Dual Pseudo-Labels Interactive Self-Training for Semi-Supervised Visible-Infrared Person Re-Identification

Jiangming Shi[1*], Yachao Zhang[3*], Xiangbo Yin[2], Yuan Xie[4, 5†],
Zhizhong Zhang[4, 5], Jianping Fan[6], Zhongchao Shi[6], Yanyun Qu[1,2†]
[1] Institute of Artificial Intelligence, [2] School of Informatics, Xiamen University,
[3] Tsinghua Shenzhen International Graduate School, Tsinghua University,
[4] East China Normal University, [5] Chongqing Institute of East China Normal University, [6] Lenovo

jiangming.shi@outlook.com, yachaozhang@sz.tsinghua.edu.cn, S_yinxb@163.com
{yxie, zzzhang}@cs.ecnu.edu.cn, {jfan1,shizc2}@Lenovo.com, yyqu@xmu.edu.cn

Abstract

Visible-infrared person re-identification (VI-ReID) aims to match a specific person from a gallery of images captured from non-overlapping visible and infrared cameras. Most works focus on fully supervised VI-ReID, which requires substantial cross-modality annotation that is more expensive than the annotation in single-modality. To reduce the extensive cost of annotation, we explore two practical semi-supervised settings: uni-semi-supervised (annotating only visible images) and bi-semi-supervised (annotating partially in both modalities). These two semi-supervised settings face two challenges due to the large cross-modality discrepancies and the lack of correspondence supervision between visible and infrared images. Thus, it is difficult to generate reliable pseudo-labels and learn modality-invariant features from noise pseudo-labels. In this paper, we propose a dual pseudo-label interactive self-training (DPIS) for these two semi-supervised VI-ReID. Our DPIS integrates two pseudo-labels generated by distinct models into a hybrid pseudo-label for unlabeled data. However, the hybrid pseudo-label still inevitably contains noise. To eliminate the negative effect of noise pseudo-labels, we introduce three modules: noise label penalty (NLP), noise correspondence calibration (NCC), and unreliable anchor learning (UAL). Specifically, NLP penalizes noise labels, NCC calibrates noisy correspondences, and UAL mines the hard-to-discriminate features. Extensive experimental results on SYSU-MM01 and RegDB demonstrate that our DPIS achieves impressive performance under these two semi-supervised settings.

*Equal contribution.
†Corresponding author.

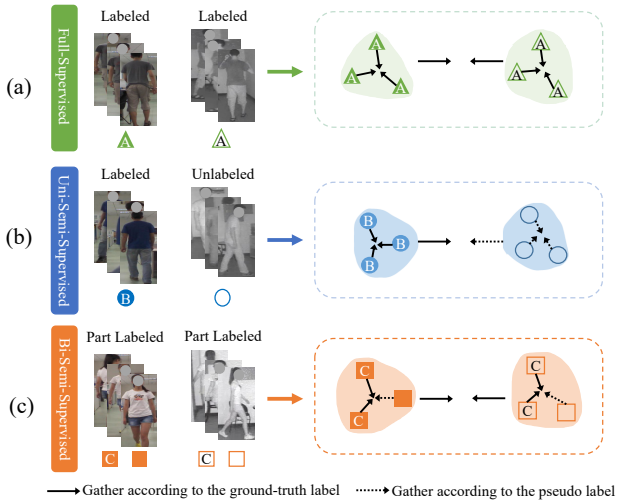


Figure 1. The figure illustrates the differences between fully supervised, uni-semi-supervised, and bi-semi-supervised learning settings, which mainly lie in the availability of labeled data.

1. Introduction

Person re-identification (ReID) aims to retrieve a person corresponding to a given query across multi-disjoint camera views[11, 43, 30]. ReID has recently gained increasing attention due to its wide range of applications in automated tracking and surveillance systems[16]. The existing ReID methods might fail to achieve encouraging results under poor illumination environments, which limits the applicability of ReID in a real-world scenario. To overcome this problem, the visible-infrared person re-identification (VI-ReID) has been proposed, which aims at retrieving infrared person images of the same identity as the given visible query and vice versa.

Though significant studies [42, 41, 37] have been made in supervised VI-ReID, they are built upon substantial labeled data. It is challenging to manually annotate every image due to the enormous number of identities. Furthermore, some images are difficult for humans to recognize, especially the color information is lost in the infrared images. For example, the scale of existing VI-ReID datasets (*e.g.*, SYSU-MM01 [36] and RegDB [25]) is relatively small due to the difficulty of annotating cross-modality images, which limits the development of VI-ReID. In this paper, we argue that it is also necessary to study a VI-ReID model trained with a small number of labels. We explore two practical semi-supervised settings: uni-semi-supervised (annotating only visible images) and bi-semi-supervised (annotating partially in both modalities). The instructions for these two settings are shown in Fig. 1 (b) and (c). Our goal is to train a VI-ReID model with limited labeled data to achieve comparable performance with the fully supervised methods.

Several pseudo-label-based methods [33, 4] involve pseudo-label generation and self-training, which show effectiveness in VI-ReID. But, all of them only focus on how to generate reliable pseudo-labels or mitigate the negative impacts of noise pseudo-labels while ignoring how to calibrate noise pseudo-labels.

In this paper, we propose a dual pseudo-label interactive self-training (DPIS) framework. Our DPIS considers both the generation of reliable pseudo-labels and the calibration of noise pseudo-labels. We employ two distinct models to generate two pseudo-labels, *i.e.*, Clustering [4] and OTLA [14]. Then, we integrate the two pseudo-labels to obtain a more accurate hybrid pseudo-label. OTLA [33] is a classifier with optimal transport, which considers both accuracy and even distribution of prediction results. This prevents most unlabeled images from being classified into a few categories, ultimately resulting in a high proportion of noise in the generated pseudo-labels. Clustering [4] is a pseudo-label refinement module designed for specific modality characteristics. Despite all this, there are inevitable noise labels in the hybrid pseudo-labels. In order to learn knowledge from noise pseudo-labels, we introduce three modules: noise label penalty (NLP), noise correspondence calibration (NCC), and unreliable anchor learning (UAL). Specifically, following the work [38], we introduce a two-component Gaussian Mixture Model (GMM) to compute confidence for each pseudo-label and divide them into reliable and unreliable anchors (these anchors will be used to construct triplet [43]). Then, we penalize the noise labels according to confidence, which eliminates the negative effects. Finally, for reliable anchors, NCC calibrates false positive samples for corresponding anchors to remove the detrimental effects of noise correspondence. For unreliable anchors, UAL utilizes the unreliable anchors that were discarded by NCC to mine hard-to-discriminate features through unsupervised contrastive learning.

