Isosceles Constraints for Person Re-Identification

Furong Xu, Bingpeng Ma, *Member, IEEE*, Hong Chang, *Member, IEEE*, and Shiguang Shan, *Senior Member, IEEE*

***Abstract*— In the existing works of person re-identification (ReID), batch hard triplet loss has achieved great success. However, it only cares about the hardest samples within the batch. For any probe, there are massive mismatched samples (crucial samples) outside the batch which are closer than the matched samples. To reduce the disruptive influence of crucial samples, we propose a novel isosceles contraint for triplet. Theoretically, we show that if a matched pair has equal distance to any one of mismatched sample, the matched pair should be infinitely close. Motivated by this, the isosceles constraint is designed for the two mismatched pairs of each triplet, to restrict some matched pairs with equal distance to different mismatched samples. Meanwhile, to ensure that the distance of mismatched pairs are larger than the matched pairs, margin constraints are necessary. Minimizing the isosceles and margin constraints with respect to the feature extraction network makes the matched pairs closer and the mismatched pairs farther away than the matched ones. By this way, crucial samples are effectively reduced and the performance on ReID is improved greatly. Likewise, our isosceles contraint can be applied to quadruplet as well. Comprehensive experimental evaluations on Market-1501, DukeMTMC-reID and CUHK03 datasets demonstrate the advantages of our isosceles constraint over the related state-of-the-art approaches.**

***Index Terms*— Person re-identification, isosceles constraint, triplet, quadruplet.**

I. Introduction

PERSON re-identification (ReID) [1] aims at retrieving designated individuals (probe) from a large amount of pedestrian images (gallery) captured by non-overlapping cameras. It is an important task to many surveillance applications such as multi-target tracking [2] and person association [3]. However, the appearance diversity of the same pedestrian and the similarity between different pedestrians make person ReID a very challenging problem. In real scenarios, the same person may experience significant appearance changes due to variations in backgrounds, occlusions, viewpoints, postures, illuminations, and so on.

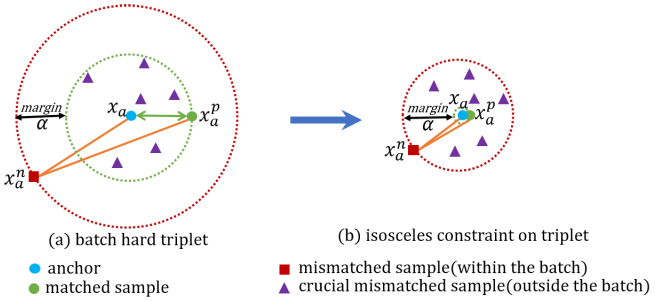


Fig. 1. Our isosceles constraint is based on batch hard triplets [4]. The dotted circle denotes the potential class boundary. The triangle represents a sample outside the batch that does not match x_a. (a) there are some crucial samples under batch hard triplet loss. (b) shows the effect after applying an isosceles constraint, matched pairs are pulled close enough, so that any mismatched samples outside the batch can be farther than the matched sample.

Recently, the triplet loss with batch hard sampling scheme [4] (batch hard triplet loss) achieves great success on ReID, which attracts extensive attention of researchers. For each anchor point x_a, the batch hard triplet loss constrains the relative distances between x_a, the farthest matched sample x_a^p and the closest mismatched sample x_a^n within a batch by a margin, as shown in Figure 1 (a). However, the batch hard triplet loss only cares about the hardest samples within the batch, and does not guarantee that the matched sample pairs (e.g., x_a and x_a^p) are close enough to each other. In fact, there are massive indeterminate mismatched samples outside the batch, some of which (denoted as purple triangles) are even closer to x_a than the matched sample x_a^p. The same situation will also appear in the model trained by quadruplet loss. These disruptive mismatched samples lead to inaccurate distance ranking, thus have a negative impact on the performance of ReID. Consequently, we call such mismatched samples as *crucial samples*.

To relieve the negative influence of crucial samples, some works [5], [6] attempt to further constrain the distance of the matched pairs or the mismatched pairs respectively with pre-defined fixed values. However, considering the large within-class and between-class variations, it is inappropriate to constrain all matched or mismatched pairs to have fixed absolute distances. Moreover, using the same fixed-value distance for all matched or mismatched images can be quite

Manuscript received August 14, 2019; revised February 20, 2020 and June 18, 2020; accepted July 8, 2020. Date of publication September 11, 2020; date of current version September 21, 2020. This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61732004, Grant 61876171, and Grant 61976203, and in part by the Fundamental Research Funds for the Central Universities. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giulia Boato. *(Corresponding author: Bingpeng Ma.)*

Furong Xu and Bingpeng Ma are with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xufurong17@mails.ucas.ac.cn; bpma@ucas.ac.cn).

Hong Chang is with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: changhong@ict.ac.cn).

Shiguang Shan is with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing 100190, China, also with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai 200031, China (e-mail: sgshan@ict.ac.cn).

Digital Object Identifier 10.1109/TIP.2020.3020648

restrictive, discouraging any distortions in the feature space. As a result, the number of crucial samples can not be effectively reduced.

Theoretically, pulling the matched samples towards x_a or pushing the mismatched samples out of the batch away from x_a during training can effectively relieve such crucial sample problem. However, since the number of mismatched samples outside the batch is huge, it is hard to guarantee that all mismatched samples are far from x_a. Therefore, it is more feasible to reduce crucial samples by shortening the distances between the matched pairs. It is worth noting that if the matched samples are close enough (Ideally, the matched samples should be centralized at one point), they should have equal distance to each mismatched sample. In this work, we further show that this is true the other way round, i.e., if the matched pairs have equal distance to multiple (at least two) mismatched samples, the matched pairs will get closer.

Inspired by this, in this paper, we propose a novel isosceles constraint to reduce the number of crucial samples. More specifically, we design an isosceles constraint for the two mismatched pairs in the each triplet. *e.g*, (x_a, x_a^n) and (x_a^p, x_a^n). The isosceles constraint together with margin constraints in the triplet constitute the Isosceles Constrained Triplet (ICT) loss. Minimizing the ICT loss function with respect to the feature extraction network makes the matched pairs closer and the mismatched pairs farther away than the matched ones. As illustrated in Figure 1 (b), after training, x_a^p is very close to x_a, and the distances between x_a and the crucial samples are larger than the distance between the matched pair. Therefore, the crucial samples can be correctly classified as mismatched samples. Similarly, the isosceles constraint can be employed directly on the mismatched pairs of quadruplets, namely Isosceles Constrained Quadruplet (ICQ) loss.

In terms of person ReID performance, extensive experiments on Market-1501, DukeMTMC-reID and CUHK03 benchmark datasets confirm the effectiveness of our isosceles constraint. Especially, compared with the batch hard triplet loss, our ICT loss gains +6.1% (79.3%-73.2%) in mAP and +5.5% (91.6%-86.1%) in Rank-1 on Market-1501. Additionally, when isosceles constrained triplet loss is combined with cross-entropy loss, we can obtain 94.2 % in Rank-1 on Market-1501.

II. Related Work

Extensive works have been reported to address the person ReID problem. These methods mainly focus on several different aspects of the issue such as designing discriminative metrics and developing robust feature descriptors. We briefly review several related works in this section.

A. Metric Learning Method

The metric learning methods [7]–[14] target on measuring the similarity between different images by designing effective distance metric. Some recent works employ triplet loss to solve the person ReID task. Ding *et al.* [15] use the standard triplet with a generation scheme to person ReID. Cheng *et al.* [5] propose an improved triplet loss which adds an additional positive-pair constraint to the original triplet loss. In [6], a novel method based on P2S similarity comparison is presented. However, the improvement in [5], [6] uses fixed values to constrain the distance of the matched or mismatched pairs respectively, which will lead to overlearning of easy samples and underlearning of hard samples. In order to select useful samples to train the network, Hermans *et al.* [4] introduce a novel the triplet loss (Batch Hard) which renders mining of hard triplets unnecessary. HAP2S [16] report a soft hard-mining scheme to make more samples involved in the gradient update. To improve the generalization performance of the model, Chen *et al.* [17] propose a quadruplet loss, which further forces the intra-class distance less than the inter-class distance between two other classes based on triplet loss. In addition, MSML [18] conducts a hard mining scheme to the quadruplets. Without loss of generality, our isosceles constraint can be applied to triplet or quadruplet to pull the matched pairs closer and push the mismatched pairs away, thus the accuracy of person ReID can be improved.

B. Feature Designing Method

The feature designing methods [19]–[28] mainly focus on extracting robust descriptor to handle the appearance variations, which mainly includes the traditional feature descriptors and deep learning features. The traditional feature extraction methods improve the performance of ReID by manually designing robust features. Ma *et al.* [29] present that the person image via covariance descriptor is robust to illumination and background variations. In [30], Liao *et al.* construct a feature descriptor which analyze the horizontal occurrence of local features and maximize the occurrence to make a stable representation against viewpoint changes. DeepReID [24] designs a filter pairing neural network (FPNN) to jointly handle multiple problems, such as misalignment, photometric transforms, occlusions, and so on. In [31], Ahmed *et al.* propose an improved deep learning framework to learn feature embedding. In [10], Zhou *et al.* propose an adaptive margin method to learn the deep features for person ReID in a siamese framework. In recent works, Sun *et al.* [32] propose a part-based network to lay emphasis on the content consistency within each part, Yao *et al.* [33] propose part loss for deep representation learning, Zheng *et al.* [34] expect to learn pose-invariant embedding, Li *et al.* [35] introduce attention mechanisms to improve the discriminability of features, Song *et al.* [36] design a mask-guided contrastive attention model to learn features separately from the body and background regions. Some works [37], [38] learn better feature by generating more data. Because of the advantages of deep convolutional neural network, we also adopt deep learning based method to extract features for person ReID.

