Research of Metric Learning-Based Method for Person Re-Identification by Intelligent Computer Vision Technology

Qiyu Rao

School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China
qiyurao2021@buaa.edu.cn

Abstract. Person re-identification technology aims to establish an efficient metric model for similarity distance measurement of pedestrian images. Candidate images captured by different camera views are ranked according to their similarities to the target individual. However, the metric learning-based method, which is commonly used in similarity measurement, often failed in person re-identification tasks due to the drastic variations in appearance. The main reason for its low identification accuracy is that the metric learning method is over-fitting to the training data. Several types of metric learning methods which differ from each other by the distribution of sample pairs were summarized in this article for analysing and easing the metric learning methods’ over-fitting problem. Three different metric learning methods were tested on the VIPeR dataset. The distributions of the distance of the positive/negative training/test pairs are displayed to demonstrate the over-fitting problem. Then, a new metric model was proposed by combining the thoughts of binary classification and multi-class classification. Related verification experiments were conducted on VIPeR dataset. Besides, the semi-supervised metric learning approach was introduced to alleviate the over-fitting problem. The experimental results reflect gap between training pairs and test pairs in the metric subspace. Therefore, reducing the difference between training data and test data is a promising way to improve the identification accuracy of metric learning method.

1. Introduction

Person re-identification technology has always been a hot spot in the field of computer vision. This technology tries to match pedestrian images cross view of cameras in non-overlapping surveillance system. Person re-identification has important research significance for its great role in many scenarios. However, person re-identification encounters several notable difficulties which face-recognition doesn’t encounter. In the application scenario of face recognition, users have relatively fixed postures and the light is under control in a close-up shooting system. While in the people re-identification, large intra-class variations and small inter-class variations due to drastic appearance variations often cause identification failure.

Early researches on person re-identification focused on ways to manually design better visual features or learn better similarity measures. Many discriminative appearance-based methods and metric learning-based methods were proposed, including appearance models Symmetry-Driven...
Accumulation of Local Features (SDALF) [1], Shape and appearance context [2], Salient Color Names Color Descriptor (SCNCD) [3], Local Maximal Occurrence (LOMO) [4], Attributes features [5], Transferring Semantic Representation [6] and metric learning models Large Margin Nearest Neighbor (LMNN) [7], Information Theoretic Metric Learning (ITML) [8]. Keep It Simple and Straight Metric (KISSME) [9], Probability Relative Distance Comparison (PRDC) [10], Relaxed Pairwise Learned Metric [11], Cross-view Quadratic Discriminant Analysis (XQDA) [12], RankSVM [13], Dual-regularized KISS Metric Learning (DRKISS) [14].

Early deep learning models also faced the drastic variations of appearance features and low identification accuracy problem [15-18]. These works mainly formulated the deep structure based on the Siamese network [15, 16]. The recently proposed deep models tried to learn efficient models to align body structures of pedestrians. Others introduced the GAN to generate samples of different pose to reduce the influence of misalignment problem. He et. al. proposed a Foreground-aware Pyramid Reconstruction (FPR) [19] to deal with the misalignment problem. It was an alignment-free deep model which utilized the reconstruction error over spatial pyramid feature for similarity measurement of pedestrian images. Karmakar et. al. [20] introduced the Generative Adversarial Networks (GAN) method combined with the pose regression and feature fusion to establish a pose invariant representation. Zhao et. al. [21] proposed a SpindleNet deep learning model, which introduced the thought of convolutional pose machine (CPM) to estimate the pose of pedestrian. The SpindleNet modeled a body structure descriptor by 14 points to obtain feature of local image patches of different body structures. Zheng et. al. [22] utilized the GAN network to generate samples to train the identification model. And the generated pedestrian images were labeled as smoothing for outliers. It improved the baseline network effectively. Qian et. al. [23] proposed a Person Transfer GAN network, which utilized the GAN network to generate 8 pedestrian images of different poses. The features of original pedestrian image and generated images were extracted together. A more robust feature representation was proposed by feature fusion.

The deep learning-based method utilizes the foreground-aware [22, 24], GAN [22, 23], and pose detection [25] approaches to reduce the impact of appearance variations. Yet, the metric learning-based method still suffers from the drastic appearance variations problem. The metric learning methods have serious over-fitting problem. In this paper, the over-fitting problem of metric learning-based method were analyzed. And the semi-supervised learning approach was introduced to solve this problem.

