Decentralised Person Re-Identification with Selective Knowledge Aggregation

Shitong Sun
shitong.sun@qmul.ac.uk

Guile Wu
guile.wu@qmul.ac.uk

Shaogang Gong
s.gong@qmul.ac.uk

Computer Vision Group,
School of Electronic Engineering and
Computer Science,
Queen Mary University of London,
London E1 4NS, UK

Abstract

Existing person re-identification (Re-ID) methods mostly follow a centralised learning paradigm which shares all training data to a collection for model learning. This paradigm is limited when data from different sources cannot be shared due to privacy concerns. To resolve this problem, two recent works [32, 41] have introduced decentralised (federated) Re-ID learning for constructing a globally generalised model (server) without any direct access to local training data nor shared data across different source domains (clients). However, these methods are poor on how to adapt the generalised model to maximise its performance on individual client domain Re-ID tasks having different Re-ID label spaces, due to a lack of understanding of data heterogeneity across domains. We call this poor 'model personalisation'. In this work, we present a new Selective Knowledge Aggregation approach to decentralised person Re-ID to optimise the trade-off between model personalisation and generalisation. Specifically, we incorporate attentive normalisation into the normalisation layers in a deep ReID model and propose to learn local normalisation layers specific to each domain, which are decoupled from the global model aggregation in federated Re-ID learning. This helps to preserve model personalisation knowledge on each local client domain and learn instance-specific information. Further, we introduce a dual local normalisation mechanism to learn generalised normalisation layers in each local model, which are then transmitted to the global model for central aggregation. This facilitates selective knowledge aggregation on the server to construct a global generalised model for out-of-the-box deployment on unseen novel domains. Extensive experiments on eight person Re-ID datasets show that the proposed approach to decentralised Re-ID significantly outperforms the state-of-the-art decentralised methods on both seen client domains and unseen novel domains.

1 Introduction

Person re-identification (Re-ID) attempts to retrieve a person of interest from a set of gallery images captured from non-overlapping camera views [10, 15, 34, 35, 36]. It plays an important role in a wide range of real-world applications, such as finding a missing person, smart city management. Existing person Re-ID methods mostly follow a centralised learning

paradigm which requires to collect all training data from different camera views or domains together for model learning. Despite significant progress has been made, centralised Re-ID learning ignores that person images contain a large amount of personal privacy information which may not be allowed to be shared to a central data collection. This limits fundamentally existing centralised learning based Re-ID in real-world applications with increasing privacy-sensitive scenarios.

To solve this problem, two recent Re-ID works [52, 41] have applied federated learning [21] to person re-identification by constructing a global generalised model (server) without access to local training data nor sharing data across different source domains (clients). This privacy-preserving Re-ID learning paradigm is known as *decentralised person Re-ID* [52]. More specifically, in decentralised person Re-ID, each local client trains its local model using its own set of local training data without sharing data with other local clients, whilst a central server generates a global model by aggregating local model weights (or local model updates) without any direct access to training data. This decentralised person Re-ID paradigm inherently protects source data privacy. Although latest decentralised Re-ID methods [52, 41] have shown encouraging performance, their focus is on how to learn a global generalised model and they have ignored the need to understand data heterogeneity across domains. As a result, they are sub-optimal when compared to local trained models for individual client domain Re-ID tasks having different Re-ID label spaces, *i.e.* showing poor 'model personalisation' on each local client domain.

In this work, we propose to optimise the trade-off between model personalisation and generalisation in decentralised person Re-ID. We present a new Selective Knowledge Aggregation (SKA) approach to facilitate feature extraction model learning in the iterative federated Re-ID learning paradigm [52, 41]. Specifically, we incorporate attentive normalisation [17] into the normalisation layers in a deep ReID model and propose to learn local normalisation layers specific to each client. These local normalisation layers are decoupled from the central server aggregation whilst the other layers in the feature extraction models are iteratively updated between the clients and the server. Learning these local specific normalisation layers helps to better represent model personalisation knowledge from instance-specific information for tuning 'downwards' the server to each participating client domain. Crucially, these local client specific normalisation layers cannot be aggregated 'upwards' to the central server, because directly updating them iteratively to the central server, e.g. by averaging as in existing decentralised Re-ID models, will degrade server generalisation to unseen novel domains. To compensate this emission of direct upwards aggregation to the central server, we introduce a dual local normalisation mechanism to additionally learn generalised normalisation layers in each local client, which are iteratively updated between the clients and the server. This improves to construct a globally generalised model for better out-of-the-box deployment on unseen novel domains.

