。LoRA通过在模型中嵌入少量可训练的低秩矩阵，在冻结原有权重的基础上，仅更新极小部分参数，从而显著降低了本地训练过程中的计算复杂度与通信成本，适用于FL场景下的大规模模型轻量化适配。同时，LoRA具备良好的迁移能力，使模型能够快速适应联邦环境中异构的下游任务，有效提升系统的可扩展性与个性化建模能力。

首先，本文应用LoRA对每个边缘设备上部署的CLIP预训练模型的图像编码器进行微调。联邦框架中基于LoRA的高效微调流程如图1所示。对边缘设备n，其图像编码器和文本编码器的预训练参数$\boldsymbol{W}\in\mathbf{R}^{d_1\times d_2}$冻结，LoRA将图像编码器的预训练权重的增量更新建模为两个小矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的乘积。$\boldsymbol{A}$、$\boldsymbol{B}$的秩为r，可基于下游任务进行调整。对于边缘设备n，对于图像样本x，其嵌入向量维度为d_1，预训练模型应用LoRA模块后的修改前向传播过程如下：

$$h=\theta_{\mathrm{v}}x+\gamma\Delta\theta_{\mathrm{v}}=\theta_{\mathrm{v}}x+\gamma\boldsymbol{BA}x \tag{5}$$

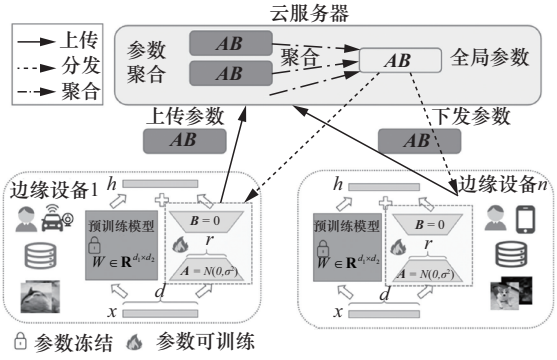


图1 联邦框架中基于LoRA的高效微调流程

其中，$\boldsymbol{A}\in\boldsymbol{R}^{r\times d_2}$，$\boldsymbol{B}\in\boldsymbol{R}^{d_1\times r}$，$\Delta\boldsymbol{\theta}_{\mathrm{v}}\in\boldsymbol{R}^{d_1\times d_2}$，其中$r$为秩，且通常$r<<\{d_1,d_2\}$，$d_1$、$d_2$分别表示神经网络的输入和输出维度，基于秩$r$分解的目的是减少矩阵的维度，同时保留其主要的特征和结构。γ为比例因子，$h\in\boldsymbol{R}^{d_2}$为输出向量。$\boldsymbol{A}$中的值通过Kaiming初始化进行随机初始化，而$\boldsymbol{B}$则用零填充进行初始化。这意味着在训练之前没有增量更新，因此输出是保持不变的。在反向传

播的过程中，因预训练参数都保持冻结，因此式（4）的交叉熵损失只更新$\boldsymbol{A}$、$\boldsymbol{B}$参数。

在FL本地训练过程中引入LoRA的高效微调，通过低秩矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的乘积建模权重的增量更新，仅需更新极少量参数，大幅减少本地训练所需要的计算资源和通信开销。这种特性使其适用于资源受限的边缘设备。同时，由于更新聚焦于边缘设备最关键的低秩子空间，模型能更快收敛到适应本地数据分布的解，提升个性化效率。

2.2　个性化增强的MoE联邦微调

为进一步提升个性化能力，本文基于MoE技术对模型的泛化和个性化参数进行解耦，并在本地进行样本级别的微调。个性化增强的MoE联邦微调框架如图2所示。具体而言，本文框架包括两个LoRA模块：全局LoRA特征提取器和个性化LoRA特征提取器。全局LoRA上传到云端，并通过联邦平均（federated averaging，FedAvg）算法进行更新；而个性化LoRA则在每个客户端本地进行微调。为了进一步提升个性化能力，框架结合了MoE技术，在本地动态更新全局和个性化LoRA模块。

