Neurocomputing 467 (2022) 229–241

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Cross-domain person re-identification with pose-invariant feature decomposition and hypergraph structure alignment

Shuanglin Yan [a], Yafei Zhang [b,c,*], Minghong Xie [b], Dacheng Zhang [b], Zhengtao Yu [b,c]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming650500, China
[c] Yunnan Key Laboratory of Artificial Intelligence, Kunming 650500, China

ARTICLE INFO

Article history:
Received 22 September 2020
Revised 27 July 2021
Accepted 25 September 2021
Available online 1 October 2021
Communicated by Zidong Wang

Keywords:
Person re-identification
Dictionary learning
Matrix factorization
Domain shift problem
Hypergraph structure alignment

ABSTRACT

Person Re-identification (Re-ID) has attracted more and more attention thanks to its great practical value in the field of video surveillance. Most works have focused on solving the problem of supervised Re-ID on a single domain and made significant progress. However, the cross-domain Re-ID is still challenging due to the domain bias between the source and target domains. To this end, we propose a dictionary learning algorithm based on matrix factorization to eliminate the influence of style and pedestrian pose information on the cross-domain Re-ID. Specifically, the proposed approach includes two novel parts: (1) the original visual feature is decomposed into pose-invariant feature space, camera-style feature space and residual feature space to extract discriminant pose-invariant feature that is not affected by style and pedestrian pose information, such that the influence of interference information between pedestrians on recognition can be eliminated; (2) considering the domain-invariance of attribute, a hypergraph structure alignment is introduced to integrate pose-invariant feature, attribute and pedestrian identity into a dictionary learning framework. The relationship between pose-invariant feature and attribute is built so that the pedestrian attribute of the target dataset can be accurately predicted during testing. Finally, the pedestrian similarity measurement can be carried out by combining the pose-invariant feature and attribute of pedestrians. The effectiveness of the proposed algorithm is verified with the experiments on several benchmark Re-ID datasets.

1. Introduction

With the development of deep learning, Re-ID based on supervised learning has achieved significant performance in recent years. However, supervised Re-ID requires a huge number of labeled samples for training, which leads to a poor practicability and scalability in real-world scenarios. Therefore, one possible solution for this issue could be using the unsupervised Re-ID model with large-scale unlabeled data. This work aims at solving the problem of unsupervised domain adaptation (UDA) for Re-ID which is also known as the cross-domain Re-ID.

Various cross-domain Re-ID methods have been proposed, which can be mainly summarized into two categories:

- The methods deal with the cross-domain Re-ID problem by narrowing the gap between the source and target domains, including domain-invariant feature learning, domain alignment and GAN based methods [1–6]. These methods consider only the inter-domain variation between the source and target domains whereas the intra-domain (different camera views) variation of a single domain has been ignored, which is an important factor affecting Re-ID performance.
- The methods mine the underlying data distribution information of the target domain for model refinement [7–11]. These methods only take the model pre-trained on the source domain as the initial model for the feature learning in the target domain. They do not make full use of the labeled source data as beneficial supervision in the adaptation process.

In this study, the intra-domain variation of each domain is carefully considered, the separation of style information and identity information is investigated, and the influence of pedestrian pose variation on recognition is considered as well. In addition, the supervision information of the source domain has been fully utilized in the whole domain adaptation process.

* Corresponding author at: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China.
E-mail address: zyfeimail@163.com (Y. Zhang).

https://doi.org/10.1016/j.neucom.2021.09.054
0925-2312/

For UDA, how to effectively transfer the knowledge of the source domain to the target domain is the key issue. The style difference between different domains widen the gap between domains and increase the difficulty of knowledge transfer. Specifically, there are great difference in camera style between different camera networks. Fig. 1 is images from DukeMTMC-reID [12] and Market1501 [13] datasets respectively. Camera-style variation is not only the main factor leading to the domain shift problem, but also might significantly change the appearance of person. The existence of camera style greatly reduces the scalability and effectiveness of Re-ID model. If camera style information can be filtered out, the domain shift problem can be effectively solved. A phenomenon can be discovered that images from the same camera view show a common visual style (as shown in each line of images in Fig. 1), and this part of information shared by all images from the same view shows strong correlation. Based on this, we impose a low-rank constraint on camera style information to emphasize the camera style shared by images from the same camera view. And we also introduce the single-view-style-invariance that the separated style features of images from single camera view should be close to each other. Moreover, pose variation (as shown in each column of images in Fig. 1) is also one of the important factors that lead to person appearance change. Taking this into account, we enforce the pose-invariance under the assumption that identity-related feature of a pedestrian image and its corresponding identity-related features of images under different poses of the same identity should be close to each other.

Based on the above aspects, we propose a cross-domain Re-ID dictionary learning algorithm based on pose-invariant feature factorization and hypergraph structure alignment. Fig. 2 shows the framework of the proposed approach. The original visual feature is decomposed into three latent spaces: 1) a pose-invariant feature space that encodes identity-related feature that is not affected by camera-style and pose variation, 2) a camera-style feature space that encodes camera style information, and 3) a residual feature space that encodes the residual information that interferes with person Re-ID. We take the feature representation of the pose-invariant feature space as the final pedestrian feature representation. Such feature representation not only fully reduces the distribution difference of camera-level sub-domains, but also processes the difference between the source and target domains at a finer level.

In addition, considering the robustness and domain-invariance of semantic attribute, a hypergraph structure alignment is introduced to establish the transformation relationship between pose-invariant feature and semantic attribute. Then, semantic attribute of the target dataset can be accurately predicted, and the gap between different domains can be further narrowed by combining pose-invariant feature space and semantic attribute space for measurement. Finally, given a Re-ID model trained on a labeled training set, we observe that the top retrieval results are always more likely to be visually related to the query. Based on such consideration, an effective self-supervised learning mechanism is introduced to further adjust the model trained on the labeled dataset to adapt the target dataset. In this work, the main contributions are as follows.

1. A dictionary decomposition model of unsupervised cross-domain Re-ID is proposed by using the low-rank characteristics and the single-view-style-invariance of style information, pose-invariance of pedestrians. Through the decomposition model, the influence of camera style and pose information on cross-domain Re-ID is eliminated, and the difference between different domains is reduced.
2. By introducing an effective hypergraph structure alignment, the transformation relationship between pose-invariant feature and semantic attribute is established, and the strength of the joint measurements is leveraged.
3. Extensive experiments are carried out on several Re-ID datasets. The experimental results show that the proposed method can achieve comparable performance and surpass most unsupervised learning methods.

The remainder of the paper is organized as follows. Some methods related to this paper are briefly reviewed in Section 2. The proposed approach and the corresponding optimization algorithm are described in Section 3. The extensive experimental results are presented in Section 4. The analysis of the proposed algorithm is performed in Section 5. some conclusions are summarized in Section 6.

2. Related work

2.1. Unsupervised Person Re-ID

The supervised Re-ID methods [16–20] have achieved very remarkable performance. However, the performance will drop sharply when testing on a new unlabeled dataset, which limits its scalability and practicability. In order to solve this problem, many unsupervised Re-ID methods have been proposed in recent years. The aim is to design classification algorithms using the information carried by unlabeled data itself. Earlier works [14,21] designed some hand-crafted features that are robust to illumination, view and pose. Matsukawa et al. [14] proposed a descriptor based on the hierarchical distribution of pixel features. The descriptor models the mean and covariance information of pixel features in each patch and region hierarchical structure at the same time, which effectively improves the discrimination ability of features. This kind of method does not need to learn and to directly measure the similarity of the extracted features, but it ignores the sample distribution information of datasets. Some methods are proposed based on saliency learning [22,23] to locate the salient parts of pedestrian images to mine pedestrian discrimination information. Zhao et al. [22] used patch-level saliency detection to extract saliency features without identity labels for unsupervised Re-ID. However, these methods are not as efficient for large datasets.

In addition, clustering algorithms are often used to generate pseudo labels for target data [24,25], and these pseudo labels are used for supervised learning to train the classifiers. Yu et al. [24] proposed an unsupervised asymmetric measurement learning method, which uses asymmetric K-means clustering to realize person view invariance. However, the pseudo labels obtained by clustering may be noisy, it may assign the same pseudo labels to similar images with different identities, making it more difficult to distinguish similar people. Thus, these methods highly depends on the clustering result, and a hard quantization loss learnt can be prone to fit the noisy labels produced by clustering. To solve the problems, some multi-label learning methods [26,27] are proposed. Lin et al. [26] proposed an unsupervised Re-ID framework via softened similarity learning, not assign each sample with a hard label to avoid quantization loss as well as provide more room for the learning algorithm, and introduce auxiliary information to guide similarity estimation. However, it is still very challenging for unsupervised Re-ID to mine identity discrimination information since the drastic changes in the appearance of the same pedestrians and the high similarity in the appearance of different pedestrians without labeled pedestrian identity labels as learning guidance.

2.2. UDA Person Re-ID

For UDA Person Re-ID, assuming that there is a set of labeled source datasets and unlabeled target datasets. The common idea

Fig. 1. Some instance images taken from the DukeMTMC-reID (left) and Market1501 (right) datasets. Each row represents some images from the same view, and each column represents some images from different views of the same identity.

is to transfer the supervision information of the source domain to the unlabeled target domain. The current methods can be divided into two groups. In the first group, the methods deal with the UDA Re-ID problem by narrowing the gap between the source and target domains. Since dictionary learning is widely used in various fields [16,17,28–30], it is proposed to learn a shared subspace or dictionary based on dictionary learning between the source and target domains [31–33]. Peng et al. [31] proposed an asymmetric multi-task dictionary learning model, which decomposes the original visual feature into data-sharing and data-specific components to realize data transformation between different domains. Moreover, some researchers focus on GAN-based style transfer methods

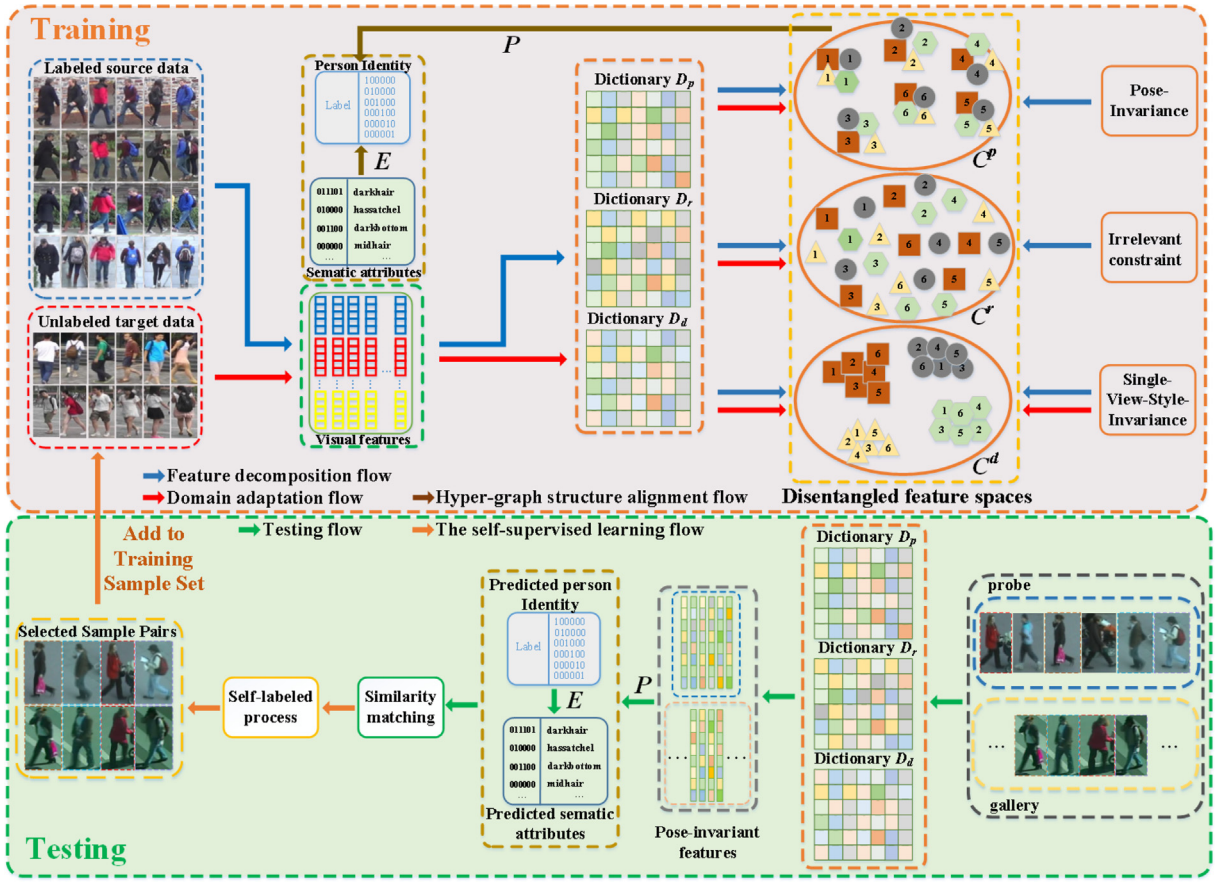

Fig. 2. Overview of the training and testing of the proposed method for the cross-domain Re-ID. Given training samples, we first extract visual features by references [14,15]. Then, visual features are disentangled by feature decomposition to generate three disentangled feature spaces $\boldsymbol{C}^p$, $\boldsymbol{C}^d$ and $\boldsymbol{C}^r$. The hypergraph structure alignment is performed to align the disentangled features $\boldsymbol{C}^p$ and attributes. And domain adaption is performed in the target domain to enhance the generalization ability of models. During testing, pose-invariant features, semantic attributes and identity are obtained by models trained during training to perform Re-ID. Moreover, a self-supervised learning process is introduced to fine-tune the model to better adapt to the target domain.