To summarize, our contributions are three-fold:

- We propose a dual pseudo-label interactive self-training framework for semi-supervised visible-infrared person ReID, which leverages the intra- and inter-modality characteristics to obtain hybrid pseudo-labels for unlabeled data.
- We introduce three modules: noise label penalty (NLP), noise correspondence calibration (NCC), and unreliable anchor learning (UAL). These modules help to penalize noise labels, calibrate noisy correspondences, and exploit hard-to-discriminate features.
- We provide comprehensive evaluations under these two semi-supervised VI-ReID. Extensive experiments on two popular VI-ReID benchmarks demonstrate that our DPIS achieves impressive performance.

2. Related Work

Visible-infrared Person ReID. Visible-infrared person ReID is more challenging than conventional person ReID only focuses on the visible modality. Most existing works have attempted to narrow the discrepancy between cross-modality images for visible-infrared ReID. A number of representation learning-based methods [32, 35, 42] designed various network structures to exploit feature discrimination. Several generation learning-based methods [6, 20] attempted to excavate modality-invariant information by image generation. In addition, many metric learning-based methods [44, 23, 52] introduced various triplet losses to mine hard samples, both cross-modality and intra-modality, which significantly improved the visible-infrared ReID performance. However, the above methods all solve the visible-infrared ReID by using the labels of two modality images as the supervision information, which limits the applications of VI-ReID. In this work, we explore two semi-supervised VI-ReID that have important implications for real-world VI-ReID deployments.

Learning with Noisy Labels. Recently, many methods have been proposed to learn deep networks with noisy labels, which can be categorized into three major types. The first type performs loss adjustment, which adjusts the loss using model predictions [27, 48]. The second type tries to separate clean samples from corrupted samples and trains the model on clean samples [19, 22, 15]. The third type attempts to reweight samples, which designs automatic learning to assign weights to training samples [19, 28]. These methods focus on the classification task. Excitedly, a few studies recently extended the learning with noise labels to other fields. The works [40, 39] demonstrated that it may not be true that negative pairs in contrastive learning correlate and designed a noise-robust contrastive loss to deal

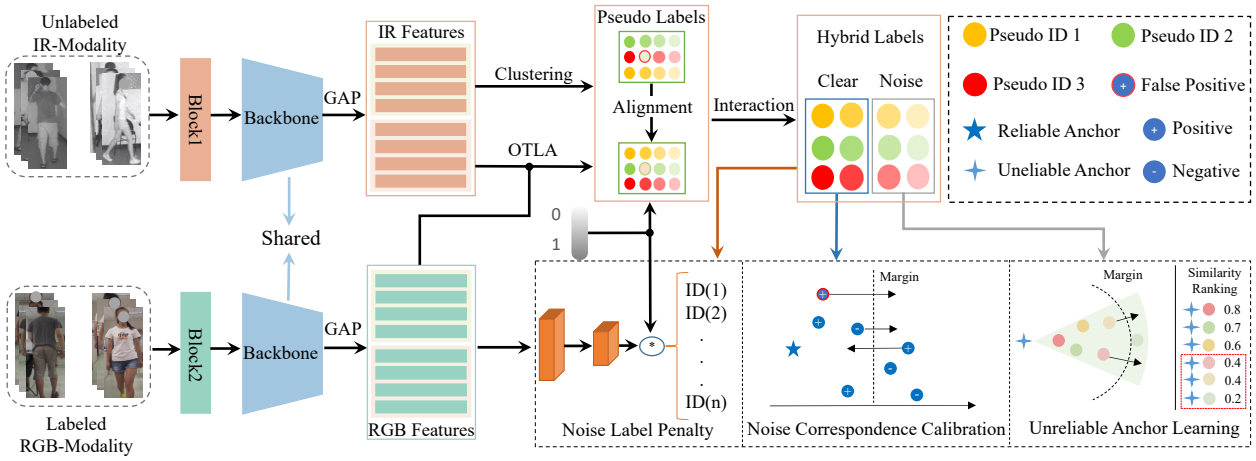


Figure 2. Framework of the proposed method. Two modality-specific block layers are utilized to capture modality-specific information for different modalities and backbone-shared layers focus on learning a multi-modality shareable space to bridge the gap between two heterogeneous modalities. Clustering is used to generate unaligned pseudo-labels of infrared (IR) images and optimal-transport label assignment (OTLA) predicts aligned pseudo-labels based on both labeled visible (RGB) and unlabeled infrared (IR) images. We align two pseudo-labels and integrate them into a hybrid label. To reduce the detrimental effects of ineluctable noise pseudo-labels, we introduce three modules: noise label penalty (NLP), noise correspondence calibration (NCC), and unreliable anchor learning (UAL). In brief, we first introduce a Gaussian Mixture Model (GMM) to compute the confidences of the generated pseudo-labels and divide all samples into reliable and unreliable anchors. Then, NLP uses these confidences to penalize the noise labels. Finally, for reliable anchors, NCC further rectifies noise correspondence to construct correct triplets, for unreliable anchors, UAL selects certain samples with the lowest similarity as negative samples for unreliable anchors.

with the problem. NCR [18] released that the correspondence of cross-modal pair may be false and proposed a novel triplet loss to handle the noisy correspondence. However, the aforementioned methods mainly focus on utilizing the high-confidence anchors to construct the triplet while abandoning the low-confidence anchors. Our DPIS also uses low-confidence anchors to construct the triplet using the paradigm of contrastive learning.

Unsupervised Domain Adaptation Person ReID. The objective of unsupervised domain adaptation (UDA) is to enhance the learning of the target domain by using the labeled source domain, whose distribution is distinct from the target domain. For the visible-infrared person ReID, this application of UDA can be viewed as an open-set category in which the classes of the two domains are inconsistent. The existing UDA methods could be divided into the following three groups: i) metric learning-based methods [45, 54, 55] tried to uncover the relationship between the source domain and the target domain in order to narrow the feature distribution discrepancy; ii) unsupervised clustering methods [10, 13, 49] have been adopted to make use of unlabeled target knowledge; iii) generation learning-based methods [8, 34, 53] would like to learn domain invariant information by mutually generating images from the source and target domains. The difference is that UDA only focuses on the performance of the target domain, while the visible-infrared person ReID is concerned with promoting the discrimination of features extracted from two modalities. In addition, the domain discrepancy between visible and infrared images is larger than UDA.

3. Methodology

In this paper, we provide a unified framework for semi-supervised person VI-ReID, which can be applied to uni-semi-supervised and bi-semi-supervised settings. We focus on the more challenging uni-semi-supervised setting and illustrate our method using this example. We propose a dual pseudo-label interactive self-training (DPIS) for semi-supervised VI-ReID as shown in Fig. 2. Our DPIS generate hybrid pseudo-label for unlabeled data. In addition, to eliminate the negative effects of noise pseudo-labels, we introduce three main components: a noise label penalty (NLP), a noise correspondence calibration (NCC), and an unreliable anchor learning (UAL). In brief, our DPIS first generates hybrid pseudo-labels and confidences for unlabeled data. Next, we divide all pseudo-labels into clear and noise labels. Subsequently, NLP utilizes the confidence to penalize the noise labels. Finally, NCC selects clear labels as anchors and rectifies noise correspondence to construct correct triplets, while UAL selects noise labels as anchors to exploit the hard-to-discriminate features in a unsupervised contrastive learning paradigm. We provide the training process of our method, as shown in Algorithm 1.