2.2.1　全局LoRA模块

全局LoRA参数通过LoRA低秩矩阵$\boldsymbol{A}$和$\boldsymbol{B}$进行更新。每个客户端在本地进行全局LoRA微调，并上传更新后的全局LoRA参数至云端。在每次训练轮次结束时，云端通过FedAvg聚合更新所有客户端的全局LoRA参数，得到新的全局LoRA参数θ^{g}：

$$\theta^{\mathrm{g}}=\sum_{n=1}^{N}\frac{w_n}{\sum_{m=1}^{N}w_m}\theta_n^{\mathrm{g}} \tag{6}$$

其中，θ_n^{g}表示客户端n所持有的全局LoRA权重，由低秩矩阵对$\boldsymbol{A}_n^{\mathrm{g}}$和$\boldsymbol{B}_n^{\mathrm{g}}$构成，即$\theta_n^{\mathrm{g}}=\{\boldsymbol{A}_n^{\mathrm{g}},\boldsymbol{B}_n^{\mathrm{g}}\}$。客户端$n$的样本量为$w_n$，该方式可以兼顾数据量大小对模型更新的影响，提升模型的全局泛化能力。

2.2.2　个性化LoRA模块

每个客户端的个性化LoRA参数θ_n^L通过LoRA低秩矩阵$\boldsymbol{A}_n^L$和$\boldsymbol{B}_n^L$进行更新。个性化LoRA模块仅在客户端本地进行微调，不会上传至云端。个性化LoRA参数的计算式如下：

$$\theta_n^L=\boldsymbol{B}_n^L\boldsymbol{A}_n^L \tag{7}$$

2.2.3　MoE动态融合

在个性化微调过程中，框架结合了MoE技术，以实现更加精细的个性化建模。门控网络根据每个数据样本$x_{n,i}$动态选择全局LoRA模块和个性化LoRA模块，从而决定如何将全局和个性化特征进行加权混合。对于每个输入样本$x_{n,i}$，全局LoRA和个性化LoRA的加权输出分别为：

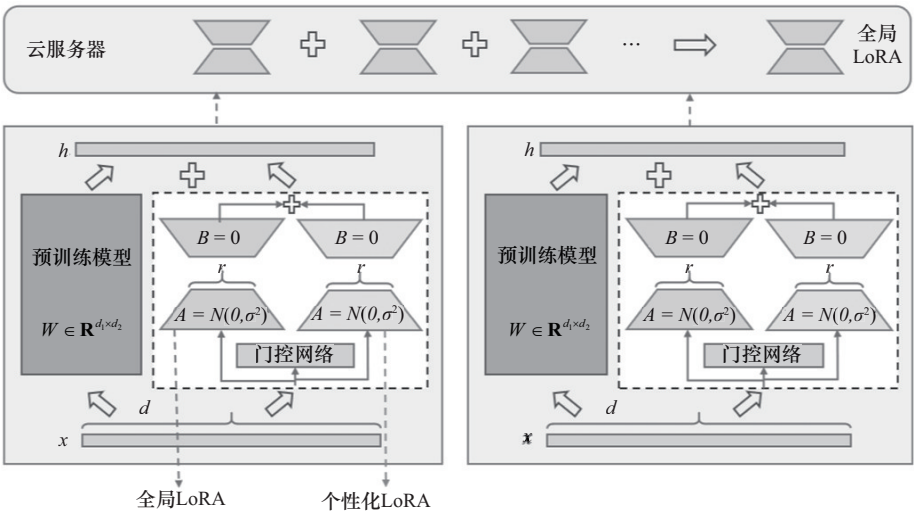


图2　个性化增强的MoE联邦微调框架

$$y_{n,i}^{\mathrm{g}}=g(x_{n,i})\cdot\theta_n^{\mathrm{g}}(x_{n,i}) \quad (8)$$

$$y_{n,i}^{\mathrm{L}}=h(x_{n,i})\cdot\theta_n^{\mathrm{L}}(x_{n,i}) \quad (9)$$

其中，$g(x_{n,i})$为全局LoRA的门控函数，表示全局LoRA模块的加权比例，$\theta_n^{\mathrm{g}}(x_{n,i})$为基于样本$x_{n,i}$，全局LoRA的输出。$h(x_{n,i})=1-g(x_{n,i})$为个性化LoRA的门控函数，表示个性化LoRA模块的加权比例，$\theta_n^{L}(x_{n,i})$为基于样本$x_{n,i}$的个性化LoRA输出。本研究使用的门控函数是单层的线性层实现轻量级的特征融合。最终，输出$\hat{y}_{n,i}$为全局和个性化特征的加权融合：

$$\hat{y}_{n,i}=y_{n,i}^{\mathrm{g}}+y_{n,i}^{L}=g(x_{n,i})\cdot\theta_n^{\mathrm{g}}(x_{n,i})+h(x_{n,i})\cdot\theta_n^{L}(x_{n,i}) \quad (10)$$

