基于特征对齐和分类器协作的个性化联邦学习

摘要

数据异构性是联邦学习中最具挑战性的问题之一，这促使人们采用各种方法为参与学习的客户学习个性化模型。在基于深度神经网络的任务中，其中一种方法是采用共享特征表示，并为每个客户端学习定制的分类器头。然而，以前的工作在局部表示学习过程中没有利用全局知识，也忽略了局部分类器头之间的细粒度协作，从而限制了模型的泛化能力。在这项工作中，本文利用全局语义知识进行显式的特征对齐，以学习更好的特征提取器。此外，本文将分类器进行线性组合，并推导出估计最优权重的优化问题。最后，在各种异构数据场景的基准数据集上进行广泛评估，证明了本文提出的方法的有效性。

关键词：个性化联邦学习；特征对齐；分类器协作

1 引言

现代学习任务通常由深度神经网络 (DNNs) 实现，这需要大量的训练数据来实现令人满意的模型性能。然而，由于数据量越来越大，收集数据的成本太高，甚至由于隐私保护而被禁止。因此，开发高效通信和保护隐私的学习算法对于充分利用客户端 (例如数据孤岛和移动设备) 中的数据具有重要意义 [1]。为此，联邦学习 (FL) 作为一种创新技术出现，在不收集原始数据的情况下，在分散的客户端上进行协作模型训练。典型的 FL 设置使用中央服务器来维护全局模型，并允许部分客户端参与不频繁的模型聚合，例如 FedAvg，当跨客户端的本地数据是独立同分布 (IID) 时，它显示出良好的性能。然而，在 FL 的背景下，客户端的数据分布通常是不相同的 (non-IID 或异构)，因为不同的设备分别生成或收集数据，并且可能有特定的偏好，包括特征分布漂移、标签分布倾斜和概念转移，这使得很难学习适用于所有客户端的单一全局模型 [2]。

为了解决这个问题，个性化联邦学习 (PFL) 已经被开发出来，其目标是为每个客户学习一个定制模型，该模型在本地数据上具有更好的性能，同时仍然受益于协作训练 [3]。深度神经网络模型通常通过特征提取器从原始数据中提取出代表性的特征，然后利用这些特征通过分类器进行分类或预测。深度学习在集中式系统和多任务学习中的成功表明，特征提取器通常扮演公共结构的角色，而分类器往往是高度任务相关的 [4]。此外，实际 FL 问题中的客户端通常处理类似的学习任务，并且在许多先前的工作中假设客户端之间存在聚类结构。因此，学习更好的全局特征表示和利用局部任务之间的相关性对于改进个性化模型具有重要意义。

在这项工作中，本文主要考虑标签分布移位的场景，即在不同的客户端上，类的数量是相同的，而每个类的数据样本数量有明显的偏移，即本地任务的标签分布是异构的。本文通过利用共享表示和客户端间分类器协作，从多任务学习的角度研究联邦学习。

具体来说，本文利用每个类的全局特征质心来正则化局部训练，这可以看作是显式的特征对齐，并且能够减少局部训练的特征提取器之间的表示多样性，从而促进全局聚合。本文还通过针对客户的线性组合进行灵活的分类器协作，鼓励相似的客户进行更多的协作，避免不相关客户的负迁移。为了估计合适的组合权重，本文利用局部特征统计和数据分布信息，通过解决二次规划问题，使每个客户端的预期测试损失最小化，从而实现最佳的偏差-方差权衡。此外，稍加修改，本文的框架仍然可以在概念转换场景下很好地工作，在这种情况下，相同的标签可能在不同的客户端具有不同的含义。

2 相关工作

2.1 基于 Non-IID 数据的联邦学习

对于 non-IID 数据的 FL 全局模型学习的改进已经有很多的研究。各种各样的工作都集中在通过利用精心设计的目标正则化来优化局部学习算法和局部偏置校正。例如，FedProx [5] 在局部训练目标中增加了一个近端项，使更新后的参数与原始下载模型保持接近;SCAFFOLD [6] 引入控制变量来纠正局部更新中的偏移;MOON [7] 采用对比损失来改进表示学习。类平衡数据重采样和损失重加权方法可以提高客户端局部数据不平衡时的训练性能。此外，还研究了数据共享机制和数据增强方法，以缓解 non-IID 数据挑战。从模型聚合的角度来看，选择对全局模型性能贡献更大的客户端也可以加快收敛速度，减轻 non-IID 数据的影响。随着公共数据的可用性，尽管数据存在异构性，但仍有可能采用知识蒸馏技术获得全局模型。与上述方法不同的是，本文的目的是针对每个客户学习定制的模型。

