Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/52/20241138

# The application of federated learning in face recognition: A systematic investigation of the existing frameworks

**Chuanzhi Xu**

School of Computer Science, University of Sydney, Camperdown NSW 2050, Australia

chxu4146@uni.sydney.edu.au

**Abstract.** This paper presents a thorough examination of the recent progress made in applying federated learning to the field of face recognition. As face recognition technology continues to gain widespread adoption across various sectors, issues related to data privacy and efficiency have taken center stage. In response, federated learning, characterized by its decentralized machine learning approach, has emerged as a promising solution to tackle these pressing concerns. This review categorises the current federated learning frameworks for face recognition into four main purposes: Training Efficiency, Recognition Accuracy, Data Privacy, and Spoof Attack Detection. Each category is explored in-depth, highlighting the principles, structures, applicability, and advantages of the frameworks. The paper also delves into the challenges faced in the integration of federated learning and face recognition, such as high computational overhead, model inconsistency, and data heterogeneity. The review concludes with recommendations for future research directions, emphasising the need for model compression, asynchronous communication strategies, and techniques to address data heterogeneity. The findings underscore the potential and challenges of applying federated learning in face recognition, paving the way for more secure and efficient facial recognition systems.

**Keywords:** Federated Learning, Face Recognition, Data Privacy.

## 1. Introduction

Face recognition, also known as facial recognition, is a technology that involves identifying and verifying individuals based on their facial features. It is a biometric method used to automatically recognise and authenticate a person's identity by analysing unique patterns and characteristics of their face. Nowadays, face recognition technology is very mature and has been widely used in many areas of daily life, including security monitoring, mobile phone unlocking, payment verification, etc.

However, the training and enhancement of face recognition intelligent models may affect users' privacy. Training such a model usually requires a large amount of face image data, which may lead to the leakage of users' private data if these data are not correctly processed and protected. In the data collection phase, the user's privacy may be violated if the user's explicit consent is not obtained. In the process of data storage and transmission, if it is not properly encrypted, it may be hacked and stolen. This also reveals the seriousness of privacy and security issues in face recognition.

DOI: 10.54254/2755-2721/52/20241138

Federated Learning is a machine learning approach in which models are trained on multiple decentralised devices holding local data samples without exchanging them. Privacy and security are ensured since the raw data does not leave the local device. Currently, Federated Learning algorithms are widely used in the fields of healthcare, finance, telecommunication, intelligent transport, and computer vision. It has the potential to be applied to many other areas. Therefore, applying Federated Learning to face recognition will be a good approach, but it is still in its infancy.

Recent advancements in the integration of federated learning and face recognition have made significant strides across various technological domains. In terms of training efficiency, the Industrial Framework for AIoT Face Recognition leverages transfer learning to enhance the efficiency of federated learning, reducing communication rounds and computational costs [1]. Additionally, the FedFace framework optimises the efficiency of pre-trained face recognition systems by harnessing additional data available on mobile devices, ensuring the retention of facial images on local devices [2]. For recognition accuracy, the FedFR framework proposed by Zhuang et al. employs a federated unsupervised domain adaptation technique to address discrepancies in data distributions between training and deployment phases [3], while Liu et al.'s iteration of FedFR aims to enhance generic face representation and optimise personalised models [4]. The FedFV framework is tailored for scenarios where data privacy concerns limit the sharing of class embeddings, achieving enhanced recognition accuracy [5]. On the privacy front, the FedAffect framework addresses the training of facial expression recognition with localised raw data, promoting better data privacy in federated learning for facial recognition [6]. The FedPad framework, on the other hand, ensures a collaborated and trained global model that respects user privacy [7]. Lastly, in detecting face spoof attacks, the FedFSAD framework by Chen et al. enhances traditional federated learning with multitask learning and manifold regularisation, showing a marked improvement in the detection of face spoof attacks [8].

This paper will revisit the studies to categorise and analyse the above frameworks they introduced and investigate the advantages of applying Federated Learning to face recognition. Furthermore, it will identify existing shortcomings and research gaps, elucidating why the application of federated learning in face recognition remains in its nascent stages.

In the subsequent sections of this paper, Section 2 will provide an overview to categorise the above frameworks. Section 3 will delve into a comparison of important frameworks and discuss their limitations, offering holistic recommendations for applying Federated Learning in face recognition. Section 4 will synthesise all perspectives and provide a systematic conclusion for this paper.

