Waref *et al. Journal of Cloud Computing* (2025) 14:24
https://doi.org/10.1186/s13677-025-00746-9

Journal of Cloud Computing

RESEARCH

Open Access

Client aware adaptive federated learning using UCB-based reinforcement for people re-identification

Dinah Waref[1*], Yomna Alayary[1], Nadeen Fathallah[2], Mohamed A. Abd El Ghany[3,4] and Mohammed A.-M. Salem[1]

Abstract

People re-identification enables locating and identifying individuals across different camera views in surveillance environments. The surveillance data contains personally identifiable information such as facial images, behavioral patterns, and location data, which can be used for malicious purposes such as identity theft, stalking, or discrimination. This raises serious ethical and privacy concerns. The communication overhead of transporting a large number of data needed to train a global model and the diverse nature of the data from different sources are serious limitations facing the development of people re-identification technologies. We address these challenges by proposing a novel three-step federated learning framework. First, we investigate the impact of data augmentation techniques on the model generalizability and explore the effectiveness of different backbone networks. Second, we use reinforcement learning-based Upper Confidence Bounds (UCB) as a client-selection strategy in the federated round that dynamically chooses devices similar to the current model state, ensuring the model is updated with relevant data and enables faster convergence. Finally, we introduce a feature-level attention mechanism focusing on discriminative features for re-identification. Extensive experiments were conducted on nine benchmark re-ID datasets. The proposed framework outperformed the federated re-ID baseline by 10% in rank-1 accuracy and achieved results comparable to the centralized approach, with a difference of 2%. This improvement over the previous state-of-the-art establishes a new benchmark for federated re-identification.

Keywords Federated learning, People re-identification, Non-IID, Contrastive learning

Introduction

Recent advancements in Artificial Intelligence (AI) have led to significant breakthroughs in various fields ranging from healthcare and retail to autonomous systems and natural language processing. These advancements also raise ethical and privacy concerns [1] with respect to collecting sensitive or personally identifiable information, which is any data that can be used to identify an individual directly or indirectly, such as facial features, a person's walking pattern, which is unique and can be used as a biometric identifier, clothing, and location. This data collection and storage, which is essential for any deep learning model, can easily trespass on individual rights. People

*Correspondence:
Dinah Waref
dinah.lamei@guc.edu.eg
[1]Media Engineering and Technology, German University in Cairo, Cairo, Egypt
[2]Institute for AI, University of Stuttgart, Stuttgart, Germany
[3]Information Engineering and Technology, German University in Cairo, Cairo, Egypt
[4]Integrated Electronic Systems Lab, TU Darmstadt, Darmstadt, Germany

re-identification (re-ID) technology, the task of locating and identifying people across different camera views, is easily integrated into existing surveillance systems [2, 3]. This ability to identify and track people can be beneficial for retail to target advertisements, healthcare to monitor patients, law enforcement to track suspects or suspicious individuals and locate missing people, and public safety. The increasing availability and affordability of surveillance technology have helped the widespread adoption of surveillance technologies in many places, such as hospitals and airports, for security to enhance security and deter crime and privacy by ensuring that only authorized personnel enter restricted zones. Although the presence of surveillance cameras in a single airport may not raise significant privacy concerns, sharing the data collected by these cameras between different airports can create significant risks. Thus, the prevalence of surveillance cameras is a huge privacy violation and goes against stringent data protection regulations such as the General Data Protection Regulation (GDPR) [4, 5]. As such, there is an increasing demand for privacy-preserving alternatives to mitigate the challenges [6] facing the development of a robust and effective re-ID system. A clear example of data misuse is the DukeMTMC dataset, samples of the dataset are shown in Fig. 1. It was retracted after repeated warnings from Human Rights Watch due to the misuse of the dataset by government entities for mass surveillance and human rights violations. The retraction emphasized the need for stricter controls on how sensitive data is used and shared, especially when it involves personal information that could be weaponized against individuals or groups.

Federated Learning (FL) offers a promising solution where it enables collaborative learning by training a local model on each edge device [8], thus eliminating the need for raw data sharing [9, 10]. Training local models across decentralized data sources (e.g., multiple cameras or institutions) and only sharing model updates with the central server, preserves the individuals' right to their own data. Despite the potential this approach holds [11], it presents unique challenges related to the data scarcity and heterogeneity naturally present in the data due to the difference in demographics, illumination, pose differences, camera view, and angle. Another dominant problem is the edge device's limited computational resources [12], which leads to poor performance and bias towards devices with more data and computational resources [13].

While federated learning provides a solid foundation for privacy-preserving people re-identification, current systems struggle with handling data heterogeneity across different clients, leading to inconsistent model performance. Moreover, existing methods do not fully exploit the potential of attention mechanisms to enhance feature learning to distinguish between similar-looking individuals. Additionally, the resource constraints in federated environments require efficient client selection and resource allocation strategies, which are overlooked in the currently available research. Based on these observations, in this study, we develop a federated learning-based person re-identification system to overcome data inconsistencies, refine feature extraction with attention mechanisms, and efficiently manage participating devices and computational resources through reinforcement learning.

The main contributions of this paper are summarized as follows:

- Develop adaptive aggregation strategies for handling data heterogeneity in federated Re-ID systems, ensuring consistent model performance across diverse clients.
- Integrate feature attention mechanisms into the federated Re-ID framework to improve discriminative feature learning and enhance model accuracy.
- Implement a UCB-based reinforcement learning approach for optimizing client selection to improve

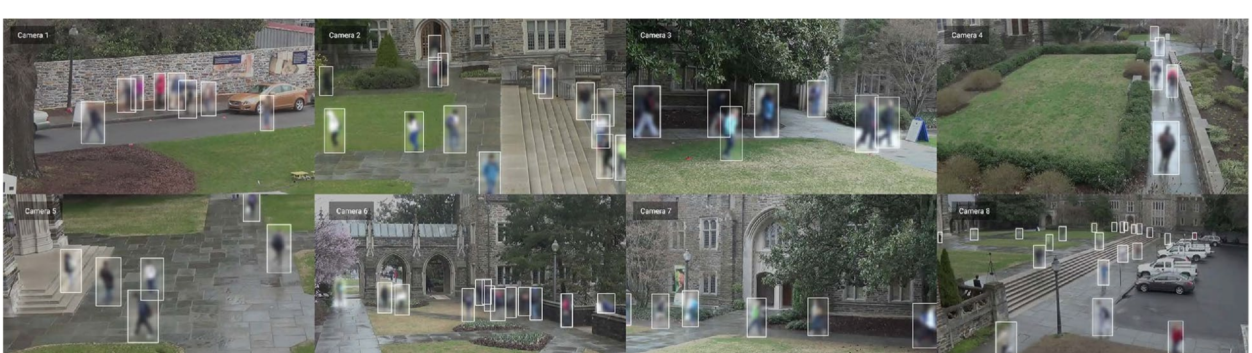

Fig. 1 A collection of images of 2,000 students in DukeMTMC [7] dataset, which was captured in their daily routines without their knowledge. This highlights the growing concern about the use of surveillance technology in public spaces

the efficiency and scalability of the federated Re-ID system.
- Evaluate the proposed framework on benchmark Re-ID datasets and compare its performance with existing state-of-the-art methods.

The rest of the paper is structured as follows: Related work section reviews related work on person re-identification and federated learning, Methods section presents the proposed federated Re-ID framework, Experiment section discusses the results and provides a comparative analysis with baseline methods and describes the experimental setup, including the datasets used and evaluation metrics, and conclusion and future work in Conclusion & future work section.

Related work

Person re-identification methods

People re-identification is an image retrieval task aiming to match individuals across non-overlapping camera views in varying environment conditions. The advancements in deep learning have enabled the extraction of more discriminative features from images, thus surpassing traditional hand-crafted feature operators [14–17]. Recent research has focused on addressing the challenges caused by varying camera angles and resolutions [18], clothing variations [19], illumination conditions [3, 20], background-bias [21], and occlusions [22–24]. However, many approaches still struggle to generalize across diverse real-world scenarios. Such limitations can reduce accuracy in dynamic environments where people may change clothing or move through differently lit areas. Attention mechanisms, triplet loss, and metric learning have been incorporated into CNN architectures to enhance feature learning [25–29]. These methods can be limited by how computationally intensive they are and may require fine-tuning to perform well in diverse environments. They also struggle with occlusions and significant appearance changes. Furthermore, the integration of multi-modal information, including gait and temporal sequences, has shown promise in improving re-identification accuracy. However, these approaches require complex data acquisition and are sensitive to environmental changes, such as variations in walking speed or camera placement. Additionally, multi-modal methods may increase computational load thus making real-time application challenging in resource-constrained settings. Synthetic data augmentation is a technique used to address the scarcity of labeled data. Generative Adversarial Networks (GANs) can generate realistic pedestrian images with diverse variations, expanding the training dataset for better feature extraction capabilities of the CNN [30–34]. While these models can produce varied appearances but fail to produce realistic environmental conditions. This can introduce bias which can reduce model performance and lead to overfitting on unrealistic features rather than genuine identity characteristics. While this work offers valuable insights into improving re-id tasks, it focuses on centralized setups where data is usually IID and available in a central data store. Several approaches were taken to protect the images from possible attacks [35, 36].

Federated learning for re-identification

Federated Learning has emerged as a promising solution for the privacy concerns related to computer vision tasks. It offers a decentralized approach to collaboratively train a deep learning model and has proved its effectiveness within a variety of tasks [37–40] and within the realm of Internet of Things (IoT) [41–43]. Recent work has explored the integration of federated learning in people re-identification where the data is inherently heterogeneous [44, 45] and the model architecture needed can differ [45]. There is a significant variation in the quality, quantity, and demographics of the data used in people re-identification. This in turn makes this approach susceptible to challenges with client heterogeneity and with varying device capabilities and data distributions that can lead to inconsistencies in model performance. Additionally, Federated learning offers privacy guarantees for the use of people's images in training, instead of centralized approaches which can be invasive. FedReID [45] introduces a decentralized training paradigm and proposes three optimization strategies aimed at enhancing model performance under statistical heterogeneity. Weng et al [46] address the challenges related to data annotation requirements and statistical heterogeneity across clients using a Coarse-to-Fine Three-Stage Strategy by establishing generic knowledge, integrating specialized knowledge, and enhancing fine-grained person representation through patch-level feature alignment, however, it is prone to domain overfitting. As a result, when deployed in a new domain with different camera angles, backgrounds, or lighting conditions, the model may fail to accurately identify individuals, as it cannot adapt to features outside of its training environment. Li et al [47] proposed an approach using domain alignment and data manipulation to address the domain shift, however, the communication efficiency and computational overhead increased significantly. Fed4ReID [48] address the Non-IID problem by integrating data augmentation techniques in edge computing environments, which helps improve model generalization across heterogeneous data.

Others [49, 50] tried to simulate unseen data scenarios and use style transformation models to enable the model to adapt to diverse conditions. However, these models often struggle to preserve identity-relevant features while transforming styles, leading to potential mismatches in real-world scenarios. Domain and Feature Hallucinating (DFH) [51], on the other hand, improves person re-identification model generalizability by generating diverse and representative features using domain-level feature statistics (DFS). Zhou et al [52] address feature shift in federated learning by augmenting the statistics (mean and standard deviation) of latent features. This can cause an increase in computational cost and varying synthetic feature quality as well as the risk of feature drift where the model becomes more attuned to hallucinated variations than actual domain-specific features. FRM [53] integrates MOON (Meta-Optimization of Optimization Networks) and a lightweight network (MB-CDNet) and provides a solution for privacy and communication issues in federated learning.

Enhancing federated learning performance

Federated Learning performance is often hindered by challenges such as statistical heterogeneity, communication inefficiencies, and model convergence issues. Several Techniques have been proposed to address these limitations including advanced aggregation strategies (e.g., FedProx, FedNova, and FedPav), knowledge distillation, and contrastive learning. Additionally, feature attention mechanisms have been explored to enhance feature representation quality and thus improve federated learning effectiveness in vision-based tasks. Recent studies, such as [45], highlight the role of adaptive weighting in aggregation to balance client contributions which improves the global model's robustness and accuracy. To improve federated learning performance, recent research has introduced adaptive model fusion techniques that dynamically assign weights to client updates based on model quality rather than sample size. Methods like FedDRL [54] use reinforcement learning to optimize client selection and aggregation to prevent the inclusion of malicious models and ensure higher global model reliability and accuracy. Moreover, adaptive federated learning strategies have been proposed to optimize the performance of ReID models by adjusting the learning process and choosing a suitable model based on the characteristics of the data and the network conditions [55–60]. While this approach can improve model personalization and efficiency, model selection strategies must be carefully designed to avoid overfitting to specific clients and ensure generalization across diverse environments. Communication efficiency is important in any federated learning setting to reduce the amount of data transmitted between clients and the central server, thereby minimizing communication overhead and improving scalability [61, 62].

