

Diversity-Authenticity Co-constrained Stylization for Federated Domain Generalization in Person Re-identification

Fengxiang Yang^{1,2}, Zhun Zhong³, Zhiming Luo^{1*}, Yifan He⁴, Shaozi Li¹, Nicu Sebe²

¹Department of Artificial Intelligence, Xiamen University, China

²Department of Information Engineering and Computer Science, University of Trento, Italy

³School of Computer Science, University of Nottingham, UK

⁴Reconova Technologies Co., Ltd., China

Abstract

This paper tackles the problem of federated domain generalization in person re-identification (FedDG re-ID), aiming to learn a model generalizable to unseen domains with decentralized source domains. Previous methods mainly focus on preventing local overfitting. However, the direction of diversifying local data through stylization for model training is largely overlooked. This direction is popular in domain generalization but will encounter two issues under federated scenario: (1) Most stylization methods require the centralization of multiple domains to generate novel styles but this is not applicable under decentralized constraint. (2) The authenticity of generated data cannot be ensured especially given limited local data, which may impair the model optimization. To solve these two problems, we propose the Diversity-Authenticity Co-constrained Stylization (DACS), which can generate diverse and authentic data for learning robust local model. Specifically, we deploy a style transformation model on each domain to generate novel data with two constraints: (1) A diversity constraint is designed to increase data diversity, which enlarges the Wasserstein distance between the original and transformed data; (2) An authenticity constraint is proposed to ensure data authenticity, which enforces the transformed data to be easily/hardly recognized by the local-side global/local model. Extensive experiments demonstrate the effectiveness of the proposed DACS and show that DACS achieves state-of-the-art performance for FedDG re-ID. Project: https://github.com/FlyingRoastDuck/DACS_official.git

1 Introduction

Person Re-identification (re-ID) aims at retrieving target pedestrian in a non-overlapped camera system, which can largely benefit the smart city construction, *e.g.*, finding lost children or escaped criminals. It is reported that deep-based methods (He et al. 2016; Huang et al. 2019; Wang et al. 2018; Ye et al. 2021; Sun et al. 2018) have drastically promoted the development of re-ID. However, these modern methods still suffer from the domain shift caused by different domains, leading to unsatisfactory performance when deployed in novel domains.

*Corresponding author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

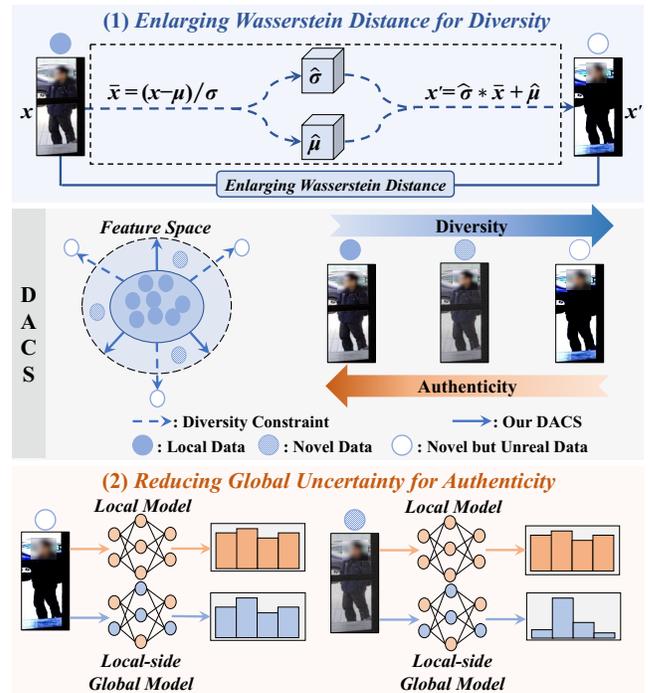


Figure 1: Schematic illustration of the proposed Diversity-Authenticity Co-constrained Stylization (DACS). Middle: We introduce a style transformation model (STM) for each domain to hallucinate novel data with two constraints. (1) Diversity Constraint: STM is encouraged to generate diverse images. (2) Authenticity Constraint: We ensure data authenticity by enforcing the transformed data to be easily / hardly recognized by local-side global model / local model.

Recent studies (Zhao et al. 2021; Song et al. 2019; Dai et al. 2021) attempt to solve domain shift issue by designing domain generalizable (DG) algorithms, where they learn generalized re-ID models by training on several labeled source domains. Despite their success, all of them require the centralization of training data, raising data privacy concerns. One promising solution is federated learning (McMahan et al. 2017), which aims to learn a generalized model by accumulating (*e.g.*, averaging) the knowledge of models independently trained on each domain. In such a learning paradigm, the data privacy issue can be largely alleviated.

Nevertheless, it is hard to achieve good generalization without centralizing training data due to the data heterogeneity under different environment conditions (Wu and Gong 2021). Moreover, since re-ID is an open-set (Panareda Busto and Gall 2017) retrieval problem, where each domain has completely different pedestrians, the optimization of federated domain generalization in re-ID (FedDG re-ID) is more challenging than other closed-set tasks like image classification (Li, He, and Song 2021) or segmentation (Liu et al. 2021).

