arXiv:2603.03007v1 [cs.LG] 3 Mar 2026

打破原型偏差循环：面向高度不平衡客户端的置信度感知联邦对比学习

Tian-Shuang Wu
TIANSHUANGWU@HHU.EDU.CN水利部水大数据技术重点实验室，计算机科学与软件工程学院，河海大学，中国南京

Shen-Huan Lyu ✉ LVSH@HHU.EDU.CN
水利部水大数据技术重点实验室，计算机科学与软件工程学院，河海大学，中国南京 计算机科学系，香港城市大学，中国香港 新型软件技术国家重点实验室，南京大学，中国南京

Ning Chen CHE-N-ING@HHU.EDU.CN
水利部水大数据技术重点实验室，河海大学计算机科学与软件工程学院，中国南京

Yi-Xiao He HEYX@NJUCM.EDU.CN
南京中医药大学人工智能与信息技术学院，中国南京

Bin Tang CSTB@HHU.EDU.CN
水利部水大数据技术重点实验室，河海大学计算机科学与软件工程学院，中国南京

Baoliu Ye
YEBL@NJU.EDU.CN新型软件技术国家重点实验室，南京大学，中国南京

Qingfu Zhang QINGFU.ZHANG@CITYU.EDU.HK
中国香港香港城市大学计算机科学系

摘要

客户端间的局部类别不平衡和数据异质性，常使基于原型的联邦对比学习陷入原型偏差循环：不平衡数据导致有偏本地原型，这些原型聚合后形成有偏全局原型，随后又被反复用作对比锚点，使误差在通信轮次中不断累积。为打破这一循环，我们提出置信度感知联邦对比学习（CAFedCL），这是一种新颖的框架，旨在改进原型聚合机制，并增强由原型引导的对比对齐。CAFedCL采用置信度感知聚合机制，利用预测不确定性降低高方差局部原型的权重。此外，我们还整合了针对少数类的生成式增强和几何一致性正则化，以稳定类间结构。从理论角度来看，我们提供了基于期望的分析，表明我们的方法能够降低估计方差，从而限制全局原型漂移并确保收敛。在类别不平衡程度和数据异质性水平各异的条件下进行的大量实验表明，CAFedCL始终优于具有代表性的联邦基线方法，在准确率和客户端公平性方面均表现更佳。

✉。通讯作者

© 预印本 2026。审阅中。请勿分发。

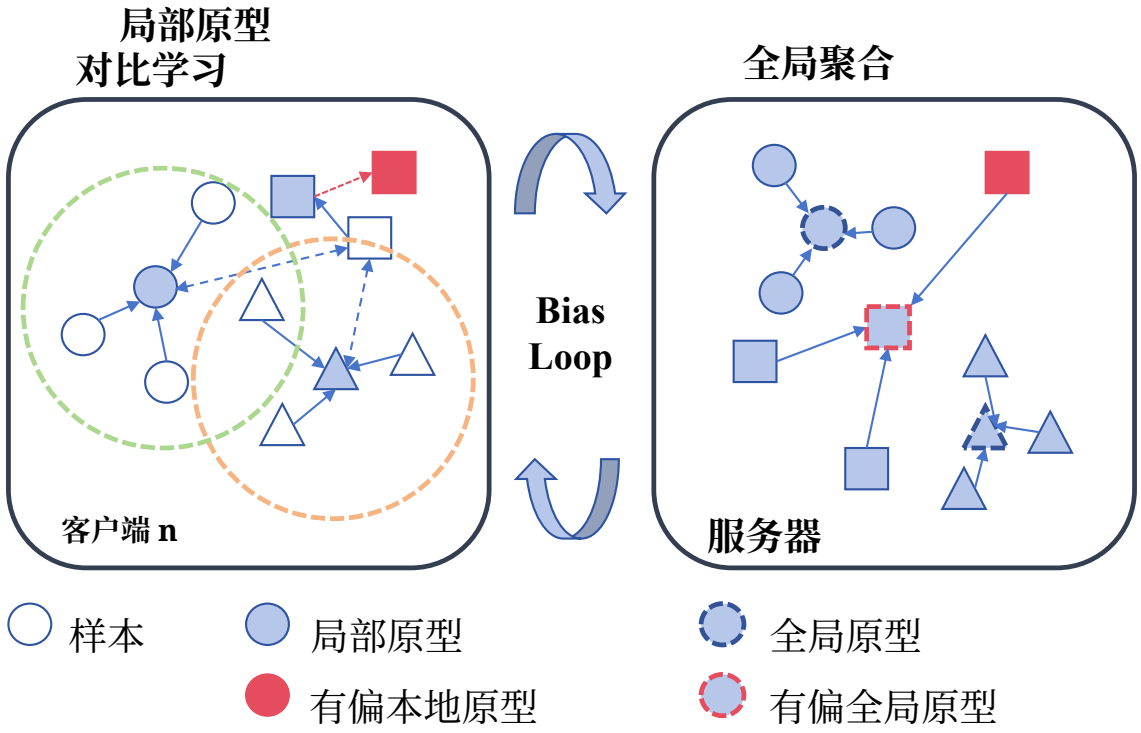

图 1：联邦对比学习中的偏差循环。

异质性水平各异的实验表明，CAFedCL在准确率和客户端公平性方面均始终优于具有代表性的联邦基线方法。

关键词： 联邦学习、对比学习、表示学习、类别不平衡

1 引言

联邦学习（FL）（McMahan等人，2017年）能够在不集中原始数据的情况下进行协作式模型训练，并已成为隐私保护分析的常见范式。然而在实践中，普通联邦学习常常受到统计异构性的挑战，尤其是标签分布偏斜和极端类别不平衡（长尾数据）（Imteaj等人，2021；Tan等人，2022a）。近期研究（Huang等人，2025）进一步表明，标签异构性经常与少数类的数据稀缺同时出现，从而导致泛化下降以及各客户端性能不均。尽管个性化联邦学习已取得进展，许多方法仍侧重于聚合通用模型参数，可能未能充分利用在偏斜局部监督下至关重要的标签结构信息（Kairouz等人，2021）。

为缓解客户端漂移，越来越多的研究从参数空间聚合转向表示空间对齐。联邦对比学习（FedCL）是其中一种代表性方向，通过正则化局部嵌入，使其趋向更加一致的全局表征。然而，标准对比学习通常依赖大量负

样本对或较大的批大小，而这在隐私保护联邦学习中难以维持。这促使研究者探索基于原型的联邦对比学习，其中原型是按类别划分的特征质心，可作为类别语义的轻量级摘要（Mu 等人，2023年； Tan 等人， 2022b）。通过将全局原型作为共享语义锚点进行通信，客户端无需传输单个样本，便能以较低的通信成本使局部表示与全局类别结构对齐（Tan 等人， 2022c）。除单原型方案外，多原型变体还能进一步刻画类内多样性，同时促进客户端间知识迁移（Qiao 等人，2023年）。

尽管这一范式颇具吸引力，但它隐含地假设聚合后的全局原型能够可靠地近似真实类别中心。我们认为，在极端不平衡和异质客户端标签质量的情况下，这一假设可能不成立；这类情况常见于医学筛查（阳性样本稀少）和工业缺陷检测（故障样本稀少）等应用。在这类情形下，原型误差可能在多个轮次中不断自我强化：有偏差或高方差的局部原型会污染聚合后的全局锚点，而被污染的锚点随后又会被用于引导后续对比更新。这一反馈循环可能逐步扭曲原型几何结构，并损害少数类判别能力。我们将这一现象称为原型偏差循环（图1）。