## 2. Categorisation and Analysis

Most frameworks below may contribute to multiple improvements of federated learning applied in facial recognition, but here, they are classified according to their most important feature.

### *2.1. Improvement in Training Efficiency*

#### *2.1.1. Industrial Federated Learning Framework for AIot's Face Recognition*

Ding et al. introduce an efficient industrial federated learning framework designed explicitly for AIoT face recognition applications [1]. The framework utilises transfer learning to lift the federated learning efficiency, reducing the communication rounds with computational costs. This framework speeds up the data training process, without the end devices' facial image sharing, to achieve the desired recognition accuracy in 20 rounds of communication on a private Asian face dataset.

DOI: 10.54254/2755-2721/52/20241138

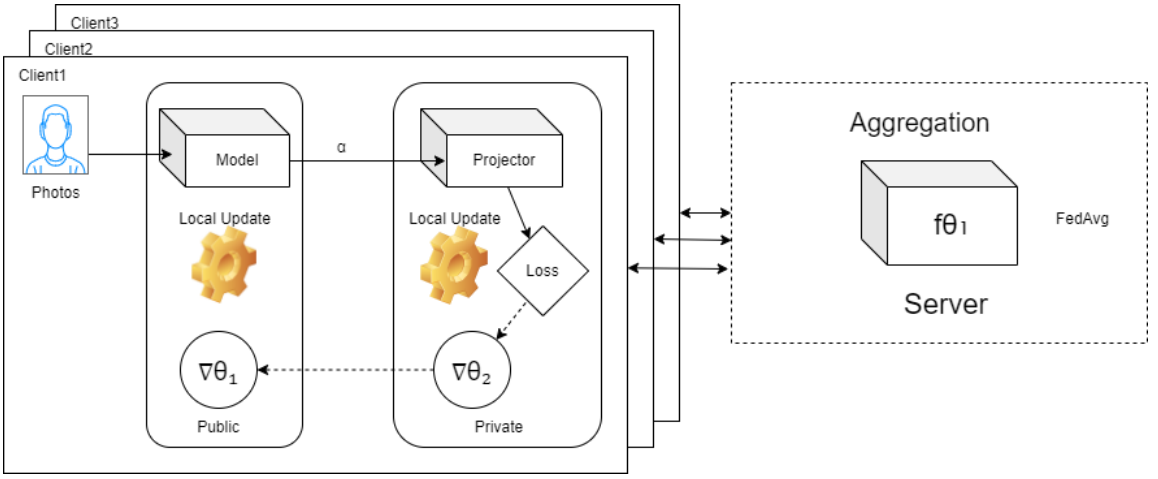


**Figure 1.** The illustration diagram of the Industrial Federated Learning Framework for AIot's Face Recognition.

Figure 1 displays the operational processes of the framework. The server sends the pre-trained to all clients. This will add a private projector. In each round of communication, the public module will spread to the server for aggregation.

### *2.1.2. FedFace*

The FedFace framework shown in Figure 2 introduced by Aggarwal et al. trains an accurate and generalisable face-recognising model with multiple clients' face images [2]. As a mobile device, the client only holds its own face image, which cannot be shared with the other clients and the central server.

