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

Fengxiang Yang[1,2], Zhun Zhong[3], Zhiming Luo[1*], Yifan He[4], Shaozi Li[1], Nicu Sebe[2]

[1]Department of Artificial Intelligence, Xiamen University, China
[2]Department of Information Engineering and Computer Science, University of Trento, Italy
[3]School of Computer Science, University of Nottingham, UK [4]Reconova Technologies Co., Ltd., China
Project: https://github.com/FlyingRoastDuck/DACS_official.git

In this appendix, we provide additional experiments and discussions for our Diversity-Authenticity Co-constrained Stylization (DACS), including sensitivity analysis (Sec.1), convergence curve (Sec.2), quality evaluation on transformed data (Sec.3), and discussion on main-stream DG methods under federated learning scenario (Sec.4).

1 Experiment Settings

Datasets & Evaluation Protocols We use four large-scale re-ID benchmarks to conduct our experiments, including Market-1501 (M) (Zheng et al. 2015), CUHK02 (C2) (Li and Wang 2013), MSMT-17 (MS) (Wei et al. 2018) and CUHK03 (C3) (Li et al. 2014). Market-1501 contains $1,501$ IDs ($32,668$ images) captured by 6 cameras. CUHK02 has $7,264$ bounding-boxes manually cropped from $1,816$ pedestrians. MSMT-17 contains $126,441$ images from $4,101$ IDs, which are obtained by 15 cameras. CUHK03 has $28,193$ pedestrian images from $1,467$ IDs. For FedDG re-ID experiments, we adopt one of the four datasets as the testing set while others as decentralized source domains. The final server-side global model is used for evaluation. We adopt the commonly used mean average precision (mAP) and rank-1 as the evaluation metrics.

Implementation Details We adopt ResNet50 (He et al. 2016) as the backbone for most of our experiments and initialize the model with ImageNet (Deng et al. 2009) pre-trained weights. We also conduct some experiments of using ViT (Dosovitskiy et al. 2021) as backbone to demonstrate our method's flexibility and effectiveness. The pool-5 features are used to compute mAP and rank-1 for evaluation. For local training of each domain, we set the training batch size $B = 64$, input image size $\{H, W\} = \{256, 128\}$. The total training epochs E is set to 40 and the local training iteration $iter$ is set to 200. We adopt SGD optimizer for the training of both re-ID models and STM during local training. The learning rate γ is set to 0.001, which will be multiplied by 0.5 at the 20th and 30th epochs. We set λ_{div} and λ_{au} in Eq. **??** to 1 and 5, respectively. We use random crop, random flip and random erasing (Zhong et al. 2020) for data augmentation.

*Corresponding author

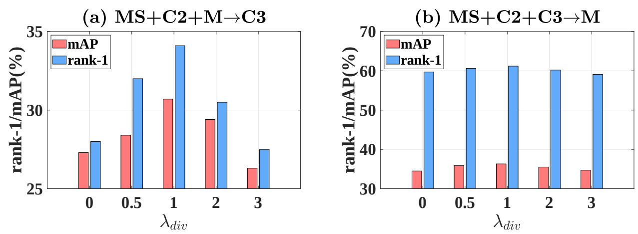

Figure 1: Sensitivity analysis for λ_{div}.

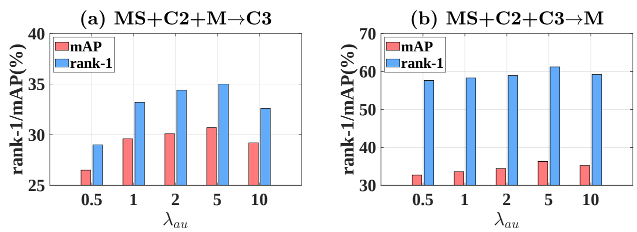

Figure 2: Sensitivity analysis for λ_{au}.

2 Sensitivity Analysis

Sensitivity for the number of clients N. In Tab. 1, we set the number of source domains N to 2 to evaluate how the number of source domains affects the final results. We conclude that: ***(1) Using more decentralized domains is beneficial to improving the FedDG re-ID accuracies***. For example, when comparing the results evaluated on Market-1501, our method achieves 31.8% and 32.0% mAP scores for "MS+C3→M" and "MS+C2→M", respectively. Both results are lower than the result in "MS+C2+C3→M", which achieves 36.3% mAP. Therefore, it is better to use more decentralized domains for FedDG re-ID. ***(2) Our method still outperforms other federated and generalized methods***. For example, we achieve 31.8% mAP for "MS+C3→M", which is better than the results of SNR (Jin et al. 2020) and FedPav (Zhuang et al. 2020). Therefore, we conclude that our method is still effective when less decentralized domains are given.

