FedIris: Towards More Accurate and Privacy-preserving Iris Recognition via Federated Template Communication

Zhengquan Luo[1,2], Yunlong Wang[2,*], Zilei Wang[1], Zhenan Sun[2], Tieniu Tan[2]
[1] University of Science and Technology of China (USTC)
[2] Institute of Automation, Chinese Academy of Sciences (CASIA)
{zhengquan.luo, yunlong.wang}@cripac.ia.ac.cn, zlwang@ustc.edu.cn, {znsun, tnt}@nlpr.ia.ac.cn

Abstract

As biometric data undergo rapidly growing privacy concerns, building large-scale datasets has become more difficult. Unfortunately, current iris databases are mostly in small scale, e.g.*, thousands of iris images from hundreds of identities. What's worse, the heterogeneity among decentralized iris datasets hinders the current deep learning (DL) frameworks from obtaining recognition performance with robust generalization. It motivates us to leverage the merits of federated learning (FL) to solve these problems. However, traditional FL algorithms often employ model sharing for knowledge transfer, wherein the simple averaging aggregation lacks interpretability, and divergent optimization directions of clients lead to performance degradation. To overcome this interference, we propose FedIris with solid theoretical foundations, which attempts to employ the iris template as the communication carrier and formulate federated triplet (Fed-Triplet) for knowledge transfer. Furthermore, the massive heterogeneity among iris datasets may induce negative transfer and unstable optimization. The modified Wasserstein distance is embedded into the FedTriplet loss to reweight global aggregation, which drives the clients with similar data distributions to contribute more mutually. Extensive experimental results demonstrate that the proposed FedIris outperforms SOLO training[1], model-sharing-based FL training, and even centralized training[2].*

1. Introduction

Iris is acknowledged to be a dominant biometric trait in next-generation identification systems. Iris recognition systems are already deployed in high-security areas such as access control, check-in inspection, surveillance systems. Unlike large-scale face datasets with millions of images or even identities (IDs) [1], publicly available iris datasets are in much smaller scale. For instance, the latest CASIA series iris dataset [2] only consists of 36,539 images from 255 subjects. In addition, privacy protection regulations over the world are constantly tightening, and privacy attacks are evolving. Some big datasets, *e.g.*, MS-Celeb-1M [3], are no longer available for public use. Hence, the traditional solution becomes nearly impossible, which collects and piles up large-scale data to train a centralized iris recognition framework. Meanwhile, the popular deep neural network (DNN) based iris recognition models are data-hungry. Abundant data are necessities for discriminative representation learning. The massive heterogeneity among current small-sized iris datasets as "data islands" further impedes centralized training. To this end, a potential solution is Federated Learning (FL) [4], which provides a decentralized learning paradigm without direct access to local raw data, but can transfer the knowledge through model sharing. For instance, FedSGD [4] updates the global parameter with the average aggregation of multiple clients' local gradients. De-

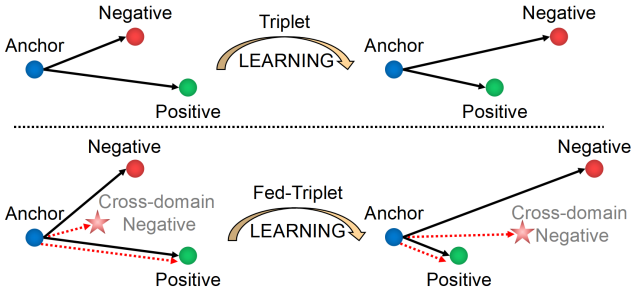


Figure 1. The goal of FedTriplet loss is that the local representations are learned to be more separate from inter-class samples and closer to intra-class samples, with the introduction of cross-client negative samples in latent space. Anchor and positive samples are from the same ID, negative samples are from other IDs but the same client, and the cross-client negative samples are from out-of-distribution IDs and other clients.

*Yunlong Wang is corresponding author
[1]Model training on a single dataset.
[2]Model training on all dataset superimposed together.

spite its great success, the recent attack methods [5] proved that raw images can be faithfully reconstructed at high resolution using the received gradients from other clients. The privacy guarantee of model-sharing-based FL is moderately broken.

Iris templates are encoded as highly abstract representations of raw iris images. They are insensitive to personal privacy and non-reusable when the feature extraction model changes. Even if the templates are leaked, it is non-trivial to reconstruct raw iris data or recognize the identity. Thus, iris templates can be employed as the knowledge carrier for FL communication. As far as we know, this work is the first to apply FL-inspired strategy into iris recognition by adopting federated template communication.

The purpose of joint training in distributed FL systems is to communicate the knowledge through client aggregation [4]. The federated template communication can inherently introduce additional constraints for the error bound of the local clients [6]. In addition, the robustness and generalization of the local model can be enhanced with the introduction of diverse templates from other clients. We provide solid theoretical foundations to support these hypotheses. More specifically, aggregated averaging empirical risk minimization (ERM) is the primary optimization strategy of model-sharing-based FL algorithms. Although the convergence guarantee is proved [7], this simple averaging aggregation still lacks interpretability in FL settings, and the performance suffers from heterogeneous distributions due to divergent optimization directions of clients. Inspired by metric learning [8], Fed-Triplet loss is proposed, as updating from the version of the local triplet loss function, for cross-client template communication. The foundation of this training strategy is that the representations share a similar training assumption in most computer vision (CV) tasks, *i.e*., these vectors are learned to be intra-class proximity and inter-class separation in latent space. *Fig.* 1 shows the goal of Fed-Triplet loss with the introduction of cross-client negative template samples. The effectiveness lies in that: 1) the discriminability of the local template is enhanced due to inter-class distance expanding; 2) the robustness and generalization ability are significantly improved due to tighter intra-class distribution.

