

# FedFR: Joint Optimization Federated Framework for Generic and Personalized Face Recognition

Chih-Ting Liu<sup>1\*</sup> Chien-Yi Wang<sup>2\*</sup> Shao-Yi Chien<sup>1</sup> Shang-Hong Lai<sup>2</sup>

<sup>1</sup> Graduate Institute of Electronics and Engineering, National Taiwan University

<sup>2</sup> Microsoft AI R&D Center, Taiwan

jackieliu@media.ee.ntu.edu.tw, chiwa@microsoft.com, sychien@ntu.edu.tw, shlai@microsoft.com

## Abstract

Current state-of-the-art deep learning based face recognition (FR) models require a large number of face identities for central training. However, due to the growing privacy awareness, it is prohibited to access the face images on user devices to continually improve face recognition models. Federated Learning (FL) is a technique to address the privacy issue, which can collaboratively optimize the model without sharing the data between clients. In this work, we propose a FL based framework called FedFR to improve the generic face representation in a privacy-aware manner. Besides, the framework jointly optimizes personalized models for the corresponding clients via the proposed Decoupled Feature Customization module. The client-specific personalized model can serve the need of optimized face recognition experience for registered identities at the local device. To the best of our knowledge, we are the first to explore the personalized face recognition in FL setup. The proposed framework is validated to be superior to previous approaches on several generic and personalized face recognition benchmarks with diverse FL scenarios. The source codes and our proposed personalized FR benchmark under FL setup are available at <https://github.com/jackie840129/FedFR>.

## Introduction

Face recognition has been an active and vital topic among computer vision community for a long time. The state-of-the-art training frameworks formulate face recognition as a metric learning problem, and employ the large-scale identity classification as the proxy task to learn face features, which could discriminate between different identities robustly. Recently, the quick evolution of softmax-based loss functions for identity classification greatly promote the performance of face recognition. However, the training of face recognition model heavily relies on centralizing a huge amount of personal face images, which are usually not accessible due to the uprising privacy concern in many countries. Therefore, it is necessary to navigate the development of face recognition under the premise of privacy preservation.

Federated learning (FL) provides a distributed and privacy-aware framework to train models where multiple clients collaboratively learn without sharing their data with

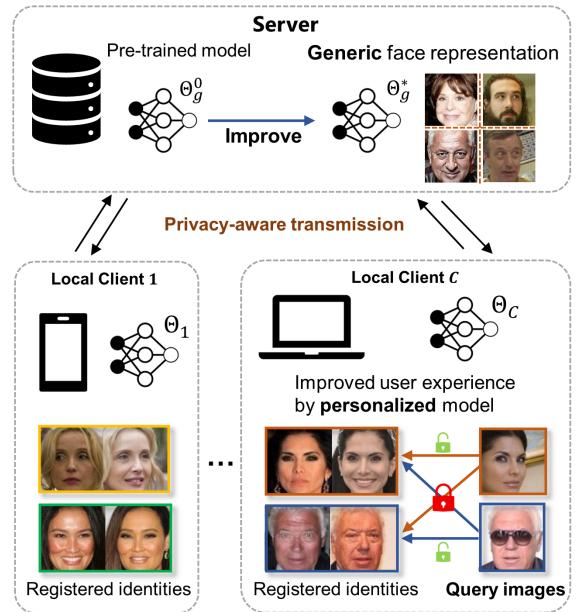


Figure 1: **The Federated Learning (FL) setup for face recognition.** Given a pre-trained face recognition model, we aim to simultaneously improve the generic face representation at the server, and produce an optimal personalized model for each client without transmitting private identities’ images or features out of the local devices.

the central server or other clients. A classical FL method called FedAvg (McMahan et al. 2017) aggregates and averages the gradients from local clients on the server, and transmit the updated model back to the clients for the next round of local optimization. In the past few years, there has been significant progress in FL (Kairouz et al. 2019) on image classification task, which boosts the performance of aggregated global model under diverse FL scenarios. However, these approaches cannot be directly applied onto face recognition due to several critical reasons: 1) Face recognition is an open-set classification task, where training and testing identity classes are different. 2) The identity classes between local clients are different, which results in different model architectures in clients. 3) In a more practical setup for face recognition (Aggarwal, Zhou, and Jain 2021), the FL train-

\*Both authors contributed equally to this work.

Copyright © 2022, Association for the Advancement of Artificial Intelligence ([www.aaai.org](http://www.aaai.org)). All rights reserved.

ing starts from a publicly available face recognition model, rather than from scratch as in traditional FL.

In order to address these aforementioned issues, a recent work FedFace (Aggarwal, Zhou, and Jain 2021) proposed an FL framework for face recognition model training in a privacy-aware manner. It tackles the challenging setup where each of the participating clients has face images of only one identity. It employs a mean feature initialization method for the local identity proxy and a spreadout regularizer (Yu et al. 2020) at the server side to ensure that the identity proxies from the local clients are well separated. However, FedFace is limited as it only addressed a single scenario. In the real-world face recognition applications, local edge devices could be registered by multiple identities. Moreover, there exists a serious privacy concern in FedFace as it requires the local device to transmit the identity proxy to the server. As shown in Vec2Face (Duong et al. 2020), the identity class proxy could be used for reconstructing the original face images which violates the FL protocol.

