# Enhancing Person Re-identification in a Self-trained Subspace

Xun Yang, Meng Wang, Richang Hong, Qi Tian, Yong Rui

***Abstract*—Despite the promising progress made in recent years, person re-identification (re-ID) remains a challenging task due to the complex variations in human appearances from different camera views. For this challenging problem, a large variety of algorithms have been developed in the fully-supervised setting, requiring access to a large amount of labeled training data. However, the main bottleneck for fully-supervised re-ID is the limited availability of labeled training samples. To address this problem, in this paper, we propose a self-trained subspace learning paradigm for person re-ID which effectively utilizes both labeled and unlabeled data to learn a discriminative subspace where person images across disjoint camera views can be easily matched. The proposed approach first constructs pseudo pairwise relationships among unlabeled persons using the k-nearest neighbors algorithm. Then, with the pseudo pairwise relationships, the unlabeled samples can be easily combined with the labeled samples to learn a discriminative projection by solving an eigenvalue problem. In addition, we refine the pseudo pairwise relationships iteratively, which further improves the learning performance. A multi-kernel embedding strategy is also incorporated into the proposed approach to cope with the non-linearity in person's appearance and explore the complementation of multiple kernels. In this way, the performance of person re-ID can be greatly enhanced when training data are insufficient. Experimental results on six widely-used datasets demonstrate the effectiveness of our approach and its performance can be comparable to the reported results of most state-of-the-art fully-supervised methods while using much fewer labeled data.**

***Index Terms*—Person Re-identification, Self-training, Semi-supervised Learning, Computer Vision**

## I. Introduction

Person Re-identification (re-ID) [1]–[6] aims to recognize an individual across spatially disjoint cameras. It has attracted much attention in recent years for its great potential in surveillance applications such as crowded scenes anomaly detection [7] and multi-cameras pedestrian tracking [8]. Although a large number of approaches have been proposed for re-ID, it remains a challenging problem since a person's appearance often undergoes dramatic changes across camera views due to changes in view angle, body pose, illumination and background clutter.

The fundamental re-ID problem is to compare a person of interest seen in a probe camera view to a gallery of candidates captured from a camera that does not overlap with the probe one. If a true match to the probe exists in the gallery, it should have a high matching/similarity score, or rank, compared to incorrect candidates. Generally, there are two basic problems: (1) feature representation and (2) metric learning. An effective feature representation [2], [3], [9], [10] is critical for person re-ID, which should be robust to complex variations in human appearances from different camera views. Several approaches have been investigated to design a feature descriptor directly based on low-level visual features. More efforts [6], [11]–[16] have been made following the second direction to learn an optimal distance or similarity function to rank the potential matches based on their relevance. Some of them directly learn a Mahalanobis distance function parameterized by a positive semi-definite (PSD) matrix to separate positive person image pairs from negative pairs. Some others formulate re-ID as a subspace learning problem by learning a low-dimensional projection. This work follows the second approach, aiming to learn a discriminative projection to map person images from disjoint camera views into a common subspace.

Despite the promising efforts made by many researchers, most existing methods are developed in the fully-supervised setting, requiring access to a large amount of labeled training image pairs. It is impractical to expect the availability of large quantities of labeled data because labeling data is very costly. The main bottleneck for fully-supervised re-ID is the limited availability of labeled training samples. When only a small number of labeled data are available, supervised methods tend to learn a distance function that is over-fitted to the labeled data, which makes the learned distance function cannot generalize well to the test set. It's of great interest to design a solution that can utilize abundant unlabeled data. Although some semi-supervised re-ID approaches [17]–[19] have been proposed, their performances are far from satisfactory.

In this work, we design a self-trained subspace learning approach for person re-ID which effectively utilizes both labeled and unlabeled data to learn a discriminative subspace where person images across disjoint camera views can be easily matched. The classic self-training strategy [20] is exploited in this work. We first learn an initial projection matrix using the available labeled data only. Using this initial projection, all unlabeled person images are projected into a low-dimensional subspace, where the low-dimensional representation has higher discriminative power than the original features. Then, to utilize the unlabeled data, we construct pseudo pairwise relationships among the unlabeled persons using k-nearest neighbors (KNN) algorithm in this low-dimensional subspace. The pseudo pairwise relationships are encoded into a graph Laplacian regularization term which is further combined with a fully-supervised discriminative term to learn a new projection. Given the newly learned projection, we refine the pseudo pairwise relationships

Xun Yang, Meng Wang, and Richang Hong are with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China ({hfutyangxun, eric.mengwang, hongrc.hfut}@gmail.com).

Qi Tian is with the Department of Computer Science, University of Texas at San Antonio (qitian@cs.utsa.edu).

Yong Rui is with Lenovo (yongrui@lenovo.com).

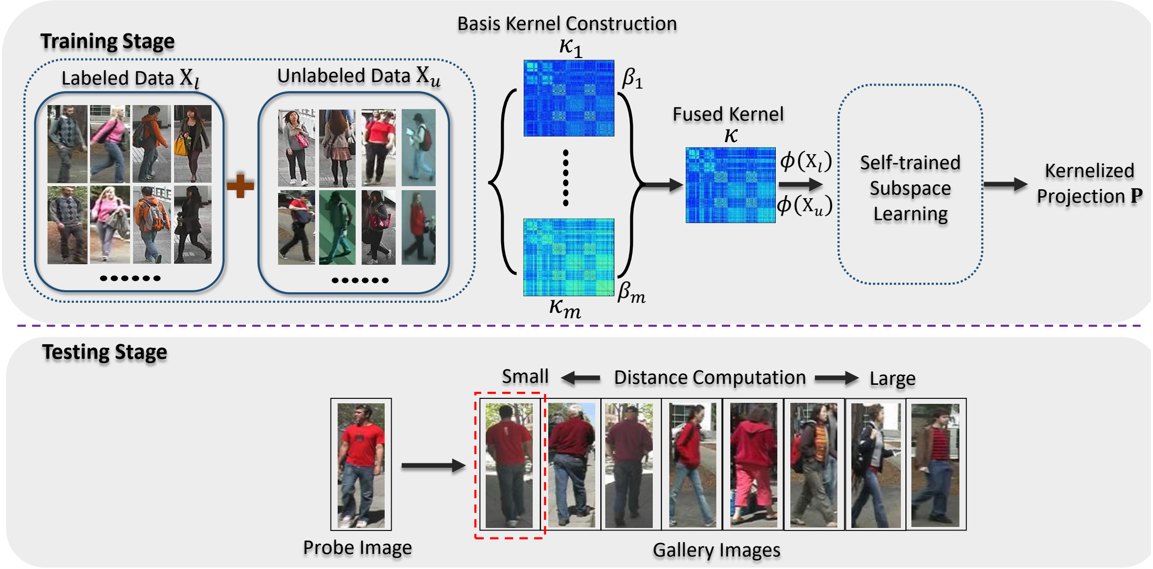


Fig. 1. Schematic illustration of the proposed person re-identification approach.

and relearned the discriminative projection with these updated pseudo pairwise relationships. This process is iterated until the pseudo pairwise relationships remain unchanged. In this way, the discriminant power of the learned subspace will be enhanced. Besides, a multi-kernel embedding strategy is incorporated into the proposed approach to cope with the non-linearity in person's appearance and explore the complementation of multiple kernels. The final person matching can be performed very efficiently by computing the Euclidean distance between a probe image and a gallery image in the self-trained subspace. A schematic illustration of the proposed approach is shown in Fig. 1.

Our main contributions are summarized as follows:

(1) We propose an effective self-trained subspace learning framework for person re-ID which is able to utilize both labeled and unlabeled person images effectively. An iterative learning strategy is included to update the pseudo pairwise relationships among unlabeled persons.

(2) We introduce a multiple kernel embedding technique into the self-trained subspace learning framework, which explores the complementary information shared by multiple kernels and handles the non-linearity in person's appearance effectively.

(3) We conduct empirical studies on widely-used person re-ID datasets. Experimental results demonstrate that the proposed method is able to achieve a performance on par with the reported results of most state-of-the-art fully-supervised methods while using much fewer labeled person samples.

## II. Related Work

### A. Person Re-ID

During the past decades, many person re-ID algorithms [2], [3], [11], [13], [14], [21]–[26] have been proposed. Mainstream works can be roughly categorized into two groups as follows.

The first group of methods focus on designing discriminative and invariant features [2], [3], [9], [10], [22], [27], [28]. Earlier works include fisher vector based local descriptors [29], color invariant features [30] and saliency learning based methods [28]. Recently, some new proposed descriptors have gained good performance, i.e., salient color names [27], local maximal occurrence (LOMO) feature [2], weighted histograms of overlapping Stripes (Whos) [10], and Gaussian of Gaussian (GOG) descriptor [3]. The GOG descriptor is used in this work. It describes a local region in a person image via hierarchical Gaussian distribution in which both means and covariances are included in their parameters. Specifically, it models the region as a set of multiple Gaussian distributions in which each Gaussian represents the appearance of a local patch. The characteristics of the set of Gaussian distributions are again described by another Gaussian distribution. The final descriptor fuses multiple GOG descriptor extracted from different color spaces. It performs well on a lot of re-ID datasets.

The second group of methods aim to learn a robust and discriminative distance function for recognizing people across views [2], [11]–[16], [24], [31]–[33]. In this group, some works aim to learn a Mahalanobis-like distance metric [16], [34], [35], while some methods focus on seeking a discriminative projection. [6], [11], [12], [36]. These two subgroups actually are closely related. We briefly introduce some well-known works as follows. Liao et al. [16] proposed a logistic metric learning approach with PSD constraints and asymmetric sample weight strategy. Zheng et al. [37] formulated re-ID as a relative distance comparison learning problem by maximizing the probability that relevant samples have smaller distance than the irrelevant ones. Kostinger et al. [13] designed a simple and effective metric learning method by computing the difference between the intra-class and inter-class covariance matrix, while the presented algorithm is very sensitive to the dimension of feature representation. As an improvement, Liao et al. [2] proposed a cross-view quadratic discriminant analysis (XQDA) method by learning a more discriminative distance metric and a low-dimensional subspace simultaneously. Pedagadi et al. [24] applied the local fisher discriminant analysis algorithm to match person images by maximizing the inter-class separability while preserving the multi-class modality, whose kernel version was presented for re-ID in [11]. Zhang et al. [12] proposed to overcome the small-sample-size problem in re-ID by learning a discriminative null space, where the within-class scatter is minimized to zero while maximizing the relative between-class separation simultaneously. Although lots of works have been developed, they are mainly developed in the fully-supervised setting. Once only a small number of labeled data are available, supervised methods are vulnerable to over-fitting. Therefore, the purpose of this work is to present a semi-supervised re-ID approach which can utilize the abundant unlabeled data to enhance the learning performance.