to narrow the gap between different domains [1,2,5,34]. For example, SPGAN [2] and PTGAN [5] first transfer image style from the source domain to the target domain, then use the transferred image for training. Recently, since the gaps of patches would be smaller than images among different datasets, Yang et al. [35] proposed a patch discriminative feature learning network (PatchNet) to learn discriminative features from patches instead of the whole images. However, the above methods try to reduce the inter-domain difference between different datasets, while ignoring intra-domain difference in a single domain, which is an important factor affecting Re-ID performance. In the second group, the methods dig the underlying data distribution information of the target domain for model refinement. Most methods [7–11,36] focus on target domain pseudo label estimation for self-training. Yu et al. [7] proposed a multilabel reference learning (MAR) method which evaluates the similarity of a pair of images by using the source dataset as a reference to mine hard negative samples. Lin et al. [9] proposed a bottom-up clustering approach, which jointly optimizes the relationship between CNN and unlabeled samples through training and merging clusters. Zhong et al. [15] proposed a framework that composes of a classification module and an exemplar memory module, which takes into account three underlying invariances in the target domain to learn robust and discriminative features, i.e., exemplar-invariance, camera-invariance, and neighborhood-invariance. Luo et al. [37] proposed a camera-aware invariance learning and cross-domain mixup scheme to takes both intra-domain and inter-domain variations into account. In references [15,37], the nearest neighbor search is performed on the target domain to mine the identity information of samples for intra-domain learning. However, the nearest neighbor search process is unreliable due to camera variations. Unlike them, our method does not require the identity information of target domain, only camera labels (known). Style invariance learning is performed in both source and target domains to remove camera style information, and the influence of pose variation is considered to jointly learn the visual feature space with pose-invariant and style-robust. However, the above methods only take the model pre-trained on the source domain as the initial model for the feature learning in the target domain and do not make full use of the labeled source datas as beneficial supervision in the adaptation process.

2.3. Attribute based Person Re-ID

The pedestrian attribute is regarded as a middle-level semantic feature representation, which includes local descriptions of pedestrians. Compared with traditional visual features with tens of thousands of dimensions, the dimension of attribute is low, usually only tens to hundreds of dimensions, so the discrimination is weak. However, due to the domain-invariance of attribute, it has attracted the attention of many scholars. Some researchers have proposed some attribute-based cross-domain Re-ID methods [3,4,38–41]. Wang et al. [4] proposed a transferable joint attribute-identity deep learning framework, which extracts identity-related attribute information through joint learning attribute and identity label data. Aiming at the problem of cross-class transfer learning, Peng et al. [38] proposed an attribute learning model based on dictionary learning, which combines user-defined semantic attribute, latent discriminant attribute, and background latent attribute to learn to deal with the problem of cross-domain Re-ID. However, these methods only regard semantic attribute as auxiliary information to assist the prediction of pedestrian identity, while semantic attribute contains local information of pedestrians, which is not fully utilized. In this paper, we use pedestrian identity information to construct hypergraph, and introduce a hypergraph structure alignment to establish the relationship between visual feature and attribute, fully mine the discriminant information in attribute, and further improve the performance of cross-domain Re-ID.

The proposed work is closely related to the UDA Person Re-ID. Different from the above methods that only consider the inter-domain variation between the source and target domains and ignore the intra-domain variation in a single domain, we regard each camera view as a domain to process the difference between the source and target domains at a finer level. We also consider the influence of pose variation on recognition and make full use of the supervision information of the source domain in the whole domain adaptation process. Moreover, due to the domain-invariance of attribute, we embed attribute into the dictionary learning framework and use them as a bridge for knowledge transfer between the source and target domains, rather than using them as auxiliary information for Re-ID.

3. The Proposed Approach

3.1. Problem Formulation

Assuming that the labeled source dataset contains K pedestrian identities and n_s labeled samples $\boldsymbol{S} = \{(\boldsymbol{x}_{ls}, \boldsymbol{a}_{ls}, \boldsymbol{y}_{ls}) | \boldsymbol{x}_{ls} \in \boldsymbol{X}_s, \boldsymbol{a}_{ls} \in \boldsymbol{A}_s, \boldsymbol{y}_{ls} \in \boldsymbol{Y}_s\}_{ls=1}^{n_s}$, where $\boldsymbol{x}_{ls} \in \mathbb{R}^d$ represents visual features of pedestrian images, $\boldsymbol{a}_{ls} \in \mathbb{R}^c$ represents semantic attributes of pedestrians, and $\boldsymbol{y}_{ls} \in \mathbb{R}^K$ represents pedestrian identity labels. Specifically, in the source dataset, $\boldsymbol{X}_s = [\boldsymbol{X}_{s,1}, \boldsymbol{X}_{s,2}, \ldots, \boldsymbol{X}_{s,i}, \ldots, \boldsymbol{X}_{s,K}] \in \mathbb{R}^{d \times n_s}$ denotes the feature sets of all images, $\boldsymbol{X}_{s,i} = [\boldsymbol{X}_{s,i,1}, \boldsymbol{X}_{s,i,2}, \ldots, \boldsymbol{X}_{s,i,v}, \ldots, \boldsymbol{X}_{s,i,V_s}] \in \mathbb{R}^{d \times n_{si}}$ denotes the feature sets of the i-th identity pedestrian, $\boldsymbol{X}_{s,i,v} = [\boldsymbol{x}_{s,i,v,1}, \boldsymbol{x}_{s,i,v,2}, \ldots, \boldsymbol{x}_{s,i,v,n_{siv}}] \in \mathbb{R}^{d \times n_{siv}}$ denotes the feature sets of the i-th identity pedestrians captured by the v-th camera, $\boldsymbol{X}_{s,v} = [\boldsymbol{X}_{s,1,v}, \boldsymbol{X}_{s,2,v}, \ldots, \boldsymbol{X}_{s,K,v}] \in \mathbb{R}^{d \times n_{sv}}$ denotes the feature sets of the pedestrians captured by the v-th camera, where K denotes the number of pedestrian identities, V_s presents the number of camera views, n_{si} denotes the number of pedestrians of the i-th identity, n_{siv} denotes the number of pedestrians of the i-th identity captured by the v-th camera, and n_{sv} denotes the number of pedestrians captured by the v-th camera in the source domain. $\boldsymbol{A}_s \in \mathbb{R}^{c \times n_s}$ and $\boldsymbol{Y}_s \in \mathbb{R}^{K \times n_s}$ are semantic attribute and identity label matrices corresponding to visual feature matrix $\boldsymbol{X}_s$, respectively. In addition, a disjointed target dataset $T = \{(x_{ut}, y_{ut}) | x_{ut} \in \boldsymbol{X}_t, \boldsymbol{y}_{ut} \in \boldsymbol{Y}_t\}_{ut=1}^{n_t}$ is given, where $Y_s \cap \boldsymbol{Y}_t = \varnothing$. In the target dataset, $\boldsymbol{X}_t = [\boldsymbol{X}_{t,1}, \boldsymbol{X}_{t,2}, \ldots, \boldsymbol{X}_{t,v}, \ldots, \boldsymbol{X}_{t,V_t}] \in \mathbb{R}^{d \times n_t}$ denotes the feature sets of all images, $\boldsymbol{X}_{t,v} = [\boldsymbol{x}_{t,v,1}, \boldsymbol{x}_{t,v,2}, \ldots, \boldsymbol{x}_{t,v,n_{tv}}] \in \mathbb{R}^{d \times n_{tv}}$ denotes the feature sets of pedestrians captured by the v-th camera, where V_t presents the number of camera views and n_{tv} denotes the number of pedestrians captured by the v-th camera in the target domain. $\boldsymbol{Y}_t$ is identity label matrix corresponding to visual feature matrix $\boldsymbol{X}_t$. For the target dataset only the visual feature $\boldsymbol{X}_t$ of the samples is known, and the identity label matrix Y_t of the samples is unknown. The target of Re-ID is to learn a classifier $f_{Re-ID} : \boldsymbol{X}_t \rightarrow \boldsymbol{Y}_t$.

3.2. Framework of Proposed Approach

The proposed framework includes three parts: feature decomposition, hypergraph structure alignment and domain adaptation.

3.2.1. Feature decomposition

In the cross-domain Re-ID, camera-style difference between different datasets is one of the main factors for the domain shift problem. Besides, pose variation among the same pedestrians is also one of the reasons that leads to difficulty in Re-ID. To solve these problems, we propose the following feature decomposition dictionary learning model. We believe that the original visual feature

is composed of different information components. Based on this consideration, we decompose the original visual feature $\boldsymbol{X}_s \in \mathbb{R}^{d\times n_s}$ into three spaces: pose-invariant feature space, camera-style feature space, and residual feature space:

$$\begin{aligned} L_{FD} = & \min_{\substack{\boldsymbol{D}_p,\boldsymbol{D}_d,\boldsymbol{D}_r,\\ \boldsymbol{C}_s^p,\boldsymbol{C}_s^d,\boldsymbol{C}_s^r}} \sum_{i=1}^{K}\sum_{v=1}^{V_s} \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{s,i,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_F^2 + \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p \\ & -\boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_* + \lambda_1(\|\boldsymbol{C}_{s,v}^d\|_{2,1} + \|\boldsymbol{C}_{s,i}^p\|_{2,1}) + \lambda_2\|\boldsymbol{Q} - (\boldsymbol{C}_s^p)^T\boldsymbol{C}_s^p\|_F^2 + \|\boldsymbol{I} \\ & - (\boldsymbol{D}_r\boldsymbol{C}_s^r)^T\boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 s.t\|\boldsymbol{d}_{p,i}\|_2^2 \\ & \le 1, \|\boldsymbol{d}_{d,j}\|_2^2 \le 1, \|\boldsymbol{d}_{r,k}\|_2^2 \le 1, \forall i,j,k \end{aligned} \tag{1}$$

where $\boldsymbol{X}_{s,i,v} \in \mathbb{R}^{d\times n_{siv}}$ represents the image feature sets of the i-th identity under the v-th view in the training dataset $\boldsymbol{S}$. $\boldsymbol{D}_p \in \mathbb{R}^{d\times d_p}, \boldsymbol{D}_d \in \mathbb{R}^{d\times d_d}$ and $\boldsymbol{D}_r \in \mathbb{R}^{d\times d_r}$ represent pose-invariant component dictionary, camera-style component dictionary and residual component dictionary respectively. While $\boldsymbol{C}_{s,i,v}^p \in \mathbb{R}^{d_p\times n_{siv}}, \boldsymbol{C}_{s,i,v}^d \in \mathbb{R}^{d_d\times n_{siv}}$ and $\boldsymbol{C}_{s,i,v}^r \in \mathbb{R}^{d_r\times n_{siv}}$ represent the pose-invariant feature, the camera-style feature and the residual feature corresponding to the original visual feature $\boldsymbol{X}_{s,i,v}$ respectively. $\boldsymbol{Q} \in \mathbb{R}^{n_s\times n_s}$ is a diagonal matrix, which is used to enhance the discriminance of the pose-invariant feature space. $\boldsymbol{Q}(i,j) = 1$ if the i-th and the j-th pose-invariant features are from the same person, otherwise $\boldsymbol{Q}(i,j) = 0$. $\boldsymbol{I} \in \mathbb{R}^{n_s\times n_s}$ is an identity matrix. $\|\cdot\|_*$ is the nuclear norm of the matrix, $\|\cdot\|_{2,1}$ represents the structured sparse norm. λ_1 and λ_2 are the regularization parameters.

