Feature-Distribution Perturbation and Calibration for Generalized ReID

Qilei Li   **Jiabo Huang**   **Jian Hu**   **Shaogang Gong**

Queen Mary University of London

Abstract

Person Re-identification (ReID) has been advanced remarkably over the last 10 years along with the rapid development of deep learning for visual recognition. However, the i.i.d. (independent and identically distributed) assumption commonly held in most deep learning models is somewhat non-applicable to ReID considering its objective to identify images of the same pedestrian across cameras at different locations often of variable and independent domain characteristics that are also subject to view-biased data distribution. In this work, we propose a Feature-Distribution Perturbation and Calibration (PECA) method to derive generic feature representations for person ReID, which is not only discriminative across cameras but also agnostic and deployable to arbitrary unseen target domains. 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 on eight person ReID datasets and the proposed PECA model outperformed the state-of-the-art competitors by significant margins.

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 [25, 49, 45]. 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 will degrade significantly when applied to unseen new target domains [29, 3, 38]. To that end, Domain Generalization (DG) [52, 53, 30], 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 in a global view to represent the samples of all domains in a common representational space. The local-based methods [4, 41, 17, 15] are usually implemented by feature perturbation and/or normalization, as shown in Figure 1 (a). However, the perturbed distributions constructed from the original data of a single source domain is subject to subtle distribution shift and also domain biased. On the other hand, the global-based approaches [3, 1, 53, 44] aim to align the feature distributions of multiple domains so that the per-domain data characteristic (*i.e.*, mean and variance of the data
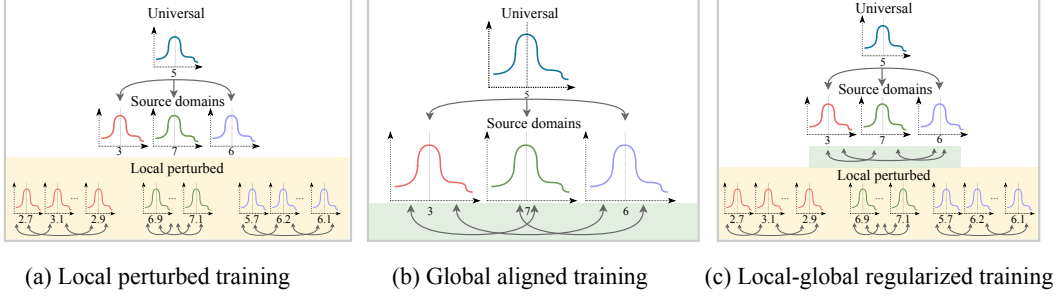
| (a) Local perturbed training | (b) Global aligned training | (c) Local-global regularized training |

Figure 1: Illustration of three training schemes in domain generalized ReID. The 'Universal' is ideally the distribution for any new target domains. The source domains are differentiated by different colors, and the perturbed distributions share the same color with the corresponding original. The number indicates the characteristics of that distribution, and similar value means a smaller domain gap, vice versa. The proposed PECA model simultaneously conducts local perturbation and global calibration to eliminate domain bias for learning a domain-agonistic representation.

distribution which is assumed to be a Gaussian distribution) is ignored when representing images of different domains, as illustrated in Figure 1 (b). They often explicitly pre-define a target distribution to be aligned towards, or implicitly learn a global consensus by training a single model with data of all the source domains. 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 with the objective to learn more generalizable discriminative representations for model deployment to unseen target domains. This is achieved by regularizing model training simultaneously with local distribution perturbation and global distribution calibration, as depicted in Figure 1(c). Specifically, on the one hand, as each source domain usually depicts limited numbers of pedestrians under certain scenarios, simply training from such data will lead to overfitting to the domain-specific inherently domain biased distribution, which harms the model's generalizability. To address this issue, we introduce 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, despite the unpredictable distribution gaps between different ReID data domains due to the undesirable scenario-sensitive information embedded in images specific to each domain, *e.g.*, the background, we consider that the features derived from different independent domains should share a high proportion of information as the universally applicable explanatory factors for domain-independent identity discrimination. In this regard, 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 the 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 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. We simultaneously regularize 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 Market1501 and the Rank-1 is by 6.0% on MSMT17.

