*Research Article*

# An Easy Way to Deploy the Re-ID System on Intelligent City: Cross-Domain Few-Shot Person Reidentification

**Xili Dai ⓘ,[1] Cheng Yang,[1] Bin Liu,[2] Haigang Gong,[1] and Xiaojun Yuan[1]**

[1]*University of Electronic Science and Technology of China (UESTC), Chengdu, China*
[2]*Southwestern University of Finance and Economics, Chengdu, China*

Correspondence should be addressed to Xili Dai; daixili_cs@163.com

An intelligent urban system relies on different types of electronic and/or sensor technologies to collect data to facilitate education, security, healthcare, etc. Person reidentification (re-ID) plays a crucial role in intelligent security, in which significant progress has been made during the past few years. One example is using re-ID systems for law enforcement tasks such as suspect identification. One common obstacle is quickly deploying a re-ID system to new city, such as data label deficiency. For example, the lacking of enough labeled data to train an excellent model in a new city, only relying on a tiny amount of criminals' pictures provided by witnesses. Fortunately, this can be modeled as a special application in unsupervised person re-ID of real-world scenarios, the study of which has become more prevalent in the re-ID community in recent years. In this paper, we first formulate our scenario as a cross-domain few-shot problem and discuss the difference between conventional supervised re-ID and unsupervised re-ID. Then, we introduce a reweighting instance method based on influence function (ReWIF) to guide the training procedure of the re-ID model. This method is motivated by the influence function, and we use two-step optimization to avoid the computation of Hessian matrices. We evaluate our proposed method on public datasets, including Market, Duke, and CUHK. Extensive experimental results show that our method can effectively address the domain bias of different datasets and the absence of labeled data on the target dataset, achieving state-of-the-art performance.

## 1. Introduction

Intelligent security is a fundamental function and need in smart and intelligent city. Person reidentification (re-ID) plays an important role in intelligent security city, especially in law enforcement areas. With the help of re-ID system, the process of suspect identification has been improved significantly. A re-ID-based video monitoring auxiliary system can greatly improve the efficiency of law enforcement tasks. However, this auxiliary system must be customized for different cities due to different illumination condition, different viewpoint, occlusion on cameras, etc. These factors make the re-ID system expensive since we need substantial pairwise labeled data across every pair of camera views. Furthermore, large-scale applications are limited as the vast majority of the data samples are unlabeled due to the prohibitive manual efforts in exhaustively labeling the

pairwise re-ID data. To address the scalability problem, some recent works focus on unsupervised re-ID by clustering on the target unlabeled data [1] or transferring the knowledge from other labeled source dataset [2, 3]. However, the performance is still not satisfactory compared to the supervised setting. Should the scalability problem of re-ID system be formulated as an unsupervised problem?

Assume that city *A* has a person re-ID system trained by labeled data sampling from city *A*, while a crime occurs in city *B* and the police obtain only a few image captures of the suspect. The city *B*'s police hope to use a person re-ID system to help them target this suspect. However, city *B* does not have a person re-ID system because of the lack of labeled data or infrastructure. So, they attempt to develop upon the re-ID system of city *A*. However, city *A* and city *B* have different illumination conditions and camera systems. The re-ID system borrowed from city *A* may not match the

requirements of city *B*. So, we expect to fine-tune the person re-ID system, relying on a few pictures of the suspect provided by the witnesses, which is a cross-domain few-shot person re-ID problem. Such scenario is a common deep learning problem and is easier than unsupervised re-ID setting since some information is accessible from the target dataset.

Fortunately, this cross-domain few-shot person re-ID problem can be regarded as a special case of the unsupervised person re-ID problem, and lots of methods about unsupervised person re-ID [4–9] can be adapted to our problem easily. However, the performance is not good and sometimes even worse since there is no mechanism to utilize the labeled few-shot data on the target dataset[2]. Another two works [10, 11] proposed small sample training methods, which can be applied to our problem directly. Unfortunately, they do not utilize the source dataset. In this paper, we first give a formal definition of the cross-domain few-shot person re-ID problem, which is a new task in person re-ID community. And then discuss the differences to other person re-ID settings. Motivated by [12], we finally propose a re-weighting instance method based on influence function (ReWIF), in which a statistical model is employed as the influence function to measure how a sample from a source domain influences the performance on the target domain. Moreover, ReWIF can be regarded as a module of most unsupervised person re-ID methods to train the pretrained model on source labeled dataset.

In this paper, we study the unsupervised person re-ID problem and relax it to a concrete and common scenario. Then, a formal definition is given to define the scenario. After that, we propose an influence function based method to solve the problem.

The contributions of this work are three-folds:

(i) The lack of generalizability for re-ID algorithm is the main issue for the scalability of re-ID system deployment. In person re-ID community, people formulate the scalability problem as purely unsupervised problem. In this paper, we find a new common application scenario for the scalability of re-ID system deployment. We formally define it as a cross-domain few-shot learning problem, which is a new task in person re-ID community. And then discuss the differences to other people's re-ID settings (see Table 1).

(ii) The new scenario can offer few-labeled data on the target dataset. In order to make full use of this benefit, we propose a method called ReWIF, which is based on an influence function with statistical analysis. It is a reweighting method on the loss function during the training. Besides that, it also can be regarded as a module of most unsupervised person re-ID methods to train the pretrained model on source labeled dataset.

(iii) We simulate the scenario and evaluate our proposed method on public datasets between Market and DuckMTMC. Extensive experimental results show

that the proposed method achieves state-of-the-art results.

## 2. Problem Formulation

With the advancement of urban infrastructure, more and more security cameras are adopted in modern cities [13]. However, deploying a re-ID system in a new city is expensive both in terms of time and financial costs, since the re-ID system usually requires customization. Not all cities have sufficient funds to cover the high cost. Thus, how to deploy a re-ID system in a new city quickly and cheaply is an emergent and meaningful task. This work considers a scenario in which we can deploy a re-ID system in a new city without large amounts of local labeled data as we mentioned before. In this paper, we use the following definitions to formulate the few-shot person re-ID problem.

*Definition 1.* Given a target testing dataset $D_t = \{x_i^t\}_{i=1}^N$, it consists of a huge amount of human images capture from the city B's camera system. A support dataset $D_r = \{(x_i^r, y_i^r)\}_{i=1}^M$ of $M$ suspects' images from the target dataset, where $M << N$, and $M$ is not a fixed constant (because we cannot guarantee the condition that each suspect appears in surveillance for the same number of times). In addition, we have an auxiliary source training dataset $D = \{D_1, D_2, \dots D_p\}$ borrowed from city $A$, where $p$ is the number of auxiliary datasets. In our scenario, each tested personnel image $x_i^r$ is produced by pedestrian detector (e.g., faster RCNN) from traffic cameras in city $B$. Each suspect's image $x_i^r$ is provided by a witness in city $B$ and has related identity $y_i^r$. Set $D$ is provided by city $A$ and has its own label space that is disjoint with $D_t$ and $D_r$. Meanwhile, dataset $D$ from city $A$ has data bias with $D_t$ and $D_r$ because of different illumination conditions and different camera systems. So, we need to search the suspect from the testing dataset $D_t$ with the help of both $D$ and $D_r$. In other words, $D$ is the source domain composed of multiple labeled training sets and $D_r$ and $D_t$ are prob set and gallery set (of the test set), respectively, from target domain. In our scenario, models can be trained on $D$ and $D_r$ and then be tested on $D_r$ and $D_t$.