To tackle the FedDG re-ID problem, recent works focus on alleviating the overfitting of each domain’s local training (Wu and Gong 2021), or adapting the vanilla federated techniques to the re-ID task (Zhuang et al. 2020). However, the direction of generating novel data through style transfer (Dumoulin, Shlens, and Kudlur 2017; Huang and Belongie 2017) is largely ignored. In the literature, data stylization is a commonly used strategy in DG (Zhou et al. 2021; Zhong et al. 2022; Tang et al. 2021), while most of them are not applicable in FedDG re-ID due to their requirement of centralizing source domains. Moreover, since the data authenticity is not explicitly ensured in these methods, unrealistic data may be generated, which will negatively impact the optimization.

In this paper, we tackle the FedDG re-ID problem in the view of data stylization and propose the Diversity-Authenticity Co-constrained Stylization (DACS) to generate diverse and authentic data for learning robust local models. Specifically, we introduce a style transformation model (STM) for each domain and jointly constrain the STM with two losses. (1) To generate diverse data, we encourage STM to produce novel data with a different distribution from the current domain. This is achieved by enlarging the simplified 2-Wasserstein distance (He et al. 2018) between the original and transformed data (see Fig. 1(1)). The larger the distance, the more diverse the generated data will be. In this way, the local model can see as diverse styles as possible during local optimization. However, the unconstrained enlargement of Wasserstein distance may lead to unrealistic stylization, which may impair the optimization. We thus propose the authenticity constraint to solve this problem. (2) Concretely, to ensure data authenticity, we require the generated data to be hardly / easily recognized by the local model / local-side global model. This is achieved by measuring and controlling the entropy produced by these two models (see Fig. 1(2)). By jointly considering the above two constraints, STM can generate diverse and authentic data, which helps us to learn more generalized local model and thereby improves FedDG re-ID accuracy. The main contributions of this paper are three-fold:

- We propose a novel data stylization approach for FedDG re-ID, which enables us to generate novel data for local training and promote FedDG re-ID accuracy.
- We design a diversity loss to enlarge the distributional discrepancies between original and generated data, which enables STM to generate novel data and avoids overfitting of local models.
- We further introduce an authenticity loss to enforce the

STM to produce authentic data, allowing the model to better benefit from the stylized images.

Extensive experiments conducted on four large-scale re-ID benchmarks demonstrate the advantage of our method in improving the generalization capability of local models. In addition, our method establishes new state-of-the-art results for FedDG re-ID.

2 Related Work

Domain Generalization. Deep neural networks are vulnerable to domain shift among different training domains. Therefore, recent studies resort to domain generalization (DG) (Li et al. 2018; Zhao et al. 2021; Jin et al. 2020; Chattopadhyay, Balaji, and Hoffman 2020) to optimize generalizable models that can be directly deployed in unseen domains. Recently, many methods are proposed to solve the DG problem in person re-ID (Jin et al. 2020; Zhao et al. 2021; Dai et al. 2021). *e.g.*, Zhao *et al.* (Zhao et al. 2021) adopt meta-learning to improve the generalization of models with interpolated features. Dai *et al.* (Dai et al. 2021) optimize an additional voting network for model aggregation to achieve generalization. Despite their success, most of them require the centralization of data from source domains, raising the risk of privacy leakage. Different from them, this paper tries to solve DG problem under the federated learning scenario (FedDG re-ID), which is a more challenging task. Please refer to supplementary for more explanations.

Federated Learning. Federated learning (McMahan et al. 2017; Karimireddy et al. 2020; Liu et al. 2021; Li, He, and Song 2021) aims at optimizing models with decentralized data to protect data privacy. FedAvg (McMahan et al. 2017) is the first federated learning algorithm, which averages locally trained models and redistributes the aggregated model to local clients for further training. Subsequently, FedProx (Li et al. 2020), MOON (Li, He, and Song 2021), and SCAFFOLD (Karimireddy et al. 2020) are proposed to prevent local overfit for better accuracies. These methods are originally designed for closed-set problems, *e.g.*, image-classification, where the training and testing sets share the same classes. To solve the open-set (Panareda Busto and Gall 2017) tasks, Zhuang *et al.* (Zhuang et al. 2020) propose FedPav, which adapts FedAvg to the re-ID by only exchanging feature extractors. Wu *et al.* (Wu and Gong 2021) first explicitly introduce the definition of FedDG re-ID and considers each source domain as an individual local client. They propose to solve the problem with model distillation. The above methods mainly focus on keeping the consistency between local and global models. Instead, this paper considers stylization for local data, which is largely overlooked.

Style Transfer. Style transfer (Dumoulin, Shlens, and Kudlur 2017; Huang and Belongie 2017; Choi et al. 2018; Zhu et al. 2017) is widely studied in the image translation, which aims to change the styles of images while retaining their semantic contents. Recently, the idea of style transfer is employed to generate data with different styles, which are used for learning generalized models. Specifically, (Zhou et al. 2021; Tang et al. 2021) interpolate styles of samples to synthesize novel ones while (Zhong et al. 2022; Wang et al.

