



Triplet Ratio Loss for Robust Person Re-identification

Shuping Hu¹, Kan Wang^{1,2}, Jun Cheng³, Huan Tan¹, and Jianxin Pang¹(✉)

¹ UBTech Robotics Corp Ltd, Shenzhen, China

{shuping.hu, kan.wang, huan.tan, walton}@ubtrobot.com

² SIAT, Chinese Academy of Sciences, Beijing, China

³ Institute for Infocomm Research, A*STAR, Singapore, Singapore
cheng_jun@i2r.a-star.edu.sg

Abstract. Triplet loss has been proven to be useful in the task of person re-identification (ReID). However, it has limitations due to the influence of large intra-pair variations and unreasonable gradients. In this paper, we propose a novel loss to reduce the influence of large intra-pair variations and improve optimization gradients via optimizing the ratio of intra-identity distance to inter-identity distance. As it also requires a triplet of pedestrian images, we call this new loss as triplet ratio loss. Experimental results on four widely used ReID benchmarks, i.e., Market-1501, DukeMTMC-ReID, CUHK03, and MSMT17, demonstrate that the triplet ratio loss outperforms the previous triplet loss.

Keywords: Person re-identification · Metric learning · Triplet ratio loss

1 Introduction

The goal of person re-identification (ReID) is to identify a person of interest using pedestrian images captured across disjoint camera views. Due to its widely deployment in real-world applications such as intelligent surveillance, ReID has become an important topic [6, 12, 14, 32, 33, 37–39, 42, 48, 52].

The key to robust ReID lies in high-quality pedestrian representations. However, due to the presence of detection errors, background occlusions and variations on poses, extracting discriminative pedestrian representations in ReID is still challenging [18, 25, 27, 31, 34]. Previous methods typically adopt two different strategies to improve the quality of pedestrian representations for ReID. The first strategy involves equipping the deep network with various modules to enhance the discriminative ability of extracted features [2, 5, 27, 30, 32]. In comparison, the second strategy leaves the network unchanged and designs loss function that directly optimizes the extracted features. Existing loss functions for ReID can be categorized into two groups: classification-based and distance-based loss functions [4, 24, 26, 28, 29], of which the triplet loss [24] is the most popular one. In brief, the triplet loss typically optimizes pedestrian features via maximizing the inter-identity distances and minimizing the intra-identity distances.

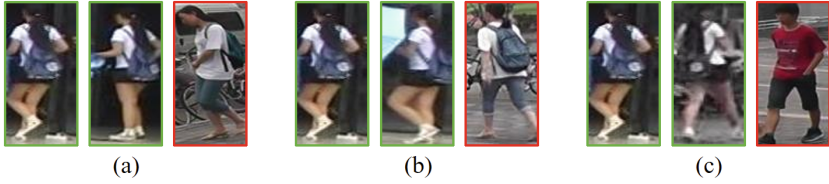


Fig. 1. Illustration of intra-pair variations. The distance of the negative pair in (a) (b) (c) is moderate, small and large, respectively. The green boxes indicate the anchor and positive images while the red boxes denote the negative images. (Color figure online)

However, the triplet loss has two inherent problems. First, the quality of the pedestrian features optimized with triplet loss is heavily affected by intra-pair variations due to the fixed margin. More specifically, first, for the three triplets presented in Fig. 1, the distance of the negative pair is moderate, small and large, respectively. This means a reasonable margin for the triplet in Fig. 1(a) is inappropriate for triplets in both Fig. 1(b) and Fig. 1(c). This is because the triplet constraint becomes too tight or loose for the two triplets, respectively. Second, the distance of the negative pair is small for triplet in Fig. 1(b); therefore triplet loss might result in a collapsed ReID model when using an improper triplet sampling strategy. From the mathematical point of view [36], this is because the triplet loss only gives slight repelling gradient¹ for hard negative image while gives large attracting gradient (see footnote 1) for hard positive image. Therefore, the embeddings of all pedestrian images will shrink to the same point.

Given the above, it is induced that an effective loss function is needed to adjust the margin according to the respective triplet and provide more reasonable gradient during the training stage. Accordingly, in this paper, we propose a novel loss called, triple ratio loss.

First, different from the triplet loss which optimizes the “difference” between intra-identity distance and inter-identity distance, in brief, the proposed triplet ratio loss directly optimizes the “ratio” of the intra-identity distance to inter-identity distance. More specifically, for a triplet, the triplet ratio loss requires the “ratio” to be smaller than a pre-defined hyper-parameter. Based on the goals and the approach, we name it as “triplet ratio loss”. Intuitively, as shown in Fig. 2(a), $\{A, P_0, N_0\}$ which denotes a triplet (A , P_0 , N_0 represents the anchor, positive, negative image in the triplet, respectively) is active but hard to be optimized for triplet loss since the intra-identity distance is larger than the decision boundary (The left boundary of the red rectangle). Besides, $\{A, P_2, N_2\}$ has no contribution to triplet loss since the intra-identity distance is already smaller than the respective decision boundary. In comparison, as illustrated in Fig. 2(b), the proposed triplet ratio loss is able to relax the tight constraint for triplet $\{A, P_0, N_0\}$ and tighten the loose constraint for triplet $\{A, P_2, N_2\}$.