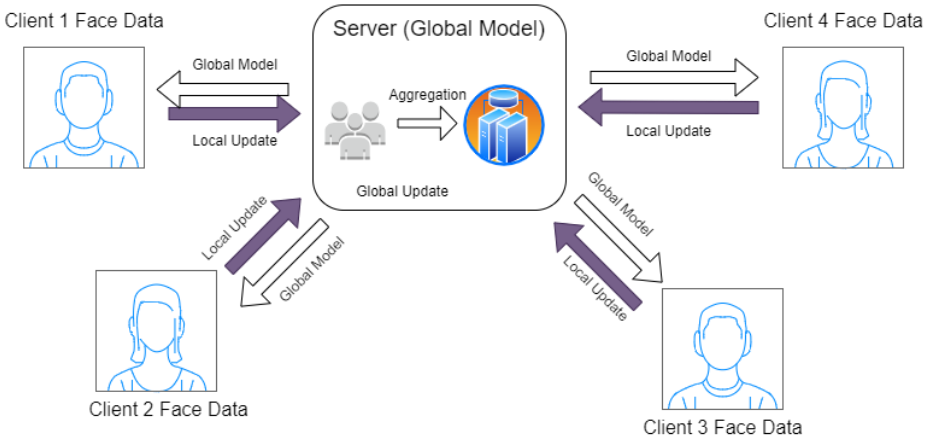


**Figure 2.** FedFace Framework Illustration Diagram.

In each communication round, the server dispatches the global model weights to all participating client nodes. These client nodes then adjust their respective local models using their individual training data. The updated local models are subsequently transmitted back to the server, where they undergo aggregation to produce a global update. The FedFace framework effectively improves the efficiency of a pre-trained face recognition system by leveraging additional data from mobile devices, all while ensuring that facial images remain stored locally on the devices.

DOI: 10.54254/2755-2721/52/20241138

### *2.2. Improvement in Recognition Accuracy*

#### *2.2.1. FedFR*

Zhuang et al. proposed a federated unsupervised domain adaption framework for face recognition called FedFR to solve the domain shift problem [3], where the model sees different data distributions between the training source domain and the deployment target domain. Their experiments show the FedFR's performance on verification accuracy in face recognition is 3%-14% higher than baseline and other classic frameworks.

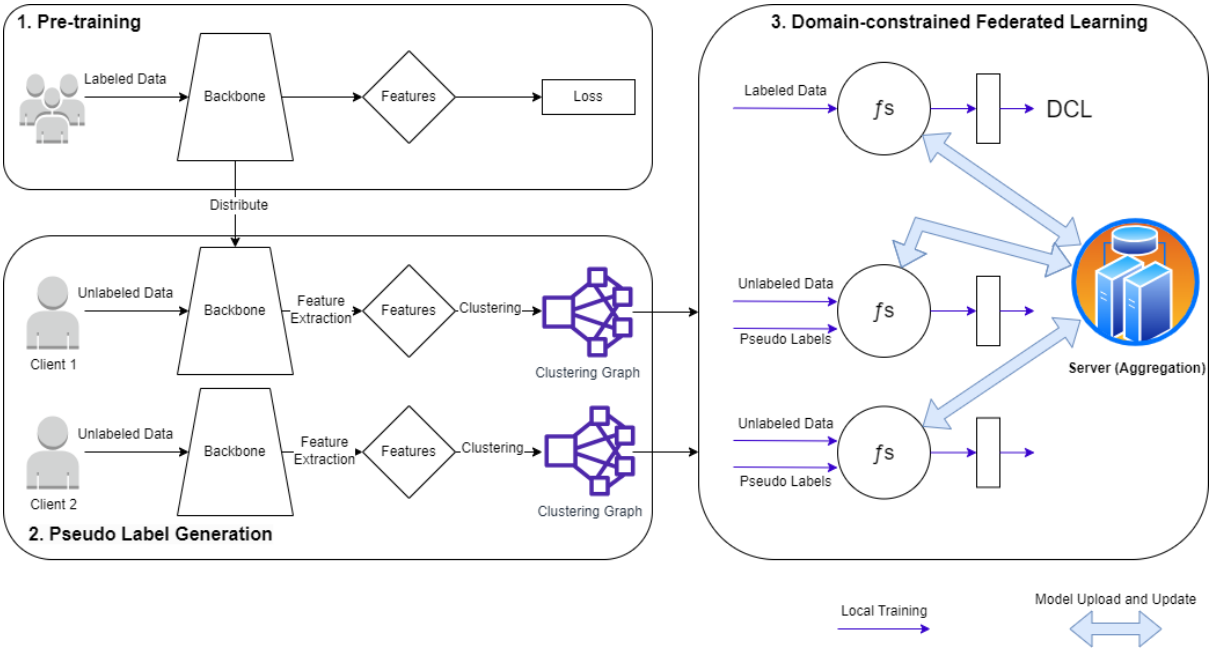


**Figure 3.** Framework Illustration of FedFR by Zhuang et al.

This FedFR framework shown in Figure 3 consists of three stages: source domain pre-training, pseudo-label generation, and domain-constrained federated learning. In the pre-training step, the labelled data from the source domain is used to train the model. In the pseudo-label generation, the unlabeled data in the target domain are assigned pseudo labels using an enhanced hierarchical clustering algorithm. In domain-constrained federated learning, the knowledge is iteratively transferred from the source domain to the target domain via federated learning, ensuring data privacy by transmitting model updates instead of raw data. A novel domain constraint loss (DCL) is introduced to guide the source domain training. These three stages form an end-to-end training process that requires data to be unshareable across the domain and requires edge devices as the client to collect the unlabeled data in the target domain.