Sensitivity for balancing factor λ_{div} of diversity loss. λ_{div} is used to adjust the importance of diversity loss L_{div} during the optimization. To investigate how λ_{div} influence the final results, we vary its value from 0 to 3 and visualize the

Table 1: Effectiveness of training with less source clients.

Methods	MS+C3→M		MS+C2→M		MS+C2+C3→M	
	mAP	rank-1	mAP	rank-1	mAP	rank-1
FedPav	24.8	48.6	25.8	50.4	25.4	49.4
SNR	29.1	54.0	29.8	54.8	30.4	56.5
Ours	**31.8**	**56.6**	**32.0**	**57.1**	**36.3**	**61.2**
Methods	MS+M→C3		MS+C2→C3		MS+C2+M→C3	
	mAP	rank-1	mAP	rank-1	mAP	rank-1
FedPav	19.9	18.5	21.4	24.5	22.5	24.3
SNR	20.4	21.0	26.1	26.5	28.5	30.0
Ours	**23.8**	**25.4**	**27.9**	**30.5**	**30.7**	**34.1**

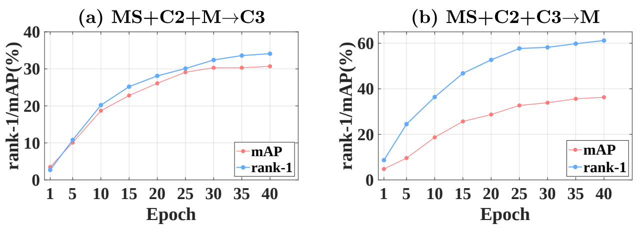

Figure 3: Convergence curve for two generalization settings.

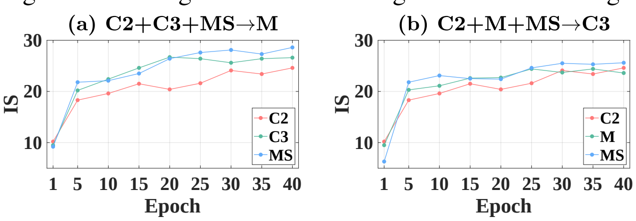

Figure 4: Inception Score (IS) for two FedDG re-ID settings.

results in Fig. 1. From the figure, we observe that the best results are achieved when $\lambda_{div} = 1$ for both experiments. Assigning higher or lower value to λ_{div} will lead to performance degradation. We thus set λ_{div} to 1.

Sensitivity for balancing factor λ_{au} of authenticity loss. λ_{au} is used to control the influence of authenticity loss L_{au}. In Fig. 2, we change its value from 0.5 to 10 to evaluate how λ_{au} affects the final results. From the results, we conclude that enlarging λ_{au} can improve generalization of re-ID model, but when its value exceed 5, the re-ID accuracies will decrease. Therefore, we suggest setting λ_{au} to 5 during the optimization.

3 Convergence Curve

We evaluate the mAP and rank-1 scores of the server-side global model on two FedDG re-ID settings every 5 epochs to check the convergence of our method. As shown in Fig. 3, the mAP and rank-1 scores become steady at the 35-th epoch for "MS+C2+M→C3", while "MS+C2+C3→M" converged at the 30-th epoch. Therefore, our method has converged within the 40 training epochs.

4 Quality Evaluation on Transformed Data

We also adopt Inception Score (IS) (Salimans et al. 2016) to evaluate the quality of transformed images for both "MS+C2+C3→M" and "MS+C2+M→C3". IS is a commonly used evaluation metric (Ma et al. 2018) that jointly considers data diversity and authenticity. Higher IS score suggests better quality of the transformed data. As shown in Fig. 4, we observe that all the IS scores of the transformed images are promoted within the 40 training epochs. Therefore, our method is capable of improving diversity and authenticity of the transformed data for better local generalization.

5 Discussion on Main-stream DG Methods for Federated Learning

Recently, there are many representative methods for training generalized re-ID models, which can achieve fabulous re-ID accuracies such as SNR (Jin et al. 2020), M^{3}L (Zhao et al. 2021), RaMoE (Dai et al. 2021), and TransMatcher (Liao and Shao 2021). Generally speaking, these works can be categorized into single-domain generalization (SDG) and multi-domain generalization (MDG). SDG methods (*e.g.*, SNR and TransMatcher) devise novel modules or network structures to improve model generalization with one source domain, while MDG methods (*e.g.*, M^{3}L and RaMoE) jointly adopt data distributions from multiple source domains for generalized training. **It is obvious that MDG methods require the centralization of source domains and are thus not applicable for FedDG re-ID.** Therefore, we only compare our method with SDG methods (SNR and TransMatcher) by deploying them to each client for local training. The results of SNR have been reported in Sec.4.2 "Comparison with State of the Art" of our main paper. We will further compare our method with TransMatcher in our supplementary.