The key ***contributions*** of this work are: We propose a new Selective Knowledge Aggregation approach to optimise the trade-off between model personalisation and generalisation in decentralised person Re-ID. To learn model personalisation knowledge from instance-specific information per local client, we incorporate attentive normalisation into the normalisation layers in a deep Re-ID model so to learn local normalisation layers specific to each client domain. To replace the direct global model aggregation in existing decentralised learning methods, we introduce a dual local normalisation mechanism which additionally learns generalised normalisation layers in each local client domain for constructing a better global generalised model for unseen novel domains.

Extensive experiments on eight person Re-ID datasets, including DukeMTMC-ReID [59],

Market-1501 [37], CUHK03-NP [14, 40], MSMT17 [30], VIPeR [4], iLIDS [38], GRID [20] and PRID [7], show that the proposed SKA model significantly outperforms the state-of-the-art decentralised methods on both seen client domains and unseen novel domains.

2 Related Work

Person Re-Identification (Re-ID) research has been very active in computer vision in recent years. It aims to match a person of interest across non-overlapping camera views [10, 33, 34, 36]. Traditional person Re-ID methods [15, 35, 37] mainly follow a supervised learning approach, which requires to collect all training data from different camera views in the target domain for model training and usually shows poor generalisation to an unseen novel domain. Recently, generalisable person Re-ID [2, 10, 25], which attempts to generalise a model with data from different domains, has attracted increasing attention and shown promising performance. For example, Song *et al*. [25] assemble data from multiple domains to optimise a domain-invariant mapping network for out-of-the-box Re-ID deployment. Choi *et al*. [2] propose to generalise batch-instance normalisation layers in a meta-learning pipeline for learning a generalisable Re-ID model. However, these generalisable person Re-ID methods require to collect training data from different domains together to learn a feature extraction model for person matching. This largely ignores data privacy and is inherently limited in situations where data from different domains cannot be shared to a central data collection. To resolve this problem, two recent works [32, 41] propose an iterative federated Re-ID learning paradigm to learn a global generalised model with the collaboration of local models but without sharing local data. This decentralised person Re-ID paradigm can inherently protect local data privacy whilst generalising a Re-ID model for deployment. However, current decentralised methods [32, 41] fail to consider how to adapt the generalised model to maximise the performance on individual domain Re-ID tasks. Our work focuses on decentralised person Re-ID model learning from multiple domains without sharing local data across domains so as to protect data privacy. But different from [32, 41], we propose to optimise the trade-off between model personalisation and generalisation by learning local normalisation layers specific to each domain and employ a dual local normalisation mechanism to facilitate aggregating a better global generalised model for both client domains and unseen novel domains.