¹ Repelling gradient denotes the gradient that pushes the features away from each other, while attracting gradient indicates the gradient that pulls the features closer.

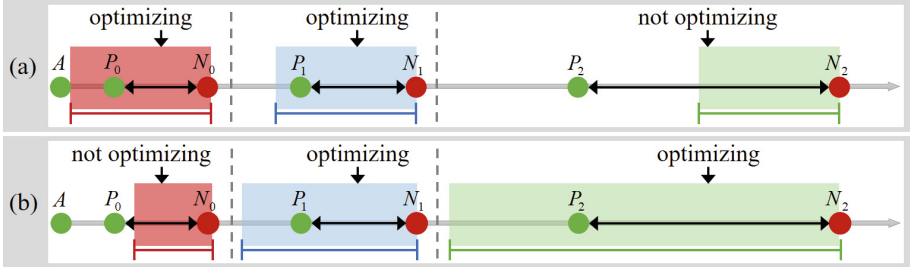


Fig. 2. During optimization, (a) For triplet loss, $\{A, P_0, N_0\}$ is active while $\{A, P_2, N_2\}$ has no contribution. (b) For triplet ratio loss, $\{A, P_2, N_2\}$ is adopted while $\{A, P_0, N_0\}$ is abandoned. The red/blue/green rectangle represents the region where positive samples need to be optimized with $N_0/N_1/N_2$. (Color figure online)

Second, the proposed triplet ratio loss improves the convergence of the ReID model. Compared with triplet loss, the triplet ratio loss drives the gradients of features in the same triplet to be adjusted adaptively in a more reasonable manner. More specifically, the triplet ratio loss adjusts the gradients for the anchor image, positive image and negative image considering both the inter-identity distance and intra-identity distance. Therefore, the triplet ratio loss gives larger repelling gradient for hard negative pair. Consequently, the proposed triplet ratio loss encourages the ReID model easier to converge and is able to prevent ReID models from shrinking to the same point [36].

The contributions of this work can be summarized as follows:

- We study the two problems that related to triplet loss for robust ReID: 1) intra-pair variations and 2) unreasonable gradients.
- We propose the triplet ratio loss to address aforementioned problems via optimizing the ratio of the intra-identity distance to inter-identity distance.
- Extensive experimental results on four widely used ReID benchmarks have demonstrated the effectiveness of the proposed triplet ratio loss.

2 Related Work

We divide existing deep learning-based ReID methods into two categories, i.e., deep feature learning or deep metric learning-based methods, according to the way adopted to improve pedestrian representations.

Deep Feature Learning-Based Methods. Methods within this category target on learning discriminative representations via designing powerful backbone. In particular, recent methods typically insert attention modules into the backbone model to enhance the representation power [2, 5]. Besides, part-level representations that sequentially perform body part detection and part-level feature extraction have been proven to be effective for ReID, as the part features contain fine-grained information [27, 30].

Deep Metric Learning-Based Methods. Methods within this category can be further categorized into two elemental learning paradigms, i.e., optimizing via class-level labels or pair-wise labels. Methods using class-level labels, e.g., the softmax loss [17, 29, 40], typically learn identity-relevant proxies to represent different identities of samples. In comparison, the methods based on pair-wise labels usually enhance the quality of pedestrian representations via explicitly optimizing the intra-identity and inter-identity distances [4, 24, 26, 29]. As one of the most popular method that using pair-wise labels, the triplet loss [24] has been proven to be effective in enlarging the inter-identity distances and improving the intra-identity compactness.

However, triplet loss has three inherent limitations [10, 24, 36]. First, the performance of triplet loss is influenced by the triplet sampling strategy. Therefore, many attempts have been made to improve the triplet selection scheme [36, 41], e.g., the distance-weighted sampling and correction. Second, the intra-identity features extracted by models are not sufficiently compact, which leads to the intra-pair variations. Accordingly, some methods [3, 11] have been proposed to constraint the relative difference between the intra-identity distance and inter-identity distance, and push the negative pair away with a minimum distance. Third, the choice of the constant margin is also vital for the learning efficiency. More specifically, too large or too small value for the margin may lead to worse performance. To this end, subsequent methods was proposed to adjust the margin based on the property of each triplet [8, 44]. However, these variants of triplet loss still methodically adopt the margin-based manner [3, 8, 11, 44].