Client selection strategies in federated learning

Effective client selection is essential in federated learning to ensure efficient training, reduce bias, and improve model convergence. Traditional methods often use random or uniform selection. This may lead to suboptimal model updates due to varying client data quality, computational resources, and network stability. To address these challenges, researchers have explored adaptive client selection strategies that prioritize high-quality and diverse data contributors while balancing participation fairness. Reinforcement learning-based approaches dynamically adjust client participation to optimize learning efficiency, whereas trust-aware mechanisms mitigate the risk of unreliable updates from low-performing clients. Client selection strategies can optimize communication efficiency by selecting a subset of clients to participate in each training round, based on factors such as data quality, computational resources, and network connectivity. These strategies can be based on reinforcement learning [63–67], or others that focus on the data quality itself [68–70]. However, this approach can be challenging due to client heterogeneity, where varying data distributions or resource capabilities may lead to imbalanced training. Additionally, improper client selection can result in slower convergence or reduced model generalization. Balancing exploration and exploitation in client selection remains a key challenge for achieving optimal performance. Other approaches have focused on transfer learning techniques to improve model performance [71]. Traditional federated learning assumes all clients are trustworthy, but real-world scenarios involve threats such as model poisoning and free-rider behaviors. Recent approaches use secure aggregation mechanisms, blockchain-based trust frameworks, and client contribution evaluation strategies to mitigate these risks, ensuring fair participation in heterogeneous environments [72]. However, they can be limited by domain mismatch, where the pre-trained model's features may not fully align with the target domain which leads to suboptimal performance. Additionally, some researchers have proposed novel model architectures specifically designed for federated learning, aiming to optimize training efficiency and communication costs [48, 53, 73–75].

Discussion

We aim to leverage these methods as a foundation for our FL approach. Since privacy is a crucial concern when dealing with personal images. We aim to leverage these methods as a foundation for our FL approach. Since privacy is a crucial concern when dealing with personal images. Currently, the state of the art in federated learning for ReID includes several approaches as discussed above. The current Federated ReID benchmark does not address the challenge of adapting to unseen domains, such as varying lighting, camera angles, or occlusions, which are common in real-world scenarios. Moreover, FedReID lacks a dynamic client selection mechanism that could improve performance based on client capabilities and data characteristics. On the other hand, other notable approaches such as Fed4ReID are limited to a baseline comparison on public datasets, without exploring domain adaptation or advanced feature extraction techniques. Another approach is The FRM framework, which relies on private datasets limits its applicability, and does not offer insights into domain adaptation or generalization across publicly available datasets. Our work advances federated learning for person re-identification by introducing several key contributions. We leverage generative learning techniques for data augmentation, enabling the model to learn robust, generalizable features in scenarios with limited real-world data. Unlike augmentation in Fed4ReID, our approach synthesizes diverse, realistic data to enhance model robustness across varied conditions. We also incorporate feature attention fusion to improve feature extraction and adapt to domain shifts like occlusions and background changes. Additionally, a reinforcement learning-based weighted aggregation strategy reduces bias in model updates, while adaptive client model selection optimizes resource use and performance across clients.

Methods

This section outlines the proposed framework by detailing the client-side re-identification model, the federated learning architecture, and the server-side mechanisms for adaptive client selection, and model aggregation. First, the problem formulation is defined and then an overall system architecture is presented to show the interactions between local models and the global server. Next, we describe the Client-Side Person Re-Identification Model with focus on its architecture, feature extractors, data preprocessing techniques that includes GAN-based synthetic augmentation and feature attention mechanisms. On the server side, we explain the Contribution-Driven Weighted Upper Confidence Bound (CDW-UCB) strategy taken for fair contribution evaluation and an adaptive model selection based on client hardware capabilities. We further refine the aggregation process by selecting only key model components for update rather than performing full model aggregation to mitigate the impact of bias and poor-performing clients.

Problem definition

In this section, we define the challenges associated with traditional federated learning algorithms in the context of person re-identification. We assume a federated learning system comprising *n* clients, where the *i*-th client is denoted as C_i, and the set of all clients is represented as $\{C_1, C_2, \ldots, C_n\}$. Each client C_i trains a local model with parameters θ_i, and the set of all client model parameters is denoted as $\{\theta_1, \theta_2, \ldots, \theta_n\}$. During each communication round, the central server selects *k* clients from the *n* available clients. The selected clients upload their locally trained models to the server, which then aggregates these models to form a global model θ_{global}. The server's primary objective is to iteratively optimize the global model, as shown in Eq. 1:

$$\min_{\theta_{\text{global}}} G(\theta_{\text{global}}) = \sum_{i=1}^{k} w_i f(\theta_i), \quad \sum_{i=1}^{k} w_i = 1, \quad w_i \geq 0 \quad (1)$$

Let w_i represent the weight assigned to the *i*-th client during aggregation. Traditional Federated Learning algorithms assume that the models uploaded by the selected clients are identically aggregated and identically sent back, which is not fair to either bad-performing clients or good-performing clients with less data. This creates a bias issue both at the server and at the client level. Specifically, the models are aggregated in a way that may not accurately represent the performance of each client. If the server fails to account for this, the global model may be compromised, as clients with poor performance may undeservedly influence the global model. The primary challenge is to ensure fairness during the aggregation process and reduce bias and maintain an optimal global model despite these inequalities.

The next challenge is to objectively evaluate the contribution of each client. This evaluation is critical for determining the weight of each client's contribution to the global model, as well as to itself, in order to balance the bias.

One critical challenge in federated learning (FL) systems is ensuring that the system adapts to the varying hardware capabilities and performance levels of clients during the aggregation process. Clients in a federated learning setup can have different hardware configurations, which affects their computational capacity, memory, and speed. Consequently, clients may perform differently during local model training, impacting the quality of the models they contribute to the global aggregation.

To formalize this challenge, we assume a federated learning system comprising *n* clients, where the *i*-th client is denoted as C_i, and the set of all clients is represented as $C_1, C_2, \ldots, C_n$. Each client C_i trains a local model with parameters θ_i, and the set of all client model parameters is denoted as $\theta_1, \theta_2, \ldots, \theta_n$.

During each communication round, the central server selects *k* clients from the *n* available clients, where the clients' local model performance varies based on hardware capabilities. The performance of each client, denoted as p_i, is computed based on factors such as the number of epochs trained, model convergence, and hardware limitations (e.g., CPU/GPU processing power). The weight assigned to each client during the aggregation, w_i, is adjusted according to both the client's performance and hardware capacity. Specifically, the weight can be represented as:

$$w_i = \frac{p_i \cdot h_i}{\sum_{i=1}^{k} p_i \cdot h_i} \tag{2}$$

Where: p_i is the performance score of client C_i, which is based on local model performance metrics (e.g., loss, accuracy). h_i is the hardware capability score of client C_i, which quantifies the client's hardware performance (e.g., CPU, GPU, memory). This ensures that clients with higher hardware capabilities and better performance receive higher weights during model aggregation, thus compensating for hardware disparities and improving the fairness of the model updates.

In each communication round, the server aggregates the models uploaded by the selected clients, forming the global model θ_{global} based on the weighted sum of local model parameters:

$$\theta_{\text{global}} = \sum_{i=1}^{k} w_i \theta_i \tag{3}$$

The primary objective of this adaptive aggregation process is to balance the influence of each client based on both their local performance and hardware capabilities. By doing so, we ensure that clients with more computational power or better performance contribute more significantly to the global model without unfairly biasing the model aggregation process. Thus, the challenge is to design an efficient weighting mechanism that accounts for the differences in client performance and hardware, ensuring a fair and robust global model aggregation while maintaining the benefits of decentralized learning. A further challenge in traditional federated learning is that all model parameters are aggregated uniformly without considering the varying contributions of different model components. This approach can be inefficient and unfair as not all parts of the model contribute equally to the learning process. Some feature extraction layers may be more critical for person re-identification, while others may introduce noise or unnecessary redundancy which leads to suboptimal model updates.

To formalize this challenge, we consider a federated learning system with *n* clients, where each client C_i trains a local model θ_i. The local model θ_i consists of two major components:

The first is the feature extractor F_i, responsible for extracting meaningful representations from input data. The second component is the classifier head H_i, responsible for mapping extracted features to predictions. Thus, the model can be represented as:

$$\theta_i = (F_i, H_i) \tag{4}$$

Traditional federated learning aggregates the entire model:

$$\theta_{\text{global}} = \sum_{i=1}^{k} w_i \theta_i \tag{5}$$

where w_i is the weight assigned to client C_i based on its contribution. However, this full aggregation does not account for the fact that different clients may have varying feature extractors due to differences in data distributions, hardware capabilities, and local training dynamics. Aggregating all parameters indiscriminately can lead to inefficiencies as some layers may generalize poorly across clients.

A key challenge in federated person re-identification is improving the discriminative power of extracted features while mitigating the effects of data heterogeneity across clients. In summary, the three key challenges are as follows: (1) How to balance the bias caused by clients' contributions to the global model during the aggregation process based on performance and data availability? (2)

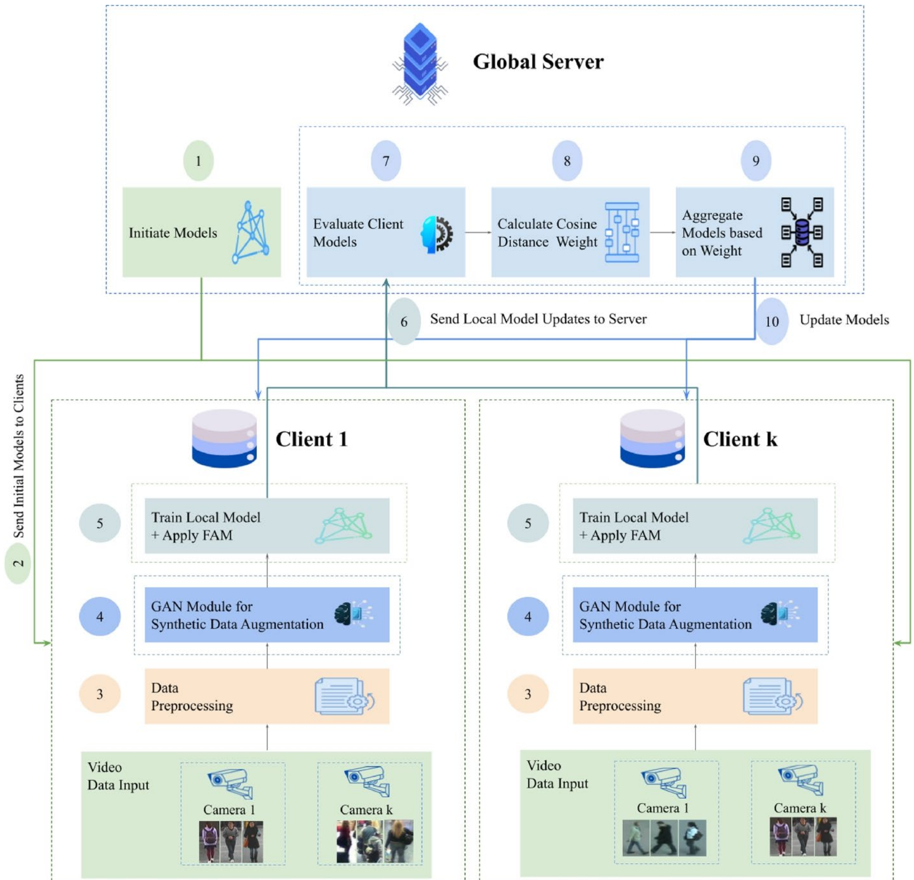

Fig. 2 Overview of the proposed system workflow

How to reduce the bias in model aggregation caused by imbalanced data or hardware capabilities of clients? (3) How can the feature extraction model's performance be enhanced using feature attention mechanisms and synthetic data augmentation to ensure a more robust global model?

Architecture design

The Federated setup assumed in this work can be defined as a federated-by-dataset scenario, or to put it differently, it is a client-edge-cloud architecture. In this setup, the edge servers act as the clients, where each edge server gathers data from a different camera in its domain and constructs its own data repository to train on. The central server manages the global models and is responsible for assigning, receiving updates from, and updating client models, aggregating the client updates into the global server, and acting as the teacher model in the knowledge distillation process between server and client (Fig. 2).

This scenario follows the existence of an edge server that has higher computational resources to train the models more effectively and thus is deemed to be more suitable and practical. In a different federated setting like federated-by-dataset, the edge devices may act as the client instead and use a client-edge-cloud architecture. Although our approach follows the federated-by-dataset setup, where each dataset is considered a client, as it is deemed more suitable, our methodology is theoretically applicable to both scenarios. In both scenarios, the data remains on the client's devices, inherently ensuring privacy preservation.

Require: Number of clients n, communication rounds T, global model θ_{global}
Server initializes θ_{global} and selects feature extractor based on client hardware.
for each communication round $t = 1, 2, \ldots, T$ **do**
  Server distributes θ_{global} to selected clients.
  for each client C_i in parallel **do**
    Client Pre-Processing: Normalize and transform local dataset.
    Data Augmentation: Use GAN-based augmentation on dataset.
    Local Training: Update model θ_i using local dataset.
    Send updated model θ_i to the server.
  end for
  Server Evaluates received models.
  Compute client contribution scores:

$$\text{reward}_i = \frac{1 - \text{cosine\_similarity}(\theta_{\text{global}}, \theta_i)}{\sum_{i=1}^{k} (1 - \text{cosine\_similarity}(\theta_{\text{global}}, \theta_i))}$$

  Global Aggregation: Perform partial model aggregation:

$$\theta_{\text{global}}^{t+1} = \sum_{i=1}^{k} w_i \cdot \theta_i$$

where w_i is based on client contribution.
  Update client models with new $\theta_{\text{global}}^{t+1}$.
end for
Output: Final global model θ_{global}.

