Person De-reidentification: A Variation-guided Identity Shift Modeling

Yi-Xing Peng[1,2,4] Yu-Ming Tang[1,4] Kun-Yu Lin[1,4] Qize Yang[3]
Jingke Meng[1,4*] Xihan Wei[3] Wei-Shi Zheng[1,2,4*]
[1]School of Computer Science and Engineering, Sun Yat-sen University, China;
[2]Peng Cheng Laboratory, Shenzhen, China; [3]Tongyi Lab, Alibaba Group;
[4]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China.
{pengyx23,tangym9,linky5}@mail2.sysu.edu.cn wszheng@ieee.org

Abstract

Person re-identification (ReID) is to associate images of individuals from different camera views against cross-view variations. Like other surveillance technologies, Re-ID faces serious privacy challenges, particularly the potential for unauthorized tracking. Although various tasks (e.g., face recognition) have developed machine unlearning techniques to address privacy concerns, such methods have not yet been explored within the Re-ID field. In this work, we pioneer the exploration of the person de-reidentification (De-ReID) problem and present its inherent challenges. In the context of ReID, De-ReID is to unlearn the knowledge about accurately matching specific persons so that these "unlearned persons" cannot be re-identified across cameras for privacy guarantee. The primary challenge is to achieve the unlearning without degrading the identity-discriminative feature embeddings to ensure the model's utility. To address this, we formulate a De-ReID framework that utilizes a labeled dataset of unlearned persons for unlearning and an unlabeled dataset of accessible persons for knowledge preservation. Instead of unlearning based on (pseudo) identity labels, we introduce a variation-guided identity shift mechanism that unlearns the specific persons by fitting the variations in their images while preserving ReID ability on other persons by overcoming the variations in images of accessible persons. As a result, the model shifts the unlearned persons to a feature space that is vulnerable to cross-view variations. Extensive experiments on benchmarks demonstrate the superiority of our method.

1. Introduction

Person re-identification (ReID) aims to match the images of the same person across different camera views based on feature similarity. While deep learning [9, 35, 43] has significantly enhanced the performance of the ReID models [10, 19, 45], there are growing public concerns about potential misuse, leading to ethical and societal implications. To avoid misuse and mitigate the potential privacy concerns, the research in surveillance has begun exploring machine unlearning techniques that limit the operational scope of these models. For instance, in face recognition, models are adjusted to "unlearn" individuals who raise privacy concerns, as mandated by regulations like the EU's General Data Protection Regulation (GDPR) [3, 5, 24, 53].

*Corresponding author

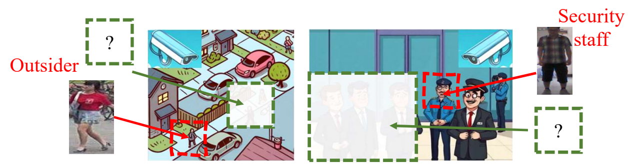

Figure 1. Person de-reidentification (De-ReID) prevents intelligent surveillance from tracking unlearned persons. For example, when using surveillance models for security (track and record the person in the red box), the residents in the community and the important customers in business (in the green box) should not be tracked for privacy and commercial secrets.

However, similar constraints on the ReID models have not yet to be explored, despite its significance for real-world applications. For instance, the residents in the neighborhood prefer to constrain the surveillance in the community to only focus on external people and not track themselves; in working buildings, security cameras should only be set for monitoring and preventing theft or accidents, not for tracking or identifying specific employees during their routine work; and homeowners expect smart home devices to track only unknown individuals for security purposes while avoiding any continuous surveillance of themselves; To meet these requirements, there is an urgent need to develop methods that limit the scope of ReID models, enabling them to "ignore" specific individuals during deployment.

In this work, we pioneer the exploration of person de-reidentification (De-ReID) problem: limiting the ReID scope by unlearning specific persons (referred to as "unlearned persons") in the ReID model while maintaining the ReID performance on other persons (referred to as "accessible persons"). In the context of image matching, a person is unlearned by a model means the model can hardly associate the images of the person under different camera views, as shown in Fig. 1. Forming a De-ReID model is challenging because only the knowledge strictly related to matching the unlearned persons should be eliminated, without destroying the identity-discriminative feature embeddings. A trivial solution is to distinguish the images of unlearned persons under different camera views for unlearning while associating the images of accessible persons based on their identity labels to preserve ReID knowledge. However, due to the large network capacity, the model may learn spurious cues or overfit the images of unlearned persons in the training set, leading to poor generalization during evaluation. Additionally, collecting identity labels for accessible persons to preserve ReID knowledge is both costly and difficult to scale, especially when the number of cameras is large [29, 31, 36].

To handle the De-ReID, we formulate a weakly-supervised unlearning framework where a labeled image set of unlearned persons and an unlabeled dataset of accessible persons are available to adapt a well-trained ReID model to De-ReID. Given that the core of ReID is to overcome cross-view variations, such as different camera viewing angles and lighting conditions, we propose variation-guided identity shift that learns De-ReID based on the variations in images. The core of our method is to guide the ReID model bias towards environmental information (e.g., lighting) when encountering unlearned persons, which shifts the unlearned persons out of the view-invariant feature space. At the same time, we keep the model remains robust to cross-view variations when querying accessible persons.

Specifically, for unlearned persons, we guide the model to **distinguish** their images and the images augmented by variations. Compared with distinguishing images under different camera views, learning to distinguish image and its augmentation results is more effective for unlearning since the augmented images are highly similar to the original image (e.g., body shape) but varies in style information (e.g., brightness and color tone). Therefore, the model is biased to encode the style information to push away these images. In contrast, for accessible persons, we **pull closer** their images and augmented versions in the feature space, expecting the model to overcome the variations when matching them. Moreover, we introduce relation regularization to characterize the desired properties of De-ReID in terms of relations.

In summary, our main contributions are as follows:

- We are the first to investigate the challenging De-ReID task and conduct a new benchmark.
- We formulate a De-ReID framework and propose variation-guided identity shift that achieves De-ReID by fitting the variations in images of interest while against the variations in others to preserve ReID knowledge.
- To further utilize the identity annotations in unlearned person images for learning, we propose relation regularization that characterizes the desired properties of ReID and De-ReID. Both qualitative and quantitative results demonstrate the effectiveness of our method.

2. Related Works

- Person re-identification. Based on deep learning, advanced ReID models automatically learn a discriminative feature embedding from a large amount of data [20]. Various network architectures [19, 21, 22, 30, 39, 42, 44, 45] and loss functions [10, 40, 41, 48] are developed to enhance feature representation learning. Besides, the challenges in ReID including cross-modality [25, 31, 51, 55], cloth-changing [23, 26], low-resolution [62], occlusion [52], and unsupervised ReID [7, 50] are also widely studied.