整体的训练过程可以表示如下。

（1）客户端在本地进行LoRA微调，基于MoE技术同时更新全局LoRA参数θ_n^{g}和个性化LoRA参数θ_n^{L}，以及对应的门控函数。

（2）每个客户端将更新后的全局LoRA参数θ_n^{g}上传至云端，个性化LoRA参数θ_n^{L}不上传。

（3）云端对全局LoRA参数进行FedAvg聚合，得到更新后的全局LoRA参数。

（4）在接下来的训练轮次中，客户端使用全局LoRA参数θ^{g}和本地微调后的个性化LoRA参数θ_n^{L}继续进行MOE融合训练。

3 实验与分析

本文对所提出的FedLRM框架相较于多种现有方法的性能表现进行了验证。实验主要从不同数据集配置下的适应能力、算法带来的性能与训练效率提升，以及关键参数对实际效果的影响等进行分析。

3.1 实验设置

本文在图像分类任务中对所提出的方法进行评估，选取Caltech101[14]和CIFAR-10[15]两个具有代表性的数据集。Caltech101数据集包含101个对象类别以及一个背景类，共计约9 000张中等分辨率图像，其中训练集和测试集分别包含约6 060张和3 086张图像。CIFAR-10是一个标准的图像分类基准数据集，由10个类别共60 000张32×32彩色图像组成，其中训练集和测试集分别包含50 000和10 000张图像。在特征提取方面，本文采用了CLIP提供的预训练的ViT-B/16模型。为保证输入一致性，所有图像在送入模型前均被标准化至224×224分辨率。

本文系统中共配置10个边缘设备，采用non-IID策略进行数据划分，以模拟实际场景中的数据异质性。具体划分方式为：首先根据类别标签对数据进行排序，并将其划分为多个数据片段；然后，每个边缘设备随机选取若干数据片段，从而保证各边缘设备拥有不同类别的数据，实现类别分布的不均衡。为满足小样本学习的设置，大部分实验中每个边缘设备每类数据抽取16个样本用于训练。

本实验在Linux操作系统环境下开展，基于Python 3.8与PyTorch 1.8.1实现，所用硬件平台为配备A100 SXM2 80 GB显存的GPU。边缘设备训练参数设置如下：批量大小为32，学习率设定为0.002，采用cosine学习率衰减策略。训练过程包含20轮全局迭代，每轮中各边缘设备执行10轮本地训练。

3.2 性能实验

本文探索了现有的先进FL方法，并通过在本地训练中配置LoRA对其进行了改进，最终选择具有代表性的FL方法：基于LoRA的联邦平均（LoRA-based federated averaging，LoRAFL）算法、基于LoRA的联邦近端（LoRA-based federated proximal，LoRAProx）算法、基于LoRA的自适应个性化联邦学习（LoRA-based adaptive personalized federated learning，LoRAAPFL）作为基准与本文提出的FedLRM进行图像分类任务的精度性能比较。

（1）LoRAFL：该方法采用基于数据量的传统加权平均策略[16]，系统中边缘设备协同训练一个

统一的全局 LoRA 模型，即为 FedLRM 移除门控网络和个性化 LoRA 模块的消融方案，用于验证 FedLRM 的性能增益。

（2）LoRAProx：该方法在 LoRAFL 的基础上，引入了一个邻近项到本地目标函数中，以约束本地更新的方向[9]，通过邻近项约束解决数据异构性的个性化方案为验证 FedLRM 框架的创新性提供可靠参照。

