Multi-source domain generalization peron re-identification with knowledge accumulation and distribution enhancement

Wanru Peng[1] · Houjin Chen[1] · Yanfeng Li[1] · Jia Sun[1]

Abstract

Domain generalization person re-identification (re-ID) is a more realistic task that aims to learn a model with multiple labeled source domains and generalize its inference to an unseen domain. However, under joint training of multi-source domains, some difficulties such as domain imbalance, domain shifts and small training batches not only influence the representation learning of the model but also lead to poor generalization performance. In this work, we enhance the generalization performance by proposing two strategies including knowledge accumulation and distribution enhancement for multi-source domain generalization person re-ID. Specifically, we encourage the learning of semantically significant features globally by establishing a simple but effective knowledge accumulation feature classifier (KAFC). In which, continuous learning guided by the label information explicitly provides stable prototype matching for each source domain, directly solving the problem that unstable parameter updates due to the small-batch data and source domain shifts. Secondly, to enhance the generalization ability and learn more robust representations, a multi-mix batch normalization (MMBN) module is introduced to generate the mixture features with cross-domain distribution information, where an approach based on the exponential moving average is adopted for capturing the domain distribution knowledge.

Extensive experiments on multiple widely-used benchmarks demonstrate that our method can effectively improve the generalization capability on unseen domains.

Keywords Person re-identification · Multi-Source Domain generalization · Continual Learning · Data Augmentation

1 Introduction

Person re-identification (re-ID) aims at matching the same person in different scenes, which has received a great deal of attention in the past decade due to its research and practical value. As one of the application scenarios, DG re-ID aims to learn a model with multiple labeled source domains and generalize its inference to an unseen domain and is more practical as it assumed less and emphasizes that the target domain is inaccessible in the training process. Generalizing the person re-identification (re-ID) algorithms to unseen domains is an important research topic for intelligent video surveillance, interest in more complicated problems has motivated more research on this task [1–5] and this paper is devoted to the DG re-ID task.

Although a large number of domain generalization methods have been proposed, such as data augmentation [6–9], representation learning [10–14], and learning strategy [15–17], many of them discuss the isomorphic tasks where the source and targets share the same label space. These isomorphic methods are difficult to directly adapt to the task of re-ID, and there are relatively few studies on DG re-ID, for there are always different classes between the source domains and target domains in open-set tasks. The approach proposed in this paper is based on the characteristics of the re-identification task, with emphasis on the class discriminability and feature generalization ability of pedestrians. Recent DG methods typically practice across multiple source domains to learn a more generalizable representation or model for new unseen target domains. In deep neural networks with strong descriptive power, if we train the single model with fixed batches from different source domains, the domain gaps will influence the generalization ability of the model and may cause overfitting and catastrophic forgetting. In this way, the multi-source DG re-ID problem is naturally

✉ Houjin Chen
  hjchen@bjtu.edu.cn

1 School of Electronic and Information Engineering, Beijing
   Jiaotong University, Beijing, China

transferred into two aspects, namely how to conduct the representation learning under the domain gap; and on the other hand, how to fully make use of the domain gap of source domains to improve the generalization on an unseen target.

To enhance the discrimination capabilities, some methods [18, 19] employ adversarial learning or disentanglement models to learn domain invariant features or achieve feature disentanglement, while the introduction of bias and loss of some domain-specific characteristics are inevitable. Other attempts take advantage of the adaptive normalization techniques or architectures. For example, Jin et al. [4] proposed a style normalization and restitution (SNR) module to reduce the domain gap caused by variations in appearance style. In [14], the combinations of different normalization techniques are presented to show that adaptively learning the normalization technique can improve DG. However, these attempts are mainly from the perspective of improving the model's adaptation to the unseen domain and fail to deeply analyze the impact of source domain shifts to feature learning in the training process, limiting the applicability of them in multi-source training. Inspired by some recent works [20] about memory-bank and continual learning, this paper considers that the way humans acquire knowledge can be effectively utilized in domain generalization by remembering the shared information of each class and integrating new knowledge into representations without forgetting previously learned knowledge. It can be beneficial in maintaining and improving representation learning under multi-domain shifts. Thus, the solution of this paper is to perform fine-grained class matches by adding the notion of continual learning into the classifier design.

In terms of exploiting the gaps in the source domain for better generalization, one of the most popular approaches is meta-learning-based models [3, 15, 16]. These models can mimic the realistic training–testing domain shifts to improve the generalization of the model, but at the same time, a significant increase in model complexity and training difficulty is inevitable. To learn a robust model, the most intuitive way is to conduct data augmentation. For example, there are some methods [7, 9] that explored Mixup [8] and AdaIN [21] techniques in original space or feature space to achieve fast stylization and in order to expose the model under the abundant domain styles, which achieved promising performance on popular benchmarks. Unlike directly mixing at the instance level, this paper adopts a probabilistic combination that mixes the bottom domain-level feature statistics for training a more robust model. All in all, we want to remain conceptually and computationally simple while the intra-domain discriminability and inter-domain variability are both taken into account in our model.

In this work, a novel multi-source DG re-ID model is proposed, which consists of two strategies, knowledge accumulation and distribution enhancement. On the one hand, to

enhance discrimination capabilities, a knowledge accumulation weight updating strategy based on the notion of continual learning is applied for classifier learning. This requires the model to accumulate prior knowledge from the seen source domains based on the memory bank and then use the knowledge to update the weights of feature classifies, which directly solves the problem that unstable parameter updates due to the small-batch data and source domain shifts.

On the other hand, this paper makes full use of the domain gaps by developing a multi-mix batch normalization (MMBN) module, where the exponential moving average (EMA) technique is used to update the domain distribution statistics, and a probabilistic mixture takes place between batch features and other domain statistics to generate the mixture embeddings. It helps improve the robustness of the model in cross-domain shifts, thereby facilitating the learning process of the whole framework. For further optimizing the model, a global contrastive loss (GCL) is designed to optimize the distance from the mixture embeddings of the original samples to their prototypes and a hybrid triplet loss (HTRI) is employed for the metric learning within small batches at different layers. Extensive experiments result on multiple widely-used benchmarks achieve satisfactory performance, which demonstrate that the knowledge accumulation and distribution enhancement can effectively improve the generalization capability on unseen domains. The contributions of this paper are summarized as follows:

(1) An effective framework for multi-source domain generalizable person re-ID is presented, where intra-domain discriminability and inter-domain variability are both considered to enhance generalization capability on the unseen domain.

(2) A novel knowledge accumulation feature classifier (KAFC) is designed, which alleviates the degrading impact of catastrophic forgetting and distributional shifts by adaptively updating previous knowledge and the old parameters.

(3) A data augmentation module MMBN is introduced to generate the mixture features with cross-domain distribution information, while GCL loss and HTRI loss are designed to further promote the representation learning and classifier learning of the model.