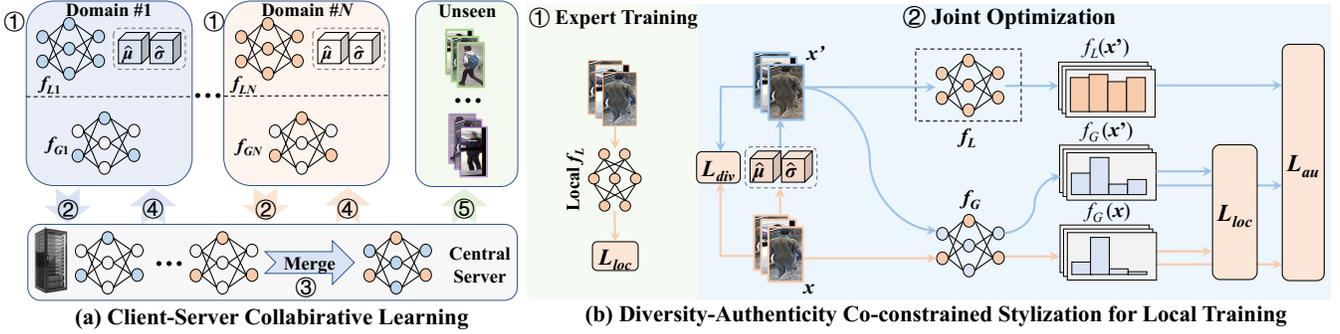


Figure 2: Illustration of our FedDG re-ID method. (a) Illustration of the overall client-server collaborative learning (CSCL), which contains five steps. ① Local Training. ② Model Upload. ③ Model Aggregation. ④ Redistribution. ⑤ Evaluation on Unseen Domains. Our paper mainly focuses on step ① local training. (b) The overall process of our DACS for local training, which is comprised of ① Expert Training and ② Joint Optimization.

2021) adopt learnable modules to directly optimize novel styles. These methods commonly require centralization of multiple domains to ensure the diversity of the generated styles, which is not applicable in federated learning. In addition, the data authenticity is not explicitly considered, which may hamper the model optimization. In this work, we devise a novel strategy to jointly ensure the diversity and authenticity, which is tailored-made for FedDG re-ID.

3 Methodology

Problem Definition. Given N labeled source domains $\mathcal{S} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_N\}$, where the i -th domain $\mathcal{D}_i = \{\mathcal{X}_i, \mathcal{Y}_i\}$ is comprised of M_i training images \mathcal{X}_i and their corresponding identity labels \mathcal{Y}_i . FedDG re-ID considers each domain as an isolated client and can not exchange data with others¹. The objective for FedDG re-ID is optimizing a generalizable re-ID model that performs well on unseen target domains by collaboratively utilizing these isolated clients with the help of central server *i.e.*, client-server collaborative learning (CSCL). The whole process formulates the cross-silo federated learning setting (Kairouz et al. 2021). As data centralization is not allowed, FedDG re-ID becomes more challenging than vanilla domain generalizable re-ID.

3.1 Overview

The overall process of CSCL for FedDG re-ID is illustrated in Fig. 2(a), which contains five steps. In step ① **Local Training**, we deploy two re-ID models and one style transformation model (STM) for each domain. The two re-ID models, including “local model” and “local-side global model”, are designed with different purposes. “Local model” is trained with only local data to retain domain-specific knowledge and will not be shared. On the contrary, “local-side global model” can be uploaded to central server. We adopt STM to generate novel data through our “diversity-authenticity co-constraint stylization” (DACS), which is achieved by using these two re-ID models as the supervision for image stylization. The generated data are subsequently leveraged for the optimization of “local-side global model”. In step ② **Model Upload**, the optimized

“local-side global model” is uploaded to the central server. In step ③ **Model Aggregation**, we aggregate the collected models to obtain the “server-side global model”. In step ④ **Redistribution**, the aggregated “server-side global model” is redistributed to each domain to update “local-side global model”. By iterating step ① to ④ until convergence, we deploy the “server-side global model” to target domains for evaluation, *i.e.*, step ⑤ **Evaluation on Unseen Domains**.

Our paper mainly focuses on step ① and tries to generate novel data with our diversity-authenticity co-constrained stylization (DACS) for local generalization. The details of our two-stage DACS for local training are demonstrated in Fig. 2(b). Concretely, in stage ① “Expert Training”, we conduct vanilla optimization on local model to maintain domain-specific knowledge. Then, in stage ② “Joint Optimization”, we devise diversity and authenticity losses to synthesize novel data for the joint optimization of STM and local-side global model. The former is designed to generate diverse data by enlarging Wasserstein distance between the original and transformed counterparts. The latter ensures the authenticity of generated images by enforcing the them to be hard/easy for the local/local-side global model to recognize through entropy function. Next, we introduce our DACS for local training in detail. Since DACS will be applied to each domain, we omit the subscript i for simplicity.

3.2 Diversity-Authenticity Constrained Stylization

Style Transformation Model. We deploy a style transformation model (STM) for each domain to generate data with novel distributions. The STM of each domain is comprised of two trainable parameters $\hat{\mu} \in \mathbb{R}^{C \times H \times W}$ and $\hat{\sigma} \in \mathbb{R}^{C \times H \times W}$, which can transform local data to novel styles through simple scaling and shifting. Specifically, given a batch of local images $x \in \mathbb{R}^{B \times C \times H \times W}$, we first compute their channel-wise data statistics μ and σ with:

$$\begin{aligned} \mu &= \frac{1}{HW} \sum_{h \in H, w \in W} x_{h,w}, \\ \sigma &= \sqrt{\frac{1}{HW} \sum_{h \in H, w \in W} (x_{h,w} - \mu)^2}, \end{aligned} \quad (1)$$

¹In this paper, “domain” and “client” are interchangeable.

where H and W are the spatial dimensions of input images, C equals to 3 for RGB images, and B is the batch size. The obtained channel-wise $\mu \in \mathbb{R}^{B \times C}$ and $\sigma \in \mathbb{R}^{B \times C}$ are utilized for stylization with the following equation:

$$x' = \phi(x; \hat{\mu}, \hat{\sigma}) = \hat{\sigma} * \frac{(x - \mu)}{\sigma + \epsilon} + \hat{\mu}, \quad (2)$$

where x' is the transformed data and ϵ is a small value to avoid zero division. $\hat{\mu}$ and $\hat{\sigma}$ are the trainable tensors in STM, which are initialized with Gaussian noise and will be optimized to explore the most plausible styles.