（3）LoRAAPFL：该方法将自适应个性化联邦学习（adaptive personalized federated learning，APFL）[17]与 LoRA 相结合，通过在本地训练中同时优化全局模型和个性化模型的低秩参数，并使用固定权重 0.5 平衡两者的更新，从而在保证参数高效性的同时提升模型在异构数据下的适应能力。该基线方案用来凸显 FedLRM 动态门控网络设计的自适应优势。

图 3 和图 4 分别展示了 Caltech101 测试集性能和 CIFAR-10 测试集性能，分析了在每类样本数分别为 1、4 和 16 时的 3 种小样本场景下，各方法的平均准确率性能。平均准确率是系统中边缘设备测试准确率平均加权。

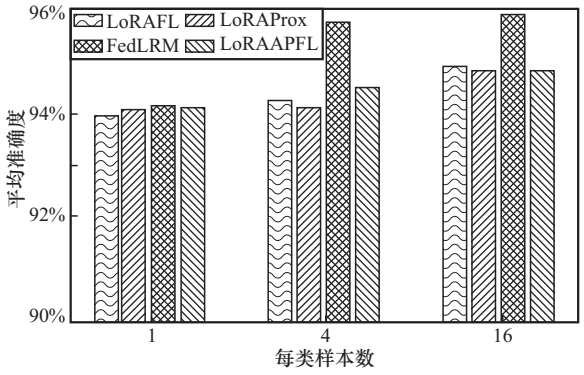


图 3　Caltech101 测试集性能

图 3 中，在 Caltech101 数据集上，本文所提方法在所有样本数量设定中均取得了最优性能，显示出较强的泛化能力与对数据异质性的适应性。在 3 种小样本场景下，FedLRM 凭借其 MoE 机制与 PEFT 策略，相对于 LoRAFL 和 LoRAProx 性能分别最高提升约 1.64 个百分点和 1.74 个百分点。相较于采用固定权重融合的 LoRAAPFL，FedLRM 通过动态门控网络实现的样本级自适应特征融合，在样本数为 4 和 16 时下分别取得了 1.3 个百分点和 1.1 个百分点的额外性能提升，这一差异验证了动态路由机制在细粒度个性化建模中的重要性。随着每类样本数量的增加，基本所有方法的准确率均有所提升，但 FedLRM 性能始终优于两种基线方案，特别是在 4 样本为 4 的设定下，FedLRM（95.76%）相对于 LoRAProx（94.11%）性能提升显著，显示出更强的模型表达能力与可迁移性。

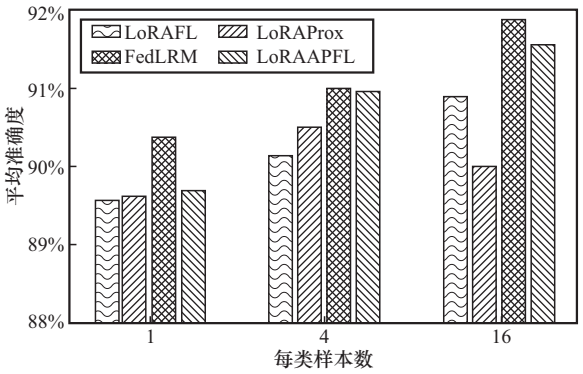


图 4　CIFAR-10 测试集性能

图 4 中，在 CIFAR-10 数据集上，FedLRM 同样展现出显著优势。在每类样本数为 1 的设定下，FedLRM 提高准确率至 90.37%，相比于 LoRAFL 的 89.56% 有明显提升，说明其在极低数据条件下也能有效捕捉类间特征差异。相较于采用固定融合权重的 LoRAAPFL，FedLRM 的动态门控机制带来了性能提升，验证了自适应特征融合对模型能力的增强作用。在样本数为 4 和 16 时，FedLRM 分别达到 91.01% 和 91.88%，尤其在 16 样本场景中，相较于 LoRAFL（90.89%）和 LoRAProx（90%）的退化趋势，FedLRM 性能最高提升 1.88 个百分点，展现出更好的稳定性与泛化性。

