Domain Generalized Person Re-Identification via Cross-Domain Episodic Learning

Ci-Siang Lin
National Taiwan University
Taipei, Taiwan
ASUS Intelligent Cloud Services
Taipei, Taiwan
Email: d08942011@ntu.edu.tw

Yuan-Chia Cheng
National Taiwan University
Taipei, Taiwan
Email: r08942154@ntu.edu.tw

Yu-Chiang Frank Wang
National Taiwan University
Taipei, Taiwan
ASUS Intelligent Cloud Services
Taipei, Taiwan
Email: ycwang@ntu.edu.tw

Abstract—Aiming at recognizing images of the same person across distinct camera views, person re-identification (re-ID) has been among active research topics in computer vision. Most existing re-ID works require collection of a large amount of labeled image data from the scenes of interest. When the data to be recognized are different from the source-domain training ones, a number of domain adaptation approaches have been proposed. Nevertheless, one still needs to collect labeled or unlabelled target-domain data during training. In this paper, we tackle an even more challenging and practical setting, *domain generalized (DG) person re-ID*. That is, while a number of labeled source-domain datasets are available, we do *not* have access to any target-domain training data. In order to learn domain-invariant features without knowing the target domain of interest, we present an episodic learning scheme which advances meta learning strategies to exploit the observed source-domain labeled data. The learned features would exhibit sufficient domain-invariant properties while not overfitting the source-domain data or ID labels. Our experiments on four benchmark datasets confirm the superiority of our method over the state-of-the-arts.
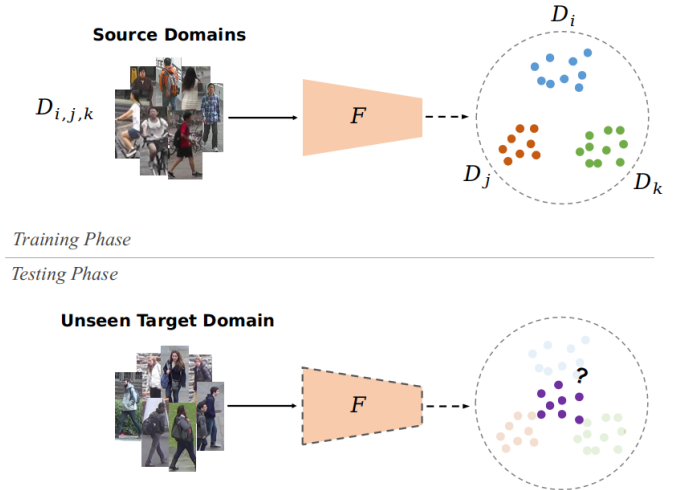
Fig. 1. Illustration of domain generalized person re-identification. During training, multiple source-domain data are observed to learn domain-invariant representations, with the goal of tackling re-ID tasks in unseen target domains.

I. INTRODUCTION

Person re-identification (re-ID) [1] has been among active research topics in computer vision due to its wide applications to person tracking [2], video surveillance systems [3] and smart cities. Given a query image containing a person of interest, re-ID aims at matching gallery images with the same identity across different camera views. A number of works [4]–[7] have been proposed to recognize the identical identity suffering from the variation of viewpoints, postures, occlusions or background clutters. However, most of these approaches require collection of a large amount of labeled image data from the scenes of interest, which is not practical for real-world applications due to limited resources and privacy issues.

Alternatively, one can utilize labeled data from one or multiple source domains, and jointly observe labeled or un-labeled target-domain data to train the re-ID models. Such domain adaptation (DA) approaches [8] have been proposed for cross-dataset re-ID. If not observing label information from target-domain data during training, the resulting unsupervised domain adaptation (UDA) setting [9]–[11] would be a more difficult task to handle. Nevertheless, for cross-dataset re-ID, one still needs to collect target-domain data for training