2.2 个性化联邦学习

在以往的研究中，流行的个性化 FL 方法包括执行局部和全局模式线性组合的加性模型混合，如 L2CD 和 APFL [8, 9]; 具有模型不相似性惩罚的多任务学习，包括 FedMTL [10]、pFedMe [11] 和 Ditto [12]。特征提取器与分类器的参数解耦，如 FedPer、LG-FedAvg 和 FedRep [13–15]；一种特殊类型的个性化 FL 方法是聚类 FL，将相似的客户端分组在一起并学习多个组内全局模型 [16]。特定于客户端的模型聚合也被研究用于细粒度联合，例如 FedFomo 和 FedAMP [17, 18]，它们与本文的方法具有相似的精神。然而，现有的特定于客户端的 FL 方法通常是通过以启发式方式评估模型相似性或验证准确性来开发的，这些技术需要在通信、计算开销和个性化的有效性之间取得良好的平衡。基于高斯过程的 FedGP [19] 和基于基于选择性知识转移的解决方案也得到了发展 [20]，但是这些方法不可避免地依赖于公共共享数据集或诱导点集。此外，还研究了由服务器端超网络启用的 pFedHN [21] 和学习多个全局模型混合的 FedEM [22]，以便为每个客户端生成定制模型。然而，pFedHN 要求每个客户端进行多次通信以学习代表性嵌入，并且 FedEM 显著增加了通信和计算/存储开销。最近，FedRoD [23] 提出使用平衡 softmax 来学习通用模型，使用 vanilla softmax 来学习个性化分类头部。FedBABU [24] 提出在特征表示学习过程中保持全局分类器不变，并通过微调进行局部采用。kNN-per [25] 将全局模型和局部 kNN 分类器集成在一起，以获得更好的个性化性能。我们的工作与 FedRep 共享最相似的学习过程，但不同之处是我们使用全局知识来指导局部表示学习，并为每个客户端执行理论上保证的分类器头部最优组合。

3 本文方法

在本节中，本文概述了通过利用更好的特征提取器和特定于用户的分类器协作来训练个性化模型的方法即 FedPAC。

3.1 问题设置

本文考虑一个有 m 个客户端和一个中央服务器的设置，其中所有客户端都与服务器通信，以协作训练个性化模型，而不共享原始私有数据。在个性化 FL 中，每个客户端 i 在 XŒY 上都有自己的数据分布 $P_{XY}^{(i)}$，其中 X 为输入空间，Y 为总共有 K 个类别的标签空间。在 FL 中通常设定 $P_{XY}^{(i)}$ 和 $P_{XY}^{(j)}$ 是一对不同的客户端 i 和 j。设 $l : X \times Y \to R_+$ 表示给定局部模型 w_i 和从 $P_{XY}^{(i)}$ 中采样的数据点的损失函数，例如交叉熵损失，则 PFL 的底层优化目标可以形式化为:

$$\min_{W} \left\{ F(W) := \frac{1}{m} \sum_{i=1}^{m} E_{(x,y)\sim P_{XY}^{(i)}} [l(w_i; x, y)] \right\} \tag{1}$$

其中 $W = (w_1, w_2, ..., w_m)$ 表示所有局部模型的集合。然而，真实的潜在分布是不可接近的，通常通过经验风险最小化 (ERM) 来实现目标。设每个客户端可以访问的来自 $P_{XY}^{(i)}$ 的数据点为 n_i，表示为 $D_i = \{(x_l^{(i)}, y_l^{(i)})\}_{l=1}^{n_i}$，对应的经验分布是 $\hat{P}_{XY}^{(i)}$ 本文假设经验边际分布 $\hat{P}_Y^{(i)}$ 与实际分布?$P_Y^{(i)}$ 相同，那么训练目标是

$$w^* = \arg\min_{w} \frac{1}{m} \sum_{i=1}^{m} [L_i(w_i) + R_i(w_i; \Omega)] \tag{2}$$

$L_i(w_i) = \frac{1}{n_i} \sum_{l=1}^{n_i} (w_i; x_l^{(i)}, y_l^{(i)})$ 是本地训练的平均损失，例如经验风险；是全局特征信息，$R_i(\cdot)$ 是一个预定义的正则化项，用于防止 w_i 过拟合。

