ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447543

META-KNOWLEDGE ENHANCED DATA AUGMENTATION FOR FEDERATED PERSON RE-IDENTIFICATION

Chunli Song[1,2,4†] *Xiaohua Chen*[2,3,4] *Wenqiu Zhu*[1] *Yucan Zhou*[2,4*] *Xiaoyan Gu*[2,3,4] *Bo Li*[2,3,4]

[1] Hunan University of Technology
[2]Institute of Information Engineering, Chinese Academy of Sciences
[3]School of Cyber Security, University of Chinese Academy of Sciences
[4]Key Lab of Cyberspace Security Defense

ABSTRACT

Recently, federated learning has been introduced into person re-identification (Re-ID) to avoid personal image leakage in traditional centralized training. To address the key issue of statistic heterogeneity in different clients, several optimization methods have been proposed to alleviate the bias of the local models. However, besides statistic heterogeneity, feature heterogeneity (e.g., various angles, different illuminations) in different clients is more challenging in federated Re-ID. In this paper, we propose a meta-knowledge enhanced data augmentation method, where the global cross semantic feature transformations are provided to each client to perform local infinite augmentation to reduce the feature difference in different clients. Specifically, to capture the cross semantic feature transformations in each client, we calculate the covariance matrix of features with the local dataset as the transferable meta-knowledge. Then, this local meta-knowledge is propagated to the server for global aggregation. Subsequently, the aggregated meta-knowledge is sent back to each client for infinite data augmentation. Moreover, since the covariance matrix indicates variations in a client, we design a variation-balanced aggregation to replace the traditional data-size-balanced aggregation. To imitate the more challenging scenario of feature heterogeneity, we focus on the federated-by-camera setting to conduct experiments, where images collected in a camera are regarded as the dataset of a client. Extensive experimental results show that our method outperforms other state-of-the-art methods. Code is available at https://github.com/songchunli1999/MEDA.

Index Terms— Person re-identification, Federated data augmentation, Statistic heterogeneity, Feature heterogeneity

1. INTRODUCTION

Person re-identification (Re-ID) is an essential process in security and surveillance, which retrieves images of an individual from a gallery set. To obtain a robust Re-ID model, images captured by different cameras are transmitted to a server for centralized training. However, with the increasingly growing social consensus on data privacy, directly transmitting sensitive personal images becomes infeasible. Fortunately, federated learning (FL) makes it possible to achieve privacy-preserving training by transmitting models rather than sensitive data.

Inspired by the idea of FL, Fedpav [1] is a pioneer in designing a federated Re-ID (FedReID) framework. Similar to the traditional FL, statistic heterogeneity is also an important issue in FedReID, where the non-independent and identically distributed local datasets will lead to a dispersed global model (i.e., the global model is dominated by a couple of local models). To alleviate the bias of the global model, Zhuang et al. [2] propose two strategies of Cosine Distance Weight (CDW) and Client Clustering (CC). CDW [2] quantifies the change degree of each local model and assigns larger weights to changing larger local models for aggregation to make the global model absorb information from all clients. CC [2] maintains multiple global models, where each is generated by aggregating similar local models to preserve information from all clients. Besides the traditional statistic heterogeneity, feature heterogeneity (e.g., various angles, different illuminations) in different clients is more challenging for FedReID since the purpose of ReID is to retrieve an individual in various conditions.

Data augmentation is a straightforward way to reduce feature heterogeneity [3, 4, 5, 6, 7, 8]. Since augmentation with local data is limited, a better option in FL is to augment new samples with information from other clients. Thus, FedMix [3] and XorMixFL [4] generalize mixup [6] into FL to generate new samples by mixing averaged data across clients in the input level. FedFTG [7] trains a local generator to model local data distribution for each client and propagates them to the server to generate pseudo-data to fine-tune the global model. However, the generated data may contain information about the original images, which is impractical for privacy-sensitive Re-ID tasks. Subsequently, FedFA [8] proposes to conduct data augmentation at the convolutional layers, where each channel of the feature is enlarged or contracted with its local feature statistics and the universal ones. This method is secure and effective, but scaling the feature map is still a simple semantic transformation. Besides, the fusion of the local and the universal feature statistics is too complicated.