The widely adopted iris datasets exhibit the characteristics of small scale, diverse collecting environments, and subject variations. These huge distribution shifts also refer to data heterogeneity. Data heterogeneity in decentralized datasets destroys the basic independent and identically distributed (i.i.d.) assumption. Under distributed FL setting, this heterogeneity between clients induces sub-optimal or detrimental performance when testing on each client's local data, *a.k.a.* negative transfer. The underlying reason is that the divergent optimization directions of client models [9], which leads task-decision models to be unstable and even non-convergent. To mitigate this problem, many reweighting strategies are proposed to measure the similarity of data distribution implicitly. For example, [10] calculated the influence level from other clients' models. The FedAMP [11] employed the model similarity to facilitate similar clients to collaborate more. In this paper, as an explicit measurement, the Wasserstein distance [12] is embedded into the Fed-Triplet loss to reweight the cross-client template samples in aggregation, which drives the clients with more similar distributions to be more contributive.

To verify the proposed framework, extensive experiments are conducted on multiple commonly adopted iris datasets. It demonstrates that FedIris outperforms SOLO training, model-sharing-based FL training (classic FedSGD [4][3]), and even centralized training. In addition, the effectiveness of the proposed FedIris to mitigate the negative transfer is carefully verified. The performance that all the clients participate in the communication rounds is much better than FedSDG and only second to the model trained on the optimal group of communication participators[4]. To sum up, the main contributions of the paper are as follows:

- FedIris is proposed as the first attempt of FL application to distributed iris recognition.
- Federated template communication transfers knowledge among clients with better interpretability and solid theoretical foundations.
- Experimental results on decentralized iris datasets demonstrate the superiority of the proposed FedIris.

2. Related Work

In this section, a review of the literature is provided mainly on federated learning and its application to biometrics.

2.1. Federated Learning

Federated learning (FL) provides a decentralized learning paradigm and fulfills privacy protection. Many techniques from other research areas are applied to FL and achieve great success, such as meta-learning [13, 14], domain adaptation [15, 16] , knowledge distillation [17, 18], multi-task learning [19, 20], etc. However, there are still many doubts about the security of model sharing. For example, Wang *et al.* [21] discusses the adversarial attacks in the form of back doors during the training process of FL. Geiping *et al.* [5] show that it is possible to faithfully

[3]FedSGD is one classic FL training method, which can be understood as FedAvg with one-step local optimization each round.

[4]The optimal group of communication participators means one possible subset of participating clients, which achieves the best performance by joint training in FedIris.

reconstruct raw images at high resolution using the communication of parameter gradients. Hongxu *et al.* [22] further recover the raw data and labels from a large training batch. Some researchers turn to other communication carriers that can be employed for knowledge transfer. Chang *et al.* [23] uploaded the output prediction to the server for improving local personalization. However, these carriers can not support great performance improvement. FedDG [24] proposed by Quande *et al.* exchanged the amplitude spectrum of local data across clients to transmit the distribution knowledge, while keeping the phase spectrum with core semantics locally for privacy protection. FedMix [25] proposed by Tehrim *et al.* exchanged locally mashed (or averaged) data for reducing a myopic bias. In this work, we attempt to make full use of iris templates for knowledge communication in the proposed framework.

2.2. Federated Biometrics

As a brand new research field, a few works have focused on the application of FL to biometrics, especially the face modality. For example, FedPAD [26] proposed by Shao *et al.* is a presentation attack detection method for face recognition system, which distributedly learned different spoof types from heterogeneous data distributions. FedFace [27] is proposed by Aggarwal *et al.* for collaborative learning of face recognition models. Meng *et al.* [28] applied rigorous differential privacy through the communication of auxiliary embedding centers for federated face recognition. Niu *et al.* [29] proposed FedGC and applied gradient correction for federated face recognition. FedAffect [30] proposed by Shome *et al.* focus on self-supervised few-shot federated learning for facial expression recognition. To our knowledge, this work is the first to apply FL for iris recognition.

3. Theoretical Foundation and Framework

This section provides the theoretical foundation of federated template communication and the framework of FedIris in detail.

3.1. Theoretical Foundation

To ensure the effectiveness of FedIris, some theoretical foundation must be guaranteed for federated template communication. Inspired by the theory of domain adaptation proved by Ben-David *et al.* [6], the introduction of the source domain can provide additional constraints for the error bound of the target domain. For brevity, we assume that there are only two data clients participating in the distributed training as source and target. The purpose is to transfer the knowledge from the source domain to the target domain through federated template communication. At first, the generalization errors of source and target domain are defined for federated template communication in latent space.