We summarize the most settings for person re-ID problem and ours in Table 1. Based on the domain information gap between training set and testing set, the person re-ID problem can be classified as domain adaptive re-ID problem (the last 2 row in Table 1) or general re-ID problem (the first 4 row in Table 1). The general re-ID problem is where the training set and testing set are from the same domain. Under such assumption, the general re-ID problem can be further classified into supervised person re-ID, few-shot person re-ID, unsupervised person re-ID, and semi-supervised person re-ID based on the label information on training set and the trainable of prob set. The last row in Table 1 is our setting, which is similar to unsupervised domain adaptation person re-ID problem but with trainable prob set.

TABLE 1: Comparisons between ours and others' settings for person re-ID problem. √ under "Prob" column means the Prob set can be used to train the model.

| Name | Train set | | Test set | |
| --- | --- | --- | --- | --- |
| | Source domain | Target domain | Prob | Gallery |
| Supervised person re-ID | | Labeled | | |
| Few-shot person re-ID | No | Labeled | √ | |
| Unsupervised person re-ID | | Unlabeled | | |
| Semisupervised person re-ID | | Few-labeled | | |
| Unsupervised domain adaptation person re-ID | Labeled | Unlabeled | | |
| (Ours) cross-domain few-shot person re-ID | Labeled | Unlabeled | √ | |

## 3. Related Work

In this section, we review the most popular works in supervised person re-ID, few-shot person re-ID, and semisupervised person re-ID. Finally, we merge the unsupervised person re-ID and unsupervised domain adaptation person re-ID into one paragraph since most related works handle them together.

### 3.1. Few-Shot Person Re-ID.
In the setting of supervised person re-ID, the identities of training set and testing set do not overlap. Two works [14, 15] propose to merge the prob set of testing set into training set and regard it as few-shot person re-ID problem. Although such modification perfectly meets the definition of few-shot learning [16], those two works do not draw much attention from the person re-ID community. The reason is that supervised person re-ID already achieves excellent performance. People in this community are more interested in the problem of unsupervised and domain adaptation. Hence, the few-shot person re-ID problem belongs to the supervised person re-ID problem, which only has a target domain. Our scenario is not a few-shot person re-ID problem.

### 3.2. Semisupervised Person Re-ID.
Different from supervised person re-ID, semisupervised person reidentification [17] assumes that few data from the training set are labeled and the rest of the training set is unlabeled. In [17], the authors proposed a novel semisupervised region metric learning method, which estimates positive neighbors to generate positive regions and learn a discriminative region-to-point metric. There are some works that focus on the few-example video-based re-ID task. Ye et al. [18] proposed a dynamic graph matching (DGM) method, which iteratively updates the image graph matching and the label estimation to learn a better feature. Liu et al. [19] updated the classifier with K-reciprocal nearest neighbors (KNN) in the gallery set and refined the nearest neighbors by apply negative sample mining with KNN in the query set. After that, Wu et al. [10] proposed a progressive learning framework to exploit the unlabeled data for semisupervised person re-ID. Wu et al. [11] proposed a multiteacher adaptive similarity distillation framework to train a user-specified lightweight student model.

### 3.3. Unsupervised Domain Adaptation Person Re-ID.
Another intuitive thought is that we can utilize auxiliary datasets because we can obtain a large amount of person re-ID datasets easily. In general, these auxiliary datasets are called source datasets, with different data distribution from the target dataset. Although advanced supervised learning methods [20–22] have achieved great performance in person re-ID. However, when testing on a new dataset, the performance drops significantly. To address cross-domain person re-ID problem, many methods [1, 3, 23–31] have been proposed. The model in [1] learns a deep re-ID model as a feature extractor on the source dataset. Next, they use unlabeled target data to refine the model by unsupervised method. Wei et al. [2] proposed a person transfer generative adversarial network (PTGAN) to reduce the gap of two domains, relieving the cost of annotations in the target dataset. The model proposed by Wang et al. [3] learns a transferable joint attribute-identity deep learning (TJ-AIDL) model by transferring the labeled information of the source dataset to a new unseen (unlabeled) target dataset. Wu et al. [23] proposed an Instance-Guided Context Rendering scheme to supervise model learning in the unlabeled target domain. The model proposed by Fu et al. [32] harnesses the similar characteristics existing in the samples from the target domain for learning to conduct person re-ID in an unsupervised manner. Huang et al. [25] concentrated on the difference of backgrounds in training and testing datasets and proposed SBSGAN to generate images with suppressed backgrounds. Wang et al. [29] proposed the SADA approach that guides the source domain images to align the target domain images by using a trained camera classifier.

All of these methods explore the relationship between the source domain and the target domain and transfer source knowledge to the target domain. They do not have any strategies to gain from a small amount of labeled data. Different from the unsupervised person re-ID setting, our scenario is a cross-domain few-shot problem since our target dataset includes a few labeled data. We expect to obtain substantially better performance based on these labeled data than the unsupervised method.

To address the cross-domain few-shot person re-ID problem, we propose a novel method to utilize auxiliary datasets in this paper. Because our support dataset $D_r$ is sampled from the same scenario with testing dataset $D_t$, they have a similar data distribution so that we can use the support dataset to represent the test dataset. Firstly, we regard the support dataset $D_r$ as a template. Each template

data sample includes the image patch $x_i^r$ and related label $y_i^r$. Then, we select helpful samples from auxiliary training datasets to update our model with the guidance of the template data. Lastly, we test our model on the testing dataset.

## 4. Methodology

In this section, we introduce our method to address the cross-domain few-shot person re-ID problem. The difficulty is that we only have a few labeled data on the target dataset. Thus, we try to exploit the assistance of other unlabeled data to improve the performance of the model on the target dataset. To address the domain bias of a different dataset, we propose a novel policy to select important samples from the source dataset to train our model. In this section, we first discuss the intuition and motivation of our method. Then, we analyze the influence function. Lastly, we propose the reweighting instance method under the framework of transfer learning.

*4.1. Intuition and Motivation.* Different from the existent setting of few-shot learning, the definition of our problem does not have the constraint of "k-way N-shot." It means that we cannot follow the solution of traditional few-shot learning. Instead, we refer to the idea of instance based transfer learning, due to two conditions: (1) cross-domain and (2) a few labeled samples from the target domain (few-shot setting) of the problem. The K-nearest neighborhood (KNN) is a simple example of instance based transfer learning. With the few labeled samples used as anchors to reweight the instances from the source domain, the KNN uses the Euclidean distance as the criterion for reweighting. In other words, KNN uses the Euclidean distance to measure the similarity and influence of samples. The influence of a sample from the source domain means how this sample influences the model performance in the target domain. The authors in [12] proposed to use a statistical model as the influence function to measure how a sample from the source domain influences a model's performance on the target domain. Motivated, by [12], we proposed our reweighting instance method based on transfer learning.