2 Related Work

2.1 Domain Generalization

The goal of domain generalization (DG) is to acquire knowledge from several related source domains and apply it to unseen target domains. The research literature on DG demonstrates great diversity and can be primarily categorized

into three aspects: (1) data augmentation. (2) representation learning (3) learning strategy. Data augmentation is a form of domain randomization that performs certain refinements or transformations to increase domain diversity thereby assisting in learning general representations. For example, Tobin et al. [22] first used this method to generate more training data from the simulated environment for generalization in the real environment. Representation learning methods [10, 19, 23] have also proven very effective, which aim to learn domain-invariant features or explicitly feature alignment between domains. From the optimization perspective, many methods [5, 15] focus on exploiting different training strategies to promote generalization capability, such as ensemble learning, meta-learning, and so on. In recent years, more multi-source DG methods have been studied for practical applications. However, most of them assume there are overlapping label spaces among multiple domains, making them unsuitable to be directly applied to the open-set person re-ID task, where the target domain has different identities and classes from the source domains.

2.2 Generalization Person Re-identification

Person re-ID has made great progress in recent years, during which plenty of methods have been introduced to address person re-ID tasks for different practical application assumptions. Specifically including fully-supervised methods, unsupervised methods, and domain generalization re-ID. The ability to generalize is crucial for person re-ID models when deployed to the unseen dataset in practical applications. To address this problem, several methods have been explored and the common goal of these is to learn more generalizable representations or models under domain shifts. The first category is about representation learning, the DG-re-ID methods based on representation learning encourage the model to disentangle person representations to learn robust domain-invariant features. For example, SNR [4] introduced IN layers to neural network architecture and disentangles the task-relevant features from the residual. To further improve the interpretability, one effort is the work of QAConv [24], which considered point-to-point image matching in deep feature maps and suggested that explicit matching is easier to generalize to the unknown domain than feature learning. Moreover, some studies [9, 23] on the multi-source DG-ReID shared the same network among multiple source domains to align distributions or had their own specific normalization layer to remove domain-specific styles. However, the benefits of these remain limited when there are significant domain shifts, small batch training data and imbalanced label distribution within and across domains. The second category that meta-learning-based methods [3, 16], although these methods proved to be very helpful for DG tasks, the complex training procedure made

model optimization difficult. Another promising solution for DG re-ID is data augmentation, recent papers [9, 25] pointed out the crucial role of the diversity of learned features from the synthetic data to prevent overfitting to the source domain in the re-ID task.

In contrast to the previous solutions, we approach the domain generalization issue from the point of view of prototype-wise matching and distribution enhancement. The common knowledge about class within the domain and the discrepancy between multi-source domains are both considered in this paper.

2.3 Continual Learning

Similar to transfer learning [26, 27], which apply existing knowledge to learn new knowledge between two different models. Continual learning is an active machine learning task for single model that seeks to continuously learn new knowledge to adjust itself and preserve most of the previously learned knowledge, just like humans learn knowledge.

Existing works can be divided into three categories, including knowledge distillation, parameter regularization and memory replay. Among them, some works prefer to rectify or post-process the output classification layer to mitigate bias caused by catastrophic forgetting and distributional shifts [28–32]. Hou et.al [29] designed a unified classifier and adopted intra-class knowledge distillation to solve the class imbalance problem between the base and new classes. Zhang et.al [30] proposed a continually evolved classifier for the few-shot incremental learning, and an adaptation module was employed to update the classifier weights at the global level. Another type of continual learning considers memory selection and generation for previous samples or gradients [20, 31], it is highly effective when storing training data is possible. [31] devised an exponential moving average framework to alleviate the degrading impact of forgetting and distributional shifts by adapting to a history of the old parameters. MemVir [20] memorizes both embedding features and class weights to utilize them as additional virtual classes. Not only utilizes augmented information for training but also alleviates a strong focus on seen classes for better generalization. There have also a number of efforts focused on the domain differences that the spatial and temporal dispersion of data brings to the task [33–35]. Yang et.al [35] first adopts a one-vs-all detector to discover persons who have been presented in previous cameras, which requires re-ID models to continuously learn informative representations without forgetting the previously learned on.

These methods give us more inspiration to fully leverage past knowledge to obtain a better representation for DG re-ID. However, we need to look further at the training characteristics of multi-source DG re-ID tasks, the overlapping of categories within domains, the shifts and imbalances between domains,
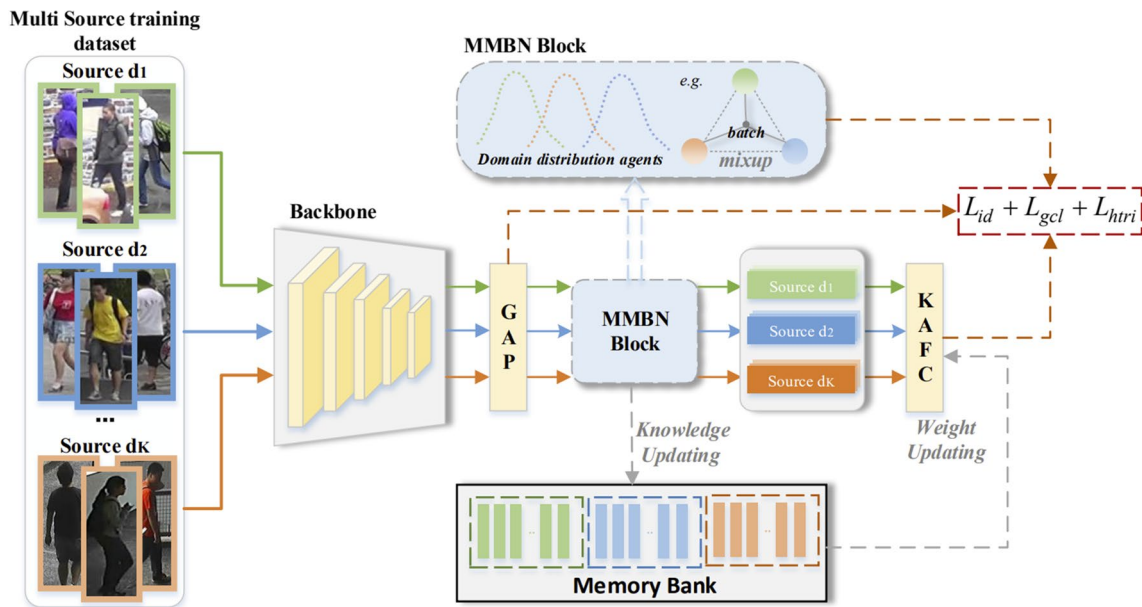
Fig. 1 The overall architecture of the proposed framework. It contains a parameters-shared backbone network to extract domain-invariant features, memory-based feature classifiers that update the weights under the global guidance of label information, and a multi-mix BN module for mixing the information across the domains

etc. In this paper, the classifier is designed based on the notion of continual learning, where class prototypes are memorized and updated continuously, meanwhile are utilized to control the optimization direction of the weights.

3 Methodology

In this section, we introduce the framework for multi-source domain generalization person re-ID. We first describe the overview of the framework in Section 3.1. Then the proposed knowledge accumulation feature classifier is presented in Section 3.2. To better conduct representation learning, a data augmentation module MMBN is adopted, which is introduced in Section 3.3.