We argue that data augmentation using the cross semantic feature transformations from other clients can generate more diversified samples, which makes the model more robust to feature heterogeneity. For example, as shown in Fig. 1, sam-

† This work was done during Chunli Song's internship at Institute of Information Engineering.
* Corresponding author: zhouyucan@iie.ac.cn.

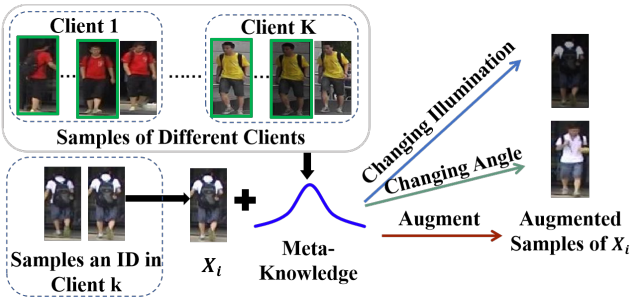

Fig. 1: Augmented samples with enriched transformations.

ples of an ID in client k are limited in illumination and angles. Since client 1 contains the angle transformation (from the back to the front) and client K contains the illumination transformation (from light to dark), we can transfer these transformations to each sample $\mathbf{X}_i$ in client k to generate new samples with different angles and illuminations.

In this paper, we propose a **M**eta-knowledge **E**nhanced **D**ata **A**ugmentation (MEDA) for FedReID, where the global cross semantic feature transformations are provided to each client for local augmentation to reduce the feature heterogeneity. Specifically, to capture the cross semantic feature transformations in each client, we calculate the covariance matrix of features with the local dataset as the transferable meta-knowledge. Then, this local meta-knowledge is proposed to the server for global aggregation. Subsequently, the aggregated meta-knowledge is sent back to each client for infinite data augmentation. Moreover, since the covariance matrix indicates variations in a client, we design a variation-balanced aggregation to replace the traditional data-size-balanced aggregation. For the experiment, there are two commonly used FedReID settings [1]: federated-by-identity [1] (which regards images from several identities as the local dataset of a client) and federated-by-camera (where images from the same camera make up a local dataset). To imitate the more challenging scenario of feature heterogeneity, we focus on the federated-by-camera setting. Extensive experiments show that our method outperforms other state-of-the-art methods. Furthermore, since the meta-knowledge only contains transformation direction at the semantic feature level, transferring it does not involve any information about the original images. The main contributions of this paper can be summarized as follows:

- We propose a meta-knowledge enhanced data augmentation method to provide each client with the global cross semantic feature transformations to alleviate the feature heterogeneity under the premise of privacy protection.
- We design a variation-balanced aggregation to replace the traditional data-size-balanced aggregation in federated learning to increase the weights of the local models with high diversity to improve the global model.
- Experimental results on commonly used Re-ID datasets have shown that our method can significantly outperform other state-of-the-art FedReID methods in federated-by-camera. Moreover, our proposed method can also achieve better results in federated-by-identity.

[1]Fedpav[1] also proposes federated-by-dataset, which is, in fact, the federated-by-identity with more data, so we use federated-by-identity to represent both federated-by-identity and federated-by-dataset here.

2. METHOD

In this section, we will describe our MEDA in detail. Given a training dataset $\mathbf{D}_k = \{\mathbf{X}_i, y_i\}_{i=1}^{N_k}$ for client $k \in [1, K]$, where $y_i \in [1, C_k]$ is the identity (ID) of $\mathbf{X}_i$, C_k and N_k are the number of IDs and instances in client k. We perform T rounds of communication in FL with E local epochs and batch size B in local training. As shown in Fig. 2, our approach contains two stages. In stage I, we perform traditional FedReID: in a communication round $t \in [1, T_1]$, the server delivers the global model $\boldsymbol{\theta}_g^{t-1}$ to each client. Then, the client trains its local model $\boldsymbol{\theta}_k^{t-1}$ and classifier $\mathcal{H}_k^{t-1}$. After that, each client passes the optimized $\boldsymbol{\theta}_k^{t-1}$ to the server for data-size-balanced aggregation to obtain $\boldsymbol{\theta}_g^t$ for the training in the next $t+1$ communication round. In stage II, we use MEDA to alleviate feature heterogeneity, which consists of meta-knowledge generation, variation-balanced aggregation, and local data augmentation. Specifically, for a communication round $t \in (T_1, T]$, each client k calculates its local meta-knowledge $\mathbf{M}_k^t$ and passes it to the server. Subsequently, the server performs variation-balanced aggregation on $\{\mathbf{M}_k^t\}_{k=1}^K$ to obtain the global meta-knowledge $\mathbf{M}_g^t$ and sends it to all clients for local data augmentation and local training.