*4.2. Revisiting of the Influence Function.* The goal of the work in [12] is to understand the effect of each training sample on the prediction. Formally, consider a model with parameter set $\theta$ and training set $\{z_1, \ldots, z_n\}$. Then, the effect of sample $z$ is defined by $\widehat{\theta}_{-z} - \widehat{\theta}$, where $\widehat{\theta}_{-z} \overset{def}{=} \operatorname{argmin}_{\theta \in \Theta} \sum_{z_i \neq z} L(z_i, \theta)$ and $\widehat{\theta}$ is the optimal parameter set for the model defined as $\widehat{\theta} \overset{def}{=} \operatorname{argmin}_{\theta \in \Theta} \sum_{z_i} L(z_i, \theta)$. The work in [12] employs influence $\widehat{\theta}_{\epsilon, z} \overset{def}{=} \operatorname{argmin}_{\theta \in \Theta} 1/n \sum_{i=1}^n L(z_i, \theta) + \epsilon L(z, \theta)$ function for an arbitrary sample $z$. The idea is to compute the parameter change if $z$ is up-weighted by some small $\epsilon$.

We briefly introduce how the work in [12] derives the closed form of $\widehat{\theta}_{-z} - \widehat{\theta}$ from influence function

$\widehat{\theta}_{\epsilon, z} \overset{def}{=} \operatorname{argmin}_{\theta \in \Theta} 1/n \sum_{i=1}^n L(z_i, \theta) + \epsilon L(z, \theta)$. Define $\triangle_\epsilon = \widehat{\theta}_{\epsilon, z} - \widehat{\theta}$. Then, $\triangle_{\epsilon = -1/n} = \widehat{\theta}_{-z} - \widehat{\theta}$. Based on the local linear hypothesis, we have $\triangle_{\epsilon = -1/n} \approx \triangle_{\epsilon = 0} + d\triangle_\epsilon/d\epsilon|_{\epsilon = 0} \cdot (-1/n)$. With the definition of $\triangle_\epsilon$ and $\widehat{\theta}_{\epsilon, z}$, we know $\triangle_{\epsilon = 0} = 0$. The next step is to determine $d\triangle_\epsilon/d\epsilon$. Note that $d\triangle_\epsilon/d\epsilon = d\widehat{\theta}_{\epsilon, z}/d\epsilon$ since $\widehat{\theta}$ is not related to $\epsilon$. The detailed derivation of (1) can be found in the appendix of [12] or in our material:

$$\mathfrak{T}_{\text{up,params}}(z) \overset{def}{=} \frac{d\widehat{\theta}_{\epsilon, z}}{d\epsilon}|_{\epsilon = 0} = -H_{\widehat{\theta}}^{-1} \nabla_\theta L(z, \widehat{\theta}). \tag{1}$$

where $H_{\widehat{\theta}} \overset{def}{=} 1/n \sum_{i=1}^n \nabla_\theta^2 L(z_i, \widehat{\theta})$ is the Hessian matrix. Equation (1) holds when $L(z_i, \widehat{\theta})$ is twice differentiable and strictly convex in $\theta$. However, the deep neural network does not satisfy the two conditions. There are two important contributions of the work in [12]: (1) approximating the Hessian matrix and speeding up the computation and (2) analyzing the feasibility under the condition of nonconvexity and nonconvergence.

## 5. Reweighting Instance Method Based on Influence Function

In this section, we introduce our reweighting instance method based on influence function (ReWIF for short), which employs the influence function and support set to guide the training of the source train set. First, we formulate our problem and derive the reweighting function through two approximations. Then, we use pretraining to approximate the Hessian matrices. Moreover, we propose a post-processing and dataset absorbing to improve the performance. Finally, we summarize our ReWIF in Algorithm 1.

*5.1. Preliminaries and Formulation.* Different from [12], we aim to maximize the model performance on the support set through determining the weights on the source training set. In other words, the determination of weights is guided by the few support set. So, we put forth the formulation as follows. Equation (2) describes how the model reweights the source training set and the guidance of support set is embodied in equation (3). $z = (z_1, z_2, \ldots z_n)$ represents $n$ source training samples and their corresponding weights are $w = (w_1, w_2, \ldots, w_n)$. $\mathbf{z}^{\text{support}} = (z_1^{\text{support}}, z_2^{\text{support}}, \ldots z_m^{\text{support}})$ denotes $m$ support samples. So, we can jointly optimize equations (2) and (3) to obtain our target model. Different from the influence function proposed in [12], the second term in equation (2) reweights all samples in the training set:

$$\widehat{\theta}_{\mathbf{w}, \mathbf{z}} \overset{def}{=} arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n L(z_i, \theta) + \sum_{i=1}^n w_i L(z_i, \theta) \tag{2}$$

$$\widehat{w} \overset{def}{=} arg \min_{\mathbf{w}} \frac{1}{m} \sum_{i=1}^m L(z_i^{\text{support}}, \widehat{\theta}_{\mathbf{w}, \mathbf{z}}), \tag{3}$$

where $\widehat{w}$ is the solution of equation (3), which satisfies $\partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w}, \mathbf{z}})/\partial w_i = 0$ if $L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w}, \mathbf{z}})$ is convex and

second-order differentiable and smooth about $w_i$. However, the exact solution is difficult to obtain. Instead, we use one-step gradient descent to approximate $\widehat{w}_i$. We approximate each $\widehat{w}_i$ as $\widehat{w}_i \approx -\eta \cdot \partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})/\partial w_i|_{w_i=0}$, where

$$\frac{\partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})}{\partial w_i}\Big|_{w_i=0} = -\nabla_\theta L(\mathbf{z}^{\text{support}}, \widehat{\theta})^T H_{\widehat{\theta}}^{-1} \nabla_\theta L(z_i, \widehat{\theta}). \tag{4}$$

### 5.2. Approximation of Hessian and the Close Form of $\widehat{w}$.
The inverse of Hessian in (4) is difficult to evaluate. To avoid the calculation of the Hessian, we first note that the Hessian results from the first term of equation (2). We approximate the optimization in equation (2) by two-stage optimization as in equations (5) and (6):

$$\widetilde{\theta} \stackrel{def}{=} arg\min_{\theta \in \Theta} \frac{1}{n}\sum_{i=1}^{n} L(z_i, \theta), \tag{5}$$

$$\widehat{\theta}_{\mathbf{w},\mathbf{z}} \stackrel{def}{=} arg\min_{\widetilde{\theta} \in \Theta} \sum_{i=1}^{n} w_i L(z_i, \widetilde{\theta}). \tag{6}$$