To enable federated learning in more realistic face recognition settings, we propose a novel framework called **FedFR**, which could jointly improve generic and personalized face representations without breaking the privacy on clients. First, we leverage the globally shared dataset to regularize the training on local clients, as the local client only has much less identities than the pre-trained dataset. With the additional transmission of the shared class embedding matrix, it can effectively prevent the local model from overfitting and also improve the generic representation at the server. Secondly, in order to reduce the computation overhead and improve the training efficiency, a novel hard negative sampling strategy is proposed to select the most critical data samples from the globally shared dataset. In addition, a contrastive loss applied on the local face representation during training could further restrict the local model drifting. Last but not least, we are interested in simultaneously optimizing the user experience on local clients, which is not explored in previous works. Although personalized FL (Kulkarni, Kulkarni, and Pant 2020) has been studied for a while, those methods are sub-optimal on the face recognition task. We propose a Decoupled Feature Customization (**DFC**) module, which consists of a feature transformation layer and one-vs-all binary classifiers. The module locally learns a customized feature space which is optimized for recognizing the registered identities at each client.

We validate FedFR on IJB-C (Maze et al. 2018) dataset for the generic recognition model performance under different FL scenarios. We also build the personalized face recognition evaluation protocol with MS-Celeb-1M (Guo et al. 2016) dataset to validate the effectiveness of the proposed DFC module. Each technique in FedFR could substantially improve both generic and personalized face representations. Our main contributions are summarized as follows:

- We propose a novel joint optimization federated learning framework FedFR, which can effectively improve both generic and personalized face recognition models under different scenarios while strictly following the privacy constraints.

- Several training techniques (hard negative sampling, contrastive regularization) are proposed and tailored for the face recognition task, and these techniques can better bridge the gap between global and local representations.
- We propose the Decoupled Feature Customization (**DFC**) module, which is the key component to enable concurrent optimization of the personalized face recognition model. The proposed binary classification objectives are also effective for optimizing the performance on each client.
- Experimental results show that our proposed solution can consistently outperform previous approaches in several challenging generic and personalized FL benchmarks.

## Related Work

**Face Recognition.** Recently, great progress has been achieved in face recognition with large-scale training data (Cao et al. 2018; Guo et al. 2016; Zhu et al. 2021), sophisticated network structures (Schroff, Kalenichenko, and Philbin 2015; He et al. 2016) and advanced designs for softmax-based loss functions (Wang et al. 2018; Deng et al. 2019; Sun et al. 2020). However, these state-of-the-art methods are not directly applicable to the federated learning setting since they assume centralized data is available on a server. Without the access to private face images from local clients, the feature learning is prohibited as the model cannot compare features between different identities. In addition, how to leverage additional identities to improve the feature incrementally based on a pre-trained face recognition model was never discussed in previous works, as they always assumed to train the model from scratch. In our federated setup, we aim to improve a publicly available pre-trained face recognition model at the server from multiple clients in a collaborative manner, while keeping the private face images and identity features at the local clients.

**Federated Learning.** Federated Learning (FL) (Li et al. 2019a; Kairouz et al. 2019; Wang et al. 2021) is a learning setup in machine learning which aims to learn a model over multiple disjoint clients while maintaining local data privacy. The most well-known and commonly used FL algorithm is FedAvg (McMahan et al. 2017), which learns a global model by averaging weight parameters across local models trained on private client datasets. Many recent works proposed to improve FedAvg from different perspectives: model convergence (Haddadpour and Mahdavi 2019; Khaled, Mishchenko, and Richtárik 2020), robustness (Bonawitz et al. 2019), communication (Konečný et al. 2016), and non-IID clients (Li et al. 2019b; Li, He, and Song 2021). Most of the previous computer vision related FL works only studied image classification tasks with small-scale datasets (e.g. MNIST, CIFAR-10). To the best of our knowledge, FedFace (Aggarwal, Zhou, and Jain 2021) is the only one which addressed the face recognition model training in the federated setup. To enhance the pre-trained FR model, it applies the spreadout regularizer (Yu et al. 2020) at the server side to ensure the identity proxies from clients are well separated. Our work differs in that we do not transmit identity prototypes as it could leak the private identity info from clients. Moreover, our work is scalable to different scenarios where each client contains more than one identity.

**Personalized Federated Learning.** Personalized FL (Kulkarni, Kulkarni, and Pant 2020) aims to learn a customized model to meet each client’s objective. Instead of training a single “general” model which is optimized for generic metric, this FL setup seeks to acknowledge the data heterogeneity among clients by constructing a “personalized” model which fits each client’s need. Many recent techniques (Liang et al. 2020; Li et al. 2021b; Chen and Chao 2021) proposed to leverage multi-task learning (MTL) (Zhang and Yang 2017) methods to incorporate clients’ task objectives into the FL framework. Another stream of approaches (Chen et al. 2018; Fallah, Mokhtari, and Ozdaglar 2020) employed meta-learning to learn a decent initial model that can be adapted to each client after some steps of local fine-tuning. Besides, (Yu, Bagdasaryan, and Shmatikov 2020) showed that conducting post-processing (e.g. fine-tuning) onto a generic FL model could achieve comparable results with other personalized methods. However, the latter two streams of approaches would require an additional stage for local adaptation. Our framework employs the MTL based approach which can optimize general and customized face recognition models simultaneously.

## Proposed Method

In this work, we build a novel FL framework for the face recognition (FR) task. In the following, we will first establish the proposed FL setup for joint generic and personalized face recognition. Next, we introduce some preliminaries of our framework, which are some basic techniques popularly employed in FR and FL respectively. Then, we will describe technical details of the proposed FedFR solution.