Federated Learning aims to learn a global model with the collaboration of multiple local models. In the seminal paper [21], McMahan *et al*. introduce a FedAvg algorithm to federated learning, which iteratively averages local model updates to construct a global model. Inspired by the success of FedAvg, there have been many promising works [5, 12, 13, 16, 18, 19, 22, 24, 28, 29] that present more effective federated learning algorithms. Due to non-IID (Independent and Identically Distributed) data distribution across clients, a global generalised model usually shows poor personalisation performance to each specific client. Recently, some works have introduced personalised federated learning [11, 26, 27] which aims to improve model personalisation specific to each client with non-IID data across clients. SCAFFOLD [11] estimates the update direction of the global model to further regularise the update direction of the local client, so as to improve local model personalisation. PFedMe [26] employs Moreau envelopes as the regularised loss for local model optimisation to facilitate personalised model learning. Our work focuses on optimising the trade-off between model personalisation and generalisation for decentralised person Re-ID. Different from contemporary federated learning, our proposed SKA method incorporates attentive
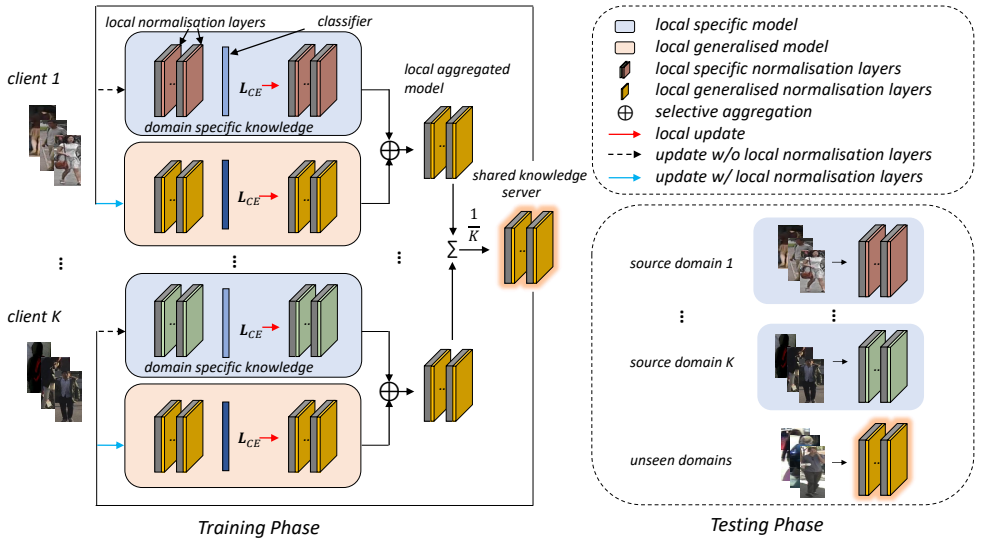
Figure 1: An overview of the Selective Knowledge Aggregation (SKA) method for decentralised person Re-ID. In each training domain, a local client model is optimised to learn local specific normalisation layers (layers in pink and green) for improving *model personalisation*, whilst a local generalised model is optimised to learn local generalisable normalisation layers (layers in yellow) for improving *model generalisation* globally. On the central server, a global generalised model is both aggregated centrally and then used to update client models locally.

normalisation into the normalisation layers and optimises local normalisation layers in a Re-ID model to explore model personalisation knowledge from instance-specific information. Moreover, a dual local normalisation mechanism is introduced to optimise global model generalisation to unseen novel domains.

3 Methodology

3.1 Overview

In this work, we study decentralised person Re-ID, where data from different domains (clients) are private and the global model (server) can only aggregate local models to learn generalised knowledge across domains. The aim of the proposed SKA method is to optimise the trade-off between model personalisation and generalisation in decentralised person Re-ID. Figure 1 shows an overview of the proposed SKA method.

Suppose there are N local client domains where each client contains a local non-shared private dataset. As shown in Figure 1, each local client builds two local models concurrently: A local specific model (the block in blue) and a local generalised model (the block in orange). The local specific model learns local normalisation layers (layers in pink and green) specific to each domain which are decoupled from the central aggregation whilst the other layers (layers in grey) in this model are transmitted to the server for central aggregation. In par-

allel, the local generalised model learns generalised normalisation layers (layers in yellow) which are all transmitted to the server for central aggregation. These two local models are trained simultaneously (for E local epochs) using a cross-entropy loss. Afterwards, a selective aggregation is conducted and generates an aggregated local model. Next, on the central server, a global generalised model is constructed with the collaboration of aggregated local models and then is transmitted to each local client to further facilitate local model learning. The local generalised models update all layers in the feature extraction network with the global generalised model, whilst the local specific models retain local normalisation layers but update other layers in the feature extraction network. Note, following [32, 41], local domain-specific classifiers are decoupled from the global model aggregation. This differs from the contemporary FedAvg [21] that aggregates all model components. After T global iterations of federated Re-ID model learning, a global generalised feature model is learnt on the server for out-of-the-box Re-ID deployments to unseen novel domains. Each local client model with local specific normalisation layers can then be combined with this global model for better model personalisation in Re-ID deployment to each client domain.

3.2 Learning Local Model Personalisation Knowledge

Existing decentralised person Re-ID works [32, 41] aggregate all layers of a feature representation network to construct a global generalised model and fail to tackle data heterogeneity across clients (domains), resulting in sub-optimal performance on participating client domains. To solve this problem, we propose to learn local 'personalisation' knowledge in each client domain to maximise model performance per domain Re-ID tasks, and this local personalisation knowledge is encoded in local normalisation layers of a deep learning network.