2. Similarity Measurement for Person Re-identification

2.1. Similarity Distance Metric Model

2.1.1. Euclidean distance. Euclidean distance is the most direct distance function to measure the difference between two points in the European Space. Assume that \( X = \{ \mathbf{x}_i \}_{1 \leq i \leq C} \) denotes a dataset, which consists of \( C \) class of data. Let \( \mathbf{x}_i \) denotes the sample of the \( i \)-th class, where \( \mathbf{x}_i \in \mathbb{R}^n \) is an \( n \)-dimensional feature vector. Then, the Euclidean distance between \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) and \( \mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jn}) \) is expressed as follows:

\[
d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2}
\]  

(1)

The corresponding vector expression of formula (1) is as follows:

\[
d(x_i, x_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}
\]  

(2)
Euclidean distance treats each dimension of sample features equally. However, different features perform differently in the step of separating the positive samples from the negative samples. Generally speaking, features are divided into main features and irrelevant features. Some irrelevant features should be ignored to reduce the interference of environmental factors.

2.1.2. Mahalanobis Distance. Mahalanobis distance corresponds to distribution of samples. Considering the sample pair \((x_i, x_j)\) of dataset \(X\), the Mahalanobis distance is defined as follows,

\[
d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)}
\]

(3)

Where \(M\) denotes the Mahalanobis matrix and \(M \geq 0\). \(x_i, x_j\) are feature vectors of two images of a sample pair. \(M\) is learned from the training data. The Mahalanobis distance can effectively avoid the interference between variables when the training data could represent the distribution of all samples. However, the metric learning method requires large number of samples. The metric model would be over-fitting to the training data when the training data are too small to describe the distribution of the population properly.

2.2. Mahalanobis Distance Metric Learning-based on Pairwise Data

2.2.1. KISSME. The KISSME statistical discriminant model introduced posterior probability on the basis of pairwise constraints to establish a Mahalanobis distance metric learning model. Define the null hypothesis \(H_0: (x_i, x_j)\) is not similar. The alternative hypothesis \(H_1: (x_i, x_j)\) is not similar. The likelihood ratio function of positive and negative sample pairs is established to measure the difference between samples \(x_i\) and \(x_j\) as follows:

\[
\delta(x_i, x_j) = \log \left( \frac{p(x_i, x_j | H_0)}{p(x_i, x_j | H_1)} \right)
\]

(4)

According to the likelihood function (4), when the likelihood of dissimilarity of the sample pair \((x_i, x_j)\) to be identified is greater than the likelihood of similarity, the value of the function is greater than 0. Then we accept the null hypothesis. \(x_i\) and \(x_j\) are not similar. Otherwise, the function value is less than 0. Then we refuse the null hypothesis. \(x_i\) and \(x_j\) are not similar. Assume that the difference vector of positive/negative pair follows the Gaussian distribution of zero mean. The probability of sample pair \((x_i, x_j)\) belongs to dissimilar/similar pair is defined as follows,

\[
p(x_i, x_j | H_0) = f_0(x_i - x_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_0|^{1/2}} e^{-\frac{1}{2}(x_i - x_j)^T \Sigma_0^{-1}(x_i - x_j)}
\]

(5)

\[
p(x_i, x_j | H_1) = f_1(x_i - x_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_1|^{1/2}} e^{-\frac{1}{2}(x_i - x_j)^T \Sigma_1^{-1}(x_i - x_j)}
\]

(6)

Where \(\Sigma_0\) denotes the covariance matrix of the positive pairs. \(\Sigma_1\) denotes the covariance matrix of the negative pairs. Then, function (4) can be rewritten as in (7)
\[ f(x_i, x_j) = (x_i - x_j)^T \Sigma^{-1}(x_i - x_j) + \log(\|\Sigma^{-1}\|) - (x_i - x_j)^T \Sigma^{-1}(x_i - x_j) - \log(\|\Sigma^{-1}\|) \] (7)

It is obvious that \(\log(\|\Sigma^{-1}\|)\) and \(\log(\|\Sigma^{-1}\|)\) are constants. Removing the \(\log(\|\Sigma^{-1}\|)\) and \(\log(\|\Sigma^{-1}\|)\), the Mahalanobis distance-based similarity distance metric model is defined as follows:

\[ d(x_i, x_j) = (x_i - x_j)^T (\Sigma^{-1} - \Sigma^{-1})(x_i - x_j) \] (8)

Dual-regularized KISS Metric Learning (DR-KISS) method introduced the regularization approach for covariance matrix estimation. The two regularized covariance matrices is defined as follows:

\[ \hat{\Sigma}_{+1} = (1 - \gamma_0)\hat{\Sigma}_0 + \gamma_0 \alpha_0 I \] (9)

\[ \hat{\Sigma}_{-1} = (1 - \gamma_1)\hat{\Sigma}_1 + \gamma_1 \alpha_1 I \] (10)

Where \(\alpha_0 = (1/d)tr(\hat{\Sigma}_0)\), \(\alpha_1 = (1/d)tr(\hat{\Sigma}_1)\). \(\gamma_0\) and \(\gamma_1\) are respectively regularization parameters of covariance matrices of the positive pairs and the negative pairs. By regularizing the covariance matrices, DR-KISSME method efficiently ease the over-fitting problem and improve the generalization ability of metric model.