Definition 3.1 *The source and target data domain of iris datasets are defined as I_S, I_T, and the raw image distributions are defined as $I_S \sim D_s$ $I_T \sim D_t$. The template extractor g maps the raw data X to iris template R in latent space.*

$$g : X \to R \tag{1}$$

Due to subject variations of various iris dataset collection, the labels are often cross-domain different. Thus, two individual classifiers map the templates to the local predictions independently for source and target as:

$$\begin{aligned} h_s : R \to Y_s, R \sim T_s \\ h_t : R \to Y_t, R \sim T_t \end{aligned} \tag{2}$$

The template distributions of source and target are defined as T_s, T_t. Besides, Y_s and Y_t are defined as the identity labels for the source and target domain. In addition, the probability measure of template distribution can be defined as:

$$Pr_T[B] \stackrel{def}{=} Pr_{D[g^{-1}(B)]} \tag{3}$$

where B is an event in latent space, Pr is the probability measure. The prediction functions of source domain and target domain can be defined as:

$$\begin{aligned} \widetilde{f}^s \stackrel{def}{=} E_{x \sim D_s}\left[\overline{h_s}(g(x))|g(x) = R, R \sim T_s\right] \\ \widetilde{f}^t \stackrel{def}{=} E_{x \sim D_t}\left[\overline{h_t}(g(x))|g(x) = R, , R \sim T_t\right] \end{aligned} \tag{4}$$

where the $\overline{h_s}$ and $\overline{h_t}$ are the predictors of source and target domain, which map the templates to local identity labels. Thus, the generalization error of the source ϵ_{T_s} and the target ϵ_{T_t} in latent space are defined as:

$$\begin{aligned} \epsilon_{T_s}(h_s) = E_{r \sim T_s}\left[E_{y \sim \widetilde{f}^s}\left[y \neq h_s(r)\right]\right] \\ \epsilon_{T_t}(h_t) = E_{r \sim T_t}\left[E_{y \sim \widetilde{f}^t}\left[y \neq h_t(r)\right]\right] \end{aligned} \tag{5}$$

A hypothesis class H is a set of functions satisfying $\forall h_s, h_t \in H, h_s, h_t : R \to Y$.

After definition, the theory of generalization bounds [31] and optimal transportation [32] are borrowed to support the generalization error bound of template communication. **Wasserstein Distance** is one of the distance measurements between two distributions p_1 and p_2 and defined as:

$$W(p_1, p_2) = \inf_{\gamma \sim \pi(p_1, p_2)} E_{(x,y) \sim \gamma}\left[\|x - y\|\right] \tag{6}$$

Generalization bound: With the same assumptions of a Reproducing Kernel Hilbert Space in [31] equipped with kernel $0 \leqslant k_l \leqslant K$. The concentration inequality [31] is

established, with probability at least $1-\delta$ for all hypotheses h the following holds:

$$P\left\{\begin{array}{c}\left|\widehat{\epsilon}_{T_s^*}(h)-\epsilon_{T_s^*}(h)\right|> \\ 2\sqrt{K/n}\left(\frac{\alpha}{n\beta\sqrt{\beta}}+\frac{1-\alpha}{n(1-\beta)\sqrt{1-\beta}}\right)+\epsilon\end{array}\right\} \leqslant exp\left\{\frac{-\epsilon^2 n}{2K\left(\frac{(1-\alpha)^2}{1-\beta}+\frac{\alpha^2}{\beta}\right)}\right\} \tag{7}$$

where $\epsilon_{T_s^*}(h)$ is the empirical combined error and the $\epsilon_{T_s^*}(h)$ is a convex combination of errors on the source and target templates as:

$$\epsilon_{T_s^*}(h)=\alpha\epsilon_{T_t}(h)+(1-\alpha)\epsilon_{T_s}(h) \tag{8}$$

With the introduction of source templates, additional constraints are provided to the generalization error bound of the target domain.

Theorem 3.2 Generalization error bound with best hypothesis: *If $\widehat{h}$ is the minimization of $\epsilon_{T_s^*}(h)$ and $h_{T_t}^*=\min_h \epsilon_{T_t}(h)$ then for any $\delta\in(0,1)$ with probability at least $1-\delta$ in latent space:*

$$\epsilon_{T_t}\left(\widehat{h}\right)\leqslant\epsilon_{T_t}\left(h_{T_t}^*\right)+2(1-\alpha)\left(W(T_s,T_t)+\lambda\right)+\theta \tag{9}$$

where

$$\begin{aligned}\lambda&=\min_h \epsilon_{T_t}(h)+\epsilon_{T_s}(h)\\ \theta&=2\sqrt{2K\left(\frac{(1-\alpha)^2}{1-\beta}+\frac{\alpha^2}{\beta}log(2/\delta)\right)/n}\\ &+4\sqrt{K/n}\left(\frac{\alpha}{n\beta\sqrt{\beta}}+\frac{1-\alpha}{n(1-\beta)\sqrt{1-\beta}}\right)\end{aligned} \tag{10}$$

Theorem 3.2 provides the effectiveness proof of federated template communication. The generalization error of the target domain, with the introduction of source templates, performs better than the best hypothesis of training only on the target templates. At least, the generalization error with the best hypothesis of mixed templates from source and target domain is no larger than the error of the best hypothesis using the target templates alone.

