

FEATURE-DISTRIBUTION PERTURBATION AND CALIBRATION FOR GENERALIZED REID

Qilei Li, Jiabo Huang, Jian Hu, Shaogang Gong

Queen Mary University of London

{q.li, jiabo.huang, jian.hu, s.gong}@qmul.ac.uk

ABSTRACT

Person Re-identification (ReID) has been advanced remarkably over the last 10 years. However, the i.i.d. (independent and identically distributed) assumption is somewhat non-applicable to ReID considering its objective to identify images of the same pedestrian across cameras at different locations. In this work, we propose a Feature-Distribution Perturbation and Calibration (PECA) method to derive generic feature representations for person ReID. Specifically, we perform per-domain feature-distribution perturbation to refrain the model from overfitting to the domain-biased distribution of each source (seen) domain by enforcing feature invariance to distribution shifts caused by perturbation. Furthermore, we design a global calibration mechanism to align feature distributions across all the source domains to improve the model’s generalization capacity by eliminating domain bias. These local perturbation and global calibration are conducted simultaneously, which share the same principle to avoid models overfitting by regularization respectively on the perturbed and the original distributions. Extensive experiments were conducted and the proposed PECA model outperformed the state-of-the-art competitors by significant margins.

Index Terms— Person ReID, Domain Generalization, Feature Perturbation, Distribution Alignment, Invariant Representation

1. INTRODUCTION

Person Re-identification (ReID) aims to identify the images of the same pedestrians captured by non-overlapping cameras at different times and locations. It has achieved remarkable success when both training and testing are performed in the same domains [1, 2]. However, the widely held i.i.d. assumption does not always hold in real-world ReID scenarios due to significantly diverse viewing conditions at different locations of biased distributions at different camera views, and more generally across different application domains. As a result, a well-trained model with i.i.d. assumption can degrade significantly when applied to unseen new target domains [3–5]. To that end, Domain Generalization (DG) [6–8], which aims at learning a domain-agnostic model, has drawn

increasing attention in the ReID community. It is a more practical and challenging problem, which requires no prior knowledge about the target test domain to achieve “out-of-the-box” deployment.

Recent attempts on generalized ReID aim to prevent models from overfitting to the training data in source domains from either a local perspective by manipulating the data distribution of each domain, or a global view by representing the samples of all domains in a common representational space. The local-based methods [9–12] are usually implemented by feature perturbation and/or normalization. However, the perturbed distributions constructed from the original data of a single source domain is subject to subtle distribution shift and domain biased. On the other hand, the global-based approaches [4, 7, 13, 14] aim to align the feature distributions of multiple domains so that the per-domain data characteristic (*i.e.*, mean and variance of the data distribution which is assumed to be a Gaussian distribution) is ignored when representing images of different domains. They often explicitly pre-define a target distribution to be aligned, or implicitly learn a global consensus by training a single model with data from all the source domains. [4, 14] However, even the domain gap is reduced by such a global regularization from restricted ‘true’ distributions, the learned representations are inherently domain-biased toward the consensus of the multiple seen training domains rather than the desired universal distribution scalable to unseen target domains given the number of domains available for training is always limited.

In this work, we present a *Feature-Distribution Perturbation and Calibration* (PECA) model to accomplish generalized ReID. This is achieved by regularizing model training simultaneously with local distribution perturbation and global distribution calibration. Specifically, on the one hand, we introduce the local perturbation module to diversify the feature distribution based on a perturbing factor estimated per domain, which enables the model to be more invariant to distribution shifts. On the other hand, we propose to simultaneously calibrate the feature distributions across all the source domains, so to eliminate the domain-specific data characteristics in feature representations that are potentially caused by identity-irrelevant redundancy. Both the proposed local perturbation and global calibration modules reinforce

the same purpose of regularizing the model training, but they are devised in different hierarchies and complementary to each other. Different from the existing methods which consider only partially from the local or global perspectives, our method handles both to promote the model in learning domain-agnostic representations.