为解决这一问题，我们提出了**CAFedCL**，这是一种针对极端不平衡和异构客户端质量设计的置信度感知联邦原型对比学习框架。其核心思想是将原型视为具有不确定性的估计，而非确定性目标。在每个客户端上，我们可选择使用条件生成对抗网络增强少数类，以缓解严重的数据稀缺；同时，采用原型引导的对比目标训练编码器，并辅以全局引导的几何正则项，以保持类间分离。在服务器端，我们以置信度加权聚合取代朴素平均，对原型和编码器参数进行聚合。每个客户端都会报告类别级置信度分数，这些分数结合了生成质量信号与基于验证的可靠性，因此在聚合时会降低不可靠的客户端-类别贡献的权重，从而获得更稳定的全局锚点。

本文的主要贡献总结如下：

- 我们识别出基于原型的联邦对比学习在极端不平衡下的一种失效机制，表明跨轮锚点复用与朴素平均如何放大原型误差，从而形成原型偏差循环。

- 我们提出CAFedCL，通过可选的尾部增强与几何正则化稳定少数类表征，并通过针对原型与模型参数的按类别置信度加权聚合缓解不可靠更新。

- 我们在多种非独立同分布和长尾设定下开展了广泛实验，结果表明，相较于强基线方法，准确率与客户端公平性均持续提升，且仅带来适度开销。

2 相关工作

2.1 联邦学习

联邦学习（FL）已成为一种隐私保护分布式学习范式，使多个客户端能够协作优化共享模型，而无需将原始数据传输至中央服务器（McMahan等人，2017年）。这种去中心化训练协议在数据敏感应用中日益重要，例如医学图像（Adnan等人，2022年；Liu等人，2021）以及物联网（IoT）系统（Wang等人，2025；Khan等人，2021）。在这些场景中，数据在资源约束下自然产生并存储于地理上分散的设备中（Imteaj等人，2021；Kairouz等人，2021）。

尽管前景广阔，现实世界中的联邦学习部署仍常受严重的系统异质性和统计异构性阻碍（Luo等人，2021），尤其是客户端数据的非独立同分布特性，这可能导致客户端漂移、目标不一致，并使泛化能力较集中式训练有所下降（Zhao等人，2018；Li等人，2020）。为应对这些挑战，研究者从优化和个性化两个角度取得了显著进展。具有代表性的解决方案包括：对本地更新进行正则化以稳定训练（Li等人，2020）、利用控制变量纠正客户端漂移（Karimireddy等人，2020），以及通过归一化聚合缓解目标不一致（Wang等人，2020；Diao等人，2024；Li 等，2021c）。此外，近期综述总结了相关进展，并指出了持续存在的开放问题，推动学界进一步开发适用于异质环境的稳健且通信高效的联邦算法（Kairouz等人，2021；Tan等人，2022a）。

2.2 联邦学习中的对比学习

对比学习是一种强大的范式，通过拉近语义相似的样本对、同时分离负样本来学习可迁移的表征，并已在自监督和监督设置中取得显著成功（Oord 等人，2018年；Chen 等人，2020年；He 等人，2020年）。受其泛化优势的启发，近期研究将对比目标引入联邦学习，以缓解非独立同分布数据下的表征错位，并提升对客户端漂移的鲁棒性（Zhang 等人，2021；Guo 等人，2023年；Zhuang 等人，2022年）。与单独优化任务特定损失相比，联邦对比学习（FedCL）提供了额外的表征层正则化，已成为异构联邦学习中一个颇具前景的研究方向。

现有的 FedCL 方法主要在通信与隐私约束下对比信号的构建与共享方式上有所不同。Tan 等人探索了基于预训练模型的联邦学习对比正则化（Tan 等人，2022c），而模型对比联邦学习则直接对比客户端模型以促进全局一致性（Li 等人，2021b）。原型中心方法进一步将类别级表征用作共享锚点（Dai 等人，2023年），包括用于原型聚合与广播的 FedProto（Tan 等人，2022b）以及采用原型对比目标的 FedProc（Mu 等人，2023年）。在单原型建模之外，Mp-FedCL 采用多原型来刻画类内多样性（Qiao 等人，2023年），而 FedTGP 引入可训练全局原型，并结合自适应间隔对比学习，以同时应对数据与模型异构性

（Zhang 等人，2024年）。此外，FedSC 提供了一种可证明的联邦自监督框架，该框架基于谱对比目标并具备差分隐私保护（Jing 等人，2024年），而 FedRCL 采用带有过度相似性惩罚的松弛对比损失，以防止在异质性下发生坍缩（Seo 等人，2024年）。这些工作凸显了对比学习在联邦表示对齐方面的有效性。

尽管取得了这些进展，但在类别不平衡和标签偏斜的设置下，FedCL 仍面临挑战，因为少数类样本稀缺，且客户端标签分布极不均衡。在此类情况下，朴素的对比学习和原型聚合可能会受到多数类特征的主导，从而产生有偏原型（Li 等人，2021a），削弱对少数类的判别能力，甚至通过迭代式全局聚合放大错误（Wang 等人，2021；Wu 等人，2023年；Li 等人，2021c）。因此，开发能够保留稳健少数类表示并稳定全局协调的不平衡感知 FedCL 机制，仍是一项重要但尚未得到充分探索的研究方向。

3 预备知识

3.1 联邦原型学习

我们考虑一个包含 K 个客户端的联邦系统，客户端索引为 $k \in \{1, \ldots, K\}$。客户端 k 持有私有数据集 $\mathcal{D}_k = \{(x_i,\ y_i)\}_{i=1}^{N_k}$，其标签为 $y_i \in \{1, \ldots, C\}$。令 $f(\cdot;\ \theta_k)$ 表示客户端编码器。联邦原型学习（FPL）交换轻量级的逐类别代表（原型），而非完整模型信息。对于类别 c，客户端 k 上的局部原型计算如下：

$$\mathbf{p}_{k,c} = \frac{1}{|\mathcal{D}_{k,c}|} \sum_{(x,y)\in\mathcal{D}_{k,c}} f(x;\theta_k), \tag{1}$$

其中，$\mathcal{D}_{k,c} \subseteq \mathcal{D}_k$ 包含类别 c 的样本，$|\mathcal{D}_{k,c}| =: n_{k,c}$。服务器聚合参与客户端的原型，以获得全局原型 $\mathcal{P}_g = \{\mathbf{p}_{g,c}\}_{c=1}^{C}$：

$$\mathbf{p}_{g,c} = \frac{\sum_{k\in S_t} n_{k,c}\mathbf{p}_{k,c}}{\sum_{k\in S_t} n_{k,c}}, \tag{2}$$

其中，S_t 是第 t 轮的参与客户端集合。当使用原型计算余弦相似度时，我们对嵌入和原型进行 ℓ_2 归一化。

3.2 不平衡条件下的估计偏差

在现实的联邦学习中，客户端数据通常呈非独立同分布且长尾分布。对于少数类，样本量 $n_{k,c}$ 可能极小，使得式（1）中的经验原型成为底层类别中心 $\boldsymbol{\mu}_{k,c}$ 的高方差估计量。

假设类内特征方差有界，则原型估计误差通常按以下规律变化

$$\mathbb{E}\big[\|\mathbf{p}_{k,c} - \boldsymbol{\mu}_{k,c}\|_2^2\big] \lesssim \frac{\sigma_{k,c}^2}{n_{k,c}}. \tag{3}$$