In comparison, the proposed triplet ratio loss introduce a novel ratio-based mechanism for optimization: optimizing the ratio of intra-identity distance to inter-identity distance. Therefore, the triplet ratio loss could adaptively adjust the margin according to respective triplet so as to weaken the influence of intra-pair variations. Besides, the optimization gradients are also improved to be more reasonable by the proposed triplet ratio loss; therefore the ReID model enjoys faster convergence and more compact convergence status.

3 Method

3.1 Triplet Loss

As one of the most popular loss function for ReID, the triplet loss [24] aims at improving the quality of pedestrian representations via maximizing the inter-identity distances and minimizing the intra-identity distances. More specifically, the triplet loss is formulated as follows:

$$\mathcal{L}_{\text{triplet}} = \sum_{a,p,n \in \mathcal{N}} [D(\mathbf{f}_i^a, \mathbf{f}_i^p) - D(\mathbf{f}_i^a, \mathbf{f}_i^n) + \alpha]_+. \quad (1)$$

Here α is the margin of the triplet constraint, and \mathcal{N} indicates the set of sampled triplets. \mathbf{f}_i^a , \mathbf{f}_i^p , \mathbf{f}_i^n represent the feature representations of the anchor image, positive image and negative image within a triplet, respectively. $D(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2$

represents the distance between embedding \mathbf{x} and embedding \mathbf{y} . $[\cdot]_+ = \max(0, \cdot)$ denotes the hinge loss. During the optimization, the derivatives for each feature representations are computed as follows:

$$\frac{\partial \mathcal{L}_{\text{triplet}}}{\partial \mathbf{f}_i^a} = 2(\mathbf{f}_i^n - \mathbf{f}_i^p), \quad \frac{\partial \mathcal{L}_{\text{triplet}}}{\partial \mathbf{f}_i^p} = 2(\mathbf{f}_i^p - \mathbf{f}_i^a), \quad \frac{\partial \mathcal{L}_{\text{triplet}}}{\partial \mathbf{f}_i^n} = 2(\mathbf{f}_i^a - \mathbf{f}_i^n). \quad (2)$$

However, the triplet loss simply focuses on obtaining correct order for each sampled triplet, it therefore suffers from **large intra-pair variations** [26] and **unreasonable repelling gradient for \mathbf{f}_i^n** [36].

3.2 Triplet Ratio Loss

To address the aforementioned two drawbacks, we propose the triplet ratio loss which optimizes triplets from a novel perspective. In brief, the triplet ratio loss directly optimizes the ‘‘ratio’’ of the intra-identity distance to inter-identity distance. More specifically, the triplet ratio loss is formulated as:

$$\mathcal{L}_{\text{tri_ratio}} = \sum_{a,p,n \in \mathcal{N}} \left[\frac{D(\mathbf{f}_i^a, \mathbf{f}_i^p)}{D(\mathbf{f}_i^a, \mathbf{f}_i^n)} - \beta \right]_+, \quad (3)$$

where $\beta \in (0, 1)$ is the hyper-parameter of the triplet ratio constraint.

During the optimization, the derivatives of the triplet ratio loss with respect to \mathbf{f}_i^a , \mathbf{f}_i^p , \mathbf{f}_i^n are:

$$\begin{aligned} \frac{\partial \mathcal{L}_{\text{tri_ratio}}}{\partial \mathbf{f}_i^a} &= \frac{-2D(\mathbf{f}_i^a, \mathbf{f}_i^n) \cdot (\mathbf{f}_i^p - \mathbf{f}_i^a) - 2D(\mathbf{f}_i^a, \mathbf{f}_i^p) \cdot (\mathbf{f}_i^a - \mathbf{f}_i^n)}{[D(\mathbf{f}_i^a, \mathbf{f}_i^n)]^2}, \\ \frac{\partial \mathcal{L}_{\text{tri_ratio}}}{\partial \mathbf{f}_i^p} &= \frac{2(\mathbf{f}_i^p - \mathbf{f}_i^a)}{D(\mathbf{f}_i^a, \mathbf{f}_i^n)}, \\ \frac{\partial \mathcal{L}_{\text{tri_ratio}}}{\partial \mathbf{f}_i^n} &= \frac{2D(\mathbf{f}_i^a, \mathbf{f}_i^p) \cdot (\mathbf{f}_i^a - \mathbf{f}_i^n)}{[D(\mathbf{f}_i^a, \mathbf{f}_i^n)]^2}. \end{aligned} \quad (4)$$