2.1. Meta-Knowledge Generation

Inspired by ISDA [9] and RISDA [10], the class semantic feature transformations can be captured by the feature covariance matrix. Following them, we take the features covariance matrix as the local meta-knowledge on the client side. Since all the IDs share the similar transformation space (e.g., angle, pose, and illumination variations), we calculated the covariance matrix with all samples in a client instead of the ID-specific one in ISDA [9] and RISDA [10]. Therefore, the client-specific covariance matrix contains more semantic transformations.

To obtain the local meta-knowledge, the server first sends the global model to each client. Then, for each sample $(\mathbf{X}_i, y_i)$ in client k, we extract its feature $\boldsymbol{f}_i$ with $\boldsymbol{f}_i = \boldsymbol{\theta}_g^t(\mathbf{X}_i)$. Since we experimentally observe that randomly selected B features $\{\boldsymbol{f}_i\}_{i=1}^B$ contains enough transformations to approximate those of the entire dataset $\mathbf{D}_k$. So, we use $\{\boldsymbol{f}_i\}_{i=1}^B$ to calculate the local meta-knowledge $\mathbf{M}_k^t$ for simplicity:

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^B \boldsymbol{f}_i}{B}, \tag{1}$$

$$\mathbf{M}_{k(m,n)}^t = \frac{\sum_{i=1}^B (\boldsymbol{f}_i^m - \boldsymbol{\mu}_k^m)(\boldsymbol{f}_i^n - \boldsymbol{\mu}_k^n)}{B-1}, \tag{2}$$

where m, n means the m-th and n-th dimension of the feature. Then, each client propagates its $\mathbf{M}_k^t$ to the server.

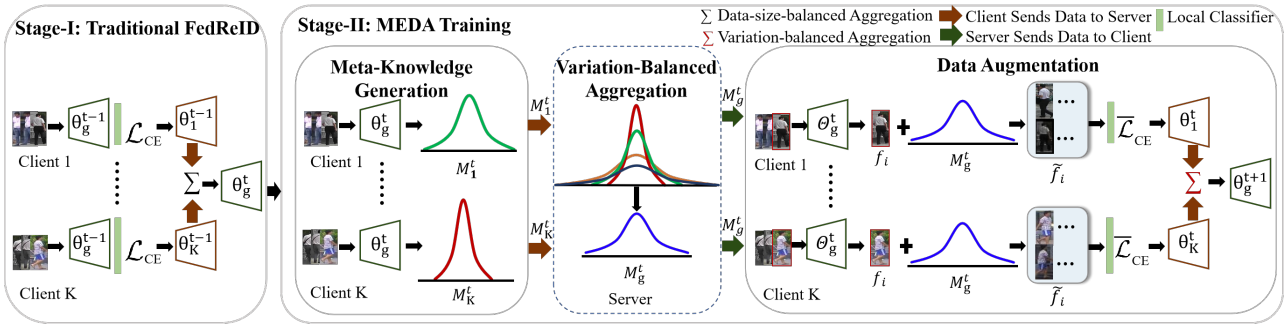

Fig. 2: The framework of MEDA. MEDA contains two stages. In stage I, we perform traditional FedReID training. In stage II, we perform MEDA training, which consists of meta-knowledge generation, variation-balanced aggregation, and data augmentation.

2.2. Variation-Balanced Aggregation

When server receives K local meta-knowledge $\{\mathbf{M}_k^t\}_{k=1}^K$, it performs aggregation to get the global meta-knowledge $\mathbf{M}_g^t$. We argue that the client with more diversified data should play a more important role in aggregation. Therefore, we propose variation-balanced aggregation, which uses the maximum eigenvalue e_k^t of $\mathbf{M}_k^t$ to evaluate the diversity degree of local data. So, the weight a_k^t for $\mathbf{M}_k^t$ in aggregation is designed as:

$$a_k^t = \frac{e_k^t}{\sum_{j=1}^K e_j^t}. \tag{3}$$

With the weight a_k^t, the sever can obtain the global $\mathbf{M}_g^t$ with:

$$\mathbf{M}_g^t = \textstyle\sum_{k=1}^K a_k^t \mathbf{M}_k^t, \tag{4}$$

where $\mathbf{M}_g^t$ contains the overall transformations in each client. Then, $\mathbf{M}_g^t$ is sent to all the clients for local data augmentation.