III. Isosceles Constraint for Triplet

In this section, we will introduce our isosceles constraint applied on triplet for person ReID problem. First, we review some related triplet-based methods. Then, we present our method with detailed formal formulations.

For clear description, we first give some definitions. Given a training mini-batch X of N samples with labels, we select triplets $\{x_a, x_a^p, x_a^n\}$ within the batch so that their labels $\{y_a, y_a^p, y_a^n\}$ satisfy $y_a^p = y_a$ and $y_a^n \neq y_a$. Meanwhile, x_a^p is the farthest matched sample from x_a, and x_a^n is the closest mismatched sample from x_a. The corresponding features are $f_\theta(x_a)$, $f_\theta(x_a^p)$ and $f_\theta(x_a^n)$, where f denotes the feature extraction network, and θ is the network parameter. To simplify the notation, we use f_a to replace $f_\theta(x_a)$, and so forth. For the distance between two samples $\{x_i, x_j\}$, we compute the Euclidean distance as follows

$$d(x_i, x_j) = \sqrt{||f_i - f_j||_2}, \tag{1}$$

where $||\cdot||_2$ is the ℓ_2-norm.

For better understanding, we give a more concrete definition of crucial sample. Suppose $\mathcal{B}$ denotes a mini-batch. During training, the set of crucial samples of each anchor $x_a \in \mathcal{B}$ is

$$\mathcal{C}_\mathcal{B}(x_a) = \{x | \forall x \notin \mathcal{B}, d(x, x_a) < d(x_a, x_a^p), y \neq y_a\}.$$

where

$$x_a^p = argmax_{x_z \in \mathcal{B}: y_z = y_a} d(x_a, x_z)$$

During testing, the set of crucial samples of each query x_q is

$$\mathcal{C}(x_q) = \{x | x_q^p = argmax_{x_z : y_z = y_q} d(x_q, x_z), \\ d(x, x_q) < d(x_q, x_q^p), y \neq y_q\}.$$

So the improvement of the performance during testing can reflect reduction in the number of crucial samples.

A. Revisit Triplet-Based Loss

Batch hard triplet loss [4] is one of the most representative metric losses, which achieves great success on ReID. It aims to force the distance between the farthest matched sample x_a^p and anchor x_a less than the distance between the closest mismatched sample x_a^n and x_a by at least a margin α. The batch hard triplet loss can be formalized as

$$\mathcal{L}_{BHT} = \frac{1}{N} \sum_{a=1}^{N} [d(x_a, x_a^p) - d(x_a, x_a^n) + \alpha]^+ \tag{2}$$

where $[.]^+ = \max(., 0)$, α is a pre-defined value, and N represents the number of triplets, which is equal to the batch size. Each image in a mini-batch only has a chance to be an anchor, then the farthest matched sample and the closest mismatched sample within the batch are selected to constitute a triplet.

B. Isosceles Constrained Triplet Loss

Since batch hard triplet loss only considers the hardest samples within a batch, there will be a lot of mismatched samples outside the batch that do not meet the margin constraint, such as the samples represented by the purple triangles shown in Figure 1 (a). In the process of training, some mismatched samples outside the batch are closer to x_a than the matched samples. These disruptive mismatched samples (crucial samples) seriously affect the performance of ReID. Therefore, reducing the number of crucial samples can effectively improve the accuracy.

Theoretically, the number of crucial samples can be reduced by pulling the matched samples closer or pushing all the crucial samples farther away. However, only the samples within the current batch can be obtained during the training, so it is extremely difficult to know which are the crucial samples. Therefore, a feasible way to reduce the number of crucial samples is to shorten the distance between matched samples.

1) The Isosceles Constraint Term: The isosceles constraint aims to make matched samples as close as possible. When the matched samples are close enough, especially when the matched samples are centralized to a point, these matched samples should have equal distance to any one of the mismatched samples. Motivated by this, we design an isosceles constraint for the two mismatched pairs in each triplet. In a mini-batch, the batch hard triplet sampling method allows each person to have at least one matched pair and two disparate mismatched samples with high probability. So when we apply isosceles constraint to each of the hardest triplets within the batch, the matched pairs will get closer.

In order to accomplish the isosceles constrain, we can directly limit the distance between two mismatched pairs to be equal, or restrain the ratio of the two. Specifically, we conceive three alternative constraints, equi-ratio constraint ($\mathcal{L}_{ICT_F}$ and $\mathcal{L}_{ICT_R}$) and equidistance constraint ($\mathcal{L}_{ICT_D}$). They are respectively defined as follows

$$\mathcal{L}_{ICT_F} = \frac{1}{N} \sum_{a=1}^{N} |1 - \frac{1}{2}(\frac{d(x_a, x_a^n)}{d(x_a^p, x_a^n)} + \frac{d(x_a^p, x_a^n)}{d(x_a, x_a^n)})| \tag{3}$$

$$\mathcal{L}_{ICT_R} = \frac{1}{N} \sum_{a=1}^{N} |\frac{d(x_a, x_a^n)}{d(x_a^p, x_a^n)} - \frac{d(x_a^p, x_a^n)}{d(x_a, x_a^n)}| \tag{4}$$

$$\mathcal{L}_{ICT_D} = \frac{1}{N} \sum_{a=1}^{N} |d(x_a, x_a^n) - d(x_a^p, x_a^n)| \tag{5}$$

where $|.|$ represents ℓ_1-norm. The above losses constrain the absolute distances of the two mismatched pairs in a triplet to be as small as possible.

$\mathcal{L}_{ICT}$ can be any one of $\mathcal{L}_{ICT_F}$, $\mathcal{L}_{ICT_R}$ and $\mathcal{L}_{ICT_D}$. $\mathcal{L}_{ICT_D}$ does not vary with distance magnitude when the absolute difference between $d(x_a, x_a^n)$ and $d(x_a^p, x_a^n)$ is the same. But $\mathcal{L}_{ICT_F}$ and $\mathcal{L}_{ICT_R}$ can get large value when $d(x_a, x_a^n)$ and $d(x_a^p, x_a^n)$ are small. This sensitivity to the distance magnitude can result in equidistant optimizations taking over once the $d(x_a, x_a^n)$ and $d(x_a^p, x_a^n)$ is small. Therefore, when the distance of mismatched pairs is small, our $\mathcal{L}_{ICT_F}$ and $\mathcal{L}_{ICT_R}$ have a limited effect. In Section V-C, we conduct experiments to test the performance of different forms of $\mathcal{L}_{ICT}$.

Our ICT loss function aims to pull the matched pairs close by constraining the two mismatched pairs in a triplet to have equal distances within the batch. In a mini-batch, the batch hard sampling method for triplet makes a person have at least one matched pair with two mismatched samples (It's very rare that two mismatched samples to be the same sample). When there are two mismatched samples ($x_a^{n_1}$ and $x_a^{n_2}$) for

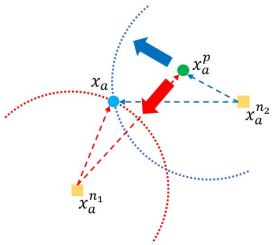


Fig. 2. Illustration of the isosceles constraint. To meet the isosceles constraint for $x_a^{n_1}$, the x_a^p can move towards the red arrow, and to meet the isosceles constraint for $x_a^{n_2}$, the x_a^p can move towards the blue arrow, thus the matched pair $\{x_a, a_a^p\}$ get closer.

a matched pair $\{x_a, x_a^p\}$, one possible schematic diagram are shown in Figure 2. In order to satisfy the isosceles constraint of $d(x_a, x_a^{n_1})$ and $d(x_a^p, x_a^{n_1})$, x_a^p can move in a direction approximate to the red arrow. Similarly, in order to satisfy the isosceles constraint of $d(x_a, x_a^{n_2})$ and $d(x_a^p, x_a^{n_2})$, x_a^p can move in a direction approximate to the blue arrow. Once the two isosceles constraints are satisfied simultaneously, the matched pair $\{x_a, x_a^p\}$ will be closer. Therefore, our ICT loss can improve performance of person ReID by pulling the matched pairs closer.

2) The Margin Constraint Term: Only isosceles constraints may lead to trivial results where all samples are aggregated to a point, so there must be a margin constraint to ensure that the mismatched pairs have relatively larger distances than the matched pairs. Our margin constraint considers the distance between two sets of mismatched and matched pairs in the triplet. In addition to the hardest samples within the batch (BH), there are semi-hard ones (BS), e.g., x_a and x_a^n are the semi-hard matched and mismatched samples for x_a^p. Batch hard triplet loss (Eqn. 2) uses a margin constraints for $d(x_a, x_a^p)$ and $d(x_a, x_a^n)$, which is helpful to identify the relatively simple samples within the batch. However, since we impose the isosceles constraint, it may cause two mismatched samples that are far apart to become closer after training. Therefore, in order to make all the mismatched pairs in the triplets farther than the matched pairs, we also impose a margin constraint to the semi-hard samples, which is expressed as follows

$$\mathcal{L}_{BST} = \frac{1}{N}\sum_{a=1}^{N}[d(x_a, x_a^p) - d(x_a^p, x_a^n) + \alpha]^+ \tag{6}$$

where α is the distance margin value between mismatched pairs and matched pairs in the triplets. To further facilitate the isosceles constraint, we use the same margin value.