There are few semi-supervised re-ID methods, except for [17]–[19], [38]. Figueira et al. [19] designed a semi-supervised method which exploits the general framework of multi-view learning with manifold regularization. It treats each person as a single class and poses re-ID as a multiple class recognition problem. However, conventional multiple class recognition methods may not be suitable for re-ID since each person usually has very limited images in re-ID. Liu et al. [17], proposed a semi-supervised coupled dictionary learning for re-ID, in which unlabeled data are used to improve the reconstructive ability of dictionaries. The authors assumed that the sparse representation coefficients of two matched images should be strictly equivalent, which is too strong to cope with dramatic changes of person's appearance. Kodirov et al. [18] presented a semi-supervised re-ID approach with graph Laplacian regularization, in which the visual similarity between a pair of unlabeled person images is computed using original low-level feature representation, which may result in a suboptimal performance. Different with [18], the proposed approach computes the similarity in a discriminative subspace learned using the available labeled data. Kodirov et al. [39] also designed an unsupervised re-ID approach by introducing a new $\ell_1$-norm based graph Laplacian term instead of the conventional squared $\ell_2$-norm in [18], which can also be extended to a semi-supervised case. Karaman et al. [38] described a semi-supervised re-ID approach. It combines discriminative models of person identity with a conditional random field to exploit the local manifold approximation induced by KNN graph. Different with our proposed approach, it is mainly designed for multi-shot scenarios where meaningful structure can be discovered easily. While, our approach can perform well in the challenging single-shot scenarios. Zhang et al. [12] introduced a semi-supervised extension for re-ID based on the self-training strategy [20]. It combines the pseudo-classes with the labeled data together into a new training set to learn the projection. Its difference with our work is that we place the pseudo pairwise information in a separate regularization term, which can reduce the negative effects of the incorrect matching pairs in the pseudo pairwise relationships and obtain more stable performance.

We also introduce a multi-kernel based extension in this work which exploits the complementation of multiple kernel representations. Some existing works [31], [40] have investigated the effects of multiple feature representations for re-ID by a score-level fusion. Different with them, we focus on a kernel-level (feature-level) fusion.

Loy et al. [41] formulated re-ID as a manifold ranking problem in an unsupervised way. It exploits the manifold structure revealed by a large quantity of gallery samples to obtain more robust ranking result. When combined with a distance metric learning method, it functions as a post-ranking approach. In this work, we have investigated the effect of this manifold ranking approach. Experimental results demonstrate that our approach and the manifold ranking method can complement each other very well.

### B. Semi-supervised Learning

Semi-supervised learning is a classic topic in the area of machine learning, which has numerous literatures. In this subsection, we only review some semi-supervised learning works which are based on self-training [20] or related to our work. Self-training [20] is a commonly used semi-supervised learning technique and probably the earliest idea about using unlabeled data [42]. It is also known as self-learning, self-labeling, or decision-directed learning. This is a wrapper-algorithm that repeatedly uses a supervised learning method. It starts by training on the labeled data only. In each step a part of the unlabeled points are labeled according to the current decision function; then the supervised method is retrained using its own predictions as additional labeled points.

[43] is one of the earliest works that applies this strategy to design an iterative reclassification procedure. [44] is a well-known example of self-training for word sense disambiguation. In [45], this strategy is used for semi-supervised clustering. Rosenberg et al. [46] applied self-training to object detection systems from images. Recently, self-training is explored in a state-of-the-art work [47] that proposes a general way to perform semi-supervised parameter estimation for likelihood-based classifiers and their estimates are never worse than the supervised solution in terms of the log-likelihood on the full training set. As a classic technique, self-training has been widely used for classification, clustering, regression, and other specific tasks in the past decades, and it still motivates researchers to develop new algorithms for specific applications nowadays.

In this paper, we focus on designing a semi-supervised learning approach for person re-ID using self-training. Note that re-ID is a retrieval problem, and there is no intersection between the persons (classes) in training and testing. Therefore, most self-training based semi-supervised methods that are mainly designed for classification problems, are not suitable for re-ID. Owing to its simplicity and effectiveness for re-ID, subspace learning is explored as the basic learning method in our approach. Thereby, our approach is also related to some semi-supervised subspace/metric learning approaches [48]–[50]. Our approach and [48], [49] follow the same way to leverage unlabeled data, which encodes the neighborhood structure of unlabeled data in a Laplacian regularizer. Different with our approach, [48] is extended from linear discriminant analysis (LDA), while our approach seems more like a semi-supervised locality preserving projections (LPP) [51]. [49] focuses on learning a PSD constrained Mahalanobis metric and it uses a general loss term of metric learning. [50] is the semi-supervised extension of local fisher discriminant analysis (LFDA) [52] which combines the supervised LFDA and the unsupervised PCA to jointly exploit labeled and unlabeled data. It doesn't exploit the neighborhood structure of unlabeled data. Besides, to better leverage unlabeled data for re-ID, our approach uses self-training to repeatedly update the neighborhood structure of unlabeled data. Experimental results in this work have demonstrated that the self-training strategy significantly enhances the learning performance. Our main contribution is that we introduce a simple and effective re-ID approach which can exploit both labeled and unlabeled data using the self-training strategy.

## III. The Proposed Approach

In this section, we describe our re-ID method. First, we briefly show how to perform re-ID in a fully-supervised subspace in Subsection III-A. Then, we introduce the proposed semi-supervised case in detail in Subsection III-B, followed by a multi-kernel based extension in Subsection III-C.

### A. Person Re-ID in a Fully-supervised Subspace

In this work, we formulate person re-ID as a subspace learning problem, which is similar with [11], [24], [36]. Assume that we are given a set of $n$ labeled training person images $\mathbf{X}_l = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{d\times n}$, and their label set $\mathbf{Y}_l = \{y_i\}_{i=1}^n \in \mathbb{R}^n$, where $d$ denotes the dimension of feature vector. The task is to learn a squared distance function $d_{\mathbf{U}}^2(\mathbf{x}_i, \mathbf{x}_j)$ which is parameterized by a low-dimensional projection $\mathbf{U}$ defined as follow:

$$d_{\mathbf{U}}^2(\mathbf{x}_i, \mathbf{x}_j) = \left\|\mathbf{U}^{\mathrm{T}}\mathbf{x}_i - \mathbf{U}^{\mathrm{T}}\mathbf{x}_j\right\|^2, \qquad (1)$$

where $\mathbf{U} \in \mathbb{R}^{d\times r}(r \ll d)$ is a low-dimensional projection matrix which maps the person images from disjoint camera views into a common subspace where person re-ID can be performed easily. $r$ is the dimension of the projected subspace. The learned distance is expected to be small if $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same person ($y_i = y_j$). Under this expectation, we formulate re-ID as

$$\begin{aligned}\mathbf{U}^* &= \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{X}_l, \mathbf{U}, \mathbf{W}_l\right) \\ &= \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n W_{ij}^l \left\|\mathbf{U}^{\mathrm{T}}\mathbf{x}_i - \mathbf{U}^{\mathrm{T}}\mathbf{x}_j\right\|^2,\end{aligned} \qquad (2)$$

where $W_{ij}^l$ is an element of a weight matrix $\mathbf{W}^l \in \mathbb{R}^{n\times n}$ which encodes the pairwise constraints information between each pair of person images

$$W_{ij}^l = \left\{\begin{array}{ll} 1 & \text{if } y_i = y_j \\ 0 & \text{otherwise} \end{array}\right. . \qquad (3)$$

The loss function $\mathcal{L}\left(\mathbf{X}_l, \mathbf{U}, \mathbf{W}^l\right)$ can be rewritten as $\mathbf{tr}\left(\mathbf{U}^{\mathrm{T}}\mathbf{X}_l\mathbf{L}^l\mathbf{X}_l^{\mathrm{T}}\mathbf{U}\right)$, where $\mathbf{tr}(\cdot)$ denotes the trace operator and $\mathbf{L}^l = \mathbf{D}^l - \mathbf{W}^l$ is known as the graph Laplacian matrix. $\mathbf{D}^l$ is a diagonal matrix whose diagonal elements equal to the sums of the row entries of $\mathbf{W}^l$, i.e., $D_{ii}^l = \sum_j W_{ij}^l$. By adding a constraint $\mathbf{tr}\left(\mathbf{U}^{\mathrm{T}}\mathbf{X}_l\mathbf{D}^l\mathbf{X}_l^{\mathrm{T}}\mathbf{U}\right) = 1$, the minimization problem in Eq. (2) can be easily solved with a generalized eigen-decomposition. The final projection $\mathbf{U}^* = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r] \in \mathbb{R}^{d\times r}$ is constituted by the resulting eigenvectors associated to the $r$ smallest eigenvalues. Usually, $r$ is set to the difference of the number of labeled persons and one.

### B. Person Re-ID in a Self-trained Subspace

The above method seeks the discriminative projection based on merely labeled data. However, the main bottleneck for fully-supervised person re-ID is the limited availability of labeled training samples. When only a small number of labeled image pairs are available, the solution of above method tends to be over-fitted to the labeled data. In this subsection, we introduce a self-trained subspace learning framework for person re-ID in a semi-supervised setting.

Given a set of $n$ labeled person images $\mathbf{X}_l = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{d\times n}$ and $u$ unlabeled person images $\mathbf{X}_u = \{\mathbf{x}_i\}_{i=n+1}^{n+u} \in \mathbb{R}^{d\times u}$, the task is to learn a projection $\mathbf{U} \in \mathbb{R}^{d\times r'}$ which has good generalization capability and discriminative power. A common way is to formulate the semi-supervised learning problem as the following general form:

$$\mathbf{U}^* = \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{X}_l, \mathbf{U}, \mathbf{W}_l\right) + \eta\mathcal{R}\left(\mathbf{X}_u, \mathbf{U}\right), \qquad (4)$$

where the first term $\mathcal{L}\left(\mathbf{X}_l, \mathbf{U}, \mathbf{W}_l\right)$ is the labeled term in Eq. (2) which only relies on labeled data, and the second term $\mathcal{R}\left(\mathbf{X}_u, \mathbf{U}\right)$ is a regularization term constructed by unlabeled data. The trade-off between these two terms is captured by a small regularization parameter ($\eta > 0$). The main problem is how to utilize the unlabeled data to construct the regularization term. A widely-used strategy in computer vision problems [18], [48], [49], [53], [54] is encoding the intrinsic geometric structure of unlabeled data into a regularizer under the manifold assumption that visually similar samples are more likely to share the same class label. A KNN graph is usually constructed to model the relationship between nearby data nodes, where an edge will be placed between two nodes if they are close. Nearest neighbors selection is performed by computing the Euclidean distance between node $i$ and node $j$ using original feature representation. However, the original feature representation has very low discriminative power due to the dramatic appearance change across camera-views in person re-ID. As a result, the constructed KNN graph may be misleading, which will degrade the learning performance.