#### *2.2.2. FedFR*

The FedFR framework introduced by Liu et al. aims to enhance the generic face representation while preserving user privacy and optimising personalised models for individual clients [4]. It employs techniques such as hard negative sampling and contrastive regularisation to bridge the gap between global and local representations, aiming for improved face recognition accuracy.

FedFR consists of three methodologies. A module called Decoupled Feature Customization (DFC) consists of a feature transformation layer and one-vs-all binary classifiers, which is designed to learn a customised feature space optimised for recognising registered identities on each client. Hard Negative Sampling is a strategy employed to select the most critical data samples from a globally shared dataset, thereby reducing computation overhead and improving training efficiency. The contrastive

DOI: 10.54254/2755-2721/52/20241138

Regularization technique is applied to local face representation during training to restrict the local model from drifting too far from the global model.

### *2.2.3. FedFV*

Liu et al. proposed a framework called Federated Face Verification (FedFV) shown in Figure 4 [5], designed for scenarios where each client can only access face images of a single class, and due to privacy concerns, class embeddings cannot be shared among clients. In the FedFV framework, the server gathers all client class embeddings and generates "equivalent class embeddings." which are then sent to clients, ensuring that each client's class embeddings are distinctly separated from others. The framework demonstrates superior performance compared to existing federated face verification techniques across several benchmarks, which contributes to better recognition accuracy of face data.

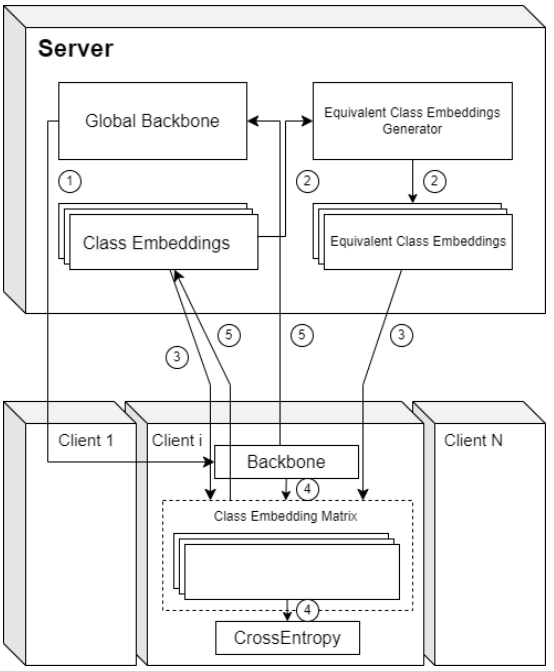


**Figure 4.** FedFV framework Illustration Diagram.

FedFV framework is specifically implemented by the following steps. The server sends the glocal backbone parameters to clients. Original Class Embeddings create the equivalent class embeddings, and then they are all dispatched to the respective client. The model is trained using client data. Parameter gradients are returned to the server. A server-side generator produces equivalent class embeddings from client class embeddings. These are then sent to selected clients, which utilise them alongside their own class embeddings to compute the Softmax Loss, ensuring distinct separation between all class embeddings.

### *2.2.4. Cascade Network*

Zhu et al. introduced a cascade network used for face detection with the mask [9]. The first level is the Dilation RetinaNet Face Location (DRFL) Network, which incorporates the Enhanced Receptive Field Context (ERFC) module, aiming to reduce network parameters and locate faces of different scales. The second level is the SRNet20 network used to fulfil embedded camera devices crafted through Neural Architecture Search (NAS).

DRFL can be used for face location. This network combines the ERFC module, which uses dilated convolutions to increase receptive fields, thereby reducing network parameters and efficiently localising

DOI: 10.54254/2755-2721/52/20241138

faces at different scales. This design facilitates accurate detection and localisation of faces in dense crowds, especially when faces may be occluded by other objects or people.