Algorithm 1 Federated learning workflow for person re-identification

The system operates in an iterative manner, following the steps outlined below.

Step 1: model initialization at the server

The global model architectures are initialized at the central server. The server selects the appropriate backbone for the model based on the hardware constraints of participating clients. The choice of feature extractor, such as ResNet or ConvNeXt, is taken at this stage.

Step 2: send model to clients

The initialized models are assigned and transmitted to the participating clients. Each client receives a model copy that will be trained on its local dataset.

Step 3: pre-processing at the client side

Each client prepares its dataset by applying data normalization and transformations. Resizing, cropping, and pixel intensity normalization are applied based on the model requirements.

Step 4: data augmentation using GANs

A Generative Adversarial Network (GAN) is used to generate synthetic training samples. This improves model robustness by diversifying the dataset, especially for underrepresented classes.

Step 5: local training at clients

Each client trains its assigned model on its private dataset. The models are trained locally.

Step 6: model transmission to server

After training, the locally updated model parameters θ_i are sent back to the server. Clients only transmit weights and gradients, maintaining data privacy.

Step 7: model evaluation at the server

The server evaluates the received models based on performance metrics such as retrieval accuracy. A weighted aggregation strategy assigns contribution scores based on cosine similarity between client and global models. The best-performing models are given higher weights in the aggregation step.

Step 8: client contribution scoring

The server calculates a reward score reward_i for each client based on:

$$\mathrm{reward}_i = \frac{1 - \mathrm{cosine\_similarity}(\theta_{\mathrm{global}}, \theta_i)}{\sum_{i=1}^{k} (1 - \mathrm{cosine\_similarity}(\theta_{\mathrm{global}}, \theta_i))} \quad (6)$$

The best-performing model receives a score of 1, while the other models are down-weighted accordingly.

Step 9: global model aggregation

The server aggregates the received models using a partial aggregation approach. Instead of fully aggregating entire models, only the feature extractor layers are updated. The aggregation strategy is discussed in Weighted partial global model aggregation section.

Step 10: client model updates

The aggregated model is weighted and redistributed to all participating clients for the next training iteration.

Client-side people re-identification model

This section describes a detailed account of the approach taken to build, train, and evaluate our model for people reidentification using federated learning. This section outlines the methods used for data pre-processing and model architecture.

Local model architecture and feature extractors

An Instance Discriminative Embedding (IDE) model is a deep learning model trained to generate a feature representation for each individual uniquely. Decomposing the term, the first part is 'Instance' which refers to a specific instance of a person in an image. The second part 'Discriminative' implies the effectiveness of the captured embedding feature in representing the person and in turn in distinguishing between different people. 'Embedding' highlights the important characteristics of the compressed representation of the person's image. Therefore, an IDE model aims to transform an image of a person into a numerical representation that effectively captures the unique visual details of that specific person and is used to transform an image of a person into a numerical representation that can capture the unique visual details of that specific person, enabling the system to match an image of a person captured by one camera with an image from another camera.

This section describes the deep learning model used for federated learning. The Instance Discriminative Embedding (IDE) model, focuses on creating discriminative representations of people in images thus enabling the system to distinguish between individuals. We decompose the IDE model into five components: the backbone network, followed by the feature attention mechanism, followed by pooling and embedding layers, ending with the loss function as shown in Fig. 3.

Backbone network The backbone network serves as the foundation of our IDE model, learning a general representation of the input image data of a person. The complexity of the backbone network architecture significantly impacts the overall performance and efficiency of the model, as more powerful backbones might extract more complex features, but they may also require higher computational resources. Another important aspect is data diversity, Clients may have varying data types and quantities. A backbone optimized for a specified type may not perform well in others.

In this work, we explore two prominent choices: the classic ResNet [76] and the recently introduced ConvNeXt [77] architecture, offered in large and tiny versions

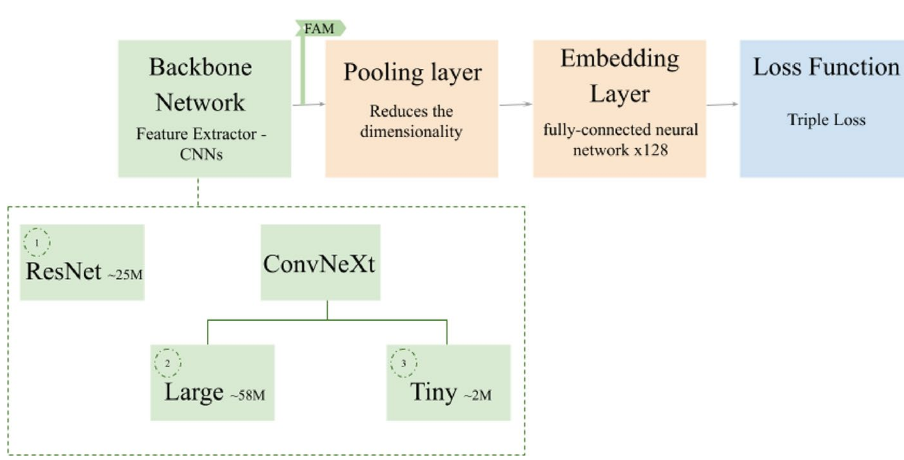

Fig. 3 Overview of the identity discriminator embedding model

to address the trade-off between accuracy and computational complexity. The pre-trained ResNet provides a strong foundation, while ConvNeXt offers new potential as shown in the comparison in Table 1. Thus we aim to evaluate the performance of both architectures for our specific federated people re-identification system.

ResNet introduced residual connections to directly add the unmodified input to the output of a convolutional block thus allowing the model to learn residual functions and mitigate the vanishing gradient problem. The ResNet block consists of Batch normalization which is used to normalize the activations of the previous layers, followed by the first convolution layer which does the feature extraction, then the output is passed through an activation function to introduce non-linearity, and followed by a second convolution layer. Its feature extraction capabilities as well as mitigating the vanishing gradient problem make it a reliable choice.

ConvNeXt represents an alternative to ResNet represents a state-of-the-art approach, achieving remarkable performance on image classification tasks. ConvNeXt relies entirely on convolutional layers but departs from the standard ResNet architecture by incorporating several key design differences. It uses depthwise separable convolutions, which break down the convolution step into two steps, a depthwise convolution for feature extraction and a pointwise convolution for channel-wise combination. This step reduces computational costs while maintaining feature quality. ConvNeXt comes in both large and tiny versions. A more concise comparison is shown in Table 2.

Data preprocessing and augmentation

The quality of a model's output is inherently linked to the quality of its input data. In essence, models are susceptible to the 'garbage in, garbage out' principle, where suboptimal or noisy data can significantly affect the accuracy and reliability of the generated results. Data augmentation and transformations are essential tools in the context of computer vision tasks, particularly for tasks like people re-identification. By synthetically expanding the training dataset through different manipulations, these techniques significantly enhance model performance and generalization.

Table 2 ConvNeXt large and tiny comparison table

Feature	ConvNeXt Large	ConvNeXt Tiny
Model Size	Larger	Smaller
Parameter Count	Higher 5̃3M	Lower 2̃M
Computational Cost	Higher	Lower
Potential Accuracy	Higher	Lower
Suitability	less suitable for resource-constrained devices	More suitable for resource-constrained devices

Data transformations techniques for data augmentations that rely on various geometric and photometric transformations to existing images are applied to the images as shown in Fig. 4. These transformations include random cropping to extract random patches from the images to mitigate occlusion, flipping, rotation, color jittering to introduce variations in brightness, contrast, and saturation in images, resizing to 256x128 for spatial integrity, and finally images were then converted into a tensor format for faster computation in the Pytorch library.

Data augmentation techniques that rely on Generative Adversarial Network (GAN) are used. We used an architecture that combines generative and discriminative learning to generate synthetic images of people as shown in Fig. 5. We aim to overcome the challenge of addressing data scarcity issues and improve model performance by addressing significant intra-class variations across different cameras. An example of the original and generated images is shown in Fig. 5. This produces an *nxn* number of images where *n* is the original number of data samples we have, increasing the size and diversity of training datasets. The general overview of the network explained in Fig. 6, consists of the generative and discriminative modules.

In Fig. 6, the figure on the left illustrates the GAN is incorporated into our work, the first segment is for unsupervised learning by leveraging the adversarial relationship between the generator and the discriminator. Additionally, the input data is fed to the backbone to train on and the generated images are used to fine-tune the backbone to improve its feature extraction capabilities. The figure on the right shows a visual representation of the generator module in the GAN. The generator

Table 1 CNN backbone comparison table

Feature	ResNet	ConvNeXt
Architecture Type	Convolutional Neural Network (CNN)	Convolutional Neural Network (CNN)
Strengths	A reliable architecture equipped to handle the vanishing gradient problem, making it a widely adopted choice	A state-of-the-art architecture with the potential to surpass current accuracy benchmarks
Weaknesses	May not be as efficient as newer architectures	Requires more training data and computational resources compared to ResNet.
Pre-trained Availability	Widely available	Less readily available, potentially requires further training

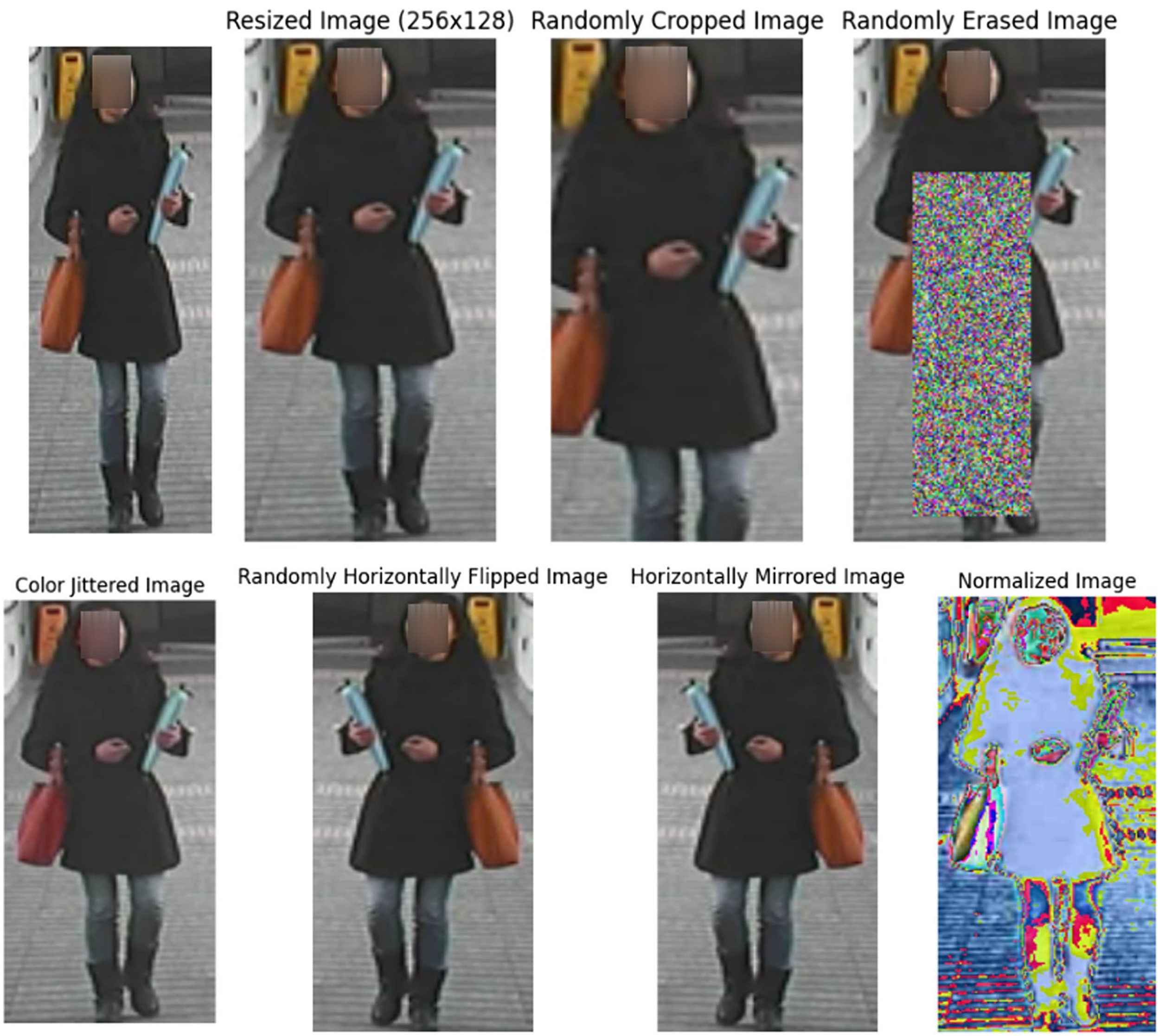

Fig. 4 A visual representation of data transformations used to augment an original image. The transformations include resizing, random cropping, random erasing, color jittering, random horizontal flipping, random mirrored imaging, and normalization

Fig. 5 Examples of synthetic images created by swapping codes of the original images