- Privacy protection for person images. Apart from pursuing high performance, protecting human privacy in human-centric tasks is also important [54, 59, 60]. A stream of research focuses on generating anonymized images by removing or obscuring identity-related information, such as faces and silhouettes, while preserving the image's utility for subsequent tasks, such as action recognition [1, 4, 15].

In the field of ReID, Ye et al. [54, 57] proposed a generative model to anonymize the person images to avoid privacy leakage while keeping the ReID utility. Wang et al. [46] add imperceptible perturbations to the images of in the database to prevent malicious use of the images. Particularly, one stream of works explores person de-identification [1, 4, 15], which obscures the identity information in images or videos, including face and silhouette. The new sensors including event cameras and LiDAR are also investigated for privacy-preserving ReID [2, 6, 14, 17].

Despite their progress, previous works mainly focus on preventing the malicious use of images while we aim to prevent the malicious use of models. Technically, previous works [54, 57] learn to anonymize person images and jointly adapt the ReID model to ensure ReID utility. Their methods cannot prevent the ReID model from retrieving specific persons, as well as the methods that explore new sensors such as event cameras for ReID [2, 6, 14, 17]. Besides, although PRIDE [46] protects the images in the database from being maliciously used, the attacker can also collect images that are not in the database to track specific persons. Moreover, some works in person de-identification only concern the identity-unrelated utility [1, 4, 15], such as action recognition, while De-ReID requires identifying accessible persons.

- Machine unlearning. De-ReID has a similar goal to machine unlearning. Machine unlearning is first proposed to eliminate the influence of specific data on the deep model as if the model never uses these data for training, which enables users to erase their personal data [3, 5, 16, 18, 37].

Kurmanji et al. [24] further apply machine unlearning for various purposes such as removing the bias from the model or eliminating the negative effects of mislabeled samples. To maximize the error in forgetting classes, they maximize the difference between the outputs of the learned model and the original model in the forgetting samples. Choi et al. [11] propose to jointly maximize the loss on the forgetting data and learn on the other data. Chen et al. [8] unlearn specific classes by adjusting their decision boundary. The forgetting samples are assigned a nearest but incorrect label for boundary shrink. Zhao et al. [58] propose GS LoRA to maintain performance on remaining classes when maximizing the loss on the forgetting data. Ye et al. [53] enforce that the model should be different from the original model in terms of the attention map of intermediate features to forget classes.

Our work is different from existing works. Firstly, even if the model does not see the unlearned person in training, the ReID model can effectively match the cross-view images of unlearned persons. Hence, the methods eliminating the influence of samples on network weights are not applicable to De-ReID. For the methods that maximize the errors in the forgetting classes, they mainly focus on classification tasks. Differently, ReID pursues discriminative features for matching. Even if the unlearned persons cannot be correctly classified as their methods desired, the model could retrieve the correct images as long as the feature distances between the query and the correct images are smaller than others.

3. De-ReID Learning

- Problem formulation. Given a well-trained ReID model denoted as $f_p(\cdot)$ and a set of unlearned persons $\mathcal{P}_u$, our goal is to learn a model $f(\cdot)$ that the model cannot match the images of unlearned persons while keeping the ReID performance for other persons that are accessible to model. $f(\cdot)$ is initialized as $f_p(\cdot)$, and $f_p(\cdot)$ is kept unchanged. An ideal De-ReID model $f(\cdot)$ is expected to satisfy the conditions:

$$\begin{cases} \phi_f(x, x_p) < \phi_f(x, x_n), \forall x \in \mathcal{P}_u, \forall x_n \text{ s.t. } y_n \neq y, \\ \phi_f(x, x_p) > \phi_f(x, x_n), \forall x \notin \mathcal{P}_u, \forall x_n \text{ s.t. } y_n \neq y, \end{cases} \tag{1}$$

where $\phi_f(\cdot,\cdot)$ is the similarity metric in the feature space of $f(\cdot)$. x is a query image with label y, and likewise y_p and y_n are labels of x_p and x_n. x_p is a different image containing the same person with x, but x_n contains a different person.

To learn $f(\cdot)$ for approximating Eq. 1, we assume a labeled image set of M_T unlearned persons denoted as $S_T = \{(x_i^t, y_i^t)\}_{i=1}^{|S_T|}$ and an unlabeled image set of accessible persons denoted as $S_O = \{x_j^o\}_{j=1}^{|S_O|}$ are available for fine-tuning. x_i^t is the i-th image in S_T with corresponding label $y_i^t \in \{1, 2, ..., M_T\}$. x_j^o is the j-th image in S_O. The persons in S_O are different from those in S_T and the persons in testing. While unlearning the specific persons based on the labeled data S_T seems straightforward (e.g., by maximizing the classification error), it is hard to ensure only the knowledge strictly subject to matching the unlearned persons is eliminated, rather than simply destroy identity discriminative feature embedding.

Considering the ubiquitous existence of cross-view variations, we propose variation-guided identity shift based on variations in images to unlearn the specific person in view-invariant feature space. Besides, we formulate the goals of De-ReID in the form of images relations and present relation regularization. An overview of our method is in Fig. 2.

3.1. Variation-guided Identity Shift

Intuitively, person ReID requires a view-invariant feature space, where image features are identity discriminative regardless of camera views. Hence, we propose to shift the unlearned persons out of the view-invariant feature space, and therefore their images from different camera views cannot be properly associated. To this end, we guide the model to fit the variations in images of unlearned persons and overcome the variations for accessible persons, forming asymmetric learning objectives. In this way, the images of unlearned persons will be encoded into distinct features affected by the variations in images while the images of accessible persons are encoded as view-invariant features.

To introduce abundant variations, we employ augmentation function $\mathcal{T}(\cdot)$ to obtain an augmented view $\hat{x}_i^t = \mathcal{T}(x_i^t)$ for each image x_i^t. To simulate the natural cross-view variations, $\mathcal{T}(\cdot)$ should not change the identity of x_i^t but properly adjust the images' color, brightness, and so forth, ensuring that $\hat{x}_i^t$ contains the same person with x_i^t but is different in image style. We ablate the design of $\mathcal{T}(\cdot)$ in experiments in our Appendix. To shift the unlearned persons out of the view-invariant feature space, the model $f(\cdot)$ should prioritize rich environmental information unrelated to identity when extracting features from the unlearned person's image x_i^t. In other words, the feature $f(x_i^t)$ should encode rich environmental information (e.g., background, image style) that cannot be applied for matching unlearned person in other cameras. To this end, we guide $f(\cdot)$ to push away x_i^t and its augmented view $\hat{x}_i^t$ in the feature space, which encourages the model to fit the variations for images of the unlearned person. Formally, the learning objective for shifting the unlearned persons out of view-invariant feature space is:

$$L_{VIS}^p = [\sigma_c - \|f(x_i^t) - f(\hat{x}_i^t)\|_2^2]_+, \tag{2}$$

where $[a]_+ = max(a, 0.0)$ and σ_c is the margin to bound the loss. Through optimization, the features of unlearned person images become *image-specific* and cannot be used for cross-view matching since the features are severely affected by the variations.