2 Related Works

Generalized Person ReID. Person ReID aims to match the same identity across disjoint cameras. However, a well-trained ReID model always significantly degrades when evaluated on novel unseen domains, caused by the domain bias between training and testing data. To obtain a robust model in achieving "out-of-the-box" deployment, recent attempts on generalized ReID avoids models from overfitting to the source domains by either local domain manipulation or global cross-domain alignment to extract domain-invariant features more resistant to domain bias.

Local domain data manipulation. It is easy to train separate local models with labeled samples respective to each source domain, with subsequently a model aggregation [4, 41]. However, these local models would overfit to the corresponding source domain, while losing generalizability to others. A natural way to solve this problem is either to diverse the local training samples for learning a knowledgeable model, or to eliminate biased information within each source domain for learning a domain-unbiased model. Both solutions fall into the category of local domain manipulation, which alters the data distribution in a per-domain manner. For data diversification, the most intuitive approach is to perform augmentation, on either the raw image [2] or feature spaces [6]. For eliminating local domain bias, normalization, which regulates the data distribution based on the data statistics, has been widely studied recently Jin *et al.* [17] introduced instance normalization (IN) for restituting the style component out of an ID representation. Jia *et al.* [15] combined batch normalization with instance normalization in a unified architecture to achieve content and style unification. However, these local diagrams consider only per-domain information during feature perturbation, and is still subject to subtle distribution shifts. In this work, we propose to perturb the per-domain feature distribution to empower the model to be agnostic to holistic domain shift. The complementary regularization provided by the global distribution calibration remedy helps the learned model being invariant against both perturbed distribution shift and real domain gap, so to extract generic yet discriminative representation for any unseen domain.

Global distribution calibration. In contrast to the local approaches, methods based on global distribution calibration consider the cross-domain association by learning a shared representational space for all domains. These methods are built based on a straightforward assumption that source invariant features are also invariant to any unseen target domains [20]. In this spirit, DEX [1] dynamically performed e space expansion towards the direction of a zero-mean normal distribution with a covariance matrix estimated from the corresponding domain. Recent works [46, 3] took the idea of meta-learning with the aim of "learning to generalize" by randomly splitting available source domains into meta-training and meta-testing sets, to mimic real-world deployment scenarios. Such a scheme implicitly aligns the cross-domain feature distributions to a shared space by randomly setting the alignment target, *i.e.*, the meta-testing set. Zhang [44] proposed learning causal invariant feature by disentangling ID-specific and domain-specific factors for all the training samples from all the source domains, which enables the disentangled feature to well-preserved ID information while sharing the same feature space for all the domains. However, even aligning among multiple 'real' source domains can reduce the domain gap, the learned representations are still biased towards the consensus of the limited seen training domains, instead of the desired universal distribution scalable to unseen target domains. In this work, we propose to associate global alignment with local perturbation to achieve hierarchical regularization to avoid the model from overfitting to the source domains, so to learn domain-agnostic representations.

Data Augmentation. The conventional paradigm of data augmentation is to diversify *data*. GANs [8] have also been extensively explored to generate new *data* samples. Yang *et al.* [40] designed an image augmentation module which helps the network to learn domain-invariant representation by distilling information learned from the augmented samples to the teacher network. More recently, feature augmentation has emerged for semantic transformations. DeepAugment [11] perturbed features via stochastic operations by forwarding images through a pre-trained image-to-image model, to generate semantically meaningful and diverse samples. Li *et al.* [22] discovered that embedding white Gaussian noise in high-dimensional feature space provides substantive statistics reflective of cross-domain variability. Li *et al.* [26] proposed to model the feature uncertainty with a multivariate Gaussian distribution to perturb hierarchical features to diversify the feature space. In this work, we explore feature distribution augmentation in each source domain to achieve per-domain feature distribution diversification rather than diversifying the data, with the objective of making the model invariant to per-domain holistic shift in order to avoid model overfitting in each source domain.
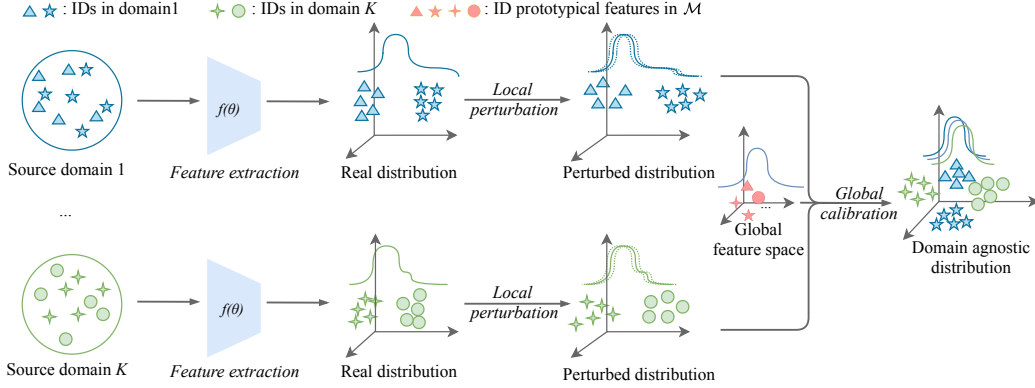
Figure 2: Overview of the proposed *Feature-Distribution Perturbation and Calibration* (PECA) model. The overall objective is to derive generic feature representation by avoiding model overfitting to the source domains, which is achieved by *Local Perturbation Module* to enforce the learned feature invariant to per-domain distribution shifts caused by perturbation, and *Global Calibration Module* to align cross-domain distribution regardless of domain annotations.