2.2.2. PCCA. Pairwise Constrained Component Analysis (PCCA) [26] models the metric learning model-based on the pairwise data. This model learns a projection matrix which projects the raw features to a low dimension feature vector. And the Euclidean distance is used to measure the similarity-based on the projection feature vector. PCCA method formulate penalty items of different kinds of sample pairs to reduce the influence of dimension decline. The optimization objective function is as follows:

\[ \min E(L) = \sum_{ij} \left[ y_{ij} \left( D^2_{ij}(x_i, x_j) - 1 \right) \right] \] (11)

\[ D^2_{ij}(x_i, x_j) = \|Lx_i - Lx_j\|^2 \] (12)

Where \(l_\beta(x) = (1/\beta)\log(1 + e^{\beta x})\) is a generalized logistic function. \(L\) is the mapping function which projects the high-dimension feature vector to low-dimension feature vector. \(y_{ij} = \{-1, +1\}\) is the label information of sample pair \((x_i, x_j)\). \(y_{ij} = 1\) means that \((x_i, x_j)\) is positive pair. Otherwise, \(y_{ij} = -1\) means that \((x_i, x_j)\) is negative pair.

2.3. Projection Subspace Learning-based Method

The Mahalanobis distance could be defined as different forms. The projection subspace learning methods formulate the person re-identification as a classification problem, which tries to learn a discriminative classification boundary separating the positive samples from negative samples. Then, the matrix \(M\) is decomposed as follows,

\[ d_{\alpha}(x_i, x_j) = \sqrt{(x_i - x_j)^T M(x_i - x_j)} = \sqrt{(x_i - x_j)^T W^T W (x_i - x_j)} \] (13)

2.3.1. Unsupervised metric learning method. The projection subspace learning method could be divided into unsupervised and supervised metric learning methods. The common unsupervised methods are Principal Component Analysis (PCA) [27], Local Retention Projection (LRP) [28], et. al.

1) PCA. measures the distribution of samples by learning a Mahalanobis distance. It aims to find \( r \) eigenvectors to express the original features linearly. Principal component analysis (PCA) algorithm preserves the original data information by maximizing the variance of projection features. The centre of samples in the original feature space is 

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

\( N \) denotes the number of the samples. The variance of samples in the projection subspace is as follows,

\[
\frac{1}{N} \sum_{i=1}^{N} (w^T \bar{x} - w^T \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{N} (w^T x_i - w^T \bar{x})^T (w^T x_i - w^T \bar{x})
\]  \( \text{(14)} \)

Where \( w \) is the projection vector which projects the image feature vector \( x_i \) to metric subspace. For convenience of calculation, the sample data is standardized by 

\[
x_i' = x_i - \bar{x}
\]

The formula (14) can be rewritten as follows:

\[
\frac{1}{N} \sum_{i=1}^{N} (w^T x_i' - w^T \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{N} (w^T x_i')(w^T x_i')^T = \frac{1}{N} \sum_{i=1}^{N} w_i x_i x_i'^T w = \frac{1}{N} w^T A w
\]  \( \text{(15)} \)

2) Local Retention Projection (LRP). characterized the local neighbourhood relationship of high-dimensional data based on Graph theory. This method learned a projection subspace which keeps local neighbourhood relations in relatively low-dimensional projection subspace. LRP model used the Euclidean distance to measure the distance between samples in the projection subspace. The distance was mapped to a Gaussian distribution. The similarity measurement model is as follows:

\[
f(x_i, x_j) = \exp \left( -\frac{||x_i - x_j||^2}{2\sigma^2} \right)
\]  \( \text{(16)} \)

According to the distance function between two samples as defined in function (16), the higher the similarity, the closer the function value is to 0. Then, the metric model is established as follows,

\[
d(x_i, x_j) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} ||x_i - x_j||^2 f(x_i, x_j)
\]  \( \text{(17)} \)

3) Large margin nearest neighbour (LMNN). introduced the thought of large margin to Mahalanobis distance learning. It defined the samples of same class to be the nearest neighbours. The samples of different classes were separated from each other by the large margin. The goal of the Mahalanobis distance model was to separate samples of different classes and gather samples of same class together. Then the objective function of the optimization model was formulated as follows,

\[
\min_{L} \{ \sum_{i=1}^{N} (1-y_{ij}) \left[ 1 + ||L(x_i, x_j)||^2 - L(x_i, x_j) \right]^2 \}
\]  \( \text{(18)} \)