Diversity Loss. To generate novel data, we attempt to enlarge the distributional difference between x and x' . Based on our STM design, we can readily estimate the distribution of x' as $\mathcal{P}(x') \sim \mathcal{N}_{novel}(\hat{\mu}, \hat{\sigma}^2)$. Similarly, the distribution of original data x can also be formulated as $\mathcal{P}(x) \sim \mathcal{N}_{ori}(\mu, \sigma^2)$, where μ and σ are obtained through Eq. 1. We thus define the diversity loss L_{div} as:

$$L_{div}(x; \phi) = -\mathbb{D}(\mathcal{P}(x), \mathcal{P}(x')), \quad (3)$$

where $\mathbb{D}(\cdot, \cdot)$ is the distributional metric with various choices like Jensen–Shannon divergence (JSD) (Lin 1991) or KL divergence (Shannon 1948). We adopt simplified 2-Wasserstein distance (He et al. 2018) as the distributional metric due to its low computational burden and simple form. Therefore, the final diversity loss can be formulated as:

$$L_{div}(x; \phi) = L_{div}(x; \hat{\mu}, \hat{\sigma}) = -\|\mu - \hat{\mu}\|_2^2 - \|\sigma - \hat{\sigma}\|_2^2. \quad (4)$$

By reducing the above diversity loss, we generate novel data x' with a different data distribution from current domain. However, the unconstrained enlargement of Wasserstein distance may generate unrealistic samples and bring negative effects on local optimization. We thus propose another constraint to ensure data authenticity.

Authenticity Loss. In each domain, we have two re-ID models: “local model” and “local-side global model”. The former is solely optimized with local data while the latter is initialized with averaged global model. Therefore, the latter model has more generalized knowledge than the former model and can be utilized to assess the authenticity of input data. To this end, we propose to adopt both models for authenticity estimation. Concretely, (1) we require the transformed data to be well recognized by the local-side global model for high authenticity (*i.e.*, keep low global uncertainty). (2) However, consistently reducing global uncertainty is not sufficient as STM may be prone to trivial solution and keep the input images unchanged. To avoid this problem, we additionally enforce the transformed data to be hardly discriminated by the “local model” (*i.e.*, increasing local uncertainty). As the “local model” contains more domain-specific information, increasing the local uncertainty enables STM to generate out-of-domain data and thus avoid the trivial solution. (3) Moreover, to facilitate the optimization of STM for data transformation, the generated counterparts should also be harder for local-side global model to recognize than the original images. We thus further constrain the global uncertainty of the transformed data to be larger than the original data.

Algorithm 1: The Process of Our Local Training.

Inputs: Training data from the i -th domain \mathcal{X}_i and labels \mathcal{Y}_i . Local iteration number $iter$. STM ϕ_i

Outputs: Feature Extractor for i -th domain θ_{G_i} .

```

1: function LOCALTRAIN(Domain  $i$ , STM  $\phi_i$ , local
   model  $f_{L_i}$ , local-side global model  $f_{G_i}$ )
2:   for  $iter\_num$  in  $iter$  do
3:     Sample a batch of training data  $\{x_i, y_i\}$ ;
4:     // Stage ①: Expert Training.
5:     Unfreeze  $f_{L_i}$ ;
6:     Optimize  $f_{L_i}$  with Eq. 7 and  $\{x_i, y_i\}$ ;
7:     // Stage ②: Joint Optimization.
8:     Freeze  $f_{L_i}$ ;
9:     Transform  $x$  to  $x'$  with  $\phi_i$  via Eq. 2;
10:    Compute Eq. 8 with  $f_{L_i}$ ,  $f_{G_i}$ , and  $\phi_i$ ;
11:    Update  $f_{G_i}$ , and  $\phi_i$ ;
12:  end for
13:  Return Feature extractor  $\theta_{G_i}$  of  $f_{G_i}$ .
14: end function

```

In this paper, we adopt entropy function $H(\cdot) = -\sum_j p_j(\cdot) \log p_j(\cdot)$ to measure the model’s uncertainty for the given data, where $p_j(\cdot)$ is the probability of classifying the input data to the j -th class. Our authenticity constraint can be formulated as:

$$H(f_G(x)) < H(f_G(x')) < H(f_L(x')), \quad (5)$$

where x and x' are original and transformed data. $f_G(\cdot)$ and $f_L(\cdot)$ are the logits predicted by “local-side global model” and “local model”, respectively. Since these inequalities can not be optimized by back-propagation, we convert Eq. 5 to the following loss:

$$L_{au}(x; \phi) = Softplus(H(f_G(x)) - H(f_G(\phi(x)))) + Softplus(H(f_G(\phi(x))) - H(f_L(\phi(x))))), \quad (6)$$

where $Softplus(\cdot) = \ln(1 + \exp(\cdot))$ is a monotonically increasing function. Minimizing the first term of Eq. 6 is equivalent to achieving $H(f_G(x)) < H(f_G(x'))$. Meanwhile, minimizing the second term of Eq. 6 means satisfying $H(f_G(x')) < H(f_L(x'))$. It should be noted that L_{au} is designed only for optimizing STM ϕ , and other models (local model f_L and local-side global model f_G) will not be updated in back-propagation.