3.1. Problem Formulation

Suppose we have a cross-modality person ReID dataset. Let $V = \{v_i\}_{i=1}^{N_v}$ denote the visible images with N_v samples, and $I = \{r_i\}_{i=1}^{N_r}$ denote the infrared images with N_r samples. In the uni-semi-supervised setting, we are only able to access the labels of visible images, and hence only v_i corresponds to an identity label $y_i^v \in [1, 2, \ldots, N_p]$, where N_p is the total number of person identities. Our goal is to train a cross-modality person ReID model without the labels of infrared images.

3.2. Dual Pseudo-label Interactive Self-training Framework

Unlike the unsupervised visible ReID problem, visible-infrared heterogeneous images have large appearance variations. Therefore, clustering methods [9, 29] are not directly suitable for grouping infrared images with visible images in the semi-supervised setting. We only use infrared images for clustering to obtain pseudo-labels $\hat{y}^{cr}$ that are not aligned with visible images.

$$\hat{y}^{cr} = \{\hat{y}_i^{cr}\}_{i=1}^{N_r} = \{Cluster(r_i)\}_{i=1}^{N_r}, \tag{1}$$

where $Cluster(\cdot)$ represents the clustering method [4].

It is challenging to construct a robust VI-ReID model without aligned infrared and visible data. One possible solution is to predict pseudo-labels for infrared images using a classifier trained on labeled visible images. Unfortunately, this method leads to the degeneration of the classifier, where most infrared samples are predicted to only a few identities, ultimately resulting in a high proportion of noise in the generated pseudo-labels. Following the method [33], we employ optimal-transport label assignment (OTLA) to predict pseudo-label $\hat{y}^{or}$ for unlabeled infrared images, which formulates the pseudo-label generation as an optimal-transport problem. Specifically, given N_r suppliers (infrared images) and N_p demanders (pseudo-labels). Suppliers supply infrared images to demanders described as a vector α, and demanders own infrared images from the suppliers described as a vector β. The purpose of the optimal-transport problem is to find an optimal transport plan $P^* \in \mathbb{R}^{N_r \times N_p}$ to minimize transport cost, which satisfies the following equation:

$$\begin{aligned} P^* = \min_P \langle P, -\log(M)\rangle + \frac{1}{\tau_{otla}} KL\left(P \| \alpha\beta^T\right), \\ \text{s.t. } P\mathbb{1}_{N_r} = \mathbb{1}_{N_r} \cdot \frac{1}{N_r}, \\ P^T\mathbb{1}_{N_p} = \mathbb{1}_{N_p} \cdot \frac{1}{N_p}, \end{aligned} \tag{2}$$

where M_{ij} denotes the cost transported from supplier i to demander j, and we adopt the softmax matrix of the identity classifier trained by visible images as the cost matrix M. P_{ij} is the plan transported from supplier i to demander j. $KL(\cdot)$ and $\langle\cdot,\cdot\rangle$ are the KL-divergence and the Frobenius dot-product, respectively. $\mathbb{1}_{N_r}$ and $\mathbb{1}_{N_p}$ denote the all-one vector of N_r and N_p dimension, respectively. Following [33], hyper-parameter τ_{otla} is set to 25.

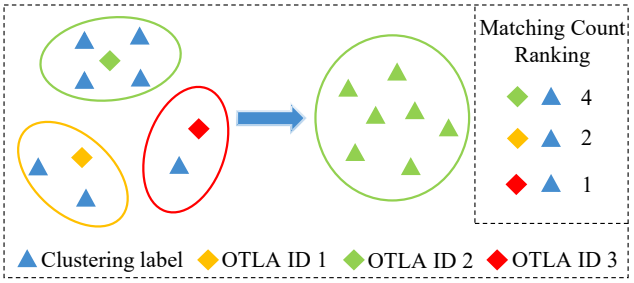


Figure 3. The alignment of matching count priority strategy. The triangles represent labels generated by clustering, and the diamonds represent pseudo-labels generated by OTLA. Since the identity of the pseudo-label is known and clustering only groups similar samples together, thus the labels can be aligned according to the matching degree of the two pseudo-labels. Specially, we align the clustered labels to pseudo-labels generated by OTLA based on the label pairs of maximum matching count.

OTLA predicts pseudo-labels for unlabeled infrared images that are not only associated with labeled visible data but also ensure the approximately even distribution of pseudo-labels. We use the iterative Sinkhorn-Knopp [7] algorithm to address the optimal transport plan P^*. We can obtain pseudo-label $\hat{y}^{or}$ of infrared images via:

$$\hat{y}^{or} = \{\hat{y}_i^{or}\}_{i=1}^{N_r} = \{Argmax\left(P_{i\cdot}^*\right)\}_{i=1}^{N_r}, \tag{3}$$

where $P_{i\cdot}^*$ denotes that the i-th row of the matrix P^*.

To obtained $\tilde{y}^{cr}$ aligned with visible data, we adopt a matching count priority scheme to align $\hat{y}^{cr}$ with $\hat{y}^{or}$ as shown in Fig. 3. However, as both two pseudo-labels are prone to containing noise, there is a risk that the same sample may have two different labels. We draw inspiration from the work [2], which has shown that deep neural networks tend to learn simple patterns before fitting noise labels. Thus, we compute the confidences of two pseudo-labels based on their classifier loss. Specifically, we introduce the two-component Gaussian Mixture Model (GMM) to fit the loss distribution from Eq. (4, 5) as follows:

$$\mathcal{L}_{cid}^i = \log P\left(\tilde{y}_i^{cr} \mid C\left(f\left(r_i\right)\right)\right), \tag{4}$$

$$\mathcal{L}_{oid}^i = \log P\left(\hat{y}_i^{or} \mid C\left(f\left(r_i\right)\right)\right), \tag{5}$$

$$p(\mathcal{L}_{id} \mid \theta) = \sum_{k=1}^{K} \pi_k \phi(\mathcal{L}_{id} \mid k), \tag{6}$$

where $f(\cdot)$ and $C(\cdot)$ are a function for extracting the features of images from different modalities and an identity

classifier, respectively. $\mathcal{L}_{id}$ is either $\mathcal{L}_{oid}$ or $\mathcal{L}_{cid}$. π_k and $\phi(\mathcal{L}_{id} \mid k)$ are the mixture coefficient and the probability density of the k-th component, respectively. We compute the posterior probability as the confidence of the i-th sample as follows:

$$w_{ci} = p\left(k \mid \mathcal{L}_{cid}^{i}\right), \tag{7}$$

$$w_{oi} = p\left(k \mid \mathcal{L}_{oid}^{i}\right), \tag{8}$$

where k and $p\left(k \mid \mathcal{L}^{i}\right)$ are the Gaussian component with smaller mean and the responsiveness of $\mathcal{L}^{i}$ at the k-th component, respectively.