To overcome the above problem, in this work, we introduce self-training to better utilize the unlabeled data $\mathbf{X}_u$. We first apply the fully-supervised method described in Subsection III-A to learn an initial projection matrix $\mathbf{U}^0$ using the labeled data $\mathbf{X}_l$ only. Then, we project the unlabeled data $\mathbf{X}_u$ into a low-dimensional subspace using $\mathbf{U}^0$. The low-dimensional representations $\mathbf{Z}_u = (\mathbf{U}^0)^{\mathrm{T}}\mathbf{X}_u$ are used to obtain the cross-view adjacency relationships among the unlabeled samples $\mathbf{X}_u$ by constructing a KNN graph. The cross-view adjacency relationships can be viewed as a kind of pseudo pairwise relationships. Given the pseudo pairwise relationships, we can

**Algorithm 1** The proposed self-trained subspace learning approach

```
Input: The labeled training data $\mathbf{X}_l$ and its weight matrix $\mathbf{W}^l$; the unlabeled training data $\mathbf{X}_u$; the parameter $\eta$; the maximal number of iterations $\mathcal{T}$.
Initialize: $\mathbf{U}^0 = \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{X}_l, \mathbf{U}, \mathbf{W}^l\right)$; $t = 1$.
Iterative: $t = 1, 2, \cdots, \mathcal{T}$
    Project $\mathbf{X}_u$ into a low-dimensional subspace through $\mathbf{Z}_u^t = (\mathbf{U}^{t-1})^{\mathrm{T}}\mathbf{X}_u$;
    Build the pseudo pairwise relationships among the unlabeled data by constructing a KNN graph using $\mathbf{Z}_u^t$;
    Encode the pseudo pairwise relationships into a weight matrix $\mathbf{W}^{u,t}$;
    Solve $\mathbf{U}^t = \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{X}_l,\mathbf{U},\mathbf{W}^l\right) + \eta\mathcal{R}\left(\mathbf{X}_u,\mathbf{U},\mathbf{W}^{u,t}\right)$ to obtain the new projection matrix;
Until $t > \mathcal{T}$ or the stop condition is met.
Output: The projection matrix $\mathbf{U}$.
```

leverage the unlabeled data in a supervised way. We encode the pseudo pairwise relationships into the following weighted matrix $\mathbf{W}^u \in \mathbb{R}^{u\times u}$ for $\mathbf{X}_u$:

$$W_{ij}^u = \begin{cases} 1 & \text{if } \mathbf{x}_i \in \mathcal{N}(\mathbf{x}_j) \text{ or } \quad \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i) \\ 0 & \qquad\qquad \text{otherwise} \end{cases}, \quad (5)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are two unlabeled person images from different views, $n+1 \leq i, j \leq n+u$. $\mathcal{N}(\mathbf{x}_i)$ denotes the nearest neighbor list of $\mathbf{x}_i$. After obtaining the weighted matrix $\mathbf{W}^u$, the regularization term in Eq. (4) is constructed as

$$\begin{aligned}\mathcal{R}\left(\mathbf{X}_u,\mathbf{U},\mathbf{W}^u\right) &= \frac{1}{2}\sum_{i=n+1}^{n+u}\sum_{j=n+1}^{n+u} W_{ij}^u \left\|\mathbf{U}^{\mathrm{T}}\mathbf{x}_i - \mathbf{U}^{\mathrm{T}}\mathbf{x}_j\right\|^2 \\ &= \mathbf{tr}\left(\mathbf{U}^{\mathrm{T}}\mathbf{X}_u\mathbf{L}^u\mathbf{X}_u^{\mathrm{T}}\mathbf{U}\right)\end{aligned} \quad (6)$$

where $\mathbf{L}^u = \mathbf{D}^u - \mathbf{W}^u$ is the Laplacian matrix for unlabeled data. $\mathbf{D}^u$ is a diagonal matrix whose diagonal elements equal to the sum of the rows entries of $\mathbf{W}^u$. Hence, by utilizing the pseudo pairwise relationships for unlabeled data, we rewrite the semi-supervised learning problem in Eq. (5) as

$$\begin{aligned}\mathbf{U}^* &= \arg\min_{\mathbf{U}} \mathcal{L}\left(\mathbf{X}_l,\mathbf{U},\mathbf{W}^l\right) + \eta\mathcal{R}\left(\mathbf{X}_u,\mathbf{U},\mathbf{W}^u\right) \\ &= \arg\min_{\mathbf{U}} \mathbf{tr}\Big(\mathbf{U}^{\mathrm{T}}\big(\mathbf{X}_l\mathbf{L}^l\mathbf{X}_l^{\mathrm{T}} + \eta\mathbf{X}_u\mathbf{L}^u\mathbf{X}_u^{\mathrm{T}}\big)\mathbf{U}\Big). \\ s.t. \quad & \mathbf{tr}\Big(\mathbf{U}^{\mathrm{T}}\big(\mathbf{X}_l\mathbf{D}^l\mathbf{X}_l^{\mathrm{T}} + \eta\mathbf{X}_u\mathbf{D}^u\mathbf{X}_u^{\mathrm{T}}\big)\mathbf{U}\Big) = 1\end{aligned} \quad (7)$$

We can easily solve the minimization problem in Eq.(7) with a generalized eigen-decomposition to obtain the projection matrix. Note that, to obtain more stable performance, the dimension $r'$ of the final subspace is fixed as the difference of the number of labeled persons and one. In other words, we do not increase the dimension of the subspace although we have exploited abundant unlabeled data.

It can be seen from the above analysis that the pseudo pairwise relationships among the unlabeled persons play a crucial role in the proposed approach. However, it inevitably includes some mismatching pairwise relationships due to the complex viewpoint variations. Therefore, given the newly learned projection, we refine the pseudo pairwise relationships and relearned the discriminative projection with these updated pseudo pairwise relationships. This procedure is iterated until the pseudo pairwise relationships remain unchanged. We summarize the proposed self-trained subspace learning algorithm in Algorithm 1.

### *C. Multi-kernel based Extension*

To better exploit the non-linearity in person's appearance and the complementary information shared by multiple kernels, we employ the multi-kernel embedding [55], [56] for the proposed approach. The task is to learn a kernelized projection $\mathbf{P}$. Given $M$ kernel matrices with the same size $\mathcal{K}_1, \cdots, \mathcal{K}_M \in \mathbb{R}^{(n+u)\times(n+u)}$ constructed using the total training set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n+u}$, the primary task is to learn a fused kernel matrix

$$\mathcal{K} = \sum_{m=1}^{M} \beta_m\mathcal{K}_m, s.t. \sum_{m=1}^{M}\beta_m = 1, \beta_m > 0, \quad (8)$$

where $\beta_m$ is a non-negative weight for the kernel matrix $\mathcal{K}_m$. In this formulation, $\{\mathcal{K}_m\}_{m=1}^M$ can be the same classic kernels with different hyper-parameters, different feature descriptors or different kernels. Let $\phi_m(\cdot)$ be the feature map for $\mathcal{K}_m$. Then $\mathcal{K}_m$ can be expressed by an inner product in the kernel space $\mathcal{K}_m = \phi_m(\mathbf{X})^{\mathrm{T}}\phi_m(\mathbf{X})$, where $\phi_m(\mathbf{X}) = [\phi_m(\mathbf{x}_1), \cdots, \phi_m(\mathbf{x}_n), \cdots, \phi_m(\mathbf{x}_{n+u})]$. We can rewrite the fused kernel matrix in Eq. (8) as

$$\mathcal{K} = \sum_{m=1}^{M}\beta_m\phi_m(\mathbf{X})^{\mathrm{T}}\phi_m(\mathbf{X}) = \phi(\mathbf{X})^{\mathrm{T}}\phi(\mathbf{X}), \quad (9)$$

where $\phi(\cdot) = \left[\sqrt{\beta_1}\phi_1(\cdot)^{\mathrm{T}}, \cdots, \sqrt{\beta_M}\phi_M(\cdot)^{\mathrm{T}}\right]^{\mathrm{T}}$ is the fused feature map. We employ the kernel alignment [57] approach to decide the kernel weights.

$$\beta_m = \frac{A(\mathcal{K}_m^{ll}, \mathcal{K}d)}{\sum_{m'=1}^{M} A(\mathcal{K}_{m'}^{ll}, \mathcal{K}d)}. \quad (10)$$

where $\mathcal{K}d \in \mathbb{R}^{n\times n}$ is the ideal kernel matrix for the labeled person samples, whose element is 1 where rows and columns correspond to the same person or 0 everywhere else. $\mathcal{K}_{m'}^{ll} \in \mathbb{R}^{n\times n}$ is a part of the base kernel matrix $\mathcal{K}_{m'}$, which corresponds to the labeled person samples. The alignment score between two kernel matrices is defined as follows

$$A(\mathcal{K}_{m'}^{ll}, \mathcal{K}d) = \frac{\langle\mathcal{K}_{m'}^{ll}, \mathcal{K}d\rangle_F}{\sqrt{\langle\mathcal{K}_{m'}^{ll}, \mathcal{K}_{m'}^{ll}\rangle_F\langle\mathcal{K}d, \mathcal{K}d\rangle_F}}, \quad (11)$$

where $\langle\mathcal{K}_{m'}^{ll}, \mathcal{K}d\rangle_F = \mathbf{tr}((\mathcal{K}_{m'}^{ll})^{\mathrm{T}}\mathcal{K}d)$.