3.2 共享特征表示

在不损失通用性的前提下，本文将深度神经网络解耦为表示层和最终决策层，其中前者也称为特征提取器，后者指分类任务中的分类器头。特征提取器为 $f : X \to R^d$ 可学习网络的参数是 θ_f，d 是特征提取输出的维度。给定一个数据点 x, 对提取的特征向量 $z = f(x)$ 进行预测可以由线性函数 $g(z)$ 产生，参数为 ϕ_g。由于每个客户端数据不足，局部学习到的特征表示容易出现过拟合，不能很好地泛化。一个合理可行的想法是通过共享相同的特征表示层来利用其他客户机上可用的数据。然而，对私有数据的多次局部更新会导致局部过拟合和客户端参数的高多样性，从而使聚合模型偏离最佳表示。为了解决这个问题，本文提出了一个新的正则化术语用于局部特征表示学习。

正则化特征对齐。客户端需要同时考虑监督学习损失和泛化误差来更新局部模型。为此，本文利用全局特征质心，并在局部训练目标中引入新的正则化项，使局部表示学习受益于全局数据。局部正则化项表示为:

$$R_i(\theta_i; c) = \frac{\lambda}{n_i} \sum_{l=1}^{n_i} \frac{1}{d} \left\| f_{\theta_i}(x_l) - c_{yl} \right\|_2^2 \tag{3}$$

$f_{\theta_i}(x_l)$ 是特征提取器，c_{yl} 是 yl 类对应的全局质心 为平衡监督损失和正则化损失的超参数。通过利用全局语义特征信息，这样的正则化术语对每个客户机都有很大的好处。直观地说，它使每个客户端能够通过显式的特征分布对齐来学习任务不变表示。

3.3 分类器协作

除了通过共享表示层来改进特征提取器之外，本文认为合并来自具有相似数据分布的其他客户端的分类器也可以提供性能提升。直观地说，当局部数据不足时，局部学习的分类器可能会有很大的方差，因此那些共享相似数据分布的客户端实际上可以通过客户间知识转移来协同训练个性化分类器。挑战在于如何评估客户之间的相似性和可转移性。为此，本文将接收到的分类器对每个客户 i 进行线性组合，以减少局部测试损失:

$$\hat{\phi}_i^{(t+1)} = \sum_{j=1}^{m} \alpha_{ij} \phi_j^{(t+1)} \quad s.t. \sum_{j=1}^{m} \alpha_{ij} = 1 \tag{4}$$

对于每个系数 $\alpha_{ij} \geq 0$ 由最小化局部预期测试损失确定，可表示为以下优化问题:

$$\alpha_i^* = \arg\min_{\alpha_i} \mathrm{E}_{(\mathrm{x},\mathrm{y})\mathrm{P}_{\mathrm{XY}}^{(\mathrm{i})}} \left[l\left(\theta, \sum_{j=1}^{m} \alpha_{ij} \phi_j; x, y\right) \right] \tag{5}$$

4 复现细节

4.1 与已有开源代码对比

本文参考原论文代码框架，并基于此框架对本地训练过程中分类器训练过程进行改进。

4.2 实验环境搭建

本文实验基于 Windows11 系统进行环境搭建且并未使用 GPU 资源，使用 pytorch 来搭建整体代码框架。

4.3 创新点

在原论文中对客户端分类器训练只进行一轮，考虑到在客户端数据量少的情况下多轮的训练很容易导致其过拟合，但原论文并未对该过程进行处理，因此在分类器训练时引入近端项对其进行改进，可以有效的提高分类器训练次数，加快其收敛过程。加入近端项后分类器优化目标函数可表示为:

$$h_k(w; w^t) = F_k(w) + \frac{\mu}{2} \|w - w^t\|^2 \tag{6}$$

5 实验结果分析

5.1 实验设置

数据集和模型。本文考虑图像分类任务，并在四个流行的数据集上评估本文的方法:(1) EMNIST (Extended MNIST) 是一个 62 类图像分类数据集，扩展了经典的 MNIST 数据集。

它包含 62 类手写字符, 包括 10 个数字, 26 个大写字母和 26 个小写字母;(2) 包含 10 类服装的 Fashion-MNIST;(3) 包含 10 类彩色图像的 CIFAR-10;(4) CINIC-10，它比 CIFAR-10 更多样化, 由两个不同的来源构建的:ImageNet 和 CIFAR-10。本文分别为 EMNIST/Fashion-MNIST 和 CIFAR-10/CINIC-10 构建了两个不同的 CNN 模型。第一个 CNN 模型由两个分别有 16 个通道和 32 个通道的卷积层构建，每个卷积层后面有一个最大池化层，在 softmax 输出之前有两个 128 和 10 个单元的完全连接层。本文使用 LeakyReLU 激活函数。第二个 CNN 模型与第一个类似，但多了一个 64 通道的卷积层。