Where \( L() \) denotes the metric function of similarity measurement. It is a Mahalanobis distance. \( (x_i, x_j) \) was positive pair. \( (x_i, x_j) \) was negative pair. \( y_{ij} \) denotes the label of the sample pair. \( y_{ij} = 1 \) means that \( (x_i, x_j) \) is positive; otherwise, \( (x_i, x_j) \) is negative when \( y_{ij} = 0 \). According to the formula (18), the total objective function of the optimization model is as follows:
\[
\min \varepsilon(L) = \sum_{i=1}^{n} \left[ \|x_i - x'_i\|^2 + \lambda \sum_{j=1}^{n} (1 - y_{ij}) \left[ 1 + \|x_i - x'_j\|^2 - \|x_i - x'_j\|^2 \right] \right]
\]  \hspace{1cm} (19)

Where \(o_i\) denotes the set of the same class to the \(i\)-th sample. By enhancing the correlation of the positive pairs and weakening the correlation of the negative pairs, the model seeks for a Mahalanobis distance metric model which minimizes the objective function value of formula (19).

2.3.2. Supervised metric learning method. The supervised methods tried to learn a projection subspace by the pairwise information of samples.

1) Zheng et. al. [10] proposed the Relative Distance Comparison (RDC) method to learn a projection subspace which separate the positive samples from the negative samples. It modelled a Mahalanobis distance function and formulated a distance comparison as follows:

\[
C_f(x^*_i, x^*_j) = \left( 1 + \exp \left[ f(x^*_i) - f(x^*_j) \right] \right)^{-1}
\]  \hspace{1cm} (20)

Where \(f(x^*_i)\) denotes the distance between positive pair and \(f(x^*_j)\) denotes distance between negative pair. The Mahalanobis distance function is defined as follows:

\[
f(x) = x'Mx
\]  \hspace{1cm} (21)

Where \(x\) denotes the difference vector of sample pairs’ feature. Then, an optimization model is proposed to learn the Mahalanobis distance by which the positive sample is closer to the instance than the negative sample. The objective function is as follows:

\[
\min r(W, O) = \sum_{i=1}^{n} \log \left[ 1 + \exp \left( \|W^t x^*_i - W^t x^*_j\|^2 \right) \right] \hspace{1cm} s.t. w^t_i w_j = 0, \forall i \neq j
\]  \hspace{1cm} (22)

2) Elastic Projections-based Metric Learning [31] proposed a pairwise similarity measurement-based on the elastic projection. Differing from the common metric learning method, this model learned two projections, including positive projection and negative projection, to improve the discrimination and robustness of the metric model. The objective function was formulated as follows:

\[
\min_{l_p, l_n} \sum_{i, j} \left[ 1 + \exp \left( \alpha_i \|l_p(x'_i - x'_j)\|^2 - \alpha_n \|l_n(x'_i - x'_j)\|^2 \right) \right] + \frac{1}{2} \|l_p\|^2 + \|l_n\|^2
\]  \hspace{1cm} (23)

Where \(M = L_p^t L_p\), \(L_n^t L_n\). \(L_p\) denotes the projection matrix of positive pair and \(L_n\) denotes the projection matrix of positive pair.

2.4. Mid-level Feature Learning-based Method

Mid-level feature learning-based method tries to learn a dictionary representation which is more robust to the appearance variations, including Semi-Supervised Coupled Dictionary Learning (SSCDL) [32], and Least Square Semi-Coupled Dictionary Learning (LSSCDL) [33] et. al.

The SSCDL method formulated a dictionary learning model to learn a mid-level feature for person re-identification. The two-dictionary representation model was established based on the pairwise relationship of probe sample and gallery sample respectively. Then, an optimization model of coupled dictionary learning was proposed as follows:
Where $D_d$ and $D_g$ are dictionaries of probe set $X$ and gallery set $Y$ respectively. $A_d$ and $A_g$ are corresponding coefficient matrices. The SSCDL method utilized the dictionary learning method to learn a robust mid-level feature. Then, the dictionary representations were used for similarity distance measurement based on Euclidean distance.