Specifically, batch normalisation (BN) [9] layers in a deep network are prevailingly used to normalise activations using data in a mini-batch which captures information about the data distribution [16]. Recently, some works [1, 16, 23] have shown that learning domain-specific batch normalisation (BN) [9] layers is beneficial to better tackle the domain shift problem across domains. In the context of federated Re-ID model learning, person images from different client domains are almost entirely in non-overlapping label spaces, which exacerbates the domain shift problem. Thus, to preserve personalisation knowledge on each participating client domain and resolve the domain shift problem, we decouple normalisation layers in each local network from central aggregation and optimise local normalisation layers specific to each client.

However, only optimising the vanilla BN layers for each client can lead to sub-optimal performance because for each local client, although person images are from the same domain, there are usually notorious misalignment of person images caused by pose change, imperfect detection, occlusion, significant variation in lighting. Thus, it is important to learn instance-specific information on each client domain whilst tackling data heterogeneity. Attention modelling is an effective solution to address this problem. To learn instance-specific channel-wise attention weights to re-calibrate features, the Squeeze-and-Excitation (SE) attention unit [8] can be incorporated into a deep network model as follows:

$$\tilde{x}^{SE} = \lambda \cdot BN(x) = (\lambda \cdot \gamma) \cdot \hat{x} + \lambda \cdot \beta \tag{1}$$

where $BN(x) = \gamma \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta$, $\mu(x), \sigma(x)$ are the mean and standard deviation computed across batch size and spatial dimensions independently for each feature channel, $\hat{x}$ is the normalised feature, λ is the learnt attentive weight, γ, β are affine parameters. Since affine

Algorithm 1 Decentralised Re-ID with Selective Knowledge Aggregation (SKA). The K clients are indexed by k, $\mathcal{N}_k$ indicates the dataset on client k, E is the number of local epochs, the number of global iteration rounds is T, η is the learning rate, w are the model weights, where w_{ls}, w_{lg} and w_l refer to model weights of a local specific model, a local generalised model and a local aggregated model respectively. Steps 1-8 are the main process, while steps 10-15, 17-23 and 25-31 are functions.

```
 1: initialize {w^k_{0,ls}, w^k_{0,lg}}^K_{k=1}
 2: for global round t = 0 to T − 1 do
 3:     for each client k = 1 to K do
 4:         w^k_{t+1,ls}, w^k_{t+1,lg} ← LocalUpdate(w^k_{t,ls}, w^k_{t,lg})
 5:         w^k_{t+1,l} ← SelectAgg(w^k_{t+1,ls}, w^k_{t+1,lg})
 6:     w_{t+1} = 1/K Σ^K_{k=1} w^k_{t+1,l}
 7:     {w^k_{t+1,ls}, w^k_{t+1,lg}}^K_{k=1} ← SelectUpdate(w_{t+1})
 8: return {w^k_{T,ls}}^K_{k=1}, w_T
 9:
10: function LOCALUPDATE(w^k_{t,ls}, w^k_{t,lg})
11:     for local epoch e = 0 to E − 1 do
12:         for mini-batch b ⊂ N_k do
13:             w^k_{t+1,ls} ← w^k_{t,ls} − η∇ℓ(w^k_{t,ls})
14:             w^k_{t+1,lg} ← w^k_{t,lg} − η∇ℓ(w^k_{t,lg})
15:     return w^k_{t+1,ls}, w^k_{t+1,lg}

16:
17: function SELECTAGG(w^k_{t+1,ls}, w^k_{t+1,lg})
18:     for layer p in w^k_{t+1,ls} do
19:         if layer p is normalisation layer then
20:             w^k_{t+1,l,p} ← w^k_{t+1,lg,p}
21:         else
22:             w^k_{t+1,l,p} ← w^k_{t+1,ls,p}
23:     return w^k_{t+1,l}

24:
25: function SELECTUPDATE(w_{t+1})
26:     for each client k do
27:         w^k_{t+1,lg} ← w_{t+1}
28:         for layer p in w_{t+1} do
29:             if p is NOT normalisation layer then
30:                 w^k_{t+1,ls,p} ← w_{t+1,p}
31:     return {w^k_{t+1,ls}, w^k_{t+1,lg}}^K_{k=1}
```

transform parameters in BN and the re-scaling parameters of the SE unit are both learning to re-calibrate features, following [17], they can be further combined as M mixture components of channel-wise affine transform:

$$\tilde{x}^{AN} = \sum_{i=1}^{M} \lambda_i \left[\gamma_i \cdot \hat{x} + \beta_i \right] \tag{2}$$

where λ_i, γ_i, β_i are the i-th mixture integrated into a compact module and learnt simultaneously, so they can learn more shared instance-specific information. We therefore combine some BN layers and attention modelling in a deep network for each local client model and decouple them from global server aggregation. This approach to learning local model personalisation knowledge brings significant benefits to model performance (see Section 4.4).