因此，尾部客户端-类别对会产生不可靠的原型。式（2）中的朴素聚合并未显式考虑这一可靠性差距，导致噪声本地原型污染全局锚点，而这些锚点会在原型引导的对比学习中被反复用于提供监督信号。

4 方法

本节提出 CAFedCL，这是一种针对极端标签不平衡和异构客户端设计的基于原型的联邦对比框架。我们的设计基于一个关键观察：在长尾客户端场景下，朴素的按类别原型聚合可能会向服务器引入高方差且有偏的锚点；这些锚点一经广播，就会在原型引导的对比学习中被反复使用，从而形成自我强化的原型偏差循环。为打破这一循环，我们引入按类别置信度机制，在原型聚合和模型聚合过程中降低不可靠客户端类别的权重，同时采用两种轻量级稳定方法来改善少数类表示。

4.1 面向联邦对齐的原型引导对比学习

在非独立同分布和长尾设定下，联邦学习常会遭遇表征错位问题，即本地编码器会逐渐偏向客户端特定特征。原型引导的联邦对比学习通过将全局原型用作轻量级语义锚点来解决这一问题：客户端将自身嵌入与共享类别参照对齐，而无需交换原始数据或完整的模型对数几率。

形式化地，对于客户端 k 上的样本 (x_i, y_i)，我们将其归一化嵌入记为 $\mathbf{z}_i = f(x_i; \theta_k)$（随后进行 ℓ_2 归一化）。服务器维护全局原型 $\mathcal{P}_g = \{\mathbf{p}_{g,c}\}_{c=1}^{C}$，其中每个 $\mathbf{p}_{g,c}$ 表示嵌入空间中的全局类别-c 中心。与纯粹的本地监督对比学习相比，当本地批次较小或某些客户端缺少部分类别时，原型引导尤其有效。

一种典型的原型引导对比目标，记为 $\mathcal{L}_{\text{I2P}}$（实例到原型），将每个嵌入与其类别原型对齐，同时排斥其他类别原型：

$$\mathcal{L}_{\text{I2P}} = -\frac{1}{|\mathcal{B}|} \sum_{(x_i,y_i)\in\mathcal{B}} \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{p}_{g,y_i})/\tau)}{\sum_{c=1}^{C} \exp(\text{sim}(\mathbf{z}_i, \mathbf{p}_{g,c})/\tau)}, \tag{4}$$

其中，$\mathcal{B}$ 是本地小批次，$\text{sim}(\cdot,\cdot)$表示余弦相似度，τ 是温度参数。最小化式（4）可将客户端嵌入拉向统一的全局锚点，从而缓解跨客户端表示漂移。

本地训练后，每个客户端根据式（**1**）计算按类别原型 $\mathbf{p}_{k,c}^{t+1}$。在计算对比相似度时，我们对嵌入和原型都进行 ℓ_2 归一化，即 $\tilde{\mathbf{z}} = \text{norm}(f(x; \theta_k))$ 和 $\tilde{\mathbf{p}}_{k,c} = \text{norm}(\mathbf{p}_{k,c})$，服务器则广播归一化全局原型 $\tilde{\mathbf{p}}_{g,c}$。这样可使对比目标与式（4）中的余弦相似度保持一致。

4.2 联邦对比学习中的原型偏差循环

第4.1 节表明，广播全局原型可以为实例与原型之间的对比提供轻量级语义锚点。然而，在极端标签不平衡且客户端质量参差不齐的情况下，局部原型可能存在噪声或偏差。聚合后，这些不完善的原型会成为全局锚点，并在公式（4）中被反复使用，因此误差可能持续存在并在多个轮次中不断累积，形成自我强化的原型偏差循环。

为明确说明这一点，我们固定一个类别 c，并跟踪全局原型 $\mathbf{p}_{g,c}^t$ 相对于底层类别中心 $\boldsymbol{\mu}_c$ 的均方误差：$E^t \triangleq \mathbb{E}\|\mathbf{p}_{g,c}^t - \boldsymbol{\mu}_c\|^2$.

假设 1（有界类内方差）。对于任意类别 c，在客户端 k上，嵌入围绕客户端特定中心 $\boldsymbol{\mu}_{k,c}$ 的方差有界：$\mathbb{E}_{x\sim\mathbf{p}_{k,\,c}}\|f_{\theta_k}(x) - \boldsymbol{\mu}_{k,c}\|^2 \le \sigma_{k,c}^2$.

在假设 1 下，局部原型的质量从根本上受有效样本量限制。回顾式（1）中的局部原型定义，可得以下估计界。

Le引理 1。在假设 *1* 下，局部原型的估计误差满足 *es*

$$\mathbb{E}\|\mathbf{p}_{k,c} - \boldsymbol{\mu}_{k,c}\|^2 \le \frac{\sigma_{k,c}^2}{n_{k,c}^{\text{eff}}}, \tag{5}$$

其中，$n_{k,\,c}^{\text{eff}} = n_{k,\,c} + \gamma m_{k,\,c}$ 为有效样本量，$m_{k,\,c}$ 为合成样本数量（如有），而 $\gamma \in [0,1]$ 用于折减其可靠性。

引理 1表明，$n_{k,\,c}^{\text{eff}}$ 较小的尾部客户端-类别对会产生高方差原型。因此，朴素聚合可能较为脆弱：即使只有少数不可靠的原型，也可能向全局锚点引入显著噪声。

这种效应会因锚点复用而进一步放大。在原型引导的 FedCL 中，全局原型$\mathbf{p}_{g,\,c}^t$ 会被广播，并在式（4）中反复复用，从而持续将局部表示拉向当前锚点。作为一阶抽象，我们用以下压缩动力学对其建模：

$$\mathbf{p}_{k,c}^{t+1} = (1-\rho)\tilde{\boldsymbol{\mu}}_{k,c} + \rho\,\mathbf{p}_{g,c}^t + \boldsymbol{\xi}_{k,c}^{t+1}, \tag{6}$$

其中，$\tilde{\boldsymbol{\mu}}_{k,\,c}$ 表示由本地数据和优化所诱导的类别-c 中心，且不复用先前的全局锚点；$\rho \in [0,1)$ 概括了式（4）中由原型复用引入的锚点吸引强度；$\boldsymbol{\xi}_{k\,c}^{t+1}$，表示零均值噪声，其满足 $\mathbb{E}\|\boldsymbol{\xi}_{k}^{t+1}{}_{,\,c}\|^2 \le \sigma_{k,\,c}^2/n_{k,\,c}^{\text{eff}}$。

在服务器上，全局原型通过对参与客户端 S_t 进行按类别加权聚合得到：

$$\mathbf{p}_{g,c}^{t+1} = \sum_{k\in S_t} \alpha_{k,c}^t\,\mathbf{p}_{k,c}^{t+1}, \quad \alpha_{k,c}^t \ge 0, \quad \sum_{k\in S_t}\alpha_{k,c}^t = 1. \tag{7}$$

命题 2。设 $E^t = \mathbb{E}\|\mathbf{p}_{g,\,c}^t - \boldsymbol{\mu}_c\|^2$。在假设 *1* 及上述动力学条件下，有

$$E^{t+1} \;\le\; \underbrace{\rho^2 E^t}_{\textit{anchor feedback}} + \underbrace{2(1-\rho)^2\,\Delta_{\text{bias}}^t}_{\textit{heterogeneity gap}} + \underbrace{2\sum_{k\in S_t}(\alpha_{k,c}^t)^2\frac{\sigma_{k,c}^2}{n_{k,c}^{\text{eff}}}}_{\textit{variance injection}}, \tag{8}$$