purposes. In this paper, we tackle an even more challenging and practical setting, *domain generalized (DG) person re-ID*. As illustrated in Fig. 1, while a number of labeled source-domain datasets are available, we do *not* have access to the target-domain data even during training. Different from cross-dataset re-ID, the goal of domain generalized person re-ID is to improve the generalization and robustness of the learned model, which is learned from multiple source domains only. While a number of works [12]–[15] have been proposed for solving DG classification tasks, the label sets across domains (including the unseen target domain) remain the same. This is very different from the setting for re-ID, in which we do *not* assume that persons of interest across domains remain the same. As a result, existing DG classification methods cannot easily tackle the re-ID tasks. For DG person re-ID, recent works like DualNorm [16] takes advantage of Batch Normalization (BN) [17] and Instance Normalization (IN) [18] to alleviate the domain difference, while DIMN [19] learns the mapping between person images and the ID classifiers but does not extended to unseen target-domain data well.

To address the challenging DG re-ID tasks without observing target-domain training data, we present an episodic learning scheme which advances meta learning strategies to exploit the observed source-domain labeled data. The learned features would exhibit sufficient domain-invariant properties while not overfitting the source-domain data or ID labels. This arms us to apply the learned model to any target domains of interest in no need of extra data collection and model updating. Compared to prior works, our experiments confirm that our proposed framework indeed improve the performance under the practically favorable setting.

We now highlight the contributions of our work below:

- We are among the first to derive domain invariant yet identity-discriminative features for re-ID without observing target-domain data during training.

- We advance meta learning strategies, allowing derivation of domain-invariant latent representation with re-ID guarantees.

- Experimental results on four benchmark datasets quantitatively verify that our approach performs preferably against state-of-the-art (cross-dataset) re-ID methods.

II. RELATED WORKS

A. Person Re-Identification

Person re-identification (re-ID) [1] has been widely studied in the literature. With the advancement of deep learning, supervised person re-ID [4], [5], [7], [20], [21] has achieved a great progress in the last decade. Existing methods typically focus on tackling the challenges of matching images with viewpoint [5] and pose variations [7], or those with background clutter [20] or occlusion presented [21]. For example, Sun et al. [5] comprehensively analyze the influence of viewpoint on re-ID by varying the rotation angle of the pedestrian relative to the camera. With the guidance of human pose maps, Ge et al. [7] develop a pose transferable GAN and derive pose invariant representations to handle pose variants. To mitigate the influence of background clutter, Tian et al. [20] apply human parsing models to extract informative features of foreground regions. With no need for part-level alignment, He et al. [21] construct spatial image pyramids that can re-identify persons accurately in the presence of heavy occlusion. While promising results have been observed, the above approaches typically requires a large amount of labeled data, thus limit themselves by scalability and practicality.

B. Cross-Dataset Person Re-Identification

Alternatively, one can utilize labeled data from one or multiple source domains, and jointly observed unlabeled target-domain data to train the re-ID models. Such approaches [9], [11], [22]–[25] aims to transfer and adapt identity-discriminative knowledge from labeled source domain to

unlabeled targert domain data. Most of the works fall into three categories: (1) image-level style transfer (2) feature-level distribution alignment (3) joint image-level and feature-level alignment. The first category [11], [22], [23] typically performs image-image translation while preserving identity information based on CycleGAN [26]. The second one [24], [25] employs constraints like Maximum Mean Discrepancy (MMD) [24] to derive a domain-invariant latent space. The last [9] achieves latent space and pixel space alignment jointly by combining techniques from both categories. While the aforementioned works are effective for cross-dataset person re-identification, one still needs to collect target-domain data for training purpose.