For stage one in equation (5), we take the standard back-propagation to find the optimal $\widetilde{\theta}$. For stage two in equation (6), we use the one-step gradient descent to approximate equation (6) by equation (7). In equation (8), the first step follows by the chain rule and the second step by using equation (6). Finally, we provide the expression of $\widehat{w}_i$ in equation (9):

$$\widehat{\theta}_{\mathbf{w},\mathbf{z}} = \widetilde{\theta} - \beta\nabla_{\widetilde{\theta}}\left(\sum_{i=1}^{n} w_i L(z_i, \widetilde{\theta})\right), \tag{7}$$

$$\begin{aligned}\frac{\partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})}{\partial w_i}\Big|_{w_i=0} &= \frac{\partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})}{\partial \widehat{\theta}_{\mathbf{w},\mathbf{z}}}\frac{\partial \widehat{\theta}_{\mathbf{w},\mathbf{z}}}{\partial w_i} \\ &= -\beta\nabla_{\widehat{\theta}_{\mathbf{w},\mathbf{z}}} L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})^T \nabla_{\widetilde{\theta}} L(z_i, \widetilde{\theta}), \end{aligned} \tag{8}$$

$$\begin{aligned}\widehat{w}_i &\approx -\eta\frac{\partial L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})}{\partial w_i}\Big|_{w_i=0} \\ &= \eta\beta\nabla_{\widehat{\theta}_{\mathbf{w},\mathbf{z}}} L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})^T \nabla_{\widetilde{\theta}} L(z_i, \widetilde{\theta}), \end{aligned} \tag{9}$$

where $\eta\,(\eta > 0)$ and $\beta\,(\beta > 0)$ are the step sizes in the two approximations.

### 5.3. Post-Processing.
In equation (9), $\widehat{w}_i$, seen as the inner product of the two gradients, is not necessarily nonnegative. As such, we propose to normalize each $\widehat{w}_i$ as

$$\widetilde{w}_i = \frac{e^{\alpha \cdot r(\widehat{w}_i)}}{\sum_i e^{\alpha \cdot r(\widehat{w}_i)}}, \tag{10}$$

where $r(\widehat{w}_i) = \max\{0, \widehat{w}_i/\sum_i\max\{0, \widehat{w}_i\}\}$ and $\alpha$ is a relaxing factor. Equation (10) means that we first clip the negative $\widehat{w}_i$ to zero and then normalize it by using the softmax function.

We now provide a heuristic explanation of equation (10). The support set is always totally different from the source training set; large angles may exist between the gradient of support set $\nabla_{\widehat{\theta}_{\mathbf{w},\mathbf{z}}} L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})^T$ and some source training samples $\nabla_{\widetilde{\theta}} L(z_i, \widetilde{\theta})$. That is, $\widehat{w}_i$ could be negative and then could be clipped by $r(\widehat{w}_i)$ to zero. From simulations, we find more than half of the elements in $\widehat{w}$ are clipped to zeros. Although we prefer the model focus on a few similar samples, many zeros in vector $r(\widehat{w})$ make the model prone to be trapped in a bad local minimum. Thus, we need to reduce the difference of weights between similar samples $(r(\widehat{w}_i) > 0)$ and dissimilar samples $(r(\widehat{w}_i) = 0)$. The dissimilar samples are useful in the sense that they can provide prior knowledge about person re-ID task and prevent overfitting. $e^{\alpha \cdot r(\widehat{w}_i)}$ prevents our model from discarding the dissimilar samples, where $\alpha$ controls the relaxation degree.

### 5.4. Absorbing Support Set into Source Training Set.
We combine the support set to the source training set to optimize equations (5) and (6). First, normalization in equation (10) eliminates the absolute relationship among $\widehat{w}$. Suppose we use equation (9) to calculate $\{\widehat{w}_1, \widehat{w}_2, \ldots, \widehat{w}_n\}$, most of them are less than zero, and the rest are less than 0.01. It indicates all the $n$ samples are dissimilar with the support set. However, the normalization equation (10) will enlarge the nonzero $\widehat{w}_i$. So, we absorb the support set into the source training set as anchors. Second, the support set share the same domain information with the target testing set, and it improves the model performance on target testing set.

### 5.5. Algorithm Flow.
We summarize ReWIF as shown in Algorithm 1. Assume that, $\mathbf{z}$ is a source training set including the support set $\mathbf{z}^{\text{support}}$. The support set $\mathbf{z}^{\text{support}}$ and test set $\mathbf{z}^{\text{test}}$ are from another domain. In stage one, we optimize $\widetilde{\theta} \stackrel{def}{=} \text{argmin}_{\theta \in \Theta} 1/n\sum_{i=1}^{n} L(z_i, \theta)$ by traditional deep learning optimizer Adam [33]. We prepare $\mathbf{z}$, $\mathbf{z}^{\text{support}}$, and $\widetilde{\theta}$ from stage one. In stage two, we sample two mini-batches $\mathbf{z}_{\text{mini}}$ and $\mathbf{z}_{\text{mini}}^{\text{support}}$ from $\mathbf{z}$ and $\mathbf{z}^{\text{support}}$. Then, calculate $\widehat{w}_i = -\eta\,\partial L(\mathbf{z}_{\text{mini}}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}_{\text{mini}}})/\partial w_i|_{w_i=0}$ based on $\mathbf{z}_{\text{mini}}$ and $\mathbf{z}_{\text{mini}}^{\text{support}}$ and equation (9). Moreover, we normalize $\widehat{w}_i$ through equation (10). Finally, we use $\widetilde{w}$ to compute $\sum_{i=1}^{\|\mathbf{z}_{\text{mini}}\|} \widetilde{w}_i L(z_i, \widetilde{\theta})$ and then update $\widetilde{\theta}$ by gradient descent. Repeat stage two (3–6 lines in Algorithm 1) until the model converges.

### 5.6. Training Time.
For our ReWIF in Algorithm 1, stage two is time-consuming and its running time is about two times that of stage one in each iteration. This can be seen from equation (4), where an extra forward and backward procedure is required to calculate $\nabla_{\widehat{\theta}_{\mathbf{w},\mathbf{z}}} L(\mathbf{z}^{\text{support}}, \widehat{\theta}_{\mathbf{w},\mathbf{z}})^T$. Fortunately, the number of iteration of stage two is much less than that of stage one.