其中，$\Delta_{\text{bias}}^t = \big\|\sum_{k\in S_t}\alpha_{k,\,c}^t\,(\tilde{\boldsymbol{\mu}}_{k,\,c} - \boldsymbol{\mu}_c)\big\|^2$ 衡量了客户端之间类别中心的非独立同分布偏移。

注释 3。命题 *2* 将全局原型误差分解为三个来源。锚点反馈 *(ρ^2E^t)*表明，一旦出现有偏全局原型，它就会传播到后续轮次，因为 $\mathbf{p}_{g,c}^t$ 会在式*(4)*中被重复用作对比锚点。异质性差距

(Δ^t_{bias})刻画了非独立同分布本地训练导致的不可避免的中心偏移。关键的是，方差注入项随$(\alpha^t_{k,c})^2/n^{\text{eff}}_{k,c}$变化，并会被具有较小$n^{\text{eff}}_{k,c}$的尾部客户端-类别对放大，使朴素聚合在长尾数据下变得脆弱。*CAFedCL*通过在原型聚合中将$\alpha^t_{k,c}$替换为类别级置信度权重$\text{conf}^t_{k,c}$，针对这一主导项进行处理，从而降低不可靠的客户端–类别原型对下一轮对比训练所用全局锚点的贡献。

所有证明均见附录。一个简单的观察是，可靠性感知加权（例如，与$n^{\text{eff}}_{k,c}$成比例）会降低方差注入项中的有效平方权重。我们在附录中给出了形式化陈述。

4.3 类别级置信度加权聚合

命题2 表明，在长尾数据下，原型偏差循环主要由方差注入项放大，而该项由不可靠的客户端–类别对主导。为抑制这一效应，CAFedCL 为每个客户端原型分配一个类别级置信度分数，并在聚合时将其用作可靠性权重。不同于纯粹基于频率的加权，我们的置信度结合了多种互补信号，因此即使客户端的样本数并非极少，只要其原型质量不确定，也可以降低该客户端的权重。

每个客户端 k 报告 $\text{conf}_{k,c}$，用于估计其对类别 c的可靠性：

$$\text{conf}_{k,c} = w_1 \,\text{conf}^{\text{data}}_{k,c} + w_2 \,\text{conf}^{\text{gen}}_{k,c} + w_3 \,\text{conf}^{\text{val}}_{k,c}, \tag{9}$$

其中 $w_1 + w_2 + w_3 = 1$。我们将 $\text{conf}^{\text{data}}_{k,c} \propto n^{\text{eff}}_{k,c}$ 设为有效样本可用性，将 $\text{conf}^{\text{val}}_{k,c} =$ 设为 $\exp(-\beta u_{k,c})$，其中使用小型验证集划分上的类别级预测不确定性$u_{k,c}$；启用尾部增强时，将 $\text{conf}^{\text{gen}}_{k,c}$,设为生成样本上的平均判别器分数（否则设为 0）。禁用生成器时，我们设定 $w_2 = 0$，并相应地重新归一化（w_1, w_3）。

服务器通过置信度加权平均更新每个全局类别原型 g:

$$\mathbf{p}^{t+1}_{g,c} = \frac{\sum_{k\in S_t} \text{conf}^t_{k,c}\, \mathbf{p}^{t+1}_{k,c}}{\sum_{k\in S_t} \text{conf}^t_{k,c} + \epsilon}. \tag{10}$$

与朴素平均相比，式（10）缩小了不可靠客户端–类别对的有效平方权重，从而抑制了式（8）中的方差注入效应，并减弱跨轮次误差累积。

除原型外，我们还使用整体客户端置信度来聚合编码器参数：

$$\text{Conf}_k = \frac{1}{C}\sum_{c=1}^{C} \text{conf}_{k,c}, \tag{11}$$

$$\theta^{t+1} = \frac{\sum_{k\in S_t} \text{Conf}^t_k\, \theta^{t+1}_k}{\sum_{k\in S_t} \text{Conf}^t_k + \epsilon}. \tag{12}$$

该聚合会下调噪声客户端的权重，并在异质长尾训练下提升全局优化的稳定性。为保持稳定性，我们对 $\text{conf}_{k,c}$ 应用逐类别裁剪和重新归一化。

算法 1 置信度感知联邦对比学习（CAFedCL） 输入：客户端 K，类别 C，轮次 T；本地训练轮次 E；损失权重 λ_{align}、λ_{geo}；置信度权重（w_1、w_2、w_3）；小常数 ϵ。 1：初始化全局编码器 θ^0 和全局原型 $\{\mathbf{p}_g^0$、$_c\}_{c=1}^C$。 2：对于 $t = 0$，1，. . .，$T - 1$ 执行 3：采样参与客户端 S_t；将 θ^t 和 $\{\mathbf{p}_g^t$、$_c\}_{c=1}^C$ 广播至 $k \in S_t$。 4：对每个客户端 $k \in S_t$ 并行执行 5：（可选）尾部增强；获得增强数据 $\mathcal{D}_k^{\text{aug}}$ （第 4.4 节）。 6：在 $\mathcal{D}_k^{\text{aug}}$ 上通过最小化式 (13) 进行 E 个轮次，以更新本地编码器 θ_k^{t+1}。 7：通过式 (1) 计算局部原型 $\{\mathbf{p}_k^{t+1}$、$_c\}_{c=1}^C$。 8：通过式 (9) 计算类别级置信度 $\{\mathbf{conf}_k^t$、$_c\}_{c=1}^C$，并设置 $\mathbf{Conf}_k^t = \frac{1}{C}\sum_{c=1}^C \mathbf{conf}_k^t$、$_c$。 9：将 θ_k^{t+1}、$\{\mathbf{p}_k^{t+1}$、$_c\}_{c=1}^C$以及 $\{\mathbf{conf}_k^t$、$_c\}_{c=1}^C$ 上传至服务器。 10：结束循环 11：通过式 (10) 更新全局原型 $\{\mathbf{p}_g^{t+1}$、$_c\}_{c=1}^C$。 12：通过式 (12) 更新全局编码器 θ^{t+1}。 13：结束循环

4.4 额外稳定器：几何正则化与尾部增强

置信度加权聚合是抑制跨轮次注入不可靠原型的主要机制。除聚合外，CAFedCL 还通过两种轻量级稳定器进一步改善少数类表征：（1）原型对齐项，用于减少局部原型与全局原型之间的坐标不匹配；（2）几何正则项，用于维持类间分离并防止类别坍塌。当尾部监督信号极其稀缺时，我们还会启用可选的 尾部增强模块，以增加有效样本量。

在每个客户端 k 上，我们使用组合目标优化编码器：

$$\mathcal{L}_{E,k} = \mathcal{L}_{\text{I2P},k} + \lambda_{\text{align}}\,\mathcal{L}_{\text{align},k} + \lambda_{\text{geo}}\,\mathcal{L}_{\text{geo},k}, \tag{13}$$

其中 $\mathcal{L}_{\text{I2P},\ k}$ 是式（4）中的实例到原型损失，$\mathcal{L}_{\text{align},\ k}$ 将局部原型柔性地对齐到全局原型空间，而$\mathcal{L}_{\text{geo},\ k}$ 则施加几何约束以防止类别坍塌。实践中，我们采用适中的（或逐渐退火的） λ_{align} ，以避免过度依赖早期有偏的锚点。

对齐项提供温和的牵引，使局部原型与广播的全局原型保持在同一坐标系中：

$$\mathcal{L}_{\text{align},k} = \sum_{c=1}^{C} \mathbb{I}[n_{k,c} > 0]\,\left\|\mathbf{p}_{k,c}^{t+1} - \mathbf{p}_{g,c}^{t}\right\|_2^2. \tag{14}$$