3.3　可训练参数量对比

本节对比了在下游任务数据集上，每个客户

端采用不同方式进行微调时的训练参数量，参数量对比见表1。其中包括本文提出的FedLRM联邦框架分别在不同低秩配置（r=1、r=4、r=16）下的训练参数量，以及传统的全量微调所需的参数量。全量微调需要对预训练模型中的所有参数进行更新，这对通信和计算资源构成了巨大压力。而在FedLRM中，仅需训练3个模块的参数，即全局LoRA、个性化LoRA自适应矩阵、门控网络，其中固定个性化LoRA秩为4，改变存在交互的全局LoRA的秩，门控网络由输出维度仅为2的线性层组成。FedLRM训练参数量远低于全量微调。以256维隐藏层为例，FedLRM在r=1时仅需36 864个参数，r=16时也仅为129 024个，仍显著低于全量微调。因此，FedLRM在不同配置下均展现出良好的参数高效性，从而极大地降低了通信开销和客户端负担。此外，FedLRM具备灵活性，可以根据实际任务需求选择合适的低秩参数r，实现模型容量与资源开销之间的平衡。

表1　参数量对比

秩	参数量
r=1	36 864
r=4	55 296
r=16	129 024
全量微调	786 432

3.4　超参实验

为了进一步探究全局LoRA模块中秩大小对模型性能的影响，本文设置了不同的秩值在Caltech101数据集上进行实验，具体包括r=1, 2, 4, 8, 16, 32，并记录了对应的平均准确率表现。LoRA秩的超参实验结果如图5所示。

从图5中可以观察到，随着LoRA秩的增加，模型的平均准确率整体呈现出显著上升的趋势。在低秩设置（如r=1和r=2）时，模型准确率相对较低，表明过低的秩限制模型对任务相关特征的建模能力。

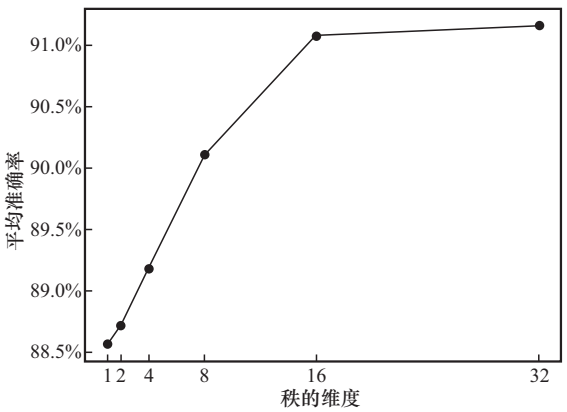


图5　LoRA秩的超参实验结果

当秩值逐步增大至r=4和r=8时，模型准确率显著提升，说明较高秩值可以增强模型在多模态特征融合方面的表达能力。进一步地，r=16时模型准确率达到了91.08%，接近饱和。此后即使秩继续提升至r=32，准确率的增幅变得非常有限，为91.16%。这表明在该任务中，LoRA秩在r=16左右已经能够充分捕获关键的子空间信息，进一步增加秩值带来的收益较小，可能还会引入额外计算开销。

综上所述，该实验表明合理选择LoRA秩值对于模型性能提升具有重要意义，在保持模型参数高效性的同时，还能显著增强个性化联邦微调后VLM的表达能力。在本实验设定中，r=16是一个性能与效率兼顾的较优选择。

4　结束语

本文聚焦于VLM在FL场景中的个性化优化问题，旨在解决大规模模型在隐私保护与资源受限环境下训练效率低、泛化能力弱的问题。针对传统FL在异构数据分布下难以兼顾个性化与泛化的问题，本文提出了FedLRM框架，一种基于VLM的MoE个性化联邦微调方法。该框架在本地构建可动态路由的专家网络，有效整合全局共享知识与本地个性化特征，从而在保证通信效率的同时，实现模型个性化性能的显著提升。

本文实验结果表明，FedLRM 在 Caltech101 和 CIFAR-10 典型数据集上均取得了较基线方案更高的性能，特别是在样本数量稀缺与客户端数据高度异构的设置下，FedLRM 准确率较其基线方案性能最高提升 1.3 个百分点和 1.88 个百分点。进一步分析表明，FedLRM 在提升个体性能的同时，仍保持较好的全局泛化能力。综上所述，本文提出的方法为构建高效、可扩展且具备鲁棒性的 VLM 个性化联邦微调优化提供了新的解决思路，为后续研究奠定了基础。

参考文献：

[1] ZHANG J Y, HUANG J X, JIN S, et al. Vision-language models for vision tasks: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5625-5644.