**Require:** a deep CNN model (we choose ResNet50) with parameter $\theta$; cross-entropy loss $L(z, \theta)$;
**Ensure:** the two stages optimized on the same CNN model;
(1) Stage one: prepare $\mathbf{z}, \theta$, then we optimize $\widetilde{\theta} \overset{def}{=} \operatorname{argmin}_{\theta \in \Theta} 1/n \sum_{i=1}^{n} L(z_i^{\text{train}}, \theta)$ by traditional deep learning optimizer Adam with loss $L(z, \theta)$;
(2) Stage two: prepare $\mathbf{z}, \mathbf{z}^{\text{support}}$ support, and $\widetilde{\theta}$;
(3) Sample mini-batch $\mathbf{z}_{\text{mini}}$ and $\mathbf{z}_{\text{mini}}^{\text{support}}$ from $z, \mathbf{z}^{\text{support}}$, respectively;
(4) Calculating the $\widehat{w}$ based on $\mathbf{z}_{\text{mini}}$ and $\mathbf{z}_{\text{mini}}^{\text{support}}$ and equation (8);
(5) Make the postprocessing by calculating equation (10).
(6) Using $\overline{w}$ to compute $\sum_{i=1}^{n} \widetilde{w}_i L(z_i, \widetilde{\theta})$ then update $\widetilde{\theta}$ by gradient descent;
(7) Repeat steps 3–6 until $\widetilde{\theta}$ converges to $\widehat{\theta}$.
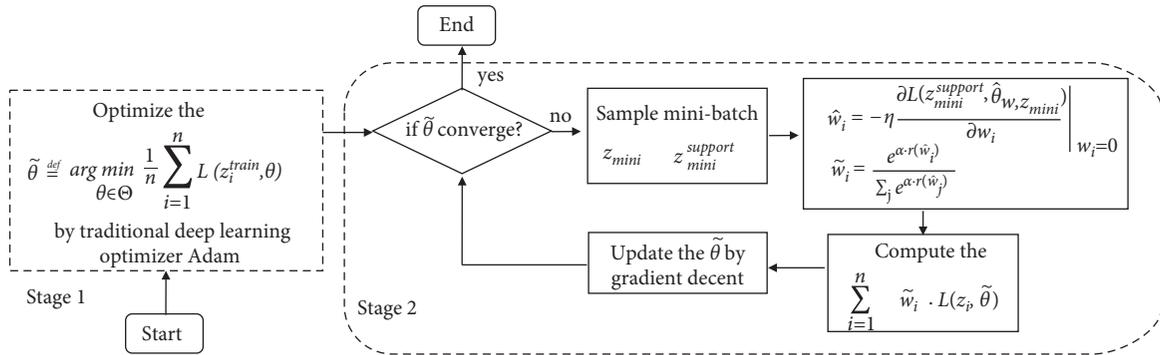
ALGORITHM 1: Framework of ReWIF.



FIGURE 1: Diagram of ReWIF. The left part is stage one and the right part is stage two.

ReWIF: We name ... and individual method.
Figure 1 shows the diagram of the ReWIF algorithm.

# 6. Experiments

The experiment contains three parts. First, some basic settings about dataset, evaluation metrics, network architecture, and training policy are introduced. Second, several experiments on the effectiveness of ReWIF are made, followed by an ablation study on two important hyperparameters. Third, we combine the ReWIF with other methods [1] and compare with their original versions.

## 6.1. Dataset and Settings

### 6.1.1. Dataset.
To simulate different domain of different cities, we use different re-ID benchmarks: Market-1501 [34], DukeMTMC-reID [35], and CUHK03 [36]. Market-1501 contains 32,668 images with 1501 identities in total. It has three parts: 12,936 images of 751 identities for training and 19,732 images of 750 identities for testing. We rename "Market-1501" to "Market" for simplicity. DukeMTM-CreID is a subset of DukeMTMC [37] dataset. It has 16,522 images of 702 identities for training and 2,228 query images and 17,661 gallery images of 702 identities for testing. We rename "DukeMTMC-reID" to "Duke" for simplicity. CUHK include 3 different datasets CUHK01 (971 identities, 3884 images, manually cropped) [38], CUHK02 (1816 identities, 7264 images, manually cropped) [39], and

CUHK03 (1360 identities, 13164 images, manually cropped + automatically detected) [40]. We follow the official splitting for training and testing.

Market, Duke, and CUHK belong to different domains that have different camera systems and illumination conditions. That is, they are suitable for our proposed scenario.

### 6.1.2. Evaluation Metrics.
As one case of our experiments, we regard Duke as city $A$ and Market as city $B$. The training set of Duke is our auxiliary source training set. We randomly choose $M$ images of K identities from Market query images as the support set. The number of each person's images is not equal. The gallery images of Market are our test set. We use the $M$ queries to retrieval on the testing set and report mean average precision (mAP) and rank-1 accuracy for our testing set. Query and gallery sets could have same camera views, but for each individual query identity, his/her gallery samples from the same camera have been excluded.

### 6.1.3. Backbone.
We employ a strong baseline [41] which has reached the state-of-the-art on Market and Duck as our backbone. The main tricks can be summarized in the following four points:

(1) We use ResNet50 as our backbone, initialize the network with pretrained parameters on ImageNet, and optimize the network with Adam. Meanwhile, we change the dimension of last fully connected layer

TABLE 2: Comparison results for different methods under a single source set. All settings are experimented repeatedly for 10 times and get the average value.

| Source —> Target | Size | Policy1 | | Policy2 | | Policy3 | | ReWIF_S | | ReWIF | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 |
| Duke—>Market | 50 | 21.2 | 47.9 | 78.3 | 89.5 | 79.7 | 91.7 | 81.9 | 93.1 | 82.2 | 95.9 |
| Duke—>Market | 10 | 18.5 | 43.2 | 61.1 | 77.7 | 76.2 | 90.3 | 80.4 | 91.7 | 81.9 | 94.8 |
| Duke—>Market | 5 | 15.5 | 44.4 | 46.9 | 62.5 | 65.5 | 82.3 | 68.6 | 85.6 | 69.9 | 88.6 |
| Duke—>Market | 1 | 13.4 | 40.0 | 32.4 | 58.6 | 61.2 | 80.4 | 66.5 | 81.2 | 67.2 | 85.9 |
| Market—> Duke | 50 | 19.3 | 40.8 | 66.0 | 79.9 | 71.7 | 89.7 | 76.7 | 91.9 | 79.6 | 93.4 |
| Market—> Duke | 10 | 18.2 | 40.0 | 43.1 | 68.6 | 70.2 | 87.1 | 75.8 | 87.1 | 77.9 | 91.1 |
| Market—> Duke | 5 | 16.8 | 37.1 | 31.9 | 58.8 | 63.2 | 80.1 | 65.9 | 81.9 | 74.8 | 87.2 |
| Market—> Duke | 1 | 15.4 | 33.3 | 27.4 | 46.6 | 60.5 | 75.9 | 62.8 | 79.3 | 70.1 | 84.1 |

to $S$, where $S$ is the number of identities in support dataset and training dataset.

(2) The input size of image is $256 \times 128$ pixels. We adopt random erasing and flipping horizontally with 0.5 probability as the data augmentation, and the cross-entropy loss is adopted.

(3) Each image is decoded into 32 bit floating point raw pixel values in $[0, 1]$. Then, we normalize RGB channels by subtracting 0.485, 0.456, and 0.406 and dividing by 0.229, 0.224, and 0.225, respectively.

(4) Randomly sample $P$ identities and $K$ images of per person to constitute a training batch. Finally, the batch size equals to $B = P \times K$. In this study, we set $P = 32$ and $K = 4$.

*6.1.4. Training Set Policy.* We combine the training set and the support set as a new training set. In each iteration, we randomly sample 128 images from the new training set and 192 images from support set. If the size of support is smaller than 192, we use all samples. Then, we fine-tune our model for 60 epochs with the guidance of the support set, and the learning rate is $3.5 \times 10^{-6}$ during whole the fine-tuning process.

*6.2. Experiment Results.* We compare our proposed ReWIF model with ReWIF_S and three baseline policies.