Contributions of this work are three-fold: **(1)** To our best knowledge, we make the first attempt to exploit jointly the local feature-distribution perturbation and the global feature-distribution calibration for improving the model’s generalizability to arbitrary unseen domains while maintaining its discrimination. **(2)** We formulate the PECA model which associates a local perturbation module (LPM) to diversify per-domain feature distribution so to refrain the model from overfitting to each source domain, and a global calibration module (GCM) to further eliminate domain bias by aligning the distribution of multiple source domains. PECA simultaneously regularizes both to strike the optimal balance between these two competing objectives. **(3)**. Extensive experiments verify the superior generalizability of the proposed PECA model over the state-of-the-art DG models on a wide range of ReID datasets by a notable margin, *e.g.*, the mAP is improved absolutely by 5.8% on average.

2. METHODOLOGY

Given K source domains $\mathcal{D} = \{D^{(k)}\}_{k=1}^K$, the objective of generalized ReID is to derive a domain-agnostic model θ which is capable of extracting domain-invariant representations for identity retrieval by a distance metric, *e.g.*, Cosine similarity or Euclidean distance, for any *unseen* target domain D^t . This is inherently challenging due to the unpredictable domain gap between training and testing data.

2.1. Overview

In this work, we propose a *Feature-Distribution Perturbation and Calibration* (PECA) model to derive domain-agnostic yet discriminative ID representations. It regularizes the model training to satisfy simultaneously both local perturbation and global calibration. The local regularization is built to perform per-domain *feature-distribution* diversification, and the global calibration is designed to achieve cross-domain *feature-distribution* alignment, as shown in Figure 1. During training, for each source domain $D^{(k)}$, a batch of samples $(x^{(k)}, y^{(k)})$, representing the pedestrian image and identity label, is fed into the network backbone to extract the feature map $e^{(k)}$. Then we perform per-domain diversification with Local Perturbation Module (LPM) to enable the local model to be invariant against per-domain shifts by training with the perturbed features $\{\hat{e}^{(k)}\}_{k=1}^k$. The balancing Global Calibration Module (GCM) further regularizes the model learning by aligning the holistic representation (the input feature of the classifier) into a common feature space constructed from

the global memory bank \mathcal{M} regardless of domain label. To distinguish the holistic representation from the intermediate representation $e^{(k)}$, we note it as $v^{(k)} \in \mathbb{R}^{B \times d}$ and its perturbed counterpart as $\hat{v}^{(k)}$ correspondingly, where d is a hyperparameter to the representation dimension, and B denotes the batch size (equal for all the domains).

2.2. Local Feature-Distribution Perturbation

Given an intermediate feature $e_i^{(k)} \in \mathbb{R}^{B \times C \times H \times W}$ extracted from the source domain $D^{(k)}$ at i -th layer, the objective of LPM is to perturb per-domain features to avoid local-domain overfitting. For notation clarity, we omit the layer index i in the following formulations. Inspired by feature augmentation [15] and Instance Normalization (IN) [16, 17], LPM performs perturbation by randomly substituting the transformation factors of IN. Specifically, we first calculate the channel-wise moments $\mu(e^{(k)}) \in \mathbb{R}^{B \times C}$ and $\sigma(e^{(k)}) \in \mathbb{R}^{B \times C}$ for IN as

$$\begin{aligned} \mu(e^{(k)}) &= \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W e_{h,w}^{(k)}, \\ \sigma^2(e^{(k)}) &= \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (e_{h,w}^{(k)} - \mu(e^{(k)}))^2. \end{aligned} \quad (1)$$