C. Person Re-Identification in Unseen Domains

Learning from data across multiple source domains for handling the associated task in unseen domains is referred to as domain generalization. To tackle domain generalization, a plethora of methods [27]–[30] have been proposed. Xu et al. [31] introduces a low-rank structure to derive the corresponding feature representation by exemplar-SVMs. [28], [30] intends to learn a domain-invariant feature space by exploiting multiple source domains. Deemed as a mix-up of above, Li et al. [29] decomposes the model into domain-specific and domain-invariant components and utilizes both to make predictions. Moreover, applying the meta-learning strategy has become a prevalent branch to address domain generalization. MLDG [13] carries out modification on MAML [32] to adapt the learning scheme to domain generalization settings. Li et al. [33] introduces an episodic training framework in which domain-specific feature extractors and classifiers are crossly trained in order to simulate interacting with a domain-specific tuned partner so that the feature extractors are able to learn robust feature representation. Despite the effectiveness of the above works, they cannot be easily extended to tackle re-id tasks since the setting of domain generalization typically assumes the label spaces are shared across domains (including the target one).

Recent works have emerged to deal with re-ID under domain generalization settings. That is, one trains the model with no access to the data from target domain and are required to recognize the unseen person of interest. DualNorm leverages both Batch Normalization (BN) [17] and Instance Normalization (IN) [18] to reduce the domain shift. DIMN [19] proposes to map gallery person images into classifier weights with the help of memory bank. Both of them do not simulate the unseen domain scenario with meta-learning. In this work, we also consider domain generalization for re-ID and propose an episodic meta learning strategy to derive domain-invariant features.

III. PROPOSED METHOD

A. Notations and Problem Formulation

We first define the notations to be used in this paper. For domain generalized person re-ID, assume that we have the access to data from N_d labeled source domains (datasets), i.e.,
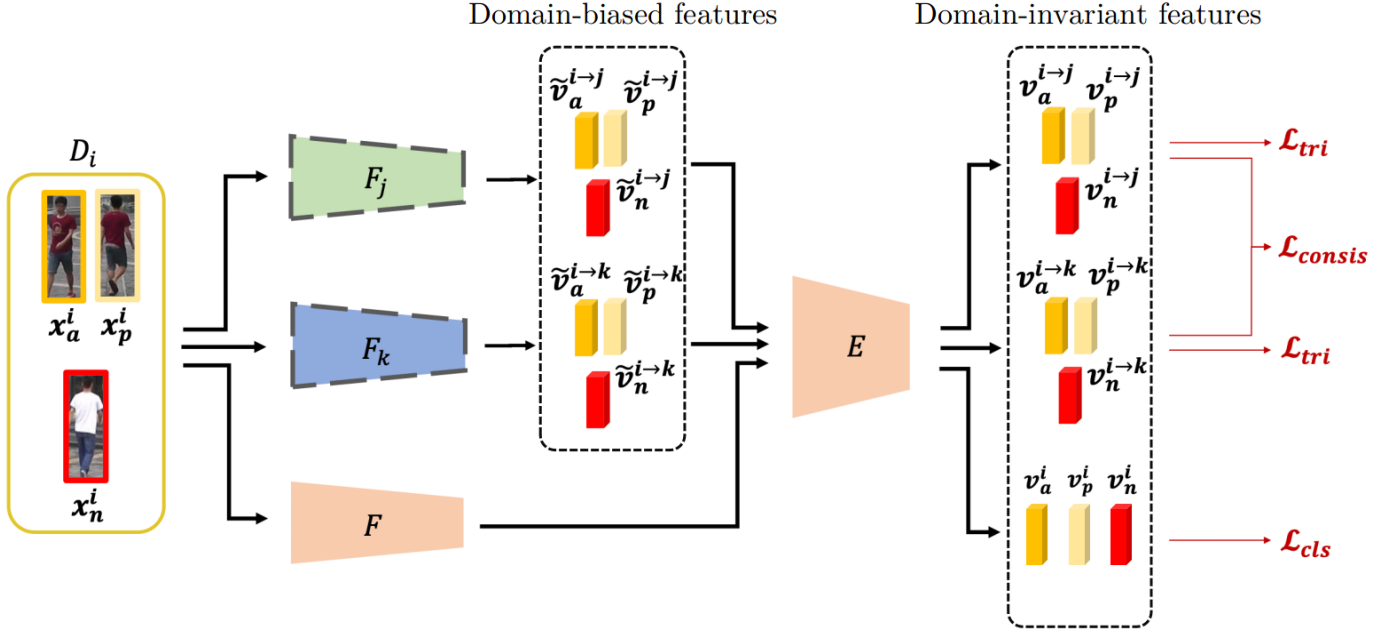
Fig. 2. Overview of our proposed framework. During training, triplet images (i.e., anchor x_a^i, and its positive x_p^i and negative neighbor x_n^i) from source domain D_i are forwarded through feature extractors F_j, F_k pretrained on domains D_j and D_k, also through a global feature extractor F. The features derived by pretrained extractors are viewed as domain-biased features $\tilde{v}$, and the global encoder E is expected to observe the resulting triplet loss $\mathcal{L}_{tri}$ and consistency loss $\mathcal{L}_{consis}$. The features derived by F are forwarded through E to guide classification losses $\mathcal{L}_{cls}$. During testing, only F and E are needed to extract as domain-invariant features v for performing re-ID.