Policy1 trains the model only on the train set during the whole training process. The total training epochs is 180. The initial learning rate is $3.5 \times 10^{-4}$ and is multiplied by 0.1 at every 60 epochs.

Policy2 learns the model only on the support dataset during the whole training process. The total training epoch is 2500. The initial learning rate is $3.5 \times 10^{-4}$ and is decreased by 0.1 at 1000, 1500, and 2000.

Policy3 [42] is also a reweighting method motivated by [12]. It reweights training samples for dealing with the imbalanced classification problem. We make necessary but trivial modifications of the method to fit our problem. The total training epochs is 180. The initial learning rate is $3.5 \times 10^{-4}$ and is multiplied by 0.1 at every 60 epochs.

ReWIF_S learns the model on the combined dataset of training dataset and support dataset during the whole training process. The total training epoch is 610. The initial learning rate is $3.5 \times 10^{-4}$ and is decreased by 0.1 at 40, 160, and 500.

ReWIF fine-tune on the ReWIF_S for 60 epochs with the guidance of support set, and the learning rate is $3.5 \times 10^{-6}$ during whole fine-tuning process.

The 2500 and 610 epochs of Policy2 and ReWIF_S are much larger than 180 epochs of Policy1. In the training procedure of Policy2 and ReWIF_S, the loss converges to a low level in early time (100 epochs for Policy2 and 50 epochs for ReWIF_S). However, the mAP still improves after the convergence of the loss. For ReWIF_S, the reason may be that the network regards the few support samples as hard samples and fits them slowly. The loss of the support set could be inundated in the large amount of training set.

*6.2.1. Results on Single Source Training Set.* The "Size" indicates the size of support set, where $Y$ denotes the number of persons. All the models are evaluated on the testing set and experimented repeatedly for 10 times and get the average value, and the results are shown in Table 2. Our proposed method obtains significant performance improvement on the testing set. Policy1 only has the mAP metric less than 20, which indicates that there is a large gap between two domains. Meanwhile, Policy2 has unsatisfactory performance because the small training set makes the model over-fit. And the performance of Policy3 is between Policy2 and RINIMI_S. The reweighting method is not suitable for our cross-domain few-shot problem. By comparison, ReWIF and ReWIF_S are better than Policy 1–3 and ReWIF is better than ReWIF_S. To prove the effectiveness of addressing few-shot problems with our method, we study the influence of the size on the support set. From the results in Table 2, the ReWIF method has larger improvement on the fewer labeled data. Especially, when the size of support sets is 5 (Market as target set), the mAP has 23 and 4.4 improvement, respectively, compare with Policy2 and Policy3. The improvement is larger when the size of the support set is 1. Finally, the performance of all methods is worse with the decrease of the identity size, but our ReWIF can keep a relatively high performance when the identity size of support set is small.

TABLE 3: Robustness results on different sizes of support sets.

| Source —> Target | Size | ReWIF | |
| | | mAP | Rank-1 |
|---|---|---|---|
| Duke—>market | 50 | 82.2 ( ± 2.4) | 95.9 (± 6.8) |
| Duke—>market | 10 | 83.9 (± 5.9) | 94.8 (± 11.7) |
| Duke—>market | 5 | 69.9 (± 8.6) | 88.6 (± 16.1) |
| Duke—>market | 1 | 67.2 (± 9.7) | 85.9 ( ± 23.9) |
| Market—> Duke | 50 | 79.6 ( ± 2.9) | 93.4 (± 5.3) |
| Market—> Duke | 10 | 77.9 (± 4.7) | 91.1 (± 9.5) |
| Market—> Duke | 5 | 74.8 (± 7.9) | 87.2 (± 15.9) |
| Market—> Duke | 1 | 70.1 (± 8.4) | 84.1 ( ± 21.8) |

TABLE 4: Experiment results of ReWIF leverages multisource sets. Three additional datasets ①, ②, and ③ will included in the source set one-by-one. (①, ②, and ③ present CUHK01 [38], CUHK02 [39], and CUHK03 [40], respectively).

| Source—> Target | Policy1 | | Policy2 | | Policy3 | | ReWIF_S | | ReWIF | |
| | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP | Rank-1 |
|---|---|---|---|---|---|---|---|---|---|---|
| Duke +①—>Market | 20.1 | 46.5 | 77.8 | 95.3 | 76.5 | 97.7 | 82.0 | 93.7 | 82.9 | 96.3 |
| Duke +①②—>Market | 17.3 | 40.9 | 60.3 | 93.0 | 79.5 | 90.9 | 82.8 | 94.7 | 84.3 | 96.6 |
| Duke +①②③—>Market | 13.5 | 40.7 | 43.9 | 90.6 | 64.9 | 96.3 | 83.7 | 95.1 | 85.1 | 97.5 |
| Market +①—>Duke | 13.2 | 29.6 | 65.8 | 88.8 | 70.9 | 89.3 | 76.9 | 92.1 | 79.9 | 94.1 |
| Market +①②—>Duke | 12.7 | 27.1 | 42.5 | 62.9 | 73.8 | 94.3 | 78.5 | 93.1 | 80.5 | 94.9 |
| Market +①②③—>Duke | 10.9 | 25.3 | 24.1 | 47.1 | 70.1 | 92.9 | 79.4 | 94.3 | 81.3 | 95.7 |

*6.2.2. Robustness Results on Support Set.* We experiment the robustness for different support sets. For each identity size (1, 5, 10, and 50), we randomly choose 10 different support sets and train the network with ReWIF. Then, we calculate the mean and bias of the mAP and rank-1 for each identity size (the value in the parenthesis means bias), and the results are shown in Table 3. With the decrease of the identity size, less robustness of the method can be seen from the bias of the mAP.

*6.2.3. Results on Multisource Training Set.* Our method can merge multidomain source datasets conveniently. We include the CUHK01 [38], CUHK02 [39], and CUHK03 [40] into the source set one-by-one. The size of support set is 50(215) for Market target set and 50(169) for Duke target set, and the results are shown in Table 4. With the increase of the number of source set, the performance of the methods ReWIF and ReWIF_S increase stable. Instead, the results of Policy1, Policy2, and Policy3 on multisource are reversal with ReWIF, even randomly. The results in Table 4 demonstrate the effectiveness of our ReWIF on the aspect of leverage cross-domain datasets.

*6.3. Ablation Study*

*6.3.1. Influence of Learning Rate.* To show the influence of the learning rate, we test different learning rates during the whole fine-tuning process, and the results are shown in Table 5. In our experiments, the learning rate, set to 3.5e-5, can achieve the best performance of 84.8 mAP.

*6.3.2. Influence of $\alpha$.* $\alpha$ is a tuning parameter in equation (10). We study the effectiveness and necessity of the alpha factor, and the results are shown in Table 5. Meanwhile, we show that a suitable alpha factor is also important. We search alpha's value ranging from 0.0 to 10.0. The results in Table 5 show that alpha equal to 1.0 can achieve the best performance of 84.8 mAP.

*6.3.3. Adaption to Other Methods.* We combine the ReWIF with one unsupervised domain adaptation person re-ID method, PUL [1], one semisupervised person re-ID method, and One-Example [10].