## Problem setup

Face recognition systems are widely applied on local user devices. Typically, the deployed model is trained on a public dataset in advanced on a server. To continuously improve the generic face representation, the intuitive way is to collect the images stored in local devices (clients) and update the model trained with augmented data. However, as mentioned previously, due to privacy issues, it is prohibited to upload any identity-related information, such as the face images and its features. Federated learning (FL) provides a framework to train models where multiple clients collaboratively learn without sharing their data with the server or with other clients. Different from typical FL setting that learns the model from scratch, in face recognition, we target on **how to enhance the generic representation of pre-trained model by leveraging the data on clients under the privacy constraint**. Besides, we also focus on the optimized user experience. Although an improved generic model can implicitly achieve it, a client-specific personalized model optimized by local objectives could achieve optimal performance on the device. Thus, we jointly consider the situation that **whether we can obtain a personalized face model which is dedicated to recognize the registered identities on each client**. To the best of our knowledge, we are the first to explore the personalized FL setup in face recognition.

## Preliminaries

**Face Recognition.** FR is an open-set problem, where the classes (identities) in training and testing are different. In the training phase, current FR methods are typically based on an identity classification objective, where the model embeds an input image into a high-dimensional representation and generates the class logits by computing the similarity between the input feature and all class embeddings (proxies). Then a softmax cross-entropy loss will be adopted to supervise the model. In our setting, the pre-trained generic face model is trained with the commonly used Cosface loss (Wang et al. 2018), which adopts an additive margin softmax. Formally, given the face embedding model  $\Theta$  and an input image  $x$  with  $y$ -th class, we can obtain its deep feature  $f = \Theta(x) \in \mathcal{R}^d$ . There is also a class embedding matrix  $\Phi \in \mathcal{R}^{d \times K}$ , where  $K$  is the total number of classes and the  $j$ -th column  $\Phi_j$  means the learned proxy of  $j$ -th class. Following Cosface loss, the original  $j$ -th logit ( $\Phi_j \cdot f + b$ ) will be simplified by ignoring the bias  $b$  and normalizing the  $\|f\|$  and  $\|\Phi_j\|$  to 1, which is just the cosine similarity  $\cos \theta_j$ . Last, the additive margin softmax cross-entropy loss for  $x$  will be computed as follows:

$$\mathcal{L}_{cos} = -\log \frac{e^{s(\cos \theta_y - m)}}{e^{s(\cos \theta_y - m)} + \sum_{j \neq y}^K e^{s \cos \theta_j}}, \quad (1)$$

where  $s$  and  $m$  are the scaling constant and the additive margin, respectively. During the testing stage, the learned face embedding model  $\Theta$  will embed the query face image into a  $d$ -dim face feature, and the system would compare the cosine distance between the query feature and pre-registered features for identity authentication.

**Federated Learning.** In our FL setup, we consider  $C$  local client nodes and one central server with the face recognition model  $\Theta_g^0$  pre-trained on a publicly available large dataset  $D_g$ , which has  $N_g$  images from  $K_g$  identities. Each local client  $i$  is initialized with  $\Theta_{l(i)}^0 = \Theta_g^0$  and registered with  $N_{l(i)}$  images from  $K_{l(i)}$  identities, which is much smaller than the public one. Our objective is to simultaneously improve the model  $\Theta_g$  for generic face representation and optimize each  $\Theta_{l(i)}$  for personalized client customization under the privacy constraints. We adopt the most commonly used FL algorithm, **FedAvg** (McMahan et al. 2017), as our baseline method. Due to the mutual exclusive classes between local clients, we follow previous FL works (Zhuang et al. 2020; Li et al. 2021a) that only send the backbone model  $\Theta$  to the server, and keep the class embedding matrix on clients. The steps for collaborative training by server and clients are as follows:

1. In the  $t$ -th communication round, the server sends the global model  $\Theta_g^t$  to all client nodes.
2. The  $i$ -th client updates the model  $\Theta_{l(i)}^t$  at round  $t$  based on  $N_{l(i)}$  local data and local learned class embedding  $W_{l(i)}$  with Cosface loss  $\mathcal{L}_{cos}$ , which is a  $K_{l(i)}$ -class classification problem.
3. The local clients only send the backbone model  $\Theta_{l(i)}^t$  to the server. The server will update the global model by