Simply learning with Eq. 2 alone will cause the model to forget the ReID knowledge, and fail to extract features *robust to variations* for the accessible persons. It will cause

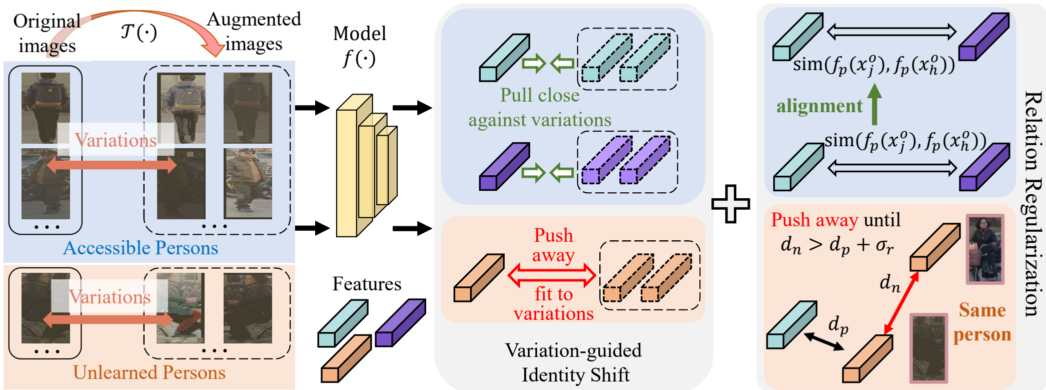

Figure 2. An overview of our method. In variation-guided identity shift, the model is guided to adapt to the variations when encountering images of unlearned persons, thereby shifting them out of the view-invariant feature space. Meanwhile, the view-invariant feature space is maintained for other accessible persons by guiding the model to overcome the variations when encountering images of these individuals. Relation regularization formulates the desired properties of De-ReID in terms of image relations and ensures the feature discriminability.

feature space collapse instead of the specific identity shift. To maintain a view-invariant feature space for re-identifying accessible persons, we guide the model to overcome the variations for images of these persons. For image x_j^o of an accessible person, we obtain its augmented view $\hat{x}_j^o = \mathcal{T}(x_j^o)$. We encourage the model to pull closer x_j^o and $\hat{x}_j^o$ to remind the model of extracting *robust* features for accessible persons. Formally, the learning function for accessible person is:

$$L_{VIS}^o = \|f(x_j^o) - f(\hat{x}_j^o)\|_2^2. \tag{3}$$

Combining Eq. 3 and Eq. 2, we form an asymmetric objective to shift the identities of unlearned persons while preserving a view-invariant feature space for accessible persons:

$$L_{VIS} = L_{VIS}^p + L_{VIS}^o. \tag{4}$$

- **Discussion on asymmetric learning scheme.** L_{VIS} alters knowledge within the ReID model by treating the variations in images asymmetrically. The two asymmetric parts in the L_{VIS} are **collaborative**, not isolated. Intuitively, one may think that guiding the ReID network against the variations (L_{VIS}^o) is merely for keeping ReID knowledge for accessible persons. **However**, L_{VIS}^o is also helpful for unlearning because it maintains a view-invariant feature space, which in turn gives the direction for shifting the identities of unlearned persons (L_{VIS}^p): out of the view-invariant space. As a result, more unlabeled images of accessible persons can improve the unlearning effect, which is not shown in other methods.

3.2. Relation Regularization

The relations between pairwise images are critical in the context of person ReID, directly reflecting the discriminative capability of the features. As formulated in Eq. 1, there should be some accessible persons spread around an unlearned person. Hence, we introduce relation regularization to facilitate De-ReID and ReID learning. Specifically, we constraints that, in the feature space, the distance between the images of the same unlearned person should be greater than the distance to the accessible persons. This TRiplet Constraint facilitates retrieving images of different persons when querying an unlearned person's image.

$$\begin{gathered} L_{TRC} = [d_p - d_n + \sigma_r]_+, \\ \text{where } d_p = \|x_i^t - x_{i,K}^o\|_2^2, d_n = \|x_i^t - x_h^t\|_2^2. \end{gathered} \tag{5}$$

Here, $x_{i,K}^o (x_{i,K}^o \in S_O)$ represents the K-th nearest image to x_i^t, and σ_r is the margin. x_h^t is an image containing the same person as x_i^t ($y_h^t = y_i^t$). L_{TRC} constrains that the distance between pairwise images of the same unlearned person should be at least larger than $d_p + \sigma_r$, where d_p is dynamically determined based on the distance to the K-th nearest accessible person. When K is larger, L_{TRC} encourages more accessible persons to spread around x_i^t. Through optimizing L_{TRC}, the top retrieval results mainly contain different persons when retrieving unlearned person images from feature space.

Complementary to variation-guided identity shift that changes the cross-view invariance in features, we employ Relation Consistent Regularization on dataset S_O to maintain the feature discriminability during De-ReID fine-tuning. The regularization is to ensure that the relations between accessible individuals remain as discriminative as those in the original model $f_p(\cdot)$ during the unlearning process:

$$L_{RCR} = \|\text{sim}(f_p(x_j^o), f_p(x_h^o)) - \text{sim}(f(x_j^o), f(x_h^o))\|_2^2. \tag{6}$$

By cooperating with L_{VIS}, the model can extract discriminative and view-invariant features for accessible persons when unlearning the specific persons.

- **Overall Learning objective.** The model $f(.)$ is trained end-to-end with an overall learning objective as:

$$L = \lambda_1 * L_{VIS} + \lambda_2 * L_{TRC} + L_{RCR}, \tag{7}$$

where λ_1 and λ_2 are trade-off parameters.

4. Experiments

4.1. Datasets and Experimental Setup

- Datasets. We conduct experiments in Market-1501 [61], MSMT17 [47], Occ-Duke [34] and SYSU-MM01 [49] datasets. Due to the limited space, we leave the results in Occ-Duke and SYSU-MM01 datasets in our Appendix. Market-1501 (abbreviated as Market) consists of 1,501 persons, and all the images are captured from 6 camera views. In the conventional ReID protocol, there are 750 persons for training and 751 persons for testing. MSMT17 is a large-scale ReID dataset, containing images of 4101 persons from 15 disjoint camera views. Conventionally, 1,041 persons are for training and 3,060 persons for testing. For evaluating De-ReID, we slightly modify the conventional protocol by referring to the experimental setups of other machine unlearning works [8, 24, 53, 58].