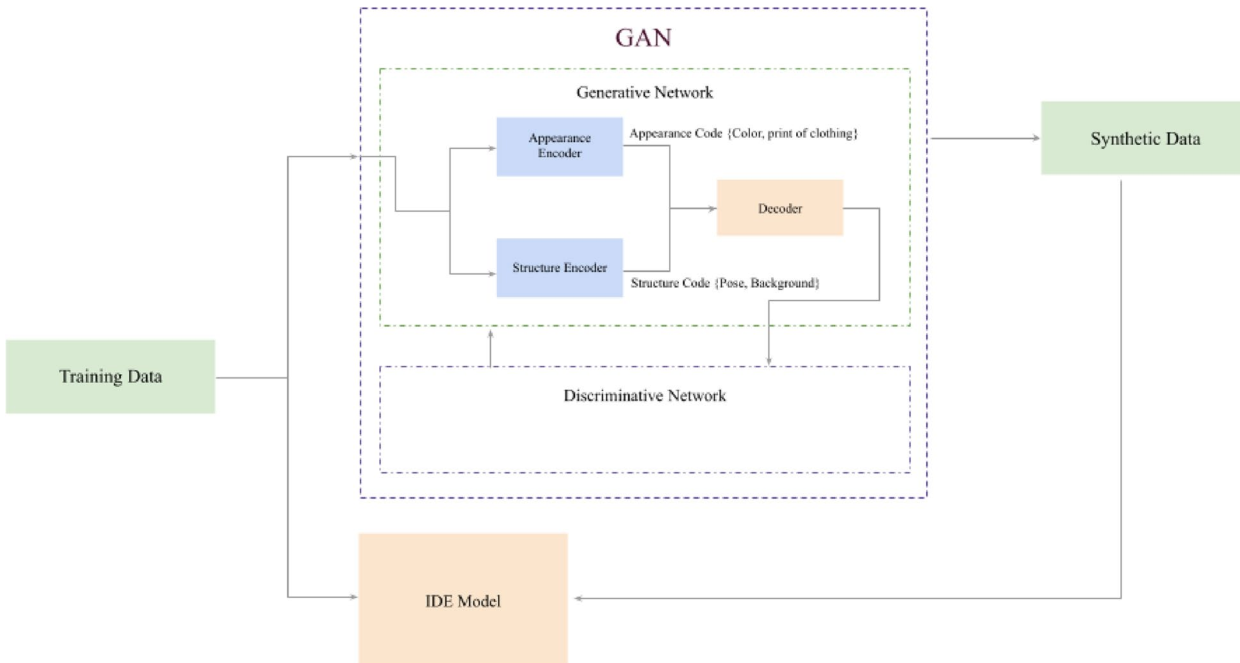

Fig. 6 Overview of the GAN used for data augmentation

uses an encoder-decoder approach, where the structure and appearance encoders extract features from the input data and the decoder synthesizes a new image based on these features. The GAN within the DG-Net framework does not require explicit pre-training, however, the appearance encoder is pre-trained on the Market dataset and fine-tuned on the other dataset at the edge. This encoder is based on ResNet50 and is initially trained on ImageNet. The structure encoder, decoder, and discriminator are trained jointly. The validity of synthetic data is ensured through joint learning, where images are decomposed into appearance and structure codes to enable cross-identity generation that refines the appearance encoder. While no explicit filtering is applied, the framework's design-including self-identity and cross-identity generation with dynamic soft labeling-ensures high-quality, realistic synthetic images.

Feature attention mechanism

Attentional Feature Fusion is used to introduce a dynamic fusion, where it calculates weights based on the input feature maps, to determine the contribution of each feature map to the final fused output. Multi-Scale Channel Attention is used to incorporate contextual information from various scales for a better understanding of the relationships between features. Point-wise convolution is used, where a 1x1 kernel is applied to each channel of the feature map. This improves the local feature representation by capturing dependencies within a channel without altering the spatial resolution. It further analyzes the image at different scales by using dilation at different rates, similar to zooming in and out of the image to try and understand the bigger picture. This technique expands the receptive field of the filters without using different filters and thus fewer parameters. The output of the convolutions at different dilation rates is then concatenated, which is passed into an average pooling layer to reduce the spatial dimensions. The output is then used to compute the attention weights.

The attention weights are then multiplied with the original feature maps, which allows features with higher weights to have a greater influence on the final fused feature map and features with lower weights will be suppressed.

The resulting feature map is fed into a standard attention module. The module consists of two fully-connected layers with non-linear ReLU activation functions, to learn a non-linear relationship between the channels. The output is passed into a Sigmoid activation function to ensure the output weights range between 0 and 1.

This final fused feature map as shown in Fig. 7, which contains information from different scales and is weighted according to importance, is then fed into the subsequent CNN layers. This process refines the feature maps to focus on more identifying features while suppressing less relevant ones.

Optimization and loss functions

To improve training, the loss function used is a combination of two separate functions, each measuring different aspects of the model's performance. The first of which is Cross-Entropy Loss [78] (CEL) shown in Eq. 7, which is

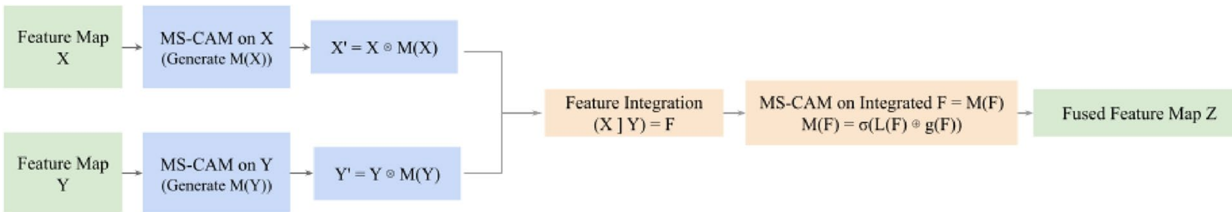

Fig. 7 Process flow of the feature attention module AFF with MS-CAM

the summation of the true labels y_{true} multiplied by the log of the predicted labels y_{pred} over all data points in a batch. CEL is a common loss function usually used for classification tasks. During training, the model receives an image and predicts an identity label (e.g., person A, person B). The CEL compares this prediction with the true identity label and helps the model to better distinguish between different people.

The second component is the triplet loss [79] shown in Eq. 8, The squared Euclidean distance is calculated between the anchor-positive and anchor-negative feature vectors with a margin that defines the minimum desired distance between the negative and anchor images compared to the positive and anchor images. It is a contrastive learning technique generally used in metric learning to learn relative distances between image embeddings in a feature space. In triplet loss, the goal is to minimize the distance between the anchor and the positive image, as in images of the same person should be close, while trying to maximize the distance between the anchor and the negative image in the feature space, where in images of different people should be further apart.

By combing the loss function into a compound loss function as shown in Eq. 9, the model learns to correctly classify identities using CEL and map features of the same identity closer together in the feature space.

$$\mathrm{CEL} = -\sum (y_{\mathrm{true}} \cdot \log(y_{\mathrm{pred}})) \tag{7}$$

$$\begin{aligned}\text{Triplet Loss} = \sum \max(0, \text{margin} - \|f(\text{Anchor}) \\ - f(\text{Positive})\|^2 + \|f(\text{Anchor}) \\ - f(\text{Negative})\|^2)\end{aligned} \tag{8}$$

$$\text{Loss} = \lambda \cdot \mathrm{CEL} + (1 - \lambda) \cdot \text{Triplet Loss} \tag{9}$$

Federated learning server-side operations

The central server plays a crucial role in coordinating federated learning by selecting clients, aggregating local models, and optimizing the global model.

Client selection strategy

Client selection is a crucial factor in federated learning systems. Not all devices are equal - some may possess more powerful hardware or have datasets that align better with the current learning objective and thus generalize better. To address this heterogeneity, we used the Upper Confidence Bound algorithm as shown in Fig. 8, an optimistic approach taken from reinforcement learning. UCB helps us identify the n clients that are the most likely to contribute valuable updates in each training round.

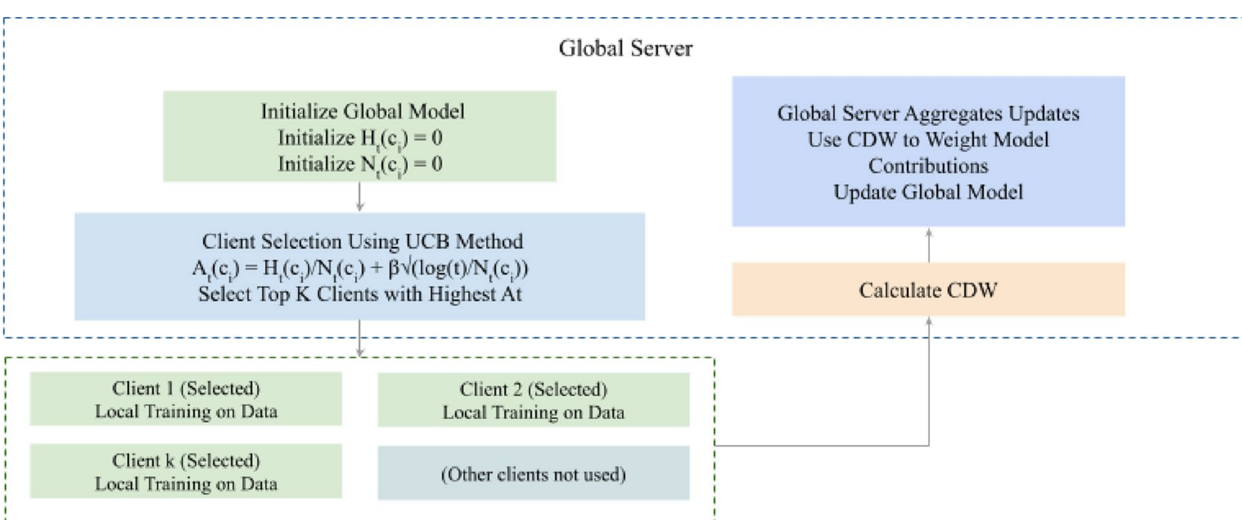

Fig. 8 Dynamic weight allocation based on client performance and upper confidence bounds algorithm

The weight is calculated based on the cosine distance between the client model and the server model. Then, the clients are given a reward based on their performance: 1 for the best model and 0 for the rest. Assuming the federated task involves *m* communication rounds, the contribution of each client is evaluated in each round. This process is formalized in the following equations. The weight for each client C_i at round *t* is calculated based on the cosine distance between the client model θ_i and the server model θ_{global}, and is given by:

$$w_i = \frac{1 - \cos(\theta_i, \theta_{\text{global}})}{\sum_{i=1}^{k} (1 - \cos(\theta_i, \theta_{\text{global}}))}, \quad (10)$$

where $\cos(\theta_i, \theta_{\text{global}})$ represents the cosine similarity between the client model θ_i and the server model θ_{global} , and w_i is the weight for client C_i. Next, the reward for each client is distributed based on its weight, where the best-performing model receives a reward of 1 and the others receive 0. This can be expressed as:

$$r_i = \begin{cases} 1 & \text{if } i = \arg\max_j w_j, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Thus, the reward function ensures that the best-performing client is rewarded with 1, while the others receive 0. The weight distribution for all clients in round *t* is calculated as follows:

$$\{C_1, C_2, \ldots, C_k\} \leftarrow \text{Weight}(\{\theta_1, \theta_2, \ldots, \theta_k\}), \quad (12)$$

where $\{C_1, C_2, \ldots, C_k\}$ represents the set of selected clients in round *t*, and $\{\theta_1, \theta_2, \ldots, \theta_k\}$ are the corresponding model parameters of these clients. Each communication round *t* is denoted as $t \in [1, m]$, where *m* is the total number of rounds.

The algorithm builds on the idea of assigning rewards to clients based on the client's model alignment with the global model. This alignment is calculated by measuring the cosine distance weight [80] (CDW), between the client's model and the global server-side model. The client with the highest calculated cosine distance weight recives a positive reward (+1), while others get 0. Initially, participating clients are chosen at random to leverage exploration, allowing the system to explore different possibilities. As the training process progresses, UCB uses the accumulated rewards to prioritize clients with a history of providing more useful updates.

Dynamic Weight Adjustment is done by calculation CDW, as shown in Eq. 13. CDW is calculated and used to dynamically allocate weights during model aggregation. The weights are calculated based on the cosine distance between client updates and the global model.

$$m_k = \frac{\sum_{i=1}^{d} w_{k,i} \cdot w_{g,i}}{\|w_k\| \cdot \|w_g\|} \quad (13)$$

CDW in Eq. 13 where m_k represents the aggregated model of the global model update, P_d is the number of participating clients in the current round, w_k, i and w_k are the model weights of the $i-th$ client and norm the client's model weights respectively, and $w_g.i$ and w_g are the global model weights at the beginning of the current round and norm of the global model weights.

Adaptive model selection strategy

To improve the fairness and performance of the federated learning system, An adaptive model selection and swapping strategy is implemented. This approach dynamically selects the most suitable feature extractor model based on each client's performance reward to ensure better alignment with client hardware capabilities. The feature extractor is chosen from three candidate models-ResNet, ConvNeXt Large, and ConvNeXt Tiny-based on the hardware constraints and then based on the reward accumulation of the client's contribution, as computed in Eq. 10. Clients with lower hardware capabilities are assigned smaller models such as ConvNeXt Tiny, while clients with more powerful hardware are assigned larger models such as ConvNeXt Large. This allows the system to adaptively balance the client contributions while optimizing the global model performance. During each communication round, the central server selects *k* clients from the *n* available clients. The selected clients upload their locally trained models to the server, which then aggregates these models to form a global model θ_{global}. The server's primary objective is to iteratively optimize the global model.