source-domain data $D_S = \{D_i\}_{i=1}^{N_d}$, and the ith source domain D_i contains a set of N_i images $X_i = \{x_u^i\}_{u=1}^{N_i}$ with the associate label set $Y_i = \{y_u^i\}_{u=1}^{N_i}$, where $x_u^i \in R^{H \times W \times 3}$ and $y_u^i \in R$ denote the uth image and its corresponding identity label from the ith source domain D_i, respectively. Note that for any pair of source domains D_i and D_j, we consider their ID labels are *disjoint*. This unique property makes our domain generalized person re-ID challenging yet practical than prior domain adaptation or domain generalization ones (recall that prior DA or DG approaches consider a joint label space across domains).

To achieve DG re-ID, we present an end-to-end meta learning framework, as illustrated in Fig. 2. Our framework aims at learning domain-invariant features v by learning feature and re-ID encoders F and E, based on feature extractors pretrained on each source domain. During each episode, three domains D_i, D_j and D_k are randomly sampled from the source domain datasets. Feature extractors F_j and F_k pretrained on D_j and D_k are used to extract the domain-biased features $\tilde{v}^{i \to j}$ and $\tilde{v}^{i \to k}$ for images x^i from domain D_i, respectively. With the proposed meta learning scheme, our encoder E would derive domain-invariant yet identity-discriminative features for re-ID purposes. The details of our proposed learning framework will be discussed in the following sub-sections.

B. Meta Learning for Domain-Invariant Representation

We now detail how we advance meta learning strategies for deriving domain-invariant features, without observing the target-domain data during training. Our proposed framework

starts with pretrained domain-specific feature extractor F_i using labeled data in each source domain D_i (e.g., using TripletLoss [4] as most re-ID works do). Thus, a total of N_d pretrained domain-specific feature extractors are obtained and will later be utilized in our meta learning framework.

As shown in Fig 2, our goal is to learn a domain-invariant feature extractor F, followed by a domain generalized encoder E, for deriving DG re-ID. To accomplish this, we present an *episodic learning* scheme that utilizes the pretrained F_i for learning the above network modules. To be more precise, in each episode during training, we randomly select data from three source domains D_i, D_j and D_k. For an input anchor image x_a^i from domain D_i, we particularly apply the pretrained domain-specific feature extractors F_j and F_k to output the associated *domain-biased* features $\tilde{v}_a^{i \to j}$ and $\tilde{v}_a^{i \to k}$. In other words, we have $\tilde{v}_a^{i \to j} = F_j(x_a^i)$ and $\tilde{v}_a^{i \to k} = F_k(x_a^i)$. Since $\tilde{v}_a^{i \to j}$ and $\tilde{v}_a^{i \to k}$ are both derived from the same image x_a^i, we require the domain generalized encoder E to output the final *domain-invariant* features $v_a^{i \to j}$ and $v_a^{i \to k}$. To enforce E to preserve the domain-invariant yet re-ID preserved information from $v_a^{i \to j}$ and $v_a^{i \to k}$, we propose to calculate the *cross-domain consistency loss* $\mathcal{L}_{consis}$ on the above feature pair:

$$\mathcal{L}_{\text{consis}} = E_{x_a^i \sim X_i} \| v_a^{i \to j} - v_a^{i \to k} \|_2. \tag{1}$$