When $\mathcal{K}$ and $(\beta_1, \beta_2, \cdots, \beta_M)$ are obtained, we denote $\mathcal{K} = \begin{bmatrix} \mathcal{K}^{ll} & \mathcal{K}^{lu} \\ \mathcal{K}^{ul} & \mathcal{K}^{uu} \end{bmatrix}$, $\mathcal{K}^l = \begin{bmatrix} \mathcal{K}^{ll} \\ \mathcal{K}^{ul} \end{bmatrix}$, and $\mathcal{K}^u = \begin{bmatrix} \mathcal{K}^{lu} \\ \mathcal{K}^{uu} \end{bmatrix}$. We denote the fused feature maps of the labeled training set and the unlabeled training set as $\phi(\mathbf{X}_l) = [\phi(\mathbf{x}_1), \cdots, \phi(\mathbf{x}_n)]$, and $\phi(\mathbf{X}_u) = [\phi(\mathbf{x}_{n+1}), \cdots, \phi(\mathbf{x}_{n+u})]$, respectively. The minimization problem in Eq. (7) can be solved by computing the following eigenvalue problem in the fused kernel space

$$\begin{aligned}&\left(\phi(\mathbf{X}_l)\mathbf{L}^l\phi(\mathbf{X}_l)^{\mathrm{T}} + \eta\phi(\mathbf{X}_u)\mathbf{L}^u\phi(\mathbf{X}_u)^{\mathrm{T}}\right)\mathbf{u} \\ =&\lambda\left(\phi(\mathbf{X}_l)\mathbf{D}^l\phi(\mathbf{X}_l)^{\mathrm{T}} + \eta\phi(\mathbf{X}_u)\mathbf{D}^u\phi(\mathbf{X}_u)^{\mathrm{T}}\right)\mathbf{u}^{\cdot}\end{aligned} \quad (12)$$

Fig. 2. Sample images from six person re-identification datasets. From left to right: VIPeR, CUHK01, PRID2011 (PRID450S), GRID, and 3DPeS.

Since the eigenvectors are the linear combinations of $\phi(\mathbf{x}_1)$, $\phi(\mathbf{x}_2)$, $\cdots$, $\phi(\mathbf{x}_{n+u})$, there exists coefficients $p_i$ such that $\mathbf{u} = \sum_{i=1}^{n+u} p_i \phi(\mathbf{x}_i) = \phi(\mathbf{X})\mathbf{p}$, where $\mathbf{p} = [p_1, p_2, \cdots, p_{n+u}]^{\mathrm{T}} \in \mathbb{R}^{n+u}$.

By simple algebra formulation, we can finally obtain the following kernelized eigenvalue problem

$$\begin{aligned}&\left(\mathcal{K}^l\mathbf{L}^l(\mathcal{K}^l)^{\mathrm{T}} + \eta\mathcal{K}^u\mathbf{L}^u(\mathcal{K}^u)^{\mathrm{T}}\right)\mathbf{p}\\ =&\lambda\left(\mathcal{K}^l\mathbf{D}^l(\mathcal{K}^l)^{\mathrm{T}} + \eta\mathcal{K}^u\mathbf{D}^u(\mathcal{K}^u)^{\mathrm{T}}\right)\mathbf{p}.\end{aligned} \tag{13}$$

Usually, it is common to apply a regularization technique for the eigenvalue problem to avoid the singularity of matrix. For simplicity, we denote $\mathbf{A} = \mathcal{K}^l\mathbf{D}^l(\mathcal{K}^l)^{\mathrm{T}} + \eta\mathcal{K}^u\mathbf{D}^u(\mathcal{K}^u)^{\mathrm{T}}$. We regularize $\mathbf{A}$ by adding an identity matrix, i.e., $\mathbf{A} = \mathbf{A} + \vartheta\frac{\mathbf{tr}(\mathbf{A})}{n+u}\mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{(n+u)\times(n+u)}$ is an identity matrix and $\vartheta$ is a small positive parameter. The final kernelized projection $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_{r'}] \in \mathbb{R}^{(n+u)\times r'}$ is constituted by the resulting eigenvectors associated to the $r'$ smallest eigenvalues.

## IV. Experimental Results and Analyses

In this section, we first introduce the datasets, the evaluation protocol, and the experimental setting. Then, we evaluate the performance of our approach on multiple re-ID datasets.

### A. Datasets, Evaluation protocol, and Setting

*1) Datasets:* In this work we use six popular person re-ID datasets: VIPeR [22], CUHK01 [58], and PRID2011 [59], PRID450S [34], GRID [60], and 3DPeS [61]. Table I provides a statistical summary of each dataset. In Table I, we indicate the number of people, bounding boxes (BBoxes), distractors, and cameras (Cam) in each dataset. Fig. 2 shows some sample images from these six datasets.

TABLE I
The characteristics of six person re-ID datasets.

| Datasets | # People | # BBoxes | # Distractors | # Cam |
|---|---|---|---|---|
| VIPeR [22] | 632 | 1264 | 0 | 2 |
| CUHK01 [58] | 971 | 3884 | 0 | 2 |
| PRID2011 [59] | 200 | 849 | 649 | 2 |
| PRID450S [34] | 450 | 900 | 0 | 2 |
| GRID [60] | 250 | 500 | 775 | 8 |
| 3DPeS [61] | 192 | 1011 | 0 | 8 |

VIPeR [22] is the most commonly-used dataset containing 632 persons in which each person has a pair of images taken from widely differing views. The large viewpoint change of 90 degrees or more as well as huge lighting variations make it one of the most challenging datasets. CUHK01 [58] is one of the largest benchmarks. It contains 971 persons from two disjoint camera views, where each person has two images in each camera view. It contains 3884 images in total. PRID2011 [59] consists of person images recorded from two cameras (camera A and camera B) captures 385 persons and 749 persons, respectively, only 200 persons appear in both camera views. PRID450S [34] is an extension of PRID2011. It contains 450 persons in which each person has a pair of images taken from two disjoint camera views. GRID [60] has 250 image pairs collected from 8 non-overlapping cameras. 775 non-paired people are also included as distractors in the gallery set, which makes it extremely challenging. GRID suffers from viewpoint variations, background clutter, occlusions and low-resolution. 3DPeS is a set of selected snapshots of the original video dataset [61], containing 192 people and 1011 images.

*2) Evaluation protocol:* For all datasets, all the individuals are randomly divided into two subsets, so that the training and testing sets contain half of the available individuals with no overlap on person identities. The single-shot experiment setting [3] is used for all datasets. We also report the multi-shot matching results for CUHK01. As random selection is involved, the evaluation procedure is repeated for 10 times and the mean results for all datasets are reported. We adopt the data splits in [3] for VIPeR, CUHK01, PRID450S, and GRID. We use the data splits in [62] for PRID2011, in which 100 persons are randomly selected for training from the 200 available persons present in both views in each data split and the remaining 100 persons of camera view A are used as probe set and the remaining 649 persons of another camera view are used as gallery set. We use the data splits in [11] for 3DPeS, in which we randomly select one image of each individual in the testing set as gallery image, and the rest are used as probe images. The Cumulated Matching Characteristics (CMC) curve is used to evaluate the performance of all methods. It provides a ranking for every image in the gallery with respect to the probe.

*3) Setting:* For the multi-kernel setting, we use the GOG descriptor [3] to construct 11 Gaussian kernels $\exp(-\frac{1}{c\mu}\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where $c$ varies from 2 to 3 with step 0.1 and $\mu$ is average squared Euclidean distance.

Parameter $\vartheta$ is set to 0.01. Two main parameters are $\eta$ in Eq. (4) and the number of nearest neighbors $k$ in KNN graph. We set them as $\eta = 1$ and $k = 2$ by cross-validation on the training set of CUHK01 and fix them for all datasets. The maximal iteration number $\mathcal{T}$ is set to 10. The iteration procedure will stop when the neighborhood structure of unlabeled data remains unchanged or changes very little.

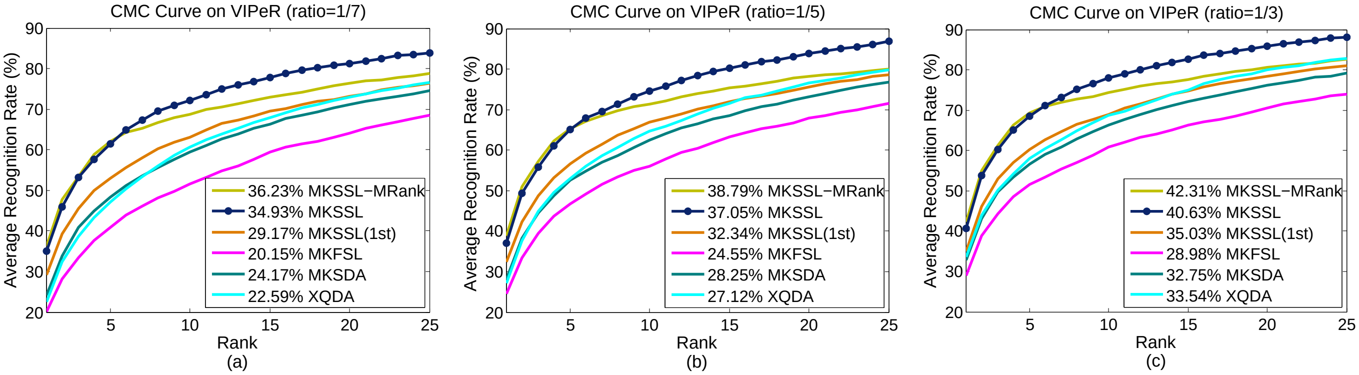


Fig. 3. Performance comparison of the proposed approach with different baselines on the VIPeR dataset with different settings of $ratio$. Rank-1 recognition rates are shown in the legends.

For the semi-supervised setting, we denote the proportion of labeled data in the training set as $ratio$ and the rest are used as unlabeled data. We evaluate the effectiveness of our approach with different settings of $ratio$. The dimension $r'$ ($r$) of the final subspace is fixed as the difference of the number of labeled persons and one.

The proposed approach is termed as MKSSL which includes a multi-kernel embedding. The baseline methods are constructed as follows:

- MKSSL(1st): The proposed approach with only one iteration.
- MKFSL: The fully-supervised subspace learning method described in Subsection III-A with the introduced multi-kernel embedding technique. Only the labeled data in the training set are used.
- MKSDA: The classic semi-supervised discriminant analysis method in [48] with the introduced multi-kernel embedding technique.
- XQDA: The state-of-the-art fully-supervised metric learning method in [2]. Only the labeled data in the training set is used.

In addition, in the testing stage the most common way is directly performing image matching using Euclidean distance in the learned distance space, which is also the default strategy of MKSSL and other compared methods. In this work, we also integrate the manifold ranking method [41] with the learned distance function of MKSSL in the testing stage to compute the similarity score between a probe image and a gallery image. We term this approach as MKSSL-MRank, in which the neighborhood graph of the manifold ranking method is constructed in the learned distance space of MKSSL instead of the original feature space. The performance of MKSSL-MRank will be evaluated in the following experiments. Note that, if not mentioned, our method and the above listed baseline methods all use the GOG descriptor in the following experiments.

### B. Results of Semi-Supervised Re-ID

In this subsection, we evaluate the performance of the proposed approach in the semi-supervised setting on the VIPeR, CUHK01, PRID2011, PRID450S, GRID, and 3DPeS datasets, compared with the baseline methods and reported results of state-of-the-art methods.

