國立臺灣大學
National Taiwan University

Microsoft

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

Chih-Ting Liu[1*], Chien-Yi Wang[2*], Shao-Yi Chien[1], Shang-Hong Lai[2]

** denotes equal contributions*

[1]Graduate Institute of Electronics Engineering, National Taiwan University

[2]Microsoft AI R&D Center, Taiwan

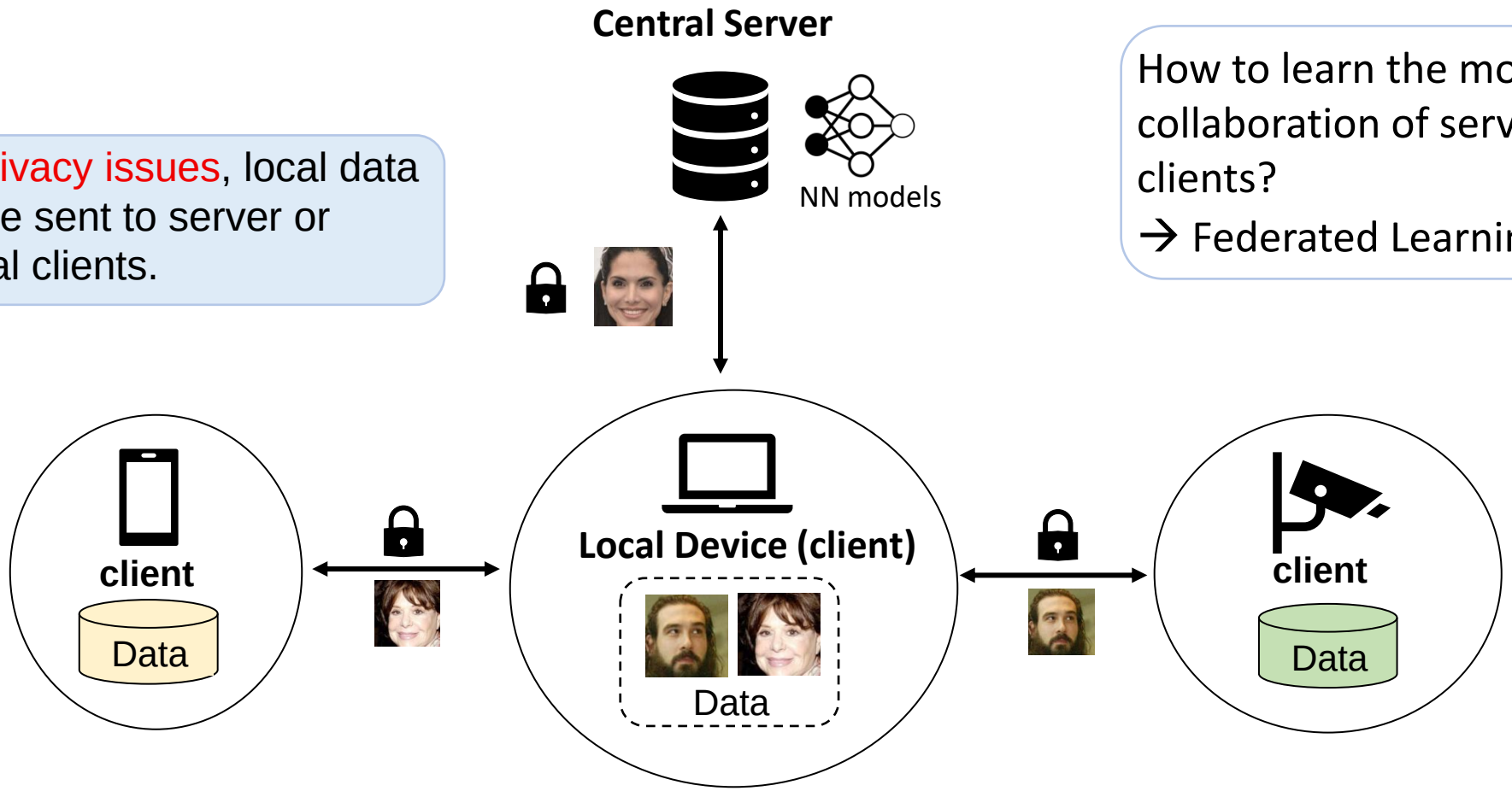
國立臺灣大學
National Taiwan University

Microsoft

Introduction of Federated Learning

What is Federated Learning (FL)

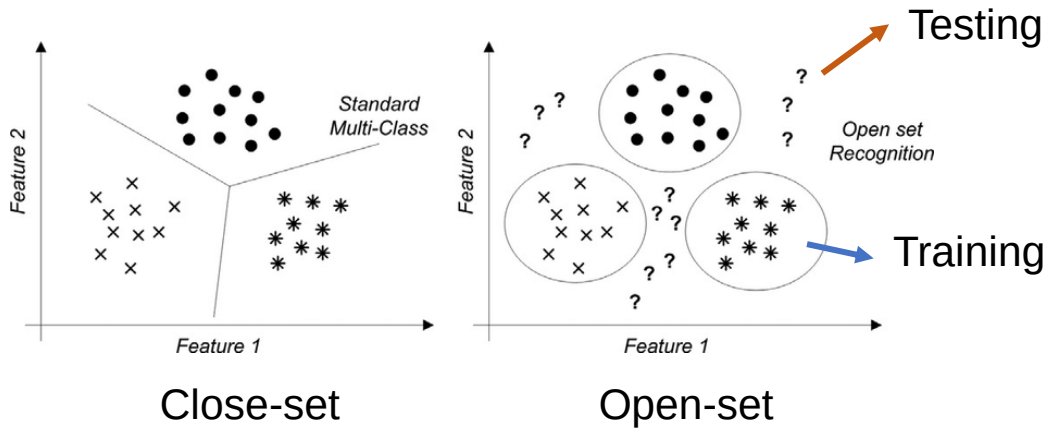
Due to privacy issues, local data **cannot** be sent to server or other local clients.



How to learn the model with the collaboration of server and local clients?
→ Federated Learning

Federated Learning for Face Recognition

- Typically, federated learning is applied on **image classification** tasks.
- What is the difference when FL is applied on **face recognition**?
  - Classification → **Close set**
  - Recognition → **Open set**



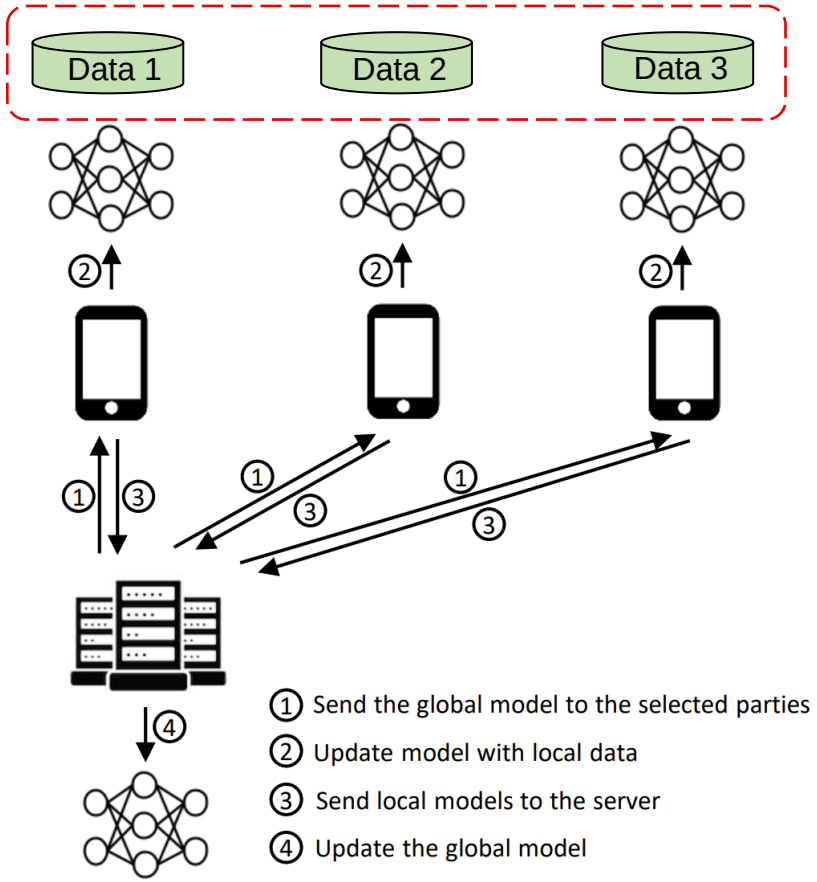
- In this paper, we formulate new FL setting and benchmark **dedicated for face recognition**.

Related Work of typical Classification Task

- Typical Pipeline of FL on Image Classification
- **Generic** Federated Learning on Image Classification
  - FedAvg [2]
  - Moon [3]
  - FL with non-IID data [4]
- **Personalized** Federated Learning on Image Classification [5,6,7]

Typical Pipeline of FL on Image Classification

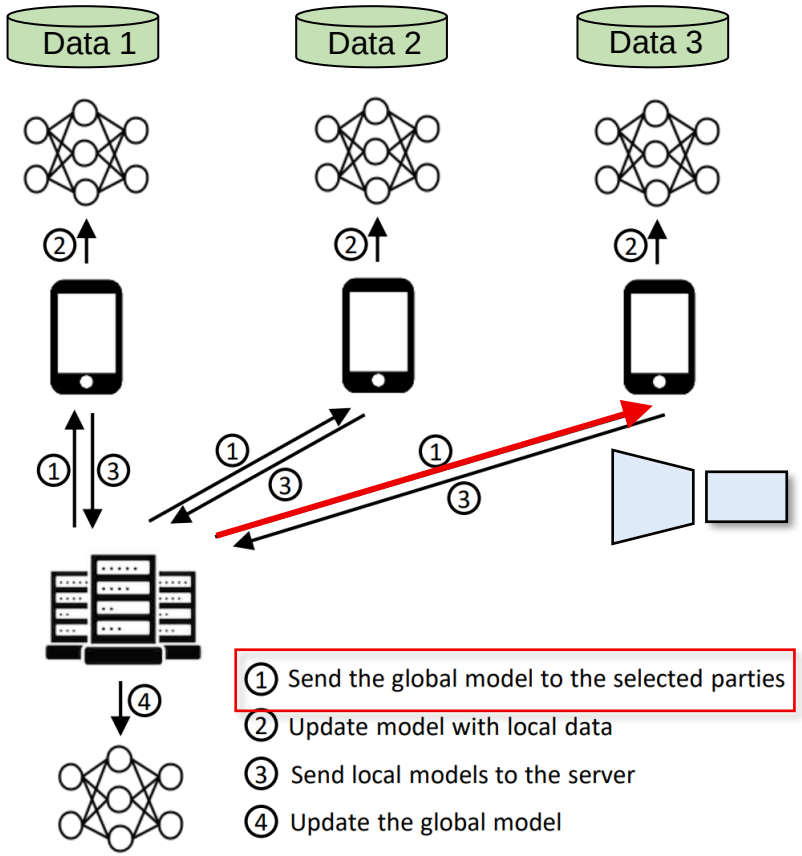
- Most papers focus on the close-set image classification task.
- A **dataset** will be split and non-IID distributed to each client (party); each contains a **fix number of classes**.



[9] Li, Qinbin, et al. "A survey on federated learning systems: vision, hype and reality for data privacy and protection." arXiv preprint arXiv:1907.09693 (2019).

Typical Pipeline of FL on Image Classification

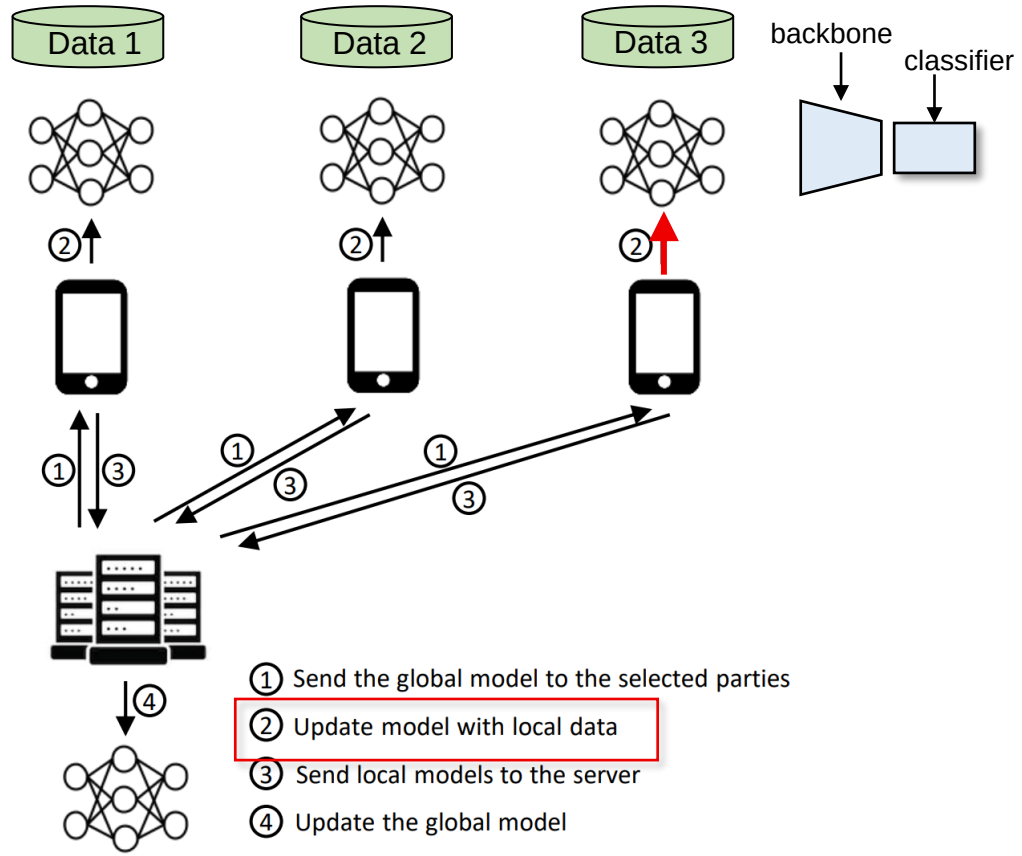
- They start from a model randomly initialized on the server.
- Send the global model to the selected parties (clients).