In addition, more constraints can be added to this target generalization error bounds by the introduction of multi-source templates, which is based on the principle of Theorem 4 derived in [33] as:

$$\left|\epsilon_{T_s^*}(h)-\epsilon_{T_t}(h)\right|\leqslant\sum_{i\in[N]}\alpha_i\left(W(T_{s_i},T_t)+\lambda_i\right) \tag{11}$$

where the convex combination of errors alters with the introduction of multi-source representations as:

$$\epsilon_{T_s^*}(h)=\sum_{i\in[N]}\alpha_i\epsilon_{T_{s_i}}(h) \tag{12}$$

and

$$\begin{aligned}\lambda_i&=\min_h \epsilon_{T_{s_i}}(h)+\epsilon_{T_t}(h)\\ \sum_{i\in[N]}\alpha_i&=1\end{aligned} \tag{13}$$

Thus, the generalization error bound of target domain, in the best hypothesis with multi-source templates introduction, can be proved as:

Theorem 3.3 Generalization error bound with multi-source templates introduction: *If h_s^* is the minimization of $\epsilon_{T_s^*}(h)=\sum_{i\in[N]}\alpha_i\epsilon_{T_{s_i}}(h)$ and $h_{T_t}^*=\min_h\epsilon_{T_t}(h)$. For any $\delta\in(0,1)$ with probability at least $1-\delta$ in latent space:*

$$\epsilon_{T_t}\left(h_s^*\right)\leqslant\epsilon_{T_t}\left(h_{T_t}^*\right)+2\sum_{i\in[N]}\alpha_i(W(T_{s_i},T_t)+\lambda_i)+\theta' \tag{14}$$

where

$$\begin{aligned}\lambda_i&=\min_h \epsilon_{T_{s_i}}(h)+\epsilon_{T_t}(h)\\ \theta'&=2\sqrt{2K\sum_{i\in[N]}\alpha_i^2 log(2/\delta)/\beta_i n}+2\sqrt{\sum_{i\in[N]}K\alpha_i/\beta_i n}\end{aligned} \tag{15}$$

Theorem 3.2 and Theorem 3.3 provide the foundations of the generalization error bound of the target with the best hypothesis. More generally, inspired by the weighted error bound for federated domain adaptation [15], the generalization error bound of the target domain with any hypothesis can be proved as:

Theorem 3.4 Generalization error bound in any hypothesis: *If h_{s_i} is the hypothesis of source domain T_{s_i} in latent space:*

$$\epsilon_{T_t}(h)\leqslant\epsilon_{T_s^*}(\sum_{i\in[N]}\alpha_i h_{s_i})+\sum_{i\in[N]}\alpha_i\left(W(T_{s_i},T_t)+\lambda_i\right) \tag{16}$$

Theorem 3.4 represents the dominant factors of target generalization error bound. If the hypothesis space is fixed, this bound depends on two items: 1) the generalization error of the multi-source templates mixture; 2) the distribution distances between multi-source domains and the target domain. This theoretical analysis explains why huge distribution shifts between clients hinder the model generalization, which even cause negative transfer. So the simple averaging aggregation is abandoned in this work, and the distribution distance is taken into consideration for global aggregation.

With regard to simple averaging aggregation, the weight is defined as the ratio of the data scale.

$$\alpha_i=\frac{n_i}{\tilde{N}},\tilde{N}=\sum_{i\in[N]}n_i \tag{17}$$

By contrast, the distribution distance is introduced to the proposed FedIris and this aggregation weight is defined as:

$$\alpha_i = \frac{n_i}{W(T_{s_i}, T_t) \sum\limits_{i \in [N]} \frac{n_i}{W(T_{s_i}, T_t)}} \tag{18}$$

In this setting, the target generalization error bound depends only on the best hypothesis of multi-source reweighting templates mixture.

Algorithm 1 FedIris: federated template communication

Input: Participating clients number N; The template extractor of each client $f_i(*)$, and the parameterθ_i; Communication rounds γ; Local learning rate η;

Output: Each client's iris template extractors after joint training in the setting of FedIris $f_i^*(*)$;

The initialization of each client's extractors.
for $k \in \{1, 2, ..., \gamma\}$ *epoch* **do**
  for $i \in \{1, 2, ..., N\}$ *clients* **do**
    The clients i: Local templates are extracted from local iris raw data by local extractor $f_i(*)$.
    The template center means the averaging of templates with the same identity, as $C_k^i, k \in [n_i]$.
    Templates and centers are sent to the server.
  end for
  The server: The calculation of reweighting Fed-Triplet loss, which details in the Algorithm 2.
  for $i \in \{1, 2, ..., N\}$ **do**
    for $j \in \{1, 2, 3...E\}$ **do**
      The clients i: the loss is received correspondingly by clients to update the local extractor as: $\theta_i^j = \theta_i^{j-1} - \eta \nabla_{\theta_i^{j-1}} L_{tc}^i$.
    end for
  end for
end for
The $f_i^*(*) = f_i(*)$

Algorithm 2 Server: reweighting Fed-Triplet loss

for $i \in \{1, 2, ..., N\}$ *clients* **do**
  Positive template pair is selected as $\langle F_i^a, F_i^p \rangle$
  for $j \in \{1, 2, ..., N\}, j \neq i$*clients* **do**
    Reweighting weight calculation $W_{ji} = \sqrt{\frac{n_j}{n_i}} \frac{1}{W(P_j, P_i)}$
    Cross-domain negative template is select as F_j^n, $L_{ji} = L_{FL}(F_i^a, F_i^p, F_j^n)$
  end for
  The final loss each client i is formed by adding cross-domain Reweighting Fed-Triplet loss and local triplet loss, as $L = \sum\limits_{j \in [N], j \neq i} W_{ji} L_{ji} + L_{triplet}$.
  Then, this loss is sent to the client i for local updating.
end for