3.3   Dual Local Normalisation for Global Generalisation Knowledge

Although decoupling normalisation layers from central aggregation helps to optimise local personalisation knowledge, it will also lose some generalisable knowledge useful in constructing a better global model for unseen novel domains. A simple solution is to average local specific normalisation layers at the end of model training, but this can lead to suboptimal global generalisation because local normalisation layers are not aggregated in the iterative global model update. To compensate this degradation in the global model learning, we introduce a dual local normalisation mechanism to explicitly learn auxiliary local generalised normalisation layers.

As shown in Figure 1, in addition to the local specific model, we optimise a local generalised model (with the same network architecture) to learn local generalised normalisation

Types	Datasets	Train ID	Train Img	Test		
				Test ID	Query Img	Gallery Img
Source client domains	Market1501	751	12936	750	3368	19732
	DukeMTMC	702	16522	702	2228	17661
	CUHK03-NP	767	7365	700	1400	5328
	MSMT17	1041	30248	3060	11659	82161
Unseen novel domains	VIPeR	-	-	316	316	316
	iLIDS	-	-	60	60	60
	GRID	-	-	125	125	900
	PRID	-	-	100	100	649

Table 1: Statistics of eight person Re-ID benchmark datasets.

layers. Then, for each client, we use local generalised normalisation layers (layers in yellow in Figure 1) from a local generalised model and other layers (layers in grey in Figure 1) from a corresponding local specific model to construct a local aggregated model. The local aggregated models from different clients are then transmitted to the server for iterative central aggregation. This helps to alleviate the degradation problem because the generalised normalisation layers are iteratively updated between the client and the server and are not specific to a certain client. As a result, the global model cumulates richer generalised knowledge from local clients that can benefit significantly model performance on unseen novel domains (see Table 3 in Section 4.3). Besides, since the dual local normalisation mechanism compensate the degradation brought by learning local personalisation knowledge, it facilitates the proposed SKA to make a trade-off between learning model personalisation and generalisation. We summarise the model training process of the proposed SKA method in Algorithm 1.

4 Experiments

4.1 Datasets and Evaluation Protocol

Datasets: To evaluate the proposed SKA model, we used four large-scale person Re-ID datasets Market1501 [37], DukeMTMC-ReID [39], CUHK03-NP [14, 40] and MSMT17 [30] as local client domains. Data on each client are only used for local model training without sharing to other clients nor the server. Furthermore, we employed four smaller person Re-ID datasets VIPeR [4], iLIDS [38], GRID [21] and PRID [7] as unseen novel domains to evaluate the generalisation performance of the global model. Following [25, 32], we generated ten random training/testing splits on each smaller dataset, and on each test split, one image of each person identity was used as the query whilst the other images were used as the gallery. Table 1 summarises the dataset statistics.

Evaluation Metrics: We used Rank-1 (R1) accuracy and mean Average Precision (mAP) for Re-ID performance evaluation.

4.2 Implementation Details

Following [32, 41], we employed ResNet-50 [6] pretrained on ImageNet [3] as the feature extraction network and used two fully-connected layers as feature vectors for classification. We used attentive normalisation [17] (pretrained on ImageNet) to replace the second vanilla batch normalisation layer on each bottleneck block. In each local client, we used SGD