3.3 Local Training

We improve the generalization of local-side global model and optimize STM with our proposed DACS during the local training. As shown in Fig. 2(b), our local training includes two stages: “expert training” and “joint optimization”.

Expert Training. In this stage (see Fig. 2(b)- ①), we train the local model with a sampled batch of local data x and their corresponding labels y with the following loss:

$$L_{loc}(x, y; f_L) = L_{tri}(\theta_L(x), y) + L_{ce}(f_L(x), y), \quad (7)$$

where θ_L is the feature extractor part of local model f_L and outputs intermediate features to compute triplet loss

Algorithm 2: The Process of Our FedDG re-ID method.

Inputs: N decentralized domains. Their corresponding training data \mathcal{X}_i and labels \mathcal{Y}_i ($1 \leq i \leq N$). Local iteration number $iter$. Total number of Epochs E .

Outputs: Generalized Feature Extractor θ_S .

```

1: // Client-Server Collaborative Learning (CSCL).
2: function FEDDGTRAIN(Training epochs  $E$ ,  $N$  decentralized domains)
3:   Initialize all  $\theta_L$  and  $\theta_G$  with
     ImageNet-pretrained weights;
4:   for  $epoch\_num$  in  $E$  do
5:     // Step ①: Local Train via DACS.
6:     for  $i$  in  $N$  do
7:        $\theta_{G_i} = \text{LocalTrain}(i, \phi_i, f_{L_i}, f_{G_i});$ 
8:     end for
9:     // Step ②: Model Upload.
10:    Upload  $\{\theta_{G_1}, \dots, \theta_{G_N}\}$  to server;
11:    // Step ③: Model Aggregation.
12:    Obtain  $\theta_S$  via Eq. 9;
13:    // Step ④: Redistribution.
14:    Redistribute  $\theta_S$  to each domain;
15:    Update  $\{\theta_{G_1}, \dots, \theta_{G_N}\}$ ;
16:  end for
17:  Return  $\theta_S$  for evaluation.
18: end function

```

(e.g., pool-5 features if we adopt ResNet-50 as the backbone). L_{tri} and L_{ce} are commonly used triplet and cross-entropy losses. This stage ensures f_L 's capability of retaining domain-specific data distribution.

Joint Optimization. In the second stage (see Fig. 2(b)- ②), we adopt the original data and their generated counterparts to optimize both f_G and STM. Meanwhile, STM will also be supervised by our diversity and authenticity losses to ensure the data quality used for joint optimization. The final loss for joint optimization is formulated as:

$$L_{DACs}(x, y; f_G, \phi) = L_{loc}(x, y; f_G) + L_{loc}(\phi(x), y; f_G, \phi) + \lambda_{div}L_{div}(x; \phi) + \lambda_{au}L_{au}(x; \phi), \quad (8)$$

where λ_{div} and λ_{au} are balancing factors. It should be noted that both diversity and authenticity losses will not affect the optimization of local-side global model f_G . Therefore, we jointly optimize STM and local-side global model in L_{DACs} . The overall process is illustrated in Alg. 1.

3.4 Subsequent Learning

Model Upload. We upload the local-side global model to the central server. Different from federated learning in image classification, each domain has different pedestrians in FedDG re-ID. Therefore, for the local-side global model $f_{G_i} = \{\theta_{G_i}, \psi_i\}$ in the i -th domain, we only share the feature extractor θ_{G_i} and keep classifier ψ_i in its own client.

Model Aggregation. We aggregate $\{\theta_{G_1}, \dots, \theta_{G_N}\}$ from all clients to obtain server-side global model θ_S :

$$\theta_S = \sum_{i=1}^N \frac{M_i}{M_{total}} \theta_{G_i}, \quad (9)$$

where $M_{total} = \sum_{i=1}^N M_i$ is the total number of images for all clients.

Redistribution. The obtained θ_S will be further redistributed to each domain to update local-side global models for the next epoch of training.

Evaluation on Unseen Domains. After iterating previous steps until convergence, the obtained “server-side global model” will be directly deployed to unseen domains for evaluation. The overall process is shown in Alg. 2.

4 Experiments

4.1 Experiment Setup

The details of all experiments, including the used datasets, evaluation protocols, and implementation details, are demonstrated in the supplementary. Note that, DukeMTMC-reID has been withdrawn and is thus not used in this work.

4.2 Comparison with State of the Art

We first compare our algorithm with state-of-the-art methods in Tab. 1. The compared algorithms can be divided into four categories. (1) Classical federated learning, such as SCAFFOLD (Karimireddy et al. 2020) and MOON (Li, He, and Song 2021). (2) Federated re-ID algorithms, including FedPav (Zhuang et al. 2020) and FedReID (Wu and Gong 2021). FedPav can be seen as the baseline for FedDG re-ID. (3) Single domain generalization (SDG) for re-ID like SNR (Jin et al. 2020) and TransMatcher (Liao and Shao 2021). *SDG is the only type of normal DG algorithm that can be used under federated scenario, please find supplementary for more explanations.* In Tab. 1, we report the results of SNR and *leave the results of TransMatcher in the supplementary.* (4) Stylization-based domain generalization, including MixStyle (Zhou et al. 2021) and CrossStyle (Tang et al. 2021). For “SNR”, we adopt its recommended hyperparameters and deploy SNR modules after each ResNet layer to ensure the best results are achieved. The independently trained models are then averaged for evaluation. For “MixStyle” and “CrossStyle”, we directly deploy them in each client to generate novel data by mixing or exchanging local styles because federated learning does not allow data centralization of source domains. “Joint” denotes training re-ID models with centralized source domains. We also report the results of using ViT (Dosovitskiy et al. 2021) as backbone and compare them with FedPav and CrossStyle.