- **Addressing the intra-pair variations.** Compared with triplet loss, the triplet ratio loss handles the intra-pair variations via adjusting the constraint for \mathbf{f}_i^p according to $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$. More specifically, first, the triplet ratio loss relatively relaxes the constraint on the intra-identity pair when $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$ is small. Second, it encourages the constraint on the intra-identity pair to be tighter, when $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$ is large. For example, when setting the value of α and β as 0.4¹, the constraint deployed on the intra-identity pair is adjustable for each triplet in triplet ratio loss, but rigid in triplet loss. More specifically, for the triplet in Fig. 1(b) where $D(\mathbf{f}_i^a, \mathbf{f}_i^n) = 0.5$, the triplet loss requires $D(\mathbf{f}_i^a, \mathbf{f}_i^p) \leq 0.1$; while the triplet ratio loss only requires $D(\mathbf{f}_i^a, \mathbf{f}_i^p) \leq 0.2$; therefore it **relaxes** the constraint on the intra-identity pair. Besides, for the triplet in Fig. 1(c) that $D(\mathbf{f}_i^a, \mathbf{f}_i^n) = 1$, the triplet ratio loss requires $D(\mathbf{f}_i^a, \mathbf{f}_i^p) \leq 0.4$. This constraint is **tightened** compared with that of triplet loss that requires $D(\mathbf{f}_i^a, \mathbf{f}_i^p) \leq 0.6$.

¹ 0.4 is an empirical value for both α [31] and β . Please refer to Sect. 4.2 for extensive evaluation on the value of β .

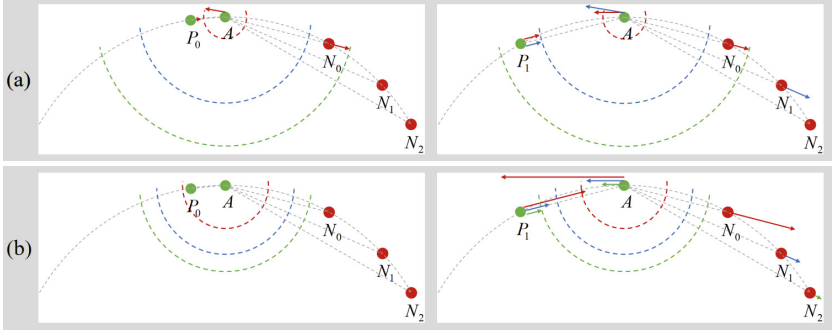


Fig. 3. Gradients for \mathbf{f}_i^a , \mathbf{f}_i^p , \mathbf{f}_i^n when respectively using (a) Triplet loss and (b) Triplet ratio loss. The green/red dot denotes the positive/negative images. The red/blue/green dotted line denotes the decision boundary for positive sample with negative sample $N_0/N_1/N_2$. The red/blue/green arrows denote the gradients of \mathbf{f}_i^a , \mathbf{f}_i^p , \mathbf{f}_i^n with negative sample $N_0/N_1/N_2$. (Color figure online)

- Addressing the gradients issue.** During the training stage, the triplet ratio loss provides \mathbf{f}_i^n a more reasonable repelling gradient. As illustrated in Eq. (4), the amplitude of repelling gradient for \mathbf{f}_i^n is inversely related to $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$. Therefore, as shown in Fig. 3, the amplitude of repelling gradient for \mathbf{f}_i^n becomes reasonably significant when $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$ is small, which is coherent to the intuition. In opposite, when $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$ is large that has almost satisfied the optimization purpose, the repelling gradient for \mathbf{f}_i^n becomes slight so that weak optimization is employed. This is intuitive since \mathbf{f}_i^n does not need much optimization in this situation. However, the triplet loss assigns counter-intuitive repelling gradient for \mathbf{f}_i^n as presented in Fig. 3.

Besides, the gradients for both \mathbf{f}_i^n and \mathbf{f}_i^p are also become more reasonable in triplet ratio loss. As illustrated in Eq. (4), gradient for \mathbf{f}_i^n is determined by both $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$ and $D(\mathbf{f}_i^a, \mathbf{f}_i^p)$, gradient for \mathbf{f}_i^p is determined by $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$. More specifically, first, the amplitude of gradient for \mathbf{f}_i^n is proportional to $D(\mathbf{f}_i^a, \mathbf{f}_i^p)$. This means that the attention on \mathbf{f}_i^n will not be significant if the intra-identity pedestrian images are similar to each other. Second, the amplitude of gradient for \mathbf{f}_i^p is inversely proportional to $D(\mathbf{f}_i^a, \mathbf{f}_i^n)$. Therefore, a triplet where the pedestrian images in the inter-identity pair are obviously dissimilar will not put emphasis on optimizing \mathbf{f}_i^p .

During the training stage, the ReID model is optimized with both cross-entropy loss and triplet ratio loss, the overall objective function can be therefore written as follows:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{tri_ratio}}. \quad (5)$$

Here \mathcal{L}_{CE} represents the cross-entropy loss, λ denotes the weight of triplet ratio loss and is empirically set to 1.