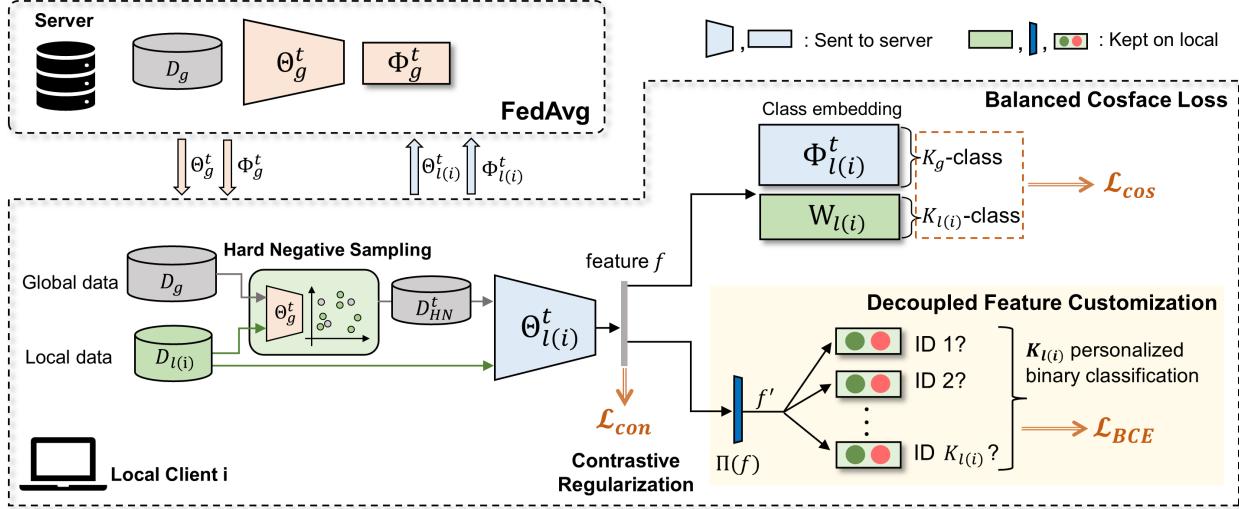


Figure 2: **FedFR.** We demonstrate the overall architecture of our method. For model on each client  $i$ , it will be optimized with balanced Cosface loss, a contrastive regularization and the binary cross entropy in our Decoupled Feature Customization branch. After training, the backbone model  $\Theta_{l(i)}$  and global class embedding  $\Phi_{l(i)}$  will be uploaded for FedAvg.

taking a weighted average of them as follows:

$$\Theta_g^{t+1} = \frac{1}{N} \sum_{i \in [C]} N_{l(i)} \cdot \Theta_{l(i)}^t, \quad (2)$$

where  $N$  is the total number of training images across all client nodes.

- Last, the updated global model will then be transmitted to each client and steps 2–4 are repeated until convergence.

FedAvg can perform well on clients with IID-distributed data. However, for our face recognition setup, the identity distributions on each client are different. Just optimizing on local data with limited number of identities to obtain  $\Theta_{l(i)}^t$  could harm the original performance of the pre-trained model (as shown in the experimental results). Furthermore, although  $\Theta_{l(i)}^t$  can improve the personalized representation for these  $K_{l(i)}$  identities, it will be continuously updated by the global model along the communication rounds, which cannot achieve optimal performance for the local users.

### FedFR: Joint Optimization Federated Framework

To tackle the issues in FedAvg, we propose a joint optimization framework **FedFR**, which can effectively improve the generic face representation at the server with the use of globally shared data, and also optimize the personalized recognition performance simultaneously at local clients. We first provide an overview of FedFR, and the system architecture is also illustrated in Figure 2. Built upon the baseline FL pipeline, we introduce several novel techniques: **1)** We employ the globally shared dataset  $D_g$  to better regularize the local model training, which could prevent the model from over-fitting on local identities. **2)** The Hard Negative Sampling strategy is introduced to select the most critical data from  $D_g$  to significantly reduce the computation on local clients. **3)** The Contrastive Regularization is employed to

control the drift of model parameters and better bridge the gap between global and local representations. **4)** To simultaneously optimize face representation for local clients, we propose the **Decoupled Feature Customization** module to transform the global representation for better fitting the local distributions. The corresponding margin-based binary classification loss  $\mathcal{L}_{BCE}$  establishes a better local objective to supervise the learning of the decoupled branch. We elaborate each technique in details as follows.

**Leveraging Globally Shared Data.** Some previous FL works on image classification (Zhao et al. 2018; Lin et al. 2020) has shown that leveraging globally shared dataset can better address the issue of heterogeneous clients. In the face recognition FL setup, the global dataset  $D_g$  which was used for pre-training the server model can be naturally shared to all the local clients. We could further regularize the training of local clients by providing the class embedding matrix  $\Phi_g$  of the shared  $K_g$  identities. As shown in Figure 2, given the shared dataset  $D_g$  on client  $i$ , the local client could build a more “balanced objective” by concatenating  $\Phi_{l(i)}^t = \Phi_g^t$  with the local private embedding matrix  $W_{l(i)}$  as a new learnable proxies and learn to classify  $K_g + K_{l(i)}$  identities with  $\mathcal{L}_{cos}$ . Thus, our balanced Cosface loss would be formulated as:

$$\mathcal{L}_{cos} = -\log \frac{e^{s(\cos \theta_y - m)}}{e^{s(\cos \theta_y - m)} + \sum_{j \neq y}^{K_g + K_{l(i)}} e^{s \cos \theta_j}}, \quad (3)$$

where the denominator is added with additional  $K_g$  negative terms. For the end of each round  $t$ , beside sending the backbone  $\Theta_{l(i)}^t$  back to server, the learned class embeddings  $\Phi_{l(i)}^t$  related to  $K_g$  global identities can also be sent back and updated by:

$$\Phi_g^{t+1} = \frac{1}{N} \sum_{i \in [C]} N_{l(i)} \cdot \Phi_{l(i)}^t. \quad (4)$$

**Hard Negative Sampling Strategy.** Jointly training with  $D_g$  can prevent model from over-fitting on local data. However, the large number of public data will also increase the computation burden on local clients, which will enlarge the training time and degrade the communication efficiency between server and clients. To obtain a better trade-off, we propose a hard negative (HN) sampling strategy to only choose a subset  $D_{HN}$  from  $D_g$ , which is critical for learning with  $D_{l(i)}$ . The proposed technique is described as follows.