The SRNet20 network is used to classify whether the face is wearing a mask or not. This network was created via Neural Architecture Search (NAS), and it was designed to efficiently identify and classify faces wearing masks. More importantly, because of privacy considerations, the SRNet20 network has carried on the training in the study, which means that each device on the local data for training uploads the model parameters only but not the actual image data. This approach not only protects the privacy of the user but also allows more training data to be obtained from multiple data sources, thus improving the accuracy of the model.

### *2.3. Improvement in Data Privacy*

#### *2.3.1. FedAffect*

Shome & Kar propose a few-shot federal learning framework named FedAffect used in facial expression recognition [6]. Utilising a small number of labelled private facial expression data, FedAffect can train local models and aggregate the weight of the local model to the centre server, making the model achieve global optimisation.

FedAffect combines two distinct neural networks - one responsible for self-learning feature representations from a large scale of unlabeled facial images and another using these representations to extract features and predict probability scores for facial expressions in few-shot learning.

FedAffect mainly deals with the training problem of recognition of facial expressions from any private and decentralised data on user devices, which promotes federated learning in facial recognition through better data privacy.

#### *2.3.2. FedPad*

Shao et al. introduce a framework called FedPad [7] shown in Figure 5, fully named "Federated Face Presentation Attack Detection", for detecting face presentation attacks in the modern face recognition pipeline. A face presentation attack detection model with good generalisation can be obtained when it is trained with face images from different input distributions and different types of spoof attacks.

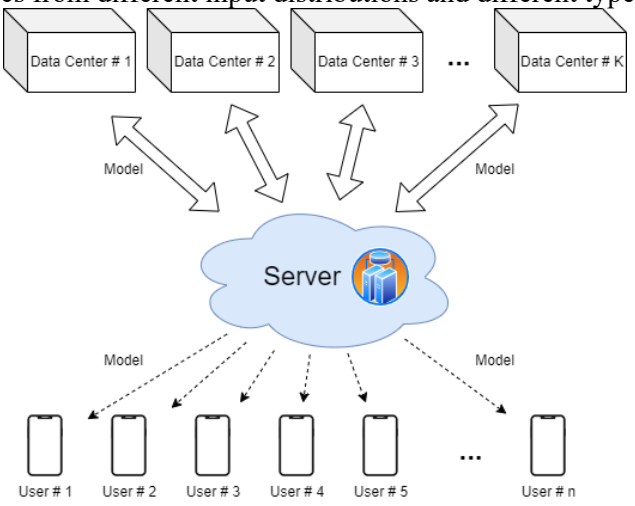


**Figure 5.** FedPad Illustration Diagram.

The FedPAD framework, categorised as traditional federated learning, allows different data centres to train their own face presentation attack detection model while preserving data privacy. A server learns a global model by iteratively aggregating model updates from all data centres without accessing their private data. After multiple rounds of communication between servers and data centres, a collaborated

DOI: 10.54254/2755-2721/52/20241138

trained and global model is achieved while ensuring data privacy. Users can easily retrieve the model from servers to their devices for detecting diverse face presentation attacks.

The main goal of this framework is to recognise facial expressions from any private and decentralised data on user devices, thus promoting federated learning in facial recognition through better data privacy.

### *2.3.3. FedGC*

Niu & Deng propose a federated learning framework specialised in facial recognition called FedGC [10], utilising backward propagation to correct gradients, aiming to solve the problem of privacy leakage caused by traditional decentralised, federated learning.

FedGC framework introduces a softmax-based regulariser by precisely injecting a cross-client gradient to correct the gradient of local softmax, being conducive to effective learning on discriminative face representations. This framework ensures that each client has a private class embed, the network updates are directed the samely as standard SGD, and the class embedding is fully unfolded between different clients, further enhancing privacy. It has been validated that the effectiveness and performance can even match other centralised training methods.

## *2.4. Improvement in detecting face spoof attacks (FedFSAD)*

Chen et al. proposed a novel framework called federated learning for face spoof attack detection (FedFSAD) [8], aiming to enhance traditional federated learning with multitask learning and manifold regularisation. Traditional methods for face spoof attack detection heavily rely on the quantity of training data. The federated learning approach of FedFSAD can alleviate this issue by efficiently and safely utilising distributed data. By using FedFSAD, the performance on detecting face spoof attacks is improved by 10%, and it shows robustness against network delays.