- The model itself contains less privacy information

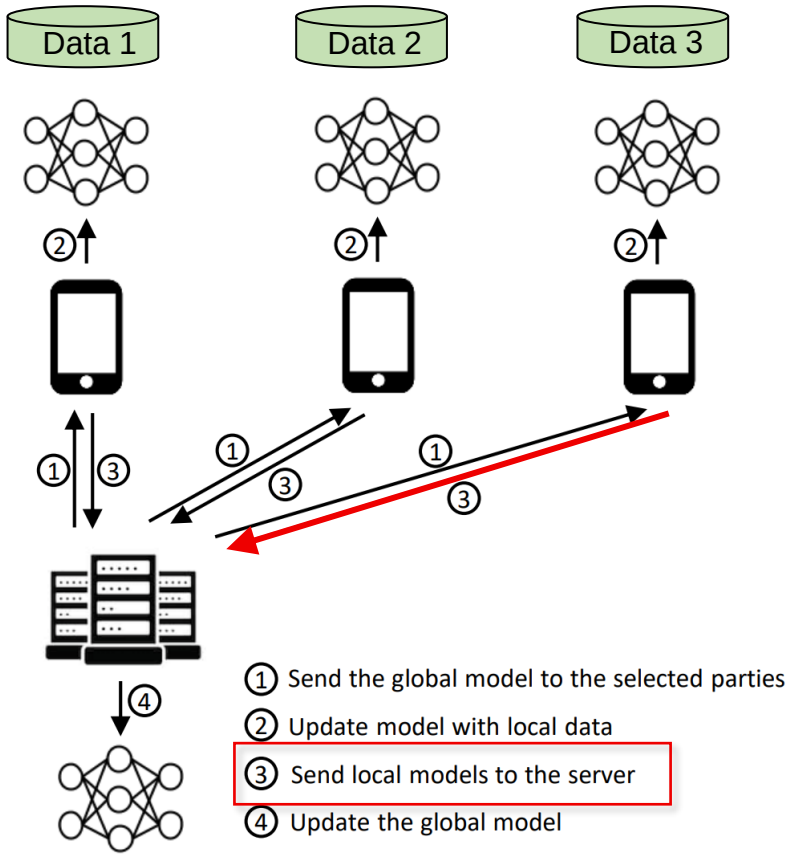
Typical Pipeline of FL on Image Classification

- Each client optimize the NN models with local data.



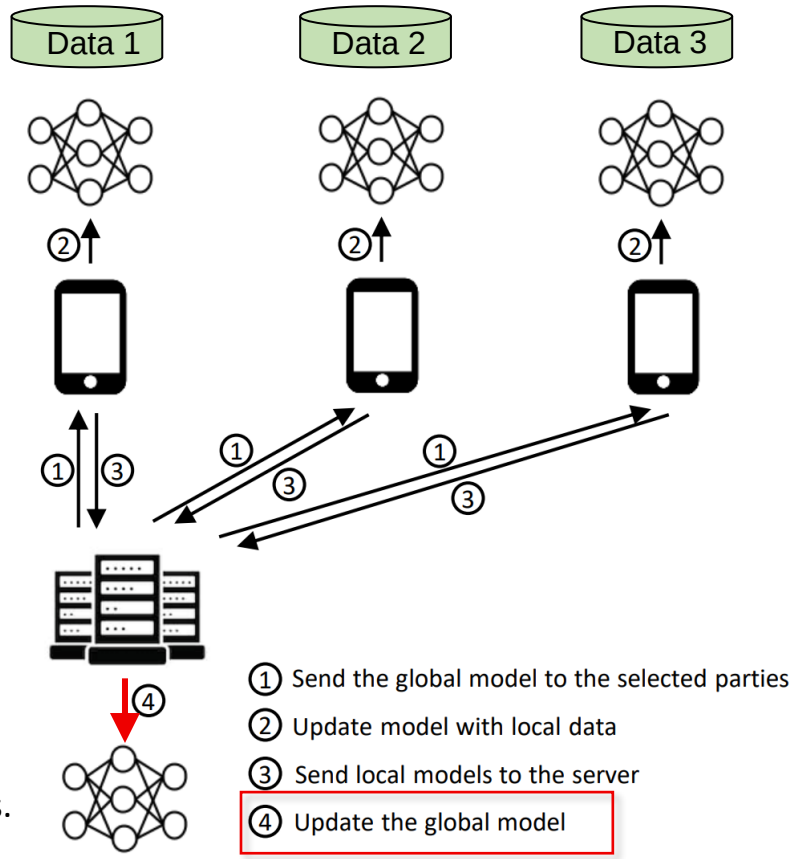
Typical Pipeline of FL on Image Classification

- Each client send the local model back to the server.



Typical Pipeline of FL on Image Classification

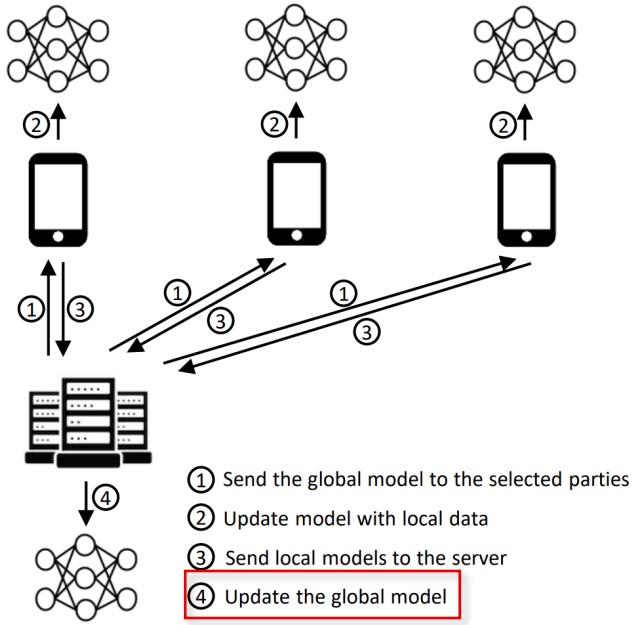
- Server aggregates the local models and updates the global model.



Repeat ① to ④ until the model converges.

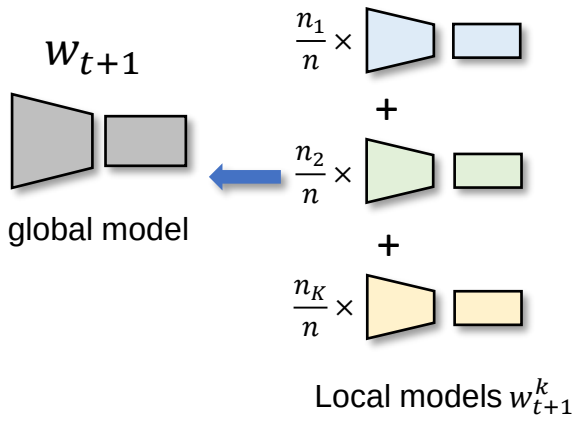
FedAvg [2]

- FedAvg is a well-known baseline FL method.
- They weighted average the local models in step 4.



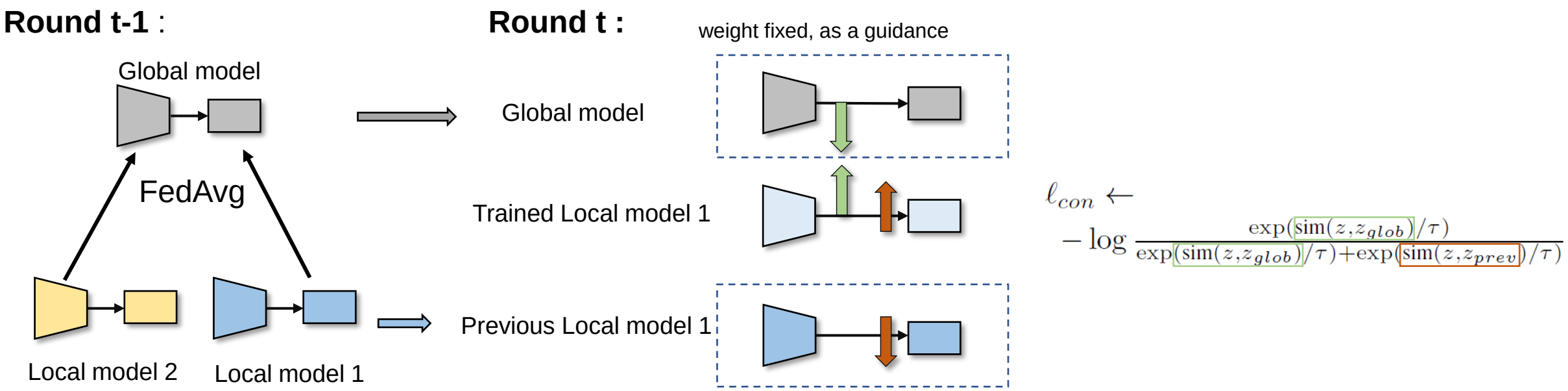
FedAvg

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$



Moon: Model Contrastive Federated Learning [3]

- To avoid overfitting on local data, this work (Moon) regularize the **feature** generated from local model to prevent it deviating too much from the global model.

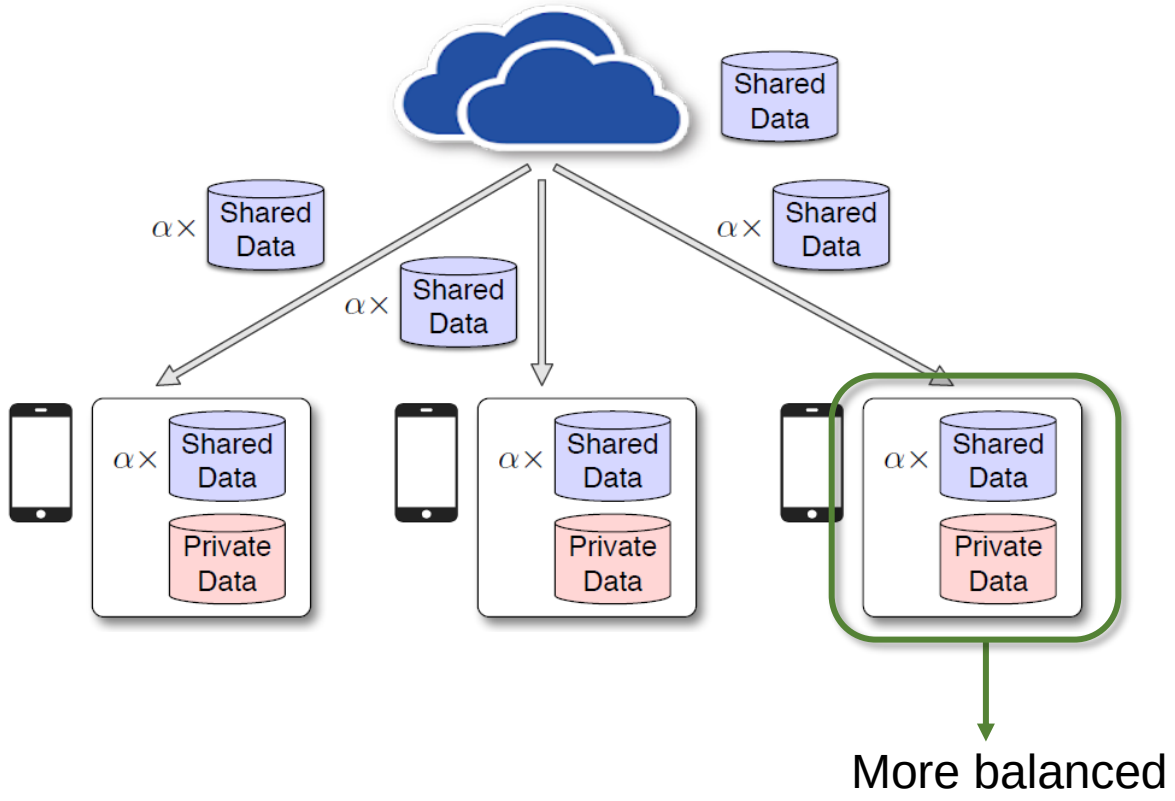


Pull the **feature** (z) close to that inferred with **global model**,

and push away the feature to that with **previous local model.**

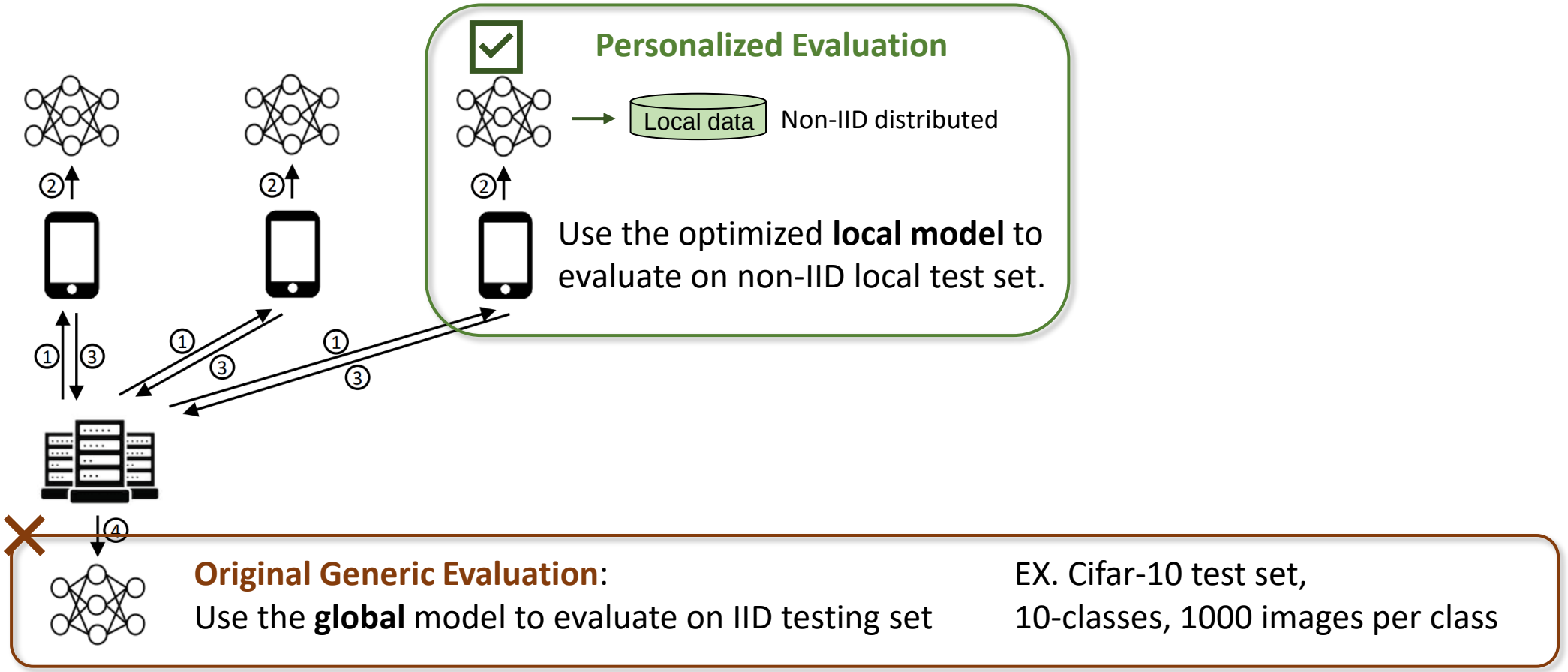
Federated Learning with Non-IID Data [4]

- If the local data is imbalanced, we can add more data!
- They leverage publicly **shared data** to make the training data more balanced.