As suggested by [11], these statistical moments encode not only style information but also certain task-relevant information dedicated to ReID. Instead of discarding all of them for style bias reduction as adopted in [12, 18], we propose to maintain the discrimination while increasing the local-domain data diversity by holistically shifting its distribution. This is achieved by perturbing the per-domain instance moments as

$$\begin{aligned} \hat{\mu}(e^{(k)}) &= \mu(e^{(k)}) + \epsilon_\mu \mathbf{h}(\mu(e^{(k)})), \\ \hat{\sigma}(e^{(k)}) &= \sigma(e^{(k)}) + \epsilon_\sigma \mathbf{h}(\sigma(e^{(k)})), \end{aligned} \quad (2)$$

where $\mathbf{h}(\cdot) \in \mathbb{R}^C$ calculate the perturbation factors which are mathematically the standard deviation, they reflect the dispersed level of the local domain, and ensures the perturbation within a plausible range, so to avoid over-perturbation which causes model collapse, or under-perturbation which cannot provide any benefit in model learning. ϵ_μ and ϵ_σ are sampled randomly from the normal distribution to introduce variability in both the direction and intensity of perturbation, to ensure a diverse set of perturbed features. We subsequently perform feature transformation by substituting the local-domain moments with broadcasting subtraction as

$$\hat{e}^{(k)} = \hat{\sigma}(e^{(k)}) \frac{e^{(k)} - \mu(e^{(k)})}{\sigma(e^{(k)})} + \hat{\mu}(e^{(k)}). \quad (3)$$

By introducing the perturbed representation $\hat{e}^{(k)}$, the per-domain feature becomes more diverse so to improve the model’s generalizability against the per-domain shift.

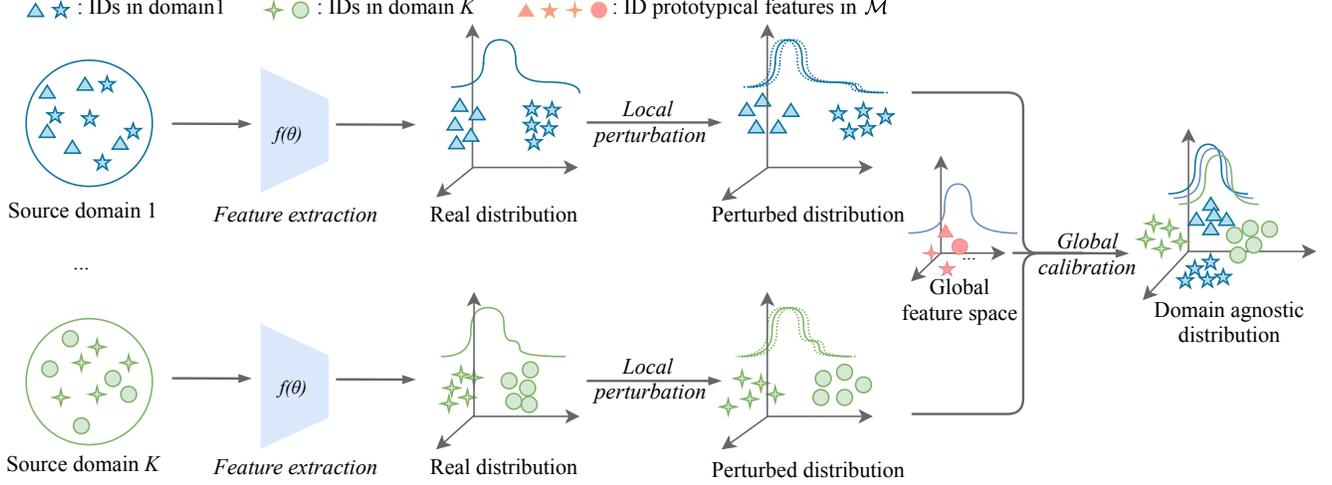


Fig. 1. Overview of the proposed *Feature-Distribution Perturbation and Calibration (PECA)* model. The overall objective is to derive generic feature representation by avoiding model overfitting to the source domains, which is achieved by *Local Perturbation Module* to enforce the learned feature invariant to per-domain distribution shifts caused by perturbation, and *Global Calibration Module* to align cross-domain distribution regardless of domain annotations.