The details of PUL are [1] as follows: it first trains a model on the source labeled dataset, uses the trained model to make predictions on the target dataset and obtains the soft-label (with many errors). Finally, we use the soft-label and data in the target dataset to retrain the model. We use ReWIF to replace the procedure of training on source labeled dataset at PUL and drive to PUL with ReWIF.

For One-Example [10], initially, a CNN model is trained on the one-example labeled samples (support set). Then, generate the pseudolabels for all unlabeled samples (target set) and select some reliable pseudolabeled data for training according to the prediction confidence. The selected subset is continuously enlarged during iterations according to a sampling strategy. At the initial stages, only the most reliable and easiest ones are included. In the subsequent stages, a growing number of pseudolabeled data are gradually selected to incorporate more difficult and diverse data. Since the One-Example method is designed on one domain, we employ the model which is trained with the One-Example

TABLE 5: Ablation results.

| Ablation experiments on learning rate | | | | | | | |
|---|---|---|---|---|---|---|---|
| Learning rate | | | 3.5e-5 | | 3.5e-6 | | 3.5e-7 |
| mAP | | | 75.2 | | 84.4 | | 82.6 |
| Ablation experiments on $\alpha$ | | | | | | | |
| $\alpha$ | 0.0 | 0.5 | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 10.0 |
| mAP | 82.4 | 82.4 | 84.8 | 84.4 | 84.0 | 83.5 | 83.1 | 81.8 |

TABLE 6: Adaptation to other methods.

| Method | Datasets | mAP | Rank-1 |
|---|---|---|---|
| PUL [1] | Duke —> Market | 61.1 | 74.9 |
| | Market —> Duke | 58.4 | 63.4 |
| One-Example [10] | Duke | 63.3 | 72.9 |
| | Market | 63.6 | 82.5 |
| PUL with ReWIF | Duke —> Market | 82.1 | 98.1 |
| | Market —> Duke | 76.2 | 94.1 |
| One-Example with REWIF | Duke | 85.2 | 100 |
| | Market | 85.9 | 100 |

TABLE 7: Comparisons of other methods with Duke [35] and Market [34]: results from [30].

| Methods | Duke —> Market | | Market —> Duke | |
|---|---|---|---|---|
| | mAP | Rank-1 | mAP | Rank-1 |
| UMDL [4] | 12.4 | 34.5 | 7.3 | 18.5 |
| PUL [1] | 20.5 | 45.5 | 16.4 | 30.0 |
| SPGAN [43] | 22.8 | 51.5 | 22.3 | 41.1 |
| ATNet [44] | 25.6 | 55.7 | 24.9 | 45.1 |
| MMFA [48] | 27.4 | 56.7 | 24.7 | 45.3 |
| TJ-AIDL [3] | 26.5 | 58.2 | 23.0 | 44.3 |
| CamStyle [45] | 27.4 | 58.8 | 25.1 | 48.4 |
| HHL [46] | 31.4 | 62.2 | 27.2 | 46.9 |
| ECN [47] | 43.0 | 75.1 | 40.4 | 63.3 |
| UDAP [49] | 53.7 | 75.8 | 49.0 | 68.4 |
| AD-Cluster [30] | 68.3 | 86.7 | 54.1 | 72.6 |
| UNRN [50] | 78.1 | 91.9 | 69.1 | 82.0 |
| ReWIF (Ours) | 55.9 | 79.1 | 51.6 | 70.9 |

method as the initial model and apply the ReWIF on it, driving to One-Example with ReWIF.

The size of support set is 50. The rank-1 and mAP of PUL and One-Example in our dataset setting are much larger than that in their paper (see Table 6). It is reasonable since the number of queries in our dataset setting is less. Meanwhile, PUL with ReWIF and One-Example with ReWIF are both better than their original versions on mAP and rank-1 due to the use of support set. It is surprising that the setting of absorbing the support set into training set improves the performance of this application.

*6.4. Comparison with Other Methods.* In this section, we compare ReWIF with state-of-the-art unsupervised person re-ID methods, including (1) UMDL [4] and PUL [1] that employed unsupervised learning; (2) eight UDA-based (Unsupervised Domain Adaptation, short for UDA)

methods, including SPGAN [43], ATNet [44], CamStyle [45], HHL [46], and ECN [47] that used GANs, MMFA [48] and TJ-AIDL [3] that used images attributes, and UDAP [49] and AD-Cluster [30] that employed clustering. Table 7 shows the person re-ID performance while adapting from Market to Duke and vice versa.

We combine the ReWIF with UDAP [49]; the details are summarized as follows: we first train a model with the UDAP method. Then, we borrow the pseudolabel of clustering results as the support set of ReWIF to fine tune the model with ReWIF. In order to follow the UDA setting, we build the support set to use the samples of training set from target domain instead of prob set of target domain.

As shown in Table 7, the UDA-based methods get better performances in most cases compare with unsupervised methods. Specifically, AD-Cluster, the state-of-the-art method, performs much better than other methods. The performance of GAN-based methods is diverse. Note that

our ReWIF achieves good performance compared with GAN-based methods. Even comparing with AD-Cluster, we get competitive results.

People might notice that the UNRN [50] is much better than other methods. UNRN borrows the uncertainty idea [51, 52] with mean teacher framework [53], which is the new state-of-the-art method on those two datasets of unsupervised domain adaptation task. However, since it is not directly to adaptive our method with it, we leave it as the future work.

## 7. Conclusion and Future Work

In this work, we proposed a special application scenario for cross-domain few-shot person re-ID problem. To address the problem of insufficient labeled data in the target dataset, we try to involve an auxiliary training set as source datasets. Meanwhile, because of the gap between the source dataset and the target dataset, the performance of the model trained on the source dataset drops drastically. So, we propose a reweighting policy that minimizes the loss of the support dataset to select helpful samples. The experiment results confirm that our reweighting policy is effective and achieves state-of-the-art performance in our problem.

*7.1. Limitations.* The ReWIF is a time-consuming method since the computation of second-order derivative (although we remove the computation of Hessian metrics). A direct drawback is that we should train the model again when the new suspect appears. Fortunately, the training time of ReWIF is less than 1.5 hours on 4 GTX 2080 GPUs. Moreover, when we accumulate enough suspects' images (labeled data) after deploying the system in the new city for about half a year, we can retrain the model as a supervised person re-ID problem

*7.2. Future Works.* The future works can be summarized in two parts. Firstly, extending our framework to methods beyond the single-network-based method, e.g., AD-Cluster, which is an adversarial framework. Another direction, e.g., video data are common used in real-world application, which need to further explore and adapt to our method. Secondly, exploring the network compression methods since the drawback mentioned above. Three works [54–56] can be explored to improve ReWIF, especially for the dictionary learning-based multifeature fusion method [54].

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

## References

[1] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised per son re-identification: clustering and fine-tuning," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 14, no. 4, pp. 1–18, 2018.

[2] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 79–88, Salt Lake City, UT, USA, June 2018.

[3] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised per son re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2275–2284, Salt Lake City, UT, USA, June 2018.