Personalized Federated Learning

- Some papers tackle this issue [5,6,7], and there is no standard architecture or setting.

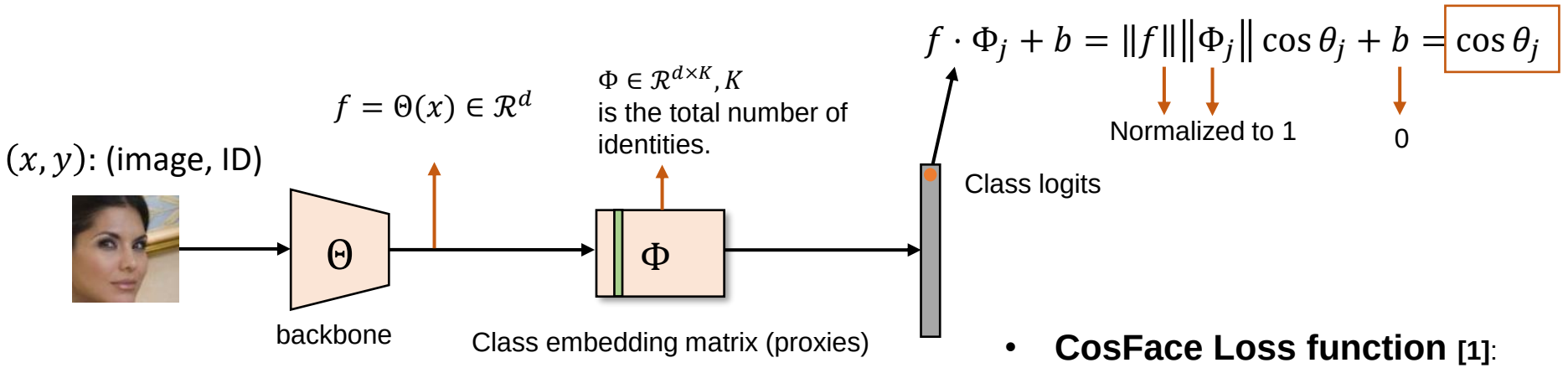


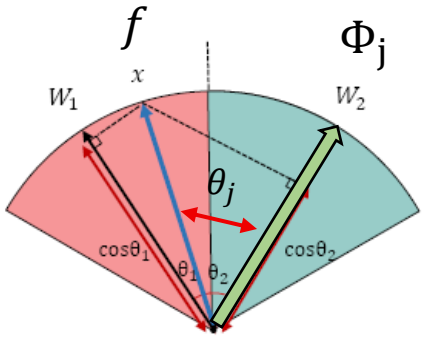
Related Work of FL on Face Recognition

- Preliminary of Face Recognition
- Federated Learning on Face Recognition
  - FedFace [8]

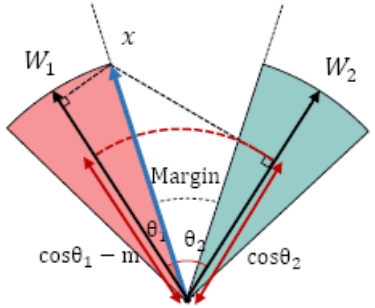
Preliminary of Face Recognition

- The training of face recognition is formulated as a classification problem





2D Hypersphere Manifold



- **CosFace Loss function [1]**:
  Cross-Entropy with Additive margin Softmax

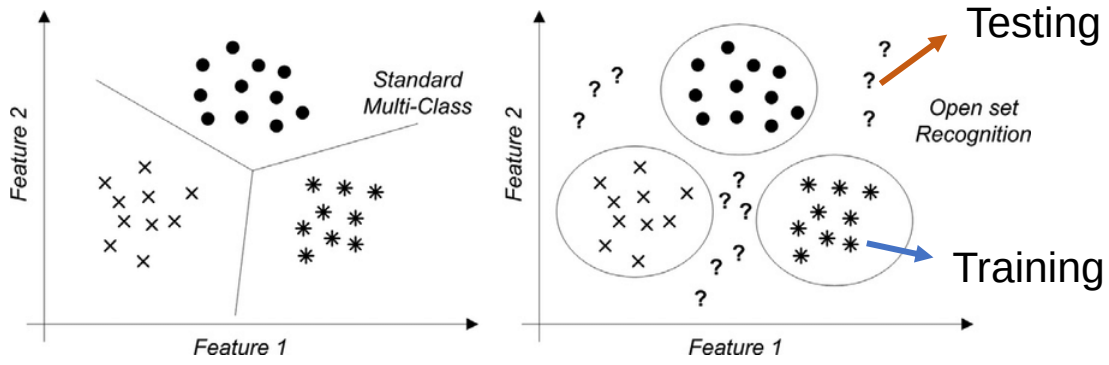
m is the margin, s is the scaling constant.

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

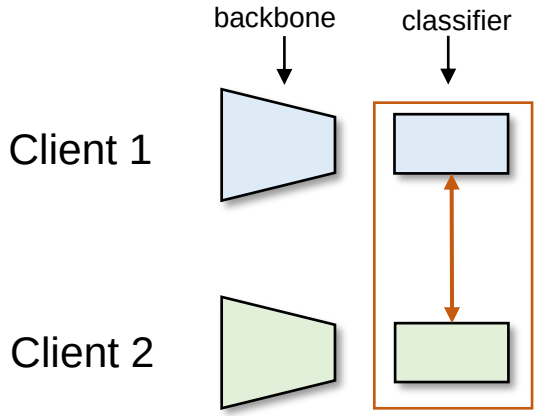
Federated Learning on Face Recognition

Typical Federated Learning cannot be directly applied on face recognition owing to :

1) Face recognition is an **open-set** problem, known classes are used for training and the unknown classes are used for testing.
2) The identity classes between local clients are different, which results in **different model architectures in clients.**



Close-set Open-set



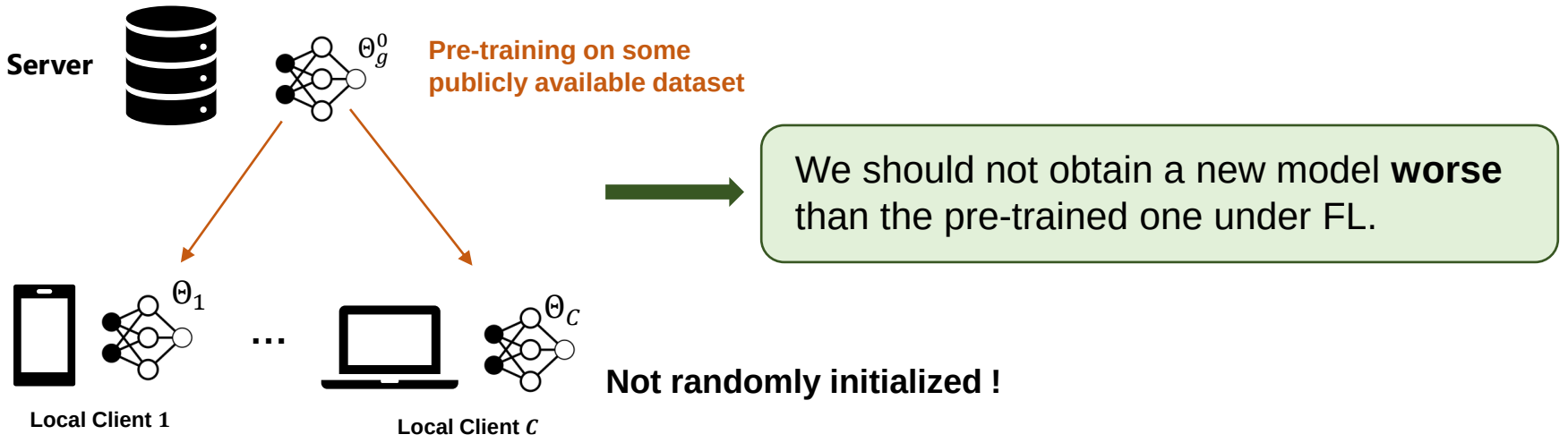
Non-overlapped IDs, the **class embeddings** are different.

FedAvg **cannot** perform on the class embedding

Federated Learning on Face Recognition

Typical Federated Learning cannot be directly applied on face recognition owing to :

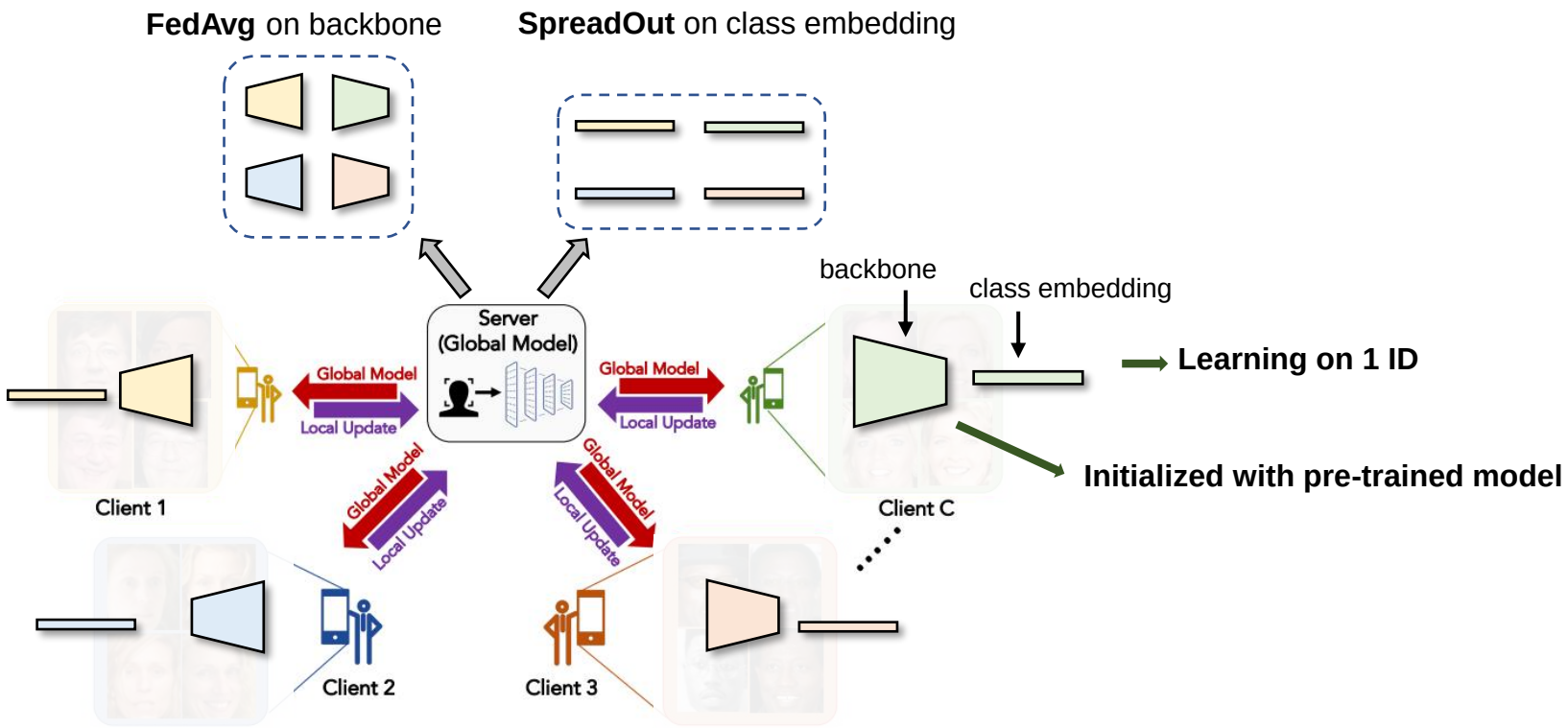
3) In a more **practical** setup for face recognition, the FL training starts from a publicly available **pre-trained model**, rather than from scratch as in traditional FL.



FedFace : Collaborative Learning of Face Recognition Model [8]

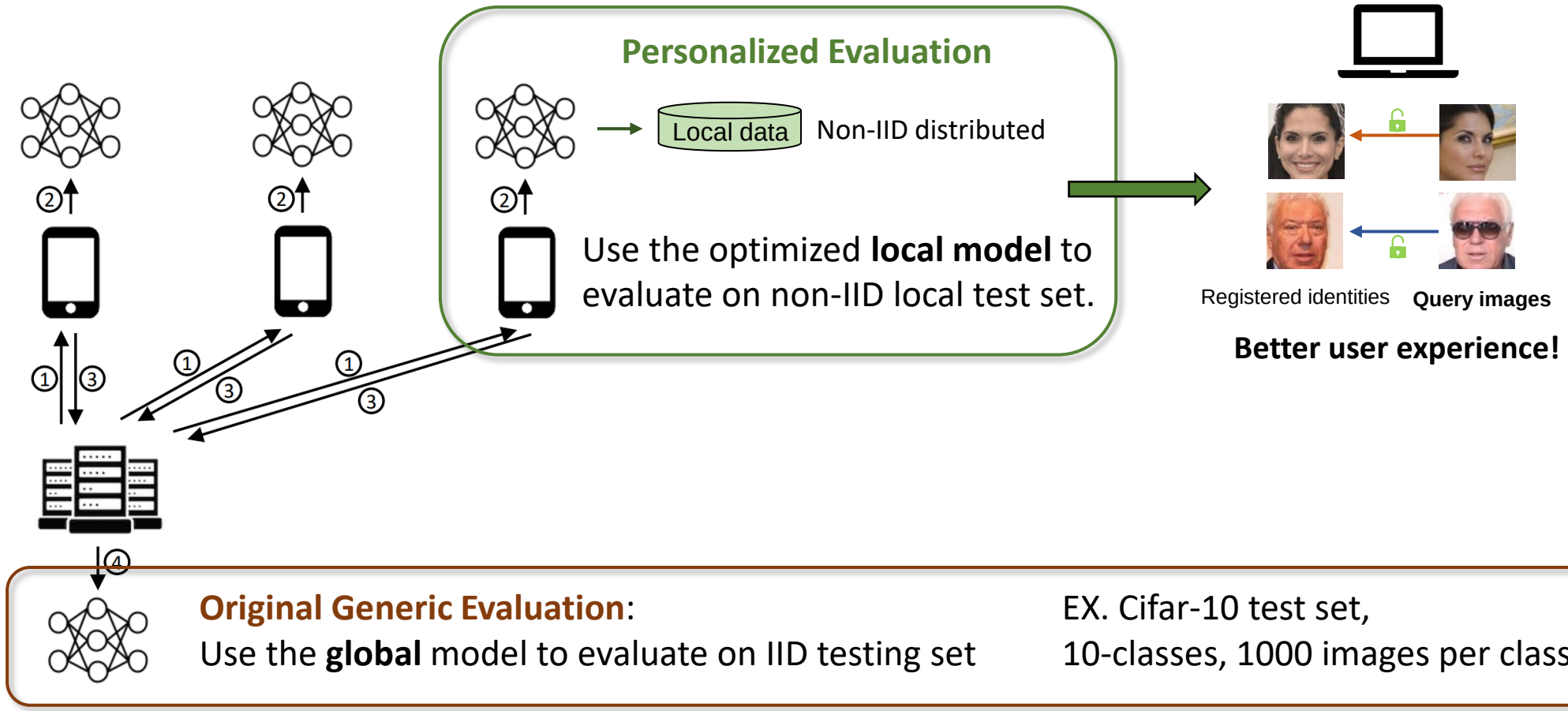
Only **one** previous work address the face recognition FL problem based on pre-trained model.