At the start of each communication round  $t$  on local client  $i$ , we first forward the global and local data to  $\Theta_g^t$  to generate their features. Then we can calculate the pair-wise cosine similarity between them. To make the training more efficient but at the same time maintain the performance, we only sample the “hard” global data for model learning, which is with similarity larger than threshold  $t_{HN}$  to any of the local data. Intuitively, with larger  $t_{HN}$ , the less global data will be used for training. We decide the threshold by leveraging the inherent feature space of the pre-trained model. As mentioned above, the pre-trained model is trained with Cosface loss, where the similarity of each sample to its proxy should be larger than those to others by a margin  $m$ . Thus, if any negative pair with similarity larger than  $t_{HN} = m$ , they should be served as a hard negative pair.

**Contrastive Regularization on Local Clients.** Inspired by the related work (Li, He, and Song 2021), which proposed a model-contrastive loss on the local training to prevent local model from deviating too much from the global model, we also apply the similar regularization on our face recognition task. Namely, we aim to decrease the distance between the face representation learned by the local model at time  $t$  ( $f = \Theta_{l(i)}^t(x)$ ) and the one learned by the global model ( $f_{glob} = \Theta_g^t(x)$ ), and increase the distance between the face representation learned by the local model at time  $t$  ( $f = \Theta_{l(i)}^t(x)$ ) and time  $t - 1$  ( $f_{prev} = \Theta_{l(i)}^{t-1}$ ). Thus, the local contrastive loss term  $\mathcal{L}_{con}$  is defined as

$$\mathcal{L}_{con} = -\log \frac{\exp(\text{sim}(f, f_{glob})/\tau)}{\exp(\text{sim}(f, f_{glob})/\tau) + \exp(\text{sim}(f, f_{prev})/\tau)}, \quad (5)$$

where “ $\text{sim}(\cdot, \cdot)$ ” measures the cosine similarity between face features, and  $\tau$  denotes a temperature hyperparameter.

**Decoupled Feature Customization.** With the contrastive regularization, the local model can avoid over-parameterizing for the local objective and continuously improve the generic face representation. However, it will go against the goal which we aim to simultaneously obtain a personalized model to improve the local user experience. Thus, as shown in Figure 2, we propose a novel Decoupled Feature Customization (DFC) module to resolve this seemly contradicting scenario. In order not to influence the feature  $f$  for generic representation, we adopt a transformation  $\Pi(f)$  with a fully-connected layer to map it to a client-specific feature space, which can recognize the  $K_{l(i)}$  identities well. To achieve this goal, there should be a local objective for optimization. Inspired by (Wen et al. 2021), we propose to adopt the binary classification on each local identity for the personalized purpose. Given the transformed feature

$f' = \Pi(f)$ , we will feed it into  $K_{l(i)}$  binary classification branches (which the total trainable weight vectors are denoted as  $\Omega_{l(i)}$ ). The  $k$ -th module contains learnable parameters which only target on classifying the positive samples from the  $k$ -th class and the negative samples from “any other” classes. Formally, we follow the loss in the related work that used margin-based binary cross-entropy ( $\mathcal{L}_{BCE}$ ) to supervise our personalized branch:

$$\begin{aligned} \mathcal{L}_{BCE} = & \frac{\lambda}{s'} \cdot \log \left( 1 + \exp(-s' \cdot (g(\cos \theta_k) - m') - b) \right) \\ & + \frac{1 - \lambda}{s'} \cdot \sum_{j \neq k} \log \left( 1 + \exp(s' \cdot (g(\cos \theta_j) + m') + b) \right), \end{aligned} \quad (6)$$

where  $\cos \theta_j$  is the cosine similarity of transformed input feature  $f'$  and the  $j$ -th weight vector  $\Omega_{l(i),j}$  in the binary classification,  $b$  is the learned bias, and the function  $g(z) = 2((z+1)^t/2) - 1$  is used to increase the empirical dynamic range of cosine similarity. The notations  $\lambda, s'$  and  $m'$  all follow those in the related work, which are the balanced factor, scaling constant and cosine margin.

It is worth mentioning that although there are only  $K_{l(i)}$  binary classification branches, not only the local data but the global data can be used to optimize our DFC module because each branch only needs to recognize “whether it is the  $k$ -th identity or not”. This objective just well-fits our personalized goal that given an unseen query image, a well-performed local face recognition system should quickly determine whether it is the registered identity or not.

**Learning Pipeline** Our overall learning framework is based on FedAvg, where there will be  $T$  communication rounds and in each round, the local clients will update the model for  $E$  epochs. In the local client training, the model will be optimized in an end-to-end manner with the total objective  $\mathcal{L}_{total}$ , which is formulated as:

$$\mathcal{L}_{total} = \alpha_1 \mathcal{L}_{cos} + \alpha_2 \mathcal{L}_{con} + \alpha_3 \mathcal{L}_{BCE}, \quad (7)$$

where all the modules  $\Theta_{l(i)}^t, \Phi_{l(i)}^t, W_{l(i)}^t, \Pi_{l(i)}^t, \Omega_{l(i)}^t$  and bias  $b$  would be updated. However, only the  $\Theta_{l(i)}^t$  and  $\Phi_{l(i)}^t$  will be sent back for globally averaged with Equation 2 and 4. In the testing phase,  $\Theta_g$  is used for generic evaluation and  $[\Theta_{l(i)}, \Pi_{l(i)}]$  is used for personalized evaluation.

## Experiments

### Experimental Setup

**Dataset** We use the MS-Celeb-1M (Guo et al. 2016) as the training dataset. To avoid the long-tail distribution, we manually select 10k identities from the dataset where each identity contains 100 face images. Within the selected subset, 6000 identities ( $K_g$ ) are used for pre-training the global model, and the other 4000 identities are equally distributed into local clients. For each identity in each local client, we use 60 images for local training, and 40 images for personalized model evaluation, respectively. Besides MS-Celeb-1M, IJB-C (Maze et al. 2018) dataset which contains 3531 identities with diverse appearance is used for evaluating the generic model performance. The selected list for FL training will be released for fair comparison in the future.