Method	Source	Duke		Market1501		CUHK03-NP		MSMT17	
		mAP	R1	mAP	. R1	mAP	R1	mAP	R1
local supervised		57.0	77.2	68.2	87.4	39.2	43.2	28.8	60.0
FedAvg[21]+D	AISTATS'17	50.9	70.8	56.6	81.3	25.2	27.9	25.9	55.0
FedProx[13]+D	MLSys'20	53.5	72.0	59.7	84.5	27.8	30.8	28.1	57.7
FedPav[41]	ACMMM'20	51.9	71.1	53.5	78.7	23.0	26.0	26.1	54.4
FedPav+AddData[41]	ACMMM'20	60.6	78.4	58.0	82.4	26.8	29.9	27.0	55.7
FedReID[32]	AAAI'21	52.1	68.0	60.1	80.2	-	-	-	48.4
MOON [12]+D	CVPR'21	53.1	72.7	58.1	83.6	26.4	28.5	27.2	56.6
FedBN[16]+D	ICLR'21	62.8	80.0	73.1	90.4	40.9	45.3	35.4	67.6
SKA	Ours	**66.6**	**83.7**	**78.2**	**92.7**	**48.1**	**53.0**	**42.9**	**73.8**

Table 2: Comparisons of decentralised learning Re-ID on four source client domains. Note that the conventional federated learning methods cannot be directly used for decentralised Re-ID, so '+D' means we implement them by decoupling domain-specific classifiers from central aggregation. 'FedPav+AddData' means using additional central unlabelled data in FedPav for knowledge distillation and weight adjustment. 'FedAvg+D' and 'FedPav' share the same principle but with different hyper-parameters.

Method	Source	VIPeR		iLIDS		GRID		PRID	
		mAP	R1	mAP	R1	mAP	R1	mAP	R1
FedAvg[21]+D	AISTATS'17	48.2	44.3	73.3	69.3	24.3	20.1	19.6	15.8
FedProx[13]+D	MLSys'20	47.3	43.2	74.9	71.1	29.1	24.8	31.2	26.8
FedPav[41]	ACMMM'20	49.5	44.9	72.8	68.8	25.5	21.7	37.0	31.9
FedPav+AddData[41]	ACMMM'20	49.6	45.3	73.1	69.0	28.7	24.2	34.4	28.5
FedReID[32]	AAAI'21	-	46.2	-	69.7	-	24.2	-	-
MOON [12]+D	CVPR'21	49.1	45.1	73.7	69.7	28.0	24.0	33.5	29.2
FedBN[16]+D	ICLR'21	47.9	43.5	72.3	68.2	25.2	21.2	31.1	26.5
SKA	Ours	**53.9**	**49.8**	**76.0**	**72.7**	**36.7**	**32.2**	**49.7**	**45.0**

Table 3: Comparisons of decentralised learning Re-ID on four unseen novel domains.

optimiser with Nesterov momentum 0.9 and weight decay 5e-4 and set the learning rate to 0.01 for the feature extraction network and 0.1 for the classifier, which were decayed by 0.1 every 40 global epochs. We empirically set batch size to 32, maximum global iterations $T = 100$, maximum local optimisation epoch $E = 1$. For fair comparison, all the reproduced methods including local supervised baseline are based on ResNet-50 backbone following same training process including optimiser, learning rate, global iteration rounds and local epochs. All of the experiments are implemented in Python and PyTorch.

4.3 Comparisons with the State-of-the-Art Decentralised Methods

We compared the proposed SKA with two state-of-the-art decentralised person Re-ID methods, namely FedReID [32] and FedPav [41], and four state-of-the-art federated learning methods, namely FedAvg [21], FedProx[13], FedBN [16] and MOON [12]. Note that the conventional federated learning methods cannot be directly used for decentralised person Re-ID, so we implement them based on the federated Re-ID backbone model [41], where domain-specific classifiers are not used for central aggregation. Table 2 gives results on source client domains and Table 3 for results on unseen novel domains.

Table 2 shows that the proposed SKA has significantly better Re-ID performance than

Components	Seen domains				Unseen domains			
	Duke		Market		VIPeR		iLIDS	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1
Baseline	50.9	70.8	56.6	81.3	48.2	44.3	73.3	69.3
Baseline+AN	58.7	78.1	64.2	87.9	54.2	50.2	77.0	74.0
Baseline+LSN	$62.8^{\dagger}$	$80.0^{\dagger}$	$73.1^{\dagger}$	$90.4^{\dagger}$	47.9	43.5	72.3	68.2
Baseline+LSN+Dual	$62.8^{\dagger}$	$80.0^{\dagger}$	$73.1^{\dagger}$	$90.4^{\dagger}$	48.8	44.6	75.2	71.3
Baseline+LSN+AN	66.6*	83.7*	78.2*	92.7*	51.8	47.8	73.1	69.3
Baseline+LSN+AN+Dual	66.6*	83.7*	78.2*	92.7*	53.9	49.8	76.0	72.7