1. They tackle the most challenge scenario : Each client contains only **one** identity. → Limited scenarios
2. FedAvg only performs on **backbone** models, not on classifiers.
3. Clients additionally transmit the **class embedding** to server for SpreadOut regularization [10]. → Privacy leakage [11]



Personalized FL + Face Recognition

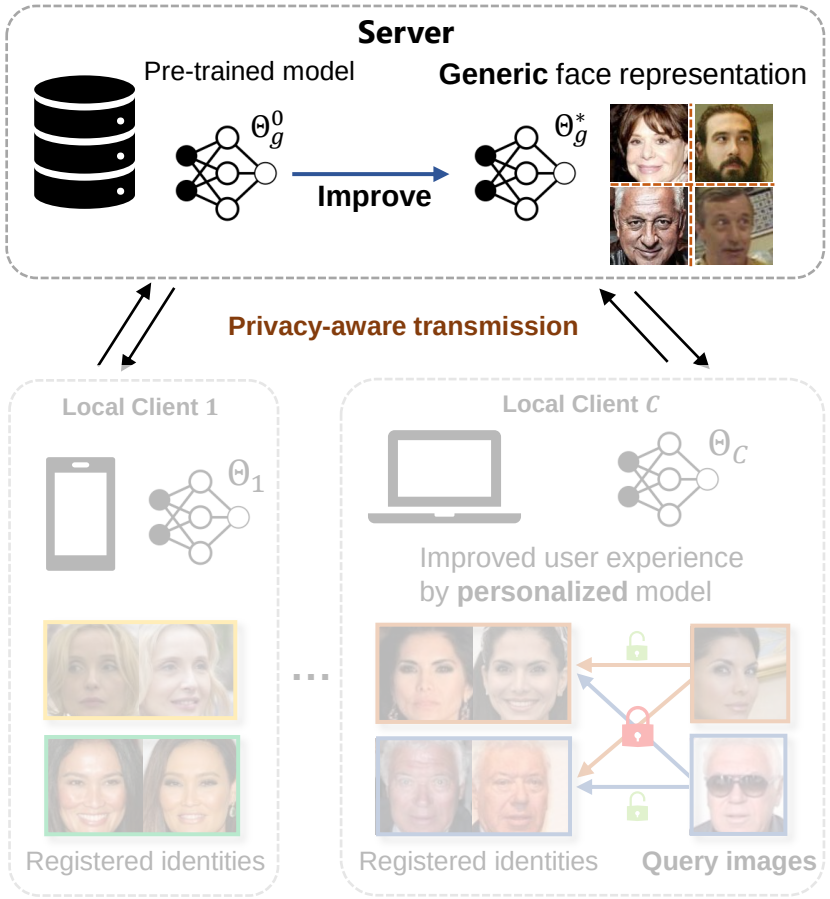
- We think that the **personalized face recognition** is also practical.



Our Problem Setup & Contributions

Our Problem Setup

- To enable a more **realistic** scenario for federated learning on face recognition, we propose a **new FL setup** that we need to jointly consider **generic** and **personalized** performance.

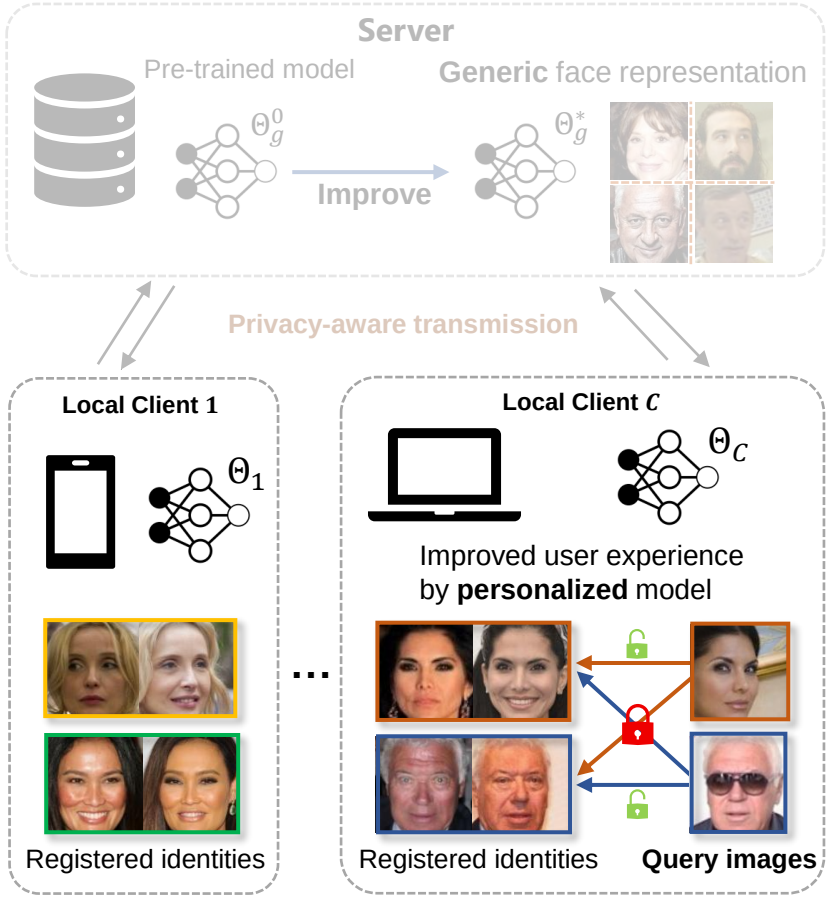


Goal 1:

How to continuously enhance the "**generic representation**" of pre-trained model under the FL environment?

Our Problem Setup

- To enable a more **realistic** scenario for federated learning on face recognition, we propose a **new FL setup** that we need to jointly consider **generic** and **personalized** performance.



To the best of our knowledge, we are the first to explore the personalized face recognition in FL setup!

Goal 2:

Given **query images** on local clients, whether we can obtain a "**personalized face model**" dedicated to recognize the registered identities. (better user experience)

Contributions

- We propose a joint optimization FL framework called **FedFR**, which can improve the generic and personalized face representation simultaneously:

Generic

1) We leverage public shared **pre-trained data** to regularize the training.
2) We propose **Hard Negative Sampling** strategy to improve the training efficiency.
3) We adopt **Contrastive Regularization** on local clients.

Personalized

4) We propose the novel **D**ecoupled **F**eature **C**ustomization (**DFC**) module, which is the key component to enable joint optimization of personalized face recognition model.
5) The proposed **binary classification** objectives are also effective for optimizing the personalized performance on each client.

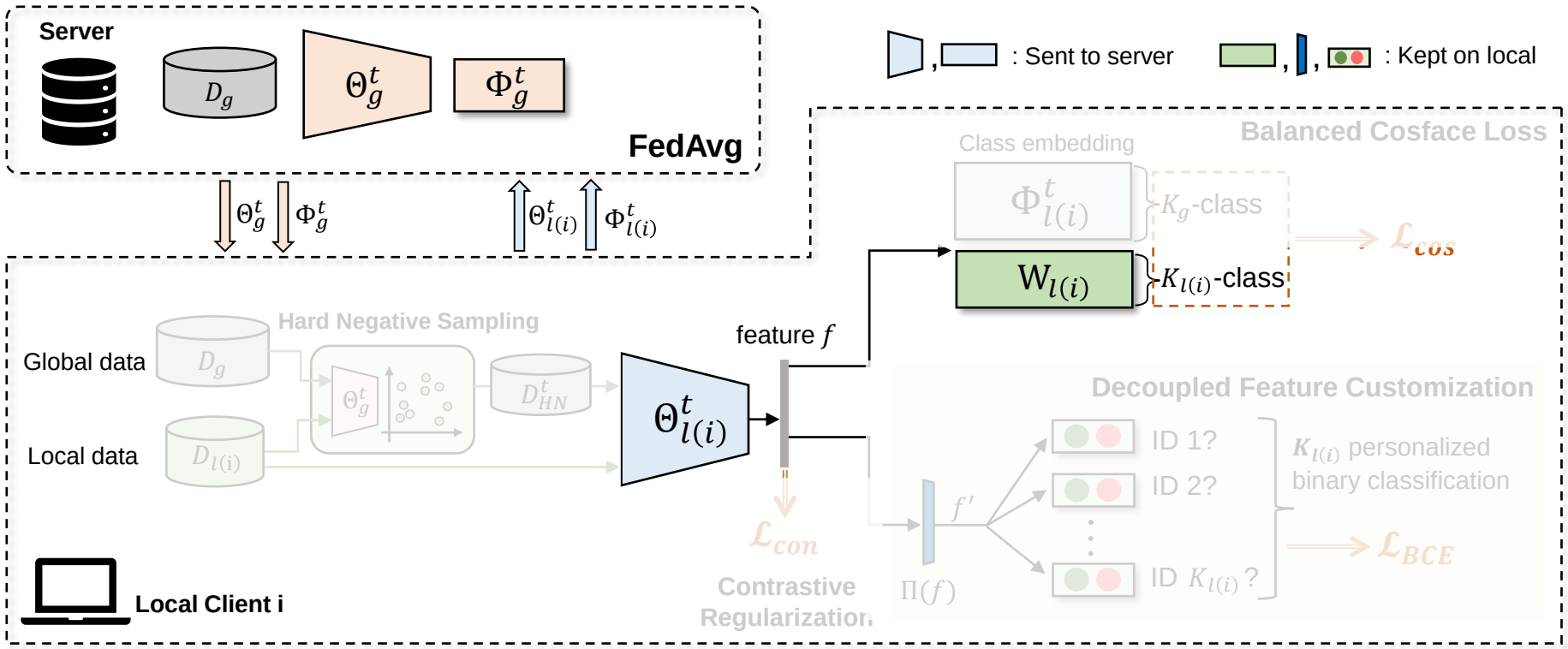
國立臺灣大學
National Taiwan University

Microsoft

Proposed Method

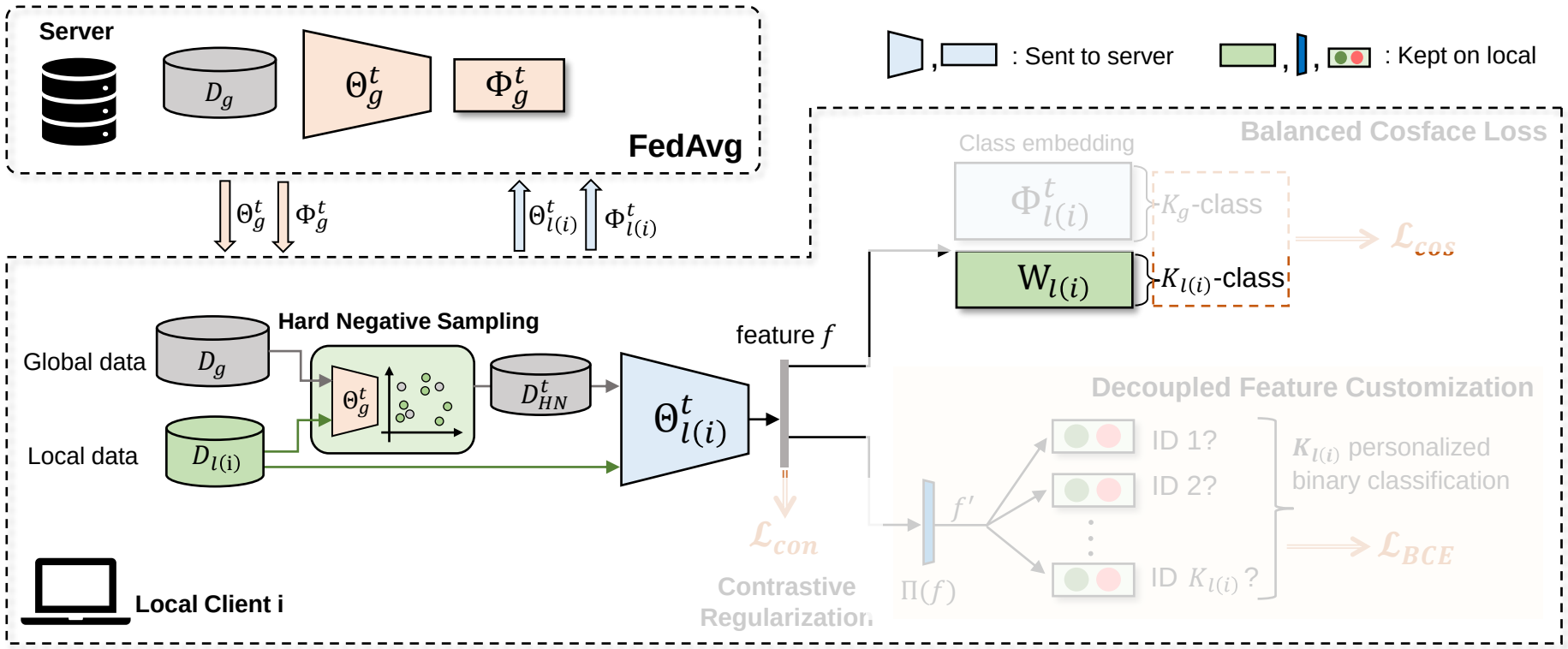
Our Overall Architecture

- Our architecture is based on FedAvg, and the pretrained model is trained with Cosface loss.
- In the t-th communication round,
  Θ_g^t : global backbone , Φ_g^t : global class embedding.
  For the i-th client, $\Theta_{l(i)}^t$: the local backbone , $W_{l(i)}$: local class embedding.