Table 1: **Ablation Studies.** We conduct FL experiments with 40 clients; each client contains 100 identities. (results are in %)

Setup	Modules			Generic Evaluation (IJB-C)				Personalized Evaluation			
	HN. sampled Global data	Contrastive	DFC. Branch	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 1e-4	1:1 TAR @ FAR 1e-6	1:N TPIR @ FPIR 1e-5	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 1e-4		
Centrally trained on 6k IDs (pre-training)				76.42	84.58	72.06	80.30	56.28	72.50	71.73	82.33
Federated Learning on 4k IDs	✗	✗	✗	73.79	83.71	67.59	78.53	67.33	85.70	82.77	92.27
	✓	✗	✗	76.79	84.64	72.76	80.76	81.75	91.91	91.97	96.09
	✓	✓	✗	77.41	85.17	73.60	81.25	77.77	89.57	89.58	94.60
	✓	✓	✓	<b>77.60</b>	<b>85.21</b>	<b>73.60</b>	<b>81.27</b>	<b>88.32</b>	<b>95.46</b>	<b>95.17</b>	<b>97.94</b>
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

**Evaluation Metrics** For the generic model evaluation, we strictly follow the IJB-C evaluation protocol, which is commonly used in the face recognition community. We report the true acceptance rates (TAR) at different false acceptance rates (FAR) for 1:1 verification protocol, and true positive identification rates (TPIR) at different false positive identification rates (FPIR) for 1:N identification protocol.

Regarding the personalized model evaluation, we carefully build up the metrics and protocols as we are the first to investigate the personalized face recognition setup. The evaluation is supposed to only focus on the face recognition user experience of the registered identities on each local client. Therefore, we establish two evaluation protocols to better measure the client-specific performance: 1) Firstly, similar to the 1:1 verification protocol in IJB-C, we establish a list of positive pairs and negative pairs for evaluation. In each client, we formulate genuine matches from local identities and build up imposter matches by pairing one local identity with a random identity from other clients. For the 40 local clients scenario where each client is registered with 100 identities, there are 7.8k positive pairs and about 630 million negative pairs in one client. We average the true acceptance rates (TAR) across all clients as the final personalized verification performance. 2) Secondly, we build up an 1:N identification protocol to estimate the login experience on a local client (device). Intuitively, the registered images from one local identity are combined to form its gallery feature. And the testing images from all clients are taken as the probe features. For the 40 local clients scenario, there are 100 gallery features and 160k probe features in one client. Similarly, we average the true positive identification rates (TPIR) across all clients as the final personalized identification performance.

**Implementation Details** For the backbone face model, we adopt the same 64-layer CNN architecture from (Liu et al. 2017; Wang et al. 2018), which outputs a 512-dimensional feature vector. The image preprocessing techniques are the same as (Deng et al. 2019), where the image is cropped to size  $112 \times 112$  and the pixel value is normalized to  $[-1, 1]$ . To simplify our network training, all hyper-parameters in  $\mathcal{L}_{cos}$ ,  $\mathcal{L}_{con}$  and  $\mathcal{L}_{BCE}$  are empirically set as the same ones in the related work, where  $m=m'=0.4$ ,  $s=s'=30$ ,  $\tau=0.5$ ,  $\lambda=0.7$  and  $t'=3$ . For  $\mathcal{L}_{total}$ , the  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  are empirically set as 1, 5 and 10. We adopt SGD optimizer with weight decay  $5 \times 10^{-4}$  and learning rate 0.001. For the FL setup, we conduct  $T=30$  communication rounds and in each round the

local clients conduct  $E=4$  epochs.

## Ablation Studies

**Effectiveness of each modules** To validate the effectiveness of each proposed module, we report the ablation studies in Table. 1. The experiments are conducted with one central server and 40 clients, where each client contains 100 identities. The performance is evaluated both on the generic and personalized benchmark. If it is under the FL setup, the global model  $\Theta_g$  will be used to test on the generic evaluation and each local model  $\Theta_{l(i)}$  will be tested on personalized data, where the shown scores are the average over all clients. Notes that for the 1:N identification in personalized evaluation, we average the feature of training images based on their identities as the gallery features in that client.

The first row is the performance of pre-trained model trained on public data with 6k classes, which is the target model that needs to be improved. Start from 2<sup>nd</sup> to 5<sup>th</sup> row, the FL setup is employed where 4k augmented IDs are added but with privacy constraints. And for the last row, it is the ideal situation that we can centrally optimize the model with data of 10k IDs. We can see that in the second row, our baseline method which directly optimizes the model with local data and perform FedAvg on the server cannot perform well. The performance is even worse than the pre-trained one owing to the over-fitting on local data. Leveraging the public data is a solution, but it may suffer from long training time and large computation overhead. With our proposed Hard Negative sampling strategy where only a subset of global data serving as negative pairs to the local data, in the 3<sup>rd</sup> row, not only the generic representation but also the personalized evaluation can be boosted. Contrastive regularization is designed for regularizing the local model from training towards the undesired local minimum. We can see that in the 4<sup>th</sup> row, the performance improves greatly on generic evaluation. However, under the same feature space parameterized by  $\Theta$ , a more generalized representation will harm the performance for recognizing specific identities on clients. Thus, in the 5<sup>th</sup> row, which is our final proposed FedFR architecture with the DFC branch, we decouple the feature from the original feature space to a new one with a transformation  $\Pi_{l(i)}$ , and optimize this space with binary cross-entropy loss tailored for the personalization. We can see that with  $\Theta_g$  for generic representation and  $[\Theta_{l(i)}, \Pi_{l(i)}]$  for personalized evaluation, both of them can achieve superior results.