Distribution Alignment. The idea of distribution alignment aims to minimize the feature discrepancy between source and target domains. However, it is impossible for DG to explicitly perform such "target-oriented" alignment due to the absence of target domains during model training. With a straightforward assumption that features which are invariant to the source domain shift should also be invariant to any unseen target domain [20], DG approaches share the spirit to minimize the discrepancy among source domains to achieve distribution alignment. There are a wide variety of statistical metrics available for minimizing, such as Euclidean distance and f-divergences. In this regard, Li *et al.* [21] proposed to minimize the KL divergence of source domain features with a Gaussian distribution. Several researchers achieved distribution alignment by minimizing a single moment (mean or variance) [31, 13] or multiple moments [5, 7] calculated over a batch of source domain samples through either a projection matrix [7] or a non-linear deep network [16]. Li *et al.* [20] minimized the MMD distance by aligning the source domain feature distributions with a prior distribution via adversarial training [8]. In this paper, the proposed global distribution calibration operates on the same principle to align the source domains in learning a domain-agnostic model. Differently, we tailor the alignment objective for person ReID considering that all samples are depicting pedestrians, rather than predefine a deterministic distribution to align, *e.g.*, Gaussian or Laplace distributions. Specifically, we constructed a common feature space upon the ID prototypical representations stored in a global memory bank, so as to eliminate domain-biased information.

3 Balancing Feature-Distribution Local Perturbation with Global Calibration

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.

3.1 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 2. During training, for each source domain D^k, a batch of samples (x^k, y^k) is fed into the network backbone to extract the feature

map e^k. Then we perform per-domain diversification with Local Perturbation Module (LPM) as

$$\{\hat{e}^k\}_{k=1}^K = \{l(e^k)\}_{k=1}^K, \tag{1}$$

where $l(\cdot)$ is the function of 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 $\mathcal{M}$ 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. This global regularization is mathematically formulated as

$$\mathcal{L}_{\mathrm{g}}(\hat{v}^k, \mathcal{M}) = ||\mathrm{dist}(\hat{v}^k), \mathrm{dist}(\mathcal{M})||_1, \tag{2}$$

where $\mathcal{L}_g(\cdot)$ is the global regularization term aiming to align the distribution of holistic ID representations $\mathrm{dist}(\hat{v}^k)$ with the global distribution $\mathrm{dist}(\mathcal{M})$.

As complementary to the LPM, GCM focus on cross-domain regularization by pulling representations into a domain-agnostic space, thus empowering the generalizability of the ReID model for any unseen novel domain. With the collaboration of LPM and GCM, the PECA model can be trained with arbitrarily conventional ReID objectives in an end-to-end manner. When deployed to an unseen novel domain, a generic distance metric (*e.g.*, Euclidean or Cosine distance) is used to measure the pairwise representational similarity between the query image against the galleries for identity retrieval.