We generate a hybrid pseudo-label by integrating $\hat{y}^{cr}$ and $\hat{y}^{or}$ and update the confidence as follows:

$$\hat{y}_i^r = \begin{cases} \tilde{y}_i^{cr}, w_{ci} \geq w_{oi} \\ \hat{y}_i^{or}, otherwise \end{cases} \tag{9}$$

$$\hat{w}_i = \begin{cases} 1 & , \tilde{y}_i^{cr} = \hat{y}_i^{or} \\ \max\left(w^{oi}, w^{ci}\right) & , otherwise \end{cases} \tag{10}$$

3.3. Noise Label Penalty

Benefiting from dual pseudo-label interaction, we have obtained hybrid pseudo-label $\hat{y}^r$ and corresponding confidence $\hat{w}$ for unlabeled infrared images. We apply $\mathcal{L}_{nlp}$ loss with confidence $\hat{w}$ to penalize the noise labels as follows:

$$\begin{aligned} \mathcal{L}_{nlp} = &- \frac{1}{N_r} \sum_{j=1}^{N_r} \hat{w}_j \log P\left(\hat{y}_j^r \mid C\left(f(r_j)\right)\right) \\ &- \frac{1}{N_v} \sum_{i=1}^{N_v} \log P\left(y_i^v \mid C\left(f(v_i)\right)\right). \end{aligned} \tag{11}$$

Triple loss [43] is a popular method for visible-infrared person ReID, which aims to decrease the distance between positive pairs and increase the distance between negative pairs. However, the correspondences might be false since the correspondences are constructed by resorting to the pseudo-label in semi-supervised learning. To overcome the problem, we divide pseudo-labels into a clear label set $S^c = \{r_i \mid w_i \geq \tau_{cn}\}$ and noise label set $S^n = \{r_i \mid w_i < \tau_{cn}\}$, where the threshold $\tau_{cn} = 0.5$ in our work. Let's assume that N_c clean labels and N_n noise labels are obtained. Subsequently, for the clear labels, the noise correspondence calibration module (NCC) is proposed to relieve the detrimental effects of noise correspondence. For the noise labels, the unreliable anchor learning module (UAL) is proposed to exploit the hard-to-discriminate features.

3.4. Noise Correspondence Calibration

In this module, we select clear label samples from S^c as anchors. Significantly, we refrain from selecting noise label samples as anchors due to the potential inaccuracies of their pseudo-labels. Thus, we don't choose corresponding positive and negative samples for unreliable anchors. We define the correspondences $y_{ap} = 1$ to indicate that the anchor a and the sample p have the same identity, thereby forming a positive pair. The correspondences $y_{an} = 0$ denote that the anchor a and the sample n have different identities and form a negative pair. NCC is designed for calibrating false correspondences. Specifically, if sample j comes from S^n, NCC will calibrate the positive pair to the negative pair. Thanks to the noise correspondence calibration, we can directly use corrected correspondences to train the visible-infrared person ReID model with $\mathcal{L}_{ncc}$ loss that is similar to the triplet loss[43]. The loss $\mathcal{L}_{ncc}$ is defined as minimizing the distance between the feature representations of the anchor and positive samples while increasing the distance between the anchor and negative samples to a certain margin, as follows:

$$\begin{aligned} \mathcal{L}_{\text{ncc}} = \sum_{i=1}^{N_c} \Big[&\left\| f\left(x_i^a\right) - f\left(x_i^p\right) \right\|_2^2 - \\ &\left\| f\left(x_i^a\right) - f\left(x_i^n\right) \right\|_2^2 + \tau_{tri} \Big]_+, \end{aligned} \tag{12}$$

where x_i^a represents anchor. x_i^p and x_i^n are positive and negative samples of anchor x_i^a, respectively. x can represent infrared images or visible images, *i.e.*, $x \in [v, r]$. $[\cdot]_+ = \max(\cdot, 0)$ and the margin τ_{tri} is a hyper-parameters.

Algorithm 1: Training Process of DPIS

Input: Initialized model parameters W, unlabeled infrared data I and labeled visible data V
Output: Optimized model parameters W

for *epoch=1:max_epoch* **do**
  $\hat{y}^{or} \leftarrow OTLA(V, I)$ and $\hat{y}^{cr} \leftarrow Cluster(I)$;
  Generate w_c and w_o using Eq. (7) and Eq. (8);
  Generate hybrid pseudo-labels $\hat{y}^r$ using Eq. (9);
  Update the confidence $\hat{w}_i$ using Eq. (10);
  $S^c = \{r_i \mid w_i \geq 0.5\}$, $S^n = \{r_i \mid w_i < 0.5\}$;
  for *batch=1:max_batch* **do**
    Calculate $\mathcal{L}_{nlp}$ using Eq. (11);
    if $anchor \in S^c$ **then**
      Calculate $\mathcal{L}_{ncc}$ using Eq. (12);
    else
      Calculate $\mathcal{L}_{ual}$ using Eq. (13);
    end
    $\mathcal{L}_{total} = \mathcal{L}_{nlp} + \mathcal{L}_{ncc} + \mathcal{L}_{ual}$;
    Optimize the model parameters W by $\mathcal{L}_{total}$.
  end
end

3.5. Unreliable Anchor Learning

The noise label samples cannot serve as anchors in NCC. However, abandoning all unreliable anchors is not always feasible, as it may limit the applicability of VI-ReID in challenging scenarios. In this paper, we argue that the unreliable label samples from S^n also act as anchors. We design an unreliable anchor learning module (UAL) to reasonably utilize the unreliable anchors abandoned by NCC. We regard UAL as unsupervised contrastive learning, which encourages the model to exploit hard-to-discriminate features. Contrastive learning (CL) has made significant progress in unsupervised learning. Unfortunately, unsupervised CL can't be directly applied to person VI-ReID since conventional CL defines samples from different instances as negative samples that contain the same identity with the corresponding anchor.

To adapt CL to VI-ReID, UAL improves the selection of negative samples for unreliable anchors. In detail, we first compute the similarity between mini-batch samples and unreliable anchors. Then, we rank the similarity and observe that it may be difficult to determine all negative samples, but we can select a proportion of negative samples. Based on this, we select τ_{sam} visible and infrared samples with the lowest similarity as the visible negative samples and the infrared negative samples for unreliable anchors, where τ_{sam} is a hyper-parameter that determines the number of negative samples to be selected. For positive samples, UAL is similar to most contrastive learning methods, we just take the argument of the anchor sample as the positive sample.