TABLE II
PERFORMANCE COMPARISON (CMC@RANK-R, %) OF OUR APPROACH WITH THE REPORTED RESULTS OF STATE-OF-THE-ART SEMI-SUPERVISED OR FULLY-SUPERVISED METHODS ON THE VIPER DATASET. A LARGER NUMBER INDICATES A BETTER RESULT.

| VIPeR | | r=1 | r=5 | r=10 | r=20 |
|---|---|---|---|---|---|
| Semi-supervised<br>$ratio = 1/3$ | MKSSL-MRank [Ours] | 42.3 | 69.4 | 74.4 | 80.6 |
| | MKSSL [Ours] | 40.6 | 68.5 | 78.1 | 85.9 |
| | LOMO+MKSSL [Ours] | 31.2 | 52.5 | 62.9 | 72.8 |
| | LOMO+LDNS [12] | 31.7 | 59.4 | 72.8 | 84.9 |
| | SSMFL [19] | 22.5 | 44.4 | 55.9 | 70.7 |
| | DLIterLap [18] | 32.5 | 61.8 | 74.3 | 84.1 |
| | SSCDL [17] | 25.6 | 53.7 | 68.1 | 83.6 |
| Fully-supervised<br>$ratio = 1$ | MKFSL [Ours] | 51.0 | 81.4 | 89.3 | 95.3 |
| | GOG+XQDA [3] | 49.7 | 79.7 | 88.7 | 94.5 |
| | LOMO+LDNS [12] | 42.3 | 71.5 | 82.9 | 92.0 |
| | LOMO+LSSCDL [63] | 42.7 | - | 84.3 | 91.9 |
| | Ensemble [31] | 45.9 | 77.5 | 88.9 | 95.8 |
| | LOMO+XQDA [2] | 40.0 | - | 80.5 | 91.1 |
| | Semantic [64] | 41.6 | 71.9 | 86.2 | 95.1 |
| | LOMO+MLAPG [16] | 40.7 | - | 82.3 | 92.4 |
| | MTL-LORAE [65] | 42.3 | 72.2 | 81.6 | 89.6 |
| | DLIterLap [18] | 38.9 | 70.8 | 78.5 | 86.1 |
| | UnL1Graph [39] | 41.5 | - | - | - |
| | MCKCCA [66] | 47.9 | - | 87.3 | 93.8 |

*1) Performance on the VIPeR Dataset:* The testing protocol for VIPeR is to randomly select 316 persons for training and 316 persons for testing. The training set is further split into two groups: one is labeled and the rest is unlabeled, according to the proportion $ratio$ of labeled data. To better evaluate the performance, we conduct multiple experiments on VIPeR by setting $ratio$ to $1/7$, $1/5$, and $1/3$, respectively.

We show the performance comparison of the proposed MKSSL approach with baseline methods in Fig. 3. We observe that our approach performs very well even when very fewer training data are labeled. When $ratio = 1/7$, MKSSL improves the performance of MKFSL by $14.78\%$ at rank-1, which indicates that the abundant unlabeled data have been well utilized. Similar improvements $12.5\%$ and $11.65\%$ at rank-1 can also be observed when $ratio = 1/5$ and $1/3$, respectively. With the increasing of $ratio$, the improvements drop gradually, since the size of unlabeled data is dropping. All the methods listed in Fig. 3 yield better performance when $ratio$ is increasing. We observe that MKSSL improves the performance of MKSSL(1st) by nearly $5\%$ at rank-1, which shows that the introduced iterative self-training strategy is, indeed,

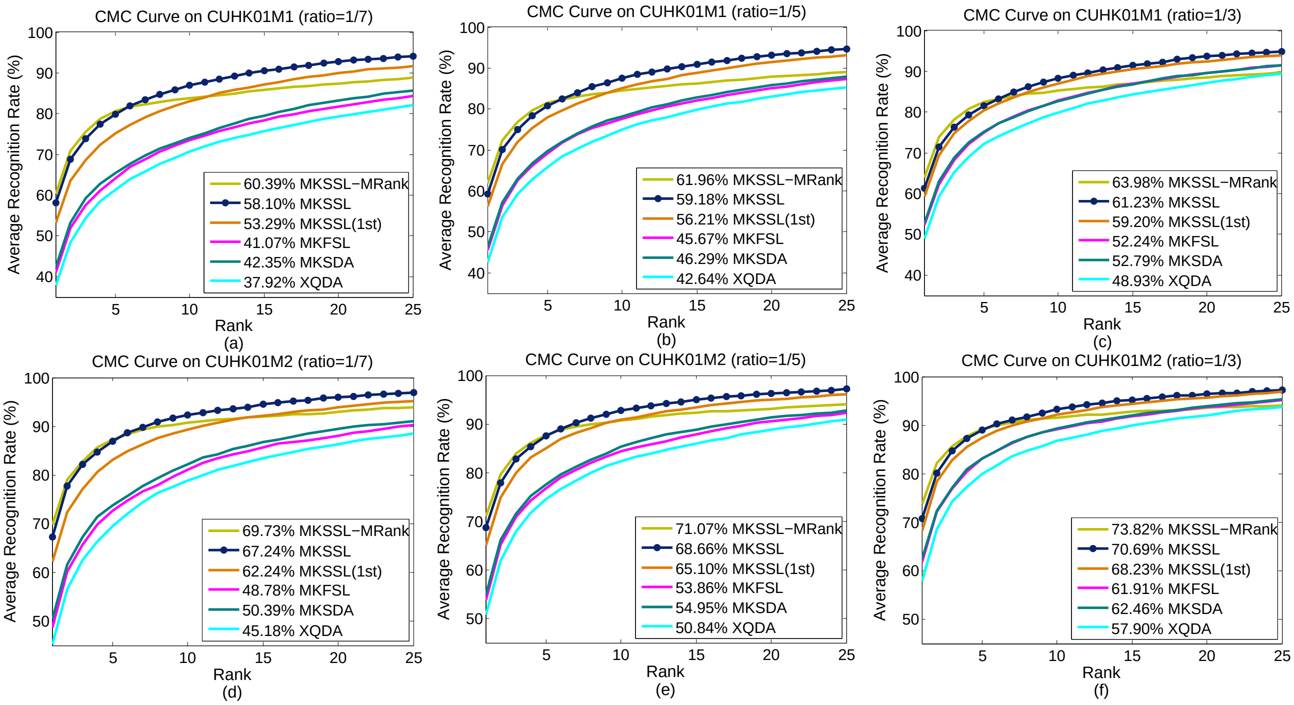


Fig. 4. Performance comparison of our approach with different baselines on the CUHK01 dataset with different settings of $ratio$. Both the single-shot matching (M=1) and the multi-shot matching (M=2) are applied. Rank-1 recognition rates are shown in the legends.

effective on VIPeR. It can also be observed that MKSSL(1st) performs better than MKSDA. The main difference between them is that the former is more like a semi-supervised extension of LPP ( [51]), while the latter is extended from LDA. The advantage of LPP on LDA has been demonstrated in [51]. By learning a discriminative subspace and a metric simultaneously, XQDA [2] performs better than MKFSL on VIPeR. Our MKSSL method surpasses XQDA by over 7% when $ratio = 1/3$ by exploiting unlabeled data. MKSSL reports the second-best rank-1 recognition rate 40.63% when $ratio = 1/3$. Benefiting from the effectiveness of the manifold ranking method [41], MKSSL-MRank improves the rank-1 recognition rate of MKSSL by 1.3%, 1.74%, and 1.68%, respectively, when $ratio = 1/7, 1/5$ and $1/3$. It demonstrates that, as a post-ranking algorithm, this unsupervised technique can complement the proposed approach very well. It refines the initial ranking scores computed in the learned distance space of MKSSL to yield a better ranking result by exploiting the manifold structure of unlabeled gallery data. In the manifold space, a higher rank will be assigned to gallery instances situated near to the probe sample, whilst locally nearby instances are encouraged to have similar ranks.

We also compare the performance of our approach with the reported results of state-of-the-art semi-supervised or fully-supervised results on VIPeR in Table II. From the results shown in Table II, we observe that MKSSL performs better than existing semi-supervised results on VIPeR. It mainly owes to the effectiveness of both the introduced self-trained subspace learning approach and the robust GOG descriptor. To compare MKSSL with LDNS [12], we also provide the result of MKSSL in Table II using the same LOMO descriptor. We observe that LOMO+LDNS [12] performs slightly better than LOMO+MKSSL on VIPeR. It may be because [12] learns a higher dimensional projection by combining pseudo classes and labeled classes into a new training set, while our method keeps unlabeled data in a Laplacian regularization term and do not increase the dimension of the final subspace although we have exploited abundant unlabeled data. In comparison, our approach is able to obtain more robust performance but at the cost of a small drop in recognition rate. It can also be observed in Table II, our semi-supervised results are almost comparable to the reported results of most state-of-the-art fully supervised methods while using much fewer labeled data.

*2) Performance on the CUHK01 Dataset:* We randomly partition the CUHK01 dataset into 486 persons for training and 485 persons for testing. The proportion of labeled data $ratio$ is set to $1/7$, $1/5$, and $1/3$ respectively. Correspondingly, there are 70, 98, and 162 persons involved respectively. To the best of our knowledge, we are the first to report semi-supervised results on CUHK01.