3.1 Overview

In multi-source domain generalization person re-ID, we are given K source domains $D_s = \{d_1, d_2, \ldots, d_K\}$, where $\{N_1, N_2, \ldots, N_K\}$ denote the classes of $\{d_1, d_2, \ldots, d_K\}$, respectively, and the total class is N. The object of this paper is to generalize well in an unknown target domain by transferring knowledge of classification learned from the source domain based on full supervision. Note that the target domain is unlabeled and can't advance access.

This paper proposes a novel multi-source DG framework that focuses on designing the feature classifier with knowledge accumulation and mixing the multi-domain information to improve generalization. As shown in Fig. 1, a backbone that extracts features from images and can be chosen from various popular networks. Secondly, the memory bank is popularly employed in most recent methods, which is used to store the global class information of source domains according to the label. The feature vectors stored in the memory bank are used to update the feature classifiers' weights in each epoch. Moreover, the BN layer is replaced after the pooling layer with an MMBN module. The MMBN module is expected to mix information among multiple domains to further enhance feature robustness and reduce the inter-domain gaps.

3.2 Knowledge Accumulated Feature Classifier (KAFC)

To extract discriminative features, we assign FC layers for each domain as classifiers for training, where the linear classifier learns a corresponding classification pattern (which can also be regarded as a prototype for each person) for each class through weights W. In our baseline, the cross-entropy loss is used to optimize the network and learn the class weights W. For a domain dataset with N_i classes, given an image, we denote y_i as the truth ID label and the prediction logits can be defined as follows:

$$p(y_i|f_i) = \frac{\exp(W_{y_i}^T f_i + b_{y_i})}{\sum_{j=1}^{N_i} \exp(W_j^T f_i + b_j)} \tag{1}$$
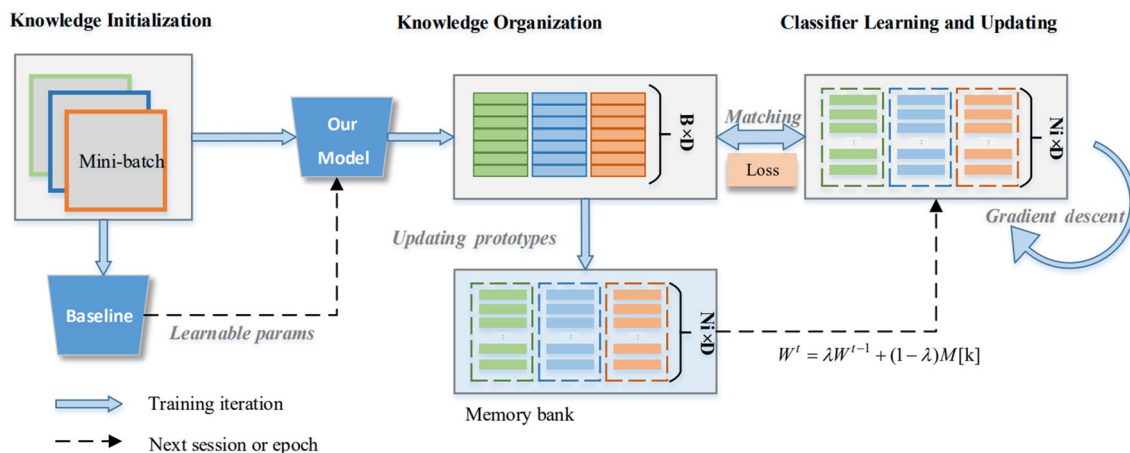
Fig. 2 Overview of KAFC, where D, N and B denote the feature dimensions, the number of classes, and the batch size, respectively. During the knowledge initialization stage, we pre-train the model by several epochs to obtain better initial representations for class prototypes. In each iteration of formal training, features are extracted from the network to update the memory bank. And we match them with the class weights and obtain the robust generalized cross-entropy loss. At the beginning of each training epoch, class prototypes in the memory bank are selected to update the current class weights

where f_i is a D-dimensional feature of the i-th image in a mini-batch, weight matrix $W \in \mathbb{R}^{N \times D}$ and bias b can be set as 0 for convenience.

Obviously, the learned weights are critical to the image prediction logits. However, due to the limitation of batch size and the shifts of domains, it is difficult to ensure the stability of the parameters when updating. Not only lead to the forgetting of prior knowledge but also affect learning new knowledge. Inspired by the idea of continual learning, we expect classifier learning can combine new knowledge with previous information at the global level. To achieve this goal, we provide classifiers with the knowledge accumulation strategy in which the class prototypes are memorized and continually updated. The main idea is to use label information and past knowledge as constraints to provide guidance for domain classifiers, which allows features and prior knowledge to be consistent and reduces the influence of unstable parameters. The knowledge accumulated strategy mainly includes three training stages: the knowledge initialization stage, the knowledge organization stage, and the classifier learning and updating stage, as shown in Fig. 2.

Knowledge Initialization. It is commonly known that in the early training process, features are not adequately expressed and the class prototypes are captured inaccurately. To reduce the bias caused by the ImageNet [36] and separate the domains, we pre-train the model by several epochs to obtain the initial class prototypes, where the class weights are initialized with the Kaiming normal initialization algorithm [37], and the ADAM optimizer is employed for gradient descent. Then, the baseline can be reused to extract features in the next stage.

Knowledge Organization. Based on the pre-trained baseline model, we continue training the model with the proposed method. A multi-domain memory bank $M = \{M_1, M_2, ..., M_K\}$ is utilized to obtain global information, where there are N class prototypes feature $F_C \in \mathbb{R}^{N \times D}$ of K source domain are stored, and N is the total class of all training domains. In each training iteration, the k-th centroid $M[k]$ is dynamically updated with the encoded features as follows:

$$M[k] \leftarrow mM[k] + (1-m)\frac{1}{|\mathcal{B}_k|} \sum_{x_i \in \mathcal{B}_k} f(x_i) \qquad (2)$$

where $f(x_i)$ is the feature vector of sample x_i, $\mathcal{B}_k$ denotes the sample belonging to source-domain class k in the mini-batch and $m \in [0,1]$ is the update ratio of the memory.

Classifier Learning and Updating. As has been noted, we have discussed how to organize and extract knowledge from the previous learning process. And then we explain how to leverage such knowledge to benefit the training. At the beginning of each training epoch, the current classifier weights W^t are updated based on the memorized prototypes in the memory bank and the past weights W^{t-1}, can be written as:

$$W^t = \lambda W^{t-1} + (1-\lambda)M[k] \qquad (3)$$

where $\lambda \in [0,1]$ is the update ratio of the weights. And then the classifier weights updates are performed using stochastic gradient descent during the training iteration.