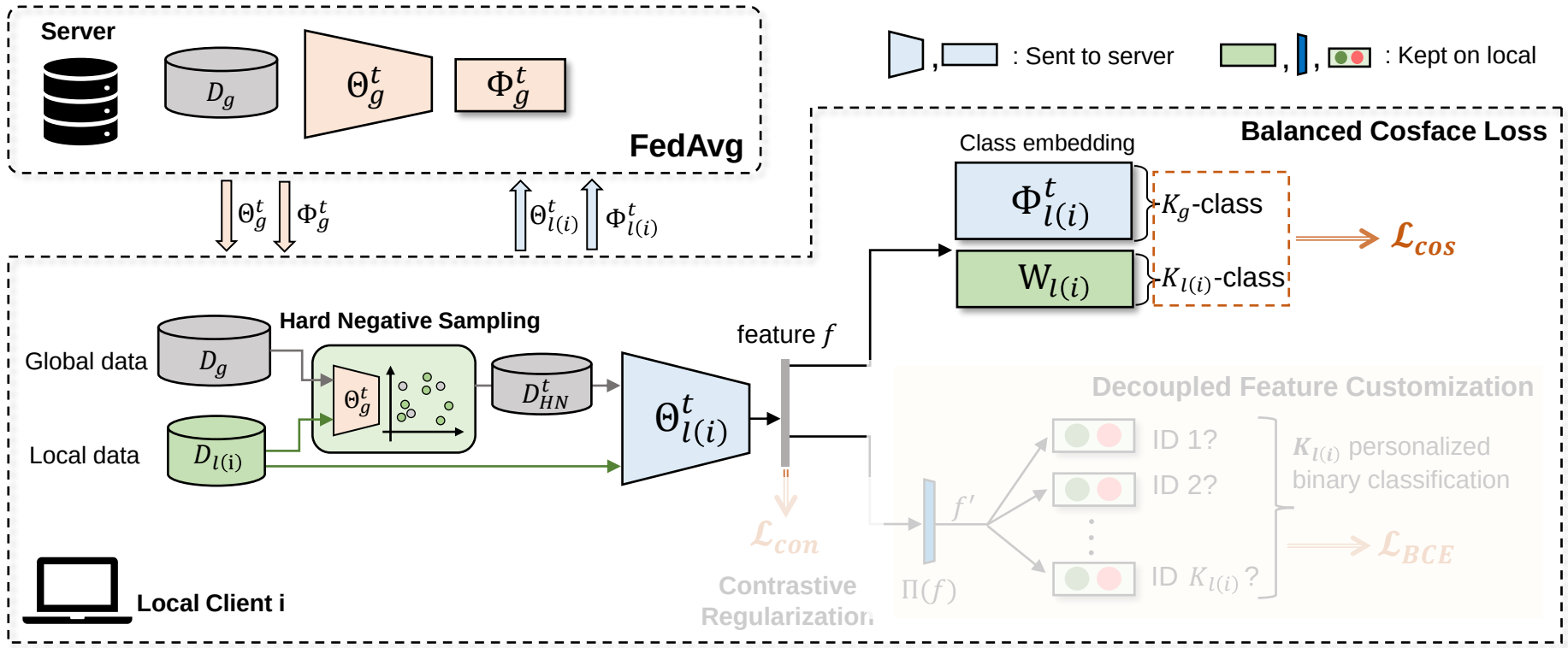
Our Overall Architecture

- Our architecture is based on FedAvg, and the pretrained model is trained with Cosface loss.
- In the t-th communication round,
  Θ_g^t : global backbone , Φ_g^t : global class embedding.
  For the i-th client, $\Theta_{l(i)}^t$: the local backbone , $W_{l(i)}$: local class embedding.



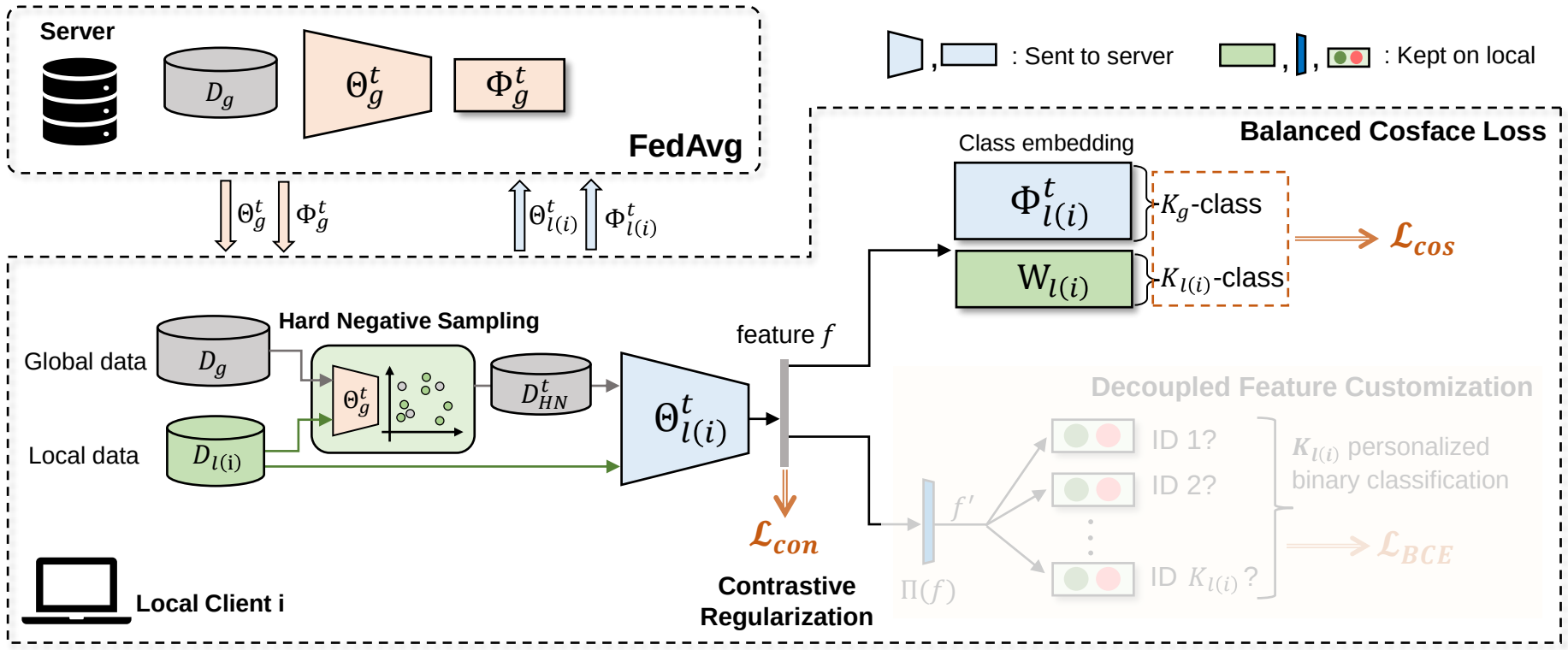
Our Overall Architecture

- Our architecture is based on FedAvg, and the pretrained model is trained with Cosface loss.
- In the t-th communication round,
  Θ_g^t : global backbone , Φ_g^t : global class embedding.
  For the i-th client, $\Theta_{l(i)}^t$: the local backbone , $W_{l(i)}$: local class embedding.



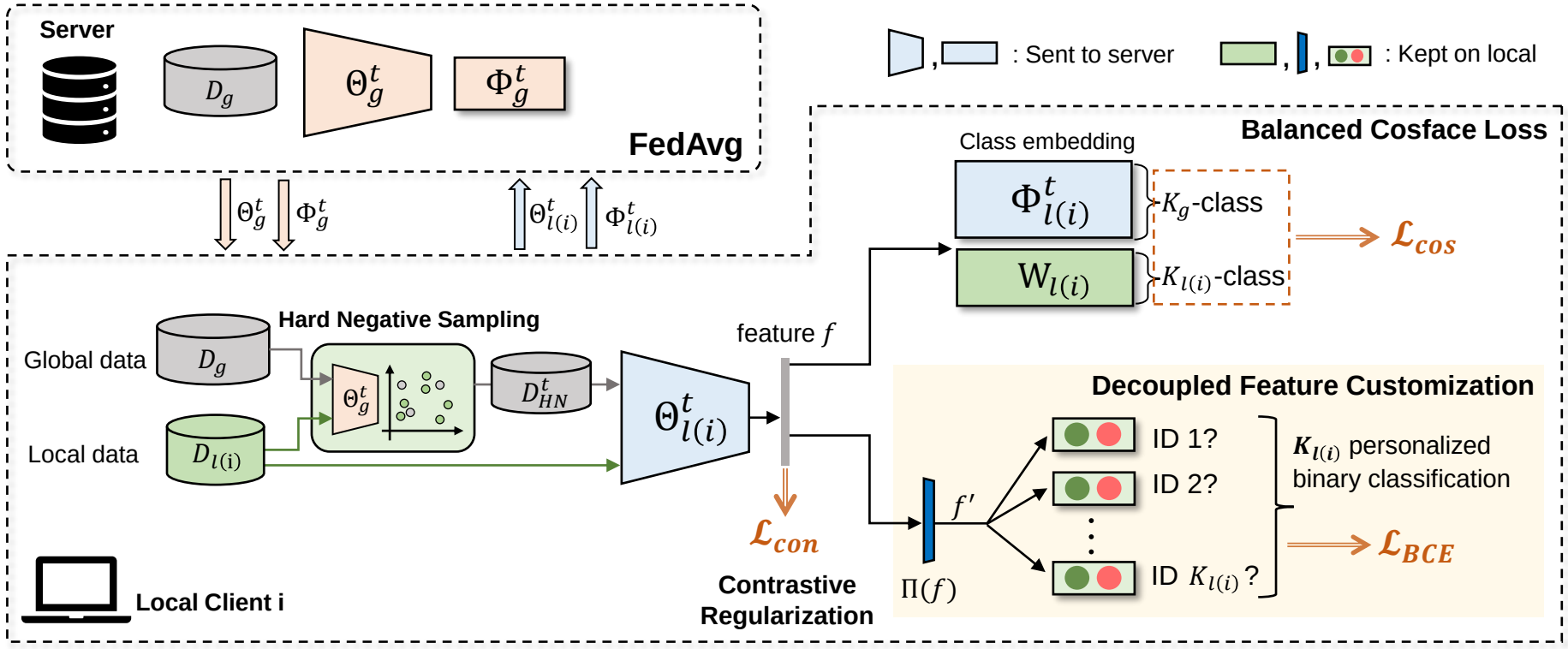
Our Overall Architecture

- Our architecture is based on FedAvg, and the pretrained model is trained with Cosface loss.
- In the t-th communication round,
  Θ_g^t : global backbone , Φ_g^t : global class embedding.
  For the i-th client, $\Theta_{l(i)}^t$: the local backbone , $W_{l(i)}$: local class embedding.



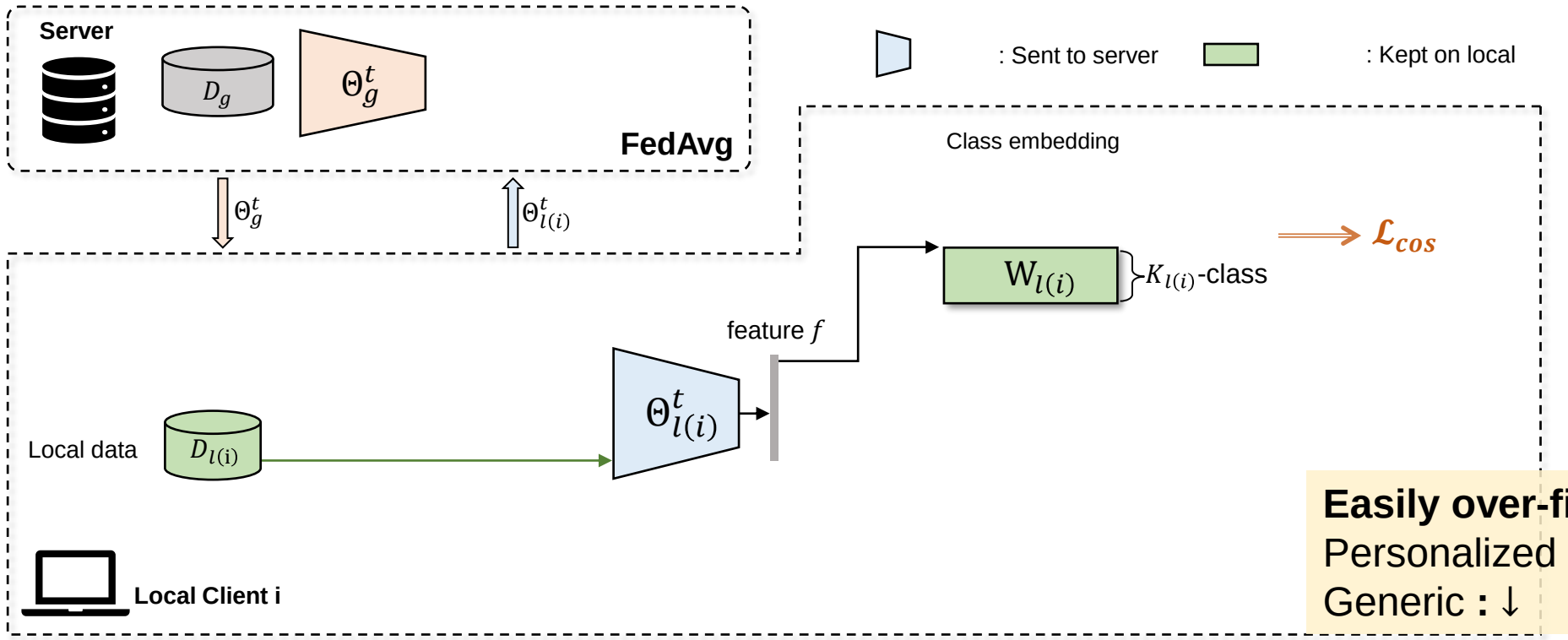
Our Overall Architecture

- Our architecture is based on FedAvg, and the pretrained model is trained with Cosface loss.
- In the t-th communication round,
  Θ_g^t : global backbone , Φ_g^t : global class embedding.
  For the i-th client, $\Theta_{l(i)}^t$: the local backbone , $W_{l(i)}$: local class embedding.