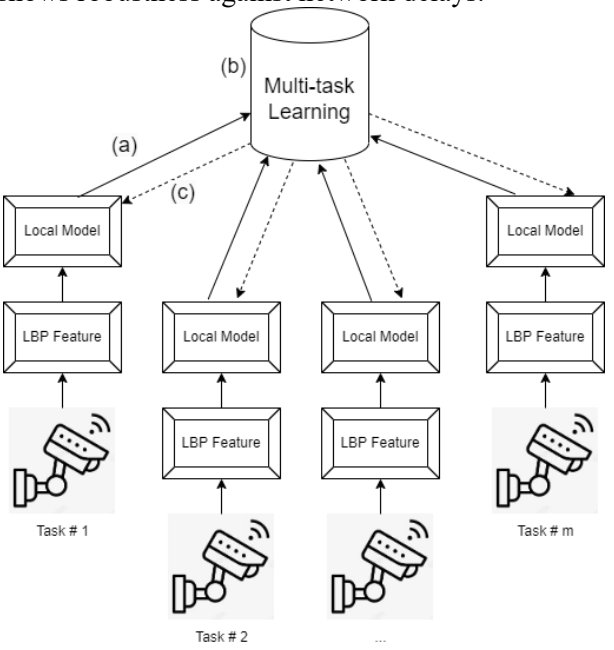


**Figure 6.** FedFSAD Illustration Diagram.

The diagram illustrates the processes of FedFSAD shown in Figure 6 in face spoof attack identification. (a) Each face recognition subsystem has a designated task for detecting these attacks. FedFSAD utilises LBP (Local Binary Pattern) features to represent facial images, which help train local models for each subsystem. (b) These individual models are sent to a central server, where they undergo a multitask learning process to develop a unified global model. (c) This global model is dispatched back to the respective subsystems, enabling them to detect face spoof attacks effectively.

DOI: 10.54254/2755-2721/52/20241138

## 3. Discussions

### *3.1. Comparison*

Section 2 explains the purposes, structure, and advantages of each of the most important frameworks or methods proposed to apply Federated Learning in Face Recognition. This part will select two from the main three purposes for comparison.

#### *3.1.1. Training Efficiency (Industrial Framework for AIot & FedFace)*

Industrial Federated Learning Framework for AIoT leverages transfer learning to enhance federated learning efficiency, reducing communication rounds and computational costs. In contrast, FedFace focuses on harnessing additional data available on mobile devices, ensuring that face images remain on local devices, thus streamlining the training process. They both contribute to better training efficiency. However, the former only suits AIot devices, while FedFace can work for all mobile devices.

#### *3.1.2. Recognition Accuracy (FedFR & FedFV)*

The FedFR proposed by Zhuang et al. employs a federated unsupervised domain adaptation technique, addressing discrepancies in data distributions between training and deployment phases to enhance recognition accuracy. In contrast, FedFV is tailored for scenarios where data privacy concerns limit the sharing of class embeddings. By generating "equivalent class embeddings", FedFV achieves enhanced recognition accuracy. Each caters to unique challenges in the face recognition model training, highlighting the versatility of federated learning solutions.

#### *3.1.3. Data Privacy (FedAffect & FedPad)*

FedAffect focuses on the model training of facial expression recognition with localised raw data, utilising two neural networks for self-learning feature representations and facial expression prediction. FedPad focuses on detecting the face presentation attack by training the model on different data centres. By iteratively aggregating model updates from all data centres without accessing their private data, FedPad ensures a collaborated and trained global model that respects user privacy. Although both aim to enhance local data privacy in training processes, they work for different facial recognition sub-domains.

### *3.2. Limitations*

#### *3.2.1. High Computational Overhead*

Federated learning requires multiple rounds of communication, including model sending, model update, and aggregation, between the server and client devices. The FedFace framework sends global model weights to each participating client node and receives local updates from them for central computation and aggregation, which is an iterative and computationally intensive process when dealing with large models and numerous client devices.

#### *3.2.2. Model Inconsistency*

Devices in a federated network may differ in terms of storage, computing power, and network connectivity. Ensuring consistent model updates on heterogeneous devices can be challenging. Different devices may have different computing power, resulting in different model update speeds. Moreover, with a large number of clients, only a few devices can perform computation in each round [11]. The inconsistency this creates will degrade model performance and slow down convergence.