2.3. Global Feature-Distribution Calibration

The global calibration module (GCM) is complementary to LPM by aligning the distribution of cross-domain features into a common feature space. GCM considers the association between the perturbed holistic representation $\hat{v}^{(k)}$ and a global memory bank \mathcal{M} . Specifically, we calculate the global statistical moments $\mu_g \in \mathbb{R}^d$ and $\sigma_g \in \mathbb{R}^d$ in each training iteration as

$$\begin{aligned} \mu_g &= \frac{1}{K} \sum_{k=1}^K \frac{1}{N^{(k)}} \sum_{n=1}^{N^{(k)}} \mathcal{M}_n^{(k)}, \\ \sigma_g^2 &= \frac{1}{K} \sum_{k=1}^K \frac{1}{N^{(k)}} \sum_{n=1}^{N^{(k)}} (\mathcal{M}_n^{(k)} - \mu_g)^2, \end{aligned} \quad (4)$$

where $\mathcal{M}_n^{(k)} \in \mathbb{R}^d$ is the prototypical feature of the n -th identity in the k -th domain. These global statistical moments depict a feature space shared by the prototypical representations on \mathcal{M} for all the identities. Subsequently, the holistic representations are calibrated into the joint feature space by

$$\mathcal{L}_g = \frac{1}{K} \sum_{k=1}^k (\|\mu(\hat{v}^{(k)}) - \mu_g\|_1 + \|\sigma(\hat{v}^{(k)}) - \sigma_g\|_1). \quad (5)$$

Here, $\mu(\hat{v}^{(k)}) \in \mathbb{R}^d$ and $\sigma(\hat{v}^{(k)}) \in \mathbb{R}^d$ are the channel-wise mean and standard deviation of the perturbed representation $\hat{v}^{(k)}$. GCM enables the extracted features to fall into a domain-invariant space. The hierarchical regularization achieved by LPM and GCM makes the model generic in extracting domain-agnostic representations.

2.4. Training Pipeline

Learning objective. The proposed PECA model is jointly trained with the identity loss \mathcal{L}_{id} and the global regularization item \mathcal{L}_g as

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{id} + \lambda \mathcal{L}_g, \quad \mathcal{L}_{id} = - \sum_{j=1}^C p_j^{(k)} \log \tilde{p}_j^{(k)}, \\ \tilde{p}^{(k)} &= \text{Softmax}(\text{MC}(\hat{v}^{(k)})), \end{aligned} \quad (6)$$

where $p^{(k)}$ is a one-hot distribution activated at $y^{(k)}$, and λ balances the importance. The function $\text{MC}(\cdot)$ stands for the memory-based classifier [18, 19], which is refreshed by Exponential Moving Average with a momentum of 0.2.

3. EXPERIMENTS

Implementation details. We used ResNet50 [20] pre-trained on ImageNet to bootstrap our feature extractor. The batch size B for each domain was set to 128, including 16 randomly sampled identities and 8 images for each identity. All images were resized to 256×128 . We randomly augmented the training data by cropping, flipping, and colorjitter. The proposed PECA was trained 60 epochs by Adam optimizer [21], and we adopted the warm-up strategy in the first 10 epochs to stabilize model training. The learning rate was initialized as $3.5e-4$ and multiplied by 0.1 at 30th and 50th epoch. The momentum for the memory update was set to 0.8. The dimension of extracted representations was conventionally set to 2048. All

Table 1. Comparison with SOTA methods. Best results are highlighted in **bold**, while the second-best results are underlined.

Method	PRID		GRID		VIPeR		iLIDs		Average	
	mAP	Rank-1								
Agg_PCB [33]	32.0	21.5	44.7	36.0	45.4	38.1	73.9	66.7	49.0	40.6
MLDG [34]	35.4	24.0	23.6	15.8	33.5	23.5	65.2	53.8	39.4	29.3
PPA [35]	45.3	31.9	38.0	26.9	54.5	45.1	72.7	64.5	52.6	42.1
DIMN [23]	52.0	39.2	41.1	29.3	60.1	51.2	78.4	70.2	57.9	47.5
DDAN [36]	67.5	62.9	50.9	46.2	<u>60.8</u>	<u>56.5</u>	81.2	78.0	65.1	60.9
DML [9]	60.4	47.3	49.0	39.4	58.0	49.2	<u>84.0</u>	77.3	62.9	53.3
META [37]	<u>71.7</u>	61.9	<u>56.6</u>	51.0	60.4	53.9	83.9	<u>79.3</u>	<u>68.2</u>	<u>61.5</u>
PECA (Ours)	72.2	<u>62.7</u>	59.4	<u>48.4</u>	70.1	61.2	85.7	79.8	71.9	63.0

the experiments were conducted on the PyTorch [22] framework with four A100 GPUs. Mean average precision (mAP) and Rank-1 accuracy are adopted as evaluation metrics.