The LSDCDL method firstly learned an SVM-based classification boundary for each instance to separate the positive samples from the negative samples. And the SSCDL-based method was used to construct an SVM classification boundary specific to each individual. The objective function is:

$$
\begin{align*}
\min_{\beta, \beta_w, M} & \left\{ x^T \Sigma_{B} x + \lambda_2 \beta^T \Sigma_{W} \beta \right\} \\
\text{s.t.} & \quad \|x^T\|_2^2 \leq 1, \quad \|\beta\|_2^2 \leq 1, \quad \forall i
\end{align*}
$$

(25)

Where $x^*$ is the probe sample. $w^*$ denotes the SVM-based classification boundaries corresponding to the samples. $D_d$ and $D_w$ are dictionaries of $X^*$ and $W^*$ respectively. $M$ is the mapping function which transform the dictionary representation $A_d$ to $A_w$. The coding vector $A_w$ was used to reconstruct the classification boundaries of $X^*$. It is obvious that LSSCDL method learned a model to estimate a classification boundary for each individual.

2.5. Classification Method for Person Re-identification

2.5.1. Fisher Discriminant Analysis. Fisher Discriminant Analysis (FDA) [29] method tried to find a set of hyperplanes in which the projection of the samples from different classes are separated as far as possible and the samples from same classes are separated as close as possible. The between-class scatter and within-class scatter is used to formulate the metric learning model. The between-class scatter was as follows,

$$
S_b = \sum_{i=1}^{C} N_i (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})^T
$$

(26)

Where $N_i$ denotes the sample number of the $i$-th class. $\bar{x}_i$ denotes mean of the $i$-th class. $\bar{x}$ denotes the mean of all samples. The within-class scatter is as follows,

$$
S_w = \sum_{i=1}^{C} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i) (x_{ij} - \bar{x}_i)^T
$$

(27)

Where $x_{ij}$ denotes the $j$-th samples of the $i$-th class. The objective function of FDA is as follows:

$$
\arg\max_w \frac{w^T r(S_b) w}{w^T r(S_w) w}
$$

(28)

According to the properties of Rayleigh entropy function, the solution of the optimization model (28) is the eigenvector of $S_b^{-1} S_w$. The eigenvectors construct a projection subspace $W = (w_1, w_2, ..., w_r)$. The Euclidean distance is used to measure the similarity between images in the projection subspace.
2.5.2. Local Fisher Discriminant Analysis. Moreover, a local fisher discriminant analysis (LFDA) [30] method was proposed by introducing the affinity matrix based on subspace learning. It computed affinity matrix by local invariance. The LFDA method set the affinity matrices of between-class scatter and within-class scatter with different scale factors. The between-class scatter and within-class scatter was rewritten as follows:

\[ S_b = \frac{1}{2} \sum_{i,j=1}^{C} A^b_{ij} (x_i - x_j) (x_i - x_j)^T, \]  
\[ S_a = \frac{1}{2} \sum_{i,j=1}^{C} A^a_{ij} (x_i - x_j) (x_i - x_j)^T, \]  

Where \( A^b_{ij} \) and \( A^a_{ij} \) are affinity matrices of between-class scatter and within-class scatter respectively. The LFDA method improves the FDA method in feature selection, which makes the model to learn more controllable in parameter setting.

2.5.3. Cross view Quadratic Discriminant Analysis. Cross view Quadratic Discriminant Analysis method formulated a classification-based metric learning method by combining the Quadratic Discriminant Analysis (QDA) and KISSME model. The metric model should be discriminative enough to distinguish the difference between individuals and robust enough to against appearance variations. QDA was used to learn a discriminative metric subspace by minimizing the within-class scatter and maximizing between-class scatter. The objective function was as follows:

\[ \max_w J(w) = \frac{w^T \Sigma_{SS} w}{w^T \Sigma_{SS} w} \]  

Where \( \Sigma_{SS} \) denotes the between-class scatter. \( \Sigma_{SS} \) denotes the within-class scatter. The optimization model (31) is transformed to the eigenvalue problem of matrix \( \Sigma_{SS} \). The solution of eigenvectors were ranked in the order of corresponding eigenvalues. The former \( r \) eigenvectors were selected to form the projection matrix \( w = (w_1,w_2,...,w_r) \).

Then, the original data is projected into the low-dimensional projection subspace. And a Mahalanobis distance of KISSME model was learned as follows:

\[ d(x_i^p, x_j^p) = (x_i^p - x_j^p)^T M (x_i^p - x_j^p), \]  

Where \( M = \left( \Sigma_{SS}^{-1} - \Sigma_{SS}^{-1} \right) \cdot \Sigma_{SS}^{-1} = w^T \Sigma_{SS}^{-1} w \) and \( \Sigma_{SS}^{-1} = w^T \Sigma_{SS}^{-1} w \).