几何正则项会明确保持原型之间的类间分离；当多数类别主导对比梯度时，这一点至关重要。我们通过对全局原型施加基于间隔的惩罚项来构造 $\mathcal{L}_{\text{geo},\ k}$：

$$\mathcal{L}_{\text{geo},k} = \sum_{c \neq c'} \max\Big(0,\ m - \left\|\mathbf{p}_{g,c} - \mathbf{p}_{g,c'}\right\|_2\Big). \tag{15}$$

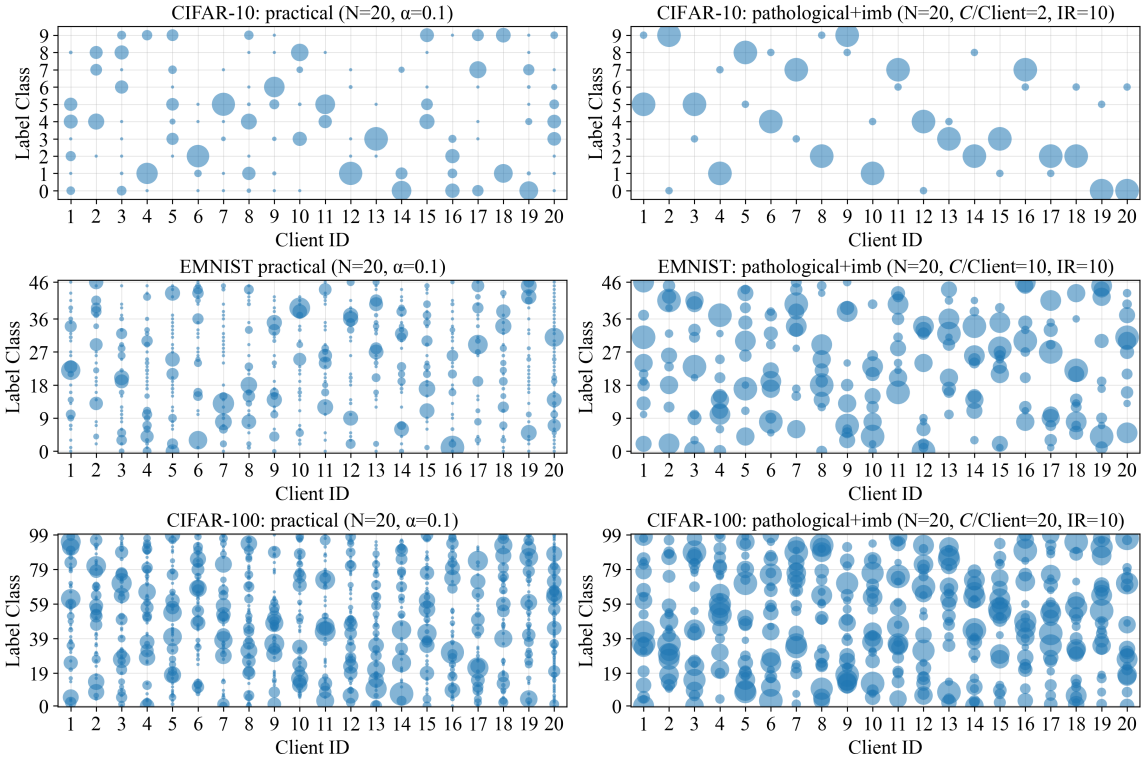

图 2：默认设置下 CIFAR-10、EMNIST 和 CIFAR-100 的数据分布。

其他几何先验（例如正交性或类单纯形排列）也可以直接替换使用。

当少数类监督信号极其稀缺时，我们可以选择使用轻量级条件生成器增强尾部类别，以提供额外的带标签样本并提高n_k^{eff}，c。该模块与 CAFedCL 解耦，也可替换为重采样或 MixUp 等更简单的策略。为求完整，我们在附录中总结对抗目标，并且仅在编码器训练期间将生成样本用作额外的监督数据。

总体而言，尾部增强通过增加有效监督信号来提高尾部原型可靠性，而几何正则化则能抵御多数类导致的挤压效应，维持判别性原型空间。与置信度加权聚合相结合，它们为下一轮式（4）中的实例到原型对比提供稳定的锚点。

5 实验

5.1 实验设置

数据集与基线方法我们采用三个标准基准进行实验：CIFAR-10、CIFAR-100（Krizhevsky 等人，2009）和 EMNIST（Cohen 等人，2017年），涵盖不同程度的

表 1：不同异质设置下三个数据集的测试准确率（%）。结果以 5 次运行的准确率和标准差报告。

方法	实用异质（$\alpha = 0.1,\ K = 20$）						病态异质（$K = 20,\ IR = 10$）					
	CIFAR-10		CIFAR-100		EMNIST		CIFAR-10		CIFAR-100		EMNIST	
	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓
FedAvg	60.32	12.03	29.74	4.08	82.95	11.59	48.73	19.15	24.47	7.58	72.38	14.96
FedProx	61.82	10.07	31.42	4.25	84.27	8.13	60.29	9.11	30.60	5.49	72.45	7.21
MOON	82.06	11.19	46.23	5.15	92.77	6.53	84.94	13.44	48.70	6.95	93.24	5.40
FedProto	88.33	8.13	52.81	6.58	95.50	1.69	89.45	7.78	52.14	5.90	94.88	2.63
FedRCL	87.26	8.79	49.99	8.29	93.57	3.33	86.09	9.30	47.73	8.21	91.25	5.01
FedProc	44.53	6.17	18.98	5.62	65.66	3.57	38.74	6.86	17.04	3.19	63.20	3.71
FedLC	78.48	7.10	44.75	4.70	90.45	7.08	80.11	10.51	46.34	6.61	90.85	7.80
MP-FedCL	86.94	8.46	49.70	6.26	94.30	3.28	87.37	7.55	47.27	5.98	90.05	4.33
FedTGP	89.24	7.03	51.30	8.59	94.47	4.54	89.57	8.46	51.65	6.97	94.59	4.03
CAFedCL	**91.15**	**4.87**	**54.01**	**3.65**	**96.84**	**1.54**	**90.36**	**5.08**	**54.57**	**2.82**	**97.00**	**2.25**

数据异质性和参与率。对于非独立同分布情形，如图 2 所示，我们考虑两种设置：（1）实际设置，通过从带有对称参数 α的狄利克雷分布中采样客户端级标签比例来模拟标签偏斜；（2）病态设置，对每个客户端可接收的标签类别数量施加硬性限制，从而造成更极端的标签分布异质性。不同于以往设置，我们还在病态划分的基础上，通过控制不平衡比率 IR（最大类别规模与最小类别规模之比）引入类别不平衡。评估时，我们使用每个数据集的完整测试集，并报告五次运行的平均 Top-1 准确率。我们将所提方法与具有代表性的联邦对比学习基线方法进行比较，包括 FedAvg（McMahan等人，2017）、FedProx（Li 等人，2020）、MOON（Li 等人，2021b）、FedProto（Tan 等人，2022b）、FedRCL（Seo 等，2024）、FedProc（Mu 等人，2023）、FedLC（Zhang 等人，2022）、MP-FedCL（Qiao 等人，2023）和 FedTGP（Zhang 等，2024）。