数据分区。本文使所有客户端具有相同的数据大小，其中 s% 的数据 (默认为 20%) 从所有类中统一采样，其余 (100 - s)% 来自每个客户端的一组主导类。本文将客户端分成多个组，每个组中的客户端共享相同的主导类，且保持本地训练数据的大小较小，以满足 FL 的需求。每个客户机上的测试数据与训练数据具有相同的分布。

训练设置。本文使用 mini-batch SGD 作为所有方法的局部优化器，并且局部训练 epoch 的数量设置 5，对于所有数据集，将全局通信轮数设置为 200。

表 1. 不同数据集上最终测试精度 (%) 的比较。对有 20 个客户的 FL 系统采用全员参与，对有 100 个客户的 FL 系统采样率为 0.3 的客户抽样。

Method	EMNIST		Fashion-MNIST		CIFAR-10		CINIC-10	
	20 clients	100 clients	20 clients	100 clients	20 clients	100 clients	20 clients	100 clients
FedPAC	88.38	89.96	91.73	93.08	80.86	84.85	74.44	77.58

5.2 实验结果

性能比较。在两个设置上进行实验，其中客户端数量分别为 20 和 100。对于后者，本文采用随机客户选择，抽样率 0.3，并在最后一轮中充分参与。除了 EMNIST 之外的所有数据集，每个客户端的训练数据大小都设置为 600,EMNIST 的大小为 1000。主要结 2-320 果如表 1 所示，很明显，本文提出的方法在小规模和大规模 FL 系统上都表现良好。对于所有数据集，FedPAC 在平均测试精度上都优于其他方法，这证明了全局特征对齐和客户端间分类器协作的有效性和优势。

消融实验。本文在 FedPAC 中有两个关键的设计组件，即特征对齐 (FA) 和分类器组合 (CC)。在这里，本文进行消融研究以验证这两个模块的有效性。分别在四个数据集上使用 FA 和 CC，并得到了 20 个客户端的平均准确性。如表 2 所示，两者都有助于提高平均准确率，并且两者的组合能够获得最满意的模型性能，这意味着在本文提出的方法下可以构建更好的全局特征提取器和更合适的个性化分类器

表 2. 消融实验。FA 表示特征对齐，CC 表示分类器协作;None 表示既不使用 FA 也不使用 CC，而 Both 表示同时使用 FA 和 CC。

Method	Design Choices in FedPAC			
	None	w/FA	w/CC	Both
EMNIST(%)	75.31	86.56	83.68	**88.38**
Fashion-MNIST(%)	88.02	91.54	90.18	**91.73**
CIFAR-10(%)	71.45	79.07	77.82	**80.86**
CINIC-10(%)	66.18	72.98	71.33	**74.44**