2.3. Data Augmentation

With the global meta-knowledge $\mathbf{M}_g^t$, each client can use the overall semantic transformations to conduct augmentation. Specifically, we model a specific sample x_i in client k as a Gaussian distribution $\tilde{\boldsymbol{f}}_i$, where the original feature $\boldsymbol{f}_i$ acts as the mean and $\mathbf{M}_g^t$ as the covariance, i.e.,

$$\tilde{\boldsymbol{f}}_i \sim N(\boldsymbol{f}_i, \beta \mathbf{M}_g^t), \tag{5}$$

where β is a coefficient to control the strength of data variation in augmentation. With $\tilde{\boldsymbol{f}}_i$, client k can generate new instances by sampling from the distribution $\tilde{\boldsymbol{f}}_i$. However, this explicit augmentation limits the diversity of the generated samples.

An ideal way is to generate as much data as possible. However, the increasing sampling frequency can lead to additional calculations. Fortunately, ISDA [9] and RISDA [10] have made infinite implicit data augmentation possible, which not only enhances the diversity of samples but also saves storage space and training time. Following them, we conduct infinite implicit data augmentation for each sample x_i from the distribution $\tilde{\boldsymbol{f}}_i$. Then, we consider all possible enhanced samples by estimating the upper bound of the cross-entropy loss with:

$$\overline{\mathcal{L}}_{CE_{\tilde{\boldsymbol{f}}_i}} = -\log \frac{e^{\hat{y}_i^{y_i}}}{\sum_{j=i}^{C_k} e^{\hat{y}_i^j + \frac{1}{2}\beta(\boldsymbol{w}_j^T - \boldsymbol{w}_{y_i}^T)\boldsymbol{M}_g^t(\boldsymbol{w}_j - \boldsymbol{w}_{y_i})}}, \tag{6}$$

where $\hat{y}_i^j$ is the j-th logits of x_i and $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_{C_k}]$ is the last fully connected layer weight matrix. So, we use implicit data augmentatin in Eq.(6) to train K local clients and then send the local model $\{\boldsymbol{\theta}_k^t\}_{k=1}^K$ back to the server.

Finally, on the server side, we also perform variation-balanced aggregation with the weight a_k^t in Eq.(3) to obtain the global model for the next communication round with:

$$\boldsymbol{\theta}_g^{t+1} = \textstyle\sum_{k=1}^K a_k^t \boldsymbol{\theta}_k^t. \tag{7}$$

3. EXPERIMENTS

3.1. Experimental Settings

Datasets. We conduct experiments on nine Re-ID datasets: MSMT17 [11], DukeMTMC-reID [12], Market-1501 [13], CUHK03-NP [14], PRID2011 [15], CUHK01 [16], VIPeR [17], 3DPeS [18] and iLIDS-VID [19].

Evaluation Metrics. We use mean precision (mAP) and cumulative matching accuracy (CMC) Rank-1, Rank-5, and Rank-10 as evaluation metrics to evaluate the performance.

Implementation Details. We adopt ResNet-50 as our backbone and initialize it with the weights pre-trained on ImageNet. We set B = 32, E = 1, T = 300, and T_1 = 150. We adopt the SGD optimizer with a momentum 0.9 and a weight decay 5e-4.

Comparing Methods. We compare our MEDA with Fedpav [1] and CDW [2]. We also report the results of centralized training (model trained with data collected from all the clients) and local training (model trained with local data).

3.2. Federated-by-Camera

In the more challenging federated-by-camera setting, images from the same camera make up a local dataset. Since some datasets only a few hundred images in each camera, which is insufficient to train ResNet-50, so we only use MSMT17 [11], Market-1501 [13], and CUHK03-NP [14] to conduct experiments. To show the effectiveness of retrieving images with large variations, the global test data are utilized which contain images from all the cameras. Then, for local training, we report the performance of the best local model on the global test data. Furthermore, we calculate the mean and variance of the results with three different seeds.

Table 1: Results on three datasets under federated-by-camera setting. $*$ means performance of the best local model.