The first item in Eq. (1) decomposes the original visual features into pose-invariant component, camera-style component and residual component, minimizing it can ensure the reconstruction ability of each component to the original visual features. To ensure the separation of different components, we impose specific constraints according to the spatial morphologies of each component. First, pedestrian images under the same view show a common visual style (as shown in each line of images in Fig. 1), thus the information shared by all pedestrians of the same view represents the camera-style information of the camera view, which shows a strong correlation, i.e., the camera-style feature space is low-rank. Based on this, a low-rank constraint (the second term in Eq. (1)) is imposed on the camera-style component to emphasize the camera style shared by images from the camera view. And pedestrian images from the same view should have similar camera-style features, which can be represented by the same atoms in $\boldsymbol{D}_d$. Thus, we introduce a single-view-style-invariance constraint(the third term in Eq. (1)) to push camera-style features of images from a single camera view close to each other. Second, in the pose-invariant feature space, pedestrian features with the same identity should be identity-related and not affected by pose variation, i.e., the pedestrian images with the same identity should have similar pose-invariant feature corresponding to $\boldsymbol{D}_p$, so we introduce the pose-invariance constraint (the fourth item in Eq. (1)) to promote the extraction of the pose-invariant feature. To ensure the preservation of identity information in the process of component decomposition, identity consistent constraint(the fifth term in Eq. (1)) is introduced, which makes the similarity between the pose-invariant features of pedestrians from the same identity as large as possible and the similarity between different identities as small as possible. Finally, the residual feature space includes some irregular interference information except pose and style information for Re-ID, irregular constrain (the sixth item in Eq. (1)) is applied to ensure the extraction of the residual features.

3.2.2. Hypergraph structure alignment

Considering the domain-invariance and robustness of attribute, we introduce attribute to help cross-domain Re-ID. Since the attribute of the target domain is unknown, only visual feature is known. If the relationship between visual feature and attribute can be established during training, we can predict attribute of the target domain by visual feature during testing. Therefore, we introduce hypergraph structure alignment to integrate pose-invariant feature space, semantic attribute space, and identity space into a unified framework. The loss function is expressed as follows:

$$\begin{aligned} & L_{HSA} = \min_{\boldsymbol{P},\boldsymbol{E}} \alpha_1(\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 + \|\boldsymbol{H} - \boldsymbol{E}\boldsymbol{A}_s\|_F^2) + \alpha_2 tr\left(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T\right) \\ & s.t\|\boldsymbol{p}_z\|_2^2 \le 1, \|\boldsymbol{e}_y\|_2^2 \le 1, \forall z,y \end{aligned} \tag{2}$$

where $\boldsymbol{H} \in \{0,1\}^{|E|\times|V|}$ represents a hypergraph association matrix. A hypergraph $G(HX, HE)$ is constructed through samples and pedestrian identities in the source domain, including a set of vertices $HX = [hx_1, hx_2, \ldots, hx_{|V|}]$ and a set of hyperedges $HE = [he_1, he_2, \ldots, he_{|E|}]$, where $|V|$ and $|E|$ represent the number of vertices and hyperedges respectively. Each vertex hx_i represents each pedestrian image in the source domain. Each hyperedge he_i represents a pedestrian identity in the source domain, which is a non-empty subset of vertices, i.e., pedestrian vertices with the same identity are located in the same hyperedge. Thus, $|E| = K$ and $|V| = n_s$. For any given hypergraph, its hyperedge can be easily transformed into an association matrix $\boldsymbol{H} \in \{0,1\}^{|E|\times|V|}$, where each row corresponds to a hyperedge and each column corresponds to a vertex. Specifically, if $hx_i \in he_j$, then $\boldsymbol{H}_{ij} = 1$; Otherwise $\boldsymbol{H}_{ij} = 0$. α_1 and α_2 are the regularization parameters. In the first and second terms of Eq. (2), pose-invariant features and attributes are aligned to the same identity structure through transformation matrices $\boldsymbol{P} \in \mathbb{R}^{K\times d_p}$ and $\boldsymbol{E} \in \mathbb{R}^{K\times c}$ respectively, which indirectly build the relationship between pose-invariant features and attributes, and their discrimination is ensured. Thus, attributes of the target domain can be predicted by visual features during testing. $tr(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T)$ represents Laplace regularization of hypergraph. If hx_i and hx_j represent the same person, we hope that the distance between the corresponding pose-invariant features $\boldsymbol{c}_{s,i}^p$ and $\boldsymbol{c}_{s,j}^p$ is as small as possible. The discrimination of pose-invariant features are further enhanced, and as our final pedestrian visual features. $\boldsymbol{L} = \boldsymbol{I}_e - \boldsymbol{W}$ represents the Laplace matrix of the hypergraph. $\boldsymbol{I}_e$ represents an identity matrix and $\boldsymbol{W}$ represents the weight matrix of the hypergraph, which is used to measure the degree of correlation between two vertices.

$$\boldsymbol{W} = \boldsymbol{D}_x^{-1/2}\boldsymbol{H}\boldsymbol{W}_e\boldsymbol{D}_e^{-1}\boldsymbol{H}^T\boldsymbol{D}_x^{-1/2} \tag{3}$$

where $\boldsymbol{D}_e$ and $\boldsymbol{D}_x$ represent the diagonal matrices of the degrees of the hyperedges and the degrees of the vertices, respectively. $\boldsymbol{W}_e$ represents a diagonal matrix of hyperedge weights (set to 1).

3.2.3. Domain adaptation

In the above learning process, only the data of the source domain is utilized. Due to the difference between the source and target domains, this may lead to a domain shift problem, making the trained model unable to adapt to the target domain well. To solve this problem, we introduce a domain adaptation term and use some unlabeled data of the target domain to participate in the training of the feature decomposition model. The loss function is expressed as follows:

$$\begin{aligned} L_{DA} = & \min_{\substack{\boldsymbol{D}_p,\boldsymbol{D}_d,\boldsymbol{D}_r,\\ \boldsymbol{C}_t^p,\boldsymbol{C}_t^d,\boldsymbol{C}_t^r}} \sum_{v=1}^{V_t} \|\boldsymbol{X}_{t,v} - \boldsymbol{D}_p\boldsymbol{C}_{t,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{t,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{t,v}^r\|_F^2 + \|\boldsymbol{X}_{t,v} - \boldsymbol{D}_p\boldsymbol{C}_{t,v}^p \\ & -\boldsymbol{D}_r\boldsymbol{C}_{t,v}^r\|_* + \lambda_1\|\boldsymbol{C}_{t,v}^d\|_{2,1} \\ & s.t\|\boldsymbol{d}_{p,i}\|_2^2 \le 1, \|\boldsymbol{d}_{d,j}\|_2^2 \le 1, \|\boldsymbol{d}_{r,k}\|_2^2 \le 1, \forall i,j,k \end{aligned} \tag{4}$$

where $\boldsymbol{X}_{t,v} \in \mathbb{R}^{d\times n_{tv}}$ represents the pedestrian image feature sets of the v-th view in the target dataset T. $\boldsymbol{C}_{t,v}^p \in \mathbb{R}^{d_p\times n_{tv}}, \boldsymbol{C}_{t,v}^d \in \mathbb{R}^{d_d\times n_{tv}}$

and $\boldsymbol{C}_{t,v}^{r} \in \mathbb{R}^{d_r \times n_{tv}}$ represent the coding coefficients of $\boldsymbol{X}_{t,v}$ corresponding to the three component dictionaries $\boldsymbol{D}_p, \boldsymbol{D}_d$ and $\boldsymbol{D}_r$ respectively. Finally, our entire objective function is expressed as:

$$L = L_{FD} + L_{HSA} + L_{DA} \tag{5}$$

3.3. Model Optimization

The overall loss function of the proposed method can be expressed as follows:

$$\begin{aligned} L \;=\; & \min_{\substack{\boldsymbol{D}_p,\boldsymbol{D}_d,\boldsymbol{D}_r,\boldsymbol{P},\boldsymbol{E},\\ \boldsymbol{C}_s^p,\boldsymbol{C}_s^d,\boldsymbol{C}_s^r,\\ \boldsymbol{C}_t^p,\boldsymbol{C}_t^d,\boldsymbol{C}_t^r}} \sum_{i=1}^{K}\sum_{v=1}^{V_s} \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{s,i,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_F^2 \\ & + \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p - \boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_* + \lambda_1(\|\boldsymbol{C}_{s,v}^d\|_{2,1} + \|\boldsymbol{C}_{s,i}^p\|_{2,1}) \\ & + \lambda_2\|\boldsymbol{Q} - (\boldsymbol{C}_s^p)^T\boldsymbol{C}_s^p\|_F^2 + \|\boldsymbol{I} - (\boldsymbol{D}_r\boldsymbol{C}_s^r)^T\boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 \\ & + \alpha_1(\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 + \|\boldsymbol{H} - \boldsymbol{E}\boldsymbol{A}_s\|_F^2) + \alpha_2 tr(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T) \\ & + \sum_{v=1}^{V_t}\|\boldsymbol{X}_{t,v} - \boldsymbol{D}_p\boldsymbol{C}_{t,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{t,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{t,v}^r\|_F^2 \\ & + \|\boldsymbol{X}_{t,v} - \boldsymbol{D}_p\boldsymbol{C}_{t,v}^p - \boldsymbol{D}_r\boldsymbol{C}_{t,v}^r\|_* + \lambda_1\|\boldsymbol{C}_{t,v}^d\|_{2,1} \\ & s.t \|\boldsymbol{d}_{p,i}\|_2^2 \le 1, \|\boldsymbol{d}_{d,j}\|_2^2 \le 1, \|\boldsymbol{d}_{r,k}\|_2^2 \le 1, \|\boldsymbol{p}_z\|_2^2 \le 1, \|\boldsymbol{e}_y\|_2^2 \le 1, \forall i,j,k,z,y \end{aligned} \tag{6}$$

The model (6) is non-convex for all variables, but convex for each of them separately. Therefore, we use alternating optimization algorithm to solve the above model.