Naturally, the updated classifiers are used to make predictions and a robust generalized cross-entropy loss is obtained, which integrates the robustness of the memory module with
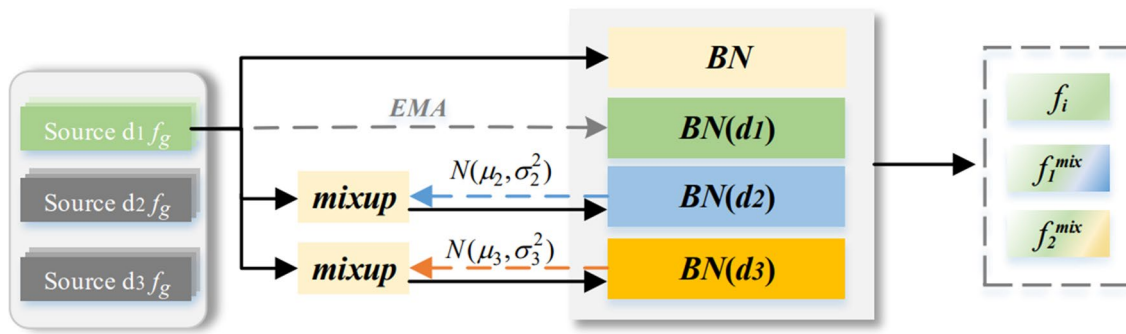
Fig. 3 Illustration of MMBN module, the number of domain K is set as 3 for convenience. We only present the feature augmentation on one domain and the same for the other domains. MMBN consists of two branches, one is the CBN layer for all domains, and the other is DSBN for specific source domains. In the training stage, the μ and σ in the CBN and the DSBN are updated by exponential moving average (EMA). The statistics of CBN are used for the testing stage, and the domain-specific properties are used for the mixup technique to mix with current domain features

the training efficiency of the traditional cross-entropy loss, which can be computed as:

$$L_{id} = -\sum_{j=1}^{N_i} y_i \log\left(\frac{\exp(W_{y_i}^t{}^T f_i)}{\sum_{j=1}^{N_i} \exp(W_j^t{}^T f_i)}\right) \quad (4)$$

3.3 Multi-Mix Batch normalization (MMBN)

Learning a domain-invariant model becomes difficult when source domains become more diverse, because of the inclusion of a lot of domain-specific style information in each domain. BN is a widely used training technique in domain generalization re-ID works. However, if the domain gap is significant, only using a common BN layer to share the parameters across multiple domains may not be conducive to generalization and robustness. Thus, we introduce a multi-mix BN module, in which domain-specific BN layers are integrated and are expected to enhance the diversity and robustness of output by mixing domain information. In MMBN, a common BN (CBN) layer is employed to conduct normalization and store the multi-domain statistics for the test stage. Meanwhile, the domain-specific BN (DSBN) layers are used to obtain the individual domain statistics and share the affine parameters with the CBN layer. The operation of MMBN is illustrated in Fig. 3.

Specifically, the features before the MMBN are denoted as f_g, and the CBN can be expressed as:

$$f_i = CBN(f_g) = \gamma \frac{f_g - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (5)$$

where $f_g \in \mathbb{R}^{B \times D}$, γ and β are the affine parameters, the ε is a small constant to avoid divide by zero. And the mean μ and variance σ within a mini-batch can be computed as follows:

$$\mu = \frac{1}{B}\sum_{b=1}^{B} f_g[b,:], \ \sigma^2 = \frac{1}{B}\sum_{b=1}^{B}(f_g[b,:] - \mu)^2 \quad (6)$$

During training, CBN estimates the mean and variance of activations across multiple domains by exponential moving average operation and are used for the testing stage. The EMA operation can be written as:

$$\overline{\mu} = (1-\alpha)\overline{\mu} + \alpha\mu \quad (7)$$

$$\tilde{\sigma}^2 = (1-\alpha)\tilde{\sigma}^2 + \alpha\sigma^2 \quad (8)$$

where α is the exponential average factor, the higher the factor, the more relevant to the current batch statistics.

Inspired by [7, 8], this paper expects to mix up domain information by using domain-specific statistics. However, directly fusing batch features for each domain is prone to incur noises. Therefore, in order to search the global domain representation for each domain, the Eqs. 7 and 8 are employed to estimate the domain statistics more stably. For the domain d_i, the domain-specific statistics are denoted as $\overline{\mu}_i$ and $\tilde{\sigma}_i$. We assume that the domain-specific statistics follow the Gaussian process, and then we can obtain K Gaussian distributions as the domain agents. A reparameterization trick is employed to randomly sample B features from each Gaussian distribution N, and obtain domain-specific features $\{f_1^{ds}, ..., f_K^{ds}\}$, the process is defined as following:

$$f_i^{ds} \sim N(\overline{\mu}_i, \tilde{\sigma}_i^2) \quad (9)$$

To maintain consistency within domains and alleviate the impact of variability among domains,

We mix the other domain information with the current domain, and the mixing process can be formulated as:

$$g_f^{mix} = \frac{1}{2} \times \theta f_j^{ds} + \left(1 - \frac{1}{2} \times \theta\right) f_g \ \ 1 \leq j \leq K-1, j \neq i \quad (10)$$

$$f_j^{mix} = BN(g_j^{mix}) = \gamma_j \frac{g_j^{mix} - \mu_z^{mix}}{\sqrt{\sigma_z^{mix2} + \varepsilon}} + \beta_j \tag{11}$$

where g_j^{mix} and f_j^{mix} reflect the characteristics after distribution fusion, μ_j^{mix} and $\sigma_j^{mix\,2}$ are the mean and var of g_j^{mix}, and $\theta \sim$ Beta(1,1) is the mixing ratio.

Loss functions for the outputs of MMBN are designed to further guarantee the semantically meaningful and robustness of the features. Specifically, a global contrastive loss based on inner product similarity is employed to reflect the relative difference in direction, while a mixed triplet loss based on Euclidean distance is adopted to reflect the absolute difference in value.

Only considering the optimization of local samples may be detrimental to the generalizability of the model, thus we design a prototype-based contrastive loss at the global optimization level named global contrastive loss (GCL). The main idea of the GCL is to gradually optimize the feature representation by contrastive learning between the mixed features and class prototypes in the multi-domain feature space.

Specifically, for feature sets $f^{mix} = \{f_1^{mix},...,f_{K-1}^{mix}\}$, computing the feature similarity matrix with $N = \{N_1, N_2, ..., N_K\}$ prototypes, where the prototypes denote the current classifier weights. Then, one positive sample and K_{neg} negative samples are selected to calculate the contrastive loss. The GCL loss of domain d_i is defined as:

$$L_{gcl} = -\frac{1}{K-1} \sum_{j=1,j\neq i}^{K} \log \frac{\exp(<f_j^{mix} \cdot w^+>/\tau)}{\exp(<f_j^{mix} \cdot w^+>/\tau) + \sum_{z=1}^{K_{neg}} \exp(<f_j^{mix} \cdot w_z^->/\tau)} \tag{12}$$

where $< \bullet >$ denotes the inner product and τ is the temperature parameter.