3. The Over-fitting Problem in Person Re-identification
Despite great improvement of person re-identification technology in recent years, metric learning-based method suffers from poor generalization due to the over-fitting problem. There are lots of differences between training data and test data due to the drastic variation of poses, viewpoints, backgrounds and illumination.

To prove the over-fitting problem of existing metric learning-based method, quantitative experiments have been conducted. In this paper, 3 typical metric learning models were tested on the VIPeR dataset and the distribution of sample pairs’ distance was analysed as shown in Table 1. The mean and standard deviation of distance of the positive training pairs, the positive test pairs, the negative training pairs, and the negative test pairs under the metric model are displayed. It is worth noting that there is obvious difference between the positive training pairs and positive test pairs. Thus, the over-fitting problem exists in the metric learning methods.
Table 1. Illustration of over-fitting problem. The distance of the positive/negative pairs of training data and test data over 3 different methods.

| Method | Distance distribution of the positive training pairs | Distance distribution of the positive test pairs | Distance distribution of the negative training pairs | Distance distribution of the negative test pairs |
|--------|------------------------------------------------------|-----------------------------------------------|-------------------------------------------------|-----------------------------------------------|
| XQDA   | 13.57±1.70                                          | 161.98±31.37                                  | 722.77±104.51                                   | 299.75±73.25                                  |
| FDA    | 0.033±0.01                                          | 0.26±0.07                                      | 1.18±0.21                                       | 0.42±0.14                                     |
| LMNN   | 2.45±1.67                                           | 1.67±0.86                                      | 12.24±2.6                                       | 4.33±1.98                                     |

4. The Binary Classification and Multi-class Classification-based Metric Learning

As aforementioned above, person re-identification can be formulated as both a multi-classification problem of different persons and binary classification problem of positive against negative pairs. In this paper, the two methods were combined together to solve the identification problem. The optimization model was as follows:

\[
\arg\max_w J(w) = \left(w^T S_w w\right) \left(w^T (S_w + S_c) w\right)
\]

According to the model in formula (33), the proposed method combines the binary classification LDA and multi-classification LDA by minimizing the difference between images of a same person in classification problem of the positive and negative pairs. It improves the discrimination of the metric model. The equivalent form of the formula (33) is as follows:

\[
\arg\max_w J(w) = w^T S_w w \quad \text{s.t.} \quad w^T (S_w + S_c) w = 1
\]

The proposed model can be solved by the solution of the eigenvalue problem as follows:

\[
\lambda w = (S_w + S_c)^{-1} S_w w
\]

Besides, semi-supervised metric learning-based method shows promising performance in improving the generalization of metric model. The semi-supervised learning firstly learned a metric model by existing methods and measured the similarity between image pairs. According to the measurement results, the similar samples were labelled as potential positive pairs. The potential positive pairs were then used to retrain the metric model to improve the generalization on test data.

5. Experimental Results

Quantitative experiments have been conducted on the VIPeR [42] dataset to demonstrate the effectiveness of the improved method. The experimental results are shown as in Table 2 and Figure 1. 316 training data were tested in the experiments. Rank 1, rank5, rank10, and rank 20 of CMC curves are used to evaluate the performance of different methods. The CMC curves of 4 methods were given in Figure 1. The Double Constraint LDA method was proposed in section 4 and the semi-supervised method was also introduced to our model. The experimental results show that the proposed Double Constraint LDA method is more discriminative than both binary LDA-based method and multi-classification method. Meanwhile, the semi-supervised method is effective in solving the over-fitting problem. From the perspective of identification accuracy, the proposed method in this paper achieves the best recognition accuracy in rank-1, rank-5 and rank-10. Especially in the rank-1 recognition accuracy, compared with the best comparison method, our model is 12.12% higher than the best comparison method (MCK-CCA) and 19.23% higher than the baseline method (XQDA).
Table 2. Comparison results on VIPeR dataset