4 Experiment

We evaluate the effectiveness of the proposed triplet ratio loss on four popular ReID benchmarks, i.e., Market-1501 [45], DukeMTMC-reID [47], CUHK03 [16], and MSMT17 [35]. We follow the official evaluation protocols for these datasets and report the Rank-1 accuracy and mean Average Precision (mAP).

Market-1501 contains 12,936 images of 751 identities in the training dataset and 23,100 images of 750 identities in the test dataset. DukeMTMC-ReID consists of 16,522 training images of 702 identities and 19,889 testing images of 702 identities. CUHK03 includes 14,097 pedestrian images of 1,467 identities. The new training/testing protocol detailed in [49] is adopted. MSMT17 is divided into a training set containing 32,621 images of 1,041 identities, and a testing set comprising 93,820 images of 3,060 identities.

4.1 Implementation Details

Experiments are conducted using the PyTorch framework. During the training stage, both offline and online strategies are adopted for data augmentation [16]. The offline translation is adopted and each training set is enlarged by a factor of 5. Besides, the horizontal flipping and the random erasing [50] with a ratio of 0.5 are utilized. All images mentioned above are resized to 384×128 .

ResNet50 [9] trained with only cross-entropy is used as the baseline. ImageNet is used for pretrain. In order to sample triplets for the triplet ratio loss, we set P to 6 and A to 8 to construct a batch (whose size is therefore 48). The value of β is set as 0.4 for CUHK03 and 0.7 for the other three datasets. The standard stochastic gradient descent (SGD) optimizer with a weight decay of 5×10^{-4} and momentum of 0.9 is utilized for model optimization. All the models are trained in an end-to-end fashion for 70 epochs. The learning rate is initially set to 0.01, then multiplied by 0.1 for every 20 epochs.

4.2 Impact of the Hyper-Parameter β

In this experiment, we evaluate the performance of triplet ratio loss with different value of β on Market-1501 and CUHK03-Label. The other experimental settings are consistently kept to facilitate the clean comparison.

From the experimental results illustrated in Fig. 4, we can make the following two observations. First, the performance of the triplet ratio loss tends to be better when the value of β increases; this is because a small value of β leads to too strict constraint on intra-identity distances. Second, the performance of the triplet ratio loss drops when the value of β further increase. This is because a large value of β brings loose constraints on intra-identity distances; therefore it harms the intra-identity compactness.

4.3 Triplet Ratio Loss vs. Triplet Loss

We show the superiority of the proposed triplet ratio loss over triplet loss by comparing them from both quantitative and qualitative perspectives.

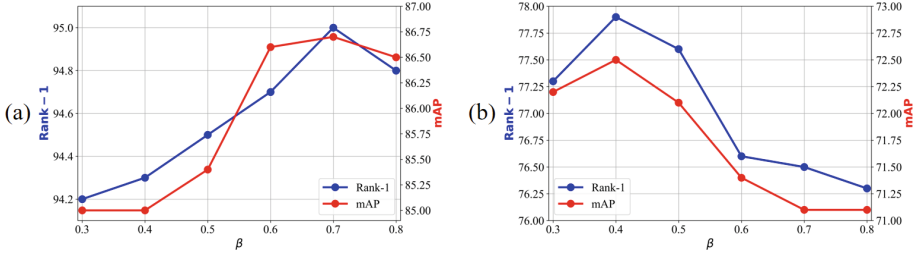


Fig. 4. The Rank-1 accuracy and mAP using different value of β on (a) Market-1501 dataset and (b) CUHK03-Label dataset.

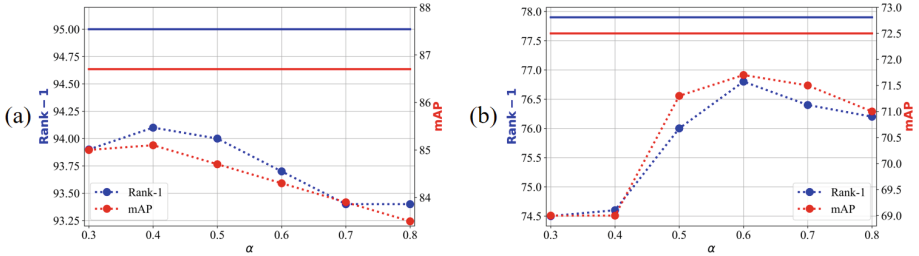


Fig. 5. The Rank-1 accuracy and mAP using different value of α on (a) Market-1501 dataset and (b) CUHK03-Label dataset. The blue and red solid lines denote the Rank-1 accuracy and mAP of using triplet ratio loss. (Color figure online)

Table 1. Evaluation of triplet ratio loss on ResNet50. R-1 in Table 1 and Table 2 denotes the Rank-1 accuracy.