Deep metric learning between samples in small batches is more appropriate to capture discriminative features. Due to the characteristic of decreasing the intra-class distance and increasing the inter-class distance, triplet loss is very suitable to train person re-ID network. Expecting for features to be more robust to inter-domain distribution differences, a hybrid triplet loss (HTRI) is adopted in the model. The hybrid triplet loss extends the optimization scope, which includes the original optimization objective f_g and the mixed features $f^{mix} = \{f_1^{mix},...,f_{K-1}^{mix}\}$.

$$L_{htri} = \frac{1}{(K-1) \times B} \sum_{k=1}^{(K-1)} \sum_{a \in f_k^{mix}} [d_{a,p} - d_{a,n} + \delta]_+ + \frac{1}{B} \sum_{a' \in f_g} [d_{a',p'} - d_{a',n'} + \delta]_+ \tag{13}$$

where d_p and d_n are feature distances of positive pair and negative pair. δ is the margin of hybrid triplet loss, and $[z]_+$ equals to $max(z,0)$. This will guide the network to pay more attention to the intra-domain feature variability after distributional fusion, meanwhile, encouraging the network to extract more discriminative features.

The model totally includes three losses as follows:

$$L_{final} = L_{id} + L_{gcl} + L_{htri} \tag{14}$$

4 Experiments

4.1 Implementation Details

We follow the general pipeline of the Multi-source DG re-ID methods to build the baseline and incorporate the proposed method on top of the baseline. These are described separately.

Baseline (Base). In the baseline settings, we treat each source domain equally, each source domain has its classification layer, and jointly trains the feature network. Specifically, images of a batch size from each domain are sent to the model sequentially, and identification loss and triplet loss are used for the optimization of the model, where the label smoothing scheme is employed for the cross-entropy loss.

This paper (Base + KAFC + MMBN). We validate the effectiveness of the proposed method on top of the baseline (Base). A memory bank is built with the centroids (each centroid is an averaged feature of each person) of IDs for all of the domains. The centroids are utilized to update the weights of the classification layers in KAFC and compute the loss function. For enhancing robustness of the mixed features of MMBN, a global contrastive loss and a hybrid triplet loss are added, to enlarge the within-identity similarity and encourage the network to learn domain-invariant representations.

Implementation details We implement the method with two common backbones, i.e., ResNet-50 [38] and IBN-Net50 [39].

For training, each mini-batch contains 32 images (8 identities with 4 images). Images are resized to 256×128 and random flipping and random cropping are used for data augmentation. For the memory, the momentum coefficient m is set to 0.2 and the temperature factor τ is set to 0.05. ID loss is label smoothing and the parameter is set to 0.1. The margin δ of triplet loss and hybrid triplet loss are 0.3 and 1 respectively. To optimize the model, we use an Adam optimizer with a weight decay of 0.0005. The learning rate is initialized to $3.5 \times 10-5$ and increases linearly to $3.5 \times 10-4$ in the first 10 epochs. Then, the learning rate is decayed by 0.1 at the 30th, 40th and 50th epochs. The total training stage takes 70 epochs.

Datasets and evaluation metrics We demonstrate the effectiveness of the proposed method on four large-scale bench-mark datasets: Market1501 [40], DukeMTMC-reID [41], CUHK03 [42] and MSMT17 [43]. Specifically,

Table 1 Statistics of four experiment datasets

Dataset	Train IDs	Train images	Test IDs	Query images	Gallery images	Cameras
Market-1501	751	12,936	750	3368	15,913	6
DukeMTMC-reID	702	16,522	702	2228	17,661	8
CUHK03-NP	767	7365	700	1400	5332	-
MSMT17	1041	32, 621	3060	11,659	82,161	15

three datasets are used as source domains for training and the remaining domain for testing. Table—shows the specific details of the four datasets. Note: for CUHK03 and MSMT17 which have multi protocols, we use MSMT17_V1 and the new protocol CUHK03-NP for both training and testing. Concisely, we denote these datasets as M, D, C-NP and MS in the following tables. Rank-n (for $n = 1$, 5, and 10) and mean average precision (mAP) are adopted to evaluate the performance of different re-ID models (Table 1).

4.2 Comparison with State-of-the-art Methods

In this section, we compared the proposed method with the current state-of-the-art methods, including QAConv [24], OSNet [44], CBN [45], SNR [4], M3L [3], MixNorm [46] and DSAF [47]. QAConv considers more on interpretable of matching process and constructs query-adaptive convolution kernels on the fly to achieve local matching, which is trained with Resnet-50. OSNet is a flexible fusion mechanism that uses features extracted from different scales, the variant OSNet-AIN is a further development of the original model that adds instance normalization layers. CBN propose a batch normalization within cameras, which normalizes the image features of each camera separately by calculating the mean and variance of the image under each camera, thus eliminating the domain differences

between different cameras. M3L improves its robustness and generalization ability through the memory-based module and meta-learning strategy. SNR and MixNorm are the data augmentation method that reduces overfitting by enhancing data diversity, but their performance still has space for improvement. Table 2. shows the experimental results of our method and previous works. With a plain backbone, ResNet-50 or IBN-Net50, the proposed method outperforms these methods on the aforementioned four benchmarks.

Results on the Market1501 and DukeMTMC-reID. We can regard Market1501 and DukeMTMC-reID as mid-scale datasets. From the Table 2, this paper achieves the best performance under different types of backbones. Specifically, when testing on Market1501, our method achieves 56.5%mAP and 80.9% Rank-1, which outperforms M3L by 4.0%and 2.6% with the same backbone IBN-Net50, respectively. On DukeMTMC-reID, we can see that there is no significant improvement in the performance of the previous methods, while our proposed shows the boosts of 3.1% and 3.6% in terms of R1 accuracy and mAP, respectively, by comparing with the suboptimal M3L. By compared with the best method MixNorm, our method stays competitive at mAP and have satisfactory results on rank1.

Table 2 Comparison with the state-of-the-art unsupervised person re-ID methods on four datasets. (BOLD: BEST)

Method	Reference	MS+C-NP+D→M		MS+C-NP+M→D		M+C-NP+D→MS		MS+M+D→C-NP	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
QAConv50 [24]	2020ECCV	39.5	68.6	43.4	64.9	10.0	29.9	19.2	22.9
OSNet [44]	2019ICCV	44.2	72.5	47.0	65.2	12.6	33.2	23.3	23.9
OSNet-AIN [44]	2019ICCV	45.8	73.3	47.2	65.6	16.2	40.2	27.1	27.4
CBN [45]	2020ECCV	47.3	74.7	50.1	70.0	15.4	37.0	25.7	25.2
SNR [4]	2020CVPR	48.5	75.2	48.3	66.7	13.8	35.1	29.0	29.1
M3L(ResNet-50) [3]	2021CVPR	51.1	76.5	48.2	67.1	13.1	32.0	30.9	31.9
M3L(IBN-Net50) [3]	2021CVPR	52.5	78.3	48.8	67.2	15.4	37.1	31.4	31.6
MixNorm [46]	2022TMM	51.4	78.9	49.9	**70.8**	**19.4**	**47.2**	29.0	29.6
DSAF [47]	2023PR	53.2	79.7	-	-	16.8	42.4	-	-
Ours((ResNet-50)	-	54.1	80.0	51.4	69.4	15.4	37.8	30.3	30.9
Ours((IBN-Net50)	-	**56.5**	**80.9**	**52.4**	70.3	17.0	41.3	**34.3**	**35.7**