[2] QI P, CHIARO D, GUZZO A, et al. Model aggregation techniques in federated learning: a comprehensive survey[J]. Future Generation Computer Systems, 2024, 150: 272-293.

[3] WEN J, ZHANG Z X, LAN Y, et al. A survey on federated learning: challenges and applications[J]. International Journal of Machine Learning and Cybernetics, 2023, 14(2): 513-535.

[4] HE H Y, CAI J F, ZHANG J, et al. Sensitivity-aware visual parameter-efficient fine-tuning[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 11791-11801.

[5] WANG S W, YU L X, LI J. Lora-ga: low-rank adaptation with gradient approximation[J]. arXiv preprint, 2024, arXiv: 2407.05000

[6] WEI K, LI J, MA C, et al. Personalized federated learning with differential privacy and convergence guarantee[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 4488-4503.

[7] HAO J F, CHEN P, CHEN J, et al. Multi-task federated learning-based system anomaly detection and multi-classification for microservices architecture[J]. Future Generation Computer Systems, 2024, 159: 77-90.

[8] YE C Y, ZHENG H, HU Z G, et al. PFedSA: personalized federated multi-task learning *via* similarity awareness[C]//Proceedings of the 2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Piscataway: IEEE Press, 2023: 480-488.

[9] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. arXiv preprint, 2018, arXiv: 1812.06127

[10] CHEN T L, CHEN X X, DU X Z, et al. AdaMV-MoE: adaptive multi-task vision mixture-of-experts[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 17300-17311.

[11] FENG Y, GENG Y L, ZHU Y F, et al. PM-MOE: mixture of experts on private model parameters for personalized federated learning[C]//Proceedings of the ACM on Web Conference 2025. New York: ACM Press, 2025: 134-146.

[12] MEI H, CAI D, ZHOU A, et al. FedMoE: personalized federated learning via heterogeneous mixture of experts[J]. arXiv preprint, 2024, arXiv:2408.11304.

[13] RADFORD A, KIM J W, HALLACY C, et al.Learning transferable visual models from natural language supervision[J]. arXiv preprint, 2021, arXiv:2103.00020

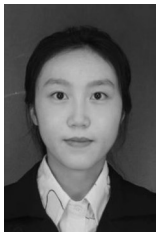
[14] BANSAL M, KUMAR M, SACHDEVA M, et al. Transfer learning for image classification using VGG19: Caltech-101 image data set[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(4): 3609-3620.

[15] KRIZHEVSKY A. Learning multiple layers of features from tiny images[D]. Toronto: University of Toronto, 2009.

[16] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the International Conference on Artificial Intelligence and Statistics, 2016.

[17] DENG Y, KAMANI M M, MAHDAVI M. Adaptive personalized federated learning[J]. arxiv preprint, 2020, arXiv: 2003.1346.

[作者简介]



丁美琳（2000– ），女，天津大学智能与计算学部在读，主要研究方向为边缘计算。

靳晓嘉（1972– ），男，博士，中国移动通信集团天津有限公司副总经理、正高级工程师，主要研究方向为 AI 大模型构建、大数据分析及网络规划等。

李荣盛（1977– ），男，中国移动通信集团天津有限公司人工智能产业研究院高级工程师，主要研究方向为自智网络、云计算、人工智能。

赵东明（1984– ），男，博士，中国移动通信集团天津有限公司人工智能产业研究院正高级工程师、技术总监/高级专家，主要研究方向为自然语言处理和大语言模型。

高菲（1997– ），女，天津大学智能与计算学部硕士生，主要研究方向为边缘计算和联邦学习。

赵云凤（1997– ），女，天津大学智能与计算学部博士生，主要研究方向为边缘计算、边缘智能和分布式机器学习。

仇超（1988– ），女，天津大学智能与计算学部副教授，主要研究方向为边缘计算、边缘智能和区块链。

王晓飞（1982– ），男，天津大学智能与计算学部教授，主要研究方向为边缘计算、边缘智能和边缘系统。