3.1. Comparisons to the State-Of-The-Art Methods

We follow the conventional benchmarking setting [9, 11, 23], by using five datasets, including Market1501 (M) [24], MSMT17 (MT) [5], CUHK02 (C2) [25], CUHK03 (C3) [26], CUHK-SYSU (CS) [27], and DukeMTMC (D) [28], as source domains, and evaluated the generalizability on four datasets of different domains not contributing to training (unseen), which are PRID [29], GRID [30], VIPeR [31], and iLIDs [32]. All the images in the source domains were used for training, regardless of the original training or testing splits. We performed 10-trial evaluations by randomly splitting query/gallery sets, and reported the averaged performance in Table 1, which shows the superiority of the proposed PECA over the state-of-the-art (SOTA) competitors.

3.2. Ablation Study

Components analysis. We investigated the effects of different components in PECA model design to study individual contributions. We trained a baseline model with only identity loss \mathcal{L}_{id} . Table 2 shows that both the LPM and GCM are beneficial individually, and the benefits become clearer when they are jointly adopted as in the PECA model. From another perspective, it also verifies that solely considering the local or global regularization is biased, and it is non-trivial that the PECA explores both in a unified framework.

Table 2. Components analysis of LPM and GCM. PECA associates both. Evaluated on mAP (%).

Setting	PRID	GRID	VIPeR	iLIDs	Average
baseline	69.1	59.0	68.9	82.5	69.9
+LPM	71.5	58.0	69.7	85.3	71.1
+GCM	69.7	59.1	69.7	85.3	71.0
PECA	72.2	59.4	70.1	85.7	71.9

Table 3. Effects of the global calibration objective, whose importance is decided by the weight λ in Eq. 6.

Metric	w/o	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$	$\lambda = 100$
mAP	71.0	71.2	71.9	70.8	33.9
Rank-1	61.9	62.2	63.0	62.1	23.9

Effects of the global calibration objective. By linearly varying λ from 0.1 to 100, we observed from Table 3 that moderately applying GCM (e.g., 0.1 or 1) is beneficial to PECA’s generalizability; further increasing λ to a larger value (e.g., 10 or 100) brings more harm than help. This is because the learning process is dominated by the calibration regularization and the model can barely learn from identity loss, hence, the resulted feature is less discriminative. Given the above observations, we set $\lambda = 1$ in practice for our PECA model.

4. CONCLUSIONS

In this work, we presented a novel *Feature-Distribution Perturbation and Calibration* (PECA) model to learn generic yet discriminative representation. PECA simultaneously conducts model regularization on local per-domain feature-distribution and global cross-domain feature-distribution to learn a better domain-invariant feature space representation. Benefited from the diverse features synthesized by local perturbation, PECA expands per-domain feature distribution for more domain shifts robustness. From the global calibration, feature distributions of different domains are represented and holistically referenced in a shared feature space with their domain-specific data characteristics being ignored, resulting in higher model generalizability.

Acknowledgements

This work was supported by the China Scholarship Council, the Alan Turing Institute Turing Fellowship, Veritone. We utilised Queen Mary’s Apocrita HPC facility, supported by QMUL Research-IT.