Method	Market-1501		Duke		CUHK03-D		CUHK03-L		MSMT17	
	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
Baseline	91.6	81.1	84.6	71.2	63.2	59.1	67.9	63.0	68.3	41.4
w/ Triplet loss	94.1	85.1	86.8	73.5	74.1	68.3	76.8	71.7	75.4	51.4
w/ Triplet ratio loss	95.0	86.7	88.6	76.4	75.4	69.3	77.9	72.5	79.3	55.2

We compare the performance of triplet ratio loss with that of triplet loss under different value of α (the margin of triplet loss) in Fig. 5. It is concluded that the triplet ratio loss consistently outperforms triplet loss in both Rank-1 accuracy and mAP. For example, the triplet ratio loss beats the best performance of triplet loss by 0.9% in Rank-1 accuracy and 1.6% in mAP on Market-1501. The above experiments demonstrate the superiority of triplet ratio loss. Besides, the experimental results listed in Table 1 show that triplet ratio loss brings consistent performance improvement for the baseline: in particular, the Rank-1 accuracy is improved by 11.0% while mAP is also promoted by 13.8% on MSMT17. These experimental results verify the effectiveness of the proposed triplet ratio loss.

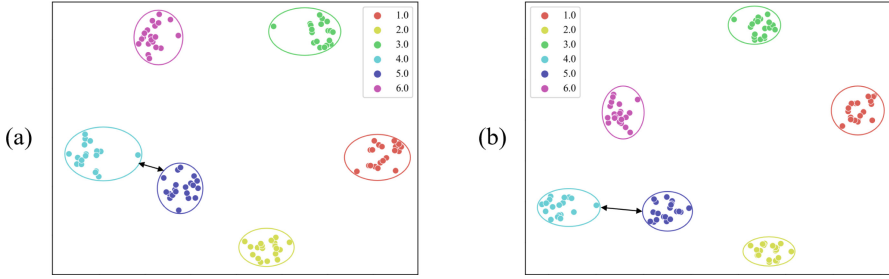


Fig. 6. Distribution of the pedestrian features extracted by ResNet-50 that trained respectively with (a) Triplet loss and (b) Triplet ratio loss.

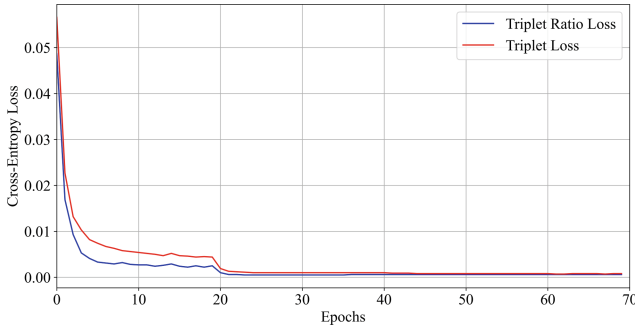


Fig. 7. The cross-entropy loss curves of the training process on Market-1501.

The conclusion is further supported by visualizing the features optimized by triplet ratio loss and triplet loss in Fig. 6, as well as the curves associated with the cross-entropy loss when adopting the two losses in Fig. 7. After assessing Fig. 6 and Fig. 7, we can make the following observations: first, the features optimized with triplet ratio loss are more compact than that learned using triplet loss; this indicates the triplet ratio loss is more effective on **addressing the intra-pair variations**. Second, the cross-entropy loss with triplet ratio loss converges faster than that with triplet loss. This is because triplet ratio loss improves optimization procedure via providing **more reasonable gradients**. The above analyses prove the superiority of triplet ratio loss.

4.4 Comparison with State-of-the-Art

We compare the proposed triplet ratio loss with state-of-the-art methods on Market-1501 [45], DukeMTMC-ReID [47], CUHK03 [16], and MSMT17 [35]. For fair comparison, we divide existing approaches into two categories, i.e., holistic feature-based (HF) methods and part feature-based (PF) methods.

Table 2. Performance comparisons on four popular benchmarks. “-” represents these results are not available.