3.2. Framework

Federated triplet loss: Although the section 3.1 provides the effectiveness guarantee of iris template communication. How to communicate the local knowledge such as discriminability of extractors, distribution patterns, and template diversity is still a problem. In traditional model-sharing-based FL methods, averaging empirical risk minimization (ERM) is widely applied as the optimization direction. However, it is recently proved that the model generalization greatly suffers from the clients' distribution shifts if optimized by ERM. Inspired by Triplet loss [8], as a classic metric-learning technique, this loss function aims to minimize the intra-class distances and maximize the inter-class distances in latent space. Similarly, this training strategy is shared among the clients and Fed-Triplet loss is proposed for federated template communication with better interpretability. The Fed-Triplet loss is defined as:

$$\begin{aligned} L_{FT} = max(&\|f_i(x_i^a) - f_i(x_i^p)\|_2^2 \\ &- \|f_i(x_i^a) - f_j(x_j^n)\|_2^2 + margin, 0) \end{aligned} \tag{19}$$

where $\langle f_i(x_i^a), f_i(x_i^p), f_j(x_j^n) \rangle, i, j \in [N], i \neq j$ represents that one *Fed-Triplet* selected two templates from one client i with the same identity, and one template from other clients with out-of-distribution identity. The $margin$ is introduced to separate the positive pair from the negative by a distance. As the visualization of the principle of Fed-Triplet in *Fig.* 1, with the insertion of enough cross-domain negative templates, the local inner-class boundary is tightened and the inter-class distance is enlarged. The former enhance the robustness of model generalization, and the later enhances the distinguishability of local identities.

Distribution similarity aggregation: In addition to the distribution knowledge transferred by template communication, the reweighting aggregation is another key point of this work to achieve a lower target generalization error bound. The modified Wasserstein distance normalized by the dataset scale is incorporated into FedTriplet loss to reweight the heterogeneous distributions. The specific reweighting matrix is defined as:

$$W = \begin{bmatrix} 1 & \sqrt{\frac{n_1}{n_2}} \frac{1}{W(P_1,P_2)} & \cdots & \sqrt{\frac{n_1}{n_N}} \frac{1}{W(P_1,P_N)} \\ \sqrt{\frac{n_2}{n_1}} \frac{1}{W(P_2,P_1)} & 1 & \cdots & \sqrt{\frac{n_N}{n_2}} \frac{1}{W(P_2,P_N)} \\ \cdots & \cdots & \cdots & \cdots \\ \sqrt{\frac{n_N}{n_1}} \frac{1}{W(P_N,P_1)} & \sqrt{\frac{n_N}{n_2}} \frac{1}{W(P_N,P_2)} & \cdots & 1 \end{bmatrix} \tag{20}$$

where n_i represents the templates number of client i. P_i presents the template distribution of the $i-th$ dataset. $W(P_i, P_j)$ represents the Earth Mover's Distance (EMD) [12] between clients i and j. Besides, the off-diagonal elements of the weight matrix is set to 1. After that, the weight matrix is normalized by the $\frac{n_i}{n_j}$ to balance scale imbalance between clients. This reweighting aggrega-

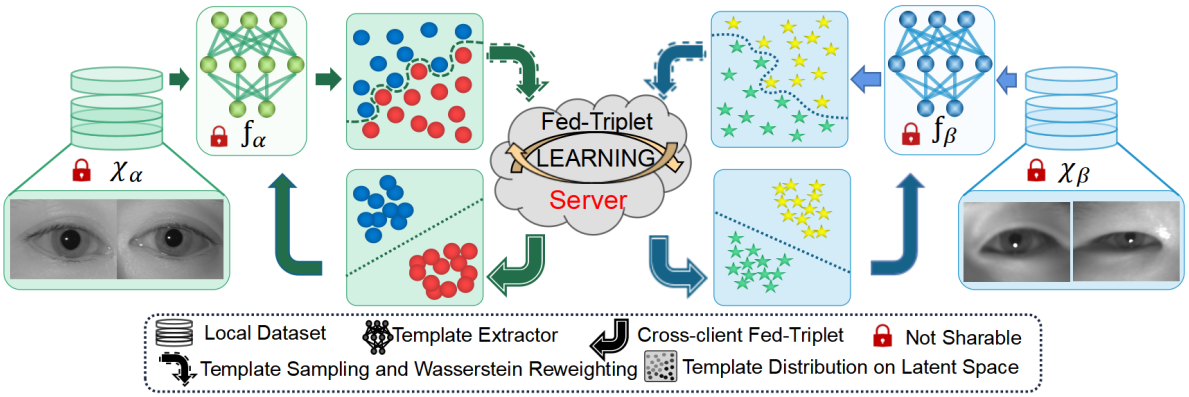


Figure 2. The overall framework of FedIris. FedIris fulfills federated template communication, Fed-Triplet loss and Wasserstein reweighting.

tion drives the clients with more similar template distributions to contribute more mutually.

Framework formulation: Fig. 2 shows the overall framework of the proposed FedIris, which employed the template-communication-based FL. The overall training process is expressed as Algorithm 1. The calculation of reweighting Fed-Triplet loss for each client is expressed as Algorithm 2 in the global server.

4. Experiments

In this section, we introduce the details of the datasets and experiments.