3.2 Local Feature-Distribution Perturbation

Given an intermediate feature representation $e_i^k \in \mathbb{R}^{B \times C \times H \times W}$ 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 [22] and Instance Normalization (IN) [14, 26], 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

$$\mu(e^k) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} e_{h,w}^k, \quad \sigma(e^k) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (e_{h,w}^k - \mu(e^k))^2. \tag{3}$$

As suggested by [17], 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 [15, 46], 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

$$\hat{\mu}(e^k) = \mu(e^k) + \epsilon_\mu \mathrm{h}(\mu(e^k)), \quad \hat{\sigma}(e^k) = \sigma(e^k) + \epsilon_\sigma \mathrm{h}(\sigma(e^k)), \tag{4}$$

where $\mathrm{h}(\cdot)$ calculate the perturbation factors, which are mathematically the standard derivation. 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 varies the perturbation intensity to guarantee the diversity of perturbed features, and both are randomly sampled from a standard normal distribution. We subsequently perform feature transformation by substituting the local-domain moments as

$$\hat{e}^k = \hat{\sigma}(e^k) \frac{e^k - \mu(e^k)}{\sigma(e^k)} + \hat{\mu}(e^k). \tag{5}$$

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.

3.3 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 $\mu_{\text{g}} \in \mathbb{R}^d$ and $\sigma_{\text{g}} \in \mathbb{R}^d$ in each training iteration as

$$\mu_{\text{g}} = \frac{1}{K} \frac{1}{N^k} \sum_{k=1}^{K} \sum_{n=1}^{N^k} \mathcal{M}_n^k, \quad \sigma_{\text{g}} = \frac{1}{K} \sum_{k=1}^{K} \sum_{n=1}^{N^k} (\mathcal{M}_n^k - \mu_g), \tag{6}$$

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

$$\mathcal{L}_{\text{g}}(\hat{v}^k, \mathcal{M}) = \frac{1}{K} \sum_{k=1}^{K} (||\mu(\hat{v}^k) - \mu_{\text{g}}||_1 + ||\sigma(\hat{v}^k) - \sigma_{\text{g}}||_1). \tag{7}$$

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 derivation 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.

3.4 Training Pipeline

Learning objective. Given the formulations of LPM and GCM, the proposed PECA can benefit from conventional learning supervisions. Specifically, the PECA model is jointly trained with a softmax cross-entropy loss $\mathcal{L}_{\text{id}}$ and the global regularization item $\mathcal{L}_{\text{g}}$ as

$$\mathcal{L} = \mathcal{L}_{\text{id}} + \lambda \mathcal{L}_{\text{g}}, \quad \mathcal{L}_{\text{id}}(x^k, y^k) = -\sum_{j=1}^{C} p_j^k \log \tilde{p}_j^k, \quad \tilde{p}^k = \text{Softmax}(\text{MC}(\hat{v}^k)). \tag{8}$$

The notations x^k and y^k are the raw input images sampled from domain D^k and its corresponding ID label, respectively, whilst p^k is a one-hot distribution activated at y^k. The function $\text{MC}(\cdot)$ stands for the memory-based classifier [51, 46], and λ decides the importance of $\mathcal{L}_{\text{g}}$ regarding the identity loss $\mathcal{L}_{\text{id}}$.

Memory bank update. In each training iteration, once the network parameters are updated according to $\mathcal{L}$ (Eq. (8)), the memory bank $\mathcal{M}$ is then refreshed by Exponential Moving Average (EMA) as

$$M_{y^k}^k = \beta M_{y^k}^k + (1 - \beta)\hat{v}^k, \quad k = \{1, \dots, K\}, \tag{9}$$

in which β is the EMA momentum. The prototypical features in the memory bank $\mathcal{M}$ is iteratively updated with the latest corresponding ID representations. Consequently, a more discriminative feature space will be yielded by $\mathcal{M}$ for global alignment.