The $\mathcal{L}_{ual}$ loss is to mine the hard-to-discriminate features by utilizing unreliable anchors as follows:

$$\mathcal{L}_{\text{ual}} = \sum_{i=1}^{N_n} \sum_{j=1}^{\tau_{sam}} \Big[\left\| f\left(r_i^{ua}\right) - f\left(x_i^{up}\right) \right\|_2^2 - \left\| f\left(r_i^{ua}\right) - f\left(x_{ij}^{un}\right) \right\|_2^2 + \tau_{tri} \Big]_+, \tag{13}$$

where r_i^{ua} represents unreliable anchor from S^n. x_i^{up} and x_i^{un} are positive and negative samples of unreliable anchor r_i^{ua}, respectively.

The total loss $\mathcal{L}_{total}$ of proposed method is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{nlp} + \mathcal{L}_{ncc} + \mathcal{L}_{ual}. \tag{14}$$

4. Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness of our method in terms of two semi-supervised settings. Actually, the uni-semi-supervised learning (USSL) setting is essentially a special unsupervised domain adaptation (UDA) problem, where we only provide labels for visible images. Therefore, we compare our approach with several state-of-the-art methods in different label-efficient VI-ReID, *i.e.*, full-supervised learning (SL), unsupervised domain adaptation (UDA), uni-semi-supervised learning (USSL), and bi-semi-supervised learning (BSSL).

4.1. Experimental Setting

Dataset. We evaluate our DPIS on two public datasets **SYSU-MM01** [36] and **RegDB** [25]. **SYSU-MM01** is a challenging visible-infrared ReID dataset. It is collected from six camera views (four visible and two infrared), including both indoor and outdoor scenes. This dataset contains 287,628 visible images and 15,792 infrared images with 491 identities. Among them, 22,258 visible images and 11,909 infrared images with 395 identities are used for training. 3,803 infrared images are used for query and 301 visible images are randomly selected to make up the gallery set. **RegDB** contains 412 identities, each of whom has 10 visible images and 10 infrared images. RegDB is randomly divided, with half for training and the remaining for testing. Note that on both two datasets, only the ground-truth labels of visible images are utilized for training in the uni-semi-supervised setting and we can utilize labels of different rates for each identity on both modalities in the bi-semi-supervised setting.

Evaluation Protocols. Cumulative matching characteristics (CMC) [44] and mean average precision (mAP) are adopted as evaluation metrics. For fair comparisons, we report the results of all-search mode and indoor-search mode with the official code on SYSU-MM01. We also report the results on RegDB by randomly splitting the training and testing sets 10 times and reporting the average results.

4.2. Implementation Details

Our proposed framework is implemented in PyTorch. The settings of two modality-specific block layers follow AGW [43], and we adopt ResNet-50 as shared-backbone to extract 2048d features, which is initialized with the ImageNet pre-trained weights. In the training stage, training images are resized as 256×128 and random horizontal flipping is used for data augmentation[41]. The total number of training epochs is 80. At each training step, we randomly sample 8 identities, of which 4 visible and 4 infrared images are selected to formulate a batch. To address the issue of error accumulation in self-training, we adopt cross-branch self-training as previous methods[21, 3].

4.3. Results and Analysis

We compare our DPIS with several state-of-the-art methods under three settings, *i.e.*, uni-semi-supervised (USSL) full-supervised learning (SL), unsupervised domain adaptation (UDA). The quantitative results are shown in Table 1. To further evaluate the generality of our DPIS, we conduct extensive experiments in bi-semi-supervised setting (BSSL) in which the labeled ratio varies from 10%,

Table 1. Comparisons with state-of-the-art methods in different label-efficient VI-ReID on SYSU-MM01 and RegDB, *i.e.*, full-supervised learning (SL), unsupervised domain adaptation (UDA), and uni-semi-supervised learning (USSL). All methods are measured by Rank-1 (%) and mAP (%). Methods marked by † denote re-implementations based on public code.

Settings			SYSU-MM01				RegDB			
			All Search		Indoor Search		Visible2Thermal		Thermal2Visible	
Type	Method	Venue	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
SL	JSIA-ReID [31]	AAAI'20	38.1	36.9	43.8	52.9	48.5	49.3	48.1	48.9
	DDAG [42]	ECCV'20	54.8	53.0	61.0	68.0	69.4	63.5	68.1	61.8
	AGW [43]	TRAMI'21	47.5	47.7	54.2	63.0	70.1	66.4	70.5	65.9
	NFS [5]	CVPR'21	56.9	55.5	62.8	69.8	80.5	72.1	78.0	69.8
	LbA [26]	ICCV'21	55.4	54.1	58.5	66.3	74.2	67.6	72.4	65.5
	CAJ [41]	ICCV'21	69.9	66.9	76.3	80.4	85.0	79.1	84.8	77.8
	MPANet [37]	CVPR'21	70.6	68.2	76.7	81.0	83.7	80.9	82.8	80.7
	DART [38]	CVPR'22	68.7	66.2	72.5	78.2	82.0	73.8	83.6	75.7
	FMCNet [47]	CVPR'22	66.3	62.5	68.2	74.1	89.1	84.4	88.4	83.9
	MID [17]	AAAI'22	60.3	59.4	64.9	70.1	87.5	84.9	84.3	81.4
	LUPI [1]	ECCV'22	71.1	67.6	82.4	82.7	88.0	82.7	86,8	81.3
	DPIS(Ours)	-	68.2	65.7	71.7	77.0	86.3	77.4	82.1	74.7
UDA	MEB-Net† [46]	ECCV'20	7.3	6.9	20.4	11.7	5.6	6.9	14.9	14.0
	D-MMD† [50]	ECCV'20	12.5	10.4	19.0	15.4	2.2	3.7	2.0	3.6
	MMT† [12]	ICLR'20	13.9	8.7	20.7	15.9	7.3	7.6	16.9	14.9
	SpCL(UDA)† [13]	NIPS'20	15.1	6.5	19.5	12.1	3.3	4.3	8.4	9.5
	GLT† [51]	CVPR'21	7.7	9.5	12.1	18.0	2.9	4.5	6.3	7.6
USSL	MAUM-50 [24]	CVPR'22	28.8	36.1	-	-	-	-	-	-
	MAUM-100 [24]	CVPR'22	38.5	39.2	-	-	-	-	-	-
	OTLA [33]	ECCV'22	48.2	43.9	47.4	56.8	49.9	41.8	49.6	42.8
	DPIS(Ours)	-	**58.4**	**55.6**	**63.0**	**70.0**	**62.3**	**53.2**	**61.5**	**52.7**

25% to 50%, and the quantitative results are shown in Table 2. If not specified, we conduct analysis experiments on SYSU-MM01 in the single-shot & all-search mode.

Comparison with Fully-supervised Methods. Surprisingly, our DPIS outperforms several fully-supervised methods [31, 42, 43, 5, 26] on SYSU-MM01 dataset. The result demonstrates that our DPIS can exploit information from unlabeled infrared images. In addition, we have to acknowledge that there is still a large gap between our DPIS and and many state-of-the-art methods.