4.1. Dataset

The publicly available iris datasets are of severe heterogeneity and small scale, which are reasonable for validating the proposed framework FedIris. The CASIA_lamp (LA) [34], CASIA_thousand (TH) [35], CASIA_irisV4 (CA) [5], CSIR (CS) [36], ICE (IC) [37] are adopted in the recognition experiments. The adopted datasets are collected for iris recognition, but have distinctive attributes in nature. There are complex differences between these datasets. For example, CASIA_Lamp and CASIA_irisV4 were collected under different illumination; CASIA_Lamp and ICE have distinct differences in image quality and blur degree; ICE and CASIA_irisV4 have huge scale inbalances. Note that there is no overlap of subjects between the partition of the training set and the testing set in each client, where the details of these iris datasets are shown in *Fig.* 3 (a). All samples are proceeded with the same preprocessing methods as [38], including detection, segmentation, and normalization.The Equal Error Rate (EER) is widely used as the metric to evaluate the performance of iris recognition. The formula is as follows:

[5]Portions of the research in this paper use the CASIA-IrisV4 collected by the Chinese Academy of Sciences, Institute of Automation (CASIA)

$$FAR = \frac{N_{FA}}{N_{IRA}} * 100\%$$
$$FRR = \frac{N_{FR}}{N_{GRA}} * 100\% \quad (21)$$

where N_{FA} is the number of false acceptances, N_{IRA} is the number of impostor pairs, N_{FR} is the number of false rejections, and N_{GRA} is the number of genuine pairs. EER is the working point where FAR equals to FRR.

4.2. Iris Recognition Experiments

The iris template extractor in these experiments is the DNN-based UniNet [39] using MindSpore [40] framework. The final recognition results are calculated by the similarities measured by Hamming distance, which is widely used in iris recognition.

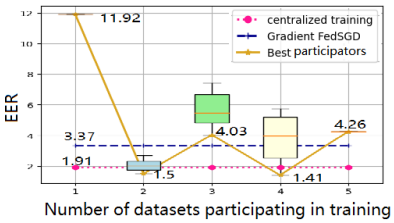
Heterogeneity influence: A group of experiments is conducted to verify the influence caused by the introduction of heterogeneous iris samples. All the possible groups of communication participators are verified. In this part, only Fed-Triplet loss is applied for federated template communication. The EERs are compared with the results of the models trained by centralized learning or FedSGD [4].

Effectiveness: To verify the effectiveness of the federated template communication in the proposed Fed-Triplet loss function, the results are compared under the following four training settings. 1) Five UniNets are initialized randomly and trained on the respective iris training set, which is named *SOLO*. 2) One UniNet is initialized randomly and trained on five training sets aggregated together into a centralized one, called *Centralized*. 3) Five UniNets are initialized randomly and trained by FedSGD [4]. 4) Five UniNets are trained by federated template communication along with the Fed-Triplet loss function. No reweighting aggregation is used here.

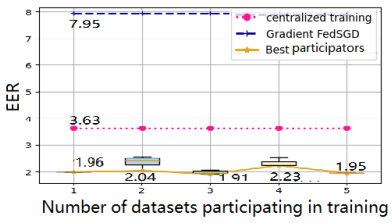
Ablation: The proposed FedIris trained with or without Fed-triplet and reweighting aggregation is validated,

Name	Train/Test	IDs	Samples
CASIA_Lamp	Train	655	12902
	Test	164	3250
CASIA_Thousand	Train	1600	15965
	Test	400	3993
ICE	Train	194	2424
	Test	48	515
CSIR	Train	640	12785
	Test	160	3199
Part of CASIA_irisV4	Train	2230	30435
	Test	557	5563

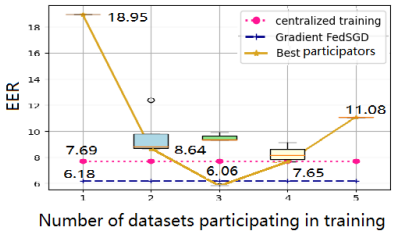
(a) Details of Iris Datasets



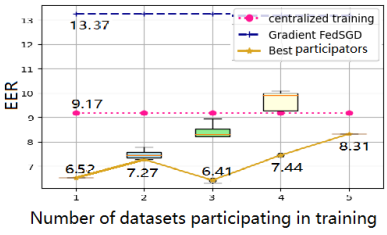
(b)EER on Thousand



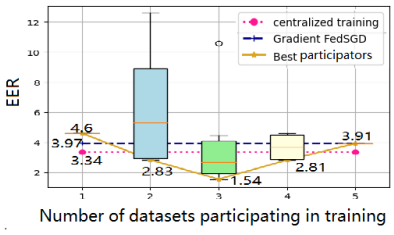
(c) EER on ICE



(d) EER on Lamp



(e) EER on CSIR



(f) EER on CASIAIRIS

Figure 3. The (a) shows the detail of iris datasets. The (b) (f) shows the EERs of the UniNets trained by FedIris with only Fed-Triplet loss function under different number of iris dataset participators. These curves verify that the common sense that more data brings better performance is not always correct and present the severe interference caused by distribution shifts between iris datasets.

which aims to verify the effectiveness and complementarity of these two proposed strategies.

Asynchronous factors: Considering that the asynchronous conditions often occur due to the differences in computing and communication capabilities of each client. The auxiliary experiment aims to represent the degree of performance degradation under asynchronous conditions. First, five UniNets are trained in the manner of SOLO on the corresponding iris dataset. Then, only one randomly selected client is trained continuously through federated template communication by fixing the parameters of the remaining four networks.

Cross-client verification: Cross-client verification is to verify the personalization of each extractor trained by FedIris. To be more concrete, the extractors from one client test on the cross-client dataset may obtain varying performances.

4.3. Results and Analysis

In this section, we show the results and analysis of the iris recognition experiments.