4 Experiments

| (a) CUHK02 | (b) CUHK03 | (c) DukeMTMC | (d) Market1501 | (e) MSMT17 |

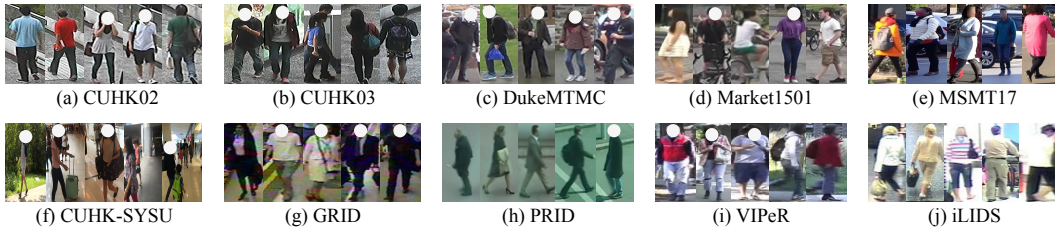| (f) CUHK-SYSU | (g) GRID | (h) PRID | (i) VIPeR | (j) iLIDS |

Figure 3: Example identity samples from different domains. Significant domain gaps are caused by the variation on nationality, illumination, viewpoints, resolution, scenario, etc.

Datasets and protocols. We conducted multisource domain generalization on a wide-range of benchmarks, including Market1501 (M) [47], DukeMTMC (D) [50], MSMT17 (MT) [38], CUHK02 (C2) [23], CUHK03 (C3) [24], CUHK-SYSU (CS) [39], and four small datasets including PRID [12], GRID [28], VIPer [9], and iLIDs [48]. The statistics of these datasets are shown in Table 1, and a few samples are visualized in Figure 3, which reveals significant domain gaps caused by variations on nationality, illumination, viewpoint, resolution, scenario, etc. Mean average precision (mAP) and CMC accuracy on Rank-1 are adopted as evaluation metrics.

Implementation details. We used ResNet50 [10] pre-trained on ImageNet to bootstrap our feature extractor. The batch size was set to 128, including 16 identities and 8 images for each. 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 optimizor [18], 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 the experiments were conducted on the PyTorch [33] framework with four A100 GPUs.

Table 1: Statistics of ReID datasets.

Datasets	Probe		Gallery	
	ID	Img	ID	Img
PRID [12]	100	100	649	649
GRID [28]	125	125	900	900
VIPeR [9]	316	316	316	316
iLIDS [48]	60	60	60	60
	Abbr.		ID	Img
Market1501 [47]	M		1,501	29,419
DukeMTMC [50]	D		1,812	36,411
MSMT17 [38]	MS		4,101	126,441
CUHK02 [23]	C2		1,816	7,264
CUHK03 [24]	C3		1,467	14,097
CUHK-SYSU [39]	CS		11,934	34,574

4.1 Comparisons to the State-Of-The-Art

Comparison under the traditional benchmark setting. Under the existing benchmark setting [4, 17, 36], five datasets (M+D+C2+C3+CS, as in Table 1) were used as source domains, and the generalizability was evaluated on four *small-scale* datasets of different domains not contributing to training (unseen), which are PRID, GRID, VIPeR, iLIDs. All the images in the source domains were used for training, without the original training or testing splits. Being consistent with existing performance evaluation protocols [17, 36], we performed 10-trail evaluations by randomly splitting query/gallery sets, and reported the averaged performance in Table 2, which shows the considerable superiority of the proposed PECA over the state-of-the-art (SOTA) competitors.

Table 2: Comparisons with the SOTA methods under traditional setting. Best results are in **bold**.

Method	PRID		GRID		VIPeR		iLIDs		Average	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
AggAlign [43]	25.5	17.2	24.7	15.9	52.9	42.8	74.7	63.8	44.5	34.9
Reptile [32]	26.9	17.9	23.0	16.2	31.3	22.1	67.1	56.0	37.1	28.0
CrossGrad [35]	28.2	18.8	16.0	9.0	30.4	20.9	61.3	49.7	34.0	24.6
Agg_PCB [37]	32.0	21.5	44.7	36.0	45.4	38.1	73.9	66.7	49.0	40.6
MLDG [19]	35.4	24.0	23.6	15.8	33.5	23.5	65.2	53.8	39.4	29.3
PPA [34]	45.3	31.9	38.0	26.9	54.5	45.1	72.7	64.5	52.6	42.1
DIMN [36]	52.0	39.2	41.1	29.3	60.1	51.2	78.4	70.2	57.9	47.5
SNR [17]	66.5	52.1	47.7	40.2	61.3	52.9	89.9	84.1	66.3	57.3
RaMoE [4]	67.3	57.7	54.2	46.8	64.6	56.6	**90.2**	**85.0**	69.1	61.5
PECA (Ours)	**72.2**	**62.7**	**59.4**	**48.4**	**70.1**	**61.2**	85.7	79.8	**71.9**	**63.0**