From Tab. 1, we have three conclusions. **(1) Optimizing with more decentralized domains can achieve better accuracies.** Here we take experiments evaluated on Market-1501 to demonstrate. When optimizing with only one domain, the best mAP score is 23.3% (“MS→M”), which is lower than all FedDG re-ID methods trained with multiple domains. Therefore, FedDG re-ID with multiple domains is worth researching and has significant meaning for optimizing robust and safe re-ID system. **(2) Our method achieves state-of-the-art performance.** Specifically, for “MS+C2+C3→M”, we achieve 36.3% in mAP and 61.2% in rank-1 accuracy when using ResNet-50 as backbone, which are the highest accuracies among previous methods. Moreover, our method also outperforms “Joint”, demonstrating

Methods	MS + C3 +C2 → M		C2 + C3 +M → MS		MS + C2 +M → C3	
	mAP	rank-1	mAP	rank-1	mAP	rank-1
Single Model (1)	23.3	47.5	2.7	8.8	18.0	18.5
Single Model (2)	13.2	31.1	1.7	6.0	21.6	22.5
Single Model (3)	18.9	41.2	3.3	10.0	10.2	11.2
MOON	26.8	51.1	4.8	14.5	20.9	22.5
SCAFFOLD	26.0	50.5	5.3	15.8	22.9	26.0
FedPav	25.4	49.4	5.2	15.5	22.5	24.3
FedReID	30.1	53.7	4.5	13.7	26.4	26.5
MixStyle	31.2	53.5	5.5	16.0	28.6	31.5
CrossStyle	32.5	59.6	4.6	14.0	27.8	28.0
SNR	32.7	59.4	5.1	15.3	28.5	30.0
Ours	36.3	61.2	10.4	27.5	30.7	34.1
SNR+Ours	37.7	65.9	11.6	29.4	33.6	37.5
Joint	32.2	58.6	5.6	16.3	29.5	29.0
FedPav (ViT)	37.4	62.6	14.6	33.7	23.7	25.0
CrossStyle (ViT)	41.4	65.8	17.9	40.8	31.0	38.4
Ours (ViT)	45.4	70.7	20.3	44.2	36.6	42.1

Table 1: Comparison with state-of-the-arts, we compare our method with state-of-the-art federated learning algorithms. M: Market-1501, C2: CUHK02, C3: CUHK03, MS: MSMT-17. Joint: the model jointly trained on all source domains without decentralization constraint. Single Model (N): the model is individually trained on the N-th source domain of each setting. “ViT”: Results of using ViT as Backbone.

Method	Attributes			MS+C2 +C3→M		MS+M +C2→C3	
	ϕ	L_{div}	L_{au}	mAP	rank-1	mAP	rank-1
Baseline	×	×	×	25.4	49.4	22.5	24.3
RS	✓	×	×	25.1	50.6	20.4	22.8
DC	✓	✓	×	31.9	57.1	29.1	30.0
AC	✓	×	✓	34.5	59.7	27.3	28.0
DACS	✓	✓	✓	36.3	61.2	30.7	34.1

Table 2: Ablation study on STM ϕ , diversity loss L_{div} , and authenticity loss L_{au} . RS: RandStyle. DC: diversity constraint. AC: authenticity constraint. DACS: Ours.

its capability of learning generalized re-ID models under federated learning constraint. Similar results can also be found in other settings. When we change the backbone to “ViT” (Dosovitskiy et al. 2021), our method can also outperform FedPav and CrossStyle, demonstrating its effectiveness. **(3) Our method is compatible with other domain generalization algorithm.** Concretely, when compared with “SNR”, the mAP scores of “SNR+Ours” are improved. In sum, the above results demonstrate the superiority and compatibility of the proposed method for FedDG re-ID.

4.3 Ablation Study

To better understand how STM ϕ , diversity loss L_{div} , and authenticity loss L_{au} affect the FedDG re-ID results, we gradually add them into the training process of “MS+C2+C3→M” and “MS+C2+M→C3” for ablation

Method	MS+C2 +C3→M		MS+M +C2→C3	
	mAP	rank-1	mAP	rank-1
Baseline	25.4	49.4	22.5	24.3
(a) The choice of distributional metric				
JSD	34.2	58.8	29.1	32.8
KL	33.6	56.8	28.0	30.4
WD (ours)	36.3	61.2	30.7	34.1
(b) Data transfer in channel or spatial level				
Channel level	30.0	55.4	25.1	25.5
Spatial level (Ours)	36.3	61.2	30.7	34.1
(c) The design of authenticity loss				
$H(f_G(x)) < H(f_G(x'))$	33.3	58.4	27.0	29.2
$H(f_G(x')) < H(f_L(x'))$	34.6	59.0	26.7	28.6
Complete (Ours)	36.3	61.2	30.7	34.1

Table 3: Further experiments. (a) The choice of distributional metric in Eq. 3. (b) The choice of channel-level or spatial-level transfer. (c) The design of authenticity loss.

study. We compare five different training schemes in Tab. 2. **Baseline:** results of vanilla federated learning. **RandStyle (RS):** results of transforming data to random styles for optimization. The parameters of STM are randomly initialized and do not have any constraint during the optimization. **Diversity Constraint (DC):** results of DACS without authenticity constraint for local training. **Authenticity Constraint (AC):** results of DACS without diversity constraint in local training step. **DACS:** results of using complete DACS.