Specifically, in each dataset, we take M_T persons from the testing set as the unlearned persons. Considering the scale of the datasets, we set $M_T \in \{25, 50, 75, 100\}$ in MSMT17 and $M_T \in \{25, 50\}$ in Market1501. For each unlearned person, we split his images into two image sets from disjoint cameras. One set is added to the training set and the other is kept in the testing set. After adding the unlearned persons to the training set, we divide the training set into two subsets. One pre-training subset consists of accessible persons with identity annotations and is used for pretraining the ReID model. There are 900 accessible persons in the pre-training subset of MSMT17 and 600 accessible persons in the pre-training subset of Market-1501.

The fine-tuning subset consists of *unlabeled images of accessible persons* and *labeled images of unlearned persons*, simulating the practical scenario where we need to efficiently fine-tune the ReID model without collecting additional labeled images of accessible persons. The dataset statistics are shown in Table 1. We emphasize that the images of the unlearned persons in the testing set under the cameras are different from those of the training set. The identities of accessible persons in the testing set are different from those in the training set. More details are in Appendix.

- Experimental setup. In each dataset, the pre-training subset is used to learn a ReID model in a supervised manner as $f_p(\cdot)$. After obtaining the well-trained ReID model, the fine-tuning subset is for learning De-ReID model $f(\cdot)$ in our main experiments.

- Evaluation metrics. We adopt the widely used Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) to evaluate our performance. Following previous works [38, 58], we also adopt H-Means to comprehensively evaluate the model's ability of ReID in accessible persons and De-ReID in unlearned persons. Formally,

Table 1. Statistics in the case of $M_T = 100$ in MSMT17 and $M_T = 50$ in Market1501. M_T is the number of unlearned persons for De-ReID. Notably, the accessible persons in training are different from those in testing. "Train" refers to the fine-tuning subset. For the unlearned persons, the images in training and those in testing are from different cameras. More details are in Appendix.

	MSMT17				Market-1501			
	M_T	$\|S_T\|$	M_O	$\|S_O\|$	M_T	$\|S_T\|$	M_O	$\|S_O\|$
Train	100	4295	141	5061	50	931	151	2284
Query	100	533	2960	10706	50	150	700	3068
Gallery	100	4136	2960	74165	50	918	700	18033

$$\text{H-Mean} = \frac{2 * \text{R-1}_O(f) * \Delta\text{R-1}_T}{\text{R-1}_O(f) + \Delta\text{R-1}_T}, \quad \text{where } \Delta\text{R-1}_T = \text{R-1}_T(f_p) - \text{R-1}_T(f). \tag{8}$$

where $\text{R-1}_O(f)$ is model f's Rank-1 accuracy in accessible persons, and $\Delta\text{R-1}_T$ is the changes in the Rank-1 accuracy in unlearned persons after De-ReID fine-tuning using different methods. In the following, we only discuss the performance of $f(\cdot)$ and therefore R-1_O means $\text{R-1}_O(f)$ by default.

4.2. Implementation Details

By default, we use ViT-B [13] as the ReID backbone $f_p(\cdot), f(\cdot)$. The images are resized to 256×128, and the data augmentation includes random crop and flip. For introducing variations in variation-guided identity shift, we use stronger augmentation derived from RandAugment [12]. We detailed the augmentation and conducted corresponding ablation studies in the Appendix. The AdamW [32] optimizer is adopted for De-ReID learning with an initial learning of 3e-4. The weight decay is set to 0. We empirically set trade-off parameters $\lambda_1 = 1.0$ and $\lambda_2 = 1.0$. The batch size for accessible persons is 48, and that for unlearned persons is 32. The above hyper-parameters are the same in all experiments. K in Eq. 5 is set to 20.

4.3. Comparison with Related Methods

- Compared methods. Since this is the first work for De-ReID, we compare our method with existing methods potentially applicable to ReID. We pretrain a vanilla ViT-B in the pretraining subset of each dataset and then apply these methods to the ViT-B in the fine-tuning subset for De-ReID. We report the performance of the machine unlearning methods, including GS-LoRA [58], LIRF [53], SCRUB [24], and Boundary Shrink [8]. Besides, we implement a naive solution: label augmentation based on cameras, which assign different labels for images containing the same person but under different cameras. This naive solution is denoted as "LabelAug", and the triplet loss and cross-entropy loss [33] are applied for training LabelAug.

Notably, previous methods usually assume that the labeled data of accessible classes are available. Since the data

Table 2. Comparisons in the MSMT17 dataset. "BS" is the Boundary Shrink [8]. "H" denotes "H-Mean". R-1_T is the Rank-1 accuracy on the unlearned persons, and R-1_O is the Rank-1 accuracy on the accessible persons. R-1_T is expected to be low for unlearning specific persons, while a higher R-1_O and H is preferred. Since we apply additional augmentation for asymmetric contrastive learning, we further apply our augmentation to LIRF, denoted as LIRF*. "RR" is the Re-Ranking [63]. The method with the best/second-best H-Mean is marked in red/ blue. Blue background means the H-Mean are improved by Re-Ranking.

Method	$M_T = 25$			$M_T = 50$		
	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$
LabelAug	54.6	73.8	38.1	56.8	73.5	32.4
+ RR	62.3	76.9	29.2	65.3	77.2	21.2
BS	63.8	78.7	27.3	54.6	74.9	35.2
+ RR	70.0	82.2	18.3	62.5	78.7	25.3
SCRUB	43.8	75.0	49.1	45.2	74.4	45.1
+ RR	52.3	78.4	41.3	53.7	78.2	36.6
LIRF	10.0	62.5	66.2	18.5	54.6	56.8
+ RR	6.9	65.4	69.2	20.5	57.3	57.2
LIRF*	18.5	68.8	65.1	35.1	69.7	52.8
GS-LoRA	56.9	77.5	35.9	60.2	80.7	28.6
+ RR	56.9	79.7	36.2	64.1	82.9	23.2
Ours	4.6	77.0	76.3	10.8	72.9	69.7
+ RR	4.6	79.0	77.3	10.4	75.1	70.9

Table 3. Comparison results in the MSMT17 dataset when $M_T = 75$ and $M_T = 100$. The notations are consistent with Table 2

Method	$M_T = 75$			$M_T = 100$		
	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$
LabelAug	61.4	72.2	29.2	62.7	71.9	27.2
+ RR	69.1	75.8	18.6	66.8	75.7	21.8
BS	57.2	69.1	33.9	54.8	68.3	36.3
+ RR	65.6	74.3	23.7	61.4	72.7	28.9
SCRUB	54.7	74.8	37.5	50.5	67.8	40.6
+ RR	62.9	79.0	27.7	55.7	72.8	35.9
LIRF	18.6	51.9	56.1	20.8	49.7	53.8
+ RR	19.6	55.1	57.5	24.8	52.4	53.5
LIRF*	38.4	65.0	50.5	15.6	50.9	56.7
GS-LoRA	62.4	79.5	28.4	62.1	76.1	28.3
+ RR	65.3	81.8	24.5	63.8	78.6	26.2
Ours	12.4	69.8	68.5	13.1	67.1	66.7
+ RR	13.1	71.6	69.0	15.3	68.8	66.4

of accessible persons is unlabeled in our experiments, we apply our L_{RCR} as a regularization for previous methods. We further apply our methods to advanced ReID models, including DCFormer [27], PAT [28], and PHA [56].