Dataset	Clients		Method	Rank-1	Rank-5	Rank-10	mAP
MSMT17	15	Reference	Centralized Training	47.35±0.16	60.66±0.03	65.91±0.17	26.81±0.11
			Local Training*	13.23±0.13	20.64±0.56	24.89±0.40	4.07±0.05
		FedReID Method	Fedpav [1]	23.76±0.67	33.53±0.95	38.75±0.93	9.12±0.33
			CDW [2]	24.48±0.36	34.69±0.55	39.80±0.35	9.51±0.23
			MEDA	**24.84±0.61**	**34.88±0.64**	**40.32±0.52**	**9.75±0.25**
Market-1501	6	Reference	Centralized Training	88.71±0.39	95.43±0.33	97.35±0.20	72.69±0.71
			Local Training*	51.97±0.77	66.14±0.76	72.49±0.98	27.04±0.54
		FedReID Method	Fedpav [1]	61.62±0.93	76.66±2.01	82.91±2.59	35.01±0.74
			CDW [2]	62.31±0.31	76.88±0.03	82.83±0.73	36.33±1.98
			MEDA	**66.12±0.46**	**79.15±0.65**	**83.63±0.45**	**41.44±0.57**
CUHK03-NP	2	Reference	Centralized Training	49.71±0.19	68.78±0.69	76.96±0.06	45.82±1.14
			Local Training*	8.21±1.05	14.64±0.87	19.42±1.01	7.47±0.64
		FedReID Method	Fedpav [1]	11.18±0.15	19.92±0.93	26.19±1.51	10.88±0.14
			CDW [2]	11.57±0.22	19.14±0.61	25.18±1.14	10.96±0.28
			MEDA	**12.59±0.49**	**20.71±0.57**	**27.33±0.60**	**11.70±0.43**

Table 2: Results on nine datasets under federated-by-identity setting.

	Method	MSMT17	DukeMTMC	Market-1501	CUHK03-NP	PRID2011	CUHK01	VIPeR	3DPeS	iLIDS-VID
Reference	Centralized Training	54.6	84.2	91.7	64.0	80.0	89.7	65.5	82.1	80.6
	Local Training	49.6	80.1	88.9	49.3	55.0	69.0	27.5	65.4	52.0
FedReID Method	Fedpav [1]	41.0	74.3	83.4	31.7	**37.7**	73.4	**48.1**	69.2	79.9
	CDW [2]	43.8	73.0	82.3	**33.6**	34.0	**78.3**	43.9	70.3	79.5
	MEDA	**44.9**	**75.7**	**84.0**	31.6	34.0	72.4	43.0	**73.1**	**83.6**

Table 1 shows the results on the three datasets. Our method achieves state-of-the-art performances in federated-by-camera setting. Especially on Market-1501, MEDA outperforms Fedpav [1] and CDW [2] by 4.50% and 3.81% in the Rank-1 accuracy, which illustrates the effectiveness of MEDA. However, the improvements on MSMT17 are limited. This is because it contains 15 clients, and some local dataset only contains a few hundred images, resulting in limited semantic transformation. Thus, the diversity of augmented data is limited.

3.3. Federated-by-Identity

For the federated-by-identity setting, we use all the nine datasets, where each dataset is a local dataset. For testing, since it is intractable to construct a global test dataset, we follow Fedpav [1] to use the local test data in each client to evaluate the performance. As shown in Table 2, we can conclude that: (1) Our MEDA performs better than Fedpav and CDW on most datasets. (2) MEDA is more effective for small datasets with an improvement of 2.8% and 4.1% on 3DPeS and iLIDS-VID, respectively. (3) The performance on PRID2011 and VIPeR drops greatly compared to FedPav. The reason may be that all the clients use the same β in Eq.(5) for augmentation. The variation may be too strong makes the generated data mixing some IDs in these two datasets, and we will explore a proper β for them in the future.

3.4. Ablation Study

To clearly show the effectiveness of data augmentation (DA) and variation-balanced aggregation (VBA) in MEDA, we report the results of baseline(Fedpav)+VBA, baseline+DA, and baseline+VBA+DA on Market-1501. As shown in Fig.3 (a), we can see that baseline+VBA and baseline+DA outperform baseline by 0.80% and 3.70%. In addition, baseline+VBA+DA outperforms baseline by a large margin of 4.39%. Fig.3 (b) shows the impact of varying values of β. It can be observed that the results exhibit slight fluctuations when different values of β are employed and the performance can achieve the best when β=1. Nevertheless, all the results surpass Fedpav, which suggests that MEDA is robust to β.