Heterogeneity influence analysis: The performances of FedIris under different groups of communication participators are shown in *Fig.* 3. The following conclusions are obtained by analyzing the results: 1) The distribution shifts between the communication participators may cause negative transfer for EER performances. It is unreasonable to train the model on aggregating more datasets while ignoring distribution shifts. 2) The EERs of the UniNet trained on the optimal group of communication participators are even better than centralized training. It indicates that the heterogeneity in the centralized datasets may have a negative effect on performance improvement. 3) The influences of different source templates' introduction on the performance are not equal due to the imbalance of the scale and quality. For example, the participation of CASIA series datasets benefits Lamp and harms ICE.

Table 1. The LA, TH, IC, CS, and CA represent the abbreviations of the datasets, respectively. "Fed-Triplet" means only applied Fed-Triplet loss function. "FedIris_Asnc" denotes training in the asynchronous condition. "FedIris_All" means the performance that all the clients participate in the communication. "FedIris_Best" implies the theoretically optimal performance as the best optimal group of communication participators can not be known before testing in practice. It can be seen that "FedIris_All" achieves superior performance and is only slightly inferior to "FedIris_Best".

Methods	LA	TH	IC	CS	CA
SOLO	18.95	11.92	1.96	6.52	4.60
Centralized	7.69	1.91	3.63	9.17	3.34
FedSGD	6.18	3.37	7.94	13.37	3.95
FedTriplet	11.08	4.26	1.95	8.31	3.91
FedIris_Asyn	7.94	1.72	2.47	7.54	2.81
FedIris_All	6.13	1.67	**1.79**	7.01	2.17
FedIris_Best	**6.06**	**1.41**	1.91	**6.41**	**1.54**

Effectiveness analysis: The results are shown in the *Tab.* 1. The results of the UniNet trained by FedSGD or FedTriplet are close to the results of centralized training, which are much better than SOLO training. It verifies the

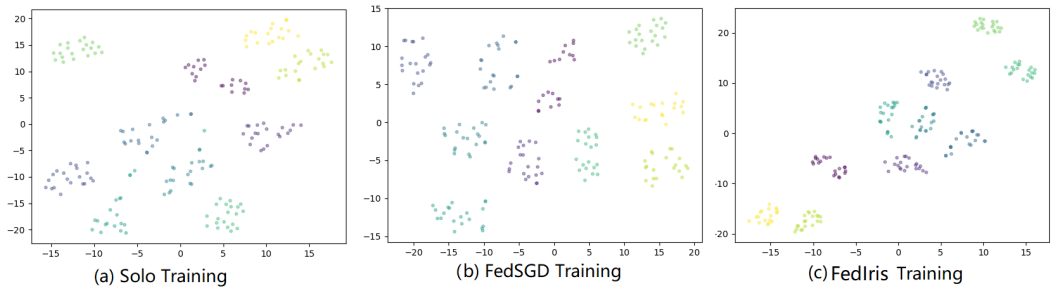


Figure 4. The figure shows the template distributions of CASIA_Thousand [35] dataset. The models are trained by SOLO, FedSGD, and FedIris.

Table 2. Cross-client verification results. EERs are varying on cross-client testing sets. It shows that FedIris can improve the personalization of each client.

FedIris	LA	TH	IC	CS	CA
LA	11.08	5.07	5.06	**5.06**	6.79
TH	6.89	4.26	9.86	14.12	4.83
IC	**4.72**	**0.96**	**1.95**	6.36	3.94
CS	6.50	1.79	2.55	8.31	**2.00**
CA	15.14	7.75	10.24	16.53	3.91

competitiveness of federated template communication. Interestingly, the EERs on CSIR indicate that the final performance of aggregating more data cannot constantly improve. In other words, the introduction of datasets with huge distribution shifts is useless or even harmful.

Ablation study: The results of ablation experiments are tabulated in *Tab.* 1. *FedIris_All* means that all the clients participate in the communication rounds. The EER performances of *FedIris_All* are much better than *Fed-Triplet*. *Fed-Triplet* removes the reweighting aggregation in the proposed FedIris. It is proven that the reweighting aggregation significantly mitigates the negative transfer caused by distribution shifts. In addition, *FedIris_Best* implies the best performance achieved by the optimal group of communication participators. However, such selections of communication participators can not be known beforehand in practice. On the other hand, it can be seen that *FedIris_All* achieves superior performance and is only slightly inferior to *FedIris_Best*. In contrast to *FedIris_Best*, *FedIris_All* is more practical and can be applied in real-world scenarios.

To visualize the distinguishability of templates in latent space, *Fig.* 4 depicts distributions under different training methods such as SOLO, FedSGD, and FedIris through T-SNE techniques. It is demonstrated that FedIris successfully reduces intra-class variance and tightens the intra-class distribution when compared with SOLO and FedSGD training.

Auxiliary analysis: *FedIris_Asnc* in *Tab.* 1 denotes training in the asynchronous condition. The EER performances only show a slight drop compared with *FedIris_All*, which are obviously better than training on centralized datasets. It manifests the potential to transfer the recognition ability by federated template communication, even under asynchronous or less cooperative situations.

Cross-client verification analysis: The results are shown in *Tab.* 2. Compared with model sharing, FedIris provided personalized model parameters for each client. Counter-intuitively, the local extractor is sometimes inferior to cross-client extractors when testing on its own dataset. The underlying reason may be that some inequities among certain clients are introduced by the proposed federated template communication.