1. Fix others, update $\boldsymbol{C}_s^p$. It is challenging if we directly solve $\boldsymbol{C}_s^p$ for the existence of $\|\boldsymbol{C}_{s,i}^p\|_{2,1}$. To solve this problem, we introduce a relaxation variable $\boldsymbol{T}_i = \boldsymbol{C}_{s,i}^p$, and represent as follows:

$$\begin{aligned} \{\boldsymbol{C}_s^p, \boldsymbol{T}_i\} \;=\; & \arg\min_{\boldsymbol{C}_s^p,\boldsymbol{T}_i} \sum_{i=1}^{K}\sum_{v=1}^{V_s} \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{s,i,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_F^2 \\ & + \|\boldsymbol{X}_{s,i,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,i,v}^p - \boldsymbol{D}_r\boldsymbol{C}_{s,i,v}^r\|_* + \lambda_1\|\boldsymbol{T}_i\|_{2,1} + \|\boldsymbol{T}_i - \boldsymbol{C}_{s,i}^p\|_F^2 \\ & + \lambda_2\|\boldsymbol{Q} - (\boldsymbol{C}_s^p)^T\boldsymbol{C}_s^p\|_F^2 + \alpha_1\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 + \alpha_2 tr(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T) \end{aligned} \tag{7}$$

the analytical solution to $\boldsymbol{T}_i$ can be obtained:

$$\boldsymbol{T}_i = (2\boldsymbol{I}_T - \lambda_1\boldsymbol{\Lambda})^{-1}(2\boldsymbol{C}_{s,i}^p) \tag{8}$$

where $\boldsymbol{I}_T \in \mathbb{R}^{n_{si}\times n_{si}}$ is an identity matrix, $\boldsymbol{\Lambda}$ is a diagonal matrix with the k-th diagonal element as $\lambda_{kk} = \frac{1}{\|(\boldsymbol{T}_i)_{k,:}\|_2}$. After obtaining $\boldsymbol{T} = [\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_K]$, we update $\boldsymbol{C}_s^p$ by solving:

$$\begin{aligned} \boldsymbol{C}_s^p \;=\; & \arg\min_{\boldsymbol{C}_s^p} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_* \\ & + \|\boldsymbol{T} - \boldsymbol{C}_s^p\|_F^2 + \lambda_2\|\boldsymbol{Q} - (\boldsymbol{C}_s^p)^T\boldsymbol{C}_s^p\|_F^2 + \alpha_1\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 + \alpha_2 tr(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T) \end{aligned} \tag{9}$$

For convenience, intermediate variables $\boldsymbol{K}_s = \boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r$ and $\boldsymbol{F}_s = (\boldsymbol{C}_s^p)^T$ are introduced, and Eq. (9) can be written as:

$$\begin{aligned} \{\boldsymbol{C}_s^p, \boldsymbol{K}_s, \boldsymbol{F}_s\} \;=\; & \arg\min_{\boldsymbol{C}_s^p,\boldsymbol{K}_s,\boldsymbol{F}_s} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{K}_s\|_* \\ & + \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r - \boldsymbol{K}_s\|_F^2 + \|\boldsymbol{T} - \boldsymbol{C}_s^p\|_F^2 + \lambda_2\|\boldsymbol{Q} - \boldsymbol{F}_s\boldsymbol{C}_s^p\|_F^2 \\ & + \|\boldsymbol{F}_s - (\boldsymbol{C}_s^p)^T\|_F^2 + \alpha_1\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 + \alpha_2 tr(\boldsymbol{C}_s^p\boldsymbol{L}(\boldsymbol{C}_s^p)^T) \end{aligned} \tag{10}$$

By alternating optimization, we first obtain $\boldsymbol{K}_s$ by using the singular value thresholding (SVT) algorithm [42]. And the analytical solution to $\boldsymbol{F}_s$ can be obtained:

$$\begin{aligned} \boldsymbol{F}_s = & (\lambda_2\boldsymbol{Q}(\boldsymbol{C}_s^p)^T + (\boldsymbol{C}_s^p)^T \\ & - \frac{1}{2}\alpha_2\boldsymbol{L}^T(\boldsymbol{C}_s^p)^T)(\lambda_2\boldsymbol{C}_s^p(\boldsymbol{C}_s^p)^T + \boldsymbol{I}_C)^{-1} \end{aligned} \tag{11}$$

where $\boldsymbol{I}_C \in \mathbb{R}^{d_p\times d_p}$ is an identity matrix. Finally, with the updated $\boldsymbol{K}_s$ and $\boldsymbol{F}_s$, the analytical solution to $\boldsymbol{C}_s^p$ can be obtained:

$$\begin{aligned} \boldsymbol{C}_s^p = & (2\boldsymbol{D}_p^T\boldsymbol{D}_p + 2\boldsymbol{I}_C + \alpha_1\boldsymbol{P}^T\boldsymbol{P} + \lambda_2\boldsymbol{F}_s^T\boldsymbol{F}_s)(2\boldsymbol{D}_p^T\boldsymbol{X}_s + \boldsymbol{T} \\ & + \alpha_1\boldsymbol{P}^T\boldsymbol{H} + \lambda_2\boldsymbol{F}_s^T\boldsymbol{Q} + \boldsymbol{F}_s^T - \boldsymbol{D}_p^T\boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_p^T\boldsymbol{D}_r\boldsymbol{C}_s^r \\ & - \boldsymbol{D}_p^T\boldsymbol{K}_s - \frac{1}{2}\alpha_2\boldsymbol{F}_s^T\boldsymbol{L}^T) \end{aligned} \tag{12}$$

2. Fix others, update $\boldsymbol{C}_s^d$, and get the following formula:

$$\boldsymbol{C}_{s,v}^d = \arg\min_{\boldsymbol{C}_{s,v}^d} \|\boldsymbol{X}_{s,v} - \boldsymbol{D}_p\boldsymbol{C}_{s,v}^p - \boldsymbol{D}_d\boldsymbol{C}_{s,v}^d - \boldsymbol{D}_r\boldsymbol{C}_{s,v}^r\|_F^2 + \lambda_1\|\boldsymbol{C}_{s,v}^d\|_{2,1} \tag{13}$$

the analytical solution to $\boldsymbol{C}_{s,v}^d$ can be obtained:

$$\boldsymbol{C}_{s,v}^d = (2\boldsymbol{D}_d^T\boldsymbol{D}_d + \lambda_1\boldsymbol{\Sigma})^{-1}(2\boldsymbol{D}_d^T\boldsymbol{X}_{s,v} - 2\boldsymbol{D}_d^T\boldsymbol{D}_r\boldsymbol{C}_{s,v}^r - \boldsymbol{D}_d^T\boldsymbol{D}_p\boldsymbol{C}_{s,v}^p) \tag{14}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with the k-th diagonal element as $\sigma_{kk} = \frac{1}{\|(\boldsymbol{C}_{s,v}^d)_{k,:}\|_2}$.

3. Fix others, we update $\boldsymbol{C}_s^r$ by introducing two variables $\boldsymbol{R}_s$ and $\boldsymbol{J}_s$, and get the objective function as:

$$\begin{aligned} \{\boldsymbol{C}_s^r, \boldsymbol{R}_s, \boldsymbol{J}_s\} \;=\; & \arg\min_{\boldsymbol{C}_s^r,\boldsymbol{R}_s,\boldsymbol{J}_s} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{R}_s\|_* \\ & + \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r - \boldsymbol{R}_s\|_F^2 + \|\boldsymbol{I} - \boldsymbol{J}_s\boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 \\ & + \|\boldsymbol{J}_s - (\boldsymbol{D}_r\boldsymbol{C}_s^r)^T\|_F^2 \end{aligned} \tag{15}$$

By alternating optimization, we can first obtain $\boldsymbol{R}_s$ by using singular value thresholding (SVT) algorithm [42], and then update $\boldsymbol{J}_s$ by solving:

$$\boldsymbol{J}_s = (2(\boldsymbol{C}_s^r)^T\boldsymbol{D}_r^T)(\boldsymbol{D}_r\boldsymbol{C}_s^r(\boldsymbol{C}_s^r)^T\boldsymbol{D}_r^T + \boldsymbol{I})^{-1} \tag{16}$$

Finally, with the updated $\boldsymbol{R}_s$ and $\boldsymbol{J}_s$, the analytical solution to $\boldsymbol{C}_s^r$ can be obtained:

$$\boldsymbol{C}_s^r = (3\boldsymbol{D}_r^T\boldsymbol{D}_r + \boldsymbol{D}_r^T\boldsymbol{J}_s^T\boldsymbol{J}_s\boldsymbol{D}_r)^{-1}(2\boldsymbol{D}_r^T\boldsymbol{X}_s + 2\boldsymbol{D}_r^T\boldsymbol{J}_s^T - 2\boldsymbol{D}_r^T\boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r^T\boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r^T\boldsymbol{R}_s) \tag{17}$$

4. Fix others, update $\boldsymbol{C}_t^p$. To solve this problem, we introduce a variable $\boldsymbol{K}_t$ to get the following formula:

$$\begin{aligned} \{\boldsymbol{C}_t^p, \boldsymbol{K}_t\} \;=\; & \arg\min_{\boldsymbol{C}_t^p,\boldsymbol{K}_t} \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 + \|\boldsymbol{K}_t\|_* \\ & + \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{K}_t\|_F^2 \end{aligned} \tag{18}$$

we can first obtain $\boldsymbol{K}_t$ by using singular value thresholding (SVT) algorithm [42]. We then update $\boldsymbol{C}_t^p$ by solving:

$$\boldsymbol{C}_t^p = \left(\tilde{D}_p{}^T\tilde{D}_p\right)^{-1}\tilde{D}_p{}^T\tilde{X}_t^p \tag{19}$$

where $\tilde{X}_t^p = \left[\boldsymbol{X}_t - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r; \boldsymbol{X}_t - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{K}_t\right], \tilde{\boldsymbol{D}}_p = [\boldsymbol{D}_p; \boldsymbol{D}_p]$

5. Fix others, update $\boldsymbol{C}_t^d$. We update $\boldsymbol{C}_t^d = [\boldsymbol{C}_{t,1}^d, \boldsymbol{C}_{t,2}^d, \ldots, \boldsymbol{C}_{t,v}^d, \ldots, \boldsymbol{C}_{t,V_t}^d]$ in the same way as Eq. (13), the analytical solution to $\boldsymbol{C}_{t,v}^d$ can be obtained:

$$\begin{aligned} \boldsymbol{C}_{t,v}^d = & (2\boldsymbol{D}_d^T\boldsymbol{D}_d + \lambda_1\boldsymbol{\Omega})^{-1}(2\boldsymbol{D}_d^T\boldsymbol{X}_{t,v} - 2\boldsymbol{D}_d^T\boldsymbol{D}_r\boldsymbol{C}_{t,v}^r \\ & - 2\boldsymbol{D}_d^T\boldsymbol{D}_p\boldsymbol{C}_{t,v}^p) \end{aligned} \tag{20}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with the k-th diagonal element as $\omega_{kk} = \frac{1}{\|(\boldsymbol{C}_{t,v}^d)_{k,:}\|_2}$.

6. Fix others, we update $\boldsymbol{C}_t^r$ by introducing a variable $\boldsymbol{R}_t$, and get the objective function as:

$$\begin{aligned} \{\boldsymbol{C}_t^r, \boldsymbol{R}_t\} \;=\; & \arg\min_{\boldsymbol{C}_t^r,\boldsymbol{R}_t} \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 + \|\boldsymbol{R}_t\|_* \\ & + \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{R}_t\|_F^2 \end{aligned} \tag{21}$$

We can first obtain $\boldsymbol{R}_t$ by using singular value thresholding (SVT) algorithm [42], and update $\boldsymbol{C}_t^r$ by solving:

$$\boldsymbol{C}_t^r = \left(\tilde{D}_r{}^T\tilde{D}_r\right)^{-1}\tilde{D}_r{}^T\tilde{X}_t^r \tag{22}$$

where $\tilde{X}_t^r = \left[\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d; \boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{R}_t\right], \tilde{\boldsymbol{D}}_r = [\boldsymbol{D}_r; \boldsymbol{D}_r]$.

7. Fix others, and update $\boldsymbol{D}_p$. To solve this problem, we introduce variables $\boldsymbol{B}_s$ and $\boldsymbol{B}_t$, and represent as follows:

$$\begin{aligned}\{\boldsymbol{D}_p, \boldsymbol{B}_s, \boldsymbol{B}_t\} &= \arg\min_{\boldsymbol{D}_p,\boldsymbol{B}_s,\boldsymbol{B}_t} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{B}_s\|_* + \|\boldsymbol{B}_t\|_* \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 + \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r - \boldsymbol{B}_s\|_F^2 \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{B}_t\|_F^2 \\ &s.t \|\boldsymbol{d}_{p,i}\|_2^2 \le \boldsymbol{1}, \forall \boldsymbol{i}\end{aligned} \tag{23}$$

where $\boldsymbol{B_s}$ and $\boldsymbol{B_t}$ can be obtained by using singular value thresholding (SVT) algorithm [42]. Thus, the solution to $\boldsymbol{D_p}$ can be obtained by solving:

$$\begin{aligned}\boldsymbol{D}_p &= \arg\min_{\boldsymbol{D}_p} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r \\ &-\boldsymbol{B}_s\|_F^2 \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 + \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{B}_t\|_F^2 \\ &s.t \|\boldsymbol{d}_{p,i}\|_2^2 \le 1, \forall i\end{aligned} \tag{24}$$

The optimal $\boldsymbol{D}_p$ can be obtained through the Lagrange dual [43].