- Results. Table 2 and Table 4 shows the comparison results. From the tables, we can observe that our method achieves the best performance in terms of H-Mean. Since the post-processing methods like Re-Ranking [63] can improve the results, we investigate whether these methods can reduce the unlearning effect when retrieving unlearned persons. The results show that existing unlearning methods usually get worse results after post-processing. For example, in Table 2, the Rank-1 accuracy of the SCRUB on unlearned persons is clearly increased from 43.8% to 52.3% when $M_T = 25$,

Table 4. Comparison results in the Market-1501 dataset.

Method	$M_T = 25$			$M_T = 50$		
	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$	R-$1_T \downarrow$	R-$1_O \uparrow$	H $\uparrow$
LabelAug	50.7	76.5	51.3	64.7	77.4	42.5
+ RR	49.3	77.6	52.8	66.0	78.9	41.3
BS	57.3	73.1	44.5	62.0	72.2	44.3
+ RR	54.7	75.3	47.4	65.3	74.0	41.4
SCRUB	50.7	76.0	51.2	54.7	70.2	50.4
+ RR	56.0	77.4	46.6	64.7	71.8	41.6
LIRF	25.3	84.8	72.9	31.3	75.9	68.7
+ RR	26.7	84.4	71.9	32.7	75.7	67.7
LIRF*	37.3	86.0	64.5	39.3	79.0	64.6
GS-LoRA	56.0	91.7	48.9	65.3	91.4	43.7
+ RR	54.7	91.3	50.2	67.3	91.3	41.3
Ours	10.7	91.1	84.4	12.0	84.4	83.2
+ RR	5.3	91.0	87.4	10.0	83.5	83.8

demonstrating that the model does not effectively unlearn the persons and there are potential cues to re-identify them.

While other methods achieve a worse H-mean after Re-ranking, our method can benefit from Re-Ranking [63] since the overall performance H-Mean is improved in most cases. For example, in the case of $M_T = 25$ in Market dataset, the H-Means of our method is increased to 87.4% from 84.4%.

Compared with the other methods, Boundary Shrink changes the classification decision boundary of the unlearning classes, i.e., the unlearned persons in this work. However, the feature representation is the core for person image matching instead of the classifier. In contrast, our method directly optimizes the feature representation for De-ReID. As a result, our method outperforms Boundary Shrink by 30.4% H-Mean in MSMT17 with $M_T = 100$.

LIRF unlearns classes by constraining the attention map of intermediate features to be significantly different before and after unlearning. Since it directly constrains the features, it obtains the second-best results. However, the attention map has lots of knowledge about extracting identity information from images. Hence, their constraints largely hurt the ReID ability on accessible persons, and our method outperforms LIRF by 12.9% at H-Mean in MSMT17 when $M_T = 100$. Besides, since we apply additional augmentation in variation-guided identity shift, we further validate the effectiveness of the additional augmentation by applying the augmentation in LIRF, denoted as LIRF*. However, the H-Mean is improved only in the case of $M_T = 100$ in MSMT17 and decreases in other cases, demonstrating that additional augmentation does not necessarily lead to higher performance.

GS-LoRA effectively protects the ReID ability on accessible persons but falls behind in unlearning specific persons. LabelAug directly guides the model in distinguishing the images of unlearned persons under different camera views, but it has an unsatisfactory performance. We speculate it is because the model overfits the images in training and fails to unlearn the specific persons under cameras unseen in training. Differently, our method employs variation-guided

Table 5. Evaluation with different ReID models. Regardless of the ReID models, our method performs the best in terms of erasing the knowledge about unlearned persons (quantified by R-1_T) and keeping ReID knowledge for accessible persons (R-1_O). “Null” indicates that no De-ReID method is applied.

ReID model	De-ReID method	R-1_T ↓	R-1_O ↑	H ↑
DCFormer	Null	81.2	84.4	—
	LabelAug	44.1	54.0	46.2
	GS-LORA	75.3	81.9	16.4
	LIRF	18.0	51.0	57.7
	Ours	**17.8**	**66.1**	**66.3**
PAT	Null	72.8	75.6	—
	LabelAug	38.3	43.3	40.1
	GS-LORA	52.5	60.9	33.5
	LIRF	16.7	42.6	49.4
	Ours	**12.0**	**55.0**	**59.0**
PHA	Null	70.2	78.7	—
	LabelAug	44.3	51.4	41.2
	GS-LORA	56.4	68.3	33.6
	LIRF	23.6	50.5	52.7
	Ours	**3.9**	**51.2**	**60.8**

identity shift (L^p_{VIS}, L^o_{VIS}) and guides the model to handle the variations adaptively. Fitting the variations in unlearned images helps the model bias to the environmental information when querying the unlearned persons. Simultaneously, the model is guided to be robust to the variations when querying accessible persons. As a result, the unlearned persons are shifted out of the view-invariant feature space while the accessible persons are retained.

- **Evaluations on other ReID methods.** Apart from the comparison on the vanilla ViT backbone, we also compared our method with related methods on the advanced ReID models [27, 28, 56]. The experimental results in Table 5 demonstrate that our method is superior to others regarding unlearning different ReID models.

4.4. Ablation Study

We conduct experiments to show the effectiveness of each component in our design. Due to the space limit, we report the experimental results in Market and $M_T \in \{25, 100\}$ in MSMT17 in Table 6, and the rest in Appendix. Regardless of M_T, our designs clearly improve H-Mean that comprehensively evaluates the effectiveness of unlearning specific persons and ReID performance on accessible persons.

- **The effectiveness of the relation regularization.** From Table 6, we first observe that the model suffers from catastrophic forgetting in ReID knowledge and collapses in the absence of L_{RCR}. Although our variation-guided identity shift pulls closer the images and the augmented images for accessible persons, it mainly preserves the view-invariant feature embedding but it fails to preserve the feature discriminability, leading to model collapses. L_{RCR} preserves the discriminability of features by preserving the relations between accessible persons.