(Bold: Best)

Table 3 Ablation study on the impact of knowledge accumulated feature classifier ($\mathcal{C}$), global contrastive loss ($\mathcal{G}$) and hybrid triplet loss ($\mathcal{H}$) for multi-Source DG re-ID

	$\mathcal{C}$	$\mathcal{G}$	$\mathcal{H}$	MS+C-NP+D→M		MS+C-NP+M→D	
				mAP	Rank-1	mAP	Rank-1
0	×	×	×	42.1	71.4	39.6	59.0
1	√	×	×	51.1	77.3	49.1	68.1
2	√	√	×	54.2	80.3	50.6	69.3
3	√	×	√	55.3	79.8	51.7	70.0
4	×	√	√	52.1	79.0	48.9	67.4
5	√	√	√	**56.5**	**80.9**	**52.4**	**70.3**
Index	$\mathcal{C}$	$\mathcal{G}$	$\mathcal{H}$	M+C-NP+D→MS		MS+M+D→C-NP	
				mAP	Rank-1	mAP	Rank-1
0	×	×	×	13.1	34.9	16.7	15.9
1	√	×	×	16.6	41.3	27.8	28.4
2	√	√	×	16.9	41.5	30.8	31.0
3	√	×	√	16.4	39.8	32.7	32.9
4	×	√	√	16.6	40.7	27.7	28.4
5	√	√	√	**17.0**	**41.3**	**34.3**	**35.7**

(Bold: Best)

Results on the MSMT17 and CUHK03-NP. MSMT17 and CUHK03-NP are the large-scale and small-scale datasets, respectively. When IBN-Net50 meets our proposed, the performance is achieved 17.0%/41.3% at mAP/rank-1 on MSMT17 and 34.3%/35.7% at mAP/rank-1 on CUHK03-NP. Furthermore, we also observe that the introduction of instance normalization layers indeed enhances the diversity and generalizable of models, the performance under the backbone that is related to IN layer is significantly better than the unrelated one, particularly on the inconsistent scale dataset like MSMT17 and CUHK03-NP. When testing on CUHK03-NP, because the domain is less small than other domains, how to extract more discriminative representations is more important than the reduction of multi-source domain gaps. The above results verify the strong domain generalization of our method on different scale datasets.

4.3 Ablation study

In this section, we have conducted comprehensive ablation experiments using IBN-Net50 as the backbone to investigate how components and hyperparameters impact the performance of our proposal. The results of the ablation study of different components on four datasets are shown in Table 3. The performance variations of hyperparameters on Market1501 are shown in Fig. 4.

Analysis of Proposed Components. In this part, we analyze the impact of knowledge accumulated feature classifier ($\mathcal{C}$), global contrastive loss ($\mathcal{G}$) and hybrid triplet loss ($\mathcal{H}$) on four datasets. Firstly, the experiment setting in Index-0 denotes our baseline, the performances on different targets are unsatisfying. Secondly, in index-1

and index-4, the knowledge accumulated feature classifiers bring great boosts to all datasets, especially on the small and mid-scale datasets such as Market-1501, the results are improved by 5.9% and 9.0% in Rank-1 and mAP. This shows that accumulating previous knowledge under the global guidance of label information is useful and play an important role for steady and continuous representation learning. Thirdly, under the settings of index-2 and index-3, the model can further improve the performance compared with only adding KAFC, which indicates the effectiveness of our proposed loss function. Moreover, the model integrated with all components outperforms the baseline on all target datasets. On
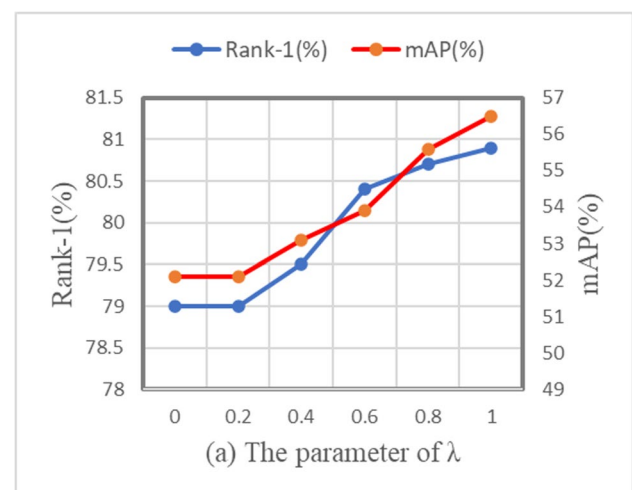

Fig. 4 Sensitivity of person re-ID accuracy to the weights update ratio λ. Rank-1 and mAP on Market-1501 are shown
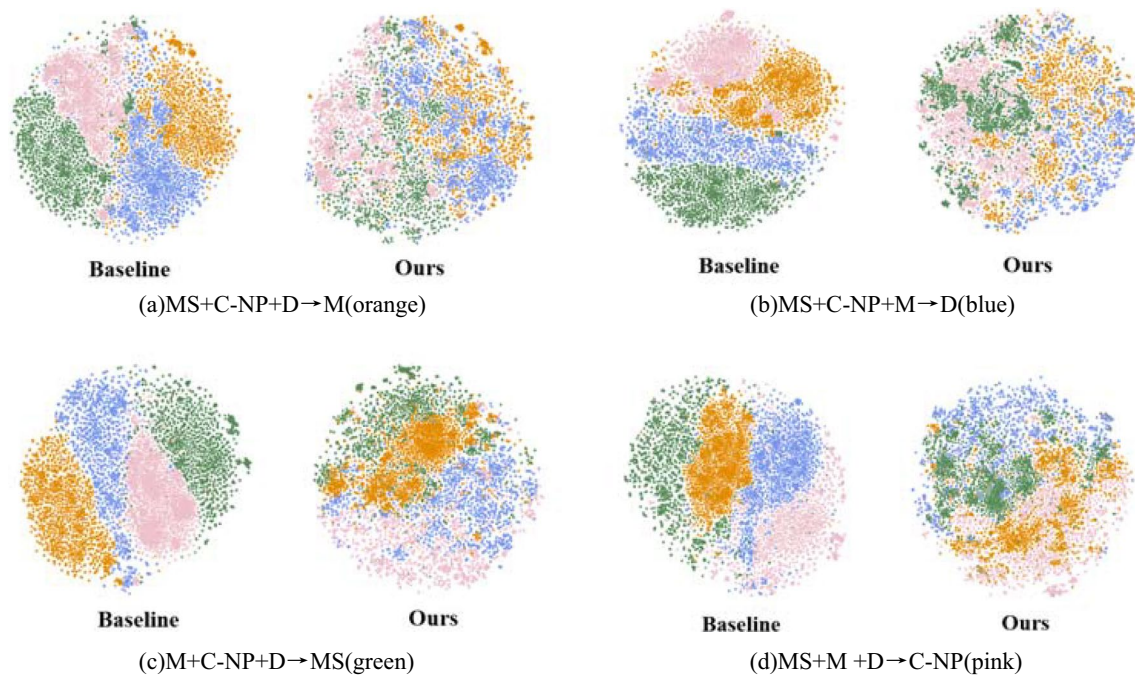
Fig. 5 t-SNE plots of the sample representations. The top line is the baseline, and the bottom is the proposed method. **a** MS+C-NP+D→M(orange). **b** MS+C-NP+M→D(blue). **c** M+C-NP+D→MS(green). **d** MS+M+D→C-NP(pink)