Weighted partial global model aggregation

The chosen federated aggregation algorithm is Federated Partial Averaging (FedPav) instead of traditional Federated Averaging (FedAvg). In a Non-Identically Distributed (non-IID) setting, where data distributions across participating clients are significantly different, the clients' local models can diverge significantly. Averaging these models during aggregation using FedAvg can slow down convergence or even harm performance. FedPav addresses this model discrepancy using partial averaging, where only a specific subset of model weights is averaged instead of all model weights during aggregation. This subset represents the core of the model, in our case the backbone feature extractor, while the rest of the model or in other terms the client-specific parts of the model remain unchanged. Which leads to faster convergence, reduced model discrepancy, improved model performance, and requires minimal modification to existing FL algorithms. FedPav enables personalized performance on the client side as user data is used for training and allows the server side to capture common patterns. Additionally, weighted averaging assigns a weight to each update received from clients based on the calculated cosine distance weight. This helps assess how well a client's local model aligns with the current global model on the server. The rationale behind this is to focus on more informative clients, where clients with local models closer to the current global model are more likely to contribute more relevant updates. This helps mitigate the problem of imbalance participation, where clients with less representative data are prevented from skewing the learning process, leading to faster convergence.

This allows for more focused learning on the relevant features that each client can contribute based on their performance. The client models are divided into components, and the feature extractor part, denoted as θ_{fe}, is aggregated. The global feature extractor model $\theta_{\text{fe, global}}$ is then computed as:

$$\theta_{\text{fe, global}} = \sum_{i=1}^{k} w_i \theta_{\text{fe},i}, \quad \sum_{i=1}^{k} w_i = 1, \quad w_i \geq 0 \tag{14}$$

Where $\theta_{\text{fe},i}$ represents the feature extractor of client C_i. This ensures that the aggregation focuses on the components that contribute most to the task at hand, improving both the accuracy and efficiency of the global model, while also reducing bias caused by irrelevant components of the models.

Experiment

This section presents the research findings, including the federated model performance with the task of people re-identification across different settings.

Experimental setup

The programming language we used for running our experiments is Python on Google Colab platform. Google Colab provided a convenient and powerful platform for our experiments by offering a cloud-based Jupyter Notebook environment with pre-installed libraries and access to free GPU resources. The federated environment is simulated on colab where each dataset is assigned to a simulated client device. Each client device trained a local model on its assigned data partition. The local model is a copy of the original global model.

The experiments were conducted on Linux 21.2 with an Intel i5-10600KF CPU, 64 GB DRAM, and Tesla T4 GPU. The simulation was performed in a Python 3.10 environment. We experimented with ResNet-50 and ConvNeXt as backbone models, both models pre-trained on ImageNet dataset [81].

The federated learning system involves 9 clients, each represented by one dataset: Market-1501, CUHK03-NP, PRID2, CUHK01, CUHK02, VIPeR, 3DPeS, iLIDS, and GRID. The number of identities (classes) in each dataset varies, with Market-1501 having 751 identities, CUHK03-NP having 767, and 3DPeS having 93, among others. The total number of training images ranges from 632 in VIPeR to 12,936 in Market-1501 as shown in Table 3. The system uses a batch size of 32 and a learning rate of 0.0001. Clients train their models for 10 epochs per round, with communication rounds set to 20. The model architectures backbones are ResNet or ConvNeXt Large or ConvNeXt Tiny.

Datasets

In this section, we will discuss the datasets used in our experiments. A variety of person re-identification datasets were used. The nine datasets used are shown in Table 3. Datasets such as CAVIARa [82] and Imagery Library for Intelligent Detection Systems (iLIDS) [83] are taken in controlled indoor environments, while others such as Market and QMUL underGround Re-IDentification (GRID) [84] are captured in outdoor settings. Datasets Chinese University of Hong Kong (CUHK) [85–87] provide variations in viewpoint and pose. CUHK01 and CUHK02 are smaller-scale datasets with fewer identities and camera views, while CUHK03 is a larger-scale dataset with more identities and camera views, providing a more challenging and realistic benchmark for person re-identification. Datasets PRID [88] and Viewpoint Invariant Pedestrian Recognition (VIPeR) [89] have limited

Table 3 Overview of datasets used in the federated person re-identification experimental setup

Dataset	# Cameras	Train # Identities	Train # Images	Test # Query	Test # Gallery
Market-1501	6	751	12,936	750	19,732
CUHK03-NP	2	767	7,365	700	5,332
PRID2	2	285	3,744	100	649
CUHK01	2	485	1,940	486	972
CUHK02	2	971	3,107	971	776
VIPeR	2	316	632	316	316
3DPeS	2	93	450	86	316
iLIDS	2	59	248	60	130
GRID	8	250	4000	207	621

image pairs per identity. This dataset variation, shown in Table 4, is to demonstrate the model's performance under a range of conditions. The datasets were chosen to include different variances to simulate the statistical heterogeneity in real-world scenarios. They include a variety in the number of images, identities, camera views, demographics, image resolution, illumination, and scenes. The statistical heterogeneity makes training even more challenging centralized training where data is IID. The unbalanced data problem is represented in the disparity of data volumes and the domain discrepancies among datasets demonstrate the non-IID. This collection of datasets is done to ensure the model's generalizability and robustness in real-world applications.

Evaluation metrics

Model Evaluation Metrics such as Rank-K Accuracy and Mean Average Precision (mAP) were used to evaluate the model performance. Rank-k calculates the model's retrieval accuracy at specific rank positions (K). A high Rank-1 accuracy, shown in Eq. 17, indicates the model's effectiveness in identifying the most likely match for each query image. Rank-1 accuracy acts as the true positive rate and therefore indicates the misclassification rate (false positives). This is done to validate that our results are actually representative of correct classifications and not affected by dataset-specific characteristics. Rank-5 Accuracy, shown in Eq. 18, assesses whether the correct match falls within the top 5 retrieved images. Rank-10 Accuracy, shown in Eq. 19, extends the evaluation to the top 10 retrieved images. mAP, shown in Eq. 15,16, provides an overview of the model's performance across all query images by calculating the average precision (AP) at each rank position (top 1, top 5, etc.) and then averaging these values. A higher mAP means better overall performance, reflecting the model's ability to retrieve correct matches at various ranking positions.

$$AP = \sum_{k=1}^{n} (R(k) - R(k-1)) \cdot P(k) \tag{15}$$

$$mAP = \frac{1}{C} \sum_{c=1}^{C} AP_c \tag{16}$$

$$\text{Rank-1 Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \delta(r_i \leq 1) \tag{17}$$

$$\text{Rank-5 Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \delta(r_i \leq 5) \tag{18}$$

$$\text{Rank-10 Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \delta(r_i \leq 10) \tag{19}$$

$P(k)$ is the precision at the k-th threshold
$R(k)$ is the recall at the k-th threshold
n is the number of thresholds
C is the number of classes
AP_c is the Average Precision for class c
N is the total number of queries
r_i is the rank of the correct item for the i-th query
$\delta(\cdot)$ is the indicator function, which is 1 if the condition is true and 0 otherwise

Experiment results

We present the experimental evaluation of the proposed approach, compared with the benchmark and centralized training. We conduct these experiments with a batch size of 32 and one local epoch. The results presented in Table 5 showcase the performance of the global model on different re-identification datasets, evaluated using rank-1, rank-5, rank-10 accuracy, and mean average precision (mAP). The global model demonstrates strong performance on the Market-1501, 3DPeS, and iLIDS datasets, with Rank-1 accuracies of 88%, 86%, and 84.3%, respectively. The high Rank-1 values indicate the model's effectiveness in correctly identifying individuals on the first attempt. This is further supported by high mAP scores, especially on 3DPeS (80.0) and iLIDS (78.2), reflecting the model's retrieval capabilities across various queries.

Table 4 Overview of the demographic characteristics of individuals included within the datasets used for the federated person re-identification experiments

Dataset	Description	Place of Origin	Demographics
Market-1501	In front of a supermarket	Tsinghua University, China	University students and staff
CUHK03-NP	Campus cameras	Chinese University of Hong Kong	Diverse population of university students and staff.
PRID	Retail store	Austria	Diverse
CUHK01	Campus cameras	Chinese University of Hong Kong	University students and staff
CUHK02	Campus cameras	Chinese University of Hong Kong	University students and staff
VIPeR	Two outdoor cameras near a gate	California, Santa Cruz	University students and staff
3DPeS	Campus cameras	University of Surrey, UK	Diverse population of university students and staff.
iLIDS	Airport arrival hall	UK	Diverse, reflecting airport travelers
GRID	Cameras in a subway station	China	Diverse, reflecting commuters

Table 5 Global model performance on each dataset

	Rank-1	Rank-5	Rank-10	mAP
Market-1501	88.2	89.7	90.8	70.3
CUHK03-NP	59.7	62.5	65.4	51.2
PRID	75.6	80.3	83.3	69.8
CUHK01	80.5	84.4	85.2	74.6
CUHK02	56.9	58.5	59.2	50.4
VIPeR	52.0	55.8	58.1	49.7
3DPeS	86.4	87.9	89.5	80.0
iLIDS	84.3	86.3	88.4	78.2
GRID	57.5	61.2	64.6	57.0

These results highlight the model's generalizability, likely due to the diversity of the data in these datasets. However, the performance on datasets such as CUHK03-NP, CUHK02, VIPeR, and GRID shows a Rank-1 accuracy drop to 59%, 56%, 52%, and 57%, respectively. The corresponding mAP scores are also lower, with CUHK02 and VIPeR showing 50.4 and 49.7, respectively. These datasets are more challenging due to smaller dataset sizes, lower image quality, and higher variability in appearance. The Rank-5 and Rank-10 accuracies do show improvement, suggesting that while the model may struggle with exact matches, it can still identify the correct individual within the top few predictions, showing a degree of resilience. These results indicate that the proposed approach can obtain a good generalized global model that is transferable to diverse scenarios and underscore the effectiveness of the global model in handling complex and varied datasets, while also highlighting areas where further refinement is necessary. The lower performance on certain datasets suggests that while the global model benefits from aggregated data, it can still be influenced by the inherent challenges of individual datasets. This reinforces the need for the weighted aggregation technique and adaptive client model selection strategies proposed in our work, ensuring that the model is not only generalized but also adaptable to specific dataset characteristics. Overall, the results validate the global model's ability to balance between specialization and generalization, addressing the core challenges in federated learning for people re-identification. Notably, the model excels on datasets like Market-1501, 3DPeS, and iLIDS, indicating its effectiveness in handling both large-scale and challenging re-identification tasks. The use of generative learning to augment datasets and the feature attention fusion technique further enhances the local models' ability to extract high-quality features, as seen in the strong performance on complex datasets like Market-1501.

The effectiveness of the proposed framework in addressing bias while ensuring data security in distributed learning systems is demonstrated in Table 6. The comparison between the local and global models across the datasets shows that the global model trained using our proposed pipeline for federated learning and using weighted aggregation, generally outperforms the local models, especially on challenging datasets like CUHK03-NP and PRID with a 14% and 17% increase in accuracy respectively. This suggests that by balancing the contributions of different clients, the server-side model is not unduly influenced by any single participant, leading to a more robust and equitable model. The local model, which is trained individually on each client's dataset, demonstrates superior performance on most datasets, particularly Market-1501, CUHK01, CUHK02, and iLIDS, with notable margins over the global model, demonstrating a margin of 1%, 8%, 3%, and 5% respectively. This suggests that the local models are better tuned to the specific characteristics of individual datasets, likely due to the absence of cross-client interference during training. However, the global model, which aggregates knowledge from all clients, outperforms the local model on datasets like CUHK03-NP, PRID, VIPeR, and 3DPeS. This indicates that the global model benefits from the diversity of data, leading to better generalization and robustness on these datasets. The Reinforcement learning-based client selection strategies used, UCB, can introduce bias by favoring certain clients over time based on reward allocation. However, in our setup, this issue was mitigated due to the sufficient number of training rounds and clients, ensuring all clients contributed effectively to the global model. The balanced participation across multiple rounds prevented any single client from being overlooked or deprioritized. Our experiments confirmed that the aggregation process remained fair and with no observed bias towards specific clients which leads to stable convergence and overall improved performance These results underline the importance of the weighted aggregation technique employed in the study, as it aims to balance the contributions of different clients, preventing any single dataset from dominating the global model's learning process. The slightly lower performance of the global model on certain datasets could be attributed to the variations in client data quality and quantity, a challenge that our adaptive client model selection strategy aims to mitigate by considering client capabilities and performance metrics.