实现细节 我们采用 ResNet-18 作为骨干网络，并将批归一化替换为组归一化。我们使用随机梯度下降从头开始训练模型，学习率设为 0.1，指数衰减参数设为 0.998，权重衰减设为 0.001，且不使用动量。本地训练轮数设为 5，并调整批大小，以确保所有实验中每轮均进行 10 次本地迭代。我们使用 PyTorch 实现所有方法，并在搭载 NVIDIA GeForce RTX 4090 GPU 的英特尔 i9-13900KF CPU 上运行实验。

5.2 性能

我们在 CIFAR 和 EMNIST 数据集上，将所提出的方法 CAFedCL 与近期的联邦对比学习基线方法及代表性的联邦学习方法进行比较。表1显示，CAFedCL 在所有数据集和异构设置下均持续优于最先进的基线方法。FedAvg 和 Fed-Prox 等经典的基于优化的方法在异质性条件下性能显著下降，且客户端间差异较大。此处的 std 表示客户端测试准确率的标准差，

表 2：异构设置下的性能（%）。结果以 5 次运行的准确率和标准差报告。

方法	狄利克雷参数				不平衡比率				每个客户端的类别数				客户端数量			
	$\alpha = 0.05$		$\alpha = 1.0$		$IR = 50$		$IR = 100$		C/客户端 $= 10$		C/客户端 $= 50$		50 个客户端		100 个客户端	
	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓	准确率↑	标准差↓
FedAvg	52.54	12.24	86.67	6.19	43.28	22.81	35.29	17.50	20.87	7.12	29.66	2.59	56.73	12.88	54.14	13.65
FedProx	55.48	14.05	87.93	6.44	48.30	11.32	37.33	13.72	28.56	6.28	32.70	2.85	58.50	10.12	55.94	10.47
MOON	65.79	10.40	90.95	5.52	54.23	15.73	42.61	18.25	56.33	4.95	25.49	5.22	79.51	11.58	78.19	10.66
FedProto	66.86	8.35	91.27	5.24	55.48	9.56	44.46	10.29	62.12	4.66	35.06	4.41	83.07	9.92	80.44	10.25
FedRCL	68.30	9.83	92.56	4.77	53.93	10.36	43.08	10.78	60.29	3.98	34.77	2.54	84.08	8.73	82.75	9.11
FedProc	35.68	9.11	84.33	6.80	30.65	8.41	25.49	11.97	29.81	4.02	15.88	4.45	42.24	7.02	41.52	7.31
FedLC	67.33	7.26	91.90	5.03	53.66	13.08	42.41	9.28	55.09	5.17	34.70	3.32	75.32	7.60	73.09	8.51
MP-FedCL	66.18	6.79	91.21	4.52	57.21	9.37	45.15	10.63	62.94	3.19	36.59	3.21	82.25	10.20	79.62	10.55
FedTGP	67.14	7.01	92.58	4.33	58.67	7.55	45.74	7.82	64.16	4.66	36.41	2.76	83.77	8.24	81.95	8.90
CAFedCL	**70.22**	**5.17**	**93.96**	**4.07**	**60.14**	**5.58**	**46.53**	**6.47**	**65.85**	**2.56**	**39.22**	**2.48**	**85.16**	**5.92**	**83.60**	**6.39**

数值越小，表示客户端之间的准确率差距越小，因此公平性越好。近期的对比学习基线方法（如 MOON 和 FedRCL）取得了显著提升，但在 CIFAR-100 等细粒度任务上的增益不够稳定。基于原型的 FedProto 在一定程度上改善了表征对齐，但在极端不平衡情况下仍容易受到原型偏差循环的影响。相比之下，CAFedCL 同时实现了最高的全局准确率和最低的客户端间标准差，表明所提出的置信度感知机制能够有效降低不可靠更新的权重，并避免牺牲难以学习的客户端。总体而言，CAFedCL 不仅具有更强的泛化能力，还显著提升了客户端公平性，因此更适用于异构且不平衡场景。

5.3 鲁棒性

为进一步评估鲁棒性，我们改变了关键异质性和规模因素，包括狄利克雷参数 α、不平衡比率 IR、每个客户端的类别数（C/客户端）以及客户端总数量。在此表中，**标准差**表示客户端准确率的标准差；因此，标准差越小，客户端之间的性能差距越小，公平性也越好。在所有测试条件下，CAFedCL 均取得了最高准确率和最低的客户端间标准差，展现出始终出色的有效性和公平性。

当标签异构性变得更加严重（例如，$\alpha = 0.05$）时，CAFedCL 仍明显比先前方法更加稳定，准确率达到 70.22，标准差为 5.17，而许多基线方法的离散程度则大得多。随着类别不平衡加剧，CAFedCL 的性能平稳下降，仍然是最可靠的方法；在 IR $= 50/100$ 下，其准确率为 60.14/46.53，标准差相对较低（5.58/6.47），而经典方法的准确率大幅下降，方差也显著增大。在不同的 C/客户端约束下，CAFedCL 始终实现最小的客户端间差距（标准差为 2.56 和 2.48），表明即使在类别稀疏性极高的场景下，也能保持较强的公平性。最后，将系统扩展到更多客户端（$50 \rightarrow 100$）时，CAFedCL 仍保持有竞争力的性能（85.16 →83.60），且标准差仅小幅上升（5.92 $\rightarrow$ 6.39），体现出良好的可扩展性。总的来说，CAFedCL 在带来稳健提升的同时，并未牺牲难以学习的客户端，因此在多种异质且不平衡的设置下，始终能取得更高的准确率和更好的客户端公平性。

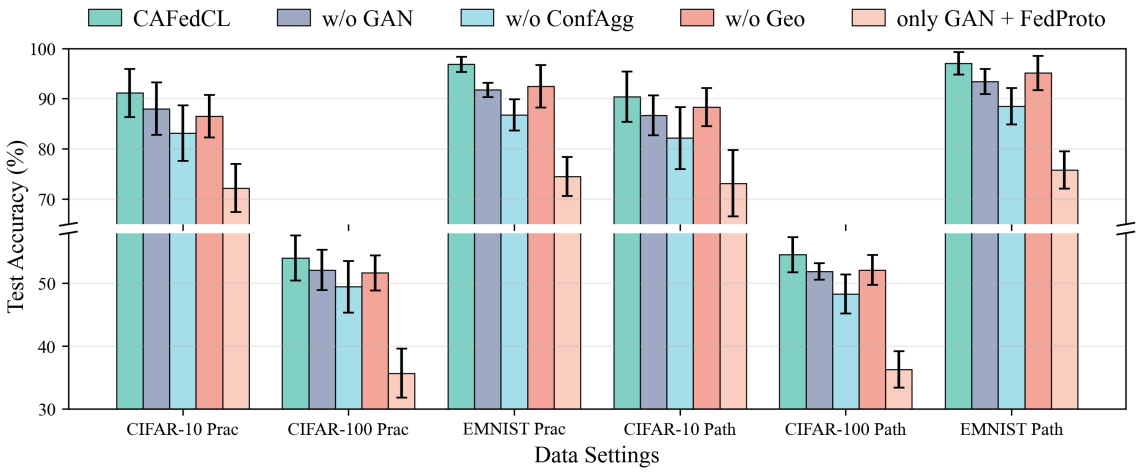

图3: CAFedCL 及其消融变体在默认设置下的准确率 .