5. Conclusions

This work is the first attempt to propose an FL-inspired framework for iris recognition, which is named FedIris. To make full use of the characteristics of iris, federated template communication is proposed for knowledge transfer. The effectiveness of federated template communication is proved with solid theoretical foundations. The optimization strategy is updated from typical averaging ERM to Fed-Triplet loss minimization, which has better interpretability. To mitigate the negative transfer caused by huge distribution shifts among the participating clients, the reweighting aggregation based on Wasserstein distance is incorporated into Fed-Triplet loss, driving the clients with similar template distribution to mutually learn more from each other. Extensive experiments verify that the proposed FedIris surpasses model-sharing FL methods, or even centralized learning in some situations.

Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No.62006225, 61906199, 62176025), sponsored by CAAI-Huawei Mindspore Open Fund.

References

[1] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, et al. 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*, pages 10492–10502, 2021. 1

[2] Junxing Hu, Leyuan Wang, Zhengquan Luo, Yunlong Wang, and Zhenan Sun. A large-scale database for less cooperative iris recognition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE, 2021. 1

[3] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pages 87–102. Springer, 2016. 1

[4] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017. 1, 2, 6

[5] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients–how easy is it to break privacy in federated learning? *arXiv preprint arXiv:2003.14053*, 2020. 2

[6] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*, pages 137–144, 2007. 2, 3

[7] 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. 2

[8] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 2, 5

[9] Durmus Alp Emre Acar, Yue Zhao, Ruizhao Zhu, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. Debiasing model updates for improving personalized federated training. In *International Conference on Machine Learning*, pages 21–31. PMLR, 2021. 2

[10] 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. 2

[11] 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. 2

[12] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000. 2, 5

[13] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020. 2

[14] Fei Chen, Mi Luo, Zhenhua Dong, Zhenguo Li, and Xiuqiang He. Federated meta-learning with fast convergence and efficient communication. *arXiv preprint arXiv:1802.07876*, 2018. 2

[15] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. *arXiv preprint arXiv:1911.02054*, 2019. 2, 4

[16] Daliang Li and Junpu Wang. Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581*, 2019. 2

[17] Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[18] Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-free knowledge distillation for heterogeneous federated learning. *arXiv preprint arXiv:2105.10056*, 2021. 2

[19] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. In *Advances in Neural Information Processing Systems*, pages 4424–4434, 2017. 2

[20] 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. 2

[21] Hongyi Wang, Kartik Sreenivasan, Shashank Rajput, Harit Vishwakarma, Saurabh Agarwal, Jy-yong Sohn, Kangwook Lee, and Dimitris Papailiopoulos. Attack of the tails: Yes, you really can backdoor federated learning. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[22] Hongxu Yin, Arun Mallya, Arash Vahdat, Jose M Alvarez, Jan Kautz, and Pavlo Molchanov. See through gradients: Image batch recovery via gradinversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16337–16346, 2021. 3

[23] Hongyan Chang, Virat Shejwalkar, Reza Shokri, and Amir Houmansadr. Cronus: Robust and heterogeneous collaborative learning with black-box knowledge transfer. *arXiv preprint arXiv:1912.11279*, 2019. 3

[24] Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1013–1023, 2021. 3

[25] Tehrim Yoon, Sumin Shin, Sung Ju Hwang, and Eunho Yang. Fedmix: Approximation of mixup under mean augmented federated learning. *arXiv preprint arXiv:2107.00233*, 2021. 3

[26] Rui Shao, Pramuditha Perera, Pong C Yuen, and Vishal M Patel. Federated face anti-spoofing. *arXiv preprint arXiv:2005.14638*, 2020. 3

[27] Divyansh Aggarwal, Jiayu Zhou, and Anil K Jain. Fedface: Collaborative learning of face recognition model. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021. 3

[28] Qiang Meng, Feng Zhou, Hainan Ren, Tianshu Feng, Guochao Liu, and Yuanqing Lin. Improving federated learning face recognition via privacy-agnostic clusters. *arXiv preprint arXiv:2201.12467*, 2022. 3

[29] Yifan Niu and Weihong Deng. Federated learning for face recognition with gradient correction. *arXiv preprint arXiv:2112.07246*, 2021. 3

[30] Debaditya Shome and Tejaswini Kar. Fedaffect: Few-shot federated learning for facial expression recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4168–4175, 2021. 3

[31] Ievgen Redko, Amaury Habrard, and Marc Sebban. Theoretical analysis of domain adaptation with optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 737–753. Springer, 2017. 3

[32] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781. 3

[33] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010. 4

[34] Zhuoshi Wei, Tieniu Tan, and Zhenan Sun. Nonlinear iris deformation correction based on gaussian model. In *International Conference on Biometrics*, pages 780–789. Springer, 2007. 6

[35] Hui Zhang, Zhenan Sun, and Tieniu Tan. Contact lens detection based on weighted lbp. In *2010 20th International Conference on Pattern Recognition*, pages 4279–4282. IEEE, 2010. 6, 8

[36] Qi Zhang, Haiqing Li, Zhenan Sun, and Tieniu Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, 13(11):2897–2912, 2018. 6

[37] P Jonathon Phillips, Kevin W Bowyer, Patrick J Flynn, Xiaomei Liu, and W Todd Scruggs. The iris challenge evaluation 2005. In *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8. IEEE, 2008. 6

[38] Caiyong Wang, Jawad Muhammad, Yunlong Wang, Zhaofeng He, and Zhenan Sun. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on Information Forensics and Security*, 15:2944–2959, 2020. 6

[39] Zijing Zhao and Ajay Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3809–3818, 2017. 6

[40] Mindspore. http://www.mindspore.cn. 6