Table 6 Local vs global model performance on each dataset

	Market-1501	CUHK03-NP	PRID	CUHK01	CUHK02	VIPeR	3DPeS	iLIDS	GRID
Local Model	89.3	45.6	58.3	88.4	59.8	57.2	85.3	79.3	60.7
Global Model	88.2	59.7	75.6	80.5	56.9	52.0	86.4	84.3	57.5

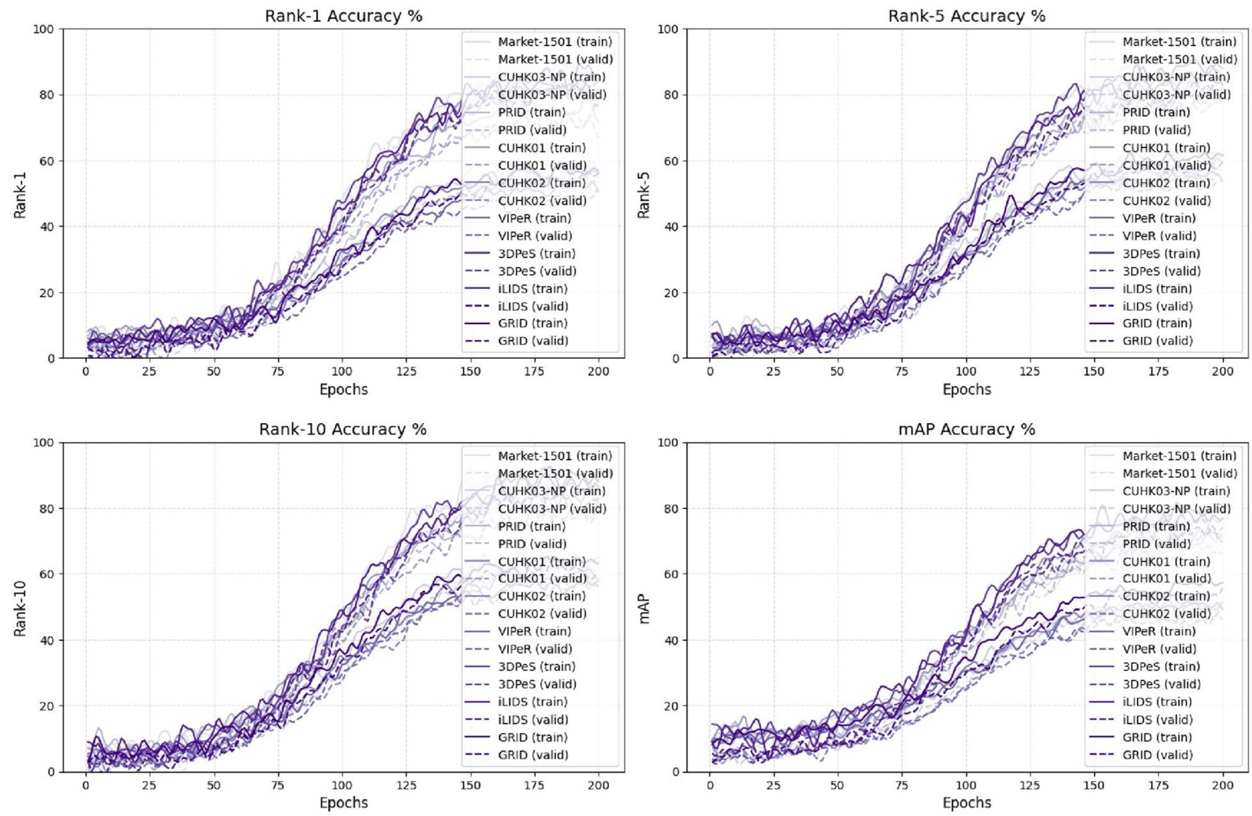

Fig. 9 The accuracy of the global model across different datasets

The graph shown in Fig. 9 shows the training progress of our person re-identification model across the different experimental datasets. The model generally improves over time, but performance varies significantly across datasets. Market-1501 achieves the highest scores, nearing 90% on rank-based metrics, while other datasets like CUHK03-NP, PRID, CUHK01, CUHK02, VIPeR, 3DPeS, ILIDS, and GRID show lower and more varied performance, which mirrors the differing dataset complexities and potential overfitting to Market-1501.

Ablation study

In this section, we evaluate the impact of the proposed enhancements to the ReID model as well as the adaptive methodologies employed in the federated learning system.

Impact of ReID model enhancements

This subsection focuses on evaluating the impact of GAN-based augmentation and the Feature Attention Mechanism (FAM) on the performance of the ReID model.

Impact of GAN synthetic data augmentation As demonstrated in Table 7, the use of GAN-generated images significantly improved performance across different datasets. For instance, in the Market-1501 dataset, mAP accuracy increased from 80.3% to 88.7%. Similar improvements were seen in larger datasets like Market, where the model's ability to generalize marginally improved with the introduction of synthetic images.

Table 7 Impact of data augmentation on model performance

Dataset	No Augmentation (mAP)	With Augmentation (mAP)
Market-1501	80.3%	88.7%
GRID	47.5%	59.3%

Impact of feature attention mechanisms As shown in Fig. 10 Feature map generated by the model pre-trained on ImageNet showing the initial feature representation before further training. Feature map produced by the model fine-tuned on the person identification dataset to show the model's learned features relevant to the target domain. The feature map extracted from the model with the feature attention mechanism applied to illustrate the model's ability to concentrate on relevant visual cues.

Table 8 compares two scenario in a federated learning setup with 3 clients in an IID scenario. The added feature attention configuration shows a significant performance boost, achieving 87.0% mAP, compared to 75.5% mAP for no added attention. This highlights the importance of attention mechanisms in improving model accuracy for person re-identification tasks.

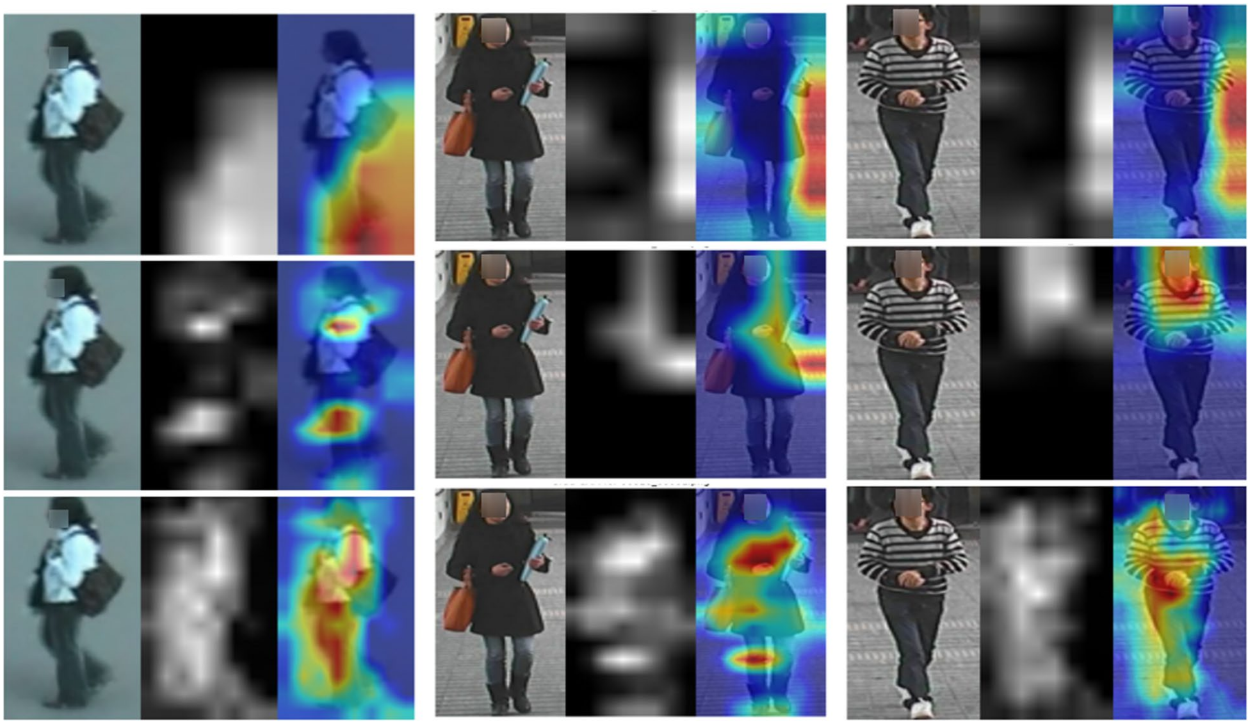

Fig. 10 Feature maps visualization to show the effect of the FAM

Table 8 Impact of FAM on model performance

Configuration	Accuracy % (mAP)
No Feature Attention	75.5
With Feature Attention	87.0

Impact of adaptive methodologies

Impact of partial aggregation This subsection compares the impact of FedPav and FedAvg on federated learning performance. We analyze the effects of client selection, model selection, and weighted partial aggregation on the global model's performance. Table 9 compares the performance of two federated learning methods, FedPav and FedAvg, across three key metrics: Rank-1 accuracy, convergence speed, and fairness. FedPav demonstrates superior performance, achieving a Rank-1 accuracy of 83.1%, significantly higher than FedAvg's 72.4%. Additionally, FedPav converges faster, requiring around 75 to 90 rounds compared to FedAvg's 85 to 100 rounds which indicates greater efficiency in reaching a stable solution. Overall, FedPav outperforms FedAvg in accuracy, efficiency, and fairness, making it a more robust and effective choice for federated learning tasks, particularly in scenarios where these factors are critical.

Table 9 Performance comparison between FedPav and FedAvg

Method	Rank-1 Accuracy	Convergence Speed
FedAvg	72.4%	Low (85–100 epochs)
FedPav	83.1%	High (75–90 epochs)

Impact of adaptive model selection As shown in Fig. 11, The adaptive model selection strategy significantly improved the learning process by assigning backbone models based on each client's dataset characteristics and computational capabilities. For clients with complex data, like Market-1501 and GRID, while initially assigned ConvNeXt Large to handle the data complexity, they were later switched to ResNet-50. This reassignment ensured that even large datasets benefitted from robust feature extraction without overwhelming resource constraints. Meanwhile, clients with smaller and simpler datasets, such as PRID, VIPeR, and iLIDS, were consistently assigned the lighter ConvNeXt Tiny model, which allowed for faster processing while maintaining strong performance suited to the lower complexity of the data.

Impact of adaptive strategies on IID and non-IID setup In this experiment, we use two datasets, 3DPeS and iLIDS, each with different numbers of identities (93 and 59, respectively), to evaluate the performance of our federated learning approach under an IID scenario with three clients. The data from both datasets are evenly distributed across the three clients, ensuring each client has an equal number of samples from each identity. This setup provides a controlled environment to test the algorithm's ability to generalize when the data distribution is uniform across clients. The following table shows the retrieval accuracy results for the IID scenario (Table 10).

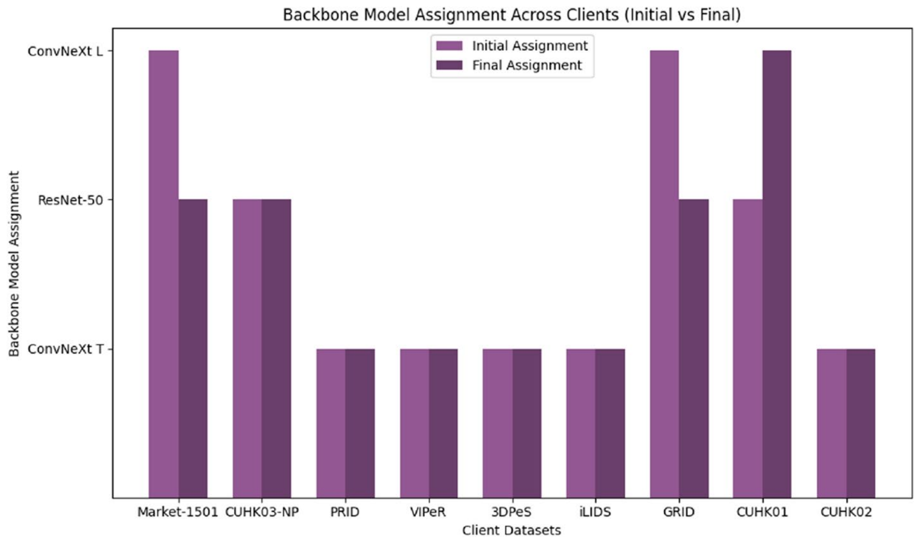

Fig. 11 Initial and final assignment of backbones for each client

Table 10 Rank-1 Accuracy in IID Scenario and non-IID Scenario

Proposed Method	3DPeS	iLIDS
IID setup (with no adaptive methods)	73.9%	68.1%
IID setup (with adaptive methods)	92.5%	93.2%
Non-IID setup (with no adaptive methods)	56.2%	47.0%
Non-IID setup (with adaptive methods)	86.2%	84.6%

The results show that in the IID setup, the federated learning model performs consistently across both datasets, with the retrieval accuracy being 82.5% for 3DPeS and 78.3% for iLIDS. This validates our algorithm in a simple scenario where the data is equally distributed across clients, demonstrating stable performance. The consistency in the results suggests that our approach is effective under controlled data conditions. By validating our algorithm in various setups, including this IID scenario, we aim to evaluate its robustness across different types of data distributions and ensure its generalizability to more complex non-IID settings.

The graph shown in Fig. 12 presents the performance comparison of the model under IID and Non-IID scenarios. In both scenarios, the model's performance generally improves over epochs. However, the IID setting shows significantly higher accuracy, particularly for the 3DPES dataset. The Adaptive strategies offer an advantage in the IID setting as well as the Non-IID setting. ILIDS datasets, both with and without adaptive learning, demonstrate lower overall accuracy compared to 3DPES across both IID and Non-IID scenarios, inherent from the small size and dataset difficulty. The fluctuations in accuracy that are especially clear in the initial epochs indicate the challenges of training, which are more evident in the Non-IID setting and the difficulties in generalizing to unseen distributions.