Methods		Market-1501		Duke		CUHK03-D		CUHK03-L		MSMT17	
		R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
PF	PCB+RPP [27]	93.8	81.6	83.3	69.2	63.7	57.5	-	-	78.2	52.5
	CASN (PCB) [46]	94.4	82.8	87.7	73.7	71.5	64.4	73.7	68.0	-	-
	Auto-ReID [22]	94.5	85.1	-	-	73.3	69.3	77.9	73.0	69.8	43.6
	MHN-6 (PCB) [1]	95.1	85.0	89.1	77.2	71.7	65.4	77.2	72.4	-	-
	BAT-net [7]	95.1	87.4	87.7	77.3	76.2	73.2	78.6	76.1	79.5	56.8
	MuDeep [21]	95.3	84.7	88.2	75.6	71.9	67.2	75.6	70.5	-	-
	MGN [30]	95.7	86.9	88.7	78.4	66.8	66.0	68.0	67.4	-	-
	DSA-reID [43]	95.7	87.6	86.2	74.3	78.2	73.1	78.9	75.2	-	-
HF	SFT [19]	93.4	82.7	86.9	73.2	-	-	68.2	62.4	73.6	47.6
	PGR [15]	93.9	77.2	83.6	66.0	-	-	-	-	66.0	37.9
	IANet [13]	94.4	83.1	87.1	73.4	-	-	-	-	75.5	46.8
	CAL [23]	94.5	87.0	87.2	76.4	-	-	-	-	79.5	56.2
	OSNet [51]	94.8	84.9	88.6	73.5	72.3	67.8	-	-	78.7	52.9
	3DSL [2]	95.0	87.3	88.2	76.1	-	-	-	-	-	-
	BDB-Cut [5]	95.3	86.7	89.0	76.0	76.4	73.5	79.4	76.7	-	-
	IDE*+AM-Softmax [29]	92.4	83.8	-	-	-	-	-	-	75.6	49.3
	IDE*+Soft [‡] +Tri [†] [20]	94.1	85.7	86.2	75.9	-	-	-	-	-	-
	IDE*+Circle loss [26]	94.2	84.9	-	-	-	-	-	-	76.3	50.2
	IDE*+Soft [‡] +Ratio	95.0	86.7	88.6	76.4	75.4	69.3	77.9	72.5	79.3	55.2
	IDE*+Soft[‡]+Ratio**	95.8	93.6	91.6	88.7	83.7	83.9	85.8	85.6	83.5	71.3

† Triplet loss ‡ Softmax loss * ResNet-50 ** Re-ranking [49]

After examining the results tabulated in Table 2, we can make the following observations. First, compared with the PF-based methods, equipping ResNet-50 with the proposed triplet ratio loss achieves comparative performance though PF-based methods extract fine-grained part-level representations. Second, equipping ResNet-50 with the proposed triplet ratio loss also achieves comparative performance when compared with the state-of-the-art HF-based methods. For example, the Rank-1 accuracy of our method is the same as that of 3DSL [2], a most recent method that requires additional 3D clues, with the mAP of our method is merely lower than that of 3DSL by 0.6% on Market-1501. Third, compared with works that explore loss functions for ReID, the proposed method achieves the best performance when using the same backbone for feature extraction. Specifically, the proposed triplet ratio loss outperforms Circle loss [26] by 3.0% and 5.0% in terms of the Rank-1 accuracy and mAP, respectively, on the MSMT17 benchmark. At last, the Re-ranking [49] further promotes the performance of triplet ratio loss: the triplet ratio loss finally achieve 95.8% and 93.6%, 91.6% and 88.7%, 83.7% and 83.9%, 85.8% and 85.6%, 83.5% and 71.3% in terms of the Rank-1 accuracy and mAP, respectively, on each dataset. The above comparisons justify the effectiveness of triplet ratio loss.

5 Conclusion

In this paper, we propose a novel triplet ratio loss to address two inherent problems: 1) heavily influenced by intra-pair variations and 2) unreasonable gradients, that associated with triplet loss. More specifically, first, the triplet ratio loss directly optimizes the ratio of intra-identity distance to inter-identity distance, therefore the margin between intra-identity distance and inter-identity distance could be adaptively adjusted according to respective triplet. Second, the triplet ratio loss adjusts the optimization gradients for embeddings considering both the inter-identity distances and intra-identity distances. The experimental results on four widely used ReID benchmarks have demonstrated the effectiveness and superiority of the proposed triplet ratio loss.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant U2013601, and the Program of Guangdong Provincial Key Laboratory of Robot Localization and Navigation Technology, under Grant 2020B121202011 and Key-Area Research and Development Program of Guangdong Province, China, under Grant 2019B010154003.