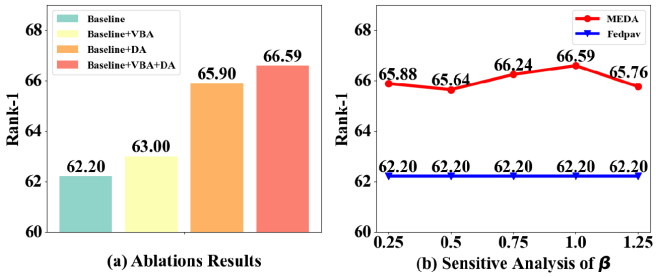

Fig. 3: Ablation study on Market-1501.

4. CONCLUSION

In this paper, we propose a meta-knowledge enhanced data augmentation method, where the global cross semantic feature transformations are provided to each client to perform local infinite augmentation to reduce the feature difference in different clients. Moreover, we design a variation-balanced aggregation to increase the weights of the local models with rich diversity to improve the global model. Extensive experimental results show that our method can significantly outperform other state-of-the-art methods. In the future, we will consider more refined data augmentation to further improve performance, such as personalized data augmentation for different clients based on local data diversity.

Acknowledgements. This work was supported by the National Natural Science Foundation of China under Grant 62006221, the Scientific Research Fund of Hunan Provincial Education Department (23A0423), the Grants 2022YFB3103503 and No. XDC02050200.

5. REFERENCES

[1] 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*, 2020, pp. 955–963.

[2] Weiming Zhuang, Xin Gan, Yonggang Wen, and Shuai Zhang, "Optimizing performance of federated person re-identification: Benchmarking and analysis," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 1s, pp. 1–18, 2023.

[3] Tehrim Yoon, Sumin Shin, Sung Ju Hwang, and Eunho Yang, "Fedmix: Approximation of mixup under mean augmented federated learning," in *International Conference on Learning Representations*, 2021.

[4] MyungJae Shin, Chihoon Hwang, Joongheon Kim, Jihong Park, Mehdi Bennis, and Seong-Lyun Kim, "Xor mixup: Privacy-preserving data augmentation for one-shot federated learning," in *Proceedings of the International Conference on Machine Learning*, 2020.

[5] Xiaohua Chen, Yucan Zhou, Dayan Wu, Chule Yang, Bo Li, Qinghua Hu, and Weiping Wang, "Area: Adaptive reweighting via effective area for long-tailed classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19277–19287.

[6] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.

[7] Lin Zhang, Li Shen, Liang Ding, Dacheng Tao, and Ling-Yu Duan, "Fine-tuning global model via data-free knowledge distillation for non-iid federated learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10174–10183.

[8] Tailin ZHOU, Jun Zhang, and Danny Tsang, "FedFA: Federated learning with feature alignment for heterogeneous data," in *International Conference on Learning Representations*, 2023.

[9] Yulin Wang, Gao Huang, Shiji Song, Xuran Pan, Yitong Xia, and Cheng Wu, "Regularizing deep networks with semantic data augmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3733–3748, 2021.

[10] Xiaohua Chen, Yucan Zhou, Dayan Wu, Wanqian Zhang, Yu Zhou, Bo Li, and Weiping Wang, "Imagine by reasoning: A reasoning-based implicit semantic data augmentation for long-tailed classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, pp. 356–364.

[11] 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*, 2018, pp. 79–88.

[12] 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*, Oct 2017.

[13] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1116–1124.

[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*, 2014.

[15] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof, "Person re-identification by descriptive and discriminative classification," in *Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17*. Springer, 2011, pp. 91–102.

[16] Wei Li, Rui Zhao, and Xiaogang Wang, "Human reidentification with transferred metric learning," in *Asian Conference on Computer Vision*. Springer, 2013, pp. 31–44.

[17] Douglas Gray and Hai Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features.," *European Conference on Computer Vision*, vol. 2008, pp. 262–275, 2008.

[18] Davide Baltieri, Roberto Vezzani, and Rita Cucchiara, "3dpes: 3d people dataset for surveillance and forensics," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, 2011, pp. 59–64.

[19] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang, "Person re-identification by video ranking," in *European Conference on Computer Vision*. Springer, 2014, pp. 688–703.