8. Fix others, and update $\boldsymbol{D}_d$ by solving:

$$\begin{aligned}\boldsymbol{D}_d &= \arg\min_{\boldsymbol{D}_d} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 \\ &s.t \|\boldsymbol{d}_{d,j}\|_2^2 \le 1, \forall j\end{aligned} \tag{25}$$

The optimal $\boldsymbol{D}_d$ can be obtained through the Lagrange dual [43].

9. Fix others, and update $\boldsymbol{D}_r$. To solve this problem, we introduce variables $\boldsymbol{M}_s, \boldsymbol{M}_t$ and $\boldsymbol{V}$ as:

$$\begin{aligned}\{\boldsymbol{D}_r, \boldsymbol{M}_s, \boldsymbol{M}_t, \boldsymbol{V}\} &= \arg\min_{\boldsymbol{D}_r,\boldsymbol{M}_s,\boldsymbol{M}_t} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{M}_s\|_* \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 + \|\boldsymbol{I} - \boldsymbol{V}^T\boldsymbol{V}\|_F^2 \\ &+\|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r - \boldsymbol{M}_s\|_F^2 + \|\boldsymbol{V} - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{M}_t\|_F^2 + \|\boldsymbol{M}_t\|_* \\ &s.t \|\boldsymbol{d}_{r,k}\|_2^2 \le 1, \forall k\end{aligned} \tag{26}$$

where $\boldsymbol{M}_s$ and $\boldsymbol{M}_t$ can be obtained by using singular value thresholding (SVT) algorithm [42]. Moreover, the solution to $\boldsymbol{V}$ can be obtained by solving:

$$\boldsymbol{V} = \arg\min_{\boldsymbol{V}} \|\boldsymbol{V} - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{I} - \boldsymbol{V}^T\boldsymbol{V}\|_F^2 \tag{27}$$

It is challenging if we directly solve the problem for the existence of $\boldsymbol{V}^T\boldsymbol{V}$. To solve this problem, we introduce a relaxation variable $\boldsymbol{G} = \boldsymbol{V}^T$, and relax Eq. (27) as follows:

$$\{\boldsymbol{V}, \boldsymbol{G}\} = \arg\min_{\boldsymbol{V},\boldsymbol{G}} \|\boldsymbol{V} - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{I} - \boldsymbol{G}\boldsymbol{V}\|_F^2 + \|\boldsymbol{G} - \boldsymbol{V}^T\|_F^2 \tag{28}$$

All terms in Eq. (28) are characterized by Frobenius norm, thus the analytical solution to $\boldsymbol{G}$ and $\boldsymbol{V}$ can be obtained:

$$\boldsymbol{G} = 2\boldsymbol{V}^T\left(\boldsymbol{I} + \boldsymbol{V}\boldsymbol{V}^T\right)^{-1} \tag{29}$$

and

$$\boldsymbol{V} = \left(\boldsymbol{G}^T\boldsymbol{G} + 2\boldsymbol{I}\right)^{-1}\left(\boldsymbol{D}_r\boldsymbol{C}_s^r + 2\boldsymbol{G}^T\right) \tag{30}$$

Finally, we update $\boldsymbol{D}_r$ by solving:

$$\begin{aligned}\boldsymbol{D}_r &= \arg\min_{\boldsymbol{D}_r} \|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_d\boldsymbol{C}_s^d - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 + \|\boldsymbol{V} - \boldsymbol{D}_r\boldsymbol{C}_s^r\|_F^2 \\ &+\|\boldsymbol{X}_s - \boldsymbol{D}_p\boldsymbol{C}_s^p - \boldsymbol{D}_r\boldsymbol{C}_s^r - \boldsymbol{M}_s\|_F^2 + \|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_r\boldsymbol{C}_t^r - \boldsymbol{M}_t\|_F^2 \\ &+\|\boldsymbol{X}_t - \boldsymbol{D}_p\boldsymbol{C}_t^p - \boldsymbol{D}_d\boldsymbol{C}_t^d - \boldsymbol{D}_r\boldsymbol{C}_t^r\|_F^2 \\ &s.t \|\boldsymbol{d}_{r,k}\|_2^2 \le 1, \forall k\end{aligned} \tag{31}$$

The optimal $\boldsymbol{D}_r$ can be obtained through the Lagrange dual [43].

10. Fix others and update $\boldsymbol{P}$ and $\boldsymbol{E}$ respectively, the subproblem can be formulated as:

$$\boldsymbol{P} = \arg\min_{\boldsymbol{P}} \alpha_1\|\boldsymbol{H} - \boldsymbol{P}\boldsymbol{C}_s^p\|_F^2 \qquad s.t \|\boldsymbol{p}_z\|_2^2 \le 1, \forall z \tag{32}$$

and

$$\boldsymbol{E} = \arg\min_{\boldsymbol{E}} \alpha_1\|\boldsymbol{H} - \boldsymbol{E}\boldsymbol{A}_s\|_F^2 \qquad s.t \|\boldsymbol{e}_y\|_2^2 \le 1, \forall y \tag{33}$$

The optimal $\boldsymbol{P}$ and $\boldsymbol{E}$ can be obtained through the Lagrange dual [43].

For clarity, we summarize the complete procedure for solving the objective function (6) in Algorithm 1.

Algorithm 1: Algorithm of cross-domain Re-ID with pose-invariant matrix factorization and hypergraph structure alignment.

Input: $\boldsymbol{X}_s, \boldsymbol{X}_t, \boldsymbol{Y}_s, \boldsymbol{A}_s$.
1: Initialize $\boldsymbol{D}_p, \boldsymbol{D}_d, \boldsymbol{D}_r, \boldsymbol{P}, \boldsymbol{E}$ randomly.
2: **while** not converged **do**
3: Update $\boldsymbol{C}_s^p$ by solving (12)
4: Update $\boldsymbol{C}_s^d$ by solving (14)
5: Update $\boldsymbol{C}_s^r$ by solving (17)
6: Update $\boldsymbol{C}_t^p$ by solving (19)
7: Update $\boldsymbol{C}_t^d$ by solving (20)
8: Update $\boldsymbol{C}_t^r$ by solving (22)
9: Update $\boldsymbol{D}_p$ by solving (24)
10: Update $\boldsymbol{D}_d$ by solving (25)
11: Update $\boldsymbol{D}_r$ by solving (31)
12: Update $\boldsymbol{P}$ by solving (32)
13: Update $\boldsymbol{E}$ by solving (33)
14: **end while**
Output $\boldsymbol{D}_p, \boldsymbol{D}_d, \boldsymbol{D}_r, \boldsymbol{P}, \boldsymbol{E}$.

3.4. Person Re-identification

In the above training process, we have obtained the pose-invariant dictionary $\boldsymbol{D}_p$, the camera-style component dictionary $\boldsymbol{D}_d$, the residual component dictionary $\boldsymbol{D}_r$, and the transformation matrices $\boldsymbol{P}$ and $\boldsymbol{E}$. With these dictionaries, we can calculate the pose-invariant feature $\boldsymbol{c}_{t,h}^p$, the camera-style feature $\boldsymbol{c}_{t,h}^d$ and the residual feature $\boldsymbol{c}_{t,h}^r$ of each test sample $\boldsymbol{x}_{t,h}$ in the target domain through the following formula:

$$\begin{aligned}\{\boldsymbol{c}_{t,h}^p, \boldsymbol{c}_{t,h}^d, \boldsymbol{c}_{t,h}^r\} &= \arg\min_{\boldsymbol{c}_{t,h}^p,\boldsymbol{c}_{t,h}^d,\boldsymbol{c}_{t,h}^r} \|\boldsymbol{x}_{t,h} - \boldsymbol{D}_p\boldsymbol{c}_{t,h}^p - \boldsymbol{D}_d\boldsymbol{c}_{t,h}^d - \boldsymbol{D}_r\boldsymbol{c}_{t,h}^r\|_2^2 \\ &+\epsilon(\|\boldsymbol{c}_{t,h}^p\|_2^2 + \|\boldsymbol{c}_{t,h}^d\|_2^2 + \|\boldsymbol{c}_{t,h}^r\|_2^2)\end{aligned} \tag{34}$$

When the pose-invariant feature $\boldsymbol{c}_{t,h}^p$ is calculated, we can predict identity representation $\boldsymbol{h}_{t,h}$ and semantic attribute $\boldsymbol{a}_{t,h}$ through Eqs. (35) and (36):

$$\boldsymbol{h}_{t,h} = \boldsymbol{P}\boldsymbol{c}_{t,h}^p \tag{35}$$

and

$$\boldsymbol{a}_{t,h} = \arg\min_{\boldsymbol{a}_{t,h}} \|\boldsymbol{h}_{t,h} - \boldsymbol{E}\boldsymbol{a}_{t,h}\|_2^2 \tag{36}$$

For the test sample, with $\boldsymbol{h}_{t,h}$ and $\boldsymbol{a}_{t,h}$, we can calculate the similarity scores Sim_h and Sim_a of pedestrian image pairs in identity space and semantic space respectively through Eq. (37):

$$Sim(\boldsymbol{z}_a, \boldsymbol{z}_b) = \frac{\boldsymbol{z}_a^T \bullet \boldsymbol{z}_b}{\|\boldsymbol{z}_a\|_2^2 \bullet \|\boldsymbol{z}_b\|_2^2 \bullet (\|\boldsymbol{z}_a - \boldsymbol{z}_b\|_2^2 + \varepsilon)} \tag{37}$$

where $\boldsymbol{z}_l, (l = a, b)$ represents a vector ($\boldsymbol{h}_{t,h}$ or $\boldsymbol{a}_{t,h}$) in identity space or semantic space. $\varepsilon > 0$ is a small constant to avoid being divided

by zero. Since the identity space is discriminative and the semantic space is domain-shared, combining multiple spaces can effectively improve the performance of cross-domain Re-ID. We calculate the similarity score by:

$$Sim_{final} = \tau Sim_a + (1-\tau)Sim_h \tag{38}$$

where $\tau \geqslant 0$ is the weight. Since the identity space has strong discrimination, while the semantic space has domain sharing, but the discrimination is slightly weak, we set $\tau = 0.3$. This work introduces a self-supervised learning strategy to select reliable sample pairs in the target domain and fine-tune the model to better adapt to the target domain. We adopt the same selection strategy as in reference [44].

4. Experiments

4.1. Datasets and Settings

In order to verify the effectiveness of the algorithm, we carry out multiple experiments on eight benchmark Re-ID datasets, including VIPeR [45], PRID2011 [46], GRID [47], PRID450s [48], CUHK01 [49], DukeMTMC-reID [12], Market1501 [13] and MSMT17 [5]. A detailed description of these datasets is provided in Table 1. For VIPeR, PRID2011, PRID450s and CUHK01, all samples of each dataset are used to train the model when they are served as the source datasets. When they are served as the target datasets, each one of them is divided into two parts with non-overlapping identities. One part is used to train the model with the source datasets, and the other is used for testing. The process is repeated 10 times, and the average value of 10 times is taken as the final rusult. For DukeMTMC-reID, Market1501 and MSMT17, we follow the separation protocol in references [5,12,13]. Only the training part of each dataset is used to train the model when the aforementioned three datasets are served as the source. When they are served as the target, the test part of each dataset is used to test the model.

In the experiment, we use the Gaussian of Gaussian (GOG) [14] descriptor as the appearance visual features of pedestrians. In the above datasets, Layne et al. [50] only provides attribute annotation for VIPeR, PRID2011, GRID. In addition, DukeMTMC-reID and Market1501 datasets carry attribute annotation. Therefore, the datasets with attribute annotation can be used as both training dataset and test dataset, while the datasets without attribute annotation can only be used as test dataset. In the proposed model, there are 8 parameters to be set, including the atomic sizes d_p, d_d, d_r of dictionaries $\boldsymbol{D}_p, \boldsymbol{D}_d, \boldsymbol{D}_r$, and the regularization term parameters $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ and ϵ. In the experiment, we set these parameters as $d_p = 600$,

$$d_d = 180, d_r = 180, \lambda_1 = 0.0001, \lambda_2 = 0.01, \alpha_1 = 0.1, \alpha_2 = 0.1$$

and $\epsilon = 0.1$ respectively. More details on how to select parameters will be shown in Section 5.4. Performance is evaluated by the Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP).