5.4 消融研究

我们在图中报告了消融研究3 ，以分别考察 CAFedCL 在六种设置（三个数据集上的实际异质性和病态异质性）下各组成部分的作用。我们考虑了四种变体：（i）w/o GAN 移除少数类的生成式增强；（ii）w/o ConfAgg 将我们的置信度感知聚合替换为朴素聚合方案；（iii）w/o Geo 移除几何一致性正则化；（iv）only GAN+ FedProto 仅保留基于 GAN 的增强，并在 FedProto 的基础上应用该增强，同时舍弃 CAFedCL 的聚合和正则化设计。总体而言，CAFedCL 始终取得最佳准确率，而所有变体的性能均有所下降，说明性能提升源于这些组成部分的协同作用。其中，w/o ConfAgg 导致的下降幅度最大且最为一致（例如，CIFAR-10：实际设置下为 91.15→83.13，病态设置下为 90.36→82.15；EMNIST：分别为 96.84→86.77 和 97.00→88.48），这凸显了置信度感知聚合对于抑制不可靠客户端、避免有偏原型在多个轮次中不断累积至关重要。移除 GAN 增强也会导致明显的性能下降，尤其是在异质性更具挑战性的情况下，这证实了 GAN 增强在提升少数类覆盖率和稳定原型估计方面的作用。w/o Geo 会造成中等但不可忽视的性能下降，表明几何正则化有助于在客户端漂移的情况下维持一致的类别结构。最后，only GAN + FedProto 在所有设置下的表现都显著逊于 CAFedCL，说明若缺少 ConfAgg 和所提出的表示层约束，仅将合成样本注入原型基线是不够的。

5.5 超参数敏感性分析

图4报告了 **CAFedCL** 在 CIFAR-10 实际异质设置下的超参数敏感性。总体而言，CAFedCL 在相当宽的范围内表现稳定，并且一贯偏好适中的参数选择。增大几何损失权重 λ_{geo}可将准确率从 87.66（$\lambda_{\text{geo}}=0$）提升至 $\lambda_{\text{geo}}=1$处的峰值 91.15，而更大的取值

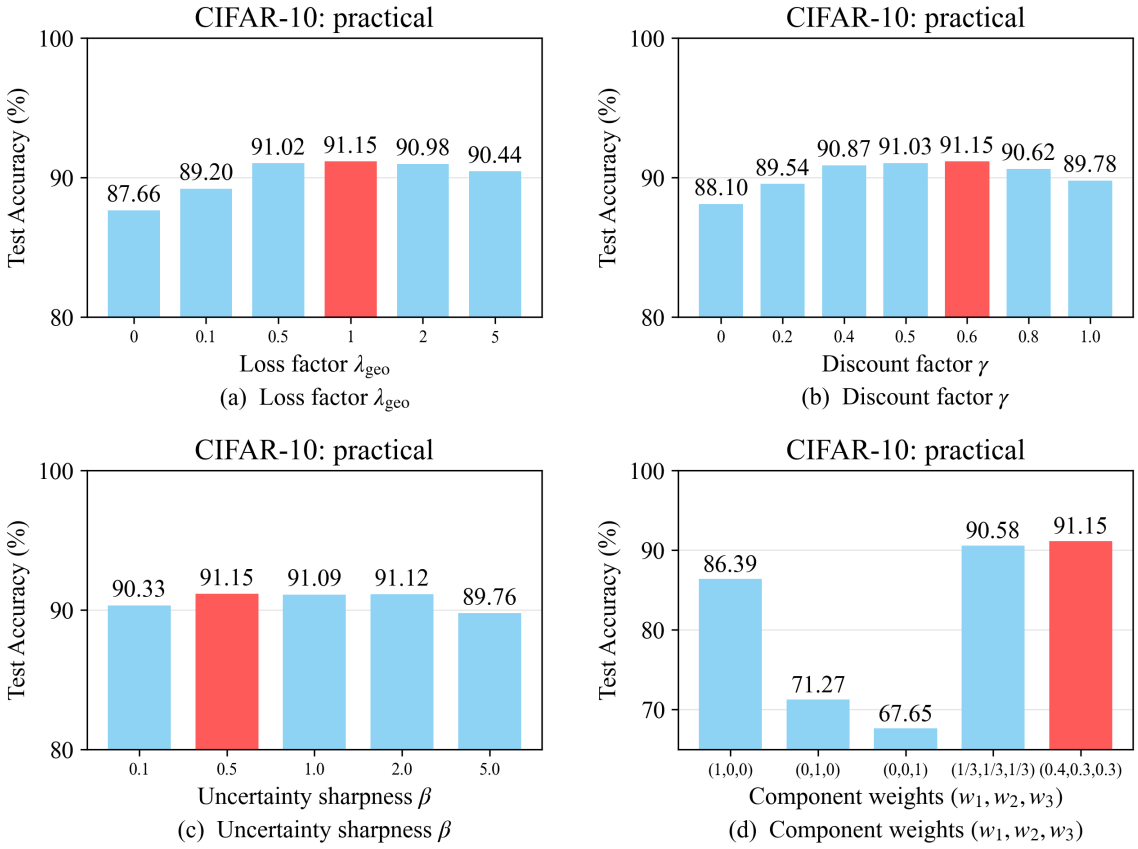

图4：CAFedCL 在不同超参数下的准确率。

会略微降低性能，这表明过强的几何正则化可能会对优化施加过多约束。合成折扣因子 γ 呈现出类似的“适中为佳”趋势，在 γ=0.6 处达到最高准确率（91.15），并随着 γ 接近 1.0 而下降，这支持对生成样本进行折扣，以避免过度信任不完美的生成结果。对于不确定性锐度 β，性能在 β=0.5 处达到峰值（91.15），且在 $\beta \in [0.5, 2]$范围内保持稳健；而 β 过小或过大则会削弱判别能力，或使聚合过于挑剔。最后，对置信度分量权重（w_1、w_2、w_3）进行调整表明，依赖单一分量并非最优（例如，(0、0、1)：67.65）；而结合互补信号则能带来强劲且稳定的性能，我们的默认混合权重（0.4、0.3、0.3）取得了最佳结果。

6 结论

本研究探讨了极端类别不平衡和客户端异质性条件下联邦对比学习面临的关键挑战。尽管基于原型的通信能够提高通信效率，但其本质上容易受到不稳定性的影响。我们正式识别并刻画了原型偏差循环：存在噪声和偏差的客户端原型聚合后形成不准确的全局锚点，随后又受到原型引导的对比

目标的进一步强化，导致表征在多个通信轮次中持续退化。为打破这一恶性循环，我们提出 CAFedCL（置信度感知联邦对比学习），包含三个协同组件：（1）按类别进行可靠性加权，动态校准原型聚合和模型聚合；（2）全局几何正则化，以维护原型空间的结构完整性；（3）尾部类别增强，明确强化少数类表征。这一集成设计在不增加通信开销的前提下，提升了系统鲁棒性和客户端间公平性。

未来工作中有多个值得探索的方向。首先，开发更加轻量且计算效率更高的类别级可靠性估计策略，有望进一步提升 CAFedCL 在大规模联邦系统中的可扩展性。其次，将置信度感知聚合与差分隐私或安全聚合等隐私保护技术相结合，对于实际部署至关重要。最后，将所提出的框架扩展到更复杂的联邦设置中，包括异构模型架构或异步通信，仍是未来值得研究的方向，尤其需要深入理解模型失配和随时间变化的客户端参与如何影响开放世界部署中的置信度校准以及原型对齐的稳定性。