Effectiveness of solely using diversity or authenticity constraint. By comparing “RS” and “DC” of Tab. 2, we note that diversity loss L_{div} is beneficial to improving the generalization of FedDG re-ID models. Specifically, in row “RS”, we do not use any constraint on STM and its parameters are randomly initialized. In this case, STM can not generate transformed data that are beneficial to the generalization, leading to the stagnant re-ID accuracies. The results in “RS” can not even outperform “baseline”, until taking diversity loss into optimization. Similarly, the comparison between “RS” and “AC” in Tab. 2 demonstrate the effectiveness of using authenticity loss in optimization. We thus conclude that either diversity or authenticity constraint is capable of improving FedDG re-ID accuracies.

Effectiveness of diversity-authenticity co-constraint. By comparing “DC”, “AC”, and “DACS” in Tab. 2, we observe that jointly using diversity and authenticity losses can further improve FedDG re-ID accuracies. For example, in “MS+C2+C3→M”, only using diversity or authenticity constraint during model optimization achieves 31.9% and 34.5% mAP scores, respectively. However, after jointly using our two constrains, the mAP score becomes 36.3%. Therefore, we conclude that the proposed diversity and authenticity constraints are complementary to each other. Using both constraints can coherently improve performance.

4.4 Further Experiments

The choice of distributional metric. In Eq. 3, we choose to enlarge Wasserstein distance between the original and

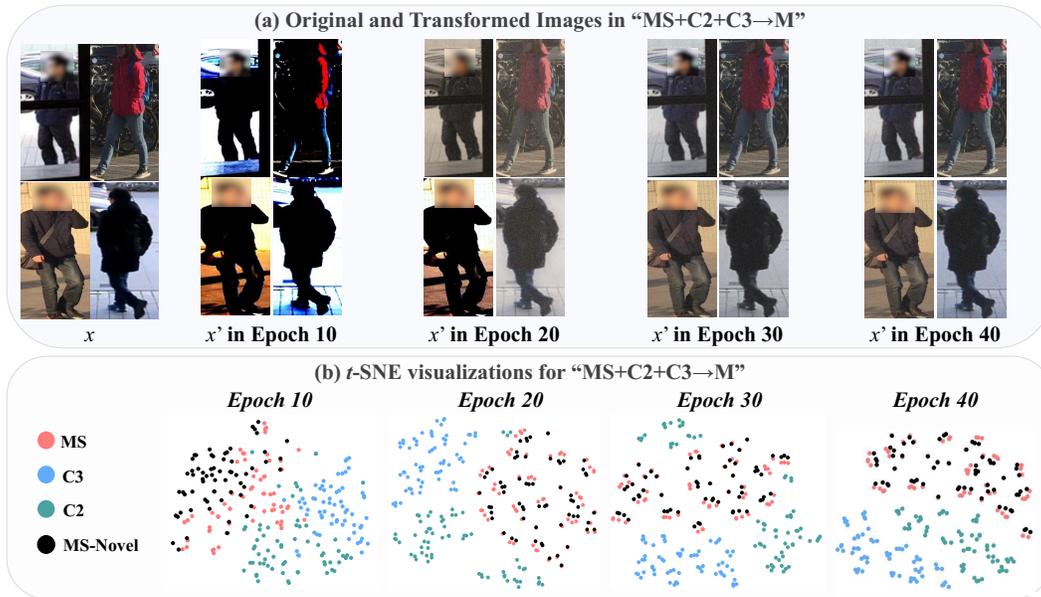


Figure 3: (a) The original and transformed data during the optimization of “MS+C2+C3→M”. (b) t -SNE visualizations for “MS+C2+C3→M”. MS: MSMT-17. C3: CUHK03. C2: CUHK02. MS-Novel: Transformed MS data.

transformed data to ensure the data diversity. However, the distributional metric has various choices, such as JSD (Lin 1991) or KL divergence (Shannon 1948). We conduct experiments of using these two metrics in “MS+C2+C3→M” and “MS+M+C2→C3” to find the best distributional metric for diversity loss. As shown in Tab. 3(a), using Wasserstein distance can achieve the best re-ID accuracies for both tasks and we thus use Wasserstein distance in diversity loss.

The choice of channel- or spatial-level style transfer. The learnable parameters ($\hat{\mu}$ and $\hat{\sigma}$) in our STM are $C \times H \times W$ tensors, which enable us to transfer styles for each pixel, *i.e.*, spatial-level style transfer. There is also another type of strategy for data transfer, which defines $\hat{\mu} \in \mathbb{R}^C$ and $\hat{\sigma} \in \mathbb{R}^C$ to uniformly transfer pixels in the same channel to the same style, *i.e.*, channel-level transfer. In Tab. 3(b), we report the results of using channel-level style transfer on two FedDG re-ID tasks. From the results, we observe that using channel-level style transfer achieves lower accuracies than our spatial-level design. We thus choose to transfer styles of input images spatially as it is a more effective manner.