Another constraint L_{TRC} attempts to push away the images between unlearned persons to prevent them from being cross-view retrieved, referring to the distance between accessible persons and unlearned persons. We observe a degraded H-Mean in the absence of L_{TRC} in all cases. For example, when deprecating L_{TRC} in the case of $M_T = 50$ in Market, the Rank-1 accuracy of the unlearned persons is increased to 24.0% from 12.0%, meaning that more unlearned person images are correctly matched and is not expected. As a result, the H-Mean drops by 9.8% compared with the full model. In the case of $M_T = 100$ in MSMT17, the Rank-1 accuracy of the unlearned persons increases to 18.3% from 13.1% without L_{TRC}, and the H-Mean drops by 3.3%.

Table 6. Ablation study in MSMT17 dataset. “W/o” means “without”, and other notations are the same as Table 2. $f_p(\cdot)$ is the initial model for De-ReID learning. The full model achieves the best H-Mean. Notably, without L_{RCR}, the model forgets ReID knowledge and collapses. “SD” denotes the “self-augmented discrimination”.

Method	MSMT17					
	$M_T = 25$			$M_T = 100$		
	R-1_T ↓	R-1_O ↑	H ↑	R-1_T ↓	R-1_O ↑	H ↑
$f_p(\cdot)$	80.3	85.4	—	79.5	85.4	—
Components in Relation regularization						
W/o L_{TRC}	5.5	73.4	74.1	18.3	65.3	63.2
W/o L_{RCR}	—	—	—	—	—	—
Components in Variation-guided Identity Shift						
W/o SD	4.6	74.5	75.1	10.5	58.7	63.4
W/o L^o_{VIS}	10.0	75.8	72.9	18.0	63.2	62.3
W/o L_{VIS}	20.5	77.4	67.5	41.3	70.0	49.4
Full model	**4.6**	**77.0**	**76.3**	**13.1**	**67.1**	**66.7**

Table 7. Ablation study in Market-1501.

Method	Market-1501					
	$M_T = 25$			$M_T = 50$		
	R-1_T ↓	R-1_O ↑	H ↑	R-1_T ↓	R-1_O ↑	H ↑
$f_p(\cdot)$	89.3	96.3	—	94.0	96.3	—
Components in Relation regularization						
W/o L_{TRC}	20.0	89.7	78.2	24.0	78.3	73.9
W/o L_{RCR}	—	—	—	—	—	—
Components in Variation-guided Identity Shift						
W/o SD	24.0	90.3	75.8	11.3	79.7	81.2
W/o L^o_{VIS}	10.7	84.5	81.4	12.7	77.3	79.2
W/o L_{VIS}	46.7	93.9	58.6	55.3	89.9	54.1
Full model	**10.7**	**91.1**	**84.4**	**12.0**	**84.4**	**83.2**

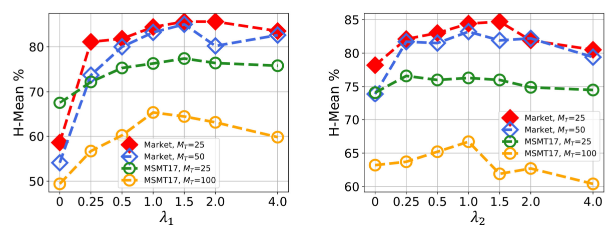

Figure 3. Evaluation results under various λ_1 and λ_2.

The hyper-parameter λ_2 controls the weight of L_{TRC}. We evaluate our model under different λ_2 and show the performance in Fig. 3. When $\lambda_2 = 0$, L_{TRC} is not employed and the performance is inferior. As λ_2 increases, the performance is improved. However, L_{TRC} should not be set to too large to avoid overwhelming other learning objectives.

- The effectiveness of the variation-guided identity shift. The variation-guided identity shift (VIS) emphasizes the style difference between images of unlearned persons in the feature space while keeping the other images robust to the style variations. Overall, comparing the results of "W/o L_{VIS}" and the full model in Table 6, we observe that the ReID model struggles to unlearn the specific persons without L_{VIS}. For example, in MSMT17 dataset, the Rank-1 accuracy on unlearned persons is 20.5%/41.3% when $M_T = 25/100$, which is 15.9%/28.2% larger than the full model. This is because ReID requires view-invariant features and our model correspondingly guides the model to be biased to encode environmental information for images of unlearned persons. As a result, we effectively prevent the images of unlearned persons from being re-identified. The hyper-parameter λ_1 controls the weight of L_{VIS} in optimization. We evaluate our model under different λ_1 and show the performance in Fig. 3.

We also studied the effectiveness of the designs in VIS. As discussed in the introduction, we argue that simply pushing away the images under different camera views may learn the spurious cues for De-ReID, and therefore we push away the image and the corresponding augmented image in our VIS. We refer to this as "self-augmented discrimination". To verify its effectiveness, we conduct an experiment that pushes away the image and the augmented image from another image, denoted as "W/o SD", and observe a significant performance drop. For example, in the case of $M_T = 25$ in Market dataset, the Rank-1 accuracy on the unlearned person increases to 24.0% from 10.7% without SD, and the H-Mean drops by 8.6%. In other cases, the H-Mean drops without using the SD, demonstrating the effectiveness of SD.

Moreover, we observe that guiding the ReID to overcome the variations for accessible persons is important for balancing the ReID ability and De-ReID of unlearned persons. Deprecating the L^{o}_{VIS} leads to $3\% \sim 4\%$ performance drops in H-Mean in different cases, and reduce the protection effect since the rank-1 accuracy of unlearned persons increases.

- The impact of the unlabeled sample. We evaluate our method in MSMT17 under different ratios between unlabeled images of accessible persons and labeled images of unlearned persons in the fine-tuning subset. The results are in Table 8.

From the results, we obtain two conclusions. (i) More unlabeled data is helpful for both unlearning specific persons and keeping ReID performance on the accessible persons. When the number of accessible persons decreases, the Rank-1 accuracy of the unlearned persons (denoted as R-1_T) increases, which means the unlearned persons are easier to retrieve by the ReID model. Simultaneously, the Rank-1 accuracy of the accessible persons (R-1_O) decreases, which means the ReID model cannot properly re-identify the accessible persons. (ii) Although the performance of our method decreases when the unlabeled images become fewer, our method can effectively utilize the publicly available unlabeled person images for unlearning, demonstrating the scalability of our method. When using additional unlabeled images from LUPerson dataset, our method can achieve similar performance even though the number of accessible persons in MSMT17 decreases from 141 to 60.