参考文献

M. Adnan、S. Kalra、J. C. Cresswell、G. W. Taylor和H. R. Tizhoosh。面向医学图像分析的联邦学习与差分隐私。*Scientific Reports*，12(1):1953，2022年。T. Chen、S. Kornblith、M. Norouzi和G. Hinton。一种用于视觉表征对比学习的简单框架。见国际机器学习会议，第1597–1607页。PmLR，2020年。G. Cohen、S. Afshar、J. Tapson和A. Van Schaik。EMNIST：将MNIST扩展到手写字母。见*2017*年国际神经网络联合会议（*IJCNN*），第2921–2926页。IEEE，2017年。Y. Dai、Z. Chen、J. Li、S. Heinecke、L. Sun和R. Xu。利用类别原型应对联邦学习中的数据异质性。见*AAAI*人工智能会议论文集，第37卷，第7314–7322页，2023年。Y. Diao、Q. Li和B. He。通过模型拼接利用联邦学习中的标签偏斜。见*AAAI*人工智能会议论文集，第38卷，第11784–11792页，2024年。Y. Guo、X. Tang和T. Lin。FedBR：通过本地学习偏差降低改进异构数据上的联邦学习。见国际机器学习会议，第12034–12054页。PMLR，2023年。K. He、H. Fan、Y. Wu、S. Xie和R. Girshick。用于无监督视觉表征学习的动量对比。见*IEEE/CVF*计算机视觉与模式识别会议论文集，第9729–9738页，2020年。

C. Huang、X. Chen、Y. Zhang和H. Wang。FedCoSR：面向非独立同分布数据中标签异构性、采用对比可共享表示的个性化联邦学习。*IEEE*控制论汇刊，2025。A. Imteaj、U. Thakker、S. Wang、J. Li和M. H. Amini。面向资源受限物联网设备的联邦学习综述。*IEEE*物联网期刊，9(1):1–24，2021。S. Jing、A. Yu、S. Zhang和S. Zhang。FedSC：基于谱对比目标、面向非独立同分布数据的可证明联邦自监督学习。见第*41*届国际机器学习会议论文集，ICML'24。JMLR.org，2024年。P. Kairouz、H. B. McMahan、B. Avent、A. Bellet、M. Bennis、A. N. Bhagoji、K. Bonawitz、Z. Charles、G. Cormode、R. Cummings等。联邦学习的进展与开放问题。基础与趋势® 机器学习，14(1–2):1–210，2021。S. P. Karimireddy、S. Kale、M. Mohri、S. Reddi、S. Stich和A. T. Suresh。SCAFFOLD：面向联邦学习的随机控制平均。见国际机器学习会议，第5132–5143页。PMLR，2020年。L. U. Khan、W. Saad、Z. Han、E. Hossain和C. S. Hong。面向物联网的联邦学习：最新进展、分类与开放挑战。*IEEE*通信综述与教程，23(3):1759–1799，2021。A. Krizhevsky、G. Hinton等。从微型图像中学习多层特征。2009。G. Li、V. Jampani、L. Sevilla-Lara、D. Sun、J. Kim和J. Kim。面向少样本分割的自适应原型学习与分配。见*IEEE/CVF*计算机视觉与模式识别会议论文集，第8334–8343页，2021a。Q. Li、B. He和D. Song。模型对比联邦学习。见*IEEE/CVF*计算机视觉与模式识别会议论文集，第10713–10722页，2021b。T. Li、A. K. Sahu、M. Zaheer、M. Sanjabi、A. Talwalkar和V. Smith。异构网络中的联邦优化。机器学习与系统会议论文集，2:429–450，2020年。X. Li、M. Jiang、X. Zhang、M. Kamp和Q. Dou。FedBN：通过局部批归一化在非独立同分布特征上进行联邦学习。*arXiv*预印本 *arXiv:2102.07623*，2021c。Q. Liu、C. Chen、J. Qin、Q. Dou和P.-A. Heng。FedDG：通过连续频率空间中的情景学习实现医学图像分割上的联邦域泛化。见*IEEE/CVF*计算机视觉与模式识别会议论文集，第1013–1023页，2021。M. Luo、F. Chen、D. Hu、Y. Zhang、J. Liang和J. Feng。无惧异质性：面向非独立同分布数据联邦学习的分类器校准。《神经信息处理系统进展》，34:5972–5984，2021。

B. McMahan、E. Moore、D. Ramage、S. Hampson和B. A. y Arcas。基于去中心化数据的深度网络通信高效学习。载于人工智能与统计学，第1273–1282页。PMLR，2017年。X. Mu、Y. Shen、K. Cheng、X. Geng、J. Fu、T. Zhang和Z. Zhang。FedProc：面向非独立同分布数据的原型对比联邦学习。《未来一代计算机系统》，143:93–104，2023年。A. v. d. Oord、Y. Li和O. Vinyals。基于对比预测编码的表示学习。*arXiv*预印本 *arXiv:1807.03748*，2018年。Y. Qiao、M. S. Munir、A. Adhikary、H. Q. Le、A. D. Raha、C. Zhang和C. S. Hong。Mp-fedcl：面向边缘智能的多原型联邦对比学习。*IEEE*物联网期刊，11(5):8604–8623，2023年。S. Seo、J. Kim、G. Kim和B. Han。面向联邦学习的松弛对比学习。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第12279–12288页，2024年。A. Z. Tan、H. Yu、L. Cui和Q. Yang。迈向个性化联邦学习。《*IEEE*神经网络与学习系统汇刊》，34(12):9587–9603，2022a。Y. Tan、G. Long、L. Liu、T. Zhou、Q. Lu、J. Jiang和C. Zhang。FedProto：面向异构客户端的联邦原型学习。载于*AAAI*人工智能会议论文集，第36卷，第8432–8440页，2022b。Y. Tan、G. Long、J. Ma、L. Liu、T. Zhou和J. Jiang。基于预训练模型的联邦学习：一种对比学习方法。《神经信息处理系统进展》，35:19332–19344，2022c。J. Wang、Q. Liu、H. Liang、G. Joshi和H. V. Poor。解决异构联邦优化中的目标不一致性问题。《神经信息处理系统进展》，33:7611–7623，2020年。L. Wang、S. Xu、X. Wang和Q. Zhu。应对联邦学习中的类别不平衡。载于*AAAI*人工智能会议论文集，第35卷，第10165–10173页，2021。R. Wang、W. Huang和W. Huang。面向机器故障诊断的轻量级联邦域泛化与全局—局部对比学习。*IEEE*物联网期刊，12(19):40750–40763，2025。doi: 10.1109/JIOT.2025.3590721。N. Wu、L. Yu、X. Yang、K.-T. Cheng和Z. Yan。Fediic：面向类别不平衡医学图像分类的稳健联邦学习。载于医学图像计算与计算机辅助干预国际会议，第692–702页。施普林格，2023年。J. Zhang、Z. Li、B. Li、J. Xu、S. Wu、S. Ding和C. Wu。通过逻辑值校准处理标签分布偏斜的联邦学习。载于国际机器学习会议，第26311–26329页。PMLR，2022年。

J. Zhang、Y. Liu、Y. Hua和J. Cao。FedTGP：面向联邦学习中数据与模型异构性的可训练全局原型与自适应间隔增强对比学习。载于*AAAI*人工智能会议论文集，第38卷，第16768–16776页，2024年。L. Zhang、Y. Luo、Y. Bai、B. Du和L.-Y. Duan。通过统一特征学习与优化目标对齐实现面向非独立同分布数据的联邦学习。载于*IEEE/CVF*国际计算机视觉会议论文集，第4420–4428页，2021。Y. Zhao、M. Li、L. Lai、N. Suda、D. Civin和V. Chandra。面向非独立同分布数据的联邦学习。*arXiv*预印本 *arXiv:1806.00582*，2018年。W. Zhuang、Y. Wen和S. Zhang。散度感知联邦自监督学习。*arXiv*预印本 *arXiv:2204.04385*，2022年。