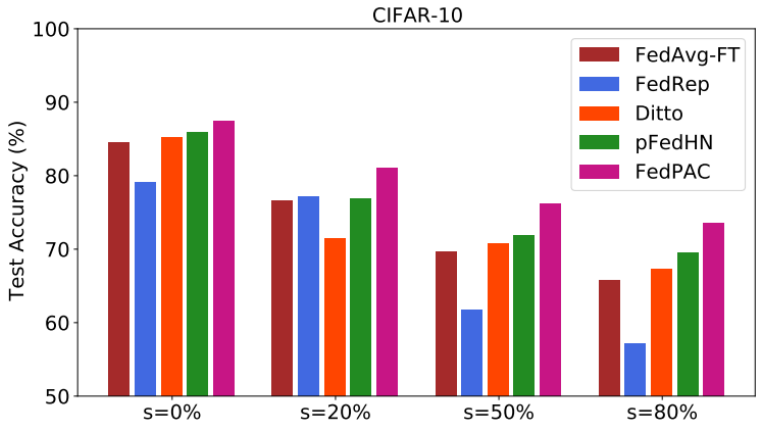

图 1. 数据异构性的影响

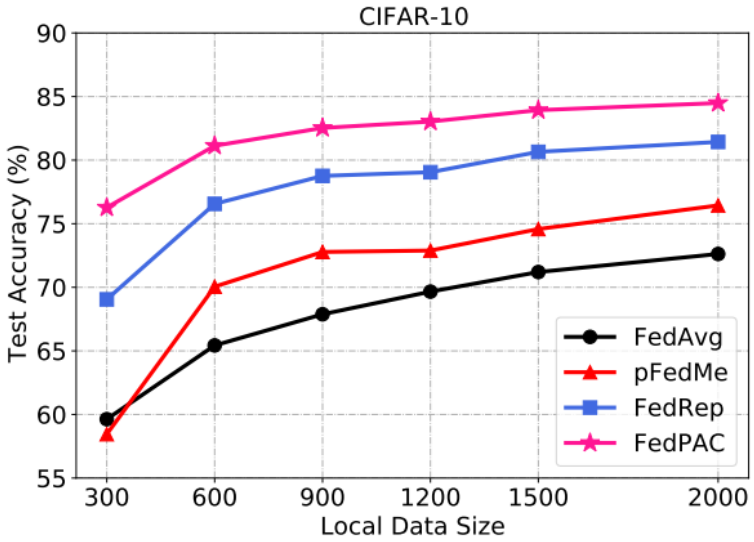

图 2. 数据集大小的影响

数据异质性和数据大小的影响。本文改变 s 的值来模拟不同程度的数据异构，s=0% 表示高度异构的情况 (non-IID)，s=80% 表示客户端间的数据更均匀。本文对 CIFAR-10 数据集进行了评估，不同方法的结果见图 1。可以发现，本文的方法始终优于其他基线，这表明了它在各种异构数据场景中的泛化性和鲁棒性。本文还用不同的本地数据大小测试了本文的

FedPAC 和 FedAvg，并记录了结果模型的准确性，如图 2 所示。结果表明，具有不同数据大小的客户端可以始终从参与 FL 中受益，并且本文的方法获得了更高的性能增益。

6 总结与展望

在本文中，介绍了用于增强表示学习的全局特征对齐和用于在 FL 中构建个性化分类器的新型分类器组合算法，为它们在异构设置中的实用性提供了理论和经验证明。未来的工作包括在更复杂的环境中分析最优模型个性化，例如，分布式系统或具有动态数据分布的客户端，并研究局部特征提取器的最优聚合。

参考文献

[1] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020.

[2] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 965–978. IEEE, 2022.

[3] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.

[4] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International conference on machine learning*, pages 2089–2099. PMLR, 2021.

[5] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[6] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

[7] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021.

[8] Filip Hanzely and Peter Richtárik. Federated learning of a mixture of global and local models. *arXiv preprint arXiv:2002.05516*, 2020.

[9] Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461*, 2020.

[10] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. *Advances in neural information processing systems*, 30, 2017.

[11] Canh T Dinh, Nguyen Tran, and Josh Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems*, 33:21394–21405, 2020.

[12] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *International Conference on Machine Learning*, pages 6357–6368. PMLR, 2021.

[13] Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

[14] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*, 2020.

[15] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International conference on machine learning*, pages 2089–2099. PMLR, 2021.

[16] Moming Duan, Duo Liu, Xinyuan Ji, Yu Wu, Liang Liang, Xianzhang Chen, Yujuan Tan, and Ao Ren. Flexible clustered federated learning for client-level data distribution shift. *IEEE Transactions on Parallel and Distributed Systems*, 33(11):2661–2674, 2021.

[17] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M Alvarez. Personalized federated learning with first order model optimization. *arXiv preprint arXiv:2012.08565*, 2020.

[18] Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. Personalized cross-silo federated learning on non-iid data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 7865–7873, 2021.

[19] Idan Achituve, Aviv Shamsian, Aviv Navon, Gal Chechik, and Ethan Fetaya. Personalized federated learning with gaussian processes. *Advances in Neural Information Processing Systems*, 34:8392–8406, 2021.

[20] Jie Zhang, Song Guo, Xiaosong Ma, Haozhao Wang, Wenchao Xu, and Feijie Wu. Parameterized knowledge transfer for personalized federated learning. *Advances in Neural Information Processing Systems*, 34:10092–10104, 2021.

[21] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *International Conference on Machine Learning*, pages 9489–9502. PMLR, 2021.

[22] Othmane Marfoq, Giovanni Neglia, Aurélien Bellet, Laetitia Kameni, and Richard Vidal. Federated multi-task learning under a mixture of distributions. *Advances in Neural Information Processing Systems*, 34:15434–15447, 2021.

[23] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning for image classification. *arXiv preprint arXiv:2107.00778*, 2021.

[24] Jaehoon Oh, Sangmook Kim, and Se-Young Yun. Fedbabu: Towards enhanced representation for federated image classification. *arXiv preprint arXiv:2106.06042*, 2021.

[25] Othmane Marfoq, Giovanni Neglia, Richard Vidal, and Laetitia Kameni. Personalized federated learning through local memorization. In *International Conference on Machine Learning*, pages 15070–15092. PMLR, 2022.