Different from SNR, DACS can not be directly compared with TransMatcher (Liao and Shao 2021) due to the following three reasons. (1) Different network structures. TransMatcher adopts an additional transformer to encode feature maps obtained by ResNet-50 for further discriminative learning, while our method does not have such a module. (2) Different loss functions. TransMatcher adopts binary cross entropy loss to match the feature maps of positive sample pairs, which does not leverage model certainty and is thus not compatible with our authenticity loss. (3) Different evaluation processes. TransMatcher adopts matching score (Liao and Shao 2020) to compute the similarities between a pair of images, which is different from the commonly used Euclidean distance of pool-5 features (Jin et al. 2020; Zhao et al. 2021).

To make a fair comparison, we apply three modifications. We first replace ResNet-50 on each local domain with TransMatcher to form "Fed-TM" as the baseline for federated learning. Based on "Fed-TM", we further deploy one FC layer on each TransMatcher to implement our DACS. We call this modified method as "Fed-TM-DACS". Finally, the matching score (Liao and Shao 2020) is used to compute feature similarities for evaluation. In our experiments, we use the default hyper-parameters and training sampler (*i.e.*, Graph Sampler (Liao and Shao 2022)) mentioned in the original TransMatcher paper to ensure the best results are

Table 2: Comparison with TransMatcher and our method. "TM (N)": TransMatcher is individually trained on the N-*th* source domain of each setting. "Avg": averaged accuracies of three "TM (N)" experiments under each generalization setting. "SAvg": the averaged accuracies of three settings for each method. "Fed-TM": TransMatcher under federated scenario. "Fed-TM-DACS": Fed-TM + Our DACS.

Methods	MS + C3 +C2 → M		C2 + C3 +M → MS		MS + C2 +M → C3		SAvg	
	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1
TM (1)	52.0	80.1	14.6	46.1	22.5	23.7	29.7	50.0
TM (2)	33.3	67.6	13.1	40.7	45.6	56.0	30.7	54.8
TM (3)	32.6	63.8	18.4	47.3	21.4	22.2	24.1	44.4
Avg	39.3	70.5	15.4	44.0	29.8	34.0	28.2	49.7
Fed-TM	46.5	74.2	16.4	45.8	41.3	48.0	34.7	56.0
Fed-TM-DACS	52.4	80.9	20.8	54.7	43.2	50.2	38.8	61.9

achieved. All the results are shown in Tab. 2.

We conclude from Tab. 2 that: **(1) Federated learning with multiple source domains has practical meaning in the real world.** We take "MS+C3+C2→M" as an example. Although the SDG algorithm "TransMatcher" can achieve higher results than "Fed-TM" in some generalization experiments (*e.g.*, 52.0% mAP in "MS→M"), it fails to achieve competitive results when using other source domains for optimization, leading to the lower averaged mAP score of three "TM (N)" experiments (39.3%) than the mAP of "Fed-TM" (46.5%). Although TransMatcher is a brilliant SDG method, its performance is sensitive to the choice of source domain for optimization. This increases the difficulty of real-world deployment because we do not know what type of source domain is beneficial to the generalization of re-ID model in unseen target domain. Therefore, it is better to use multiple domains for generalized training. Moreover, federated learning further alleviates the privacy issue, which increases the practicality of using multiple domains for optimization. **(2) Our DACS helps improve the accuracies of TransMatcher under federated scenario.** As shown in "Fed-TM-DACS" of Tab. 2, after integrating our DACS into "Fed-TM", the re-ID accuracies of three settings are all improved. Furthermore, by comparing "SAvg" in Tab. 2, we note that "Fed-TM-DACS" achieves the highest accuracies (mAP=38.8%, rank-1=61.9%), demonstrating the effectiveness of our method.

References

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*.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A Large-Scale Hierarchical Image Database. 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*.

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*.

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*.

Li, W.; and Wang, X. 2013. Locally Aligned Feature Transforms across Views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Li, W.; Zhao, R.; Xiao, T.; and Wang, X. 2014. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Liao, S.; and Shao, L. 2020. Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting. In *European Conference on Computer Vision*.

Liao, S.; and Shao, L. 2021. TransMatcher: Deep Image Matching Through Transformers for Generalizable Person Re-identification. In *Advances in Neural Information Processing Systems*.

Liao, S.; and Shao, L. 2022. Graph sampling based deep metric learning for generalizable person re-identification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Ma, L.; Sun, Q.; Georgoulis, S.; Van Gool, L.; Schiele, B.; and Fritz, M. 2018. Disentangled person image generation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; and Chen, X. 2016. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

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*.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the International Conference on Computer Vision*.

Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; and Yang, Y. 2020. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

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*.