4.2. Results and Discussion

In this section, we compare the proposed method with some popular methods in recent years to verify its effectiveness. The performance of the method on several Re-ID datasets and the comparison results with different methods are discussed.

4.2.1. Experiments on VIPeR

We use PRID2011 and GRID as the source datasets and VIPeR as the target dataset for experiments. Some methods are selected to compare with our methods, including Adversarial (2016) [51], UMDL (2016) [31], SDC (2017) [52], CAMEL (2017) [24], UJSDL (2018) [32], AIESL (2020) [44], and SSAE (2020) [53]. The comparison results are shown in Table 2. As can be seen from Table 2, our method is superior to other methods in Rank-1, Rank-5 and Rank-10, including some deep learning methods Adversarial (2016) [51], CAMEL (2017) [24], which proves the effectiveness and superiority of the proposed method.

4.2.2. Experiments on PRID2011

VIPeR is the source dataset and PRID2011 is the target dataset. The proposed method is compared with some state-of-the-art methods like GL (2016) [54], UMDL (2016) [31], SSDAL (2016) [55], TJ-AIDL (2018) [4], JSLAM (2018) [38], MMFA (2018) [3], PTGAN(2018) [5], ATNet (2019) [41], AIESL (2020) [44], SSAE (2020) [53], UDA-TP(2020) [8], ACT(2020) [56] and MMT(2020) [10]. The results are shown in Table 3. As shown in Table 3, our method reaches 38.20% in Rank-1 accuracy, is superior to other methods, 11.4%, 3.1% and 14.2% higher in Rank-1 accuracy than some methods based on deep learning TJ-AIDL (2018) [4], MMFA (2018) [3] and ATNet (2019) [41], respectively.

4.2.3. Experiments on PRID450s

VIPeR is the source dataset and PRID450s is the target dataset. Some recent unsupervised cross-domain Re-ID methods are compared, including AdaRSVMs (2015) [57], cMAT-DCA (2016) [33], UMDL (2016) [31], TSR (2017) [58], UJSDL (2018) [32], and SSAE (2020) [53]. As shown in Table 4, the proposed method achieves the optimal performance on the PRID450s dataset, which outperforms the second best SSAE (2020) [53] an improvement of 2.52%, 3.40%, and 0.97% at Rank-1, Rank-5, and Rank-10 respectively.

4.2.4. Experiments on CUHK01

VIPeR is the source dataset and CUHK01 is the target dataset. Compared with some methods in recent years, such as UDML (2016) [31], TSR (2017) [58], CAMEL (2017) [24], DAS (2018) [1], UJSDL (2018) [32], PN-GAN(2018) [59], AIESL (2020) [44] and MFAGL (2020) [60], the comparison results are shown in Table 5. From the results, it can be seen that the proposed method achieves

Table 1
A detailed description of some datasets for experiment.

Dataset	IDs	Cams	Imgs	TrainIDs/Imgs	TestIDs	query	gallery
VIPeR [45]	632	2	1264	316/632	316	316	316
PRID2011 [46]	200	2	949	100/200	100	100	649
GRID [47]	250	2	1275	125/250	125	125	900
PRID450s [48]	450	2	900	225/450	225	225	225
CUHK01 [49]	971	2	3884	486/972	485	970	970
Market1501 [13]	1501	6	32217	751/12936	750	3368	19732
DukeMTMC-reID [12]	1812	8	36441	702/16522	702	2228	17661
MSMT17 [5]	4101	15	126441	1041/32621	3060	11659	82161

Table 2
Performance comparison with some competing methods on the VIPeR for cross-dataset person Re-ID. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
Adversarial (2016) [51]	22.80	38.60	50.30	**63.90**
UMDL (2016) [31]	31.50	-	-	-
SDC (2017) [52]	25.80	-	-	-
CAMEL (2017) [24]	30.90	-	-	-
UJSDL (2018) [32]	20.30	38.04	49.11	60.38
AIESL (2020) [44]	28.92	40.41	46.58	52.44
SSAE (2020) [53]	26.84	39.72	49.27	60.38
Proposed	**31.80**	**46.30**	**54.81**	63.26

the optimal performance. Moreover, compared with the deep learning methods, the proposed method does not need a large number of training samples, which proves the advantages of the proposed method.

4.2.5. Experiments on Large-scale Datasets

In order to further prove the effectiveness of the proposed model, we also evaluate the proposed method on several large-scale Re-ID datasets, i.e. Market1501 (Market), DukeMTMC-reID (Duke), MSMT17. The purpose of the self-supervised learning strategy is to select reliable positive sample pairs in the target domain to adjust the model. However, for large-scale datasets, the number of positive sample pairs accounts for a very small proportion of the total. In this case, self-supervised learning strategies may play a negative role. Therefore, we remove the self-supervised learning strategy for large-scale datasets. We conduct experiments on Market with Duke as the source dataset, Duke with Market as the source dataset and MSMT17 with Market or Duke as the source dataset, and combine GOG with deep features learned from reference [15] as the feature representation of pedestrians (Baseline). We compare the proposed method with recent popular cross-domain Re-ID methods through the datasets Market, Duke and MSMT17, the results are summarized in Table 6 and Table 7. On the Market dataset, the mAP and Rank-1 accuracy of the proposed method are 77.20% and 43.49%, respectively. It improves the Baseline by 5.50% and 1.62% in terms of Rank-1 accuracy and mAP. The proposed method outperforms all the methods compared. Besides, it achieves 65.48% and 40.82% on Rank-1 accuracy and mAP,respectively, which surpasses the Baseline by 9.18% and 3.90% in terms of Rank-1 accuracy and mAP on the Duke dataset. The proposed method achieves 30.84% and 25.40% of the Rank-1 accuracy on Duke→MSMT17 and Market→MSMT17. To summarize, the proposed method achieves promising results on the three public datasets, and the experiment results on large-scale datasets validate the its superiority.

Table 3
Performance comparison with some competing methods on the PRID2011 for cross-dataset person Re-ID. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
GL (2016) [54]	25.00	-	-	-
UMDL (2016) [31]	24.20	-	-	-
SSDAL (2016) [55]	20.10	-	-	-
TJ-AIDL (2018) [4]	26.80	-	-	-
JSLAM (2018) [38]	25.60	-	-	-
MMFA (2018) [3]	35.10	-	-	-
PTGAN(2018) [5]	33.50	-	-	-
ATNet (2019) [41]	24.00	-	-	-
AIESL (2020) [44]	33.70	51.10	57.30	65.70
SSAE (2020) [53]	29.10	48.40	55.70	64.80
UDA-TP(2020) [8]	22.00	-	-	-
ACT(2020) [56]	24.00	-	-	-
MMT(2020) [10]	25.00	-	-	-
Proposed	**38.20**	**56.30**	**64.10**	**72.00**

Table 4
Performance comparison with some competing methods on the PRID450s for cross-dataset person Re-ID. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
AdaRSVMs (2015) [57]	27.78	44.98	53.69	63.89
cMAT-DCA (2016) [33]	12.31	12.31	39.16	53.56
UMDL (2016) [31]	31.50	66.22	76.36	84.93
TSR (2017) [58]	29.00	49.40	58.40	69.80
UJSDL (2018) [32]	45.78	68.00	78.22	85.78
SSAE (2020) [53]	61.64	76.36	84.89	91.56
Proposed	**64.16**	**79.76**	**85.86**	**91.56**

5. Analysis of Proposed Algorithm

5.1. Convergence Analysis

As mentioned in Section 3.4, for Eq. (6), the model is non-convex for all variables, but when we fix all the others and change one of them, the model is convex. Therefore, we adopt alternating optimization algorithm to make the above model converge to the optimal solution. Fig. 3 shows the convergence curve of each variable $\{\boldsymbol{D}_p, \boldsymbol{D}_d, \boldsymbol{D}_r, \boldsymbol{P}, \boldsymbol{E}\}$ on the VIPeR dataset. As can be seen from Fig. 3, when the number of iterations reaches 20, all variables can converge to a stable solution. Therefore, in the experiment, we set the training iteration number of Algorithm 1 to 20.

5.2. Complexity Analysis

The complexity of the proposed method mainly comes from solving the nuclear norm through singular value thresholding (SVT) algorithm. During training, Eqs. (10), (15), (18), (21), (23) and (26) all involve the solution of the nuclear norms. In addition, the number of training iterations is K_1, so the time complexity during training is $O(K_1(d^2n_s + d^2n_t))$. During the testing, the complexity of the algorithm is mainly determined by the similarity measurement between samples, and its complexity is $O(pq)$, p and q are the numbers of pedestrians in the probe and gallery set, respectively. In addition, we also introduce a self-supervised learning process. Except for the first training and testing, each self-supervised learning process includes a training process and a testing process. We assume that the number of self-supervised learning is K_2. Therefore, we need $K_2 + 1$ training and testing, so the total time complexity of the proposed algorithm is $O((K_2 + 1)(K_1(d^2n_s + d^2n_t) + pq)$.

Table 5
Performance comparison with some competing methods on the CUHK01 for cross-dataset person Re-ID. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
UDML (2016) [31]	27.10	-	-	-
TSR (2017) [58]	22.40	35.90	47.90	64.50
CAMEL (2017) [24]	57.30	-	-	-
DAS (2018) [1]	54.90	-	-	-
UJSDL (2018) [32]	27.74	48.81	57.71	66.85
PN-GAN(2018) [59]	27.58	49.17	59.57	–
AIESL (2020) [44]	63.26	81.63	87.36	91.84
MFAGL(2020) [60]	58.10	-	-	-
Proposed	**64.76**	**82.68**	**89.20**	**93.63**

Table 6
Performance comparison with some recent methods on the Market1501 and DukeMTMC-reID datasets. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Duke→Market			Market→Duke		
	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP
Baseline	71.70	83.63	41.87	56.30	70.58	36.92
UDML (2016) [31]	34.50	52.60	12.40	18.50	31.40	7.30
PTGAN(2018) [5]	38.60	-	-	27.40	-	-
SPGAN(2018) [2]	51.50	70.10	22.80	41.10	56.60	22.30
TJ-AIDL (2018) [4]	58.20	74.80	26.50	44.30	59.60	23.0
CamStyle (2019) [61]	58.80	78.20	27.40	48.40	62.50	25.10
ECN(2019) [15]	75.10	87.60	43.0	63.30	75.80	40.40
CDIL (2020) [62]	57.20	73.0	27.40	-	-	-
CSGLP (2020) [63]	61.20	77.50	31.50	47.80	62.30	27.10
UADA-SD(2021) [64]	57.40	72.40	30.20	45.30	57.80	30.30
3D-GAT(2021) [65]	59.40	75.20	28.60	45.10	59.30	26.10
STReID(2021) [66]	62.30	79.10	31.60	52.30	65.90	29.20
EDAAN(2021) [67]	64.50	83.0	35.40	57.80	72.20	39.60
CAC–CSP(2021) [68]	69.40	82.80	36.90	57.50	71.20	37.0
Proposed	**77.20**	**89.43**	**43.49**	**65.48**	**77.15**	**40.82**

Table 7
Performance comparison with some recent methods on the MSMT17 dataset. The cumulative matching rate(%) are listed. '-' denotes that no reported result is available.