Fig. 4 shows the performance comparison of our approach with different baseline methods. From the single-shot results illustrated by Fig. 4 (a), Fig. 4 (b), and Fig. 4 (c), we can observe that MKSSL outperforms MKFSL by 17.03%, 13.51%, and 8.99% at rank-1 respectively. It reveals that the performance of the fully-supervised MKFSL method can be improved significantly by utilizing unlabeled data. With the increasing of unlabeled data, the contribution of unlabeled data becomes more obvious. Besides, MKSSL improves the rank-1 recognition rates of MKSSL(1st) by 4.81%, 2.97%, and 2.03%, respectively, when M=1. It shows that the iterative self-training strategy has enhanced the learning performance effectively. Our approach also outperforms the classic semi-supervised MKSDA method and the fully-supervised XQDA method. MKSSL-MRank improves the rank-1 recognition rate of MKSSL by over 2% on CUHK01. Similar improvements

TABLE III

Performance comparison (CMC@rank-r, %) of our approach with the reported results of state-of-the-art fully-supervised methods on the CUHK01 dataset. Both the single-shot matching (M=1) results and the multi-shot matching (M=2) results are shown. A larger number indicates a better result.

| **CUHK01** | | M=1 | | | M=2 | | |
|---|---|---|---|---|---|---|---|
| | | r=1 | r=5 | r=20 | r=1 | r=5 | r=20 |
| Semi-supervised $ratio = 1/3$ | MKSSL-MRank [Ours] | 64.0 | 82.4 | 88.6 | 73.8 | 89.1 | 93.5 |
| | MKSSL [Ours] | 61.2 | 81.6 | 93.7 | 70.7 | 89.0 | 96.4 |
| Fully-supervised $ratio = 1$ | MKFSL [Ours] | 62.0 | 82.9 | 94.2 | 72.2 | 89.4 | 97.0 |
| | GOG+XQDA [3] | 57.8 | 79.1 | 92.1 | 67.3 | 86.9 | 95.9 |
| | LOMO+LDNS [12] | - | - | - | 65.0 | 85.0 | 94.4 |
| | LOMO+LSSCDL [63] | - | - | - | 66.0 | - | - |
| | LOMO+XQDA [2] | 49.2 | 75.7 | 90.8 | 63.2 | 84.0 | 93.7 |
| | CPDL [67] | 59.5 | 81.3 | 93.1 | - | - | - |
| | MCPCNN [68] | 53.7 | 91.0 | 96.3 | - | - | - |
| | DeepRank [69] | 50.4 | 84.0 | 91.3 | - | - | - |
| | Deep [70] | 47.5 | 80.0 | - | - | - | - |
| | LOMO+MLAPG [16] | - | - | - | 64.2 | 85.4 | 94.9 |
| | Ensemble [31] | 53.4 | 76.4 | 90.5 | - | - | - |
| | MCKCCA [66] | 56.6 | - | 92.0 | 69.5 | - | 96.2 |

can also be observed in Fig. 4 (d), Fig. 4 (e), and Fig. 4 (f), when performing multi-shot matching.

We also compare the performance of our approach with the reported results of state-of-the-art fully-supervised methods on CUHK01 in Table III. From the results shown in Table III, we observe that the results of our semi-supervised approach outperform the reported results of most state-of-the-art fully-supervised methods while using very fewer labeled data. It indicates that we may not need to require all the training data to be labeled for a large re-ID dataset. We can obtain satisfactory performance by only labeling a small proportion of training data if using an effective semi-supervised learning strategy.

*3) Performance on the PRID2011, PRID450S, GRID, and 3DPeS Datasets:* For the evaluations on the PRID2011, PRID450S, GRID, and 3DPeS datasets, we use 100, 225, 125, and 95 individuals, respectively, for training. The remaining available individuals are used for testing. The proportion of labeled individuals in the training set is set to $ratio = 1/3$ for the four datasets. Fig. 5 shows the CMC performance of the proposed approach and baseline methods on the four datasets with $ratio = 1/3$. As shown in Fig. 5, by leveraging the abundant unlabeled data, our MKSSL method improves the rank-1 recognition rates of MKFSL by 10.8%, 12.05%, 7.12%, and 3.16%, respectively, on the four datasets. MKSSL improves the performances of MKSSL(1st) at rank-1 by 3.3%, 1.03%, 2%, and 0.12%, respectively. We observe that the iterative learning strategy of MKSSL yields nearly no improvement on MKSSL(1st) in Fig. 5 (d). Since 3DPeS is a multi-shot dataset in which the number of images for each person varies from 2 to 26 images. Compared to other five datasets, which only have 2 or 4 images for one individual, the neighborhood structure in 3DPeS can be discovered more easily by only one iteration.

We compare the performance of our approach with the reported results of state-of-the-art semi-supervised or fully-supervised methods on the four datasets in Table IV, Table V, Table VI, and Table VII, respectively. We observe that, by using the GOG descriptor, our semi-supervised results can outperform the reported results of most fully-supervised methods, while using much fewer labeled data.

TABLE IV

Performance comparison (CMC@rank-r, %) of our approach with the reported results of state-of-the-art semi-supervised or fully-supervised methods on the PRID2011 dataset. A larger number indicates a better result.

| **PRID2011** | | r=1 | r=5 | r=10 | r=20 |
|---|---|---|---|---|---|
| Semi-supervised *ratio* = 1/3 | MKSSL-MRank [Ours] | 31.4 | 53.5 | 63.2 | 73.0 |
| | MKSSL [Ours] | 29.2 | 52.4 | 62.1 | 72.8 |
| | LOMO+MKSSL [Ours] | 21.1 | 42.3 | 58.1 | 71.2 |
| | DLIterLap [18] | 22.1 | 45.3 | 56.5 | 66.3 |
| | LOMO+LDNS [12] | 24.7 | 46.8 | 58.2 | 68.2 |
| Fully-supervised *ratio* = 1 | MKFSL [Ours] | 34.2 | 58.0 | 66.6 | 78.2 |
| | GOG+XQDA [3] | 35.9 | 60.1 | 68.5 | 78.1 |
| | LOMO+LDNS [12] | 29.8 | 52.9 | 66.0 | 76.5 |
| | Ensemble [31] | 17.9 | 39.0 | 50.0 | 62.0 |
| | Mahalanobis [34] | 16.0 | - | 41.0 | 51.0 |
| | DLIterLap [18] | 25.2 | 51.9 | 62.9 | 71.6 |
| | UnL1Graph [39] | 30.1 | - | - | - |
| | LOMO+XQDA [2] | 26.7 | 49.9 | 61.9 | 73.8 |
| | MCKCCA [66] | 26.7 | - | 62.1 | 73.3 |

TABLE V

Performance comparison (CMC@rank-r, %) of our approach with the reported results of state-of-the-art semi-supervised or fully-supervised methods on the PRID450S dataset. A larger number indicates a better result.

| **PRID450S** | | r=1 | r=5 | r=10 | r=20 |
|---|---|---|---|---|---|
| Semi-supervised *ratio* = 1/3 | MKSSL-MRank [Ours] | 63.4 | 86.8 | 89.6 | 92.9 |
| | MKSSL [Ours] | 61.6 | 85.7 | 92.6 | 96.7 |
| Fully-supervised *ratio* = 1 | MKFSL [Ours] | 66.9 | 89.3 | 94.2 | 97.7 |
| | GOG+XQDA [3] | 68.4 | 88.8 | 94.5 | 97.8 |
| | LOMO+XQDA [2] | 62.6 | 85.6 | 92.0 | 96.6 |
| | LOMO+LSSCDL [63] | 60.5 | - | 88.6 | 93.6 |
| | MirrorKMFA [71] | 55.4 | 79.3 | 87.8 | 93.9 |
| | MEDVL [72] | 45.9 | 73.0 | 82.9 | 91.1 |
| | Transfer [64] | 44.9 | 71.7 | 77.5 | 86.7 |
| | Struct [73] | 44.4 | 71.6 | 82.2 | 89.8 |
| | SCNCD [27] | 41.6 | 68.9 | 79.4 | 87.8 |
| | MCKCCA [66] | 55.8 | - | 90.8 | 95.5 |

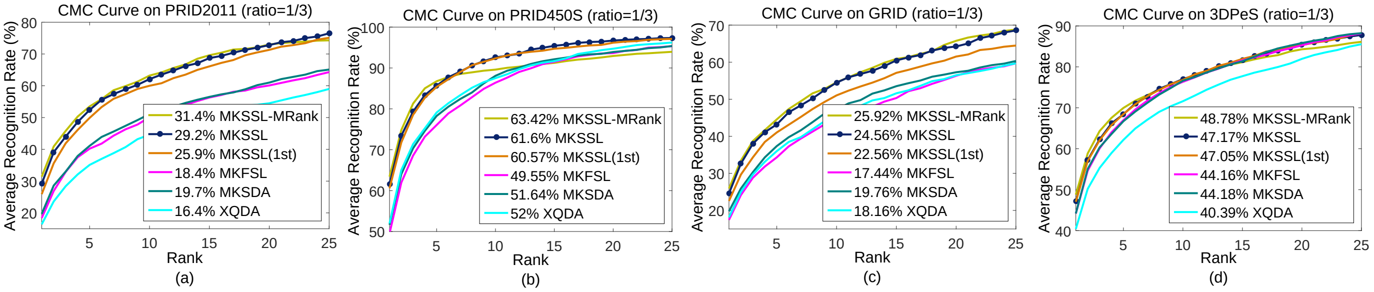


Fig. 5. Performance comparison of the proposed approach with different baselines on the six datasets: (a) PRID2011, (b) PRID450S, (c) GRID, and (d) 3DPeS. Rank-1 recognition rates (%) are shown in the legends.

TABLE VI
PERFORMANCE COMPARISON (CMC@RANK-R, %) OF OUR APPROACH WITH THE REPORTED RESULTS OF STATE-OF-THE-ART SEMI-SUPERVISED OR FULLY-SUPERVISED METHODS ON THE GRID DATASET. A LARGER NUMBER INDICATES A BETTER RESULT.

| GRID | | r=1 | r=5 | r=10 | r=20 |
|---|---|---|---|---|---|
| Semi-supervised<br>*ratio* = 1/3 | MKSSL-MRank [Ours] | 25.7 | 44.2 | 53.8 | 65.4 |
| | MKSSL [Ours] | 24.6 | 43.2 | 54.5 | 64.2 |
| Fully-supervised<br>*ratio* = 1 | MKFSL [Ours] | 26.4 | 47.2 | 55.8 | 67.8 |
| | GOG+XQDA [3] | 24.7 | 47.0 | 58.4 | 69.0 |
| | LOMO+LSSCDL [63] | 22.4 | - | 51.3 | 61.2 |
| | LOMO+XQDA [2] | 16.6 | - | 41.8 | 52.5 |
| | LOMO+MLAPG [16] | 15.6 | - | 40.5 | 52.5 |

TABLE VII
PERFORMANCE COMPARISON (CMC@RANK-R, %) OF OUR APPROACH WITH THE REPORTED RESULTS OF THE STATE-OF-THE-ART SEMI-SUPERVISED OR FULLY-SUPERVISED METHODS ON THE 3DPES DATASET. A LARGER NUMBER INDICATES A BETTER RESULT.

| 3DPeS | | r=1 | r=5 | r=10 | r=20 |
|---|---|---|---|---|---|
| Semi-supervised<br>*ratio* = 1/3 | MKSSL-MRank [Ours] | 48.8 | 69.9 | 76.9 | 84.3 |
| | MKSSL [Ours] | 47.2 | 68.4 | 77.0 | 85.4 |
| | SVM+CRF [38] | 45.5 | - | - | - |
| Fully-supervised<br>*ratio* = 1 | MKFSL [Ours] | 54.4 | 77.9 | 86.7 | 94.1 |
| | Ensemble [31] | 53.3 | - | - | - |
| | KLFDA [11] | 54.0 | 77.7 | 85.9 | 92.4 |

### C. On the Iterative Self-training Strategy

In this subsection we investigate the effect of the iterative self-training strategy and evaluate the robustness of our approach. We use the output of each iteration to perform person matching on the test set with different settings of $ratio$. The VIPeR and CUHK01 datasets are used as two examples. We show the average rank-1 recognition rates of MKSSL at each iteration in Fig. 7.