Baseline FedAvg pipeline

- Local client only optimize the $\Theta^t_{l(i)}$ and $W_{l(i)}$ on local data $D_{l(i)}$ with $N_{l(i)}$ images.
- Server conduct FedAvg only on **local backbones** : $\Theta_g^{t+1} = \frac{1}{N} \sum_{i \in [C]} N_{l(i)} \cdot \Theta^t_{l(i)}$

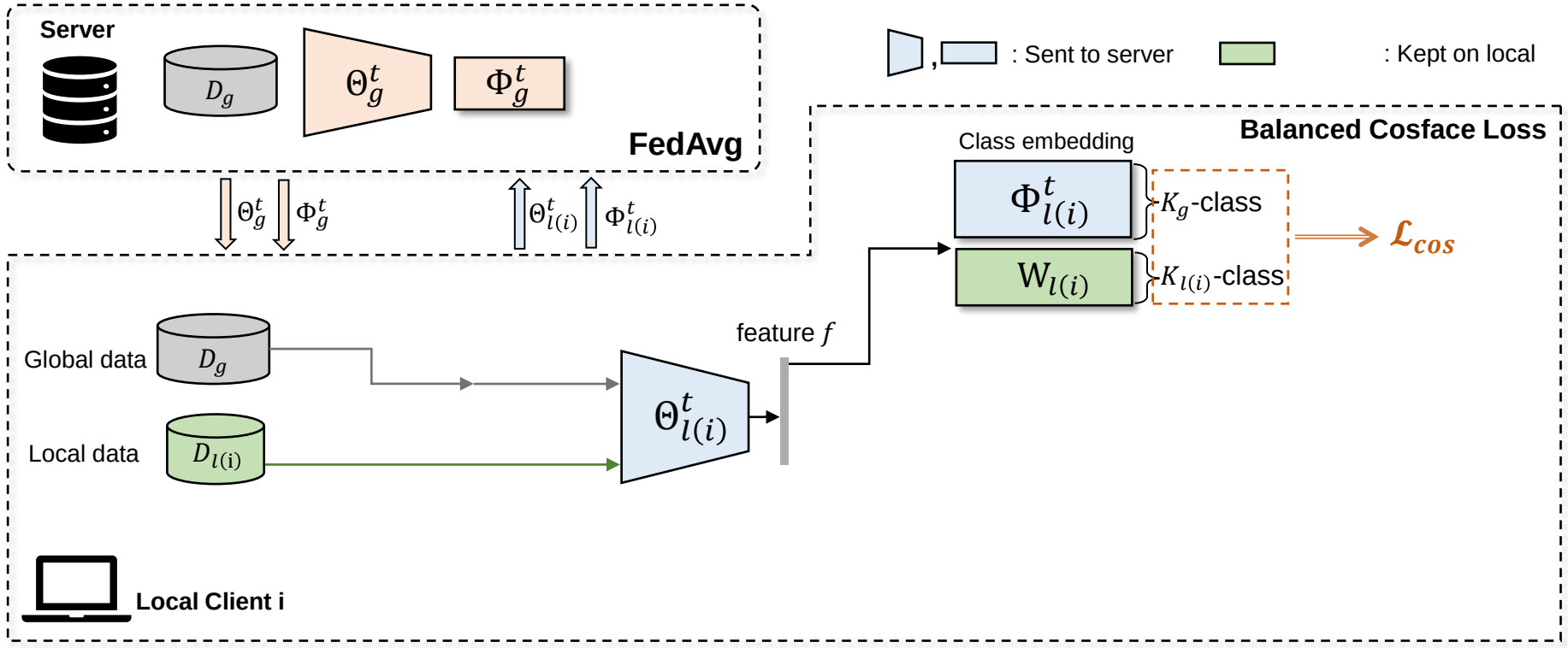


Easily over-fit on local data !
Personalized : ↑
Generic : ↓

Leverage Globally Shared Data

- Inspired by [4], besides only transmitting the global data D_g , we send the **class embedding** Φ_g^t to clients, which are all without privacy concern.

- The local objective can be more **balanced** with the new Cosface loss : $\mathcal{L}_{cos} = -\log \frac{e^{s(\cos\theta_y - m)}}{e^{s(\cos\theta_y - m)} + \sum_{j \neq y}^{K_g + K_{l(i)}} e^{s\cos\theta_j}}$,

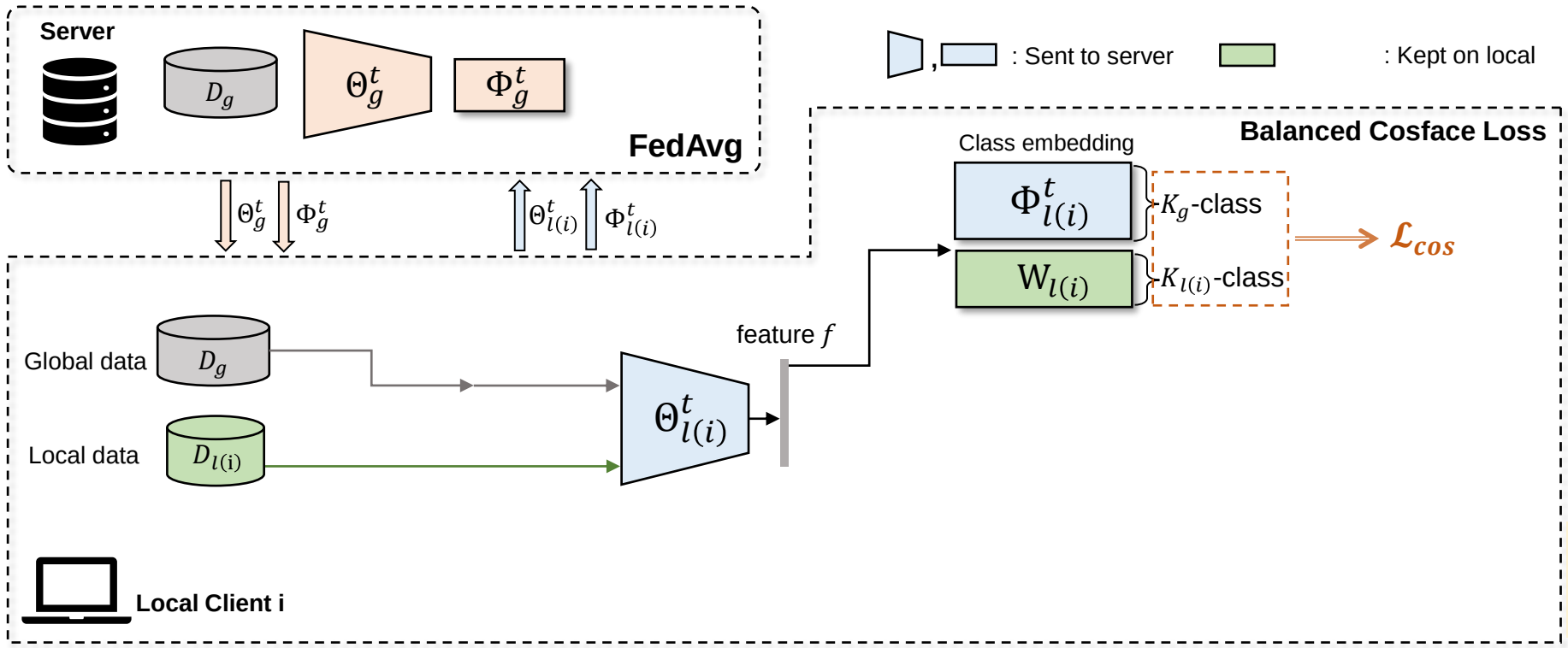


Leverage Globally Shared Data

- Server conduct **FedAvg** on both backbones and K_g global class embeddings with:

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

- However, optimizing on D_g+ $D_{l(\mathrm{i})}$ is very **time-consuming** and **not efficient.**

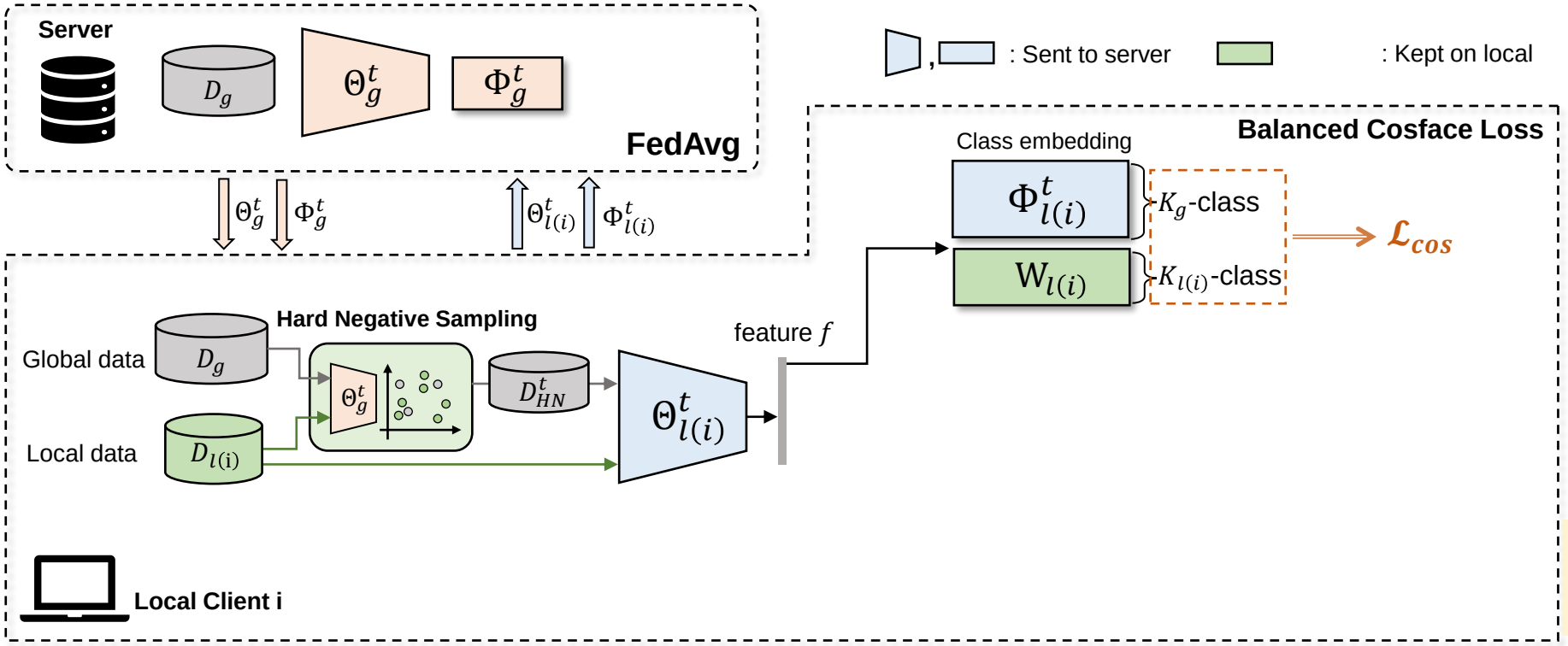


Personalized : ↑
Generic : ↑

Hard Negative Sampling Strategy

- To obtain a better trade-off, we propose a hard negative (**HN**) sampling strategy.
- We only sample a "**hard**" subset D_{HN} from D_g, which is with **feature similarity** larger than a threshold t_{HN} to any of the local data $D_{l(\mathrm{i})}$.

* The experiments of choosing suitable and reasonable t_{HN} are in our paper.

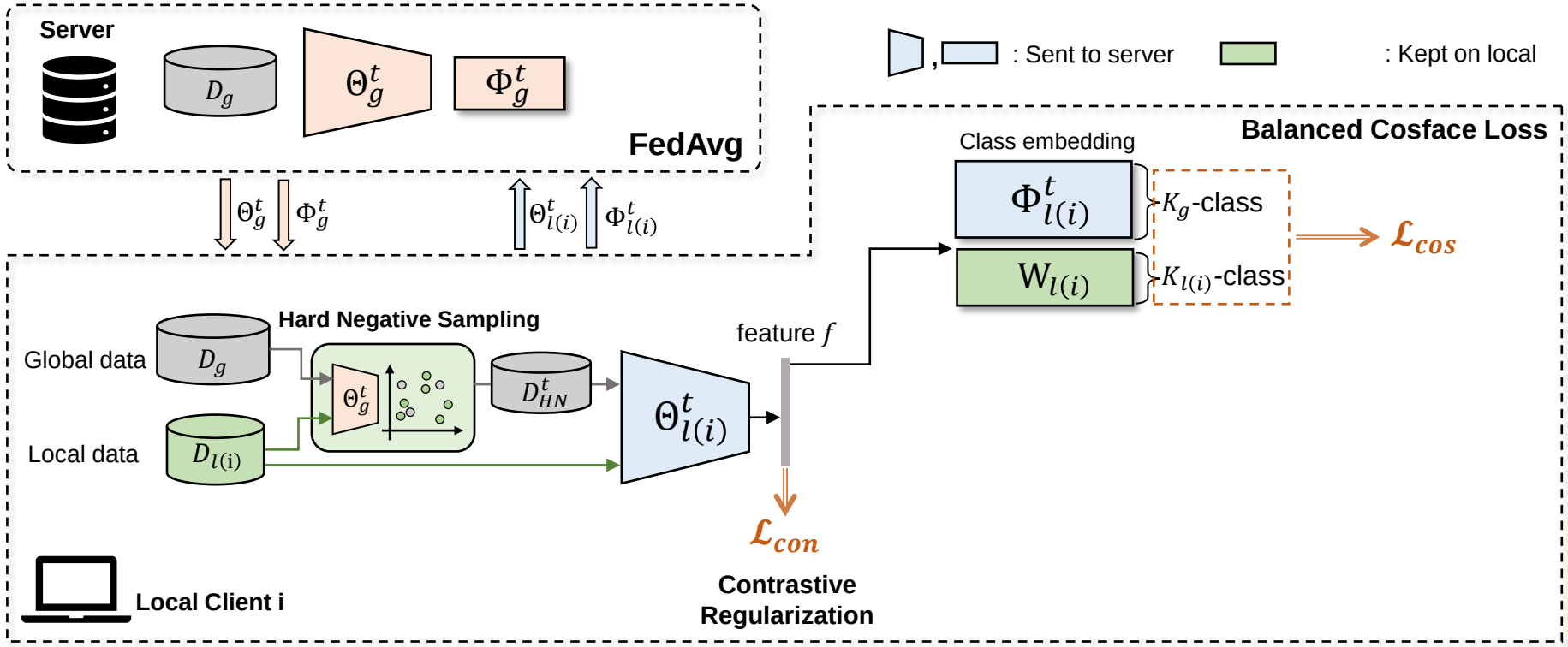


Personalized : ↑
Generic : ↑

Contrastive Regularization

- To reduce the gap between global and local model more, just as the previous work [3], we add the contrastive loss on the feature f.