Further study on the design of authenticity loss. The authenticity loss in Eq. 6 is proposed to achieve two inequalities, *i.e.*, $H(f_G(x)) < H(f_G(x'))$ and $H(f_G(x')) < H(f_L(x'))$. We conduct experiments of independently using the two inequalities as constraints and compare them with the results of using complete L_{div} in Tab. 3(c). From the results, we note that independently using each inequality can bring improvements on the re-ID accuracies, while jointly using both can achieve the best results. Therefore, we conclude that both inequalities are important for L_{div} .

4.5 Visualization

We visualize the original / transformed images and features during the optimization of “MS+C2+C3→M” to better un-

derstand our algorithm. Specifically, given a batch of MS images x , we obtain their transformed counterparts x' with STM and show them in Fig. 3(a). Moreover, these data, combined with data from other source domains (C2 and C3), are forwarded to server-side global model θ_S for feature extraction and t -SNE (Van der Maaten and Hinton 2008) visualization in Fig. 3(b). From these two figures, we observe that at the early stage of training (the first 10 epochs), transformed images x' are quite different from their original counterparts x in both feature-level and image-level. In image-level (see Fig. 3(a)), the transformed images have high contrast and unrealistic illumination, which may lead to the performance degradation of re-ID model. In feature-level (see Fig. 3(b)), the transformed data have large discrepancies with their original images in the feature space. However, these discrepancies are gradually reduced at the later stage of training (after 20th epoch). Based on these observations, we conjecture that DACS first focuses on the diversity of the transformed data and gradually improves the data authenticity for further local training. The step-by-step transformation ensures DACS’s effectiveness.

5 Conclusion

In this paper, we propose a diversity-reality co-constrained stylization (DACs) method for FedDG re-ID task. Specifically, we adopt STM to generate novel data by jointly consider diversity and authenticity constrains. The diversity loss requires the generated data to be different from local domain by enlarging Wasserstein distance. The authenticity loss enforces the transformed data to be hard / easy for local model / local-side global model to recognize for data authenticity. Extensive experiments show the efficacy of our method.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62276221, No. 62376232), the Natural Science Foundation of Fujian Province of China (No. 2022J01002), and the Science and Technology Plan Project of Xiamen (No. 3502Z20221025).

References

- Chattopadhyay, P.; Balaji, Y.; and Hoffman, J. 2020. Learning to balance specificity and invariance for in and out of domain generalization. In *European Conference on Computer Vision*.
- Choi, Y.; Choi, M.; Kim, M.; Ha, J.-W.; Kim, S.; and Choo, J. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- Dai, Y.; Li, X.; Liu, J.; Tong, Z.; and Duan, L.-Y. 2021. Generalizable person re-identification with relevance-aware mixture of experts. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representation*.
- Dumoulin, V.; Shlens, J.; and Kudlur, M. 2017. A learned representation for artistic style. In *International Conference on Learning Representation*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- He, R.; Wu, X.; Sun, Z.; and Tan, T. 2018. Wasserstein CNN: Learning invariant features for NIR-VIS face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Huang, G.; Liu, Z.; Pleiss, G.; Van Der Maaten, L.; and Weinberger, K. 2019. Convolutional networks with dense connectivity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; and Zhang, L. 2020. Style normalization and restitution for generalizable person re-identification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*.
- Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2018. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Li, Q.; He, B.; and Song, D. 2021. Model-Contrastive Federated Learning. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*.
- Liao, S.; and Shao, L. 2021. TransMatcher: Deep Image Matching Through Transformers for Generalizable Person Re-identification. In *Advances in Neural Information Processing Systems*.
- Lin, J. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*.
- Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021. FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*.
- Panareda Busto, P.; and Gall, J. 2017. Open set domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*.
- Song, J.; Yang, Y.; Song, Y.-Z.; Xiang, T.; and Hospedales, T. M. 2019. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *European Conference on Computer Vision*.
- Tang, Z.; Gao, Y.; Zhu, Y.; Zhang, Z.; Li, M.; and Metaxas, D. N. 2021. CrossNorm and SelfNorm for Generalization Under Distribution Shifts. In *Proceedings of the International Conference on Computer Vision*.
- Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*.
- Wang, G.; Yuan, Y.; Chen, X.; Li, J.; and Zhou, X. 2018. Learning discriminative features with multiple granularities for person re-identification. In *Proceedings of the ACM International Conference on Multimedia*.
- Wang, Z.; Luo, Y.; Qiu, R.; Huang, Z.; and Baktashmotlagh, M. 2021. Learning to diversify for single domain generalization. In *Proceedings of the International Conference on Computer Vision*.
- Wu, G.; and Gong, S. 2021. Decentralised Learning from Independent Multi-Domain Labels for Person Re-Identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; and Hoi, S. C. H. 2021. Deep Learning for Person Re-identification: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhao, Y.; Zhong, Z.; Yang, F.; Luo, Z.; Lin, Y.; Li, S.; and Sebe, N. 2021. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Zhong, Z.; Zhao, Y.; Lee, G. H.; and Sebe, N. 2022. Adversarial style augmentation for domain generalized urban-scene segmentation. In *Advances in Neural Information Processing Systems*.

Zhou, K.; Yang, Y.; Qiao, Y.; and Xiang, T. 2021. Domain generalization with mixstyle. In *International Conference on Learning Representation*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the International Conference on Computer Vision*.

Zhuang, W.; Wen, Y.; Zhang, X.; Gan, X.; Yin, D.; Zhou, D.; Zhang, S.; and Yi, S. 2020. Performance optimization of federated person re-identification via benchmark analysis. In *Proceedings of the ACM International Conference on Multimedia*.