3) The Overall Isosceles Constrained Triplet Loss: Combining the isosceles loss and the margin losses, we get the overall isosceles constrained triplet loss:

$$\mathcal{L}_{iso-tri} = (\mathcal{L}_{BHT} + \mathcal{L}_{BST}) + \lambda\mathcal{L}_{ICT}, \tag{7}$$

where λ decides the relative weight between $\mathcal{L}_{ICT}$ and $\mathcal{L}_{BHT} + \mathcal{L}_{BST}$. The isosceles constrained triplet loss adopts one isosceles and two margin constraints on the hardest triplets within a batch. On the premise of meeting the margin constraints (Eqn. 2 and 6), our isosceles constraint (Eqn. 3, 4 or 5) can pull the matched pairs closer and push the mismatched pairs farther away, thus reduce the number of crucial samples, and effectively improve the performance of ReID. We may refer to Figure 1 (b) again to see the optimized triplets.

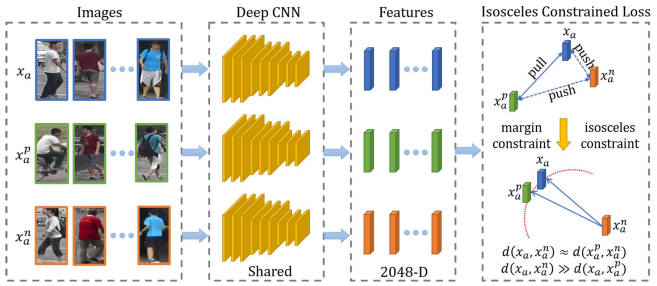


Fig. 3. The overall architecture of our method based on isosceles constrained loss.

In addition to batch hard triplet loss, traditional triplet loss also has a large number of crucial samples, which greatly affects the performance of ReID. Therefore, we can also directly apply the isosceles constraint to the triplet loss.

C. The Person ReID Network

1) Multi-Loss ReID: For deep learning based ReID methods, both identity loss, *e.g.*, cross-entropy (CE) and metric loss, *e.g.*, triplet loss, are usually used. Identity loss makes full use of label information to classify samples according to their identities, while the metric loss employs weak label information about sample pairs. In addition, these two types of losses have different optimization objectives. For example, the cross-entropy loss can enlarge the inter-class variance, while the triplet loss aims to widen the gap between matched pairs and mismatched pairs. Some works [16], [39] use both the triplet loss and the cross-entropy loss to make the model achieve higher matching accuracy. In this work, we combine ICT loss with cross-entropy loss to get better performance of person ReID.

2) Network Architecture: The overall architecture of our person ReID method is depicted in Figure 3. As the figure shows, we adopt the ResNet-50 [40] architecture pretrained on ImageNet [41] as our backbone. To adapt to the ReID task, we discard the last fully connected (FC) layer and add batch normalization (BN) layer after global average pooling (GAP) layer, so we get 2,048 dimensional features. For ICT loss learning, we use the 2,048-D features to calculate the Euclidean distance between sample pairs. For cross-entropy loss learning, we connect a FC layer after BN as the classifier, whose output units is equal to the number of class. In the test stage, we use the 2,048-D features to calculate the similarity between samples.

IV. Isosceles Constraint for Quadruplet

In this section, we will introduce our isosceles constraint applied on quadruplet for person ReID problem. First, we review some related quadruplet-based methods. Then, we present our method with detailed formal formulations.

For quadruplet, in addition to x_a, x_a^p and x_a^n in triplet, we choose an extra sample x_n^n that is closest to x_a^n, whose label y_n^n satisfies $y_n^n \neq y_a$ and $y_n^n \neq y_a^n$.

A. Revisit Quadruplet-Based Loss

Quadruplet loss [17] is modified based on the triplet loss. For any probe, it simultaneously considers two mismatched samples. The quadruplet loss can be formalized as

$$\mathcal{L}_{quad} = \frac{1}{N}\sum_{a=1}^{N}\{[d(x_a, x_p) - d(x_a, x_n) + \alpha_1]^+ + [d(x_a, x_p) - d(x_n, x_n') + \alpha_2]^+\} \quad (8)$$

where x_p matches x_a, x_n does not match x_a, x_n' does not match either x_a or x_a^n, α_1 and α_2 are pre-defined values, and N represents the number of quadruplets. To mine hard samples for better training, Xiao *et al.* [18] propose a margin sample mining loss, which choose the hardest matched pair and mismatched pair within a batch to constitute a quadruplet. In other words, there is only one quadruplet in a batch.

B. Isosceles Constrained Quadruplet Loss

Quadruplet loss selects four samples to calculate loss every time. Traditional quadruplet loss [17] choose all quadruplets while MSML [18] choose one quadruplet within a batch, which are not good for training. Inspired by batch hard sampling scheme for triplet [4], we propose a analogous sampling method for quadruplet.

1) The Margin Constraint Term: In order to ensure that the matched and mismatched pairs can be distinguished correctly, the batch hard margin constraint based on quadruplets can be expressed as

$$\mathcal{L}_{BHQ} = \frac{1}{N}\sum_{a=1}^{N}\{[d(x_a, x_a^p) - d(x_a, x_a^n) + \alpha]^+ + [d(x_a, x_a^p) - d(x_a^n, x_n^n) + \alpha]^+\} \quad (9)$$

where α is a pre-defined value, and N is equal to the batch size. Since both x_a^n and x_n^n are the hardest mismatched samples within the batch, different from the traditional quadruplet loss, we use the same margin α.

2) The Isosceles Constraint Term: There are naturally two samples x_a^n and x_n^n in the quadruplets that do not match the matched pair $\{x_a, x_a^p\}$. Therefore, just like the isosceles constraint for the triplets, we can construct the following isosceles constraint $\mathcal{L}_{ICQ}$ for quadruplets

$$\mathcal{L}_{ICQ_F} = \frac{1}{N}\sum_{a=1}^{N}\{|1 - \frac{1}{2}(\frac{d(x_a, x_a^n)}{d(x_a^p, x_a^n)} + \frac{d(x_a^p, x_a^n)}{d(x_a, x_a^n)})| + |1 - \frac{1}{2}(\frac{d(x_a, x_n^n)}{d(x_a^p, x_n^n)} + \frac{d(x_a^p, x_n^n)}{d(x_a, x_n^n)})|\} \quad (10)$$

$$\mathcal{L}_{ICQ_R} = \frac{1}{N}\sum_{a=1}^{N}\{|\frac{d(x_a, x_a^n)}{d(x_a^p, x_a^n)} - \frac{d(x_a^p, x_a^n)}{d(x_a, x_a^n)}| + |\frac{d(x_a, x_n^n)}{d(x_a^p, x_n^n)} - \frac{d(x_a^p, x_n^n)}{d(x_a, x_n^n)}|\} \quad (11)$$

$$\mathcal{L}_{ICQ_D} = \frac{1}{N}\sum_{a=1}^{N}\{|d(x_a, x_a^n) - d(x_a^p, x_a^n)| + |d(x_a, x_n^n) - d(x_a^p, x_n^n)|\} \quad (12)$$

3) The Overall Isosceles Constrained Quadruplet Loss: Combining the isosceles constraint and the margin constraint, we get the overall isosceles constrained quadruplet loss:

$$\mathcal{L}_{iso-qud} = \mathcal{L}_{BHQ} + \lambda\mathcal{L}_{ICQ}, \quad (13)$$

Similarly, when isosceles constrained quadruplet loss acts as a loss function to supervise model learning, we adopt the network shown in Figure 3. In addition, the optimization process can be referred to the analysis of isosceles constrained triplet loss. Since the matched pair $\{x_a, x_a^p\}$ in the quadruplet has two mismatched samples x_a^n and x_n^n, our isosceles constraint apply to quadruplet can pull the matched pair closer.

V. Experiments

In this section, we describe the experimental details and testify the effectiveness of ICT and ICQ loss function on three widely used ReID datasets.

A. Datasets and Settings

We evaluate our method on three public benchmark datasets, namely Market-1501 [22], DukeMTMC [42] and CUHK03 [24]. All the three datasets contain more than one thousand identities and the large numbers of images are closer to the practical application.

1) Market-1501: The dataset contains 32,668 images of 1,501 identities, which are labeled by bounding boxes with a DPM detector [43]. It is split into 751 identities for training and 750 identities for testing. Each identity is captured by six cameras at most and two cameras at least. In the testing phase, following the same setting of [22], 3,368 images are selected as probe set to query the correct identities across the testing set.

2) DukeMTMC-reID: This dataset contains 36,411 images of 1,812 identities captured by 8 high-resolution cameras, which are manually cropped from multi-camera tracking dataset DukeMTMC [42]. Following [44], 16,522 images of 702 identities are used for training, 2,228 images of another 702 identities are used as query images, and the remaining 17,661 images are gallery images.