Methods	Source	MSMT17			
		Rank-1	Rank-5	Rank-10	mAP
Baseline	Duke	27.13	38.85	44.27	9.41
PTGAN(2018) [5]	Duke	11.80	-	27.40	3.30
ECN(2019) [15]	Duke	30.20	-	-	10.20
UADA-SD(2021) [64]	Duke	24.20	33.60	37.80	**11.70**
Proposed	Duke	**30.84**	**42.38**	**47.87**	10.29
Baseline	Market	22.30	32.81	37.89	7.73
PTGAN(2018) [5]	Market	10.20	-	24.40	2.90
ECN(2019) [15]	Market	25.30	-	-	8.50
UADA-SD(2021) [64]	Market	21.30	30.60	35.30	**11.30**
Proposed	Market	**25.40**	**36.40**	**42.07**	8.50

5.3. Ablation Study

In order to prove the effectiveness of each regularization term in the proposed model, we compare the performance of the sub-model after deleting one of the terms with that of the final model on the VIPeR dataset, and evaluate the comparison results through CMC curves. The model (6) includes the following regularization terms: Pose-Invariance Constraint $\|\boldsymbol{C}^p_{s,i}\|_{2,1}$ (PI), Single-View-Style-Invariance Constraint $\|\boldsymbol{C}^d_{s,v}\|_{2,1}$ (SVSI), Identity Consistent Constraint $\|\boldsymbol{Q} - (\boldsymbol{C}^p_s)^T\boldsymbol{C}^p_s\|^2_F$ (IC), Residual Component Constraint $\|\boldsymbol{I} - (\boldsymbol{D}_r\boldsymbol{C}^r_s)^T\boldsymbol{D}_r\boldsymbol{C}^r_s\|^2_F$ (RC), and Laplace Constraint $tr(\boldsymbol{C}^p_s\boldsymbol{L}(\boldsymbol{C}^p_s)^T)$ (LC). The comparison results are shown in Fig. 4. As can be seen from Fig. 4, the performance will decline more or less after deleting each item, proving that each regularization item has a certain positive effect on the final performance. Among them, Single-View-Style-Invariance Constraint(SVSI) plays a great role, which may be due to the fact that SVSI Constraint can greatly promote the separation of camera-style information, narrow the gap between different domains, and effectively alleviate the domain shift problem.

5.4. Parameters Selection

In the proposed model (6), there are 8 parameter $d_p, d_d, d_r, \lambda_1, \lambda_2, \alpha_1, \alpha_2$ and ϵ need to be set. In the experiment part, we give the specific value of each parameter. In this section, we

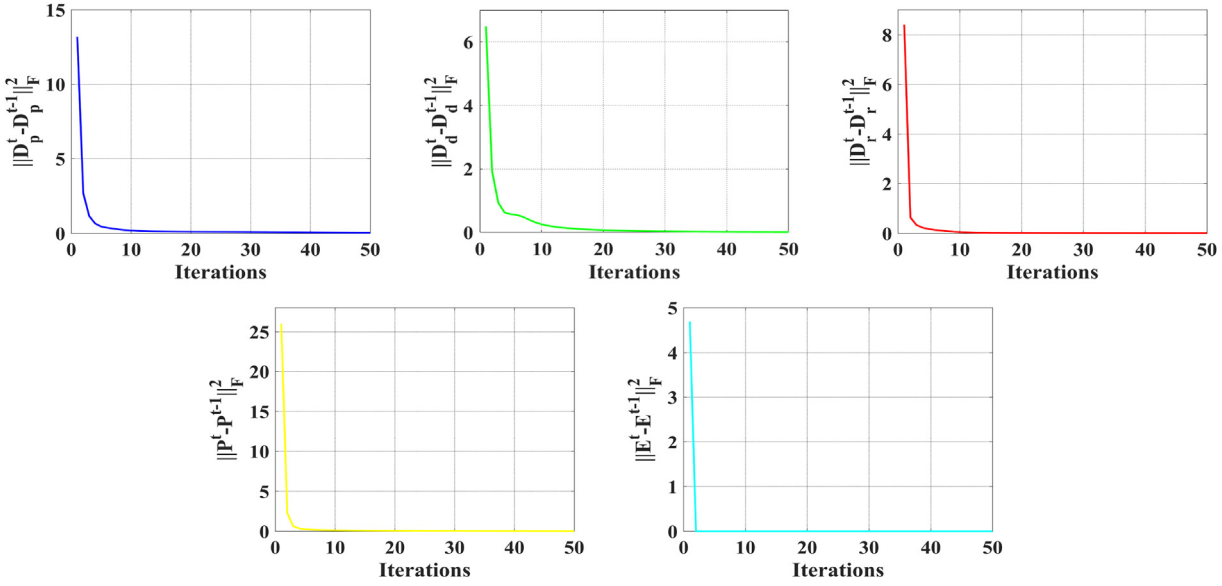

Fig. 3. The convergence analysis of the proposed method on VIPeR dataset.

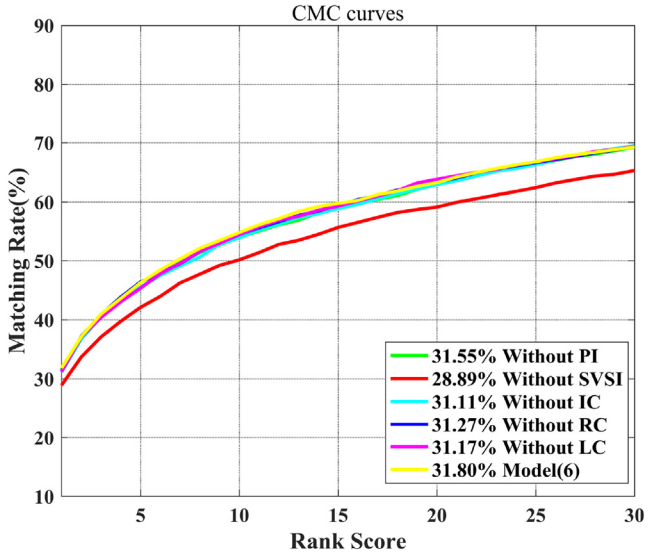

Fig. 4. Ablation analysis of the proposed model; Comparisons of CMC curves of different models on VIPeR dataset.

will discuss how the values of these parameters are selected and the role of each parameter. We discuss the influence of parameters by changing one of the parameters and fixing other parameters. Specifically, we use PRID2011 and GRID as the source dataset and VIPeR as the target verification set. For each parameter, we set different values within a certain range to observe its impact on performance and fix other parameters to the values given in Section 4.1. Fig. 5 shows the range of each parameter setting, as well as the Rank-1 identification accuracy and CMC curve corresponding to each value.

As can be seen from Fig. 5, the atomic sizes d_p, d_d and d_r of dictionaries $\boldsymbol{D}_p, \boldsymbol{D}_d$ and $\boldsymbol{D}_r$ have little influence on the performance within the corresponding parameter range, but when $d_p = 600, d_d = 180$ and $d_r = 180$ respectively, the proposed method performs better. In addition, parameters λ_2 and α_1 have little influence on the performance of the model. We set $\lambda_2 = 0.01$ and $\alpha_1 = 0.1$ respectively. However, the parameters λ_1, α_2 and ϵ have great influence on the performance. Only when these parameters are set within a certain range ($\lambda_1 \leq 0.01, \alpha_2 \leq 1, 0.01 \leqslant \epsilon \leq 1$) can the proposed method obtain better performance. In order to achieve satisfactory performance, we set these parameters to $\lambda_1 = 0.0001, \alpha_2 = 0.1, \epsilon = 0.1$. Therefore, all parameter settings are consistent with Section 4.1.

6. Conclusion

Aiming at the problem of cross-domain Re-ID, we proposed an effective dictionary learning algorithm. In order to solve the domain shift problem caused by the difference of camera styles, we formulated the cross-domain Re-ID problem as a feature

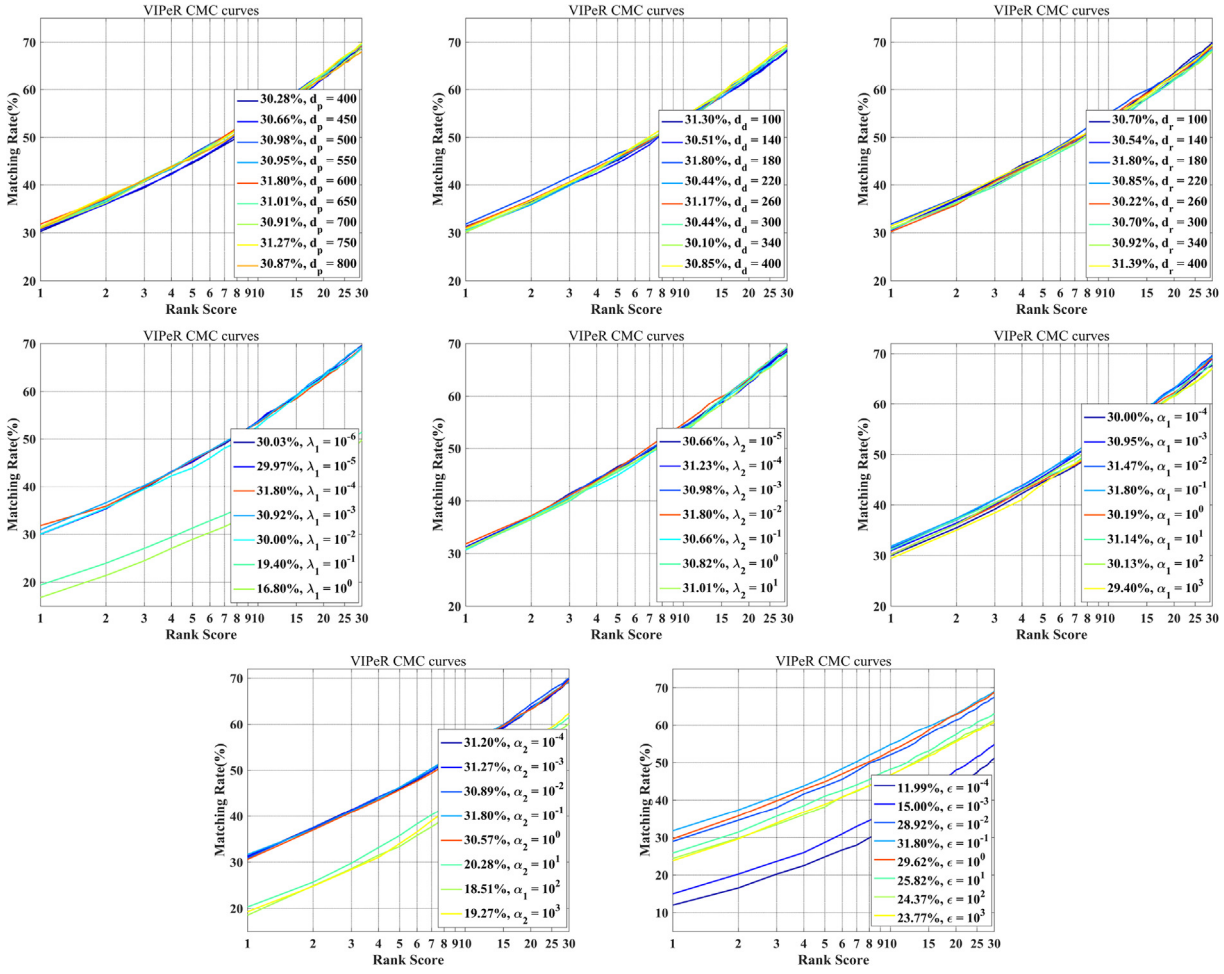

Fig. 5. Parameter sensitivity analysis on the VIPeR dataset.

decomposition problem and realized the separation of pedestrian identity related information, camera style information, pedestrian pose and other interference information. Then robust pedestrian feature for Re-ID was extracted. Moreover, attribute was introduced to further mitigate the domain shift problem. The experimental results showed that the proposed model outperforms the related unsupervised cross-domain Re-ID models in the effectiveness and practicability. Although the results imply that the proposed approach is superior to many cross-domain person Re-ID methods, including some methods based on deep learning, it lacks flexibility and is difficult to solve the problems of pedestrian occlusion and clothes change. For future study, we may focus on the Re-ID problem under occlusion or clothes changes within a deep learning framework.

CRediT authorship contribution statement

Shuanglin Yan: Conceptualization, Methodology, Software, Investigation, Data curation, Writing - original draft. **Yafei Zhang:** Conceptualization, Methodology, Supervision, Writing - review & editing, Funding acquisition. **Minghong Xie:** Methodology, Investigation, Validation, Formal analysis. **Dacheng Zhang:** Writing - review & editing, Visualization, Investigation. **Zhengtao Yu:** Writing - review & editing, Supervision.