Market1501 and DukeMTMC-reID datasets, the results increased by 10.1%/11.6% R1 accuracy and 14.3%/12.6% mAP respectively. When testing on MSMT17 and CUHK03-NP, the results increased by 6.4%/18.8% R1 accuracy and 3.9%/17.6% mAP respectively. These results demonstrate that our proposed components are effective and mutually beneficial to improving the model generalization. But we can also observe that it is unrealistic and inevitable to expect that data augmentation will enable the transformed training data to cover the distribution of all test data. When only conducting the distribution mixing, the performance improvement is not significant, especially on the large-scale dataset MSMT17 with abundant distributions.

The impact of weight update ratio λ. λ is the weight update ratio of feature classifiers in the KAFC.

Large λ will enhance the guidance of the memory bank, but if the gradient direction is constrained by a large λ completely during iterations, the gradient descent may be too slow. In the proposed method, we only constrain at the beginning of each epoch and still use stochastic gradient descent in each training iter to guarantee the learning ability of the network. In this section, we compare the performance with different settings of λ on Market-1501, which vary from 0 to 1. As seen in Fig. 4(a), when λ is set to 1, the best precision is obtained.

4.4 Visualization

Figure 5 visualizes the feature representation of the baseline and the proposed method on the four datasets. We can observe that in the baseline, there are big domain gaps between multi-source domains and target domains. While in our method, features of different domains in the feature space tend to fuse better, which illustrates that the proposed method is effective to reduce the domain gaps and learn the domain-invariant representations.

5 Conclusions

In this paper, a multi-source domain generalization method is presented for person re-ID, which aims to improve discriminating and generalization capabilities on both seen and unseen domains by formulating two novel strategies including a knowledge accumulation strategy and a distribution enhancement strategy. In particular, we design a new knowledge accumulation feature classifier to adaptively update previously learned knowledge and the old parameters to alleviate the degrading impact of catastrophic forgetting and distributional shifts. For enhancing the robustness of the model under the multi-domain shifts, the MMBN module is introduced to capture and mix the domain-specific statistics. Moreover, in order to better optimize feature representations at the global and local level, we introduce the global

contrastive loss and hybrid triplet loss. Finally, our method is evaluated on four public benchmark datasets, extensive experiments show that our method's effectiveness and superiority over the other state-of-the-art methods.

In the future, we will jointly leverage the semi-supervised and weakly-supervised algorithms to reduce the reliance on labels for large datasets and improve the performance of person re-ID by exploring the knowledge of unlabeled data and the data with weaker labels.

Authors contribution Author 1 (First Author): Conceptualization, Methodology, Software, Investigation, Validation, Formal Analysis, Writing—Original Draft;

Author 2(Corresponding Author): Conceptualization, Funding Acquisition, Resources, Supervision, Writing—Review & Editing.

Author 3: Supervision, Writing—Review & Editing;

Author 4: Investigation, Formal Analysis, Writing—Review & Editing.

Data availability Dataset derived from public resources and made available with the article. The datasets analyzed during the current study are available in the [repository name Market1501] [https://doi.org/10.1109/ICCV.2015.133] reference number [40], [repository name DukeMTMC-reID] reference number [41], repository name CUHK03-NP] reference number [42] and [repository name MSMT17] [https://doi.org/10.1145/3394171.3413904] reference number [43].

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wang J, Lan C, Liu C, Ouyang Y, Qin T, Lu W et al (2022) Generalizing to unseen domains: a survey on domain generalization. Knowl Data Eng Trans. https://doi.org/10.1109/TKDE.2022.3178128
2. Song J, Yang Y, Song Y, Xiang T et al (2019) Generalizable person re-identification by domain-invariant mapping network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). CVPR, pp 2288–2295
3. Zhao Y, Zhong Z, Yang F, Luo Z, Lin Y, Li S et al (2021) Learning to generalize unseen domains via memory-based multi-source meta learning for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.00621
4. Jin X, Lan C, Zeng W, Chen Z, Zhang L (2020) Style normalization and restitution for generalizable person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR42600.2020.00321
5. Dai Y, Li X, Liu J, Tong Z, Duan L (2021) Generalizable person re-identification with relevance-aware mixture of experts. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.01588
6. Li P, Li D, Li W, Gong S, Fu Y, Hospedales TM (2021) A simple feature augmentation for domain generalization. In: International Conference on Computer Vision (ICCV), pp 8886–8895. https://doi.org/10.1109/ICCV48922.2021.00876
7. Zhou K, Yang Y, Qiao Y, Xiang T (2021) Domain generalization with mixstyle. In: International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.2205.07211
8. Luo C, Song C, Zhang Z (2020) Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 224–241. https://doi.org/10.1007/978-3-030-58555-6_14
9. Wang W, Liao S, Zhao F, Kang C, Shao L (2021) Domainmix: Learning generalizable person re-identification without human annotations. In: British Machine Vision Conference (BMCV). https://doi.org/10.48550/arXiv.2011.11953
10. Dou Q, Castro DC, Kamnitsas K, Glocker B (2019) Domain generalization via model-agnostic learning of semantic features. In: Advances in Neural Information Processing Systems (NIPS). https://doi.org/10.48550/arXiv.1910.13580
11. J Jia, Q Ruan, TM Hospedales (2019) Frustratingly easy person re-identification: generalizing person Re-ID. In: Practice. In: British Machine Vision Conference (BMVC). https://doi.org/10.48550/arXiv.1905.03422
12. D Ulyanov, A Vedaldi, V Lempitsky (2016) Instance normalization: the missing ingredient for fast stylization. arXiv:1607.08022. https://doi.org/10.48550/arXiv.1607.08022
13. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal co-variate shift. In: International Conference for Machine Learning (ICML). https://doi.org/10.48550/arXiv.1502.03167
14. Fan X, Wang Q, Ke J, Yang F, Gong B, Zhou M (2021) Adversarially adaptive normalization for single domain generalization. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 8208–8217. https://doi.org/10.48550/arXiv.2106.01899
15. Seokeon C, Taekyung K, Minki J, Hyoungseob P, Changick K (2021) Meta batch-instance normalization for generalizable person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.00343
16. Yu S, Zhu F, Chen D et al (2021) Multiple domain experts collaborative learning: multi-source domain generalization for person re-identification. arXiv:2105.12355. https://doi.org/10.48550/arXiv.2105.12355
17. Wu G, Gong S (2021) Collaborative optimization and aggregation for decentralized domain generalization and adaptation. In: International Conference on Computer Vision (ICCV), pp 6484–6493. https://doi.org/10.1109/ICCV48922.2021.00642
18. Shao R, Lan X, Li J et al (2019) Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2019.01026
19. Zhang H, Zhang Y, Liu W et al (2021) Towards principled disentanglement for domain generalization. arXiv:2111.13839. https://doi.org/10.48550/arXiv.2111.13839
20. Ko B, Gu G, Kim HG (2021) Learning with memory-based virtual classes for deep metric learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.48550/arXiv.2103.16940
21. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. In: International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2017.167

22. Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P (2017) Domain randomization for transferring deep neural networks from simulation to the real world. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 23–30. https://doi.org/10.1109/IROS.2017.8202133

23. Zheng K, Liu J, Wu W et al (2021) Calibrated feature decomposition for generalizable person re-identification. arXiv:2111.13945. https://doi.org/10.48550/arXiv.2111.13945

24. Liao S, Shao L (2020) Interpretable and generalizable person rei-dentification with query adaptive convolution and temporal lifting. In: European Conference on Computer Vision (ECCV), pp 456–474. https://doi.org/10.1007/978-3-030-58621-8_27

25. Ang E, Shan L, Kot AC (2021) DEX: domain embedding expansion for generalized person re-identification. arXiv:2110.11391. https://doi.org/10.48550/arXiv.2110.11391

26. Yassine H et al (2023) Video surveillance using deep transfer learning and deep domain adaptation: Towards better generalization. Eng Appl Artif Intell 119. https://doi.org/10.1016/j.engappai.2022.105698

27. K Hamza, Y Himeur, AI Awad (2023) Deep transfer learning for intrusion detection in industrial control networks: a comprehensive review. J Netw Comput Appl 220:103760

28. Pu N, Chen W, Liu Y et al (2021) Lifelong person re-identification via adaptive knowledge accumulation. arXiv:2103.12462. https://doi.org/10.48550/arXiv.2103.12462

29. Hou S, Pan X, Chen CL, Wang Z, Lin D (2019) Learning a unified classifier incrementally via rebalancing. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 831–839. https://doi.org/10.1109/CVPR.2019.00092

30. Zhang C, Song N, Lin G et al (2021) Few-shot incremental learning with continually evolved classifiers. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.01227

31. Simon C, Faraki M, Tsai YH et al (2022) On generalizing beyond domains in cross-domain continual learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.48550/arXiv.2203.03970

32. Rebuffi SA, Kolesnikov A, Sperl G, Lampert CH (2017) iCaRL: incremental classifier and representation learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.587

33. Wu G, Gong S (2021) Generalising without forgetting for lifelong person re-identification. Proc AAAI Conf Artif Intell 35(4):2889–2897

34. Ge W, Du J, Wu A et al (2022) Lifelong person re-identification by pseudo task knowledge preservation. Proc AAAI Conf Artif Intell 36(1):688–696

35. Yang Z, Wu D, Li B et al (2022) Joint plasticity learning for camera incremental person re-identification. arXiv:2210.08710

36. Deng J, Dong W, Socher R, Li LJ, Li K, Li F (2009) Imagenet: A large-scale hierarchical image database. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 248–255. https://doi.org/10.1109/CVPR.2012.6248055

37. He K, Zhang X, Ren S et al (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2015.123

38. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770–778. https://doi.org/10.1109/CVPR.2016.90

39. Pan X, Luo P, Shi J, Tang X (2018) Two at once: Enhancing learning and generalization capacities via ibn-net. In: European Conference on Computer Vision (ECCV), pp 484–500

40. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: The IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2015.133

41. Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C (2016) Performance measures and a data set for multi-target, multi-camera tracking. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham

42. Li W, Zhao R, Xiao T, Wang X (2014) Deep-reid: Deep filter pairing neural network for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 152–159. https://doi.org/10.1109/CVPR.2014.27

43. Liu X, Zhang S (2020) Domain adaptive person re-identification via coupling optimization. In: Proceedings of the 28th ACM International Conference on Multimedia, pp 547–555. https://doi.org/10.1145/3394171.3413904

44. K Zhou, Y Yang, A Cavallaro et al (2021) Learning generalisable omni-scale representations for person re-identification IEEE Trans Softw Eng. https://doi.org/10.48550/arXiv.1910.06827

45. Zhuang Z, Wei L, Xie L et al (2020) Rethinking the distribution gap of person re-identification with camera-based batch normalization. European Conference on Computer Vision (ECCV). https://doi.org/10.1109/TCSVT.2021.3058111

46. Qi L, Wang L, Shi Y, Geng X (2022) A novel mix-normalization method for generalizable multi-source person re-identification. IEEE Trans Multimed (IEEE TMM). https://doi.org/10.48550/arXiv.2201.09846

47. L Qi, J Liu, L Wang et al (2023) Unsupervised generalizable multi-source person re-identification: A domain-specific adaptive framework, pattern recognition (PR)
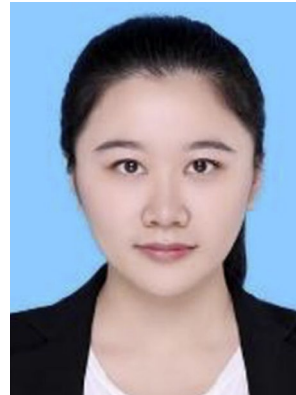
Wanru Peng was born in qin-huangdao, China in 1995. She is currently studying for the Ph.D. degree in information and communication engineering at Beijing Jiaotong University, Beijing, China. She received the bachelor's degree in electronic information engineering from Capital Normal University, China, in 2019. Her current research interests include computer vision and pattern recognition.

Yanfeng Li was born in Hebei, China in 1988. She received the B.S. degrees in communication engineering from Beijing Jiaotong University, Beijing, China, in 2010 and the Ph.D. degree in circuit and system from Beijing Jiaotong University, Beijing, China, in 2015. Since 2015, she has been an Associate Professor with School of Electronic and Information Engineering, Beijing Jiaotong University.

Houjin Chen was born in Anhui, China in 1965. He received the B.S. degree in signal engineering from Lanzhou Jiaotong University, Lanzhou, China, in 1986 and M.S. and Ph.D. degrees in communication and control from Beijing Jiaotong University, Beijing, China, in 1989 and 2003. Since 2000, he has been a Professor with School of Electronic and Information Engineering, Beijing Jiaotong University.

Jia Sun was born in Liaoning, China in 1995. She is currently studying for the Ph.D. degree in communication engineering at the school of electronic and information engineering, Beijing Jiaotong university. She received the B.S. degrees in electronic science and technology from Beijing Jiaotong University, Beijing, China, in 2017. Her current research interests include image processing, pattern recognition and deep learning.