3) CUHK03: This dataset is constituted by 14,096 images of 1,467 identities captured by several surveillance cameras. The dataset provides two types of bounding boxes annotations, including the manually annotated bounding boxes and automatically detected bounding boxes by the DPM detector, and we use the latter in this paper. Following [45], we adopt the new training/testing protocol to split the dataset into two balanced parts: 767 identities are in the training set and the rest 700 identities are in the testing set (gallery).

4) Evaluation Protocol: The dataset is separated into the training set and the testing set, in which images of the same person can only appear in either set. The testing set is further divided into probe set and gallery set, and the two sets contain different images of the same person. We evaluate the quality

TABLE I

COMPARISON WITH OTHER TRIPLET-BASED METHODS. $*$ MEANS THAT THE METHOD ADOPTS BATCH HARD SAMPLING METHOD [4]. ICT_F, ICT_R AND ICT RESPECTIVELY REPRESENT THAT OUR ISOSCELES CONSTRAINT IS IMPLEMENTED ACCORDING TO EQU. 3, 4 AND 5

Method	Market-1501				DukeMTMC-reID				CUHK03			
	mAP	r=1	r=5	r=10	mAP	r=1	r=5	r=10	mAP	r=1	r=5	r=10
Triplet [15]	50.9	70.1	86.5	91.1	58.5	74.9	86.5	90.0	42.2	43.7	65.0	73.9
Improved Triplet [5]	52.8	72.5	87.2	92.3	58.9	75.4	87.0	90.1	43.8	45.1	66.3	75.6
P2S [6]	53.9	75.5	89.3	93.2	59.2	75.1	86.8	90.0	42.9	44.6	65.8	74.7
Batch Hard [4]	73.2	86.1	94.2	96.1	66.5	81.7	90.6	93.0	59.4	62.4	78.4	85.8
Improved Triplet*	73.6	87.5	94.8	96.3	66.9	82.1	90.8	93.1	60.2	62.9	78.8	86.3
P2S*	73.9	87.8	95.2	96.4	67.2	82.5	90.8	93.2	60.6	63.1	79.1	86.3
HAP2S [16]	74.1	87.4	94.9	96.6	67.6	82.8	90.9	93.0	60.4	62.8	78.6	85.9
ICT_F	76.1	89.3	95.4	96.8	68.3	83.1	91.1	93.2	61.9	65.0	81.1	86.7
ICT_R	78.8	90.9	**96.7**	**97.7**	69.9	**84.0**	**92.0**	**94.0**	62.3	65.3	**81.6**	**87.9**
ICT_D	**79.3**	**91.6**	96.3	**97.7**	**70.5**	83.9	91.3	93.5	**64.4**	**67.4**	81.3	87.8

of different person ReID models using Cumulative Matching Characteristic (CMC) [46] curves and Mean Average Precision (mAP). All experiments are performed in single query setting.

B. Implementation Details

We adopte the network architecture shown in Figure 3, and set the output units of the last FC layer to 751, 702 and 767 for Market-1501, DukeMTMC-reID and CUHK03 respectively.

In training phase, we resize all input images to 256×128 and adopt horizontal flipping for data augmentation. We use Adam [47] optimizer to train the network and 4 images from each of 16 persons are randomly chosen as a 64-size mini-batch. We set the parameters of both ICT and ICQ loss with margin $\alpha = 0.3$ and weight $\lambda = 1.0$. The model is trained for 60 epoches. The initial learning rate is set as 3.0×10^{-4} and reduced by 10 times at the 20-th and 40-th epoch. In testing phase, we do not apply any data augmentations on account of efficiency. For all evaluation, we take the output of the BN layer as the intermediate feature embedding (2,048-D) and use the Euclidean distance to compute the similarity between query and gallery.

C. Comparison With Triplet-Based Losses

We compare our ICT with some triplet-based methods reported in recent person ReID works, including triplet loss [15], improved triplet loss [5], P2S triplet loss [6], batch hard triplet loss [4] and HAP2S loss [16]. To be fair in comparison, we reproduce these methods by applying the same mini-batch configuration and tune the parameters to optimum for each loss. In addition, we also compared improved triplet loss [5] and P2S triplet loss [6] with batch hard sampling scheme, namely Improved Triplet* and P2S*.

The experimental results on three datasets are presented in Table I. From the table we can see that our ICT loss using any $\mathcal{L}_{ICT_F}$, $\mathcal{L}_{ICT_R}$ and $\mathcal{L}_{ICT_D}$ consistently outperforms all other competitors on three datasets. All loss function in Table I are based on triplet. Improved Triplet [5] and P2S [6] with fixed-value constraints to facilitate triple loss have some effect. But the same restriction on the distance of all matched or mismatched pairs will discourage distortions in the feature. Our proposed isoscelses constraint adaptively optimizes the distance between samples through relative restriction. Therefore, ICT can pull the matched pairs close while avoiding feature space distortions, thus further reducing thethe number of crucial samples.

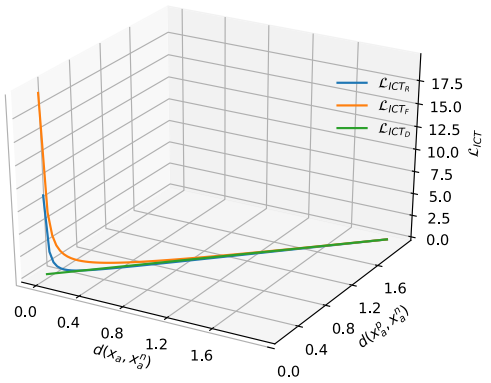


Fig. 4. Three forms of $\mathcal{L}_{ICT}$. When $d(x_a, x_a^n)$ and $d(x_a^p, x_a^n)$ are small, the slight difference between the two causes the $\mathcal{L}_{ICT_F}$ and $\mathcal{L}_{ICT_R}$ to get large loss.

We also observe that the performance improvement of $\mathcal{L}_{ICT_F}$, $\mathcal{L}_{ICT_R}$ and $\mathcal{L}_{ICT_D}$ shows an increasing trend. The differences among the three are shown in Figure 4. From the figure, we can see that the three forms $\mathcal{L}_{ICT}$ have different values when the difference between $d(x_a, x_a^n)$ and $d(x_a^p, x_a^n)$ is the same, especially if the distance itself is small. The experimental results and curve graph indicate that when the distance of mismatched pairs is small, the performance of our isosceles constraint is limited. Therefore, our isosceles constraint needs to be used with the margin constraint.

As shown in Table I, our ICT achieves a significant improvement in mAP, e.g., +6.1% (79.3-73.2) on Market-1501, +4.0% (70.5-66.5) on DukeMTMC-reID, and +5.0% (64.4-59.4) on CUHK03, which means that all samples matching the probe rank higher on the retrieve list. The experimental results prove that our isosceles constraint really closes the distance of all matched samples, thus effectively reducing the number of crucial samples. Uniformly, the practical $\mathcal{L}_{ICT_D}$ employed for triplet is adopted in all subsequent experiments (notes for ICT) to represent our isosceles constraint.

D. Comparison With State-of-the-Arts

We compare our ICT with the recent state-of-the-art methods on three datasets, including HA-CNN [35],

TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS

Method		Market-1501		DukeMTMC-reID		CUHK03	
		mAP	r=1	mAP	r=1	mAP	r=1
HA-CNN [35]	CVPR18	75.7	91.2	63.8	80.5	38.6	41.7
PCB+RPP [32]	ECCV18	81.6	93.8	69.2	83.3	-	-
PABR [48]	ECCV18	76.0	90.2	64.2	82.1	-	-
MLFN [49]	CVPR18	74.3	90.0	62.8	81.0	47.8	52.8
MGCAM [36]	CVPR18	74.3	83.8	-	-	46.9	46.7
Random Walk [50]	CVPR18	82.5	92.7	66.4	80.7	-	-
DGRW [50]	CVPR18	82.5	92.7	66.4	80.7	-	-
SGGNN [51]	ECCV18	82.8	92.3	68.2	81.1	-	-
SVDNet [52]	ICCV17	62.2	82.3	56.8	76.7	37.3	41.5
Re-ranking [45]	CVPR17	63.6	77.1	-	-	37.4	34.7
HAP2S [16]	ECCV18	69.8	84.6	60.6	76.1	-	-
AOS [53]	CVPR18	70.4	86.5	62.1	79.2	43.3	47.1
AACN [54]	CVPR18	66.9	85.9	59.3	76.8	-	-
DuATM [55]	CVPR18	76.6	91.4	64.6	81.8	-	-
EB [56]	CVPR18	80.5	-	-	-	-	-
EDKM [57]	CVPR18	75.3	90.1	63.2	80.3	-	-
AWTL [58]	CVPR18	79.7	89.5	63.4	79.8	-	-
GCSLDC [59]	CVPR18	81.6	93.5	69.5	84.9	-	-
SPReID [60]	CVPR18	83.4	93.7	73.3	86.0	-	-
BraidNet-CS+SRL [61]	CVPR18	69.5	83.7	59.5	76.4	-	-
DaRe [62]	CVPR18	69.3	86.4	57.4	75.2	53.7	55.1
HACNN*+DHA-Net [21]	TIP19	76.0	91.3	64.1	81.3	-	-
ICT		79.3	91.6	70.5	83.9	64.4	67.4
ICT (RE)		81.3	91.6	71.7	85.1	65.1	68.5
ICT+CE		83.7	94.2	73.5	86.6	66.8	69.3
ICT+CE (RE)		**84.9**	**94.4**	**75.3**	**88.1**	**67.6**	**70.2**

PCB+RPP [32], PABR [48], MLFN [49], MGCAM [36], Random Walk [50], DGRW [50], SGGNN [51], SVDNet [52], Re-ranking [45], HAP2S [16], AOS [53], AACN [54], DuATM [55], EB [56], EDKM [57], AWTL [58], GCSLDC [59], SPReID [60], BraidNet [61], DaRe [62] and DHA [21]. Random erasing (RE) [63] is an adaptable data augmentation, we also test its effectiveness in our approach. For the sake of testing efficiency, we apply neither test-phase augmentation nor post-ranking. The experimental results are shown in Table II. From the table, we can get the following conclusions:

Firstly, our ICT is better than most of state-of-the-art methods. Our proposed isosceles constraint specifically constrains the distance of the matched pair, so that the images of the same identity can be ranked in front during retrieval. When compared with a recently reported model MGCAM [36], which combine metric learning with identify loss and using external information (mask), the proposed method achieves an improvement of +5.0% in mAP and +7.8% in Rank-1 on Market-1501, +17.5% in mAP and +20.7% in Rank-1 on CUHK03. The experimental results prove that our method is reliable and has certain performance advantages on person ReID.

Secondly, our multi-loss method can achieve the best mAP and Rank-1 accuracy on all three datasets, especially the mAP outperforms most the methods by a large margin. We respectively gain 83.7%, 73.5%, 66.8% in mAP and 94.2%, 86.6%, 69.3% in Rank-1 on Market-1501, DukeMTMC-reID and CUHK03. We also notice that our method is only a little better than SPReID [60], but SPReID uses a larger size image (748×246) and external information (mask), which is computationally time consuming. Our method only improves on the loss function, which may be further improved if combined with other training techniques.

TABLE III
MULTI-LOSS RESULTS ON MARKET-1501

Method	mAP	r=1	r=5	r=10
CE [1]	78.9	93.1	97.2	97.4
Triplet [15]+CE	79.1	93.2	97.6	98.1
Improved Triplet [5] +CE	79.3	93.5	97.8	98.3
P2S [6]+CE	79.2	93.2	97.7	98.2
Batch Hard [4] +CE	82.4	93.6	97.8	98.6
Improved Triplet*+CE	82.5	93.5	97.5	98.3
P2S*+CE	82.2	93.2	97.6	98.4
HAP2S [16]+CE	82.3	93.7	97.6	98.5
ICT+CE	**83.7**	**94.2**	**97.9**	**98.7**

In addition, we test the effect of RE in combination with our method. After using RE, our method can be further improved, *e.g.,* +2.0% in mAP on Market-1501 when our ICT adopts RE. This shows that RE is effective, and at the same time, our method can co-work with RE.

E. Further Analysis

1) Effect of Multi-Loss: Since the metric loss only focuses on the disparity between matched pairs and mismatched pairs, but not on the variances between different categories. We add identity loss (cross-entropy) on the basis of the metric loss. For a clear description, we mark the multi-loss methods as $*$+CE, where $*$ represents the original metric learning method. To test the effect of the two loss combinations, we conduct experiments on Market-1501 dataset. The experimental results are reported in Table III.

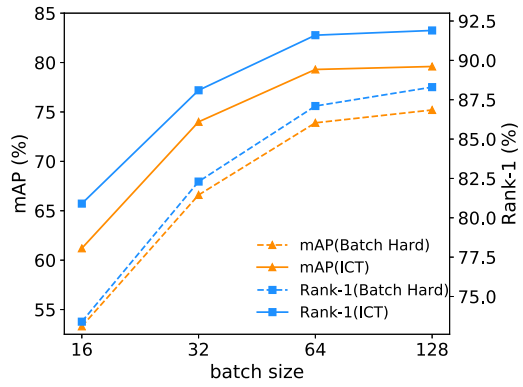


Fig. 5. Effect of batch size on batch hard triplet loss and ICT.

From the table, we can see that the mAP of the multi-loss is significantly improved compared with the cross-entropy loss. When compared with triplet-based methods recorded in Table I, the Rank-k of the multi-loss is improved obviously. This experimental phenomenon shows that multi-loss properly reduces the intra-class variances introduced by identity loss, and increases inter-class variances caused by metric learning loss. In all variants of triplet loss combined with cross-entropy, our ICT+CE achieves the best results. We gain 83.7% in mAP and 94.2% in Rank-1, which outperforms cross-entropy by 4.8% (83.7-78.9) in mAP and outperforms ICT by 2.6%(94.2-91.6) in Rank-1. The performance improvement shows that muli-loss can make up for the deficiency of identity loss and metric loss.

2) Effect of Batch Size: In order to test the impact of batch size on performance, we perform experiments on our ICT and the most relevant batch hard triplet loss with varying batch sizes, the experimental results are shown in Figure 5. From the figure, we can see that, compared with batch hard triplet loss, our ICT achieves consistent improvements in both mAP and Rank-*k*, with the extent of improvement decreases with the increase of batch size. Actually, the number of crucial samples is reduced with the increase of batch size and our ICT can further reduce the number of crucial samples in various batch size.

3) Generalization Ability: The datasets about person ReID are usually collected in different scenarios, such as Market-1501 collected on the Tsinghua University campus and DukeMTMC-reID collected from Duke University. When the model is used to solve the person ReID problem in the actual application scenario, the wide data collection difference requires the model to have a good generalization ability. In order to evaluate the generalization ability of the model for cross-scenario, we test the results of transferring between Market-1501 and DukeMTMC-reID.

The experimental results are reported in Table IV. Compared with Table I on DukeMTMC-reID, we can see a dramatic performance drop on the same dataset. However, our ICT loss has the best generalization ability compared to other methods, we obtain 21.9% in mAP and 36.4% in Rank-1 on DukeMTMC-reID, which is 4.7% (21.9-17.2) higher in mAP and 5.7% (36.4-30.7) in Rank-1 compared with the second best method HAP2S [16]. Similarly, our ICT has better accuracy when training on DukeMTMC-reID but testing on Market-1501. The experimental results show that our ICT has better generalization ability than other triplet-based losses.

TABLE IV

RESULTS WHEN THE MODEL IS TRAINED ON MARKET-1501 BUT TESTED ON DUKEMTMC-REID (MARKET→DUKE) AND WHEN THE MODEL IS TRAINED ON DUKEMTMC-REID BUT TESTED ON MARKET-1501 (DUKE→MARKET)

Method	Market→ Duke		Duke→Market	
	mAP	r=1	r=5	r=10
Triplet [15]	7.7	14.7	13.1	33.8
Improved Triplet [5]	13.2	24.8	17.6	37.9
P2S [6]	11.7	22.2	14.5	34.7
Batch Hard [4]	16.6	29.5	19.8	40.6
Improved Triplet*	17.1	30.3	20.7	44.9
P2S*	16.7	29.8	20.1	44.2
HAP2S [16]	17.2	30.7	21.4	46.8
ICT	**21.9**	**36.4**	**25.6**	**54.8**
CamStyle [37]	25.1	48.4	27.4	58.8
HHL [64]	27.2	46.9	31.4	62.2
ECN [65]	40.4	63.3	43.0	75.1
LIAM(GPP) [66]	54.4	74.0	63.8	84.1

TABLE V

TEST RESULTS FOR DIFFERENT CONSTRAINT COMBINATIONS ON MARKET-1501, THE TOTAL LOSS IS OBTAINED BY ADDING CONSTRAINTS WITH EQUAL WEIGHT

Test	$\mathcal{L}_{BHT}$	$\mathcal{L}_{BST}$	$\mathcal{L}_{ICT}$	mAP	r=1	r=5	r=10
1	✓			73.2	86.1	94.2	96.1
2		✓		68.2	84.1	93.1	95.5
3			✓	4.7	12.6	26.7	34.4
4	✓	✓		73.1	86.5	94.5	96.0
5	✓		✓	77.9	90.0	96.1	97.5
6		✓	✓	77.1	89.8	96.0	97.1
7	✓	✓	✓	**79.3**	**91.6**	**96.3**	**97.7**

In addition, we compare our ICT with some unsupervised domain adaptation methods in Table IV. From the table we can see that some methods are better than our ICT. But they adopt some transfer learning skills, and our method only improve loss function. The results show that generalization ability can be improved from different levels.

4) Comparison of Different Constraint Combinations: To evaluate the contribution of the isosceles constraint between two mismatched pairs in the triplet, we design seven experiments to test the performance of ReID on Market-1501, in which the triplets meet the constraints of different combinations. Considering the fairness of comparison, all experiments use batch hard sampling scheme [4]. The experimental results are shown in Table V.

As can be seen from the table, Test #5, 6, 7 with an isosceles constraint is significantly better than Test #1, 2, 4, which shows that applying the isosceles constraint on triplets while meeting the margin constraint can effectively reduce the number of crucial samples. In addition, we notice that Test #3 with only isosceles constraint has very low precision. This is because when there is only an isosceles constraint, all samples will be clustered to one point, resulting in the inability to correctly distinguish matched pairs from mismatched pairs. Therefore, our isosceles constraint pulls the matched

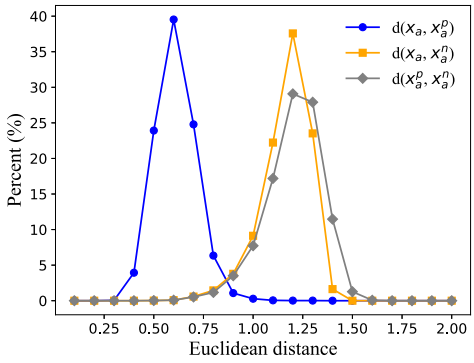


Fig. 6. Distance distribution of all batch hard triplets on training set of Market-1501.

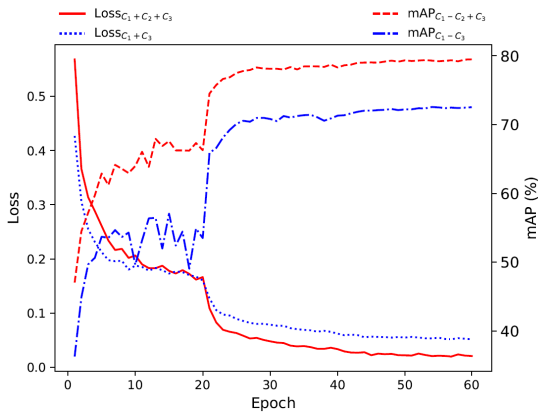


Fig. 7. Convergence comparisons on Market-1501. Learning rate dropped 10 times at 20 and 40 epoch respectively.

pairs closer, margin constraint can push the mismatched pairs away, and their combined effect can effectively reduce the number of crucial samples, thus improving the performance of ReID.

5) Effect of Add $\mathcal{L}_{BST}$ to $\mathcal{L}_{BHT}$ Constraint: From Table V, we can see that there is no accuracy improvement after adding $\mathcal{L}_{BST}$ to $\mathcal{L}_{BHT}$ constraint. Intuitively, when the triplets satisfy both $\mathcal{L}_{BHT}$ and $\mathcal{L}_{BST}$ constraints, person ReID should have better performance than only satisfying $\mathcal{L}_{BHT}$ constraint. In fact, the sampling method of batch hard triplet loss determines that $d(x_a, x_a^n)$ is the hardest distance for x_a within the batch, while $d(x_a^p, x_a^n)$ is semi-hard for x_a^p, thus $\mathcal{L}_{BST}$ constraint is more likely to satisfy compared with $\mathcal{L}_{BHT}$ constraint. Therefore, adding $\mathcal{L}_{BST}$ constraint can not further affect network optimization. To further demonstrate the impact of $\mathcal{L}_{BST}$, we calculate the distance distribution of all batch hard triplets involved in training when the network converges. The statistical results on Market-1501 are shown in Figure 6. From the distance distribution diagram, in general, we can see that the distances of some mismatched pairs $\{x_a^p, x_a^n\}$ are larger than the distance of $\{x_a, x_a^n\}$ after model convergence.

6) Effect of Add $\mathcal{L}_{BST}$ to $\mathcal{L}_{BHT} + \mathcal{L}_{ICT}$ Constraint: From Table V, we can see that after adding the isosceles constraint $\mathcal{L}_{ICT}$ to $\mathcal{L}_{BHT}$, and then increasing the $\mathcal{L}_{BST}$ constraint, the performance is improved. This is because the existence of the isosceles constraint may make the semi-hard $d(x_a^p, x_a^n)$ close to the hard $d(x_a, x_a^n)$, thus not meeting the margin constraint. Therefore, in order to ensure that the distances of all mismatched pairs in the triplets are larger than that of matched pairs, margin constraint should be applied to both. The comparison of loss and mAP on Market-1501 is shown in the Figure 7. We can see that our isosceles constrained triplet loss obtains the better convergence than $\mathcal{L}_{BHT}+\mathcal{L}_{ICT}$ constraint.

7) Effect of Our Isosceles Constraint on Different Backbones: With the popularity of deep learning, many backbones have been designed for visual tasks, *e.g.,* GoogLeNet [67], ResNet [40], DenseNet [68]. To verify that our isosceles constraint is backbone independent, we choose different backbones for feature extraction. The result are see in Table VI. From the table, we can see consistent improvements in our ICT under different backbones.

TABLE VI

RESULTS WHEN OUR ICT IS TRAINED WITH DIFFERENT BACKBONES ON MARKET-1501

Backbone	Method	mAP	r=1	r=5	r=10
ResNet-50	Batch Hard [4]	73.2	86.1	94.2	96.1
	ICT	79.3	91.6	96.3	97.7
GoogLeNet	Batch Hard [4]	70.2	84.7	93.2	95.8
	ICT	73.2	87.7	94.5	96.7
DenseNet-121	Batch Hard [4]	61.1	79.9	92.2	94.3
	ICT	63.0	81.4	93.1	94.4

TABLE VII

RESULTS OF COMBINING ICT IN DIFFERENT FORMS ON MARKET-1501

$\mathcal{L}_{ICT_F}$	$\mathcal{L}_{ICT_R}$	$\mathcal{L}_{ICT_D}$	mAP	r=1	r=5	r=10
$\frac{1}{2}$	$\frac{1}{2}$	0	78.6	90.7	96.1	97.4
$\frac{1}{2}$	0	$\frac{1}{2}$	77.5	89.0	95.8	97.1
0	$\frac{1}{2}$	$\frac{1}{2}$	78.7	90.9	96.0	97.4
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	78.8	91.1	96.3	97.6

8) Comparison of Combination Among $\mathcal{L}_{ICT_F}$, $\mathcal{L}_{ICT_R}$ and $\mathcal{L}_{ICT_D}$: To further evaluate the effects of different forms of isosceles constraints, we test several combinations. The result are see in Table VII. From the table, we can see that the combination of the three has a better effect, but not more than just $\mathcal{L}_{ICT_D}$ (see Table I). In addition, when the worst $\mathcal{L}_{ICT_F}$ is combined with $\mathcal{L}_{ICT_R}$ or $\mathcal{L}_{ICT_D}$, the performance can be improved to a certain extent compared to $\mathcal{L}_{ICT_F}$. When the best $\mathcal{L}_{ICT_D}$ is combined with $\mathcal{L}_{ICT_R}$ or $\mathcal{L}_{ICT_F}$, the performance decreases. The experimental result shows that the improvement of our isosceles constraint is limited at the distance of the mismatched pair is small.

9) Parameter Analysis: We evaluate the impact of λ in Equ. (7) on Market-1501. λ controls the relative importance of the proposed isosceles constraint. As shown in Figure 8, the proposed constraint is proven effective when compared to $\lambda = 0$ but a larger λ does not bring more gains in accuracy. Specifically, when $\lambda = 1.5$ yields the best mAP and $\lambda = 1.0$ yields the best Rank-1. But there is a very small performance difference in the interval [1, 1.5]. To make a fair comparison, we set $\lambda = 1.0$ for all experiments.

TABLE VIII

COMPARISON WITH QUADRUPLET-BASED METHODS. ∗ MEANS THAT THE METHOD ADOPTS THE BATCH HARD SAMPLING METHOD

Method	Market-1501				DukeMTMC-reID				CUHK03			
	mAP	r=1	r=5	r=10	mAP	r=1	r=5	r=10	mAP	r=1	r=5	r=10
Quadruplet [17]	61.2	79.9	91.3	92.9	60.2	76.3	88.1	90.9	50.3	53.2	75.0	83.3
MSML [18]	70.6	85.2	93.7	95.3	63.9	78.4	87.0	90.1	56.8	59.3	75.1	83.8
Quadruplet*	75.3	87.8	94.7	96.4	65.6	80.1	88.8	91.3	59.5	61.9	77.8	85.9
ICQ_F	77.2	89.9	95.3	96.7	68.5	82.7	91.1	93.1	61.9	63.4	80.1	87.2
ICQ_R	**79.6**	**91.0**	**96.3**	**97.8**	69.6	**83.8**	**91.6**	93.7	62.7	64.6	81.0	**87.8**
ICQ_D	78.5	**91.0**	**96.3**	97.7	**70.4**	83.4	91.5	**93.9**	**63.5**	**66.3**	**81.1**	87.3

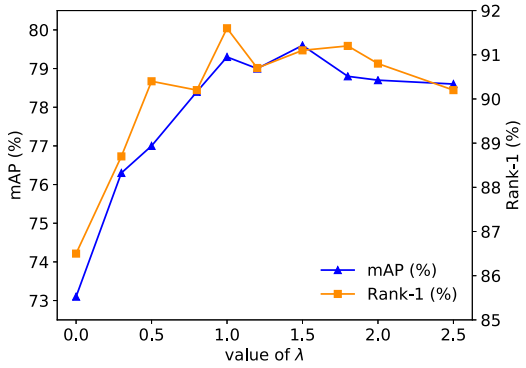


Fig. 8. Influence of the parameter λ in Equ. 7 and 13 on Market-1501.

F. Comparison With Quadruplet-Based Losses

To prove the universality of isosceles constraint, we conduct experiments for quadruplet on Market-1501, DukeMTMC-reID and CUHK03. We compare our ICQ with some quadruplet-based methods, including Quadruplet [17], MSML [18] and Batch Hard Quadruplet loss (Quadruplet*) improved by us. The experimental results are shown in Table VIII. From the table we can see consistent improvements on all three datasets after adding isosceles constraint. Specifically, compared with Quadruplet*, ICQ gains +3.2%, +4.8% and 4.0% in mAP on Market-1501, DukeMTMC-reID and CUHK03 respectively. In addition, we notice that ICQ and ICT finally achieve proximity accuracy, which shows that as long as there are multiple mismatched samples for one matched pair during the optimization process, the isosceles constraint can improve the accuracy by shortening the distance of the matching pair.

G. Isosceles Constraint for Fine-Grained Classification

Fine-grained

classification aims to differentiate subordinate classes of a common superior class, which is much necessary to obtain discriminative features. To verify our isosceles constraint on fine-grained classification task, we conduct experiments on commonly used fine-grained classification datasets. *e.g.* Caltech-UCSD Birds (CUB-200-2011) [69] and Stanford Cars (Cars196) [70]. CUB-200-2011 is a bird classification task with 11,788 images from 200 wild bird species, 5,994 images for training and the rest 5,794 images for testing. Cars196 dataset contains 16,185 images over 196 classes, and each class has a roughly 50-50 split, 8,144 images for training and 8,041 images for testing.

TABLE IX

EXPERIMENTAL RESULTS FOR FINE-GRAINED CLASSIFICATION ON CUB-200-2011 AND CARS196

Method	Accuracy(%)	
	CUB−200−2011	Cars196
CE	84.5	90.1
CE+Batch Hard [4]	86.2	92.1
CE+ICT	88.1	93.8

TABLE X

EXPERIMENTAL RESULTS FOR VEHICLE REID ON VERI AND VEHICLEID. BATCH HARD IS THE BASELINE IN THE PAPER

Dataset	Method	mAP	r=1	r=5	r=10
VeRi	Batch Hard [4]	64.6	87.8	94.6	96.6
	ICT	67.8	89.0	95.2	96.9
VehicleID	Batch Hard [4]	66.3	81.2	93.9	96.7
	ICT	68.7	83.5	95.6	97.4

The experimental results are shown in Table IX. From the table, we can see that metric learning loss function (Batch Hard Tripelt) combined with identification loss (CE) has a certain improvement over identification loss. When the metric learning loss function is replaced with our ICT, we further achieve consistent improvement on both CUB-200-2011 and Cars196. The improvement shows that our isosceles constraint is effective on fine-grained classification.

H. Isosceles Constraint for Vehicle ReID

1) Datasets and Setting: To evaluate the effectiveness of our isosceles constraint on vehicle ReID benchmarks, we conduct experiments on commonly used vehicle datasets, *e.g.* VeRi [71] and VehicleID [72]. VeRi includes 40,000 bounding box annotations of 776 cars across 20 cameras in traffic surveillance scenes, 37,778 images of 576 classes for training and the others for testing. VehicleID has 221,763 bounding boxes of 26,267 identities, captured across various surveillance cameras in a city. We use 113,346 images of 13,164 classes for training and 19,777 images of 2400 classes for testing.

The experimental results are shown in Table X. From the table, we can see that our ICT can achieve consistent improvement on both VeRi and VehicleID. Specially, for VeRi, we achieve +3.2% in mAP and +1.2% in Rank-1, for VehicleID, we achieve +1.6% in mAP and +2.3% in Rank-1. The improvement shows that our isosceles constraint works on vehicle ReID.

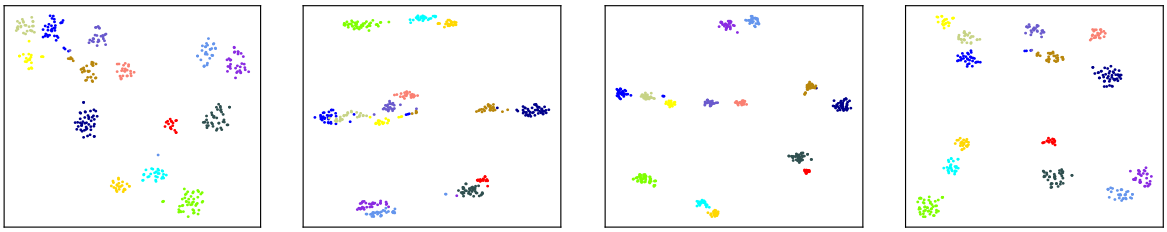
I. Generic Deep Metric Learning

1) Datasets and Setting: To further evaluate the effectiveness of our isosceles constraint, we apply our

TABLE XI

COMPARISON WITH THE OTHER METRIC LEARNING METHODS ON CUB-200-2011 AND CARS196 DATASET. THE BEST AND SECOND BEST RESULTS ARE MARKED BY RED AND BLUE COLORS, RESPECTIVELY

Method	CUB-200-2011					Cars196				
	NMI	R@1	R@2	R@4	R@8	NMI	R@1	R@2	R@4	R@8
Lifted Struct [73]	56.5	43.6	56.6	68.6	79.6	56.9	53.0	65.7	76.0	84.3
N-pairs [74]	57.2	45.4	58.4	69.5	79.5	57.8	53.9	66.8	77.7	86.4
Clustering [75]	59.2	48.2	61.4	71.8	81.9	59.0	58.1	70.6	80.3	87.8
Proxy NCA [76]	59.5	49.2	61.9	67.9	72.4	64.9	73.2	82.4	86.4	88.7
Smart Mining [77]	59.9	49.8	62.3	74.1	83.3	59.5	64.7	76.2	84.2	90.2
HDC [78]	-	53.6	65.7	77.0	85.6	-	73.7	83.2	89.5	93.8
Angular Loss [79]	61.0	54.7	66.3	76.0	83.9	63.2	71.4	81.4	87.5	92.1
HAP2S [16]	63.4	56.1	68.4	79.2	86.9	63.1	74.1	83.5	89.9	94.1
ICT	65.2	59.1	71.1	80.8	88.1	77.4	91.2	94.2	96.9	98.1
ICQ	65.7	59.0	70.1	79.8	87.5	76.2	87.6	93.1	96.0	97.7



(a) (b) (c) (d)

Fig. 9. Visualization of deeply-learned features by (a) CE loss [1], (b) batch hard triplet loss [4], (c) our ICT, (d) ICT combination of CE loss. The points with different colors denote features from different identities. (Best viewed in color).

ICT and ICQ to the popular deep metric learning benchmarks CUB-200-2011 [69] and Cars196 [70]. The CUB-200-2011 dataset consists of 11,788 bird images of 200 species, following the same training/testing setting described in [73], the first 100 species (5,864 images) are used for training and the others for testing. The Cars196 dataset contains 16,185 car images of 196 classes, where the first 98 classes (8,054 images) for training and the rest for testing. We use two standard metrics, the normalized mutual information (NMI) [80] and Recall@K [81], to measure the performances. We adopt the same network architecture as person ReID task only with ICT or ICQ (Figure 3) for training and testing. The initial learning rate is $1.0 \times e^{-4}$, $\alpha = 0.1$, all images are cropped to 224×224 and 8 images from each class are selected to form the 64(8*8) mini-batch.

2) Comparison With State-of-the-Arts: We compare our ICT and ICQ with state-of-the-art methods as shown in Table XI. From the table we can see that both ICT and ICQ outperform the other methods on the CUB-200-2011 and Cars196 datasets. Specifically, our ICQ obtains best NMI and ICT gets best Rank-k on CUB-200-2011. It is worth noting that our method performs well on the Cars196, ICT gain +12.5% (77.4-64.9) in NMI and +15.0% (89.1-74.1) in R@1 than the previous best method, which indicates that our method has better clustering quality and retrieval performance than other losses. Therefore, isosceles constraint can act as a general constraint on a variety of retrieval tasks. From the table, we can see that isosceles constraint applied on triplet is better than that on quadruplet, we guess that quadruplet itself has the stronger constraint than triplet, so the performance improvement of our isosceles constraint on quadruplet grew weaker than triplet.

J. Visualization Analysis

1) Visualization Analysis of Features: We adopt the t-SNE [82] tool to visualize the feature embeddings learned by the losses. We randomly choose 15 identities from the testing set of Market-1501. The visualization results of the features are plotted in Figure 9. As shown in Figure 9 (a), cross-entropy loss can separate different categories, but there are significant intra-class variances. Since person ReID is an open-set task, large intra-class variances may introduce crucial samples. In Figure 9 (b), batch hard triplet loss does not directly optimize the distance of matched samples, which makes it difficult to distinguish some mismatched samples from matched samples. In Figure 9 (c), our isosceles triplet loss achieves smaller intra-class variances, which can effectively reduce the negative influence of crucial samples. Therefore, our isosceles triplet loss can make the ordering of matched samples higher in the retrieval results, thus improving mAP and Rank-k. For better retrieval performance, our multi-loss method combined with isosceles triplet loss and cross-entropy loss achieves the effect shown in Figure 9 (d), which has smaller intra-class variances than (a) and larger inter-class variances than (c).

2) Visualization Analysis of Retrieval Results: To intuitively describe the advantages of our approach, we visualize the retrieval results of the probe on the testing set of Market-1501. The query results of the probe are shown in Figure 10. For each prob, we visualize the top 6 images in the ranking lists. The row (a) and (b) are optimized by batch hard triplet [4] and Quadruplet*, (c) and (d) are the result of our ICT and ICQ. From the retrieval results, we can see that our ICT/ICQ is able to place the real matched samples to probe at the top of the list, which shows that our isosceles constraint applied

Fig. 10. Illustration of the retrieval results on Market-1501. The green rectangle represents a true match, and the red rectangle represents a negative match. The results of row (a), (b), (c) and (d) are obtained by training of batch hard triplet loss, batch hard quadtruplet loss, isosceles constrained tripelt loss and isosceles constrained quadtruplet loss respectively.

to triplet and quadtruplet can indeed pull the matched pairs closer, thus improving the performance of person ReID.

VI. Conclusion

In this paper, we propose a novel isosceles constraint to reduce the number of crucial samples, which can be applied to both triplet and quadruplet. When we employ isosceles constraint to each two mismatched pairs in the triplet or quadruplet, the matched pairs will be pulled closer. Meanwhile, to ensure that the distance of all mismatched pairs are larger than the matched pair, the margin constraints between the mismatched pairs and the matched pair are indispensable. The isosceles and margin constraints together pull the matched pairs closer and push the mismatched pairs farther away than matched ones. By this way, crucial samples are effectively reduced and the performance on ReID is improved greatly. Experiment results suggest that the proposed method is very effective on the person ReID dataset, Market1501, DukeMTMC-reID and CUHK03. In addition, our ICT and ICQ can also apply to other retrieval tasks besides person ReID.

References

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," 2016, *arXiv:1610.02984*. [Online]. Available: http://arxiv.org/abs/1610.02984

[2] S. Zhang, J. Wang, Z. Wang, Y. Gong, and Y. Liu, "Multi-target tracking by learning local-to-global trajectory models," *Pattern Recognit.*, vol. 48, no. 2, pp. 580–590, Feb. 2015.

[3] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008.

[4] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*. [Online]. Available: http://arxiv.org/abs/1703.07737

[5] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1335–1344.

[6] S. Zhou, J. Wang, J. Wang, Y. Gong, and N. Zheng, "Point to set similarity based deep feature learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3741–3750.

[7] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.

[8] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 1–16.

[9] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 709–720.

[10] S. Zhou, J. Wang, Q. Hou, and Y. Gong, "Deep ranking model for person re-identification with pairwise similarity comparison," in *Proc. Pacific Rim Conf. Multimedia*, 2016, pp. 84–94.

[11] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 34–39.

[12] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1846–1855.

[13] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A siamese long short-term memory architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 135–153.

[14] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2666–2672.

[15] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.

[16] R. Yu, Z. Dou, S. Bai, Z. Zhang, Y. Xu, and X. Bai, "Hard-aware point-to-set deep metric for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 188–204.

[17] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 403–412.

[18] Q. Xiao, H. Luo, and C. Zhang, "Margin sample mining loss: A deep learning based method for person re-identification," 2017, *arXiv:1710.00478*. [Online]. Available: http://arxiv.org/abs/1710.00478

[19] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 144–151.

[20] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 68.1–68.11.

[21] Z. Wang, J. Jiang, Y. Wu, M. Ye, X. Bai, and S. Satoh, "Learning sparse and identity-preserved hidden attributes for person re-identification," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 2013–2025, Oct. 2020.

[22] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1116–1124.

[23] D. Cheng, X. Chang, L. Liu, A. G. Hauptmann, Y. Gong, and N. Zheng, "Discriminative dictionary learning with ranking metric embedded for person re-identification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 964–970.

[24] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.

[25] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1095–1108, May 2015.

[26] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3586–3593.

[27] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2530–2539.

[28] Q. Zhou *et al.*, "Robust and efficient graph correspondence transfer for person re-identification," *IEEE Trans. Image Process.*, early access, May 8, 2019, doi: 10.1109/TIP.2019.2914575.

[29] B. Ma, Y. Su, and F. Jurie, "BiCov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 57.1–57.11.

[30] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2197–2206.

[31] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3908–3916.

[32] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 480–496.

[33] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, and Q. Tian, "Deep representation learning with part loss for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2860–2871, Jun. 2019.

[34] L. Zheng, Y. Huang, H. Lu, and Y. Yang, "Pose-invariant embedding for deep person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4500–4509, Sep. 2019.

[35] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2285–2294.

[36] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1179–1188.

[37] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "CamStyle: A novel data augmentation method for person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1176–1190, Mar. 2019.

[38] Y. Huang, J. Xu, Q. Wu, Z. Zheng, Z. Zhang, and J. Zhang, "Multi-pseudo regularized label for generated data in person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1391–1403, Mar. 2019.

[39] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, "Deep-person: Learning discriminative deep features for person re-identification," 2017, *arXiv:1711.10658*. [Online]. Available: http://arxiv.org/abs/1711.10658

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[42] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 17–35.

[43] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[44] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*," 2017, *arXiv:1701.07717*. [Online]. Available: http://arxiv.org/abs/1701.07717

[45] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-Reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1318–1327.

[46] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, Oct. 2007, pp. 1–7.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[48] Y. Suh, J. Wang, S. Tang, T. Mei, and K. M. Lee, "Part-aligned bilinear representations for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 402–419.

[49] X. Chang, T. M. Hospedales, and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2109–2118.

[50] Y. Shen, H. Li, T. Xiao, S. Yi, D. Chen, and X. Wang, "Deep group-shuffling random walk for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2265–2274.

[51] Y. Shen, H. Li, S. Yi, D. Chen, and X. Wang, "Person re-identification with deep similarity-guided graph neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 486–504.

[52] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3800–3808.

[53] H. Huang, D. Li, Z. Zhang, X. Chen, and K. Huang, "Adversarially occluded samples for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5098–5107.

[54] J. Xu, R. Zhao, F. Zhu, H. Wang, and W. Ouyang, "Attention-aware compositional network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2119–2128.

[55] J. Si *et al.*, "Dual attention matching network for context-aware feature sequence based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5363–5372.

[56] M. Tian *et al.*, "Eliminating background-bias for robust person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5794–5803.

[57] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, "End-to-end deep Kronecker-product matching for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6886–6895.

[58] E. Ristani and C. Tomasi, "Features for multi-target multi-camera tracking and re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6036–6046.

[59] D. Chen, D. Xu, H. Li, N. Sebe, and X. Wang, "Group consistent similarity learning via deep CRF for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8649–8658.

[60] M. M. Kalayeh, E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah, "Human semantic parsing for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1062–1071.

[61] Y. Wang, Z. Chen, F. Wu, and G. Wang, "Person re-identification with cascaded pairwise convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1470–1478.

[62] Y. Wang *et al.*, "Resource aware person re-identification across multiple resolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8042–8051.

[63] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 13001–13008.

[64] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 172–188.

[65] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 598–607.

[66] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Learning to adapt invariance in memory for person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 28, 2020, doi: 10.1109/TPAMI.2020.2976933.

[67] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[68] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
[69] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-UCSD birds-200-2011 dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep., 2011.
[70] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 554–561.
[71] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proc. Int. Conf. Multimedia Expo*, 2016, pp. 1–6.
[72] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2167–2175.
[73] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4004–4012.
[74] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1857–1865.
[75] H. O. Song, S. Jegelka, V. Rathod, and K. Murphy, "Deep metric learning via facility location," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5382–5390.
[76] Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, "No fuss distance metric learning using proxies," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 360–368.
[77] B. Harwood, V. Kumar B. G., G. Carneiro, I. Reid, and T. Drummond, "Smart mining for deep metric learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2821–2829.
[78] Y. Yuan, K. Yang, and C. Zhang, "Hard-aware deeply cascaded embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 814–823.
[79] J. Wang, F. Zhou, S. Wen, X. Liu, and Y. Lin, "Deep metric learning with angular loss," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2593–2601.
[80] C. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[81] H. Jégou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 117–128, Jan. 2011.
[82] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

Furong Xu received the B.S. degree from Northeast Normal University, Changchun, China, in 2017, and the M.S. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2020. Her research interests include machine learning and computer vision. She specially focuses on metric learning for fine-grained retrieval.

Bingpeng Ma (Member, IEEE) received the B.S. degree in mechanics and the M.S. degree in mathematics from the Huazhong University of Science and Technology, in 1998 and 2003, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, China, in 2009. He was a Postdoctoral Researcher with the University of Caen, France, from 2011 to 2012. He joined the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, in March 2013, where he is currently an Associate Professor. His research interests include computer vision, pattern recognition, and machine learning. He especially focuses on person re-identification, face recognition, and the related research topics.

Hong Chang (Member, IEEE) received the bachelor's degree from the Hebei University of Technology, Tianjin, China, in 1998, the M.S. degree from Tianjin University, Tianjin, in 2001, and the Ph.D. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2006, all in computer science. She was a Research Scientist with Xerox Research Centre Europe. She is currently a Researcher with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her main research interests include algorithms and models in machine learning, and their applications in pattern recognition and computer vision.

Shiguang Shan (Senior Member, IEEE) received the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2004. He has been a Full Professor with the Chinese Academy of Sciences since 2010, and currently the Deputy Director of the CAS Key Laboratory of Intelligent Information Processing. His research interests include computer vision, pattern recognition, and machine learning. He has published more than 300 articles, with totally more than 20,000 Google scholar citations. He served as the area chairs for many international conferences, including CVPR, ICCV, AAAI, IJCAI, ACCV, ICPR, and FG. He was/is an associate editor of several journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, *Neurocomputing*, CVIU, and PRL. He was a recipient of the China's State Natural Science Award in 2015 and the China's State S&T Progress Award in 2005, for his research work.