Comparison with Unsupervised Domain Adaptation Methods. We compare our DPIS with five state-of-the-art UDA methods under the uni-semi-supervised setting, and UDA results are disappointing. Although marginal improvement has been achieved, it still indicates existing UDA methods cannot help the model alleviate the modality discrepancy. Our DPIS is able to help the model alleviate the modality discrepancy and achieve superior performances.

Comparison with Uni-semi-supervised Methods. OTLA [33] is the only one we know of that has the same experimental setting as ours. However, OTLA was trained in a simple self-training manner, which would accumulate the model's errors. In addition, OTLA focused on the generation of pseudo-labels, while ignoring the inevitable calibration of noise labels. Fig. 4 shows the limitation of OTLA in pseudo-label accuracy. Compared with OTLA, our DPIS improves Rank-1 accuracy and mAP by 10.2% and 11.7% on SYSU-MM01 (all-search mode), respectively. MAUM-50 and MAUM-100 indicate that MAUM only uses 50 and 100 IR identities to train the VI-ReID model. Our DPIS requires little data annotation, and the performance is better than MAUM-50 and MAUM-100.

Comparison with Bi-semi-supervised Methods. Since our work is to first explore the bi-semi-supervised setting for VI-ReID, we compare our DPIS with five fully-supervised methods under the same bi-semi-supervised setting. We re-implement the results using the official code. As shown in Table 2, our model achieves 85.58% in Rank-1 and 76.73% in mAP under Visible2thermal modes on RegDB with 50% labels. Obviously, the results show our DPIS achieves impressive performance and outperforms other state-of-the-art methods in the same setting.

The above results demonstrate our DPIS not only achieves outstanding performance in the uni-semi-supervised setting but also achieves encouraging results in the bi-semi-supervised setting.

4.4. Ablation Study

We conduct ablation experiments to evaluate the contribution of each component. We train a model only supervised by labels of visible images as our baseline. The ablation experiment is conducted on SYSU-MM01 under the

Table 2. Comparisons with five advanced methods on SYSU-MM01 and RegDB under the bi-semi-supervised setting, and all methods are measured by Rank-1 (%) and mAP (%). All results are reimplemented based on public code. The 1^{st} and 2^{nd} best results are indicated by the red and blue colors, respectively.

Settings				SYSU-MM01				RegDB			
				All Search		Indoor Search		Visible2thermal		Thermal2visible	
Rate	Method	Venue	Year	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
10%	DDAG [42]	ECCV	2020	29.75	27.93	33.56	42.11	44.03	38.64	43.98	37.83
	AGW [43]	TPAMI	2021	32.43	33.41	37.97	49.22	22.82	18.34	20.85	19.35
	CAJ [41]	ICCV	2021	35.55	37.16	38.13	47.06	21.31	17.13	21.24	17.06
	LbA [26]	ICCV	2021	30.31	29.91	32.91	43.17	41.70	36.70	42.09	36.91
	DART [38]	CVPR	2022	37.90	35.44	40.78	49.13	40.72	33.42	37.66	31.55
	DPIS(Ours)	-	-	57.70	55.00	62.86	69.84	54.95	50.79	55.05	49.64
25%	DDAG [42]	ECCV	2020	39.13	36.99	43.52	51.73	52.48	48.09	52.86	47.89
	AGW [43]	TPAMI	2021	36.88	37.69	43.08	54.34	57.72	49.83	57.14	48.99
	CAJ [41]	ICCV	2021	48.82	48.02	57.06	64.70	52.57	43.78	48.83	40.90
	LbA [26]	ICCV	2021	49.37	47.42	54.04	62.58	51.31	43.74	47.52	43.74
	DART [38]	CVPR	2022	50.22	47.93	53.55	60.98	70.97	59.60	70.24	61.09
	DPIS(Ours)	-	-	65.20	61.75	70.50	75.55	77.28	67.74	75.34	66.27
50%	DDAG [42]	ECCV	2020	42.26	38.66	48.96	55.50	59.37	53.22	57.04	51.23
	AGW [43]	TPAMI	2021	43.99	40.63	48.64	53.80	70.78	62.71	68.20	61.41
	CAJ [41]	ICCV	2021	59.37	56.99	62.58	64.26	77.43	66.75	74.89	63.05
	LbA [26]	ICCV	2021	53.36	52.18	58.91	62.64	67.62	61.73	65.78	60.04
	DART [38]	CVPR	2022	56.92	54.43	61.25	68.27	80.63	69.79	80.34	70.98
	DPIS(Ours)	-	-	67.89	64.16	71.78	76.60	85.58	76.73	81.41	74.13

Table 3. Ablation studies on SYSU-MM01 (all-search & single-shot mode) under the uni-semi-supervised setting. Here, R indicates Rank.

Method					SYSU-MM01			
					All Search			
Baseline	DPI	NLP	NCC	UAL	R-1	R-10	R-20	mAP
✓					3.80	23.15	38.02	6.31
✓	✓				36.56	75.40	86.05	35.61
✓	✓	✓			51.90	87.70	93.87	49.38
✓	✓	✓		✓	53.03	87.93	93.78	50.07
✓	✓	✓	✓		54.24	88.40	94.59	51.83
✓	✓	✓	✓	✓	58.40	90.32	95.72	55.57

uni-semi-supervised setting. The results are summarized in Table 3, each component is revealed and plays an important role in our DPIS. DPI indicates we rain the baseline method only with hybrid pseudo-labels. Even achieving good performance using NLP, our NCC and UAL still can gain $1\% \sim 2\%$ improvement. When both NCC and UAL are utilized, greater improvements can be achieved. This result shows that NCC and UAL can complement each other, which is consistent with our motivation.

4.5. Pseudo-label Analysis

We conduct an analysis experiment to evaluate the accuracy of pseudo-labels. The results are summarized in Fig. 4. Compared with OTLA [33], ours has a better convergence value, which indicates that we provide a better pseudo-label constraint strategy.