Table 8. Evaluations under different numbers of accessible persons in MSMT17 when $M_T = 100$. M_O is the number of accessible persons in the fine-tuning subset. 'With LUP' means using the images in LUPerson dataset as auxiliary unlabeled data and the number in parentheses denotes the performance gain. By default, we do not use LUPerson dataset.

M_O	Without LUP			With LUP		
	R-$1_T\downarrow$	R-$1_O\uparrow$	H$\uparrow$	R-$1_T\downarrow$	R-$1_O\uparrow$	H$\uparrow$
141	13.1	67.1	66.7	-	-	-
100	15.8	61.3	62.5	14.1	68.2	**66.8 (+4.3%)**
80	15.8	54.3	58.6	15.4	67.7	**65.9 (+7.3%)**
60	17.4	51.4	57.6	15.4	65.6	**64.8 (+7.2%)**

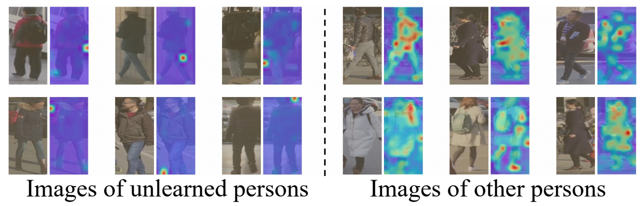

Figure 4. Attention map on images of the unlearned persons and the images of accessible persons.

4.5. Visualization

We visualize the attention map on different persons in the testing set in Fig. 4. From the figure, for the images of unlearned persons, we can observe that the model's attention mainly focus on the background which contains abundant environmental information. At the same time, the model pays diverse attention to the informative parts of images for extracting discriminative features for accessible persons. These results demonstrate the effectiveness of our method.

5. Conclusion

To address the privacy concerns in person ReID, we pioneer the exploration of the De-ReID problem that guides the ReID model to forget the knowledge about matching specific persons. We formulate a framework and propose a novel variation-guided identity shift method for De-ReID. Our method unlearns the specific person shifting them out of the view-invariant feature space and simultaneously keeps the ReID knowledge by overcoming the cross-view variations for accessible persons. We further introduce relation regularization to characterize the desired properties of De-ReID. Extensive experiments conducted on a new benchmark show the superiority of our method.

6. Acknowledgment

This work was supported partially by NSFC(U21A20471, 92470202), by the National Key Research and Development Program of China (2023YFA1008503), Guangdong NSF Project (No. 2023B1515040025)

References

[1] Prachi Agrawal and P. J. Narayanan. Person de-identification in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011. 2

[2] Shafiq Ahmad, Pietro Morerio, and Alessio Del Bue. Person re-identification without identification via event anonymization. In *2023 IEEE/CVF International Conference on Computer Vision*, 2023. 2

[3] Lucas Bourtoule, Varun Chandrasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine unlearning. In *IEEE Symposium on Security and Privacy*, 2021. 1, 2

[4] Karla Brkic, Ivan Sikiric, Tomislav Hrkac, and Zoran Kalafatic. I know that person: Generative full body and face de-identification of people in images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. 2

[5] Jonathan Brophy and Daniel Lowd. Machine unlearning for random forests. In *International Conference on Machine Learning*, 2021. 1, 2

[6] Chengzhi Cao, Xueyang Fu, Hongjian Liu, Yukun Huang, Kunyu Wang, Jiebo Luo, and Zheng-Jun Zha. Event-guided person re-identification via sparse-dense complementary learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[7] Hao Chen, Yaohui Wang, Benoit Lagadec, Antitza Dantcheva, and Francois Bremond. Learning invariance from generated variance for unsupervised person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2

[8] Min Chen, Weizhuo Gao, Gaoyang Liu, Kai Peng, and Chen Wang. Boundary unlearning: Rapid forgetting of deep networks via shifting the decision boundary. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 3, 5, 6

[9] Tianshui Chen, Tao Pu, Lingbo Liu, Yukai Shi, Zhijing Yang, and Liang Lin. Heterogeneous semantic transfer for multi-label recognition with partial labels. *International Journal of Computer Vision*, 2024. 1

[10] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: A deep quadruplet network for person re-identification. In *IEEE/CVF International Conference on Computer Vision and Pattern Recognition*, 2017. 1, 2

[11] Dasol Choi and Dongbin Na. Towards machine unlearning benchmarks: Forgetting the personal identities in facial recognition systems. In *AAAI 2024 workshop on Privacy-Preserving Artificial Intelligence*, 2023. 3

[12] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 5

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 5

[14] Lijie Fan, Tianhong Li, Rongyao Fang, Rumen Hristov, Yuan Yuan, and Dina Katabi. Learning longterm representations for person re-identification using radio signals. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[15] Ivan Filković, Zoran Kalafatić, and Tomislav Hrkać. Deep metric learning for person re-identification and de-identification. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2016. 2

[16] Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[17] Wenxuan Guo, Zhiyu Pan, Yingping Liang, Ziheng Xi, Zhi Chen Zhong, Jianjiang Feng, and Jie Zhou. Lidar-based person re-identification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2

[18] Varun Gupta, Christopher Jung, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Chris Waites. Adaptive machine unlearning. *Advances in Neural Information Processing Systems*, 2021. 2

[19] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *IEEE/CVF International Conference on Computer Vision*, 2021. 1, 2

[20] Weizhen He, Yiheng Deng, Shixiang Tang, Qihao Chen, Qingsong Xie, Yizhou Wnag, Lei Bai, Feng Zhu, Rui Zhao, Wanli Ouyang, Donglian Qi, and Yunfeng Yan. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2

[21] Bingliang Jiao, Lingqiao Liu, Liying Gao, Guosheng Lin, Lu Yang, Shizhou Zhang, Peng Wang, and Yanning Zhang. Dynamically transformed instance normalization network for generalizable person re-identification. In *European Conference on Computer Vision*, 2022. 2

[22] Jiayu Jiao, Yu-Ming Tang, Kun-Yu Lin, Yipeng Gao, Andy J Ma, Yaowei Wang, and Wei-Shi Zheng. Dilateformer: Multi-scale dilated transformer for visual recognition. *IEEE Transactions on Multimedia*, 25:8906–8919, 2023. 2

[23] Xin Jin, Tianyu He, Kecheng Zheng, Zhiheng Yin, Xu Shen, Zhen Huang, Ruoyu Feng, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. Cloth-changing person re-identification from a single image with gait prediction and regularization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2

[24] Meghdad Kurmanji, Peter Triantafillou, Jamie Hayes, and Eleni Triantafillou. Towards unbounded machine unlearning. In *Conference on Neural Information Processing Systems*, 2023. 1, 2, 5

[25] Siyuan Li, Li Sun, and Qingli Li. Clip-reid: Exploiting vision-language model for image re-identification without concrete text labels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. 2