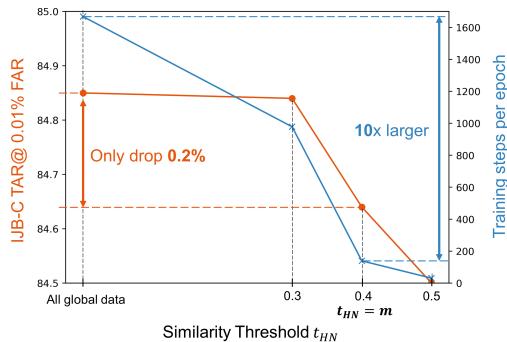


Figure 3: The generic model performance and the model training efficiency under different Hard Negative thresholds.

**Analysis of the  $t_{HN}$  in Hard Negative Sampling** In our experiments, we choose  $t_{HN}$  equals to the margin  $m=0.4$  in Cosface used in pre-training the model. To validate the effectiveness, as shown in Figure 3, we demonstrate the global performance on IJB-C and its training efficiency under different hard negative thresholds with 100 IDs per client. The training efficiency is measured in terms of the training steps per epoch. We can see that with  $t_{HN}=0.4$ , the number of sampled global data can be largely reduced by 10 times but with only 0.2% drop of the global performance, which is the best trade-off configuration in our experiments.

### Comparison with FedFace

To compare the results with FedFace (Aggarwal, Zhou, and Jain 2021), as shown in Figure 4, we construct the FL setting with diverse identities per client under total 100 clients, which is from 40 to 1. We demonstrate the results of the pre-trained model, ideal central training (upper bound), FedFace and our proposed FedFR. Because FedFace cannot be adopted on multiple IDs in a client and their FL dataset is not released, we re-implement their method on our setting that uses Cosface loss as the local objective if the number of ID is larger than 1, and also apply spreadout regularizer at the server side to separate the class proxies from clients. From the comparison on the generic model performance, FedFace could easily over-fit on local dataset and performs inferior to the pre-trained model in these scenarios. In contrast, our proposed FedFR can still improve the generic face representation under the most challenging scenario where there is only one identity in the client.

### Comparison with other Personalized FL methods

To validate the effectiveness of our Decoupled Feature Customization (DFC) module, we compare with the latest personalized FL method (Yu, Bagdasaryan, and Shmatikov 2020), which is a two-stage local adaptation approach. For fair comparison, we re-implement the “Fine-tune” and “KD” local adaptation methods, which were shown to be effective in image classification tasks, in our face recognition setup. In the first stage, the server and clients collaboratively learn to obtain a great generic model, where we use the proposed hard negative sampling strategy and the contrastive regularization in the experiments. Then, in the second stage, each

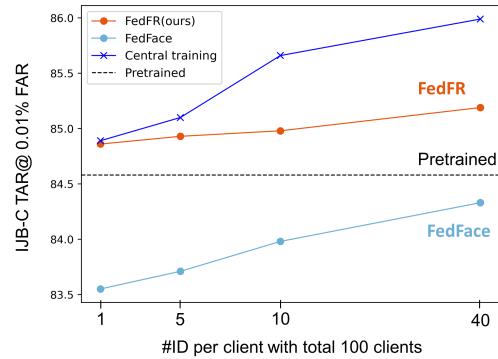


Figure 4: Generic face recognition model performance compared to FedFace. We fix the number of clients to 100 and conduct 4 scenarios of different #IDs in one client.

Table 2: Comparison of other personalized techniques. It is conducted on 40 clients with 100 IDs per each.

Method	Modules	Personalized Evaluation			
		1:1 TAR @ FAR 1e-6	1:N TPIR @ FPIR 1e-5	1:N TPIR @ FPIR 1e-5	1:N TPIR @ FPIR 1e-4
Yu et al. 2020	Fine-tune KD	73.81 75.82	86.21 87.65	88.37 89.50	93.90 94.67
<b>Ours</b> (w/ branch)	Cosface <b>BCE</b>	82.93 <b>88.32</b>	91.88 <b>95.46</b>	90.67 <b>95.17</b>	95.59 <b>97.94</b>

client separately optimizes its local model for personalization. For the “Fine-tune” method, we directly optimize each model with Cosface loss with the local and sampled global data. For the “KD” method, it is with a Knowledge Distillation technique that besides the original Cosface loss, we also supervise the output logits of local model (student) by the logits generated from original global model (teacher) with KL-Divergence loss. As illustrated in Table. 2, our proposed one-stage personalization method can outperform the two local adaptation strategies. In addition, we also conduct a variant of our method, which is also a decoupled branch but adopts a Cosface loss with multi-class classification for supervision. It is clearly verified that the proposed binary classification objective better fits the need for the personalized face recognition on clients.