| methods                      | Recognition Accuracy Rate (%) |
|------------------------------|-------------------------------|
|                              | Rank-1 | Rank-5 | Rank-10 | Rank-20 |
| ITML [8]                     | 7.91   | 18.03  | 23.42   | 33.54   |
| RDC [10]                     | 11.71  | 25.32  | 35.44   | 45.57   |
| KISSME [9]                   | 18.67  | 47.15  | 61.71   | 75.63   |
| MVS-LD2L [32]                | 20.79  | 45.08  | 61.24   | 81.36   |
| LMNN [7]                     | 25.63  | 56.96  | 71.2    | 85.44   |
| KCCA [34]                    | 30.16  | 62.69  | 76.04   | 86.8    |
| Improved Deep [16]           | 34.81  | 63.5   | 75      | 80      |
| SCIR [35]                    | 35.76  | -      | -       | -       |
| XQDA [12]                    | 40.09  | 69.15  | 81.49   | 92.09   |
| JLDM [36]                    | 40.32  | -      | 82.31   | 92.15   |
| NFST-LOMO [38]               | 40.73  | 69.94  | 82.34   | 92.37   |
| KEPLER [39]                  | 42.28  | 71.46  | 82.94   | 92.06   |
| LSSCDL [42]                  | 42.66  | -      | 84.27   | 91.93   |
| IPMLLSL [40]                 | 46.5   | 69.3   | 80.7    | 86.5    |
| MCK-CCA [41]                 | 47.2   | -      | 87.3    | 94.7    |
| Multi Classification LDA     | 38.92  | 67.97  | 80.03   | 90.76   |
| Double Constraint LDA        | 41.11  | 70.32  | 82.41   | 92.41   |
| Our Semi-supervised          | 59.32  | 76.11  | 85.27   | 94.93   |

The result of the aforementioned semi-supervised method is also greater than the identification result of SCIR, the deep learning-based model. In summary, the solution to the over-fitting problem possesses important research value in improving the identification accuracy of the metric learning-based method.

Figure 1. CMC curves of different methods on VIPeR dataset.

6. Conclusion

The person re-identification methods of different types were reviewed in this paper, including appearance feature-based methods, deep learning-based methods, and metric learning-based methods. The appearance feature-based method and metric learning-based method corresponds to the two parts
of the person re-identification task, appearance feature extraction and similarity measurement. While deep learning model establishes an end-to-end network, which formulates the two parts into a unified structure. Our taxonomy provides a guide for person re-identification. In recent years, deep learning-based method has made a significant improvement by introducing the body structure detection network. The improvement indeed alleviated the over-fitting problem, which is the main factor that influence the generalization ability of the metric model. The semi-supervised learning approach shows promising performance in improving the generalization ability of metric learning-based method. Besides, quantitative experiments were conducted in this paper to demonstrate the over-fitting problem and effectiveness of the semi-supervised method.

References
[1] Farenzena M, Bazzani L L, Perina A, et al., Person re-identification by symmetry-driven accumulation of local features, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 2360-2367.
[2] Wang X, Doretto G, Sebastian T, et al., “Shape and appearance context modeling,” in: IEEE International Conference on Computer Vision.” 2007, pp. 1-8.
[3] Yang Y, Yang J, Yan J, Liao S, Yi D, and Li S Z, “Salient color names for person re-identification,” IEEE European Conference on Computer Vision, 2014, pp. 536-551.
[4] Liao S, Hu Y, Zhu X, and Li S Z, “Person re-identification by local maximal occurrence representation and metric learning,” IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp.2197-2206.
[5] Layne R, Hospedales T M, Gong S, “Towards person identification and re-identification with attributes.” in: European Conference on Computer Vision, 2012, pp. 402-412.
[6] Shi Z, Hospedales T M, Xiang T, “Transferring a semantic representation for person re-identification and search,” in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.” 2015, pp. 4184–4193.
[7] Weinberger K Q and Saul L K. “Distance Metric Learning for Large Margin Nearest Neighbor Classification.” Journal of Machine Learning Research, 2009, vol.10, no.1, pp.207-244.
[8] Davis J V, Kulis B, Jain P, et al., “Information-theoretic metric learning.” International Conference on Machine Learning, 2007, pp.209-216.
[9] Koestinger M, Hirzer M, Wohlhart P, Roth P M, Bischof H, “Large scale metric learning from equivalence constraints.” Computer Vision and Pattern Recognition, 2012, pp.2288-2295.
[10] Zheng W S, Gong S, Tao X. “Person re-identification by probabilistic relative distance comparison.” IEEE Conference on Computer Vision and Pattern Recognition. 2011, pp. 649-656.
[11] Guillaumin M, Verbeek J, and Schmid C. “Is that you? Metric learning approaches for face identification.” International Conference on Computer Vision, 2009, pp. 498-505.
[12] Liao S, Hu Y, Zhu X, and Li S Z. “Person re-identification by local maximal occurrence representation and metric learning.” IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[13] Prosser B J, et al., “Person re-identification by support vector ranking.” British Machine Vision Conference, 2010, pp. 21.1-11.
[14] Tao D, Guo Y, Song M, Li Y, Yu Z, Tang Y Y. “Person re-identification by dual-regularized kiss metric learning.” IEEE Transation on Image Processing,2016, 25(6), 2726-2738.
[15] Li W, Zhao R, Xiao T, Wang X. “DeepReID: Deep Filter Pairing Neural Network for Person Re-identification.” IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152-159.
[16] Ahmed E, Jones M, “An improved deep learning architecture for person re-identification.” IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp.3908-3916.
[17] Yi D, Lei Z, Liao S, Li S Z, et al. “Deep metric learning for person re-identification.”
International Conference on Pattern Recognition. 2014: 34-39.