Note that the pretrained feature extractors F_j and F_k are fixed, which would *not* be updated by back-propagated gradients calculated from $\mathcal{L}_{consis}$. This is the *key* technique for each domain-specific feature extractors to preserve its own domain-specific properties, while allowing the domain generalized

Algorithm 1: Training of the Proposed Episodic Learning Scheme

---

1 Input: N_d source domains $D_S = \{D_i\}_{i=1}^{N_d}$

2 Stage 1: Pretraining

3 for *each domain D_i in D_S* **do**

4   Pretrain a domain-specific feature extractor F_i using data (X_i, Y_i) and TripletLoss

5 end

6 Stage 2: Episodic Training

7 θ_F, $\theta_E \leftarrow$ initialize

8 for *num. of iterations* **do**

9 D_i, D_j, $D_k \leftarrow$ randomly sampled from D_S

10 x_a^i, x_p^i, $x_n^i \leftarrow$ randomly sampled from D_i

11 F_j, $F_k \leftarrow$ pretrained in domain D_j and D_k

12 $\tilde{v}_a^{i \to j}$, $\tilde{v}_a^{i \to k} \leftarrow$ obtained from $F_j(x_a^i)$, $F_k(x_a^i)$

13 $v_a^{i \to j}$, $v_a^{i \to k} \leftarrow$ obtained from $E(\tilde{v}_a^{i \to j})$, $E(\tilde{v}_a^{i \to k})$

14 $\mathcal{L}_{consis} \leftarrow$ calculated by (1)

15 $\tilde{v}_p^{i \to j}$, $\tilde{v}_n^{i \to j} \leftarrow$ obtained from $F_j(x_p^i)$, $F_j(x_n^i)$

16 $v_p^{i \to j}$, $v_n^{i \to j} \leftarrow$ obtained from $E(\tilde{v}_p^{i \to j})$, $E(\tilde{v}_n^{i \to j})$

17 $\mathcal{L}_{tri} \leftarrow$ calculated by (4)

18 $\mathcal{L}_{cls} \leftarrow$ calculated by (5)

19 $\mathcal{L}_{total} = L_{cls} + \lambda_{tri} \cdot L_{tri} + \lambda_{consis} \cdot L_{consis}$

20 $\theta_{E,F} \leftarrow \theta_{E,F} - \eta \cdot \bigtriangledown L_{total}$

21 end

22 Output: Global feature extractor F and domain generalized encoder E

Dataset	Cameras	IDs	Images
Market1501 [35]	6	1,501	29,419
DukeMTMC-reID [36]	8	1,812	36,411
CUHK02 [37]	5	1,816	7,264
CUHK03 [38]	6	1,467	14,097
Total	25	6,596	87,191

Dataset	Pr. IDs	Ga. IDs	Pr. Images	Ga. Images
GRID [39]	125	900	125	1,025
i-LIDS [40]	60	60	60	60
PRID [41]	100	649	100	649
VIPeR [42]	316	316	316	316

where $v_a^{i \to j}$, $v_p^{i \to j}$ and $v_n^{i \to j}$ denote the domain-invariant latent features of x_a^i, x_p^i and x_n^i derived from the encoder E. With the above definitions, we have the *domain generalized triplet loss*, $\mathcal{L}_{tri}$, which is calculated as:

$$\mathcal{L}_{tri} = E_{(x_a^i, y_a^i) \sim (X_i, Y_i)} \max(0, m + d_p - d_n), \qquad (4)$$

where $m > 0$ is the margin enforcing the separation between positive and negative image pairs.