### Conclusion

In this paper, we address the face recognition model training under the practical federated learning setting, where each client is initialized with the pre-trained model. We propose a novel joint optimization framework FedFR, which can improve the generic face representation of the global model and at the same time enhance the personalized user experience. While the proposed hard negative sampling and contrastive regularization can efficiently bridge the gap between global and local training, the Decoupled Feature Customization (DFC) module is another novel component to enable concurrent optimization of the personalized face recognition model. The effectiveness of the proposed solution is verified on several challenging generic and personalized face recognition benchmarks. We hope that the work and the release of the personalized FR benchmark can facilitate the future research on the federated learning for face recognition.

## Acknowledgment

This research was supported in part by the Ministry of Science and Technology of Taiwan (MOST 109-2218-E-002 -026), National Taiwan University (NTU-108L104039), Intel Corporation, Delta Electronics and Compal Electronics.

## References

- Aggarwal, D.; Zhou, J.; and Jain, A. K. 2021. FedFace: Collaborative Learning of Face Recognition Model. *arXiv preprint arXiv:2104.03008*.
- Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; McMahan, H. B.; et al. 2019. Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046*.
- Cao, Q.; Shen, L.; Xie, W.; Parkhi, O. M.; and Zisserman, A. 2018. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*.
- Chen, F.; Luo, M.; Dong, Z.; Li, Z.; and He, X. 2018. Federated meta-learning with fast convergence and efficient communication. *arXiv preprint arXiv:1802.07876*.
- Chen, H.-Y.; and Chao, W.-L. 2021. On Bridging Generic and Personalized Federated Learning. *arXiv preprint arXiv:2107.00778*.
- Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*.
- Duong, C. N.; Truong, T.-D.; Luu, K.; Quach, K. G.; Bui, H.; and Roy, K. 2020. Vec2Face: Unveil Human Faces From Their Blackbox Features in Face Recognition. In *CVPR*.
- Fallah, A.; Mokhtari, A.; and Ozdaglar, A. 2020. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*.
- Guo, Y.; Zhang, L.; Hu, Y.; He, X.; and Gao, J. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV*.
- Haddadpour, F.; and Mahdavi, M. 2019. On the convergence of local descent methods in federated learning. *arXiv preprint arXiv:1910.14425*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2019. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*.
- Khaled, A.; Mishchenko, K.; and Richtárik, P. 2020. Tighter theory for local SGD on identical and heterogeneous data. In *International Conference on Artificial Intelligence and Statistics*. PMLR.
- Konečný, J.; McMahan, H. B.; Yu, F. X.; Richtárik, P.; Suresh, A. T.; and Bacon, D. 2016. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*.
- Kulkarni, V.; Kulkarni, M.; and Pant, A. 2020. Survey of personalization techniques for federated learning. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*.
- Li, C.; Niu, D.; Jiang, B.; Zuo, X.; and Yang, J. 2021a. Meta-HAR: Federated Representation Learning for Human Activity Recognition. *arXiv preprint arXiv:2106.00615*.
- Li, Q.; He, B.; and Song, D. 2021. Model-Contrastive Federated Learning. In *CVPR*.
- Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; Li, Y.; Liu, X.; and He, B. 2019a. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *arXiv preprint arXiv:1907.09693*.
- Li, X.; Huang, K.; Yang, W.; Wang, S.; and Zhang, Z. 2019b. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*.
- Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; and Dou, Q. 2021b. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*.
- Liang, P. P.; Liu, T.; Ziyin, L.; Allen, N. B.; Auerbach, R. P.; Brent, D.; Salakhutdinov, R.; and Morency, L.-P. 2020. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*.
- Lin, T.; Kong, L.; Stich, S. U.; and Jaggi, M. 2020. Ensemble distillation for robust model fusion in federated learning. *arXiv preprint arXiv:2006.07242*.
- Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; and Song, L. 2017. Sphereface: Deep hypersphere embedding for face recognition. In *CVPR*.
- Maze, B.; Adams, J.; Duncan, J. A.; Kalka, N.; Miller, T.; Otto, C.; Jain, A. K.; Niggel, W. T.; Anderson, J.; Cheney, J.; et al. 2018. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR.
- Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *CVPR*.
- Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; and Wei, Y. 2020. Circle loss: A unified perspective of pair similarity optimization. In *CVPR*, 6398–6407.
- Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; and Liu, W. 2018. Cosface: Large margin cosine loss for deep face recognition. In *CVPR*.
- Wang, J.; Charles, Z.; Xu, Z.; Joshi, G.; McMahan, H. B.; Al-Shedivat, M.; Andrew, G.; Avestimehr, S.; Daly, K.; Data, D.; et al. 2021. A Field Guide to Federated Optimization. *arXiv preprint arXiv:2107.06917*.
- Wen, Y.; Liu, W.; Weller, A.; Raj, B.; and Singh, R. 2021. SphereFace2: Binary Classification is All You Need for Deep Face Recognition. *arXiv preprint arXiv:2108.01513*.
- Yu, F.; Rawat, A. S.; Menon, A.; and Kumar, S. 2020. Federated learning with only positive labels. In *ICML*.

Yu, T.; Bagdasaryan, E.; and Shmatikov, V. 2020. Salvaging federated learning by local adaptation. *arXiv preprint arXiv:2002.04758*.

Zhang, Y.; and Yang, Q. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*.

Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; and Chandra, V. 2018. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*.

Zhu, Z.; Huang, G.; Deng, J.; Ye, Y.; Huang, J.; Chen, X.; Zhu, J.; Yang, T.; Lu, J.; Du, D.; et al. 2021. WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10492–10502.

Zhuang, W.; Wen, Y.; Zhang, X.; Gan, X.; Yin, D.; Zhou, D.; Zhang, S.; and Yi, S. 2020. Performance optimization of federated person re-identification via benchmark analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*.