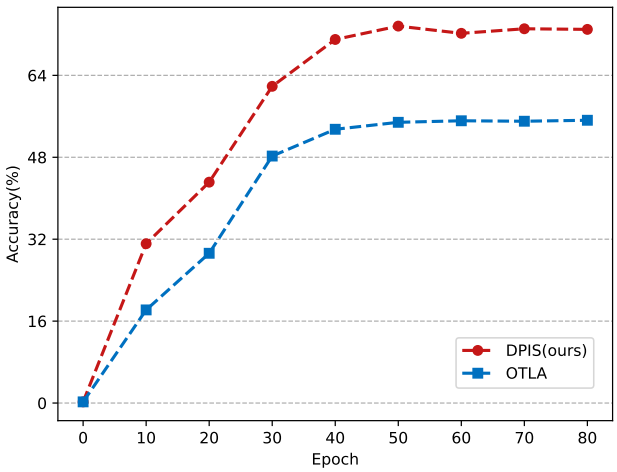
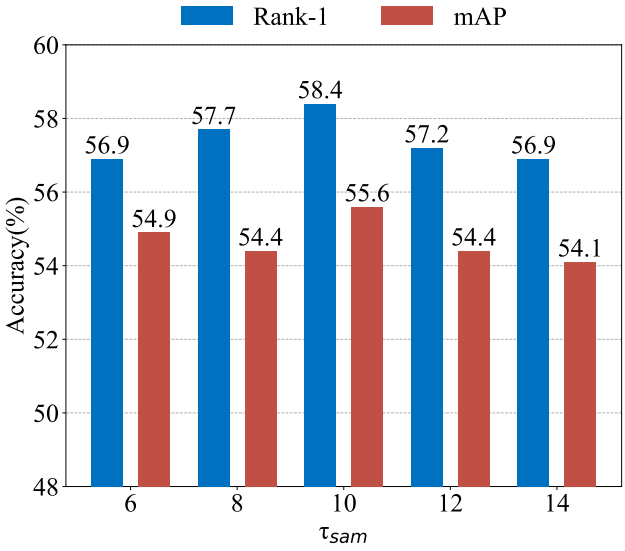


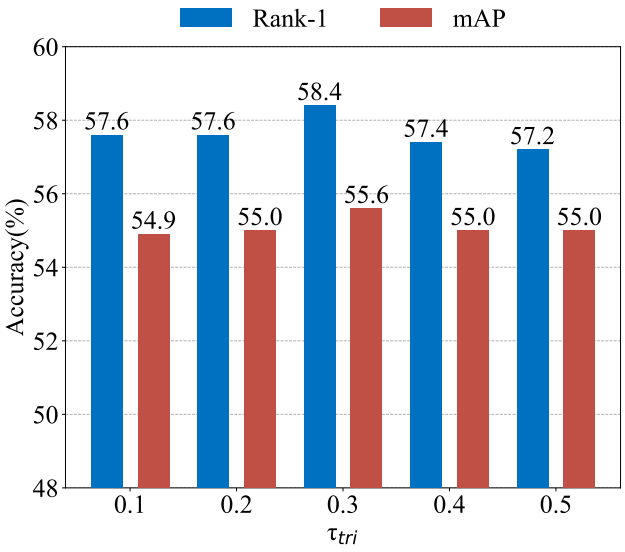
Figure 4. The comparison of pseudo-label accuracy between our DPIS and OTLA on the SYSU-MM01 dataset under the uni-semi-supervised setting. The x-axis represents the training epochs, and the y-axis reflects the accuracy.

4.6. Analysis of Hyper-parameters

We analyzed hyper-parameters under the condition that only one hyper-parameter is selected as the variable, while all other hyper-parameters are kept constant. The Fig. 5 experimentally analyze the impact of hyper-parameters τ_{sam} and τ_{tri} on SYSU-MM01. In Fig. 5 (a), we adjust hyper-parameter τ_{sam} in unreliable anchor learning loss $\mathcal{L}_{ual}$ from 6 to 14. We observe that our DPIS achieves better perfor-



(a) Hyper-paramete τ_{sam}



(b) Hyper-paramete τ_{tri}

Figure 5. Hyper-parameters analysis on SYSU-MM01 dataset.

mance when the τ_{sam} a raising and we achieve the best performance when τ_{sam} was set to 10. In Fig. 5 (b), we adjusted τ_{tri} from 0.1 to 0.5 and found that our DPIS achieved the best performance when τ_{tri} was set to 0.3.

5. Conclusions

In this paper, we discuss two novel semi-supervised settings for VI-ReID and propose a dual pseudo-label interactive self-training for these two semi-supervised settings. In addition, we introduce three modules: noise label penalty (NLP), noise correspondence calibration (NCC), and unreliable anchor learning (UAL) to eliminate the negative effects brought by the noise pseudo-labels. Our DPIS achieves promising performance, *i.e.*, 58.40% in terms of Rank-1 accuracy on SYSU-MM01 (all-search mode) under the uni-semi-supervised setting. The results are equally excellent under the bi-semi-supervised setting, our DPIS achieves 85.58% in Rank-1 and 76.73% in mAP with 50% labels on RegDB (visible-to-thermal mode).

6. Acknowledgments

This work is supported by the National Key Research and Development Program of China (No. 2020AAA0108301), National Natural Science Foundation of China (No. 62176224, 62222602, 62106075, 62176092), Natural Science Foundation of Shanghai (23ZR1420400), Natural Science Foundation of Chongqing (CSTB2023NSCQ-JQX0007), China Postdoctoral Science Foundation (No. 2023M731957), CCF-Lenovo Blue Ocean Research Fund and CAAI-Huawei MindSpore Open Fund.

References

[1] Mahdi Alehdaghi, Arthur Josi, Rafael M. O. Cruz, and Eric Granger. Visible-infrared person re-identification using privileged intermediate information. In *ECCV*, pages 720–737, 2022. 7

[2] Devansh Arpit, Stanislaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron C. Courville, Yoshua Bengio, and Simon Lacoste-Julien. A closer look at memorization in deep networks. In *ICML*, pages 233–242, 2017. 4

[3] David Berthelot, Nicholas Carlini, Ian J. Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *NIPS*, pages 5050–5060, 2019. 6

[4] Jun Chen Bin Yang, Mang Ye and Zesen Wu. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification. In *ACM MM*, 2022. 2, 4

[5] Yehansen Chen, Lin Wan, Zhihang Li, Qianyan Jing, and Zongyuan Sun. Neural feature search for rgb-infrared person re-identification. In *CVPR*, pages 587–597, 2021. 7

[6] Seokeon Choi, Sumin Lee, Youngeun Kim, Taekyung Kim, and Changick Kim. Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification. In *CVPR*, pages 10254–10263, 2020. 2

[7] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NeurIPS*, pages 2292–2300, 2013. 4

[8] Weijian Deng, Liang Zheng, Qixiang Ye, Yi Yang, and Jianbin Jiao. Similarity-preserving image-image domain adaptation for person re-identification. *CoRR*, abs/1811.10551, 2018. 3

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996. 4

[10] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *ICCV*, pages 6111–6120, 2019. 3

[11] Wenhang Ge, Chunyan Pan, Ancong Wu, Hongwei Zheng, and Wei-Shi Zheng. Cross-camera feature prediction for intra-camera supervised person re-identification across distant scenes. In *ACM MM*, pages 3644–3653, 2021. 1

[12] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *ICLR*, 2020. 7

[13] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In *NeurIPS*, 2020. 3, 7

[14] Zheng Ge, Songtao Liu, Zeming Li, Osamu Yoshie, and Jian Sun. OTA: optimal transport assignment for object detection. In *CVPR*, pages 303–312, 2021. 2