[18] Xiao T, Li H, Ouyang W, et al. “Learning deep feature representations with domain guided dropout for person re-identification.” IEEE Conference on Computer Vision and Pattern Recognition. 2016, pp.1249-1258.

[19] He L, Wang Y, Liu W, Zhao H, Sun Z, and Feng J, “Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification,” in IEEE/CVF International Conference on Computer Vision, pp. 8449–8458, 2019.

[20] Karmakar A and Mishra D, "Pose invariant person re-identification using robust pose-transformation gan." in arXiv preprint arXiv:2105.00930, 2021.

[21] Zhao H, Tian M, Sun S, et al. "Spindle net: Person re-identification with human body region guided feature decomposition and fusion." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp.1077-1085.

[22] Zheng Z, Zheng L, Yang Y, “Unlabeled samples generated by GAN improve the person re-identification baseline in vitro.” arXiv preprint arXiv:170107717, 2017.

[23] Wei L H, Zhang S L, Gao W, et al. "Person Transfer GAN to Bridge Domain Gap for Person Re-identification." IEEE Transactions on Image Processing, 2019, vol.30, pp.1-1.

[24] Liu Y, Zhou W, Liu J, et al. An End-to-End Foreground-Aware Network for Person Re-Identification.” IEEE Transactions on Image Processing, 2019, pp.2060-2071.

[25] Sun Y, Xu Q, Li Y, et al. “Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification.” IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp.393-402.

[26] Zheng L, Bie Z, Sun Y F, Wang J D, Su C, Wang S J, et al. “Mars: a video benchmark for large-scale person re-identification.” European Conference on Computer Vision. Cham: Springer, 2016, pp.868–884.

[27] Wold S, Esbensen K, Geladi P, “Principal component analysis.” Chemometrics and Intelligent Laboratory Systems, vol. 2, no.1-3, pp. 37-52, 1987.

[28] He X, Niyogi P, “Locality preserving projections.” Advances in Neural Information Processing Systems. 2016, pp. 153-160.

[29] Fisher R A, “The use of multiple measurements in taxonomic problems”, Ann. Eugen. Vol. 7, no. 2, 1936, pp.179-188.

[30] Pedagadi S, Orwell J, Velastin S, Boghossian B, “Local fisher discriminant analysis for person re-identification,” in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp.3318-3325.

[31] Li X, Liu L and Lu X, "Person Reidentification-based on Elastic Projections," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 4, pp. 1314-1327.

[32] Jing X, Zhu X, Wu F, You X, Liu Q, Yue D, Hu R, and Xu B. "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning." IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 695-704.

[33] Ying Z, Li B, Lu H, et al. “Sample-Specific SVM Learning for Person Re-identification.” IEEE Conference on Computer Vision & Pattern Recognition, 2016, pp.1278-1287.

[34] Lisanti G, Masi I, and Bimbo Del A, “Matching people across camera views using kernel canonical correlation analysis,” in International Conference on Distributed Smart Cameras, pp. 1–6, 2014.

[35] Wang F, Zuo W, Lin L, Zhang D, and Zhang L, “Joint learning of single-image and cross-image representations for person reidentification,”IEEE Conference on Computer Vision and Pattern Recognition, pp. 1288–1296, 2016.

[36] Liu Z, Lu H, et al. Person reidentification by joint local distance metric and feature transformation.[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019.

[37] Liao S and Li S Z, “Efficient psd constrained asymmetric metric learning for person re-identification,"IEEE Conference on Computer Vision and Pattern Recognition, pp. 3685–3693, 2015.
[38] Zhang L, Xiang T, and Gong S, “Learning a discriminative null space for person re-identification,” IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239-1248.

[39] Martinel N, Micheloni C, and Foresti G L, “Kernelized saliency-based person re-identification through multiple metric learning,” IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5645–5658, 2015.

[40] Zhao Z, Zhao B, and Su F, “Person re-identification via integrating patch-based metric learning and local salience learning,” Pattern Recognition, vol. 75, pp. 90–98, 2017.

[41] Lisanti G, Karaman S, and Masi I, “Multichannel-kernel canonical correlation analysis for cross-view person reidentification,” vol. 13, no. 2, p. 13, 2017.

[42] Gray D, Brennan S, Tao H. “Evaluating appearance models for recognition, reacquisition, and tracking.” IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007, pp.1-7.