[4] P. Peng, T. Xiang, Y. Wang et al., "Unsupervised cross-dataset trans fer learning for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1306–1315, Las Vegas, NV, USA, June 2016.

[5] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, "A bottom up clustering approach to unsupervised person re identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8738–8745, 2019.

[6] H.-X. Yu, W.-S. Zheng, A. Wu, X. Guo, S. Gong, and J.-H. Lai, "Unsupervised person re-identification by soft multi label learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2148–2157, Long Beach, CA, USA, June 2019.

[7] J. Lv, W. Chen, Q. Li, and C. Yang, "Unsupervised cross dataset person re-identification by transfer learning of spatial-temporal patterns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7948–7956, Salt Lake City, UT, USA, June 2018.

[8] M. Li, X. Zhu, and S. Gong, "Unsupervised person re identification by deep learning tracklet association," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 737–753, 2018.

[9] M. Ye, X. Lan, and P. C. Yuen, "Robust anchor embedding for unsupervised video person re-identification in the wild," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 170–186, Munich, Germany, September 2018.

[10] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, and Y. Yang, "Progressive learning for person re-identification with one example," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2872–2881, 2019.

[11] A. Wu, W.-S. Zheng, X. Guo, and J.-H. Lai, "Distilled per son re-identification: towards a more scalable system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1187–1196, Long Beach, CA, USA, June 2019.

[12] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proceedings of the International Conference on Machine Learning, PMLR*, pp. 1885–1894, 2017.

[13] vip.stock.fifinance.sina, "Hangzhou hikvision digital technology co. ltd. 2020 annual report," 2021, http://vip.stock.fifinance.sina.com.cn/corp/view/vCB_AllBulletinDetail.php?stockid=002415&id=7108424.

[14] L. Wu, Y. Wang, H. Yin, M. Wang, and L. Shao, "Few-shot deep adversarial learning for video-based person reidentification," *IEEE Transactions on Image Processing*, vol. 29, pp. 1233–1245, 2020.

[15] A. Rahimpour and H. Qi, "Attention-based few-shot person re-identification using meta learning," 2018, https://arxiv.org/abs/1806.09613.

[16] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: a survey on few-shot learning," *ACM Computing Surveys*, vol. 53, no. 3, pp. 1–34, 2021.

[17] J. Li, A. J. Ma, and P. C. Yuen, "Semi-supervised region metric learning for person re-identification," *International Journal of Computer Vision*, vol. 126, no. 8, pp. 855–874, 2018.

[18] M. Ye, A. J. Ma, L. Zheng, J. Li, and P. C. Yuen, "Dynamic label graph matching for unsupervised video reidentification," in *Proceedings of the IEEE international conference on computer vision*, pp. 5142–5150, Venice, Italy, October 2017.

[19] Z. Liu, D. Wang, and H. Lu, "Stepwise metric promotion for unsupervised video person re-identification," in *Proceedings of the IEEE international conference on computer vision*, pp. 2429–2438, Venice, Italy, October 2017.

[20] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[21] Z. Feng, J. Lai, and X. Xie, "Learning view-specific deep net works for person re-identification," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3472–3483, 2018.

[22] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, https://arxiv.org/abs/1703.07737.

[23] Y. Chen, X. Zhu, and S. Gong, "Instance-guided context rendering for cross-domain person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[24] A. Wu, W.-S. Zheng, and J.-H. Lai, "Unsupervised person reidentification by camera-aware similarity consistency learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[25] Y. Huang, Q. Wu, J. Xu, and Y. Zhong, "Sbsgan: suppression of inter-domain background shift for person reidentification," in *Proceedings of the IEEE/CVF Inter national Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[26] Y.-J. Li, C.-S. Lin, Y.-B. Lin, and Y.-C. F. Wang, "Cross dataset person re-identification via unsupervised pose disentanglement and adaptation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[27] Y. Lin, L. Xie, Y. Wu, C. Yan, and Q. Tian, "Unsupervised person re-identification via softened similarity learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.

[28] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao, "A novel unsupervised camera-aware domain adaptation frame work for person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[29] G. Wang, J.-H. Lai, W. Liang, and G. Wang, "Smoothing adversarial domain attack and p-memory reconsolidation for cross-domain person reidentification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.

[30] Y. Zhai, S. Lu, Q. Ye et al., "Ad-cluster: augmented discriminative clustering for do main adaptive person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.

[31] X. Zhang, J. Cao, C. Shen, and M. You, "Self-training with progressive augmentation for unsupervised cross domain person reidentification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[32] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. S. Huang, "Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), October 2019.

[33] D. P. Kingma and J. Ba, "A method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[34] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: a benchmark," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, December 2015.

[35] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2138–2147, Long Beach, CA, USA, June 2019.

[36] L. Wei, Z. Rui, X. Tong, and X. G. Wang, "Deepreid: deep filter pairing neural network for person re-identification," in *Proceedings of the Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014.

[37] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proceedings of the European conference on computer vision*, pp. 17–35, Springer, Amsterdam, The Netherlands, October 2016.

[38] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proceedings of the 11th Asian Conference on Computer Vision*, Daejeon, Korea, November 2012.

[39] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, June 2013.

[40] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: deep filter pairing neural network for person re-identification," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014.

[41] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, June 2019.

[42] M. Ren, W. Zeng, B. Yang, and R. Urtasun, "Learning to reweight examples for robust deep learning," in *Proceedings of the International Conference on Machine Learning, PMLR*, pp. 4334–4343, 2018.

[43] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self similarity and domain-dissimilarity for person reidentification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 994–1003, Salt Lake City, UT, USA, June 2018.

[44] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person reidentification," in

*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7202–7211, Long Beach, CA, USA, June 2019.

[45] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Cam style: a novel data augmentation method for person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1176–1190, 2019.

[46] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 172–188, Munich, Germany, September 2018.

[47] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: exemplar memory for domain adaptive person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 598–607, Long Beach, CA, USA, June 2019.

[48] S. Lin, H. Li, C.-T. Li, and A. C. Kot, "Multi-task mid level feature alignment network for unsupervised cross dataset person re-identification," 2018, https://arxiv.org/abs/1807.01440.

[49] L. Song, C. Wang, L. Zhang et al., "Unsupervised domain adaptive re-identification: theory and practice," *Pattern Recognition*, vol. 102, Article ID 107173, 2020.

[50] K. Zheng, C. Lan, W. Zeng, Z. Zhang, and Z.-J. Zha, "Exploiting sample uncertainty for domain adaptive person re-identification," 2020, https://arxiv.org/abs/2012.08733.

[51] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[52] Z. Zheng and Y. Yang, "Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1106–1120, 2021.

[53] A. Tarvainen and H. Valpola, "Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[54] R. Zhou, X. Chang, L. Shi, Y.-D. Shen, Y. Yang, and F. Nie, "Person reidentification via multi-feature fusion with adaptive graph learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1592–1601, 2020.

[55] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semi-supervised feature analysis for video semantic recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 648–660, 2018.

[56] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semi-supervised recurrent convolutional at tention model for human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1747–1756, 2020.