Comparison under large-scale benchmark setting. We further evaluated our model on four *large-scale* datasets (M+D+C3+MS) with the 'leave-one-out' strategy, namely taking three datasets used as source domains for model training, and one left out as an unseen target domain. Under this setting, The original train splits in the three source domains were used for training, while the test split on the unseen target domain was used for testing, same as in [46]. The evaluation results in Table 3 show that PECA outperforms the SOTA competitors by a compelling margin, Specially, on the more challenging datasets CUHK03 and MSMT17 with larger domain gaps to the other datasets, all methods give relatively poorer generalization performances. In comparison, our PECA model gains greater advantage over the other methods especially on Rank-1 scores. This suggests PECA's better scalability with greater potential in real-world deployment to different unseen target domains.

Table 3: Comparisons with the SOTA generalized person ReID models on large-scale datasets.

Method	Market-1501		DukeMTMC		CUHK03		MSMT17		Average	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
QAConv$_{50}$ [27]	39.5	68.6	43.4	64.9	19.2	22.9	10.0	29.9	28.0	46.6
M3L [46]	51.1	76.5	48.2	67.1	30.9	31.9	13.1	32.0	35.8	51.9
M3L(IBN) [46]	52.5	78.3	48.8	67.2	31.4	31.6	15.4	37.1	37.0	53.5
PECA (Ours)	**58.3**	**81.4**	**49.8**	**70.0**	**34.1**	**35.5**	**17.7**	**43.1**	**40.0**	**57.5**

4.2 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 $\mathcal{L}_{id}$, and then incorporated it with either LPM or GCM as well as both (PECA). Table 4 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 to learn a more generic representation.

Table 4: Components analysis of LPM and GCM. PECA incorporates both in a unified framework.

Setting	Market-1501		DukeMTMC		CUHK03		MSMT17		Average	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
baseline	54.1	78.5	49.0	68.1	31.1	31.9	14.9	38.1	37.3	54.1
+LPM	57.9	80.4	49.4	69.4	32.7	33.2	**17.7**	42.8	39.4	56.5
+GCM	55.0	79.5	49.0	68.5	32.6	33.6	16.1	39.4	38.2	55.2
PECA	**58.3**	**81.4**	**49.8**	**70.0**	**34.1**	**35.5**	**17.7**	**43.1**	**40.0**	**57.5**

Setting	PRID		GRID		VIPeR		iLIDs		Average	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
baseline	69.1	59.0	59.0	48.4	68.9	60.1	82.5	74.5	69.9	60.5
+LPM	71.5	61.2	58.0	**48.5**	69.7	60.9	85.3	78.7	71.1	62.3
+GCM	69.7	59.5	59.1	**48.5**	69.7	60.7	85.3	78.7	71.0	61.8
PECA	**72.2**	**62.7**	**59.4**	48.4	**70.1**	**61.2**	**85.7**	**79.8**	**71.9**	**63.0**

Discrimination and generalization trade-off. There is a trade-off between being discriminative to the source domains, and being generalized to the target domains [42]. We quantitatively assessed the proposed PECA model in this regard. The results in Table 5 indicate the baseline method fails to generalize well to the target domains but yielded compelling discrimination capacity in the source domains, which is likely due to overfitting. As a comparison, our PECA gains notable improvements in generalization ability with only slight performance drops in the source domains. This implies that PECA can effectively balance the generalization and discrimination of feature representations, so to be applied to any novel unseen domains.

Table 5: Local discrimination and global generalization trade-off.