References

1. Chen, B., Deng, W., Hu, J.: Mixed high-order attention network for person re-identification. In: ICCV, pp. 371–381 (2019)
2. Chen, J., et al.: Learning 3D shape feature for texture-insensitive person re-identification. In: CVPR, pp. 8146–8155 (2021)
3. 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: CVPR, pp. 1335–1344 (2016)
4. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: CVPR, pp. 539–546 (2005)
5. Dai, Z., Chen, M., Gu, X., Zhu, S., Tan, P.: Batch dropblock network for person re-identification and beyond. In: ICCV, pp. 3690–3700 (2019)
6. Ding, C., Wang, K., Wang, P., Tao, D.: Multi-task learning with coarse priors for robust part-aware person re-identification. TPAMI **44**(3), 1474–1488 (2022)
7. Fang, P., Zhou, J., Roy, S.K., Petersson, L., Harandi, M.: Bilinear attention networks for person retrieval. In: ICCV, pp. 8029–8038 (2019)
8. Ha, M.L., Blanz, V.: Deep ranking with adaptive margin triplet loss. arXiv preprint [arXiv:2107.06187](https://arxiv.org/abs/2107.06187) (2021)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
10. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint [arXiv:1703.07737](https://arxiv.org/abs/1703.07737) (2017)
11. Ho, K., Keuper, J., Pfrendt, F.J., Keuper, M.: Learning embeddings for image clustering: an empirical study of triplet loss approaches. In: ICPR, pp. 87–94 (2021)
12. Hou, R., Chang, H., Ma, B., Huang, R., Shan, S.: BiCnet-TKS: learning efficient spatial-temporal representation for video person re-identification. In: CVPR, pp. 2014–2023 (2021)

13. Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., Chen, X.: Interaction-and-aggregation network for person re-identification. In: CVPR, pp. 9317–9326 (2019)
14. Huang, T., Qu, W., Zhang, J.: Continual representation learning via auto-weighted latent embeddings on person ReID. In: PRCV, pp. 593–605 (2021)
15. Li, J., Zhang, S., Tian, Q., Wang, M., Gao, W.: Pose-guided representation learning for person re-identification. TPAMI **44**(2), 622–635 (2022)
16. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: CVPR, pp. 152–159 (2014)
17. Liu, W., Wen, Y., Yu, Z., Yang, M.: Large-margin softmax loss for convolutional neural networks. In: ICML, vol. 2, p. 7 (2016)
18. Liu, X., Yu, L., Lai, J.: Group re-identification based on single feature attention learning network (SFALN). In: PRCV, pp. 554–563 (2021)
19. Luo, C., Chen, Y., Wang, N., Zhang, Z.: Spectral feature transformation for person re-identification. In: ICCV, pp. 4975–4984 (2019)
20. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPR W (2019)
21. Qian, X., Fu, Y., Xiang, T., Jiang, Y.G., Xue, X.: Leader-based multi-scale attention deep architecture for person re-identification. TPAMI **42**(2), 371–385 (2020)
22. Quan, R., Dong, X., Wu, Y., Zhu, L., Yang, Y.: Auto-ReID: searching for a part-aware convnet for person re-identification. In: ICCV, pp. 3749–3758 (2019)
23. Rao, Y., Chen, G., Lu, J., Zhou, J.: Counterfactual attention learning for fine-grained visual categorization and re-identification. In: ICCV, pp. 1025–1034 (2021)
24. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: CVPR, pp. 815–823 (2015)
25. Shu, X., Yuan, D., Liu, Q., Liu, J.: Adaptive weight part-based convolutional network for person re-identification. Multimedia Tools Appl. **79**, 23617–23632 (2020). <https://doi.org/10.1007/s11042-020-09018-x>
26. Sun, Y., et al.: Circle loss: a unified perspective of pair similarity optimization. In: CVPR, pp. 6398–6407 (2020)
27. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). In: ECCV, pp. 480–496 (2018)
28. Tao, D., Guo, Y., Yu, B., Pang, J., Yu, Z.: Deep multi-view feature learning for person re-identification. TCSVT **28**(10), 2657–2666 (2017)
29. Wang, F., Xiang, X., Cheng, J., Yuille, A.L.: Normface: L2 hypersphere embedding for face verification. In: ACM MM, pp. 1041–1049 (2017)
30. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACM MM, pp. 274–282 (2018)
31. Wang, K., Ding, C., Maybank, S.J., Tao, D.: CDPM: convolutional deformable part models for semantically aligned person re-identification. TIP **29**, 3416–3428 (2019)
32. Wang, K., Wang, P., Ding, C., Tao, D.: Batch coherence-driven network for part-aware person re-identification. TIP **30**, 3405–3418 (2021)
33. Wang, P., Ding, C., Shao, Z., Hong, Z., Zhang, S., Tao, D.: Quality-aware part models for occluded person re-identification. arXiv preprint [arXiv:2201.00107](https://arxiv.org/abs/2201.00107) (2022)
34. Wang, W., Pei, W., Cao, Q., Liu, S., Lu, G., Tai, Y.W.: Push for center learning via orthogonalization and subspace masking for person re-identification. TIP **30**, 907–920 (2020)
35. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: CVPR, pp. 79–88 (2018)

36. Wu, C.Y., Manmatha, R., Smola, A.J., Krahenbuhl, P.: Sampling matters in deep embedding learning. In: ICCV, pp. 2840–2848 (2017)
37. Wu, Y., