Ranking Loss: A Novel Metric Learning Method for Person Re-identification

Min Cao[1,3], Chen Chen[1,3](✉), Xiyuan Hu[1,3], and Silong Peng[1,2,3]

[1] Institute of Automation Chinese Academy of Sciences, Beijing, China
{caomin2014,chen.chen}@ia.ac.cn
[2] Beijing Visytem Co., Ltd., Beijing, China
[3] University of Chinese Academy of Sciences, Beijing, China

Abstract. Person re-identification is the problem of matching pedestrians under different camera views. The goal of person re-identification is to make the truly matched pedestrian pair rank as the first place among all pairs, with the direct translation in math language, which equals that the distance of matched pedestrian pair is the minimum value of the distances of all pairs. In this paper, we propose a novel metric learning method for person re-identification to learn such an optimal feature mapping function, which minimizes the difference between the distance of matched pair and the minimum distance of all pairs, namely *Ranking Loss*. Furthermore, we develop an improved version of ranking loss by using p-norm as a smooth approximation of minimum function, with the advantage of manipulating parameter p to control the distance margin between matched pair and unmatched pair to benefit the re-identification accuracy. We also present an efficient solver using only a small portion of pairs in computation, achieving almost the same performance as using all. Compared with other loss function, the proposed ranking loss optimizes the ultimate ranking goal in the most direct and intuitional way, and it directly acts on the whole gallery set efficiently instead of comparatively measuring in small subset. The detailed theoretical discussion and experimental comparisons with other loss functions are provided, illustrating the advantages of the proposed ranking loss. Extensive experiments on two datasets also show the effectiveness of the proposed method compared to state-of-the-art methods.

1 Introduction

Person re-identification (re-id), which addresses the problem of identifying the same person captured from different non-overlapping cameras, is a valuable research subject for building the intelligent video monitoring system. It is also a challenging research subject due to the significant intra-class variations on visual appearance caused by the change in illumination, occlusion and person pose across different views. Nevertheless, person re-id has made great progress in recent years thanks to the continuous efforts of the researchers.

Most existing person re-id studies focus on metric learning in the supervised setting [1–4]. The training set with labelled matching pairs for each pair of

C. V. Jawahar et al. (Eds.): ACCV 2018, LNCS 11362, pp. 381–397, 2019.
https://doi.org/10.1007/978-3-030-20890-5_25

camera views is fully utilized to learn an optimal mapping function from the original feature space to the new feature space, so that positive pair (i.e. a pair of people from different views sharing the same identity) has a smaller distance than negative pair (i.e. a pair of people from different views sharing the different identity) in the new feature space.

Towards this end, the most direct way is to learn a mapping function, with which the distance of positive pair equals the minimum value of distances of all pairs with the given query person. Based on this, we propose a novel loss function for person re-id, minimizing the difference between the distance of positive pair and the minimum value of distances of all pairs. Other loss functions for person re-id are modeled with different format of setting constraints on distances, such as, the distance of positive pair is lower than a given threshold and the distance of negative pair is greater than the threshold [5,6], or the distance of positive pair is lower than the distance of negative pair with a margin [1,2,7]. Although these loss functions ultimately optimize towards the same goal as the proposed loss function, i.e. positive pair has a smaller distance than negative pair, they still achieve the ranking goal in a relatively indirect way, for instance, measuring comparative similarity in a huge number of small subset (e.g. triplet loss or quadruplet loss) or setting a insufficiently necessary hard threshold (e.g. binary classification loss). The proposed loss function works toward the ranking goal directly to push the positive pair rank as the first place among all, and can obtain better performance in person re-id than other loss functions. In addition, since the proposed loss function acts directly on the whole set efficiently, the sample imbalance problem occurred in implementing other loss functions does not exist in the context of the proposed method.

To achieve better model optimization, considering the properties of non-smooth and non-differentiable at some points for the minimum function in the proposed loss function, p-norm, an analytic function, with great differentiable property, is utilized as a smooth approximation of minimum function ($p < 0$). The introduced parameter p is set by theoretic analysis rather than choosing carefully like the other loss functions. And it is worth mentioning that, the relatively larger value for the parameter p (i.e. $p \rightarrow 0$) results in a strong constraint on a larger margin between the distances of positive pairs and negative pairs, that is, further enlarging inter-class variations and reducing the intra-class variations, which is effective for improving the accuracy of model in person re-id. Furthermore, we also develop an efficient solver for the model where only a small part of pairs are used in the computation. It can obtain almost the same performance on the testing set compared to the normal solver by using all negative pairs.

In summary, we make three contributions in this paper as follows:

(1) We propose a novel ranking loss function for metric learning based person re-id, which optimizes the ultimate re-id ranking in the most direct way.

(2) We utilize p-norm to develop a smooth and continuously differentiable version of the proposed ranking loss function, with the advantage of controlling the distance margin by manipulating parameter p to achieve better generalization ability.

(3) We propose an efficient optimization algorithm by using a small portion of pedestrian pairs instead of using all pairs in computation.

2 Related Work

In general, person re-id problem is solved by three crucial steps: feature extraction [8–10], metric learning [1,3,5–7,11] and re-ranking [12–14]. Extensive methods have been proposed focusing on one or more of the steps. For a detailed review of person re-id, the interested readers can refer to [15]. In this section, we only briefly review some representative metric learning based methods that are related to our work.

We can divide metric learning based person re-id methods into two broad categories: closed-form solution based and iterative-learning based. The closed-form solution based methods [9,11,16] are usually related to Linear Discriminative Analysis technology and the optimal solution can be obtained by the generalized eigenvalue decomposition. However, because the number of samples is generally lower than the dimension of samples' feature vector in person re-id, there is usually the singular problem in this kind of methods. In the iterative-learning based methods [1,3,5–7], an objective function is usually constructed and solved by iterative optimization algorithms to satisfy the constraints on samples of the training set. According to the objective function, the iterative-learning methods can be summarized as three main types: the binary classification loss based [5,6], the triplet loss based [1,7] and the quadruplet loss based [2]. In PCCA [5], the objective function is optimized so that the distances of positive pairs are lower than a given threshold and the distances of negative pairs are greater than the threshold. Compared with the binary classification loss with a fixed threshold, a generalized similarity metric for person re-id was proposed with an adaptive threshold [6]. Ding et al. [7] proposed a deep neural network where the relative distances between positive pairs and negative ones are maximized. Zhou et al. [4] presented a novel set to set (S2S) loss layer in deep learning framework that focuses on maximizing the margin between the intra-class set and inter-class set, while preserving the compactness of intra-class samples. In addition to the triplet loss based methods, some methods have been proposed to combine and jointly optimize the binary classification loss and the triplet loss for pursuing better performance [17,18]. Recently, Chen et al. [2] proposed a quadruplet deep network by introducing a quadruplet ranking loss to achieve a smaller intra-class variation and a larger inter-class variation.

3 Ranking Loss Function for Person Re-identification

3.1 Problem Description

Given a query person q from one camera view and a candidate set with N people $\mathcal{G} = \{\mathrm{g}_i \,|i = 1, 2, \ldots, N\}$ from another camera view. The distance between q and g_i is measured by

$$d_L(q, g_i) = \|L(\boldsymbol{x_q} - \boldsymbol{x_{g_i}})\|^2, \tag{1}$$

where $\boldsymbol{x_q}$ and $\boldsymbol{x_{g_i}}$ are the feature vector of the person q and g_i obtained by the feature extraction step. L is a feature transformation matrix.

A ranking list $\mathcal{L}(q, \mathcal{G}) = \{g_1^r, g_2^r, \ldots, g_N^r\}$ can be obtained by the distances between the query person and the candidates in ascending order, i.e. $d(q, \mathrm{g}_1^r) < d(q, \mathrm{g}_2^r) < \ldots < d(q, \mathrm{g}_N^r)$. For person re-id, we hope that the candidate having the same identity with the query person q (also known as the positive sample) is closer to the top of the ranking list. Certainly, it will be perfect in the case where the positive sample ranks first. To do this, we develop a model to learn an optimal feature transformation L by the training set.

3.2 Modeling

Without loss of generality, we introduce our method in this subsection based on the assumption that there is the case of single-shot, i.e one query person shares the same identity with only one gallery person in the candidate.

For each query person q_i (i $= 1, \ldots, M$), in the candidate set $\mathcal{G}$, there are always one person denoted by g^{i+} having the same identity with q_i, and all the rest of people having the different identity with q_i denoted by g_j^{i-} ($j = 1, 2, \ldots, N-1$). With the goal of prioritizing the positive sample g^{i+} on the ranking list, we find an optimal feature transformation L by which the distance between feature vectors of q_i and positive sample g^{i+} is the minimum value of distances between feature vectors of q_i and each sample in the set $\mathcal{G}$. So, the objective function is deduced:

$$\min_L f^*(L) = \sum\nolimits_{i=1}^{M} [d_L(q_i, g^{i+}) - \min_{g_n \in \mathcal{G}} d_L(q_i, g_n)]. \tag{2}$$

For the convenience of expression, we simplify the notation $d_L(q_i, g^{i+})$ as $d_{q_i, g^{i+}}$ and the simplifications of all other $d_L(\cdot, \cdot)$ are similar to this below.

When $\min\limits_{g_n \in \mathcal{G}} d_{q_i, g_n} = d_{q_i, g^{i+}}$ for each term in $f^*(L)$, that is, all positive samples rank first in the corresponding ranking list, the objective function reaches its the lower bound value of zero. However, the minimum function in Eq. 2 is non-smooth and non-differentiable at some points. As a result, it is an intractable problem to solve the model in Eq. 2. For this, we use p-norm as a smooth approximation of the minimum function:

$$(\sum\nolimits_{g_n \in \mathcal{G}} (d_{q_i, g_n})^p)^{\frac{1}{p}} \approx \min_{g_n \in \mathcal{G}} d_{q_i, g_n}. \tag{3}$$

Then, the new objective function is given:

$$\min_L f(L) = \sum\nolimits_{i=1}^{M} [d_{q_i, g^{i+}} - (\sum\nolimits_{g_n \in \mathcal{G}} (d_{q_i, g_n})^p)^{\frac{1}{p}}]. \tag{4}$$

This model in Eq. 4 not only inherits the advantage of the one in Eq. 2, i.e. learning the transformation L by a direct way with request of the positive pair being in the top rank, but also is favorable for solving the optimal transformation L. In addition, the introduced parameter p controls the degree of margin between distances of positive pair and negative pair, resulting in more flexibility for model to obtain a better performance.

Now, we elaborate on why the model in Eq. 4 has these advantages. For this, we reformulate p-norm in Eq. 3 as:

$$\begin{aligned}(\textstyle\sum_{g_n \in \mathcal{G}} (d_{q_i,g_n})^p)^{\frac{1}{p}} &= [(d_{q_i,g^{i+}})^p + (d_{q_i,g_1^{i-}})^p + \ldots + (d_{q_i,g_{N-1}^{i-}})^p]^{\frac{1}{p}} \\ &= [(d_{q_i,g^{i+}})^p (1 + (\frac{d_{q_i,g_1^{i-}}}{d_{q_i,g^{i+}}})^p + \ldots + (\frac{d_{q_i,g_{N-1}^{i-}}}{d_{q_i,g^{i+}}})^p]^{\frac{1}{p}}.\end{aligned} \quad (5)$$

Substituting Eq. 5 into Eq. 4, we can find that the minimum value of the objective function $f(L)$ is reached if and only if $(\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}})^p = 0$ $(\forall t = 1, \ldots, N-1)$ for each person q_i. Since we want to achieve $d_{q_i,g^{i+}} < d_{q_i,g_t^{i-}}$ $(\forall t = 1, \ldots, N-1)$ for each person q_i, meaning $\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}} > 1$ for all terms, it is possible to satisfy $(\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}})^p = 0$ only if $p < 0$. It follows that the model in Eq. 4 with a smooth and continuously differentiable loss function can achieve same functionality as the one in Eq. 2, as long as we choose an appropriate parameter of p.

Based on the basic constraint of $p < 0$, there are two cases where $(\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}})^p = 0$ is satisfied as follows:

(1) when $p \rightarrow -\infty$, the value of $\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}}$ just need to be slightly larger than 1. It means that the margin between distances of positive pairs and negative pairs is very small;

(2) when $p \rightarrow 0$, the value of $\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}}$ need to be much larger than 1. It means that the margin between distances of positive pairs and negative pairs is big.

It is obvious that when we optimize the objective function and search its minimum solution in Eq. 4, the value of p controls the degree of margin between distances of positive pairs and negative pairs. We have the flexibility of choosing p based on the demand for the margin.

We argue that the performance of model on the testing set can be improved by further enlarging the inter-class variations and reducing the intra-class variations [19]. Therefore, it is beneficial to set p as a relative large value (i.e. set $p \rightarrow 0$) in the experiments. However, it is imperative to note here that, with p getting closer and closer to zero, the correspondingly learnt transformation L results in $\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}} \rightarrow \infty$ for satisfying $(\frac{d_{q_i,g_t^{i-}}}{d_{q_i,g^{i+}}})^p = 0$, which means the denominator $d_{q_i,g^{i+}} \rightarrow 0$ and the constraint on the value of the numerator $d_{q_i,g_t^{i-}}$ is weaken in the process of optimization. To avoid two undesirable consequences: (1) the model overfitting, (2) lose the control of the margin between distances of positive pairs and negative pairs, we propose a safer choice to set $p \rightarrow 0$ but not getting closer to the value of zero. In the experiments, we set $p = -5$ and provide a detailed validation for the performance of the proposed method with different value of p, to justify our analysis above.

3.3 Optimization

It is worth noting that we improve the model from Eq. 2 to Eq. 4 with a smooth and continuously differentiable objective function at the cost of increased computational complexity. Specifically, all corresponding negative pairs for each positive pair are need to be traversed for optimizing the model in Eq. 4. On the contrary, only one negative pair for each positive pair involves in the computation for optimizing the model in Eq. 2.

Therefore, in the process of optimizing the model in Eq. 4, we simplify the computation and only the first k sample pairs with smallest distances for each positive pair are used in the computation. We use the gradient descent scheme with line search to solve the model and the simplified objective function at t-th iteration is formulated as:

$$f_t(L) = \sum\nolimits_{i=1}^{M} [d_{q_i,g^{i+}} - (\sum\nolimits_{g_n \in \Omega_t^{i,j,k}} (d_{q_i,g_n})^p)^{\frac{1}{p}}], \tag{6}$$

where $\Omega_t^{i,j,k} \subseteq \mathcal{G}$ includes first k sample pairs with smallest distances obtained by the feature transformation L_{t-1}. If $k = N$, we have $\Omega_t^{i,j,k} = \mathcal{G}$ with which the optimization process is equivalent to the one in Eq. 4; if $k = 1$, the model reverts to the one in Eq. 2.

The corresponding gradient at t-th iteration is derived as follows:

$$\frac{\partial f_t(L)}{\partial L} = \sum\nolimits_{i=1}^{M} [\pi_L(q_i, g^{i+}) - \rho(q_i, g^{i+}) \sum\nolimits_{g_n \in \Omega_t^{i,j,k}} (d_{q_i,g_n})^{p-1} \pi_L(q_i, g_n)], \tag{7}$$

where $\pi_L(\cdot, *)$ is the derivative of distance measurement d_L with respect to L:

$$\pi_L(\cdot, *) = 2L(\boldsymbol{x}_{\cdot} - \boldsymbol{x}_*)(\boldsymbol{x}_{\cdot} - \boldsymbol{x}_*)^T, \tag{8}$$

and $\rho(q_i, g^{i+})$ is a constant:

$$\rho(q_i, g^{i+}) = (\sum\nolimits_{g_n \in \Omega_t^{i,j,k}} (d_{q_i,g_n})^p)^{\frac{1}{p}-1}. \tag{9}$$

With each iteration to approximate the optimal solution, the positive pair (q_i, g^{i+}) is constantly pushed into the first rank with a minimum distance among the distances of all sample pairs (q_i, g_n), where $n \in \mathcal{G}$. It follows that although we discard some sample pairs in the process of optimization, it's still equivalent to solving the model with all sample pairs. In experiments, we set $k = 2$, and the related comparison experiments are carried out and prove that the proposed method with $k = 2$ can achieve almost the same performance and shorter running time compared to the one with $k = N$.

For convenience, we name the proposed method as R-Loss below. The overall optimization process of R-Loss is shown in Algorithm 1.

3.4 Model Extension

In this section, we extend our method to the general case: multi-shot, i.e for a query person, there are more than one positive sample in the candidate set.

We formally express this case. For each query person q_i (i $= 1, \ldots, M$), there are several positive samples denoted by $\mathcal{G}^{i+} = \{g_j^{i+} \in \mathcal{G}\,|j = 1, 2, \ldots, N^{i+}\}$, and a majority of negative samples denoted by $\mathcal{G}^{i-} = \{g_j^{i-} \in \mathcal{G}\,\big|j = 1, 2, \ldots, N^{i-}\}$, where $N^{i+} + N^{i-} = N$.

For the case of single-shot, we hope that the positive sample can rank first in the ranking list for the test set, so the distance of positive pair is the minimum of the distances of all pairs is our objective in the training process. Similarly, for the case of multi-shot, we certainly hope that one of the positive samples can rank first for the test set. For obtain a better performance in the test set, we propose a stronger constraint on the training set that is all positive sample pairs be in the front of the ranking list. For this, we learn the optimal transformation L, so that the distance of each positive pair is the minimum of the distances of this positive pair and all negative pairs with the same query person. Therefore, the objective function is deduced:

$$\min_{L} f_{ex}(L) = \sum\nolimits_{i=1}^{M} \sum\nolimits_{g_j^{i+} \in \mathcal{G}^{i+}} [d_{q_i, g_j^{i+}} - (\sum\nolimits_{g_n \in \{g_j^{i+}\} \cup \mathcal{G}^{i-}} (d_{q_i, g_n})^{p})^{\frac{1}{p}}]. \quad (10)$$

Algorithm 1. The R-Loss gradient descent algorithm

Input:

The query set $\mathcal{Q}$ from one camera and the candidate set $\mathcal{G}$ from another camera, with label information.

Initialize $L_0 = I$ (unit matrix), $\eta = 10^{-4}$ and $t = 1$.

Output:

The optimal transformation matrix L.

```
 1: Compute $f_0(L_0)$ and $\frac{\partial f_0(L)}{\partial L}|_{L=L_0}$ according to Eq.6 and Eq.7, respectively.
 2: do
 3:    Update $L_t = L_{t-1} - \eta \frac{\partial f_{t-1}(L)}{\partial L}|_{L=L_{t-1}}$ .
 4:    Compute $f_t(L_t)$ according to Eq.6.
 5:    if $f_t(L_t) \geq f_{t-1}(L_{t-1})$
 6:       $\eta \leftarrow 0.9\eta$.
 7:       if $\eta < 10^{-20}$
 8:          $L = L_t$.
 9:          break.
10:       end
11:       Return the 3-th step.
12:    else
13:       $\eta \leftarrow 1.1\eta$.
14:    end
15:    if $f_{t-1}(L_{t-1}) - f_t(L_t) < 10^{-5}$
16:       $L = L_t$.
17:       break.
18:    end
19:    Compute $\frac{\partial f_t(L)}{\partial L}|_{L=L_t}$ according to Eq.7.
20:    Update $t \leftarrow t + 1$.
21: end
22: return  The learnt optimal transformation $L$.
```

When $N^{i+} = 1$, the model in Eq. 10 reverts to the one in Eq. 4. We use the gradient descent scheme to solve the model in Eq. 10. The solving process is similar to Algorithm 1 and is not be repeated here.

4 Discussion About Different Loss Functions

In this section, we present a detailed discussion about the differences between the proposed ranking loss function and the existing three classical loss functions: the binary classification loss, the triplet loss and the quadruplet loss. Figure 1 illustrates how the relationships of distances between positive pairs and negative pairs are constrained for these loss functions.

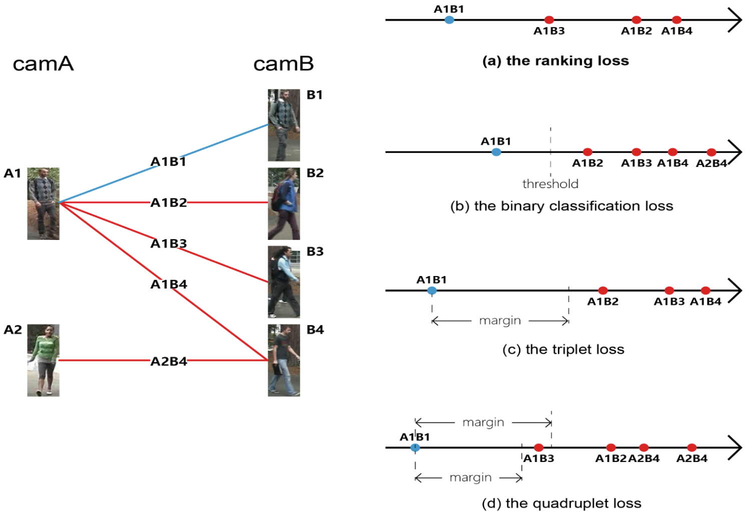

Fig. 1. The relationships between positive pairs' distances and negative pairs' distances for loss functions. The blue line and point represent the positive pair and the red ones represent the negative pair. Better viewed in colour. (Color figure online)

First of all, we formalize the three loss function with the general case of multi-shot as follows:

(1) the binary classification loss function,

$$\min_{L} f^{*}_{binary}(L) = \sum\nolimits_{i=1}^{M} \sum\nolimits_{j=1}^{N} \max(0, y_{ij}(d_{q_i,g_j} - c)), \tag{11}$$

where $y_{ij} = \begin{cases} 1 & g_j \in \mathcal{G}^{i+} \\ -1 & g_j \in \mathcal{G}^{i-} \end{cases}$ is a sign function and c is a margin threshold. Considering the non-smooth and non-differentiable properties of hinge function

$h(x) = \max(0, x)$, there is usually another form of the binary classification loss function used for person re-id,

$$\min_{L} f_{binary}(L) = \sum\nolimits_{i=1}^{M} \sum\nolimits_{j=1}^{N} l_{\beta}(y_{ij}(d_{q_i,g_j} - c)), \tag{12}$$

where $l_{\beta}(x) = \frac{1}{\beta}\log(1 + e^{\beta x})$ is a smooth approximation of hinge function with $\lim\limits_{\beta \to \infty} l_{\beta}(x) = \max(0, x)$.

(2) the triplet loss function,

$$\min_{L} f_{triplet}(L) = \sum\nolimits_{i=1}^{M} \sum\nolimits_{g_j^{i+} \in \mathcal{G}^{i+}} \sum\nolimits_{g_j^{i-} \in \mathcal{G}^{i-}} \max(0, d_{q_i,g_j^{i+}} - d_{q_i,g_j^{i-}} + c), \tag{13}$$

where c is a margin threshold.

(3) the quadruplet loss function,

$$\begin{aligned} \min_{L} f_{quadruplet}(L) = \sum\nolimits_{i=1}^{M} \sum\nolimits_{g_j^{i+} \in \mathcal{G}^{i+}} (\sum\nolimits_{g_j^{i-} \in \mathcal{G}^{i-}} \max(0, d_{q_i,g_j^{i+}} - d_{q_i,g_j^{i-}} + c_1) \\ + \sum\nolimits_{v \neq i, g_n \in \mathcal{G}^{v-}, g_n \notin \mathcal{G}^{i+}} \max(0, d_{q_i,g_j^{i+}} - d_{q_v,g_n} + c_2)), \end{aligned} \tag{14}$$

where c_1 and c_2 are the margin thresholds.

For the binary classification loss function, just the upper bound of positive pairs' distances and the lower bound of negative pairs' distances are constrained by the threshold value of c, and the margin between the positive pair and negative pair is not constrained. Instead, the value of p controls the margin in the proposed loss function and we can set the value of p resulting in a large margin between positive pair and negative pair in the new learnt space, which is profitable for performance in person re-id.

For the triplet loss function, the value of c directly determines the margin between positive pair and negative pair in the new learnt space. A small value of c results in the small margin between positive pair and negative pair and vice versa. However, in generally we don't know what the value of margin should be, so it is difficult to choose a certain value of margin (i.e. set the value of c). Instead, the value of p controls the degree of the margin in the proposed loss function, and we know that a large margin in the new learnt is advantage for performance, and based on this, we can set the value of p.

For the quadruplet loss function, a stricter constraint on margin is established. And similar to the triplet loss function, it is difficult to choose the appropriate values of margin for the quadruplet loss function. Even worse is that there are two parameters c_1 and c_2 which are need to be set.

Our proposed ranking loss function, as well as the three loss functions, all aim to drive the distance of positive pair to the minimum of the distances of all pairs for each query person. But for different loss functions, there are different ways to reach the goal. And it is the most direct way for the proposed ranking loss function, which is more accordant with person re-id problem. In addition, from Eqs. 11–14 we can see that all negative pairs involve in the computation for each positive pair, resulting in the high demand of computing. Therefore, researchers proposed to solve the model with one positive pair vs. a fraction of negative pairs for the binary classification loss function, one positive pair vs. a hardest negative pair (i.e. a negative pair with the smallest distance among all negative pairs) for the triplet loss function and quadruplet loss function. Even so, there is the problem of sample imbalance in the binary classification loss function, and the improved models for the triplet loss function and quadruplet loss function are intractable to solve due to the operation of finding the hardest negative pair. Instead, there does not exist the problem of sample imbalance in the proposed ranking loss function since we just impose constraints on the rank of positive pair, meanwhile, only two pairs for each positive pair are involved in the computation for our model, with almost the same performance and higher computational efficiency compared to using all pairs.

5 Experiments

In this section, we conduct experiments from three aspects: (1) evaluate the performance of the proposed method with different settings; (2) compare the proposed method with other loss function based methods; (3) compare the proposed method with state-of-the-art methods. Before starting, we make a detailed introduction for the settings relevant to the experiments.

5.1 Experimental Settings

Datasets. Two publicly available datasets are used in experiments: VIPeR [20] and CUHK01 [21]. The **VIPeR** dataset includes 632 pedestrian image pairs captured by two different cameras in an outdoor environment with only one image per person for each view. The **CUHK01** dataset has 971 identities taken from two camera views in a campus environment with 2 images of every person under each camera view.

Feature. There are three feature descriptors used in the experiments: Local Maximal Occurrence (LOMO) [9], Gaussian Of Gaussian (GOG) [22] and Histogram of Intensity Pattern & Histogram of Ordinal Pattern (HIPHOP) [23]. For the LOMO descriptor, the horizontal occurrence of local features described by the Scale Invariant Local Ternary Pattern (SILTP) and HSV histogram are maximized to make a stable representation against viewpoint changes. For the GOG descriptor, the person image is described by cascaded Gaussian distributions in which both means and covariances are considered. For HIPHOP descriptor, it describes the person image based on AlexNet [24] convolution neural network.

Evaluation Protocol. We evaluate the performance of person re-id methods by Cumulative Matching Characteristics (CMC) where Rank r represents the expectation of correct match at r-th in the ranking list. The reported results are the average results of 10 times of the partition of training set and test set. For each partition, similar to most publications, the identities are randomly divide into two equal parts used for training set and test set in VIPeR dataset respectively; 485 identities and 486 ones are used for training and testing with single-shot (M1) and multi-shot (M2) setting in CUHK01 dataset.

5.2 Analysis of the Proposed Method

p-norm vs. min. We aim to learn an optimal feature transformation L by which the distance of positive pair is the minimum value of distances of all sample pairs with the same query person in the new feature space. The p-norm is used as a smooth approximation of the minimum function in Eq. 3. The superiority of p-norm compared with the minimum function has been analyzed theoretically in Sect. 3.2. In this subsection, we validate the derived theoretical results by the experiment with GOG feature descriptor employed. The CMC curves of R-Loss(p-norm) and R-Loss(min) on VIPeR and CUHK01 datasets are reported in Fig. 2. We can observe that, R-Loss(p-norm) improves the performance of R-Loss(min) by a large margin on all datasets. By using p-norm as a smooth approximation of minimum function, the Rank 1 increases 24.55%, 17.53% and 20.02% on VIPeR, CUHK01(M1) and CUHK01(M2) datasets, respectively. It reveals that the optimization algorithm by using the gradient descent scheme in the R-Loss(p-norm) model can converge to a better solution than in the R-Loss(min) model.

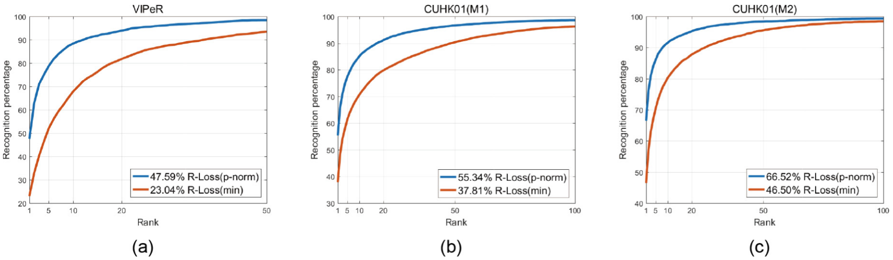

Fig. 2. Comparison of CMC curves for the proposed method by using p-norm function and minimum function.

On the Parameter p. In Sect. 3.2, we present a detailed analysis about the selection of parameter p and draw a conclusion that a larger value of p but not getting closer to the value of zero based on a basic constraint $p < 0$ is beneficial to the accuracy of person re-id. In this subsection, we compare the performance of the proposed R-Loss method with different values of p on VIPeR and CUHK01 datasets, and LOMO descriptor and GOG descriptor are used as the feature representation of person, respectively. As shown in Fig. 3, when $p \rightarrow -\infty$, the accuracy of Rank1 gradually degrades on all datasets. More specifically, when $p < -10$, the accuracy rapidly degrades on all datasets. It indicates that a lower value of p with $p \rightarrow -\infty$ is disadvantage to the performance. Besides, we can also see that the accuracy of Rank1 degrades rapidly with $p = -1$ on CUHK01(M1) and CUHK01(M2) datasets. Although the proposed method with $p = -1$ has well performance on VIPeR dataset, setting the value of p close to the value of zero is a riskier choice. Therefore, we set $p = -5$ in the experiments.

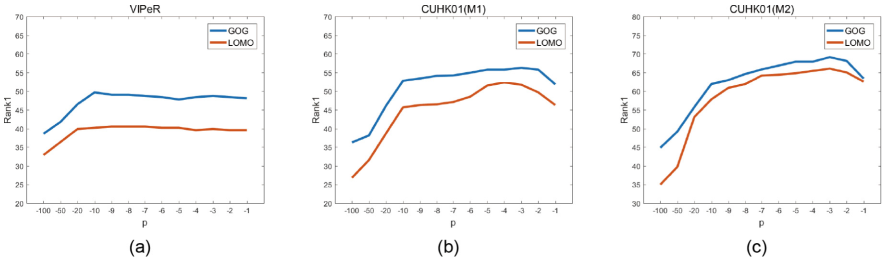

Fig. 3. The performances of Rank1 for the proposed method with different value of p.

On the Parameter k. In the process of optimizing the model in Eq. 4, we introduce a simplified algorithm to find the optimal solution. At each iteration of optimization algorithm, the first k sample pairs with smallest distances for each positive pair are used in the computation. In this subsection, we investigate the effects of k on the performance and running time of the proposed R-Loss method. Here, we use the LOMO and GOG feature descriptors for representing the person image, respectively, and the experiments are carried out on VIPeR and CUHK01 datasets. We show the results in Table 2 by varying k from 2 to N. We can see that the proposed method is kind of invariant to k on performance (fluctuate around 2%), and the running time gradually decreases with the decrease of k. When $k = N$, all samples are used for optimizing model so that the optimization process become much more complex, the performance might slightly be affected. As a result, $k = 2$ is an optimal choice.

5.3 Comparison with Other Loss Function Based Methods

In this paper, we propose a novel metric learning based person re-id method by optimizing a ranking loss function. Compared with the binary classification loss based [5,6], the triplet loss based [1,7] and the quadruplet loss based methods [2], there is no need to select the parameters carefully and no problem of sample imbalance in the proposed ranking loss based method. In the following, we will verify a more critical advantage of the proposed method compared with classical loss function based methods: a better generalization ability on the testing set (Table 1).

For a fair comparison, GOG feature descriptor is used in this comparison experiment. Similar to most of methods, we set $c = 1$ for both the binary classification loss function and the triplet loss function, and $c_1 = 1$, $c_2 = 0.5$ for the triplet loss function. The comparison results are shown in Table 3 the proposed

Table 1. Effects of k on the performance and running time of the proposed R-Loss method. Running time refers to one iteration time of optimization algorithm. $k = N$ represents the case for using all samples in optimization.

k	VIPeR				CUHK01(M1)				CUHK01(M2)			
	GOG		LOMO		GOG		LOMO		GOG		LOMO	
	r = 1	T(s)	r = 1	T(s)	r = 1	$T(s)$	r = 1	$T(s)$	r = 1	$T(s)$	r = 1	$T(s)$
2	47.59	0.39	39.49	0.36	55.34	9.07	51.55	9.35	67.90	9.34	64.81	9.29
5	47.82	0.77	39.49	0.37	55.98	9.28	51.86	9.44	69.14	9.50	65.84	9.49
10	48.82	0.81	39.75	0.38	56.19	9.57	51.65	9.58	68.52	9.64	66.26	9.65
N	46.61	1.48	38.32	1.42	54.33	32.84	50.41	32.70	65.84	32.39	64.20	32.63

Table 2. Comparison with methods based on other loss function on VIPeR and CUHK01 datasets. The best results (%) are respectively shown in red. Better viewed in colour.

Methods	VIPeR				CUHK01(M1)				CUHK01(M2)			
	r = 1	r = 5	r = 10	r = 20	r = 1	r = 5	r = 10	r = 20	r = 1	r = 5	r = 10	r = 20
Binary	46.23	77.34	87.78	93.64	32.33	51.32	60.91	71.71	37.02	59.63	69.84	80.12
Binary (smooth)	46.46	78.42	88.42	94.43	51.44	76.14	84.16	90.63	62.61	84.73	90.16	94.67
Triplet	30.09	62.41	76.93	87.09	45.54	69.25	78.05	85.93	54.69	78.07	86.13	91.91
Quadruplet	33.01	66.11	77.91	88.54	36.00	59.14	68.75	77.91	44.14	69.03	78.48	86.26
R-Loss	47.59	78.86	88.42	93.96	55.34	77.51	85.14	91.16	66.52	86.13	91.58	95.29

R-Loss method outperforms other loss function based methods on all datasets. Specifically, our method achieves about 3.9% gain at Rank 1 compared with the second result from the binary classification loss (smooth) based method on CUHK01 dataset with single-shot and multi-shot setting.

5.4 Comparison with State-of-the-Art Methods

In this section, we evaluate our method against the state-of-the-art person re-id methods. The experiment results of our method are presented by a simple fusion of scores obtained by the three types of features mentioned above.

Result on VIPeR Dataset. We compare the proposed method with 13 existing person re-id methods. From the results shown in Table 4, we can see that our method achieves the highest performance at Rank1, Rank5 and Rank20. More specifically, our method beats the closest competitor EBG [25] by 1.11% at Rank1. EBG [25] as a feature extraction based method where a deep neural network was proposed to solve the background bias problem, still yield the poorer results compared with our proposed method based on classical metric learning technology. It indicates that the deep learning based methods cannot currently work well on the small person re-id dataset.

Result on CUHK01 Dataset. The proposed R-Loss method is compared with traditional methods and deep learning based methods in CUHK01 dataset with

Table 3. Comparison with the state-of-the-art methods on VIPeR dataset. The best results for deep learning based and traditional methods (%) are respectively shown in boldface and red. Better viewed in colour.

	Method	Reference	r = 1	r = 5	r = 10	r = 20
Deep learning	IDLA	2015CVPR [26]	34.81	63.60	75.63	84.49
	Deep Ranking	2016TIP [27]	38.40	69.20	81.30	90.40
	TCP	2016CVPR [28]	47.80	74.70	84.80	91.10
	PDC	2017ICCV [29]	51.27	74.05	84.18	91.46
	DeepAlign	2017ICCV [30]	48.70	74.70	85.10	**93.00**
	EBG	2018CVPR [25]	**51.90**	74.40	84.80	90.20
	MLS	2018CVPR [31]	50.10	73.10	84.35	–
	MC-PPMN	2018AAAI [32]	50.13	**81.17**	**91.46**	–
Traditional	WARCA	2016ECCV [33]	40.22	68.16	80.70	91.14
	TMA	2016ECCV [34]	48.19	–	87.65	93.54
	PatchM&LocalM	2017PR [1]	46.50	69.30	80.70	–
	MVLDML+	2018TIP [35]	50.00	79.20	88.50	94.70
	GCT	2018AAAI [36]	49.40	77.60	87.20	94.00
	R-Loss	Our	53.01	83.07	90.82	96.27

Table 4. Comparison with the state-of-the-art methods on CUHK01 dataset. The best results for deep learning based and traditional methods (%) are respectively shown in boldface and red. Better viewed in colour.

	Method	Reference	r = 1	r = 5	r = 10	r = 20
Deep learning	IDLA	2015CVPR [26]	47.50	71.60	80.30	87.50
	TCP	2016CVPR [28]	53.70	84.30	91.00	96.30
	Deep Ranking	2016TIP [27]	50.40	70.00	84.80	92.00
	DeepAlign	2017ICCV [30]	75.00	93.50	95.70	**97.70**
	MC-PPMN	2018AAAI [32]	**78.95**	**94.67**	**97.64**	–
Traditional	WARCA	2016ECCV [33]	65.64	85.34	90.48	95.04
	PatchM&LocalM	2017PR [1]	53.50	82.50	91.20	96.10
	GCT	2018AAAI [36]	61.90	81.90	87.60	92.80
	MVLDML+	2018TIP [35]	61.40	82.70	88.90	93.90
	R-Loss	Our	73.66	90.64	94.14	96.87

multi-shot setting. Table 4 shows that the best results are from the deep learning based method MC-PPMN [32] where deep feature representations for semantic-components and color-texture distributions are learned based on pyramid person matching network. However, the proposed method is competitive among the traditional methods at all Rank.

6 Conclusions

In this paper, we propose a novel ranking loss function from a new perspective to solve the person re-id problem. The loss function is optimized aiming to make the distance of positive pair be the minimum value of distances of all pairs with the given query person, which is more accordant with person re-id problem. Moreover, we propose to use p-norm to approximate the minimum function to obtain a smooth and continuously differentiable loss function, which is favorable for model solution. In addition, just the first two sample pairs with smallest distances are used in the process of model optimization for improving the efficiency of model solution, with almost the same accuracies as the model using all pairs. Extensive experiments with thorough analysis demonstrate that the proposed R-Loss method achieves superior performance than state-of-the-art methods on VIPeR and CUHK01 datasets. In future research, we will explore to apply the ranking loss to deep learning scheme so that a competitive performance can be achieved on the large datasets.

Acknowledgment. This work is supported by the National Key R&D Program of China under Grant 2017YFC0803505. We appreciate Hao Dou's help on this paper.

References

1. Zhao, Z., Zhao, B., Su, F.: Person re-identification via integrating patch-based metric learning and local salience learning. Pattern Recogn. **75**, 90–98 (2017)
2. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: Computer Vision and Pattern Recognition, vol. 2 (2017)
3. Sun, C., Wang, D., Lu, H.: Person re-identification via distance metric learning with latent variables. IEEE Trans. Image Process. **26**(1), 23–34 (2016)
4. Zhou, S., Wang, J., Shi, R., Hou, Q., Gong, Y., Zheng, N.: Large margin learning in set to set similarity comparison for person re-identification. IEEE Trans. Multimed. **PP**(99), 1–1 (2017)
5. Jurie, F., Mignon, A.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: Computer Vision and Pattern Recognition, pp. 2666–2672 (2012)
6. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3610–3617 (2013)
7. Ding, S., Lin, L., Wang, G., Chao, H.: Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn. **48**(10), 2993–3003 (2015)
8. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Computer Vision and Pattern Recognition, pp. 2360–2367 (2010)
9. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)

10. Su, C., Li, J., Zhang, S., Xing, J., Gao, W., Tian, Q.: Pose-driven deep convolutional model for person re-identification. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 3980–3989. IEEE (2017)
11. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: Computer Vision and Pattern Recognition, pp. 3318–3325 (2013)
12. Cao, M., Chen, C., Hu, X., Peng, S.: From groups to co-traveler sets: pair matching based person re-identification framework. In: IEEE International Conference on Computer Vision Workshop, pp. 2573–2582 (2017)
13. Chen, C., Cao, M., Hu, X., Peng, S.: Key person aided re-identification in partially ordered pedestrian set. In: Conference the British Machine Vision Conference (2017)
14. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3652–3661. IEEE (2017)
15. Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., Radke, R.J.: A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. IEEE Trans. Pattern Anal. Mach. Intell., 1 (2016)
16. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: Computer Vision and Pattern Recognition, pp. 1239–1248 (2016)
17. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: Joint learning of single-image and cross-image representations for person re-identification. In: Computer Vision and Pattern Recognition, pp. 1288–1296 (2016)
18. Zhou, S., Wang, J., Wang, J., Gong, Y., Zheng, N.: Point to set similarity based deep feature learning for person re-identification. In: Computer Vision and Pattern Recognition, pp. 5028–5037 (2017)
19. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
20. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS) (2007)
21. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 31–44. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37331-2_3
22. Matsukawa, T., Okabe, T., Suzuki, E., Sato, Y.: Hierarchical Gaussian descriptor for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1363–1372 (2016)
23. Chen, Y.-C., Zhu, X., Zheng, W.-S., Lai, J.-H.: Person re-identification by camera correlation aware feature augmentation. IEEE Trans. Pattern Anal. Mach. Intell. **40**(2), 392–408 (2018)
24. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1097–1105 (2012)
25. Tian, M., et al.: Eliminating background-bias for robust person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018

26. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: Computer Vision and Pattern Recognition, pp. 3908–3916 (2015)
27. Chen, S.Z., Guo, C.C., Lai, J.: Deep ranking for person re-identification via joint representation learning. IEEE Trans. Image Process. **25**(5), 2353–2367 (2016)
28. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Computer Vision and Pattern Recognition, pp. 1335–1344 (2016)
29. Su, C., Li, J., Zhang, S., Xing, J., Gao, W., Tian, Q.: Pose-driven deep convolutional model for person re-identification. In: The IEEE International Conference on Computer Vision (ICCV), October 2017
30. Zhao, L., Li, X., Zhuang, Y., Wang, J.: Deeply-learned part-aligned representations for person re-identification. In: The IEEE International Conference on Computer Vision (ICCV), October 2017
31. Guo, Y., Cheung, N.-M.: Efficient and deep person re-identification using multi-level similarity. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
32. Mao, C., Li, Y., Zhang, Y., Zhang, Z., Li, X.: Multi-channel pyramid person matching network for person re-identification (2018)
33. Jose, C., Fleuret, F.: Scalable metric learning via weighted approximate rank component analysis. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9909, pp. 875–890. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46454-1_53
34. Martinel, N., Das, A., Micheloni, C., Roy-Chowdhury, A.K.: Temporal model adaptation for person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 858–877. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_52
35. Yang, X., Wang, M., Tao, D.: Person re-identification with metric learning using privileged information. IEEE Trans. Image Process. **PP**(99), 1 (2018)
36. Zhou, Q., et al.: Graph correspondence transfer for person re-identification (2018)