[26] Weijia Li, Saihui Hou, Chunjie Zhang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Yao Zhao. An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[27] Wen Li, Cheng Zou, Meng Wang, Furong Xu, Jianan Zhao, Ruobing Zheng, Yuan Cheng, and Wei Chu. Dc-former: Diverse and compact transformer for person re-identification. In *AAAI Conference on Artificial Intelligence*, 2023. 6, 7

[28] Yulin Li, Jianfeng He, Tianzhu Zhang, Xiang Liu, Yongdong Zhang, and Feng Wu. Diverse part discovery: Occluded person re-identification with part-aware transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021. 6, 7

[29] Kun-Yu Lin, Jiaming Zhou, Yukun Qiu, and Wei-Shi Zheng. Adversarial partial domain adaptation by cycle inconsistency. In *European Conference on Computer Vision*, pages 530–548. Springer Nature Switzerland Cham, 2022. 2

[30] Kun-Yu Lin, Jia-Run Du, Yipeng Gao, Jiaming Zhou, and Wei-Shi Zheng. Diversifying spatial-temporal perception for video domain generalization. *Advances in Neural Information Processing Systems*, 36:56012–56026, 2023. 2

[31] Kun-Yu Lin, Henghui Ding, Jiaming Zhou, Yu-Ming Tang, Yi-Xing Peng, Zhilin Zhao, Chen Change Loy, and Wei-Shi Zheng. Rethinking clip-based video learners in cross-domain open-vocabulary action recognition. *arXiv preprint arXiv:2403.01560*, 2024. 2

[32] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 5

[33] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 5

[34] Jiaxu Miao, Yu Wu, Ping Liu, Yuhang Ding, and Yi Yang. Pose-guided feature alignment for occluded person re-identification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 542–551, 2019. 5

[35] Tao Pu, Tianshui Chen, Hefeng Wu, Yukai Shi, Zhijing Yang, and Liang Lin. Dual-perspective semantic-aware representation blending for multi-label image recognition with partial labels. *Expert Systems with Applications*, page 123526, 2024. 1

[36] Tao Pu, Qianru Lao, Hefeng Wu, Tianshui Chen, Ling Tian, Jie Liu, and Liang Lin. Category-adaptive label discovery and noise rejection for multi-label recognition with partial positive labels. *IEEE Transactions on Multimedia*, pages 1–12, 2024. 2

[37] Ayush Sekhari, Jayadev Acharya, Gautam Kamath, and Ananda Theertha Suresh. Remember what you want to forget: Algorithms for machine unlearning. *Advances in Neural Information Processing Systems*, 2021. 2

[38] Takashi Shibata, Go Irie, Daiki Ikami, and Yu Mitsuzumi. Learning with selective forgetting. In *International Joint Conference on Artificial Intelligence*, 2021. 5

[39] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline). In *European Conference on Computer Vision*, 2018. 2

[40] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[41] Yu-Ming Tang, Yi-Xing Peng, and Wei-Shi Zheng. Learning to imagine: Diversify memory for incremental learning using unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9549–9558, 2022. 2

[42] Yu-Ming Tang, Yi-Xing Peng, and Wei-Shi Zheng. When prompt-based incremental learning does not meet strong pretraining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1706–1716, 2023. 2

[43] Yu-Ming Tang, Yi-Xing Peng, Jingke Meng, and Wei-Shi Zheng. Rethinking few-shot class-incremental learning: Learning from yourself. In *European Conference on Computer Vision*, pages 108–128. Springer, 2024. 1

[44] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. In *ACM International Conference on Multimedia*, 2018. 2

[45] Haochen Wang, Jiayi Shen, Yongtuo Liu, Yan Gao, and Efstratios Gavves. Nformer: Robust person re-identification with neighbor transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1, 2

[46] Lin Wang, Wanqian Zhang, Dayan Wu, Fei Zhu, and Bo Li. Attack is the best defense: Towards preemptive-protection person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2022. 2

[47] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 79–88, 2018. 5

[48] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision*, 2016. 2

[49] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 5

[50] Bin Yang, Jun Chen, and Mang Ye. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *IEEE/CVF International Conference on Computer Vision*, 2023. 2

[51] Bin Yang, Jun Chen, and Mang Ye. Shallow-deep collaborative learning for unsupervised visible-infrared person re-identification. In *IEEE/CVF International Conference on Computer Vision*, 2024. 2

[52] Jinrui Yang, Jiawei Zhang, Fufu Yu, Xinyang Jiang, Mengdan Zhang, Xing Sun, Ying-Cong Chen, and Wei-Shi Zheng. Learning to know where to see: A visibility-aware approach for occluded person re-identification. In *IEEE/CVF international conference on computer vision*, 2021. 2

[53] Jingwen Ye, Yifang Fu, Jie Song, Xingyi Yang, Songhua Liu, Xin Jin, Mingli Song, and Xinchao Wang. Learning with recoverable forgetting. In *European Conference on Computer Vision*, 2022. 1, 3, 5

[54] Mang Ye, Wei Shen, Junwu Zhang, Yao Yang, and Bo Du. Securereid: Privacy-preserving anonymization for person re-identification. *IEEE Transactions on Information Forensics and Security*, 2024. 2

[55] Mang Ye, Zesen Wu, Cuiqun Chen, and Bo Du. Channel augmentation for visible-infrared re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[56] Guiwei Zhang, Yongfei Zhang, Tianyu Zhang, Bo Li, and Shiliang Pu. Pha: Patch-wise high-frequency augmentation for transformer-based person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 6, 7

[57] Junwu Zhang, Mang Ye, and Yao Yang. Learnable privacy-preserving anonymization for pedestrian images. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2022. 2

[58] Hongbo Zhao, Bolin Ni, Haochen Wang, Junsong Fan, Fei Zhu, Yuxi Wang, Yuntao Chen, Gaofeng Meng, and Zhaoxiang Zhang. Continual forgetting for pre-trained vision models. 2024. 3, 5

[59] Zhilin Zhao, Longbing Cao, and Kun-Yu Lin. Revealing the distributional vulnerability of discriminators by implicit generators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8888–8901, 2022. 2

[60] Zhilin Zhao, Longbing Cao, and Kun-Yu Lin. Out-of-distribution detection by cross-class vicinity distribution of in-distribution data. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 2

[61] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Computer Vision, IEEE International Conference on*, 2015. 5

[62] Wei-Shi Zheng, Jincheng Hong, Jiening Jiao, Ancong Wu, Xiatian Zhu, Shaogang Gong, Jiayin Qin, and Jianhuang Lai. Joint bilateral-resolution identity modeling for cross-resolution person re-identification. *International Journal of Computer Vision*, 2022. 2

[63] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 6