Comparative analysis

The comparison shown in Tables 11, 12, and 13, compares the performance of the proposed model against

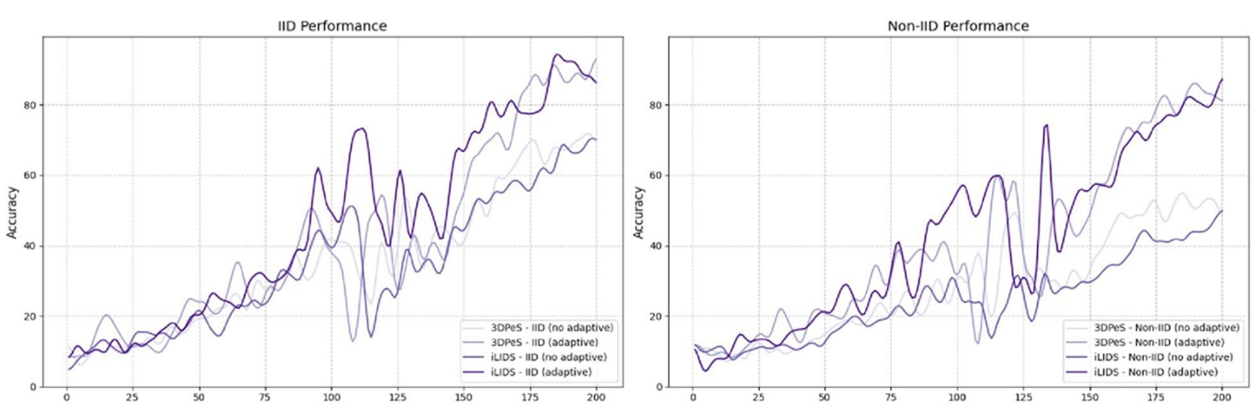

Fig. 12 The impact of adaptive strategies on 3DPES and ILIDS datasets under IID and Non-IID settings

Table 11 Proposed model performance vs federated state-of-the-art on 9 edge setup

	Rank-1	Rank-5	Rank-10	mAP
FedReID Baseline	51.46	77.07	82.37	54.34
Proposed Framework	71.23	74.06	76.05	64.58

Table 12 Proposed model performance vs federated re-identification frameworks on 2 edge setup

	Rank-1	Rank-5	Rank-10	mAP
Fed4ReID	67.20	82.34	77.88	63.80
FRM	70.67	77.0	81.25	61.60
Proposed Framework	85.33	89.02	91.21	72.30

three federated learning state-of-the-art methods-FedReID Baseline [45], Fed4ReID [48], and FRM [53]-on people re-identification tasks, measured by Rank-1, Rank-5, Rank-10 accuracies, and mean Average Precision (mAP).

The proposed framework demonstrates the highest Rank-1 accuracy at 71% on 9 clients and 85% on 2 clients, outperforming the other models. This indicates that our model is the most effective at correctly identifying individuals on the first attempt, showcasing its superior feature extraction and generalization capabilities. Additionally, the proposed framework achieves the highest mAP score of 64.58 on 9 clients and 72 on 2 clients, which is a critical metric in re-identification tasks, as it reflects the model's overall retrieval performance across all ranks. This suggests that the proposed framework not only excels in the first-rank identification but also maintains consistent performance across different queries, providing reliable and accurate results.

In comparison, the FedReID Baseline lags behind with a Rank-1 accuracy of 51% and an mAP of 54, highlighting the limitations of traditional federated learning approaches when dealing with the complexities of people re-identification. Although Fed4ReID improves on the baseline with a Rank-1 accuracy of 67% and a mAP of 63, it still falls short of the performance demonstrated by our proposed framework. FRM, while showing a slightly better Rank-1 accuracy than Fed4ReID at 70%, has a lower mAP of 61, indicating that it may struggle with consistent retrieval accuracy across varying conditions. While the proposed framework leads in Rank-1 and mAP scores, it has slightly lower Rank-5 and Rank-10 accuracies compared to Fed4ReID and FRM. This suggests that while our model excels in first-rank identification, the other models might be more effective at identifying the correct individual within the top five or ten predictions. However, the marginal differences in these metrics do not undermine the overall superiority of our proposed approach.

Table 14 Proposed model performance vs centralized state-of-the-art

	Rank-1	Rank-5	Rank-10	mAP
Centralized Baseline	72.39	75.67	81.67	68.92
Proposed Framework	71.23	74.06	76.05	64.58

The results reinforce the value of our proposed enhancements, including the weighted aggregation technique and adaptive client model selection strategies, in pushing the boundaries of federated learning for real-world applications.

Table 14 shows a comparison between our approach and a centralized learning approach. The centralized baseline model outperforms the proposed framework with a Rank-1 accuracy of 72.39% and an mAP of 68%. These results suggest that the centralized model, which has access to all data in a unified training process, is slightly more effective in accurately identifying individuals on the first attempt and in overall retrieval performance. The slight drop in Rank-1 and mAP performance for the proposed framework can be attributed to the trade-offs made to preserve data privacy and security, as well as the need to aggregate diverse client models in a distributed environment. However, the close proximity of these metrics to the centralized baseline underscores the success of our federated approach in maintaining high accuracy and retrieval performance, even in the absence of centralized data access. While the centralized baseline achieves slightly higher accuracy rates across all metrics, the performance gap between the proposed framework and the centralized baseline is relatively small. This suggests that the federated learning approach, despite the challenges of distributed training and data privacy, can achieve results comparable to a centralized setting. In comparison, the proposed framework, which operates within a federated learning environment, achieves a Rank-1 accuracy of 71% and an mAP of 64. While these metrics are slightly lower than those of the centralized baseline, they are still competitive, especially considering the inherent challenges of federated learning, such as data decentralization, privacy preservation, and potential variations in client data quality.

Table 13 Rank-1 accuracy proposed model performance vs federated state-of-the-art

	Market-1501	CUHK03-NP	PRID	CUHK01	CUHK02	VIPeR	3DPeS	iLIDS	GRID
FedReID Baseline	85.1	34.24	41.3	78.3	43.3	47.8	72.76	80.6	77.1
Fed4ReID	-	-	68.0	72.8	-	38.6	77.1	79.5	-
FRM	84.01	57.32	-	-	-	-	-	-	-
Proposed Framework	88.2	59.7	75.6	80.5	56.9	52.0	86.4	84.3	57.5

The results presented in Table 14 demonstrate that the proposed federated learning framework can achieve performance comparable to a centralized baseline. While the centralized approach may have a slight edge in terms of accuracy, the federated framework offers advantages such as data privacy and scalability, making it a viable option for distributed learning scenarios. Additionally, The adaptive model selection allows each client to use the most suitable model for its resources from a pool of models. This results in lower computational demands and, consequently, reduced CO_2 emissions compared to centralized training. A centralized approach would require a single large model, with the highest parameter count, to train on the full dataset, leading to significantly higher resource consumption. By leveraging models that better fit client-specific needs, our federated approach demonstrates environmental and computational efficiency.

Conclusion & future work

In this paper, we propose a novel federated learning framework for people re-identification that incorporates adaptive strategies, feature attention mechanisms, and UCB-based reinforcement learning into the federated learning framework. We investigate the inherent challenges of data heterogeneity and non-IID distributions in federated settings, the proposed adaptive strategies enhance model performance and convergence. GANs are used to synthesize and augment training data for a more robust model. The use of feature attention mechanisms further refines the feature extractor and the discriminative power of the model. As the number of parameters in person ReID models may make it difficult to deploy on edge devices with limited computing power, we propose adaptive methods for model selection. Additionally, the UCB-based reinforcement learning strategies optimize client selection to ensure efficient model training and deployment. Finally, we conducted extensive experiments on an edge-cloud scenario using nine datasets. Our experiments demonstrate the superiority of the proposed algorithm by outperforming other state-of-the-art federated learning algorithm benchmarks in terms of accuracy with a Rank-1 accuracy of 71% and comparable results to centralized with a margin of 2% in Rank-1 accuracy.

In the future, different GAN architectures and data labeling strategies that may further improve the quality of generated synthetic data can be explored. GANs can also be used to synthesize images for the public unlabeled dataset in a real world simulation along with publicly available datasets or webcams that can be shared among clients. Additionally, trying other aggregation strategies, such as combining FedPav with FedNova, could lead to better handling of heterogeneous data. Fairness-aware techniques should be explored in future research to further enhance the ethical implications of this approach. Finally, we test the proposed approach in a federated-by-device setting and use hierarchical aggregation so that the accuracy and generality of the trained model can be further improved.

Authors' contributions
D.W. wrote the main manuscript. Y.A. and N.F. co-edited the manuscripts and provided suggestions and mentorship to the work. M.S. and M.A. co-supervised the work. All authors reviewed the manuscript.

Funding
Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Received: 15 December 2024 / Accepted: 24 March 2025
Published online: 06 May 2025

References

1. Palumbo G, Carneiro D, Alves V (2024) Objective metrics for ethical AI: a systematic literature review. Int J Data Sci Anal. https://doi.org/10.1007/s41060-024-00541-w
2. Gamal A, Shoukry N, Salem MAM (2021) Long-term person re-identification model with a strong feature extractor. In: 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS). IEEE, Cairo, pp 74–79
3. Shoukry N, Abd El Ghany MA, Salem MAM (2022) Multi-modal long-term person re-identification using physical soft bio-metrics and body figure. Appl Sci 12(6):2835
4. Voigt P, Von dem Bussche A (2017) The EU General Data Protection Regulation (GDPR): A Practical Guide. 1st Ed. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-57959-7
5. Truong N, Sun K, Wang S, Guitton F, Guo Y (2021) Privacy preservation in federated learning: An insightful survey from the gdpr perspective. Comput Secur 110:102402
6. von Hein J, Bizer A (2018) Social media and the protection of privacy: Current gaps and future directions in european private international law. Int J Data Sci Analytics 6(3):233–239
7. Harvey J Adam LaPlace (2021) Exposing.ai. https://exposing.ai. Accessed 19 Apr 2024
8. McMahan HB, Moore E, Ramage D, y Arcas BA (2017) Communication-efficient learning of deep networks from decentralized data. In: The International Conference on Artificial Intelligence and Statistics, vol. 54. JMLR: W&CP, Fort Lauderdale, pp 1273–1282
9. Waref D, Salem M (2022) Split federated learning for emotion detection. In: 2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES). IEEE, Giza, pp 112–115
10. Abdelgaber YE, Ahmed YA, Salem MAM, Salem MAG (2023) Federated learning for resource management in edge computing. In: 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS). IEEE, pp 102–109
11. Harasic M, Keese FS, Mattern D, Paschke A (2024) Recent advances and future challenges in federated recommender systems. Int J Data Sci Analytics 17(4):337–357
12. Varlamis I, Sardianos C, Chronis C, Dimitrakopoulos G, Himeur Y, Alsalemi A, Bensaali F, Amira A (2023) Using big data and federated learning for