Declaration of Competing Interest

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

Acknowledgment

This work is partly supported by the National Key Research and Development Plan Project (Nos. 2018YFC0830105, 2018YFC0830100), National Natural Science Foundation of China under Grant (61762056, 61966021, 62161015, 61562053, 61563025, 61763020), and Yunnan Natural Science Funds under Grant (2017FB094).

References

[1] S. Bak, P. Carr, J. Lalonde, Domain adaptation through synthesis for unsupervised person re-identification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 193–209.
[2] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, J. Jiao, Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 994–1003.
[3] S. Lin, H. Li, C. Li, A. Kot, Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification, in: The British Machine Vision Conference (BMVC), 2018.
[4] J. Wang, X. Zhu, S. Gong, W. Li, Transferable joint attribute-identity deep learning for unsupervised person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2275–2284.
[5] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer gan to bridge domain gap for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 79–88.
[6] J. Song, Y. Yang, Y. Song, T. Xiang, T.M. Hospedales, Generalizable person re-identification by domain-invariant mapping network, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 719–728.
[7] H. Yu, W. Zheng, A. Wu, X. Guo, S. Gong, J. Lai, Unsupervised person re-identification by soft multilabel learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2143–2152.
[8] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, X. Wang, Unsupervised domain adaptive re-identification: Theory and practice, Pattern Recognition 102 (2020) 107173.
[9] Y.L. andX. Dong, L. Zheng, Y. Yan, Y.Yang, A bottom-up clustering approach to unsupervised person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), Vol. 33, 2019, pp. 8738–8745.
[10] Y. Ge, D. Chen, H. Li, Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification, in: International Conference on Learning Representations (ICLR), 2020, pp. 5157–5166.
[11] Y. Zhao, H. Lu, Neighbor similarity and soft-label adaptation for unsupervised cross-dataset person re-identification, Neurocomputing 388 (2020) 246–254.
[12] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3774–3782.
[13] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1116–1124.
[14] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1363–1372.
[15] Z. Zhong, L. Zheng, Z. Luo, S. Li, Y. Yang, Invariance matters: Exemplar memory for domain adaptive person re-identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 598–607.
[16] H. Li, W. Zhou, Z. Yu, B. Yang, H. Jin, Person re-identification with dictionary learning regularized by stretching regularization and label consistency constraint, Neurocomputing 379 (2020) 356–369.
[17] H. Li, J. Xu, Z. Yu, J. Luo, Jointly learning commonality and specificity dictionaries for person re-identification, IEEE Transactions on Image Processing 29 (2020) 7345–7358.
[18] Z. Chang, Z. Qin, H. Fan, H. Su, H. Yang, S. Zheng, H. Ling, Weighted bilinear coding over salient body parts for person re-identification, Neurocomputing 407 (2020) 454–464.
[19] F. Ma, X. Zhu, Q. Liu, C. Song, X. Jing, D. Ye, Multi-view coupled dictionary learning for person re-identification, Neurocomputing 348 (2019) 16–26.
[20] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5157–5166.
[21] S. Liao, Y. Hu, S.Z. Xiangyu Zhu, Li, Person re-identification by local maximal occurrence representation and metric learning, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2197–2206.
[22] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3586–3593.
[23] H. Wang, S. Gong, T. Xiang, Unsupervised learning of generative topic saliency for person re-identification, British Machine Vision Association (2014) 1–11.
[24] H. Yu, A. Wu, W. Zheng, Cross-view asymmetric metric learning for unsupervised person re-identification, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 994–1002.
[25] H. Yu, A. Wu, W. Zheng, Unsupervised person re-identification by deep asymmetric metric embedding, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (4) (2020) 956–973.
[26] Y. Lin, L. Xie, Y. Wu, C. Yan, Q. Tian, Unsupervised person re-identification via softened similarity learning, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3387–3396.
[27] D. Wang, S. Zhang, Unsupervised person re-identification via multi-label classification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10978–10987.
[28] H. Li, J. Xu, J. Zhu, D. Tao, Z. Yu, Top distance regularized projection and dictionary learning for person re-identification, Information Sciences 502 (2019) 472–491.
[29] H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognition 79 (2018) 130–146.
[30] H. Li, Y. Wang, Z. Yang, R. Wang, X. Li, D. Tao, Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion, IEEE Transactions on Instrumentation and Measurement 69 (4) (2020) 1082–1102.
[31] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, Unsupervised cross-dataset transfer learning for person re-identification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1306–1315.
[32] L. Qi, J. Huo, X. Fan, Y. Shi, Y. Gao, Unsupervised joint subspace and dictionary learning for enhanced cross-domain person re-identification, IEEE Journal of Selected Topics in Signal Processing 12 (2018) 1263–1275.
[33] X. Wang, W. Zheng, X. Li, J. Zhang, Cross-scenario transfer person re-identification, IEEE Transactions on Circuits and Systems for Video Technology 26 (8) (2016) 1447–1460.
[34] X. Liu, H. Tan, X. Tong, J. Cao, J. Zhou, Feature preserving gan and multi-scale feature enhancement for domain adaption person re-identificatio, Neurocomputing 364 (2019) 108–118.
[35] Q. Yang, H. Yu, A. Wu, W. Zheng, Patch-based discriminative feature learning for unsupervised person re-identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3628–3637..
[36] H. Tang, Z. Li, Z. Peng, J. Tang, Blockmix: meta regularization and self-calibrated inference for metric-based meta-learning, in: Proceedings of the 28th ACM International Conference on Multimedia (ACM MM), 2020, pp. 610–618.

[37] C. Luo, C. Song, Z. Zhang, Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup, in: Proceedings of the European Conference on Computer Vision(ECCV), 2020, pp. 224–241.
[38] P. Peng, Y. Tian, T. Xiang, Y. Wang, M. Pontil, T. Huang, Joint semantic and latent attribute modeling for cross-class transfer learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (7) (2018) 1625–1638.
[39] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Y. Yang, Improving person re-identification by attribute and identity learning, Pattern Recognition 95 (2019) 151–161.
[40] J. Liu, Z. Zha, H. Xie, Z. Xiong, Y. Zhang, Ca3net: Contextual-attentional attribute-appearance network for person re-identification, in: Proceedings of the 26th ACM international conference on Multimedia (ACM MM), 2018, pp. 737–745.
[41] J. Liu, Z. Zha, D. Chen, R. Hong, M. Wang, Adaptive transfer network for cross-domain person re-identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7195–7204.
[42] J. Cai, E. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, Siam Journal on Optimization 20 (4) (2008) 1956–1982.
[43] H. Lee, A. Battle, R. Raina, A. Ng, Efficient sparse coding algorithms, in: International Conference on Neural Information Processing Systems (NeurIPS), Vol. 19, 2006, pp. 801–808.
[44] H. Li, S. Yan, Z. Yu, D. Tao, Attribute-identity embedding and self-supervised learning for scalable person re-identification, IEEE Transactions on Circuits and Systems for Video Technology (2020) 1.
[45] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, in: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, 2007, pp. 1–7.
[46] M. Hirzer, C. Beleznai, P. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, in: Scandinavian Conference on Image Analysis, 2011, pp. 91–102.
[47] C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, in: IEEE International Conference on Image Processing (ICIP), 2013, pp. 3567–3571.
[48] P. Roth, M. Hirzer, M. Köstinger, C. Beleznai, H. Bischof, Mahalanobis distance learning for person re-identification, in: Person re-identification, Springer, 2014, pp. 247–267.
[49] W. Li, R. Zhao, X. Wang, Human re-identification with transferred metric learning, in: Asian Conference on Computer Vision (ACCV), 2012, pp. 31–44..
[50] R. Layne, T. Hospedales, S. Gong, Attributes-based re-identification, in: Person re-identification, Springer, 2014, pp. 93–117.
[51] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, Journal of Machine Learning Research 17 (59) (2016) 1–35.
[52] R. Zhao, W. Ouyang, X. Wang, Person re-identification by saliency learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2) (2017) 356–370.
[53] H. Li, Z. Kuang, Z. Yu, J. Luo, Structure alignment of attributes and visual features for cross-dataset person re-identification, Pattern Recognition 106 (2020) 107414.
[54] E. Kodirov, X. Tao, Z. Fu, S. Gong, Person re-identification by unsupervised l1 graph learning, Hydrobiologia 415 (11) (2016) 178–195.
[55] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 475–491.
[56] F. Yang, K. Li, Z. Zhong, Z. Luo, X. Sun, H. Cheng, X. Guo, F. Huang, R. Ji, S. Li, Asymmetric co-teaching for unsupervised cross-domain person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2020, pp. 12597–12604.
[57] C. Qin, S. Song, G. Huang, L. Zhu, Unsupervised neighborhood component analysis for clustering, Neurocomputing 168 (2015) 609–617.
[58] Z. Shi, T.M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4184–4193.
[59] X. Qian, Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y. Jiang, X. Xue, Pose-normalized image generation for person re-identification, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 650–667.
[60] R. Zhou, X. Chang, L. Shi, Y. Shen, Y. Yang, F. Nie, Person reidentification via multi-feature fusion with adaptive graph learning, IEEE Transactions on Neural Networks and Learning Systems 31 (5) (2020) 1592–1601.
[61] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camstyle: A novel data augmentation method for person re-identification, IEEE Transactions on Image Processing 28 (3) (2019) 1176–1190.
[62] Y. Yuan, W. Chen, T. Chen, Y. Yang, Z. Ren, Z. Wang, G. Hua, Calibrated domain-invariant learning for highly generalizable large scale re-identification, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020, pp. 3578–3587.
[63] C. Ren, B. Liang, P. Ge, Y. Zhai, Z. Lei, Domain adaptive person re-identification via camera style generation and label propagation, IEEE Transactions on Information Forensics and Security 15 (2020) 1290–1302.
[64] G. Tang, X. Gao, Z. Chen, H. Zhong, Unsupervised adversarial domain adaptation with similarity diffusion for person re-identification, Neurocomputing 442 (2021) 337–347.
[65] H. Zhang, Y. Li, Z. Zhuang, L. Xie, Q. Tian, 3d-gat: 3d-guided adversarial transform network for person re-identification in unseen domains, Pattern Recognition 112 (2021) 107799.
[66] Y. Chong, C. Peng, J. Zhang, S. Pan, Style transfer for unsupervised domain-adaptive person re-identification, Neurocomputing 422 (2021) 314–321.
[67] A. Khatun, S. Denman, S. Sridharan, C. Fookes, End-to-end domain adaptive attention network for cross-domain person re-identification, IEEE Transactions on Information Forensics and Security (2021) 1.
[68] H. Li, J. Pang, D. Tao, Z. Yu, Cross adversarial consistency self-prediction learning for unsupervised domain adaptation person re-identification, Information Sciences 559 (2021) 46–60.

Shuanglin Yan is currently pursuing a doctor's degree in computer science and technology at the School of Computer Science and Engineering at Nanjing University of Science and Technology. His research interests include machine learning and computer vision.

Yafei Zhang received the Ph.D. degree in signal and information processing from Institute of electronics, Chinese Academy of Sciences, Beijing, China, in 2008. She is currently an associate professor at College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. Her main research interests include image processing and pattern recognition.

Minghong Xie received the Ph.D. degree in signal and information processing from Institute of electronics, Chinese Academy of Sciences, Beijing, China, in 2009. He is currently a senior engineer at College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. His research interests include remote sensing image processing and information fusion.

Dacheng Zhang received his Ph.D. degree in Control Systems from Université Grenoble Alpes in 2018, Master degree in Electrical & Electronic Engineering from Joseph Fourier University in 2014 and Bachelor degree in Nuclear Engineering from both Grenoble Institute of Technology and North China Electric Power University in 2009. His research interests include signal processing, stochastic modeling of system, performance deterioration and lifetime assessment.

Zhengtao Yu received his Ph.D degree in computer application technology from Beijing Institute of Technology, Beijing, China, in 2005. He is currently a professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, China. His main research interests include natural language process, image processing and machine learning.