First, we can observe in Fig. 6 that the iterative self-training strategy significantly improves the performance on VIPeR and CUHK01, which demonstrates its effectiveness. A remarkable performance improvement can be observed on VIPeR when $ratio \in \{1/20, 1/10\}$ and on CUHK01 when $ratio \in \{1/40, 1/30, 1/20, 1/10\}$. The performance increases relatively slowly from $ratio = 1/10$ to $ratio = 1/3$. With the increasing of $ratio$, the performance starts to become stable. Therefore, we can empirically conclude that very large labeled training sets may not be necessary for re-ID if applying the proposed approach in this work. When $ratio$ is very small, e.g. $ratio = 1/40$ on VIPeR or $ratio = 1/70$ on CUHK01, the effect of the iterative self-training strategy is not significant. Because in our setting, the subspace is initialized by the available labeled data. When $ratio = 1/40$, there are only ten labeled persons on VIPeR, which is not sufficient. Therefore, $ratio$ is not suggested to be set as a very small value. The value that $ratio >= 1/20$ is recommended.

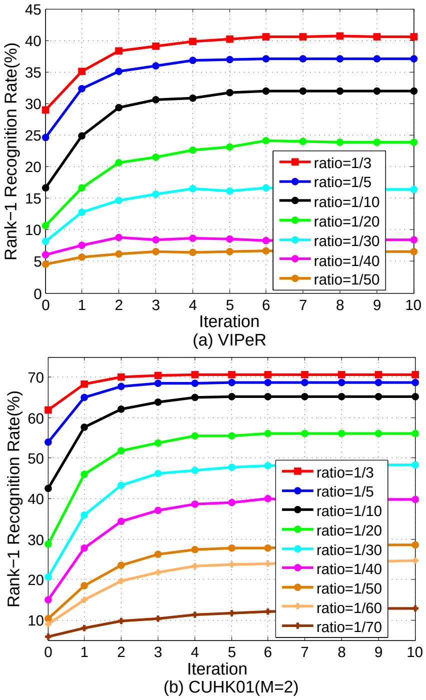
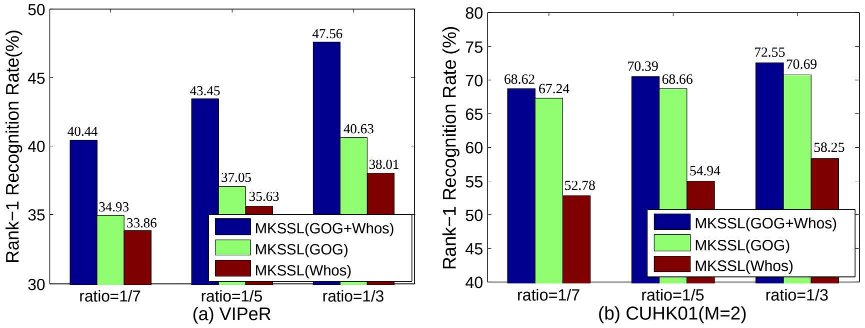


Fig. 6. Rank-1 recognition rates of MKSSL versus iteration number with different settings of $ratio$. (a) Experiments on VIPeR; (b) Experiments on CUHK01 (M=2). Note that the rank-1 recognition rates of MKFSL are illustrated at the starting point as a comparison.

Secondly, it can also be observed in Fig. 6 that the iteration procedure converges fast in most cases. Usually it only takes three to six iterations. The larger the $ratio$ is, the faster the iteration procedure converges. The convergence time mainly depends on the size of unlabeled training data.

Besides, it should be mentioned that our approach also suffers from the model drift problem. It is a common problem in self-training based methods. Specifically, error accumulation is usually inevitable during self-training iteration procedure. This problem can be observed in Fig. 6 (a) in which the performance degrades after 2 iterations when $ratio = 1/40$. A very small $ratio$ results in a poor initial projection, which makes the model drift easily. As shown in Fig. 6, our approach shows empirical robustness with a large $ratio$ (e.g. 1/10) in most cases. The risk of model drift seems to have been well controlled in our approach. Because a large $ratio$ brings a



(a) VIPeR

(b) CUHK01(M=2)

Fig. 7. The effect of multiple feature fusion with different settings of $ratio$. (a) Experiments on VIPeR; (b) Experiments on CUHK01 (M=2).

good initialization, resulting in more accurate neighborhood structure. Here, the labeled term in Eq. (4) can prevent our model from going too far off a reasonable solution. By the way, to avoid the model shift, in this work we fix the dimension of the self-trained subspace as the difference of the number of labeled persons and one. We do not increase the dimension of subspace although we have exploited abundant unlabeled data, since the pseudo pairwise relationships may inevitably contain a few mismatching pairs. We may be able to obtain a better performance in a higher-dimensional self-trained subspace but at a higher risk of model drift.

### D. On the Complementation of Multiple Features

In Subsection III-C, we introduce the multi-kernel embedding technique [55] into the proposed approach. In the experiment results and analyses reported in Subsection IV-B, we only utilize one feature descriptor. In this subsection, we exploit the introduced kernelization technique to consider multiple feature descriptors and evaluate the effect of this feature fusion strategy.

We use the 5138-D Whos descriptor [10], together with the default 27622-D GOG descriptor used above, to learn a kernelized projection using the introduced kernelization technique. For each descriptor, 11 Gaussian kernels are constructed using the same setting in Subsection IV-A3. As shown in Fig. 7, we compare the performance of MKSSL using two features (MKSSL(GOG+Whos)) with that of MKSSL using only one feature (MKSSL(GOG), MKSSL(Whos)) on the VIPeR and CUHK01 datasets, respectively. We observe that, by exploiting this feature fusion strategy, MKSSL(GOG+Whos) achieves a state-of-the-art rank-1 recognition rate 47.56% on VIPeR when $ratio = 1/3$, surpassing MKSSL(GOG) by nearly 7% and MKSSL(Whos) by over 9%. Similar improvements are also observed when $ratio = 1/5$ or $ratio = 1/7$. It demonstrates that these two feature descriptors strongly complement each other on VIPeR. On the CUHK01 dataset, MKSSL(GOG+Whos) improves the multi-shot rank-1 recognition rate of MKSSL(GOG) by nearly 2% and that of MKSSL(Whos) by over 14%, when $ratio = 1/3$. We can conclude that the introduced multi-kernel embedding strategy is flexible and effective, which can significantly enhance the performance by exploiting the complementation of multiple feature representations.

### E. On the Sensitivity of Parameters

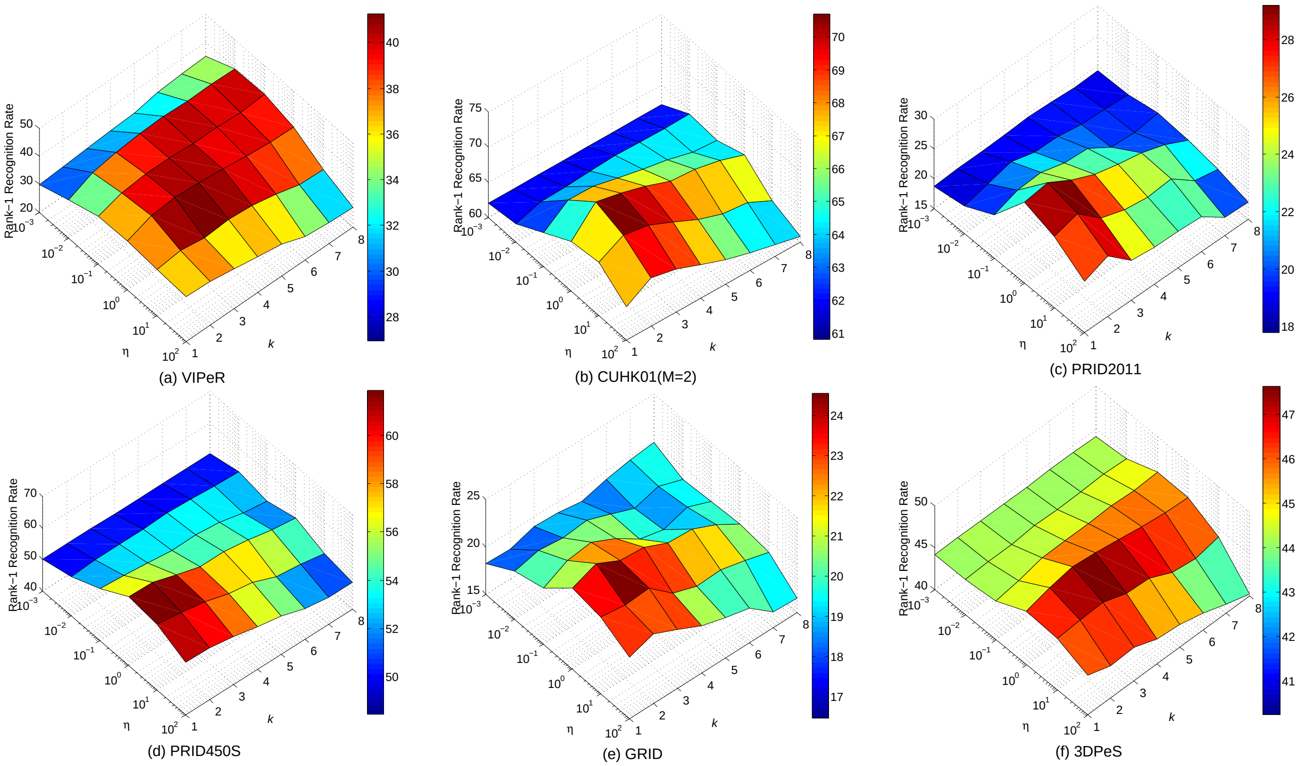
In this subsection, we evaluate the effects of parameters used in this work. There are three tuning parameters in our approach. Parameter $\vartheta$ is a small positive parameter used to regularize the matrix at the right side of Eq. (13) to avoid the singularity of matrix. It is a commonly-used regularization technique in the eigenvalue problems. We empirically set $\vartheta$ to 0.01. Here, we mainly analyze the effects of the two main parameters: $\eta$ and $k$.