Take computational feasibility into consideration, we follow [4] and select only the hardest positive and negative samples to calculate the triplet loss. By minimizing $\mathcal{L}_{tri}$, the encoder E manages to capture the identity information from domain specific features $\tilde{v}_a^{i \to j}$ to extract domain invariant but identity-discriminative features $v_a^{i \to j}$. It is worth repeating that, the pretrained domain specific feature extractors F_j and F_k will *not* be updated by this loss.

To make the training process more stable and have the encoder E describe global identity features, we additionally train a global (or domain-invariant) feature extractor F using training data from *all* source domains (datasets). To further enforce the re-ID capability, an additional classifier C is integrated into the encoder E for identity prediction, with the cross-entropy loss calculated as:

$$\mathcal{L}_{cls} = E_{(x,y) \sim D_S} - \log p(y|x), \qquad (5)$$

where $p(y|x)$ denotes the prediction probability indicating the image x belongs to its corresponding identity label y.

With the above meta learning framework together with the introduced losses, the total loss $\mathcal{L}_{total}$ of our model is

$$\mathcal{L}_{total} = L_{cls} + \lambda_{tri} \cdot L_{tri} + \lambda_{consis} \cdot L_{consis}, \qquad (6)$$

where λ_{tri} and λ_{consis} are the hyperparameters. To perform person re-ID on the unseen domain in the testing phase, the query and gallery images are forwarded through feature extractor F and the encoder E to derive domain-invariant

encoder E to extract domain-invariant features during the episodic learning process. It is worth repeating that, during learning of domain-invariant features, no target-domain data are observed.

We also note that, we choose *not* to apply adversarial learning techniques (e.g., DANN [34]) for deriving domain-invariant features. This is because that, the person identities are *disjoint* across source domains. If one applies adversarial learning or similar domain adaptation techniques for eliminating the domain differences, it is likely that the person ID or pose information is also confused by such learning strategy, which is not desirable for re-ID tasks.

C. Domain Generalized Person Re-ID

To utilize label information observed from source-domain data for learning domain-invariant features for re-ID, we adopt the triplet loss on the derived domain invariant space (i.e., feature space output by domain generalized encoder E). That is, for each domain invariant feature $v_a^{i \to j}$ derived from the input anchor image x_a^i, a triplet tuple is composed of $v_p^{i \to j}$ with the same identity label as x_a^i and $v_n^{i \to k}$ with different identity label as x_a^i. Then, the distances d_p and d_n for such positive and negative pairs are defined as:

$$d_p = \|v_a^{i \to j} - v_p^{i \to j}\|_2, \qquad (2)$$

$$d_n = \|v_a^{i \to j} - v_n^{i \to j}\|_2, \qquad (3)$$

TABLE III
PERFORMANCE COMPARISONS OF DOMAIN GENERALIZATION BASED
METHODS IN TERMS OF AVERAGED RANK-1 ACCURACY. NOTE THAT
TARGET-DOMAIN DATA ARE ONLY SEEN DURING TESTING.

Target	GRID	i-LIDS	PRID	VIPeR	Avg.
DIMN [19]	23.4	44.8	13.1	29.9	27.8
DualNorm [16]	29.2	58.3	54.3	**38.6**	45.1
Ours	**33.0**	**62.3**	**57.6**	38.5	**47.8**

TABLE IV
PERFORMANCE COMPARISONS OF DOMAIN ADAPTATION BASED METHODS
IN TERMS OF AVERAGED RANK-1 ACCURACY. NOTE THAT BASELINE AND
DANN OBSERVE ONLY SOURCE-DOMAIN DATA DURING TRAINING
WITHOUT THE ACCESS TO ANY INFORMATION FROM TARGET DOMAIN.

Target	GRID	i-LIDS	PRID	VIPeR	Avg.
Baseline	18.8	52.5	14.8	32.0	29.5
DANN [34]	29.0	57.2	56.8	37.8	45.2
Ours	**33.0**	**62.3**	**57.6**	**38.5**	**47.8**

TABLE V
ABLATION STUDIES ANALYZING THE IMPORTANCE OF EACH INTRODUCED
LOSS FUNCTION.