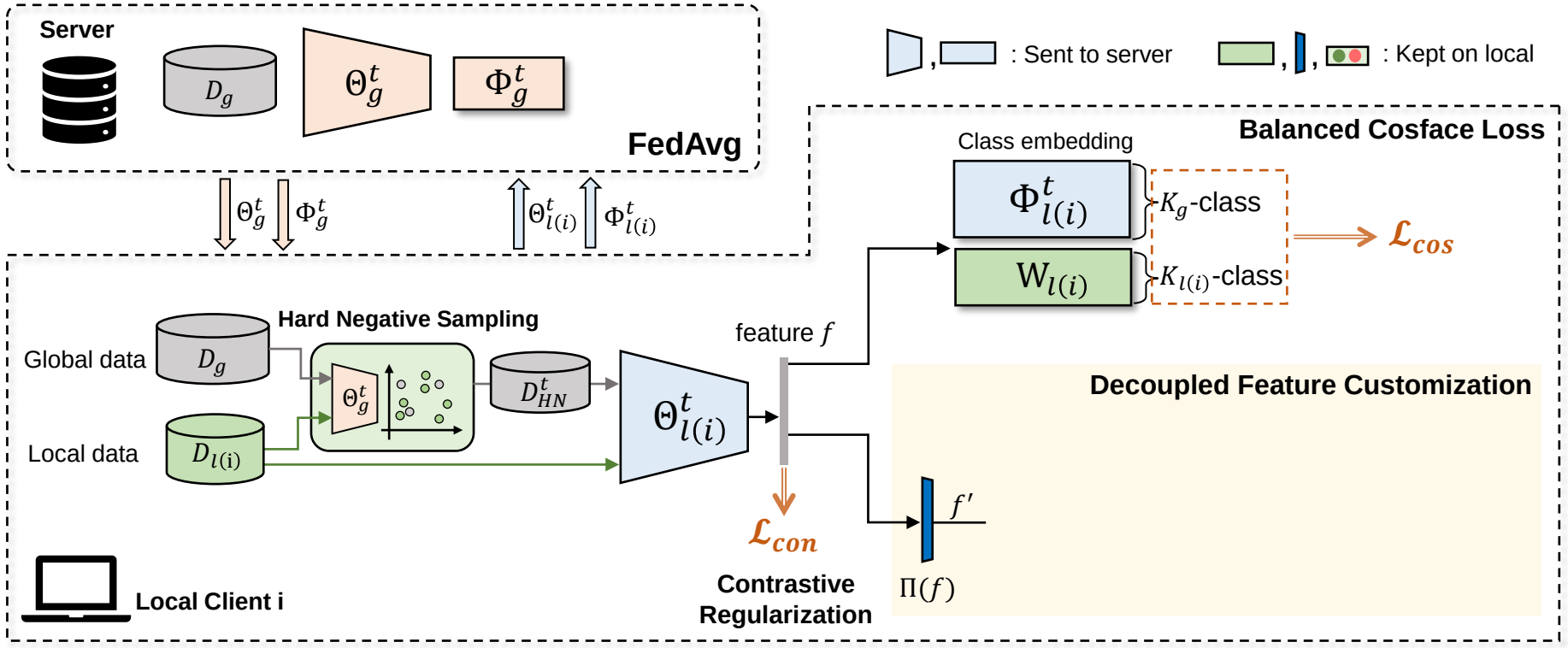
$$\mathcal{L}_{con} = -\log \frac{\exp(\mathrm{sim}(f, f_{glob})/\tau)}{\exp(\mathrm{sim}(f, f_{glob})/\tau) + \exp(\mathrm{sim}(f, f_{prev})/\tau)},$$



Personalized : ↓?
Generic : ↑

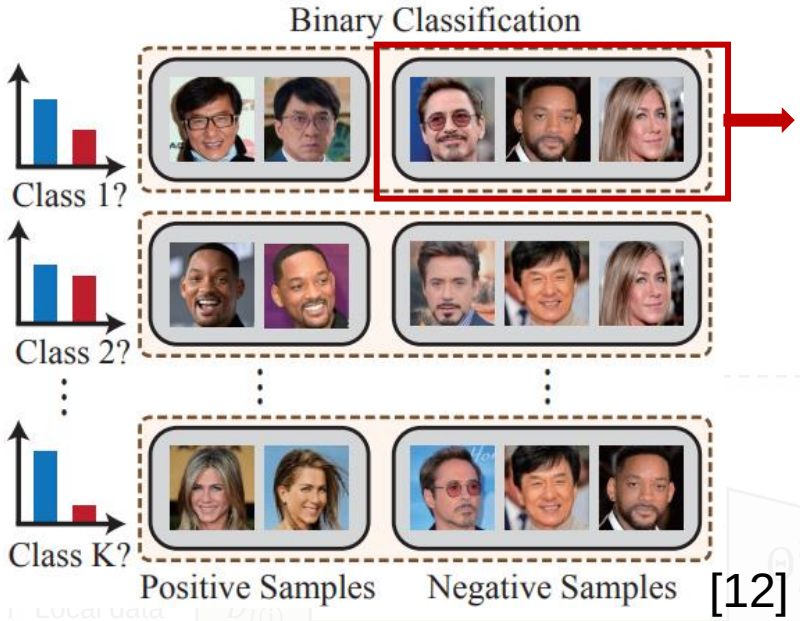
Decoupled Feature Customization

- To make the learning of global and local objective **separately,** we propose **DFC** module.
- Contains a transformation $\boldsymbol{\Pi}(\boldsymbol{f})$ that maps original feature to a new space specific for client i
  $\boldsymbol{f}$: generic feature representation ; $\boldsymbol{f'}$: personalized feature



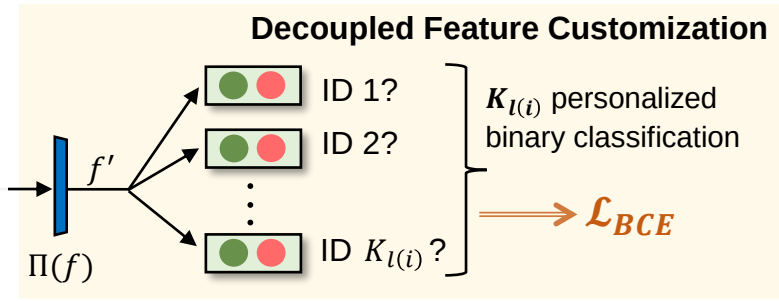
Decoupled Feature Customization

- Learning objective ? A : $K_{l(i)}$ **binary classification tasks.**
- We only need to focus on whether the input image belongs to ID 1? or ID 2? ... ID $K_{l(i)}$?
- Inspired by [12], the loss formulation is a summation of $K_{l(i)}$ margin-based Binary Cross-Entropy loss $\mathcal{L}_{BCE}$



[12]

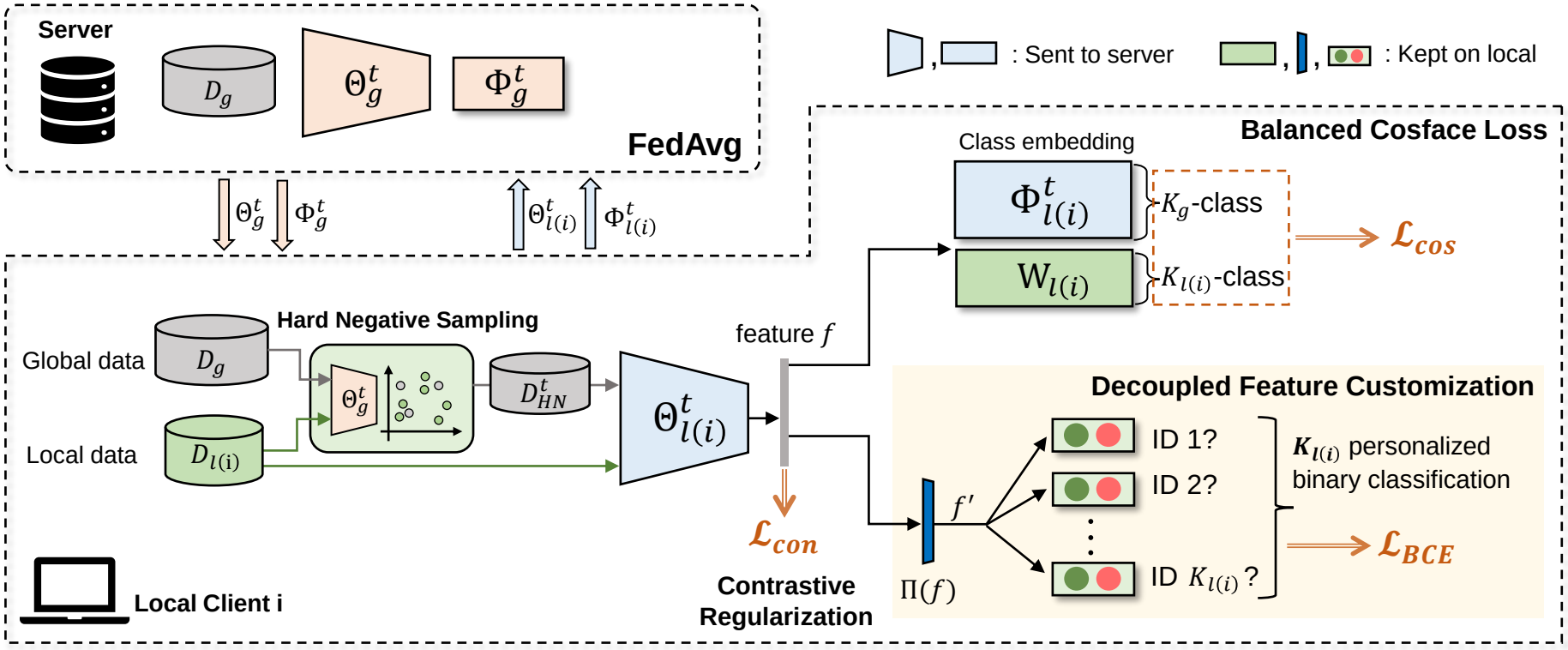
All images (including **global data**) not with the Class 1 will be classified to negative.



Personalized :↑↑
Generic : ↑

Our Overall Architecture

- Optimized end-to-end with the total loss : $\mathcal{L}_{total} = \alpha_1 \mathcal{L}_{cos} + \alpha_2 \mathcal{L}_{con} + \alpha_3 \mathcal{L}_{BCE},$
- In the testing phase, $\boldsymbol{\Theta}_{\boldsymbol{g}}^{\boldsymbol{t}}$ is used for generic evaluation,
  $[\boldsymbol{\Theta}_{\boldsymbol{l(i)}}, \boldsymbol{\Pi}_{\boldsymbol{l(i)}}]$ is used for personalized evaluation.



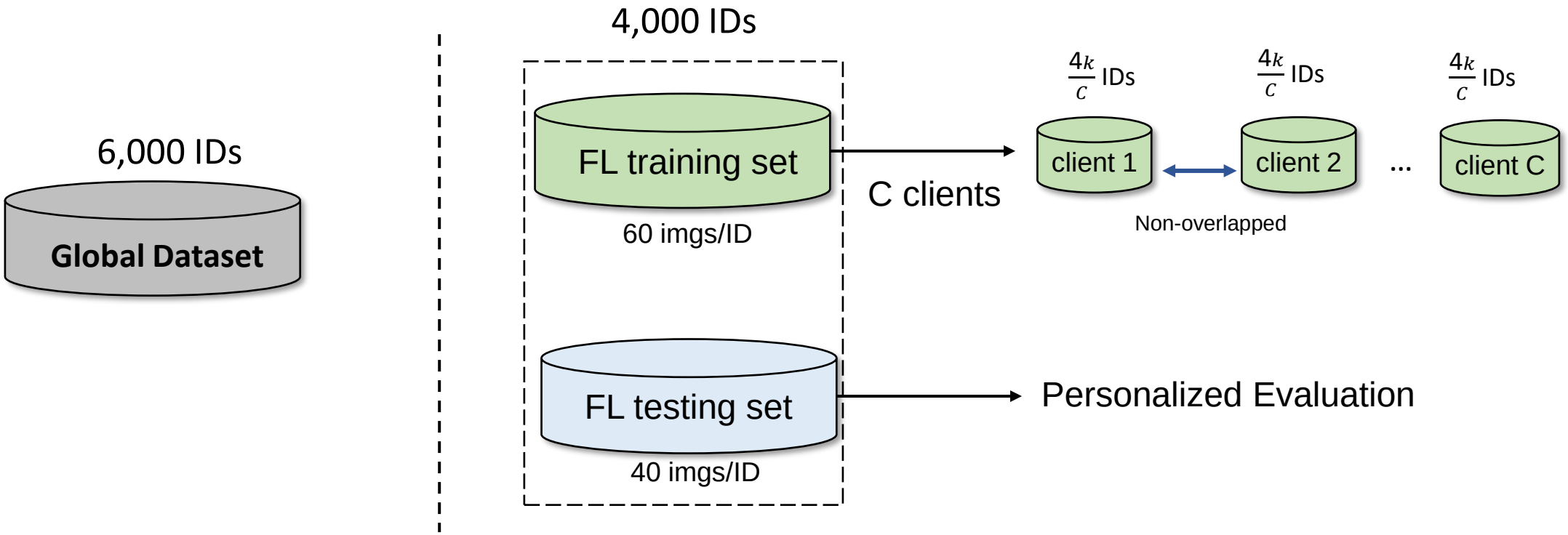
國立臺灣大學
National Taiwan University

Microsoft

Experiment Results

Dataset

- We split a subset from commonly used **MS**-Celeb-**1M** dataset [13].
- We sample total 10,000 IDs with ~100 images per class.
- **6,000** IDs for pre-trained (global dataset), **4,000** IDs for federated learning (6 : 4 = train : test).



Evaluation Metrics

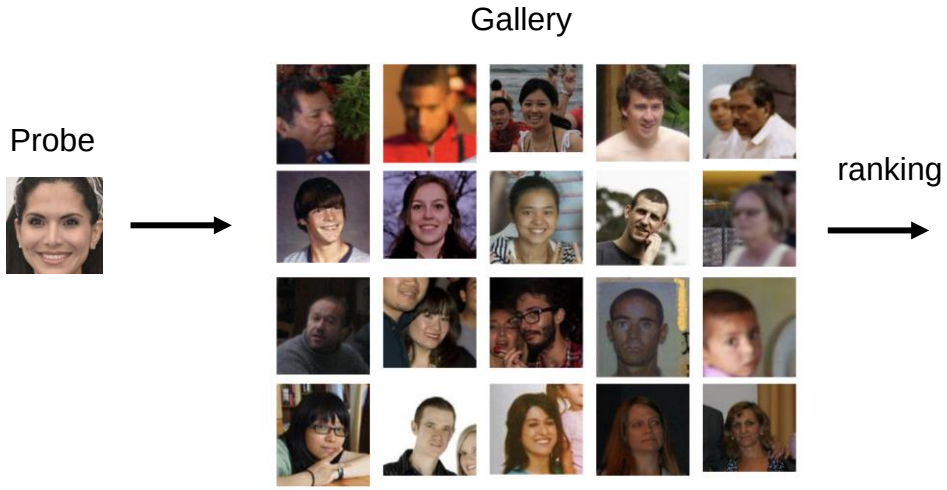
- **Generic Evaluation**

  We evaluated on IJB-C dataset [14] and follow their protocol.