#### *3.2.3. Weak Client Devices*

Most of the current face recognition devices are mobile or embedded devices [11]. Each device holds only limited computing and communication power and a small amount of data, making it difficult to perform a large number of computations to update a large face recognition model locally.

DOI: 10.54254/2755-2721/52/20241138

#### *3.2.4. Data Heterogeneity*

In federated learning, data is distributed across multiple devices, each of which may have a unique data distribution. For example, the FedFR framework's approach proposed by Liu et al. needs to handle different data distributions from different clients, which may complicate the learning process and impact the model's performance.

### *3.3. Improvements*

Therefore, in order to further improve the application of federated learning in face recognition in light of the above limitations, the following research directions can be focused. Further compressing the model parameters to be updated can reduce the cost of communication and local training, which adapts to the poor performance of some local devices. Investigation of asynchronous communication and model update strategies can allow devices to upload and download model updates at different points in time, also accommodating different devices' computing power. Communication efficiency is also important. Reducing the number of communications and the amount of data between the client and the central server is key to improving the efficiency of federated learning. More efficient model aggregation methods and communication protocols can be investigated. Research on how to deal with local facial data heterogeneity can enhance the training efficiency of federated learning. Utilising meta-learning or transfer learning techniques can adapt to different data distributions.

## 4. Conclusion

This paper reviews and purposefully categorises some innovative frameworks for applying federated learning to face recognition, briefly introducing their principles and structures and presenting their applicability and advantages. The discussion section selects frameworks for comparison and explains some shortcomings. The current federated learning frameworks for face recognition can be broadly categorised into Training Efficiency, Recognition Accuracy, Data Privacy, and Spoof Attack detection in terms of their purpose. The limitations that prevent these frameworks from being further applied in the industry include High Computational Overhead, Model Inconsistency, Weak Client Devices, and Data Heterogeneity. This paper also proposes avenues for advancing federated learning in the realm of facial recognition, including the reduction of communication expenses and the enhancement of overall efficiency. In conclusion, the application of federated learning in face recognition holds significant promise and offers substantial research opportunities.

## References

[1] Ding Y et al 2022 An Efficient Industrial Federated Learning Framework for AIoT: A Face Recognition Application ArXiv.org. https://doi.org/10.48550/arXiv.2206.13398

[2] Aggarwal D Zhou J and Jain A K 2021 FedFace: Collaborative Learning of Face Recognition Model ArXiv:2104.03008 [Cs]. https://arxiv.org/abs/2104.03008

[3] Zhuang W et al 2022 Federated Unsupervised Domain Adaptation for Face Recognition IEEE Xplore https://doi.org/10.1109/ICME52920.2022.9859587

[4] Liu C-T et al 2022 FedFR: Joint Optimization Federated Framework for Generic and Personalized Face Recognition ArXiv.org. https://doi.org/10.48550/arXiv.2112.12496

[5] Liu L Zhang Y Gao H Yu X and Cheng J 2022 FedFV: federated face verification via equivalent class embeddings Multimedia Systems https://doi.org/10.1007/s00530-022-00927-5

[6] Shome D and Kar T 2021 FedAffect: Few-shot federated learning for facial expression recognition IEEE Xplore https://doi.org/10.1109/ICCVW54120.2021.00463

[7] Shao R Perera P Yuen P C and Patel V M 2020 Federated Face Presentation Attack Detection ArXiv.org. https://doi.org/10.48550/arXiv.2005.14638

[8] Chen Y Chen L Hong C and Wang X 2022 Federated Multitask Learning with Manifold Regularization for Face Spoof Attack Detection. Mathematical Problems in Engineering e7759410. https://doi.org/10.1155/2022/7759410

DOI: 10.54254/2755-2721/52/20241138

[9] Zhu R Yin K Xiong H Tang H and Yin G 2021 Masked Face Detection Algorithm in the Dense Crowd Based on Federated Learning. Wireless Communications and Mobile Computing e8586016 https://doi.org/10.1155/2021/8586016
[10] Niu Y and Deng W 2021 Federated Learning for Face Recognition with Gradient Correction ArXiv.org. https://doi.org/10.48550/arXiv.2112.07246
[11] Shang E et al 2022 FedFR: Evaluation and Selection of Loss Functions for Federated Face Recognition 95–114 https://doi.org/10.1007/978-3-031-24383-7_6