generating energy efficiency recommendations. Int J Data Sci Analytics 16(3):353–369
13. Düsing C, Cimiano P, Paaßen B (2024) Leveraging Local Data Sampling Strategies to Improve Federated Learning (Extended Abstract). In: 11th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2024. IEEE, San Diego, 2024, pp. 1–2. https://doi.org/10.1109/DSAA61799.2024.10722778
14. Ming Z, Zhu M, Wang X, Zhu J, Cheng J, Gao C, Yang Y, Wei X (2022) Deep learning-based person re-identification methods: A survey and outlook of recent works. Image Vis Comput 119:104394
15. Wang H, Du H, Zhao Y, Yan J (2020) A comprehensive overview of person re-identification approaches. IEEE Access 8:45556–45583
16. Bedagkar-Gala A, Shah SK (2014) A survey of approaches and trends in person re-identification. Image Vis Comput 32(4):270–286
17. Leng Q, Ye M, Tian Q (2019) A survey of open-world person re-identification. IEEE Trans Circ Syst Video Technol 30(4):1092–1108
18. Akan S, Varlı S, Bhuiyan MAN (2024) An enhanced swin transformer for soccer player reidentification. Sci Rep 14(1):1139
19. Gao Z, Wei S, Guan W, Zhu L, Wang M, Chen S (2023) Identity-guided collaborative learning for cloth-changing person reidentification. IEEE Trans Pattern Anal Mach Intell 46(5):2819–2837
20. Zhang X, Luo H, Fan X, Xiang W, Sun Y, Xiao Q, Jiang W, Zhang C, Sun J (2017) AlignedReID: Surpassing human-level performance in person re-identification. arXiv preprint arXiv:1711.08184. https://arxiv.org/abs/1711.08184
21. Yang W, Li D, Chen X, Huang K (2020) Bottom-up foreground-aware feature fusion for person search. In: Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA. pp 3404–341 https://doi.org/10.1145/3394171.3413991
22. Ning E, Wang Y, Wang C, Zhang H, Ning X (2024) Enhancement, integration, expansion: Activating representation of detailed features for occluded person re-identification. Neural Netw 169:532–541
23. Islam A, Tsun MTK, Theng LB, Chua C (2024) Region-based convolutional neural networks for occluded person re-identification. Int J Adv Intell Inform 10(1):49-63. https://doi.org/10.26555/ijain.v10i1.1125
24. Zhu K, Guo H, Zhang S, Wang Y, Liu J, Wang J, Tang M (2023) Aaformer: Auto-aligned transformer for person re-identification. IEEE Trans Neural Netw Learn Syst 35(12):17307–17317
25. Lavi B, Ullah I, Fatan M, Rocha A (2020) Survey on reliable deep learning-based person re-identification models: Are we there yet? arXiv preprint arXiv:2005.00355. https://arxiv.org/abs/2005.00355
26. Wang Y, Zhang P, Wang D, Lu H (2024) Other tokens matter: Exploring global and local features of vision transformers for object re-identification. Comp Vision Image Underst 244:104030
27. Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K (2021) Attentional feature fusion. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV), Waikoloa, pp 3560–3569. https://doi.org/10.1109/WACV48630.2021.00360
28. Zhuang X, Wei D, Liang D, Jiang L (2024) Feature attention fusion network for occluded person re-identification. Image Vis Comput 143:104921
29. Saribas H, Cevikalp H, Köpüklü O, Uzun B (2022) Trat: Tracking by attention using spatio-temporal features. Neurocomputing 492:150–161
30. Zheng Z, Yang X, Yu Z, Zheng L, Yang Y, Kautz J (2019) Joint discriminative and generative learning for person re-identification. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 2138–2147. https://doi.org/10.1109/CVPR.2019.00224
31. Yang F, Zhong Z, Luo Z, Lian S, Li S (2019) Leveraging virtual and real person for unsupervised person re-identification. IEEE Trans Multimedia 22(9):2444–2453
32. Chen H, Wang Y, Lagadec B, Dantcheva A, Bremond F (2021) Joint generative and contrastive learning for unsupervised person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2021, virtual, pp 2004–2013. https://doi.org/10.1109/CVPR46437.2021.00204
33. Khaldi K, Nguyen VD, Mantini P, Shah S (2024) Unsupervised person re-identification in aerial imagery. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp 260–269
34. Li H, Ye M, Zhang M, Du B (2024) All in one framework for multimodal re-identification in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, pp 17459–17469. https://doi.org/10.1109/CVPR52733.2024.01653
35. Sun W, Jiang X, Dou S, Li D, Miao D, Deng C, Zhao C (2023) Invisible backdoor attack with dynamic triggers against person re-identification. IEEE Trans Inf Forensic Security 19:307-319.
36. Kansal K, Wong Y, Kankanhalli M (2024) Privacy-Enhancing Person Re-identification Framework – A Dual-Stage Approach. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, 2024, pp. 8528–8537. https://doi.org/10.1109/WACV57701.2024.00835
37. KhoKhar FA, Shah JH, Khan MA, Sharif M, Tariq U, Kadry S (2022) A review on federated learning towards image processing. Comput Electr Eng 99:107818
38. Blanco-Justicia A, Domingo-Ferrer J, Martínez S, Sánchez D, Flanagan A, Tan KE (2021) Achieving security and privacy in federated learning systems: Survey, research challenges and future directions. Eng Appl Artif Intell 106:104468
39. Li L, Fan Y, Tse M, Lin KY (2020) A review of applications in federated learning. Comput Ind Eng 149:106854
40. Himeur Y, Varlamis I, Kheddar H, Amira A, Atalla S, Singh Y, Mansoor W (2023) Federated learning for computer vision. arXiv preprint arXiv:2308.13558. https://arxiv.org/abs/2308.13558
41. Pham QV, Dev K, Maddikunta PKR, Gadekallu TR, Huynh-The T, et al (2022) Fusion of federated learning and industrial internet of things: A survey. Computer Networks 212:109048
42. Dong C, Zhou J, An Q, Jiang F, Chen S, Pan L, Liu X (2023) Optimizing performance in federated person re-identification through benchmark evaluation for blockchain-integrated smart uav delivery systems. Drones 7(7):413
43. Ji S, Tan Y, Saravirta T, Yang Z, Liu Y, Vasankari L, Pan S, Long G, Walid A (2024) Emerging trends in federated learning: From model fusion to federated x learning. Int J Mach Learn Cyber 15:3769–3790. https://doi.org/10.1007/s13042-024-02119-1
44. Zhuang W, Gan X, Wen Y, Zhang S (2023) Optimizing performance of federated person re-identification: Benchmarking and analysis. ACM Trans Multimedia Comput Commun Appl 19(1s):1–18
45. Zhuang W, Wen Y, Zhang X, Gan X, Yin D, Zhou D, Zhang S, Yi S (2020) Performance optimization of federated person re-identification via benchmark analysis. In: Proceedings of the 28th ACM International Conference on Multimedia, Virtual Event, Seattle, pp 955–963. https://doi.org/10.1145/3394171.3413814
46. Weng J, Hu K, Yao T, Wang J, Wang Z (2022) Robust knowledge adaptation for federated unsupervised person reid. In: 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, Sydney, pp 1–8
47. Li Y, Wang X, Zeng R, Donta PK, Murturi I, Huang M, Dustdar S (2023) Federated domain generalization: A survey. arXiv preprint arXiv:2306.01334. https://arxiv.org/abs/2306.01334
48. Zhang C, Liu X, Xiang M, Yao A, Fan X, Li G (2023) Fed4reid: Federated learning with data augmentation for person re-identification service in edge computing. In: 2023 IEEE International Conference on Web Services (ICWS). IEEE, Chicago, pp 64–70
49. Yang F, Zhong Z, Luo Z, He Y, Li S, Sebe N (2024) Diversity-authenticity co-constrained stylization for federated domain generalization in person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 38. AAAI Press, Vancouver, pp 6477–6485. https://doi.org/10.1609/aaai.v38i6.28468
50. Liu F, Ye M, Du B (2024) Domain generalized federated learning for person re-identification. Comput Vis Image Underst 241:103969
51. Yang F, Zhong Z, Luo Z, Li S, Sebe N (2022) Federated and generalized person re-identification through domain and feature hallucinating. arXiv preprint arXiv:2203.02689. https://arxiv.org/abs/2203.02689
52. Zhou T, Konukoglu E (2023) Fedfa: Federated feature augmentation. In: The Eleventh International Conference on Learning Representations, Kigali. https://openreview.net/pdf?id=U9yFP90jU0
53. Tang B, Xu X, Dai F, Wang S (2023) Person re-identification withmodel-contrastive federated learning in edge-cloud environment. Intell Autom Soft Comput 38(1):35–55. https://doi.org/10.32604/iasc.2023.036715
54. Chen L, Zhang W, Dong C, Huang Z, Nie Y, Hou Z, Qiao S, Tan CW (2024) Fed-drl: Trustworthy federated learning model fusion method based on staged reinforcement learning. Comput Inform 43(1):1–37
55. Ghari PM, Shen Y (2024) Budgeted online model selection and fine-tuning via federated learning. arXiv preprint arXiv:2401.10478. https://arxiv.org/abs/2401.10478
56. Abbasi A, Dong F, Wang X, Leung H, Zhou J, Drew S (2024) Fedgreen: Carbon-aware federated learning with model size adaptation. IEEE International

Conference on Communication Workshops (ICC Workshops), Denver, pp. 1352–1358
57. Sun R, Li Y, Shah T, Sham RW, Szydlo T, Qian B, Thakker D, Ranjan R (2022) Fedmsa: a model selection and adaptation system for federated learning. Sensors 22(19):7244
58. Ilhan F, Su G, Liu L (2023) Scalefl: Resource-adaptive federated learning with heterogeneous clients. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE CVPR, Vancouver, pp 24532–24541. https://doi.org/10.1109/CVPR52729.2023.02350
59. Quentin P, Swan S, Hugo W, Léo R, Siba H, Antoun Y (2023) Balancing accuracy and training time in federated learning for violence detection in surveillance videos: A study of neural network architectures. J Comput Sci Technol 39:1029-1039. https://doi.org/10.1007/s11390-024-3702-7
60. Liu Y, Huang A, Luo Y, Huang H, Liu Y, Chen Y, Feng L, Chen T, Yu H, Yang Q (2020) Fedvision: An online visual object detection platform powered by federated learning. In: Proceedings of the AAAI conference on artificial intelligence, EAAI 2020, vol 34. AAAI Press, New York, pp 13172–13179. https://doi.org/10.1609/aaai.v34i08.7021
61. Shaheen M, Farooq MS, Umer T (2024) Ai-empowered mobile edge computing: inducing balanced federated learning strategy over edge for balanced data and optimized computation cost. J Cloud Comput 13(1):52
62. Gu W, Zhang Y (2023) Fedeem: a fairness-based asynchronous federated learning mechanism. J Cloud Comput 12(1):154
63. Emara S, Liu D, Wang F, Li B (2024) Cascade: Enhancing reinforcement learning with curriculum federated learning and interference avoidance-a case study in adaptive bitrate selection. IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Vancouver, pp. 01-06. https://doi.org/10.1109/INFOCOMWKSHPS61880.2024.10620892
64. Chen W, Zhou X (2021) Ucbfed: Using reinforcement learning method to tackle the federated optimization problem. In: IFIP International Conference on Distributed Applications and Interoperable Systems, vol 12718. Springer, Cham, pp 99–105. https://doi.org/10.1007/978-3-030-78198-9_7
65. Sutton RS (2018) Reinforcement learning: an introduction. A Bradford Book. Cambridge, 262039249.
66. Wang Z, Chen S, Bai L, Gao J, Tao J, Bond RR, Mulvenna MD (2023) Reinforcement learning based task scheduling for environmentally sustainable federated cloud computing. J Cloud Comput 12(1):174
67. Zhang D, Sun W, Zheng ZA, Chen W, He S (2023) Adaptive device sampling and deadline determination for cloud-based heterogeneous federated learning. J Cloud Comput 12(1):153
68. Nagalapatti L, Mittal RS, Narayanam R (2022) Is your data relevant? Dynamic selection of relevant data for federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 36. EAAI 2022 Virtual Event, AAAI Press, pp 7859–7867. https://doi.org/10.1609/aaai.v36i7.20755
69. Xie P, Gao X, Li F, Xing L, Zhang Y, Sun H (2024) Communication-efficient zeroth-order adaptive optimization for federated learning. Mathematics 12(8):1148
70. Venkatesha Y, Kundu S, Panda P (2024) Assortment of attention heads: Accelerating federated PEFT with head pruning and strategic client selection. SSRN Electronic Journal. https://ssrn.com/abstract=4790569
71. Girija S, Baker T, Ahmed N, Khedr AM, Al Aghbari Z, Jha A, Sobolev K, Asl SA, Phan AH (2023) Attribute recognition for person re-identification using federated learning at all-in-edge. Internet Things 22:100793
72. Chen L, Zhao D, Tao L, Wang K, Qiao S, Zeng X, Tan CW (2024) A credible and fair federated learning framework based on blockchain. IEEE Trans Artif Intell 6(2):301–316. https://doi.org/10.1109/TAI.2024.3355362
73. Zhang L, Gao G, Zhang H (2023) Spatial-temporal federated learning for lifelong person re-identification on distributed edges. IEEE Trans Circ Syst Video Technol 35(2):1884–1896. https://doi.org/10.1109/TCSVT.2023.3281983
74. Zhuang W, Wen Y, Zhang S (2021) Joint optimization in edge-cloud continuum for federated unsupervised person re-identification. In: Proceedings of the 29th ACM International Conference on Multimedia, China. pp 433–441. https://doi.org/10.1145/3474085.3475182
75. Wu G, Gong S (2021) Decentralised learning from independent multi-domain labels for person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 35. AAAI Press, pp 2898–2906. https://doi.org/10.1609/aaai.v35i4.16396
76. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
77. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S (2022) A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE CVPR, New Orleans, pp 11976–11986. https://doi.org/10.1109/CVPR52688.2022.01167
78. Mao A, Mohri M, Zhong Y (2023) Cross-entropy loss functions: Theoretical analysis and applications. In: International conference on Machine learning. PMLR, Honolulu, pp 23803–23828
79. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society, Boston, pp 815–823. https://doi.org/10.1109/CVPR.2015.7298682
80. Kryszkiewicz M (2014) The cosine similarity in terms of the euclidean distance. In: Encyclopedia of Business Analytics and Optimization. vol. IV. IGI Global, pp 2498–2508
81. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1106–1114
82. Cheng DS, Cristani M, Stoppa M, Bazzani L, Murino V, et al (2011) Custom pictorial structures for re-identification. In: Bmvc, vol 1. Citeseer, Dundee, p 6
83. Wang T, Gong S, Zhu X, Wang S (2014) Person re-identification by video ranking. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer, Berlin, pp 688–703
84. Loy CC, Xiang T, Gong S (2010) Time-delayed correlation analysis for multi-camera activity understanding. Int J Comput Vis 90:106–129
85. Li W, Zhao R, Wang X (2013) Human reidentification with transferred metric learning. In: Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11. Springer, Berlin, pp 31–44
86. Li W, Wang X (2013) Locally aligned feature transforms across views. In: Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE Computer Society, Portland, pp 3594–3601. https://doi.org/10.1109/CVPR.2013.461
87. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: Deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society, Columbus, pp 152–159. https://doi.org/10.1109/CVPR.2014.27
88. Hirzer M, Beleznai C, Roth PM, Bischof H (2011) Person re-identification by descriptive and discriminative classification. In: Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17. Springer, Berlin, pp 91–102
89. Gray D, Tao H (2008) Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10. Springer, Berlin, pp 262–275

Publisher's Note