Setting	Source Average		Target: M		Source Average		Target: D	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
baseline	**58.6**	**73.8**	54.1	78.5	**63.0**	**77.4**	49.0	68.1
LPGC	57.9	73.0	**58.3**	**81.4**	61.9	76.3	**49.8**	**70.0**

Setting	Source Average		Target: C		Source Average		Target: MS	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
baseline	**64.2**	**82.3**	31.1	31.9	**69.9**	**79.3**	14.9	38.1
LPGC	63.9	82.2	**34.1**	**35.5**	68.8	78.7	**17.7**	**43.1**
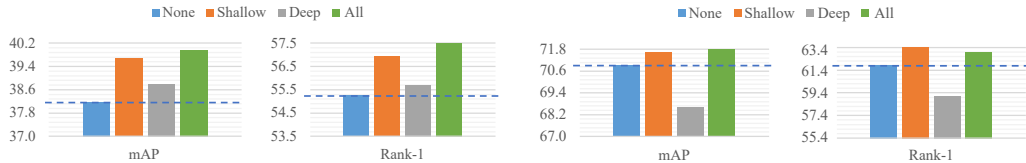
(a) Averaged performance under large-scale setting.   (b) Averaged performance under traditional setting.

Figure 4: Effects of distribution perturbation on different layers.

Effects of distribution perturbation on different layers. We studied the effects of perturbing the input distributions of various layers in our backbone network, including the 'Shallow' layers (the first convolution layer and a following residual block), and 'Deep' layers (the last two residual blocks). The results are shown in Figure 4. It is not a surprise that perturbing the shallow layers consistently improves the performance under both the traditional and large-scale settings, as perturbations in earlier stages helps enhance the invariance of most layers to distribution shift. However, solely

perturbing the deep layers exhibit distinct behaviors under different benchmark settings. This is because the training data under the traditional setting is relatively smaller with restricted diversity and perturbations in later stages tends to affect a limited part of the network that is insufficient to improve the model's generalization ability. Based on these observations, we propose to perturb all the layers to improve the robustness of the PECA model regardless of the dataset scale.

Effects of the global calibration objective. The importance of the global calibration objective for avoiding the model from overfitting to source domains is determined by the hyperparameter λ in Eq. (8). By linearly varying λ from 0.1 to 100, we observed from Table 6 that moderately applying GCM (*e.g.*, 0.1 or 1) is beneficial to PECA's generalizability; fur-

Table 6: Effects of the global calibration objective whose importance is decided by the weight λ in Eq. 8. Averaged performances are reported.

Setting	Traditional setting		Large-scale setting	
	mAP	Rank-1	mAP	Rank-1
PECA (*default*, $\lambda = 1$)	**71.9**	**63.0**	**40.0**	**57.5**
PECA w/o GPM	71.0	61.9	37.5	54.3
PECA w/ $\lambda = 0.1$	71.2	62.2	39.4	56.5
PECA w/ $\lambda = 10$	70.8	62.1	39.2	56.5
PECA w/ $\lambda = 100$	33.9	23.9	38.5	55.5

ther 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 the identity loss, hence, the resulted feature is less discriminative. We also observed that the traditional setting is relatively more sensitive to λ, as it holds much less training data for learning a robust model, and a similar phenomenon is shown in Figure 4. Given the above observations, we set $\lambda = 1$ in practice for our PECA model.

5   Conclusions

In this work, we presented a novel *Feature-Distribution Perturbation and Calibration* (PECA) model to learn generic yet discriminative representation in multiple source domains generalizable to arbitrary unseen target domains for more accurate unseen domain person ReID. PECA simultaneously conducts model regularization on local per-domain feature-distribution and global cross-domain feature-distribution in order to learn a better domain-invariant feature space representation. Benefited from the diverse features synthesized by local perturbation, PECA expands per-domain feature distribution to enable more robust to domain shifts. 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 (*i.e.*, mean and variance of feature distributions) being ignored, resulting in higher model generalizability. Experiments on extensive ReID datasets show the performance advantages of the proposed PECA model over a wide range of state-of-the-art competitors. Extensive ablation studies further provided in-depth analysis of the individual components designed in PECA model.

References

[1] Eugene PW Ang, Lin Shan, and Alex C Kot. Dex: Domain embedding expansion for generalized person re-identification. In *BMVC*, 2021.

[2] Feng Chen, Nian Wang, Jun Tang, Dong Liang, and Hao Feng. Self-supervised data augmentation for person re-identification. *Neurocomputing*, 415:48–59, 2020.

[3] Seokeon Choi, Taekyung Kim, Minki Jeong, Hyoungseob Park, and Changick Kim. Meta batch-instance normalization for generalizable person re-identification. In *CVPR*, 2021.