[15] Chunming He, Kai Li, Yachao Zhang, Yulun Zhang, Zhenhua Guo, Xiu Li, Martin Danelljan, and Fisher Yu. Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. *arXiv preprint arXiv:2308.03166*, 2023. 2

[16] Fabian Caba Heilbron, Victor Escorcia, Bernard Ghanem, and Juan Carlos Niebles. Activitynet: A large-scale video benchmark for human activity understanding. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 961–970, 2015. 1

[17] Zhipeng Huang, Jiawei Liu, Liang Li, Kecheng Zheng, and Zheng-Jun Zha. Modality-adaptive mixup and invariant decomposition for rgb-infrared person re-identification. In *AAAI*, pages 1034–1042, 2022. 7

[18] Zhenyu Huang, Guocheng Niu, Xiao Liu, Wenbiao Ding, Xinyan Xiao, Hua Wu, and Xi Peng. Learning with noisy correspondence for cross-modal matching. In *NeurIPS*, pages 29406–29419, 2021. 3

[19] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *ICML*, pages 2309–2318, 2018. 2

[20] Diangang Li, Xing Wei, Xiaopeng Hong, and Yihong Gong. Infrared-visible cross-modal person re-identification with an X modality. In *AAAI*, pages 4610–4617, 2020. 2

[21] Junnan Li, Richard Socher, and Steven C. H. Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *ICLR*, 2020. 6

[22] Junnan Li, Pan Zhou, Caiming Xiong, and Steven C. H. Hoi. Prototypical contrastive learning of unsupervised representations. In *ICLR*, 2021. 2

[23] Haijun Liu, Xiaoheng Tan, and Xichuan Zhou. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification. *IEEE Trans. Multim.*, pages 4414–4425, 2021. 2

[24] Jialun Liu, Yifan Sun, Feng Zhu, Hongbin Pei, Yi Yang, and Wenhui Li. Learning memory-augmented unidirectional metrics for cross-modality person re-identification. In *CVPR*, pages 19344–19353, 2022. 7

[25] Dat Tien Nguyen, Hyung Gil Hong, Ki-Wan Kim, and Kang Ryoung Park. Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*, page 605, 2017. 2, 6

[26] Hyunjong Park, Sanghoon Lee, Junghyup Lee, and Bumsub Ham. Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences. In *ICCV*, pages 12026–12035, 2021. 7, 8

[27] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *CVPR*, pages 2233–2241, 2017. 2

[28] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *ICML*, pages 4331–4340, 2018. 2

[29] Shokri Z. Selim and M. A. Ismail. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 81–87, 1984. 4

[30] 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 *ECCV*, pages 501–518, 2018. 1

[31] Guan'an Wang, Yang Yang, Tianzhu Zhang, Jian Cheng, Zengguang Hou, Prayag Tiwari, and Hari Mohan Pandey. Cross-modality paired-images generation and augmentation for rgb-infrared person re-identification. *Neural Networks*, pages 294–304, 2020. 7

[32] Guan'an Wang, Tianzhu Zhang, Jian Cheng, Si Liu, Yang Yang, and Zengguang Hou. Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment. In *ICCV*, pages 3622–3631, 2019. 2

[33] Jiangming Wang, Zhizhong Zhang, Mingang Chen, Yi Zhang, Cong Wang, Bin Sheng, Yanyun Qu, and Yuan Xie. Optimal transport for label-efficient visible-infrared person re-identification. In *ECCV*, pages 93–109, 2022. 2, 4, 7, 8

[34] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer GAN to bridge domain gap for person re-identification. In *CVPR*, pages 79–88, 2018. 3

[35] Xing Wei, Diangang Li, Xiaopeng Hong, Wei Ke, and Yihong Gong. Co-attentive lifting for infrared-visible person re-identification. In *ACM*, pages 1028–1037, 2020. 2

[36] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *ICCV*, pages 5390–5399, 2017. 2, 6

[37] Qiong Wu, Pingyang Dai, Jie Chen, Chia-Wen Lin, Yongjian Wu, Feiyue Huang, Bineng Zhong, and Rongrong Ji. Discover cross-modality nuances for visible-infrared person re-identification. In *CVPR*, pages 4330–4339, 2021. 2, 7

[38] Mouxing Yang, Zhenyu Huang, Peng Hu, Taihao Li, Jiancheng Lv, and Xi Peng. Learning with twin noisy labels for visible-infrared person re-identification. In *CVPR*, pages 14288–14297, 2022. 2, 7, 8

[39] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1055–1069, 2023. 2

[40] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *CVPR*, pages 1134–1143, 2021. 2

[41] Mang Ye, Weijian Ruan, Bo Du, and Mike Zheng Shou. Channel augmented joint learning for visible-infrared recognition. In *ICCV*, pages 13547–13556, 2021. 2, 6, 7, 8

[42] Mang Ye, Jianbing Shen, David J. Crandall, Ling Shao, and Jiebo Luo. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification. In *ECCV*, pages 229–247, 2020. 2, 7, 8

[43] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 2872–2893, 2022. 1, 2, 5, 6, 7, 8

[44] Mang Ye, Zheng Wang, Xiangyuan Lan, and Pong C. Yuen. Visible thermal person re-identification via dual-constrained top-ranking. In *IJCAI*, pages 1092–1099, 2018. 2, 6

[45] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In *CVPR*, pages 2148–2157, 2019. 3

[46] Yunpeng Zhai, Qixiang Ye, Shijian Lu, Mengxi Jia, Rongrong Ji, and Yonghong Tian. Multiple expert brainstorming for domain adaptive person re-identification. In *ECCV*, pages 594–611, 2020. 7

[47] Qiang Zhang, Changzhou Lai, Jianan Liu, Nianchang Huang, and Jungong Han. Fmcnet: Feature-level modality compensation for visible-infrared person re-identification. In *CVPR*, pages 7339–7348, 2022. 7

[48] Yachao Zhang, Yanyun Qu, Yuan Xie, Zonghao Li, Shanshan Zheng, and Cuihua Li. Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In *ICCV*, pages 15520–15528, 2021. 2

[49] Yachao Zhang, Yuan Xie, Cuihua Li, Zongze Wu, and Yanyun Qu. Learning all-in collaborative multiview binary representation for clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 3

[50] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *ECCV*, pages 526–544, 2020. 7

[51] Kecheng Zheng, Wu Liu, Lingxiao He, Tao Mei, Jiebo Luo, and Zheng-Jun Zha. Group-aware label transfer for domain adaptive person re-identification. In *CVPR*, pages 5310–5319, 2021. 7

[52] Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, Yi Yang, and Qi Tian. Person re-identification in the wild. In *CVPR*, pages 3346–3355, 2017. 2

[53] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero- and homogeneously. In *ECCV*, pages 176–192, 2018. 3

[54] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR*, pages 598–607, 2019. 3

[55] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Learning to adapt invariance in memory for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 2723–2738, 2021. 3