Target	GRID	i-LIDS	PRID	VIPeR	Avg.
Ours w/o $\mathcal{L}_{tri}$	31.3	59.0	55.8	37.3	45.8
Ours w/o $\mathcal{L}_{consis}$	30.6	60.3	55.7	**40.1**	46.7
Ours	**33.0**	**62.3**	**57.6**	38.5	**47.8**

re-ID features, which are applied for matching query/gallery images via nearest neighbor search in Euclidean distances.

IV. EXPERIMENTS

A. Datasets and Experimental Settings

To evaluate our proposed method, following [19] we use existing large-scale re-ID datasts as the source domains and test the performance on several target datasets which are not observed during training. To be specific, the source domains include Market1501 [35], DukeMTMC-reID [36], CUHK02 [37] and CUHK03 [38], with a total of 6596 identities and 87191 images. The target datasets include GRID [39], i-LIDS [40], PRID [41] and VIPeR [42]. We follow the single-shot setting with the number of probe/gallery images set as: GRID: 125/1025; i-LIDS: 60/60; PRID 100/649; VIPeR 316/316 respectively. Detailed data statistics are summarized in Table I and II. The average rank-1 accuracy over 10 random splits is reported based on the standard evaluation protocol and the cumulative matching curve (CMC). To be clear, we reproduce all methods and performances due to the lack of legal access to certain datasets used in original papers.

B. Implementation Details

We implement our method using PyTorch. We use the backbone model proposed in [16] as our global feature extractor F and use ResNet-50 [43] pretrained on ImageNet for domain specific modules F_i. The global encoder E consists of two fully-connected (FC) layers with BatchNorm [17] layers while the global classifier C is a single FC layer. We resize all the input images to $256 \times 128 \times 3$ (denoting height, width and channel, respectively). Random clipping and cropping are adapted for data augmentation. The margin m is set as 0.3 and we fix λ_{tri} and λ_{consis} as 0.2 and 0.01, respectively. We train our model for 150 epochs with the SGD optimizer. The initial learning rate is set as 0.01 and is decreased to 0.001 at 100 epochs. Label smoothing is used to prevent overfitting.

C. Comparisons Against State-of-the-Art

In Table III, we compare our proposed method with two state-of-the-arts [16], [19] which attempted the DG setting for re-ID. From this table, we see that our method performed favorably well and observed performance margins over the state-of-the-art methods. We achieved the averaged **Rank-1 accuracy** of **47.8%** on the four target datasets. Compared to DIMN [19], our method learned domain invariant representations, while DMIN learned a mapping between a person

image and its identity classifier weight without deriving a domain invariant latent space, and thus failed to generalize to target domain. Compared to DualNorm [16] which simply adopted BN and IN layers to alleviate the domain shift, our method meta-learned to capture identity information under the proposed episodic scheme, and thus our averaged Rank-1 accuracy was higher by **2.7%**. From the experiment, the effectiveness of our model for domain generalized person re-ID was quantitatively verified.

As mentioned in Sec. III-B, we did *not* apply the adversarial training strategy to derive domain invariant representations since person identities are disjoint for any two different re-ID datasets, and thus the adversarial training strategy might produce pose-invariant or camera-invariant features instead of domain-invariant ones. To verify this, we compare our method to a simple baseline and DANN [34], which derived domain invariant features via adversarial loss and is a popular UDA method. For the baseline, we simply trained a single ResNet-50 to predict the person identities for all domains. Although DANN was originally proposed for UDA, we repurposed it for domain generalization by adding an auxiliary classifier to the baseline model with a gradient-reversal layer for domain confusion. As shown in Table IV, our averaged Rank-1 accuracy was higher than DANN by **2.6%**. This demonstrated our cross-domain consistency loss is practically more preferable than the adversarial loss.