  - **1:1 verification TAR @ FAR** :
    True acceptance rates (TAR) at different false acceptance rates (FAR) for 1:1 **verification** protocol.
  - **1:N identification TPIR @ FPIR :**
    True positive identification rates (TPIR) at different false positive identification rates (FPIR) for 1:N **identification** protocol. (Probe → Gallery ranking)

positive

negative



Evaluation Metrics

- **Personalized Evaluation**
  - We carefully build up the metric by ourselves. (we are the first to investigate this setup)
  - The evaluation is supposed to only focus on the user experience of the **registered identities** on each local client.
  - We also establish **1:1 verification protocol** and **1:N identification protocol**.

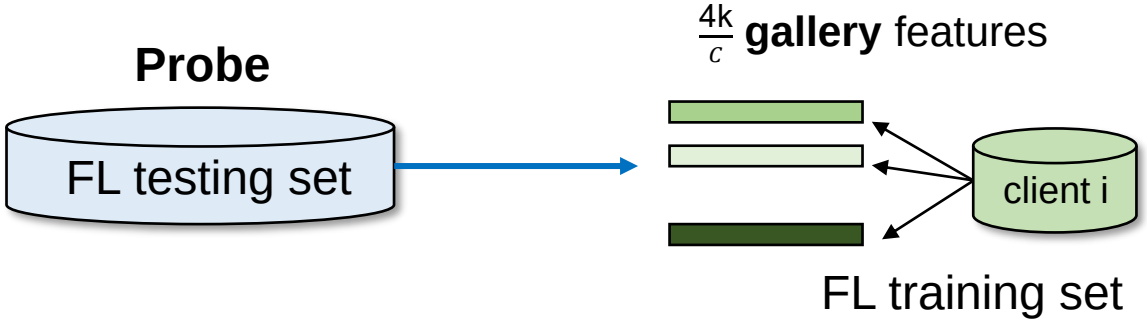
Ex. For the i-th client contains registered $\frac{4k}{c}$ IDs :

1:1 verification protocol

Positive : $\frac{4k}{c}$ IDs

Negative : $\frac{4k}{c}$ IDs $\leftrightarrow$ $(4000 - \frac{4k}{c})$ IDs

1:N identification protocol



Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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
	✓	✓	✓	**77.60**	**85.21**	**73.60**	**81.27**	**88.32**	**95.46**	**95.17**	**97.94**
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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				
	✓	✗	✗	76.79	84.64	72.76	80.76				
	✓	✓	✗	77.41	85.17	73.60	81.25	77.77	89.57	89.58	94.60
	✓	✓	✓	**77.60**	**85.21**	**73.60**	**81.27**	**88.32**	**95.46**	**95.17**	**97.94**
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

Poor personalized performance

Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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
	✓	✓	✓	**77.60**	**85.21**	**73.60**	**81.27**	**88.32**	**95.46**	**95.17**	**97.94**
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

Baseline **FedAvg**

Degrade ↓

Improve↑

Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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	8		39.57	89.58	94.60
	✓	✓	✓	**77.60**	**85.21**	**73.60**	8		95.46	**95.17**	**97.94**
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

All improve ↑

Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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
	✓	✓	✓	77.60	8		81.27	88.32	9		97.94
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.38	99.40

Improve↑

Degrade ↓

Ablation Studies

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	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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
	✓	✓	✓	**77.60**	**85.21**	**73.60**	**81.27**	**88.32**	**95.46**	**95.17**	**97.94**
Centrally trained on 10k				77.56	85.99	73.30		7.39		98.58	99.40

Our **FedFR**

The best result

Ablation Studies

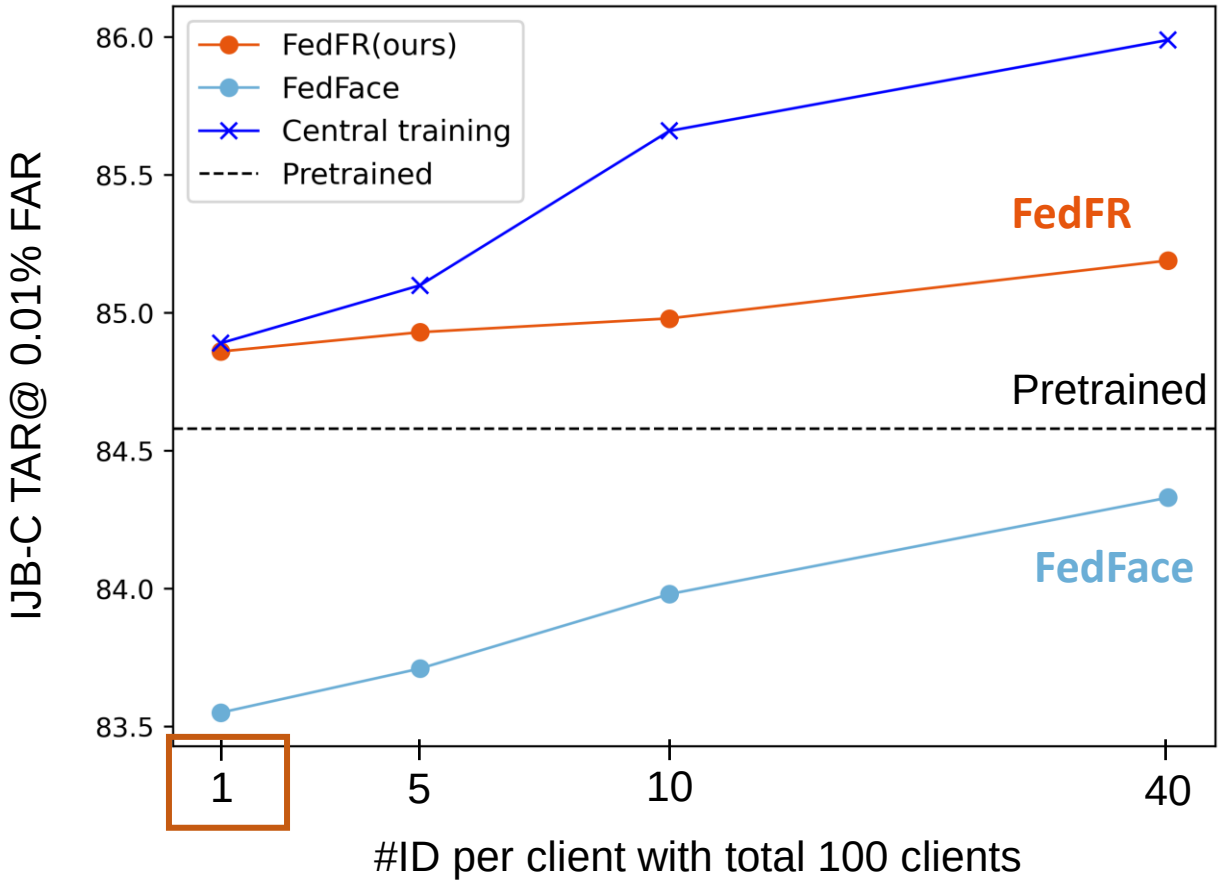
Table 1: **Ablation Studies.** We conduct FL experiments with 40 clients; each client contains 100 identities. (results are in %)

Setup	Modules			Generic Evaluation (IJB-C)				Personalized Evaluation			
	HN. sampled Global data	Contrastive	DFC. Branch	1:1 TAR @ FAR 1e-5	1:1 TAR @ FAR 1e-4	1:N TPIR @ FPIR 1e-2	1:N TPIR @ FPIR 1e-1	1:1 TAR @ FAR 1e-6	1:1 TAR @ FAR 1e-5	1:N TPIR @ FPIR 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
	✓	✓	✓	**77.60**	**85.21**	**73.60**	**81.27**	**88.32**	**95.46**	**95.17**	**97.94**
Centrally trained on 10k IDs				77.56	85.99	73.30	82.14	93.72	97.39	98.58	99.40

Upper bound

Comparable performance!

Comparison with FedFace [8]



Comparison with other Personalized FL methods

- Yu et al. [7] proposed **two-stage** personalized method.
  1. Train with FedAvg or other typical FL method.
  2. Tune the local model independently

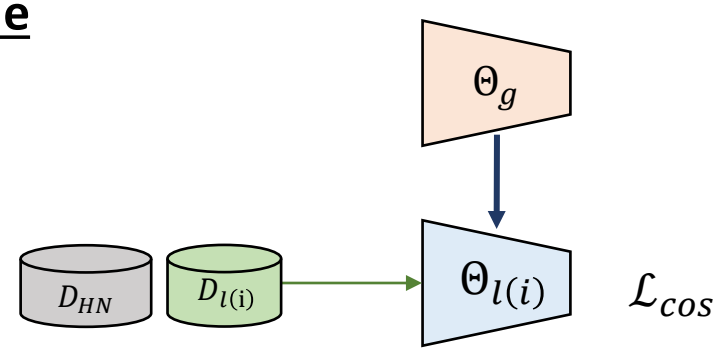
- We re-implement two techniques in [7] on face recognition based on our FedFR w/o a DFC branch.

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

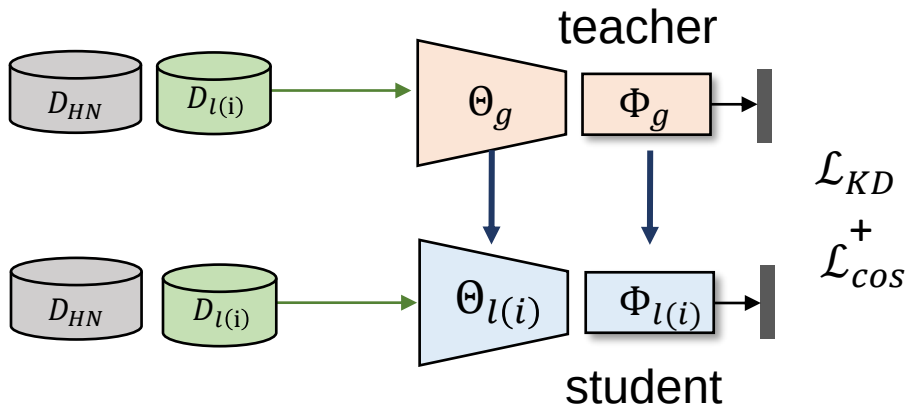
After last round of our **FedFR w/o DFC**:

Fine-tune



Fine-tune several epoch

Knowledge Distillation (KD)

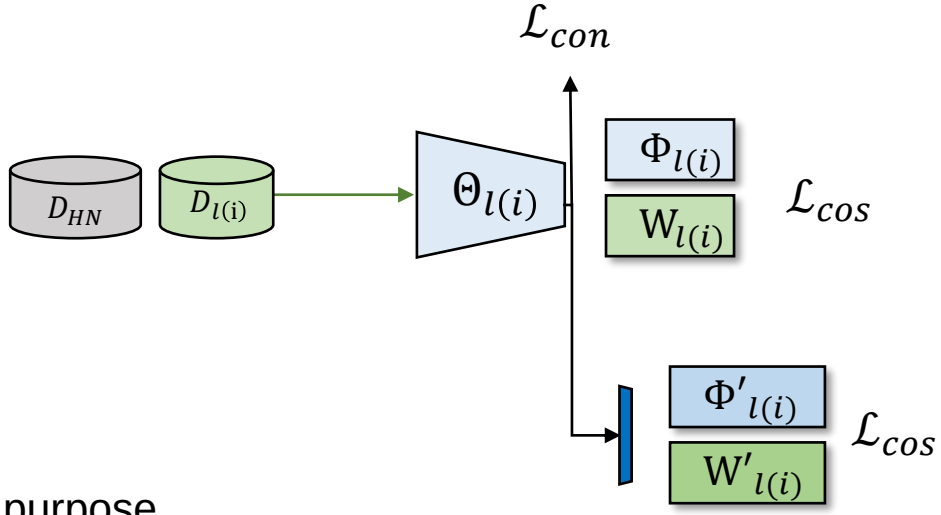


Comparison with other Personalized FL methods

- We also conduct a method **with customization branch** but with normal Cosface loss.

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



Validate that our **BCE loss** is suitable for personalized purpose

Conclusion

Conclusion

- We address the face recognition model training under the practical federated learning setting, where each client is initialized with the **pre-trained** model.

- We are the first to explore the **personalized** face recognition in FL setup.

- We propose **FedFR**, which contains
  1. "**Hard negative sampling**" and "**contrastive regularization**". They can efficiently bridge the gap between global and local training.
  2. **Decoupled Feature Customization** (DFC) module can enable concurrent optimization of the personalized face recognition model.

Reference

[1] Wang, Hao, et al. "Cosface: Large margin cosine loss for deep face recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[2] Li, Xiang, et al. "On the convergence of fedavg on non-iid data." *arXiv preprint arXiv:1907.02189* (2019).

[3] Li, Qinbin, Bingsheng He, and Dawn Song. "Model-Contrastive Federated Learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[4] Zhao, Yue, et al. "Federated learning with non-iid data." *arXiv preprint arXiv:1806.00582* (2018).

[5] Arivazhagan, Manoj Ghuhan, et al. "Federated learning with personalization layers." *arXiv preprint arXiv:1912.00818* (2019).

[6] Tan, Alysa Ziying, et al. "Towards personalized federated learning." *arXiv preprint arXiv:2103.00710* (2021).

[7] Yu, Tao, Eugene Bagdasaryan, and Vitaly Shmatikov. "Salvaging federated learning by local adaptation." *arXiv preprint arXiv:2002.04758* (2020).

[8] Aggarwal, Divyansh, Jiayu Zhou, and Anil K. Jain. "FedFace: Collaborative Learning of Face Recognition Model." *arXiv preprint arXiv:2104.03008* (2021).

[9] Li, Qinbin, et al. "A survey on federated learning systems: vision, hype and reality for data privacy and protection." arXiv preprint arXiv:1907.09693 (2019).

[10] Yu, Felix, et al. "Federated learning with only positive labels." International Conference on Machine Learning. PMLR, 2020.

[11] Duong, Chi Nhan, et al. "Vec2Face: Unveil Human Faces From Their Blackbox Features in Face Recognition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[12] Wen, Yandong, et al. "SphereFace2: Binary Classification is All You Need for Deep Face Recognition." arXiv preprint arXiv:2108.01513 (2021).

Reference

- [13] Guo, Yandong, et al. "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition." *European conference on computer vision*. Springer, Cham, 2016.
- [14] Maze, Brianna, et al. "Iarpa janus benchmark-c: Face dataset and protocol." *2018 International Conference on Biometrics (ICB)*. IEEE, 2018.

國立臺灣大學
National Taiwan University

Microsoft

Thank you for your listening