[4] Yongxing Dai, Xiaotong Li, Jun Liu, Zekun Tong, and Ling-Yu Duan. Generalizable person re-identification with relevance-aware mixture of experts. In *CVPR*, 2021.

[5] Sarah Erfani, Mahsa Baktashmotlagh, Masud Moshtaghi, Xuan Nguyen, Christopher Leckie, James Bailey, and Rao Kotagiri. Robust domain generalisation by enforcing distribution invariance. In *IJCAI*, 2016.

[6] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *ICCV*, 2019.

[7] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE TPAMI*, 39(7), 2016.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014.

[9] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 2008.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[11] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *ICCV*, 2021.

[12] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, 2011.

[13] Shoubo Hu, Kun Zhang, Zhitang Chen, and Laiwan Chan. Domain generalization via multidomain discriminant analysis. In *UCI*, 2020.

[14] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017.

[15] Jieru Jia, Qiuqi Ruan, and Timothy M Hospedales. Frustratingly easy person re-identification: Generalizing person re-id in practice. In *BMVC*, 2019.

[16] Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Feature alignment and restoration for domain generalization and adaptation. *arXiv*, 2020.

[17] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *CVPR*, 2020.

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[19] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018.

[20] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, 2018.

[21] Haoliang Li, YuFei Wang, Renjie Wan, Shiqi Wang, Tie-Qiang Li, and Alex Kot. Domain generalization for medical imaging classification with linear-dependency regularization. *NeurIPS*, 2020.

[22] Pan Li, Da Li, Wei Li, Shaogang Gong, Yanwei Fu, and Timothy M Hospedales. A simple feature augmentation for domain generalization. In *ICCV*, 2021.

[23] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, 2013.

[24] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.

[25] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, 2018.

[26] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and Ling-Yu Duan. Uncertainty modeling for out-of-distribution generalization. *ICLR*, 2022.

[27] Shengcai Liao and Ling Shao. Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting. In *ECCV*, 2020.

[28] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 2010.

[29] Chuanchen Luo, Chunfeng Song, and Zhaoxiang Zhang. Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. In *ECCV*, 2020.

[30] Divyat Mahajan, Shruti Tople, and Amit Sharma. Domain generalization using causal matching. In *ICML*, 2021.

[31] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, 2013.

[32] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv*, 2018.

[33] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[34] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille. Few-shot image recognition by predicting parameters from activations. In *CVPR*, 2018.

[35] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. *arXiv*, 2018.

[36] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *CVPR*, 2019.

[37] Yifan Sun, Liang Zheng, Yali Li, Yi Yang, Qi Tian, and Shengjin Wang. Learning part-based convolutional features for person re-identification. *IEEE TPAMI*, 43(3), 2019.

[38] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, 2018.

[39] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv*, 2016.

[40] Fu-En Yang, Yuan-Chia Cheng, Zu-Yun Shiau, and Yu-Chiang Frank Wang. Adversarial teacher-student representation learning for domain generalization. In *NeurIPS*, 2021.

[41] Shijie Yu, Feng Zhu, Dapeng Chen, Rui Zhao, Haobin Chen, Shixiang Tang, Jinguo Zhu, and Yu Qiao. Multiple domain experts collaborative learning: Multi-source domain generalization for person re-identification. *arXiv*, 2021.

[42] Pengchuan Zhang, Qiang Liu, Dengyong Zhou, Tao Xu, and Xiaodong He. On the discrimination-generalization tradeoff in gans. In *ICLR*, 2018.

[43] Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, and Jian Sun. Alignedreid: Surpassing human-level performance in person re-identification. *arXiv*, 2017.

[44] Yi-Fan Zhang, Hanlin Zhang, Zhang Zhang, Da Li, Zhen Jia, Liang Wang, and Tieniu Tan. Learning domain invariant representations for generalizable person re-identification. *arXiv*, 2021.

[45] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *CVPR*, 2020.

[46] Yuyang Zhao, Zhun Zhong, Fengxiang Yang, Zhiming Luo, Yaojin Lin, Shaozi Li, and Nicu Sebe. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In *CVPR*, 2021.

[47] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015.

[48] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *BMVC*, 2009.

[49] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *CVPR*, 2019.

[50] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017.

[51] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR*, 2019.

[52] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization in vision: A survey. *arXiv*, 2021.

[53] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *ECCV*, 2020.