Table 4: Evaluating component effectiveness on seen source domains and unseen novel domains. *,† 'Dual' does not affect model personalisation on seen source domains as local generalised normalisation layers are only learnt to construct the global generalised model.

all other models on individual client domains. Specifically, state-of-the-art decentralised Re-ID methods (FedReID [32] and FedPav [41]) perform slightly worse on average than the baseline local supervised model which trains the backbone model on each local domain in a conventional centralised learning way. In contrast, SKA significantly improves over the baseline local supervised model by 5%~10% in R1 and 10+% in mAP on all test datasets. Compared with the state-of-the-art federated learning methods (FedAvg [21], FedProx [13], FedBN [16] and MOON [12]), SKA performs better than FedBN in R1 by 4.98% and mAP by 5.9% on average, and significantly outperforms FedAvg in R1 by 17.05% and mAP by 19.3% on average. Table 3 also shows that the proposed SKA achieves notable generalisation performance improvement on unseen novel domains by outperforming the second best model in R1 by 3.6% (VIPeR), 1.6%(iLIDS), 7.4%(GRID), 13.1% (PRID) respectively, and in mAP by 4.3% (VIPeR), 1.1%(iLIDS), 7.6%(GRID), 12.7% (PRID) respectively. These results show a very clear superiority of the proposed SKA model over the state-of-the-art methods. Besides, although centralised and decentralised methods are not directly comparable, the proposed SKA method performs closely to some centralised generalisable Re-ID methods. For example, compared with the state-of-the-art centralised SNR [10] method, on Duke, SKA yields 66.6% vs. SNR yields 73.2% in terms of mAP; on VIPeR, SKA yields 53.9% vs. SNR yields 55.1% in terms of mAP.

4.4 Ablation Studies

Component Effectiveness Evaluation. In Table 4, 'Baseline' means the federated Re-ID learning baseline model [32, 41], 'AN' means incorporating Attentive Normalisation into the vanilla BN layers in the feature extraction model, 'LSN' means local specific vanilla BN normalisation, 'Dual' means the dual local normalisation mechanism. On unseen novel domains, 'Baseline+LSN' and 'Baseline+LSN+AN' are by averaging the normalisation layers at the end of model training to construct the global generalised model. Table 4 shows: (1) LSN can significantly improve model personalisation performance on seen source domains. However, it degrades clearly model generalisation performance on unseen novel domains; (2) AN can improve model personalisation and generalisation performance on both seen and unseen domains; (3) Learning local personalisation knowledge with LSN+AN in SKA brings the best model personalisation performance on seen domains; (4) The dual local normalisation mechanism can compensate the degradation brought by learning local personalisation knowledge, resulting in better global model generalisation. It improves 'Base-

Variants	Duke	Market
SKA	83.7	92.7
SKA w/o LSN	78.1	87.9
SKA w/o AN	80.0	90.4
SKA w/o AN + SE	82.0	91.2
SKA w/o AN + CBAM	80.5	91.1

Table 5: Evaluating variants of learning local personalisation knowledge (R1).

Variants	VIPeR	iLIDS
SKA w/ Dual	49.8	72.7
SKA w/o Dual + Avg	47.8	69.3
SKA w/o Dual + FeatConcat	47.2	71.8
SKA w/o Dual + RandImg	6.6	21.8

Table 6: Evaluating variants of dual local normalisation (R1).

line+LSN' by approximately 1% and 2% on VIPeR and iLIDS respectively, while improves 'Baseline+LSN+AN' by approximately 2% on both VIPeR and iLIDS. And we can see that, on unseen domains, 'Baseline+LSN+AN+Dual' performs closely to 'Baseline+AN' which does not consider learning local personalisation knowledge, while on seen source domains, 'Baseline+LSN+AN+Dual' performs significantly better than 'Baseline+AN'. This further verifies that 'Dual' can help SKA to optimise the trade-off between model personalisation and generalisation.

Evaluating Variants of Learning Local Personalisation Knowledge. In Table 5, we tested some variants of learning local personalisation knowledge. 'SE' means using the SE [8] unit in the backbone model and 'CBAM' means using the CBAM [51] unit in the backbone model. Table 5 shows that learning local specific normalisation and attentive knowledge facilitates better model personalisation. Overall, SKA performs better than other variants.