5. REFERENCES

- [1] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *CVPR*, 2019. 1
- [2] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *CVPR*, 2020. 1
- [3] C. Luo, C. Song, and Z. Zhang, "Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup," in *ECCV*, 2020. 1
- [4] S. Choi, T. Kim, M. Jeong, H. Park, and C. Kim, "Meta batch-instance normalization for generalizable person re-identification," in *CVPR*, 2021. 1
- [5] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *CVPR*, 2018. 1, 4
- [6] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization in vision: A survey," *arXiv*, 2021. 1
- [7] K. Zhou, Y. Yang, T. Hospedales, and T. Xiang, "Learning to generate novel domains for domain generalization," in *ECCV*, 2020. 1
- [8] D. Mahajan, S. Tople, and A. Sharma, "Domain generalization using causal matching," in *ICML*, 2021. 1
- [9] Y. Dai, X. Li, J. Liu, Z. Tong, and L.-Y. Duan, "Generalizable person re-identification with relevance-aware mixture of experts," in *CVPR*, 2021. 1, 4
- [10] S. Yu, F. Zhu, D. Chen, R. Zhao, H. Chen, S. Tang, J. Zhu, and Y. Qiao, "Multiple domain experts collaborative learning: Multi-source domain generalization for person re-identification," *arXiv*, 2021. 1
- [11] X. Jin, C. Lan, W. Zeng, Z. Chen, and L. Zhang, "Style normalization and restitution for generalizable person re-identification," in *CVPR*, 2020. 1, 2, 4
- [12] J. Jia, Q. Ruan, and T. M. Hospedales, "Frustratingly easy person re-identification: Generalizing person re-id in practice," in *BMVC*, 2019. 1, 2
- [13] E. P. Ang, L. Shan, and A. C. Kot, "Dex: Domain embedding expansion for generalized person re-identification," in *BMVC*, 2021. 1
- [14] Y.-F. Zhang, H. Zhang, Z. Zhang, D. Li, Z. Jia, L. Wang, and T. Tan, "Learning domain invariant representations for generalizable person re-identification," *arXiv*, 2021. 1
- [15] P. Li, D. Li, W. Li, S. Gong, Y. Fu, and T. M. Hospedales, "A simple feature augmentation for domain generalization," in *ICCV*, 2021. 2
- [16] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *ICCV*, 2017. 2
- [17] X. Li, Y. Dai, Y. Ge, J. Liu, Y. Shan, and L.-Y. Duan, "Uncertainty modeling for out-of-distribution generalization," *ICLR*, 2022. 2
- [18] Y. Zhao, Z. Zhong, F. Yang, Z. Luo, Y. Lin, S. Li, and N. Sebe, "Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification," in *CVPR*, 2021. 2, 3
- [19] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *CVPR*, 2019. 3
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016. 3
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015. 3
- [22] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017. 4
- [23] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *CVPR*, 2019. 4
- [24] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015. 4
- [25] W. Li and X. Wang, "Locally aligned feature transforms across views," in *CVPR*, 2013. 4
- [26] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014. 4
- [27] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "End-to-end deep learning for person search," *arXiv*, 2016. 4
- [28] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *ICCV*, 2017. 4
- [29] M. Hirzer, C. Belezni, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian conference on Image analysis*, 2011. 4
- [30] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *IJCV*, 2010. 4
- [31] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *ECCV*, 2008. 4
- [32] W.-S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *BMVC*, 2009. 4
- [33] Y. Sun, L. Zheng, Y. Li, Y. Yang, Q. Tian, and S. Wang, "Learning part-based convolutional features for person re-identification," *IEEE TPAMI*, vol. 43, no. 3, 2019. 4
- [34] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *AAAI*, 2018. 4
- [35] S. Qiao, C. Liu, W. Shen, and A. L. Yuille, "Few-shot image recognition by predicting parameters from activations," in *CVPR*, 2018. 4
- [36] P. Chen, P. Dai, J. Liu, F. Zheng, M. Xu, Q. Tian, and R. Ji, "Dual distribution alignment network for generalizable person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 2, 2021, pp. 1054–1062. 4
- [37] B. Xu, J. Liang, L. He, and Z. Sun, "Mimic embedding via adaptive aggregation: Learning generalizable person re-identification," in *ECCV*, 2022. 4