D. Ablation Studies

To further analyze the importance of each introduced loss function, we conduct ablation studies shown in Table V. Without the domain generalized triplet loss $\mathcal{L}_{tri}$, our model would not properly capture the identity information for unseen domain, resulting in 2% performance drop. Also, when $\mathcal{L}_{consis}$ is turned off, our model would fail to generalize to unseen domain since there is no explicit constraint for learning domain invariant representations. From the above experiment,

we confirmed that each introduced loss function is vital and beneficial to domain generalized person re-ID.

V. CONCLUSIONS

In this paper, we addressed the challenging domain generalized person re-ID problem, in which target-domain data is not available during training. With the proposed meta learning framework, we utilized episodic training strategy with pretrained domain specific feature extractors, and learn domain-invariant yet identity-discriminative features with re-ID performance guarantees. Our experiments on multiple benchmark datasets confirmed that our approach performed favorably against state-of-the-art domain adaptation and domain generalization methods on this challenge task.

ACKNOWLEDGMENT

REFERENCES

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016.
[2] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
[3] F. M. Khan and F. Brémond, "Person re-identification for real-world surveillance systems," *arXiv preprint arXiv:1607.05975*, 2016.
[4] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.
[5] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
[6] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Multimedia*, 2018.
[7] Y. Ge, Z. Li, H. Zhao, G. Yin, S. Yi, X. Wang *et al.*, "Fd-gan: Pose-guided feature distilling gan for robust person re-identification," in *Advances in neural information processing systems*, 2018.
[8] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. S. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
[9] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *The European Conference on Computer Vision*, 2018.
[10] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
[11] Y. Chen, X. Zhu, and S. Gong, "Instance-guided context rendering for cross-domain person re-identification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
[12] M. Ghifary, W. Bastiaan Kleijn, M. Zhang, and D. Balduzzi, "Domain generalization for object recognition with multi-task autoencoders," in *International Conference on Computer Vision*, 2015.
[13] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *AAAI*, 2018.
[14] Y. Balaji, S. Sankaranarayanan, and R. Chellappa, "Metareg: Towards domain generalization using meta-regularization," in *NeurIPS*, 2018.
[15] D. Li, J. Zhang, Y. Yang, C. Liu, Y.-Z. Song, and T. M. Hospedales, "Episodic training for domain generalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
[16] J. Jia, Q. Ruan, and T. M. Hospedales, "Frustratingly easy person re-identification: Generalizing person re-id in practice," *arXiv preprint arXiv:1905.03422*, 2019.
[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015.
[18] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv*, 2016.
[19] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
[20] M. Tian, S. Yi, H. Li, S. Li, X. Zhang, J. Shi, J. Yan, and X. Wang, "Eliminating background-bias for robust person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[21] L. He, Y. Wang, W. Liu, H. Zhao, Z. Sun, and J. Feng, "Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification," in *ICCV*, 2019.
[22] S. Bak, P. Carr, and J.-F. Lalonde, "Domain adaptation through synthesis for unsupervised person re-identification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
[23] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.
[24] S. Lin, H. Li, C.-T. Li, and A. C. Kot, "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification," *arXiv preprint arXiv:1807.01440*, 2018.
[25] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[26] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017.
[27] G. Blanchard, G. Lee, and C. Scott, "Generalizing from several related classification tasks to a new unlabeled sample," in *NeurIPS*, 2011.
[28] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *ICML*, 2013.
[29] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
[30] H. Li, S. Jialin Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[31] Z. Xu, W. Li, L. Niu, and D. Xu, "Exploiting low-rank structure from latent domains for domain generalization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.
[32] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017.
[33] D. Li, J. Zhang, Y. Yang, C. Liu, Y.-Z. Song, and T. M. Hospedales, "Episodic training for domain generalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
[34] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *JMLR*, 2016.
[35] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
[36] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
[37] W. Li and X. Wang, "Locally aligned feature transforms across views," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
[38] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
[39] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
[40] W.-S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *The British Machine Vision Conference*, 2009.
[41] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *SCIA*.
[42] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitio*.