Evaluating Variants of Dual Local Normalisation. In Table 6,we tested different strategies to construct normalisation layers in the global generalised model. 'Avg' means averaging local specific normalisation layers at the end of model training. 'FeatConcat' means using each local specific normalisation layers to extract features and concatenating these features for person matching. 'RandImg' means using ImageNet pretrained weights to re-initialise normalisation layers. Table 6 shows that the proposed dual local normalisation mechanism performs better than other variants for constructing a global generalised model.

5   Conclusions

In this work, we proposed a new Selective Knowledge Aggregation (SKA) approach to decentralised person Re-ID. The key idea is to learn local batch and attentive normalisation layers specific to each domain for improving model personalisation in Re-ID on seen client domains, and to use a dual local normalisation mechanism for improving model generalisation in Re-ID on unseen novel domains. Extensive experiments on eight person Re-ID datasets show the superiority of the proposed SKA approach over the state-of-the-art decentralised methods on both seen source client domains and unseen novel domains.

References

[1] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 7354–7362, 2019.

[2] Seokeon Choi, Taekyung Kim, Minki Jeong, Hyoungseob Park, and Changick Kim. Meta batch-instance normalization for generalizable person re-identification. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2021.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009.

[4] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, pages 262–275. Springer, 2008.

[5] Chaoyang He and Murali Annavaram. Group knowledge transfer: Federated learning of large cnns at the edge. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, (33), 2020.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[7] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. Person Re-Identification by Descriptive and Discriminative Classification. In *Proc. Scandinavian Conference on Image Analysis (SCIA)*, 2011.

[8] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[9] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456. PMLR, 2015.

[10] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3143–3152, 2020.

[11] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pages 5132–5143. PMLR, 2020.

[12] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10713–10722, 2021.

[13] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Conference on Machine Learning and Systems*, 2020.

[14] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014.

[15] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2018.

[16] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*, 2021.

[17] Xilai Li, Wei Sun, and Tianfu Wu. Attentive normalization. *arXiv preprint arXiv:1908.01259*, 2019.

[18] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*, 2020.

[19] Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. *arXiv preprint arXiv:2006.07242*, 2020.

[20] Chen Change Loy, Chunxiao Liu, and Shaogang Gong. Person re-identification by manifold ranking. In *2013 IEEE International Conference on Image Processing*, pages 3567–3571. IEEE, 2013.

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

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

[23] Seonguk Seo, Yumin Suh, Dongwan Kim, Jongwoo Han, and Bohyung Han. Learning to optimize domain specific normalization for domain generalization. *arXiv preprint arXiv:1907.04275*, 3(6):7, 2019.

[24] Tao Shen, Jie Zhang, Xinkang Jia, Fengda Zhang, Gang Huang, Pan Zhou, Kun Kuang, Fei Wu, and Chao Wu. Federated mutual learning. *arXiv preprint arXiv:2006.16765*, 2020.

[25] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 719–728, 2019.

[26] Canh T Dinh, Nguyen Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems*, 33, 2020.

[27] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *arXiv preprint arXiv:2103.00710*, 2021.

[28] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. In *International Conference on Learning Representations*, 2020.

[29] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in Neural Information Processing Systems*, 33, 2020.

[30] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 79–88, 2018.

[31] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[32] Guile Wu and Shaogang Gong. Decentralised learning from independent multi-domain labels for person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[33] Guile Wu, Xiatian Zhu, and Shaogang Gong. Spatio-temporal associative representation for video person re-identification. In *British Machine Vision Conference*, page 278, 2019.

[34] Guile Wu, Xiatian Zhu, and Shaogang Gong. Tracklet self-supervised learning for unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12362–12369, 2020.

[35] Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, and Wanli Ouyang. Attention-aware compositional network for person re-identification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2119–2128, 2018.

[36] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[37] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015.

[38] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *BMVC*, volume 2, pages 1–11, 2009.

[39] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3754–3762, 2017.

[40] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1318–1327, 2017.

[41] Weiming Zhuang, Yonggang Wen, Xuesen Zhang, Xin Gan, Daiying Yin, Dongzhan Zhou, Shuai Zhang, and Shuai Yi. Performance optimization of federated person re-identification via benchmark analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 955–963, 2020.