Parameter $\eta$ in Eq. (4), Eq. (5), and Eq. (13) modulates the effects of the regularization term $\mathcal{R}(\mathbf{X}_u, \mathbf{U}, \mathbf{W}^u)$ constructed using unlabeled data. The number of nearest neighbors $k$ is vital in the kNN graph. In this work, we set them as $\eta = 1$ and $k = 2$ by cross-validation on the training set of CUHK01 and fix them for all datasets. To fairly investigate the effects of $\eta$ and $k$, we observe the change of rank-1 recognition rate of MKSSL on the six datasets by varying $\eta$ and $k$ simultaneously. Specifically, the value of $\eta$ is chosen from the set $\{0.001, 0.01, 0.1, 1, 10, 100\}$, and $k$ is increased from 1 to 8 with step 1. The experimental results are illustrated in Fig. 8.

As observed in Fig. 8, our approach performs well on most datasets when $0.1 \leq \eta \leq 10$ and $k \in \{2, 3, 4\}$. On the PRID2011 and PRID450S datasets, our approach yields a good performance when $k = 1$. On the 3DPeS dataset, a high rank-1 recognition rate can be observed when $k = 4$, since it is a multi-shot dataset. Overall speaking, our approach can obtain good performance under a wide range of the parameter values.

### F. Time Complexity Analysis

The main computational cost of our MKSSL method is dominated by two parts. The first part comes from the computation of the initial projection learned using labeled data only. Its complexity is stemmed from solving a kernelized eigenvalue problem, which is approximated by $O\left(n^3 + rn^2\right)$, where $n$ is the number of labeled training images, which equals to the size of the kernel matrix constructed using labeled data, and $r$ is the dimension of the initial low-dimensional subspace. The second part is solving the kernelized eigenvalue problem in Eq. (13) using both labeled and unlabeled data. This procedure is repeated several times by applying the iterative self-training strategy. Its complexity is approximated by $O\left(T\left((n+u)^3 + r'(n+u)^2\right)\right)$, where $r'$ is the dimension of the final subspace, $u$ is the number of unlabeled training images, and $T$ is the iteration number. Therefore, the



(a) VIPeR

(b) CUHK01(M=2)

(c) PRID2011

(d) PRID450S

(e) GRID

(f) 3DPeS

Fig. 8. The sensitivity analyses of $\eta$ and $k$ by choosing $\eta$ from the set $\{0.001, 0.01, 0.1, 1, 10, 100\}$ and $k$ from 1 to 8 with step 1 on six datasets: (a) VIPeR, (b) CUHK01, (c) PRID2011, (d) PRID450S, (e) GRID, and (f) 3DPeS. To fairly investigate the effects of $\eta$ and $k$, we vary these two parameters simultaneously to observe the change of rank-1 recognition rate of MKSSL. Overall, our approach can perform well under a wide range of the parameter values.

total complexity is $O\Big(n^3+rn^2+T\left((n+u)^3+r'(n+u)^2\right)\Big)$, which is mainly determined by the number of training images and the iteration number. The proposed MKSSL approach is implemented in Matlab on a 2.9GHz CPU PC with 32G RAM.

Here, we present the practical runtime of MKSSL on the whole VIPeR dataset and the whole CUHK01 dataset with $ratio = 1/3$. The average training and testing time on VIPeR are 2.88s and 0.02s respectively. The average iteration number on VIPeR is $5.8$. The average training and testing time on CUHK01 (M=2) are 28.06s and 0.09s respectively. The average iteration number on CUHK01 (M=2) is $4.1$.

## V. Conclusion

In this work, we propose an effective semi-supervised approach for person re-ID which can leverage both labeled and unlabeled data. It formulates re-ID as a subspace learning problem by learning a discriminative projection to map the person images from disjoint camera views into a common subspace where person matching can be easily performed. It presents a self-training based subspace learning strategy in which the unlabeled person images are exploited by constructing the pseudo pairwise relationships. An iterative learning strategy is introduced to refine the pseudo pairwise relationships, which significantly enhances the learning performance. It is also able to explore the complementary characteristic of multiple feature representations for re-ID. Experimental results on multiple challenging datasets have demonstrated the effectiveness and robustness of the proposed approach.

## References

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016.

[2] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197–2206.

[3] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1363–1372.

[4] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *IEEE International Conference on Computer Vision*, 2015, pp. 1116–1124.

[5] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification," in *European Conference on Computer Vision*, 2016, pp. 868–884.

[6] Y.-C. Chen, W.-S. Zheng, J.-H. Lai, and P. Yuen, "An asymmetric distance model for cross-view feature mapping in person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[7] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.

[8] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, and S. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 663–671, 2006.

[9] R. R. Varior, G. Wang, J. Lu, and T. Liu, "Learning invariant color features for person reidentification," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3395–3410, 2016.

[10] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1629–1642, 2015.

[11] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *European Conference on Computer Vision*, 2014, pp. 1–16.

[12] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space

for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[13] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2288–2295.

[14] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3610–3617.

[15] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1268–1277.

[16] S. Liao and S. Z. Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in *IEEE International Conference on Computer Vision*, 2015, pp. 3685–3693.

[17] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3550–3557.

[18] E. Kodirov, T. Xiang, and S. Gong, "Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification," in *BMVC*, vol. 3, 2015, p. 8.

[19] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2013, pp. 111–116.

[20] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 3, no. 1, pp. 1–130, 2009.

[21] L. Zheng, H. Zhang, S. Sun, M. Chandraker, and Q. Tian, "Person re-identification in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[22] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *European Conference on Computer Vision*, 2008, pp. 262–275.

[23] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1741–1750.

[24] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3318–3325.

[25] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," *arXiv preprint arXiv:1701.07717*, 2017.

[26] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang, "Improving person re-identification by attribute and identity learning," *arXiv preprint arXiv:1703.07220*, 2017.

[27] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, "Salient color names for person re-identification," in *European Conference on Computer Vision*, 2014, pp. 536–551.

[28] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586–3593.

[29] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *Workshops on European Conference on Computer Vision*. Springer, 2012, pp. 413–422.

[30] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1622–1634, 2013.

[31] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1846–1855.

[32] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3656–3670, 2014.

[33] A. Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2666–2672.

[34] P. M. Roth, M. Hirzer, M. Koestinger, C. Beleznai, and H. Bischof, *Mahalanobis distance learning for person re-identification*. Springer, 2014, pp. 247–267.

[35] X. Yang, M. Wang, L. Zhang, and D. Tao, "Empirical risk minimization for metric learning using privileged information," in *IJCAI*, 2016.

[36] L. An, S. Yang, and B. Bhanu, "Person re-identification by robust canonical correlation analysis," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1103–1107, 2015.

[37] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[38] S. Karaman, G. Lisanti, A. D. Bagdanov, and A. Del Bimbo, "Leveraging local neighborhood topology for large scale person re-identification," *Pattern Recognition*, vol. 47, no. 12, pp. 3767–3778, 2014.

[39] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Person re-identification by unsupervised\ ell _1 graph learning," in *European Conference on Computer Vision*, 2016, pp. 178–195.

[40] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 144–151.

[41] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *IEEE International Conference on Image Processing*. IEEE, 2013, pp. 3567–3571.

[42] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.

[43] G. J. McLachlan, "Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation in discriminant analysis," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 365–369, 1975.

[44] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, 1995, pp. 189–196.

[45] S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised clustering by seeding," in *International Conference on Machine Learning*, 2002.

[46] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *IEEE Workshops on Application of Computer Vision*, 2005, pp. 29–36.

[47] M. Loog, "Contrastive pessimistic likelihood estimation for semi-supervised classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 462–475, 2016.

[48] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–7.

[49] S. C. Hoi, W. Liu, and S.-F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–7.

[50] M. Sugiyama, T. Id, S. Nakajima, and J. Sese, "Semi-supervised local fisher discriminant analysis for dimensionality reduction," *Machine learning*, vol. 78, no. 1-2, pp. 35–61, 2010.

[51] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, 2003, p. None.

[52] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *International Conference on Machine Learning*. ACM, 2006, pp. 905–912.

[53] Y. Zhang and D.-Y. Yeung, "Transfer metric learning with semi-supervised extension," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, p. 54, 2012.

[54] J. Yu, M. Wang, and D. Tao, "Semisupervised multiview distance metric learning for cartoon synthesis," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4636–4648, 2012.

[55] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 3013–3024, 2014.

[56] L. Shao, L. Liu, and M. Yu, "Kernelized multiview projection for robust action recognition," *International Journal of Computer Vision*, vol. 118, no. 2, pp. 115–129, 2016.

[57] S. Qiu and T. Lane, "A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 6, no. 2, pp. 190–199, 2009.

[58] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 31–44.

[59] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian conference on Image analysis*. Springer, 2011, pp. 91–102.

[60] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 106–129, 2010.

[61] D. Baltieri, R. Vezzani, and R. Cucchiara, "3dpes: 3d people dataset for surveillance and forensics," in *International ACM Workshop on Multimedia access to 3D Human Objects*, Scottsdale, Arizona, USA, Nov. 2011, pp. 59–64.

[62] G. Lisanti, I. Masi, and A. Del Bimbo, “Matching people across camera views using kernel canonical correlation analysis,” in *International Conference on Distributed Smart Cameras*. ACM, 2014, p. 10.

[63] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan, “Sample-specific svm learning for person re-identification,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[64] Z. Shi, T. M. Hospedales, and T. Xiang, “Transferring a semantic representation for person re-identification and search,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[65] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, “Multi-task learning with low rank attribute embedding for person re-identification,” in *IEEE International Conference on Computer Vision*, 2015, pp. 3739–3747.

[66] G. Lisanti, S. Karaman, and I. Masi, “Multi channel-kernel canonical correlation analysis for cross-view person re-identification,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2017.

[67] S. Li, S. Ming, and Y. Fu, “Cross-view projective dictionary learning for person re-identification,” in *IJCAI*, 2015.

[68] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, “Person re-identification by multi-channel parts-based cnn with improved triplet loss function,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1335–1344.

[69] S. Z. Chen, C. C. Guo, and J. H. Lai, “Deep ranking for person re-identification via joint representation learning,” *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2353–2367, 2016.

[70] E. Ahmed, M. Jones, and T. K. Marks, “An improved deep learning architecture for person re-identification,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[71] Y.-C. Chen, W.-S. Zheng, and J. Lai, “Mirror representation for modeling view-specific transform in person re-identification,” in *IJCAI*, 2015, pp. 3402–3408.

[72] Y. Yang, Z. Lei, S. Zhang, H. Shi, and S. Z. Li, “Metric embedded discriminative vocabulary learning for high-level person representation,” in *International Joint Conference on Artificial Intelligence*, 2016.

[73] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, “Person re-identification with correspondence structure learning,” in *IEEE International Conference on Computer Vision*, 2015.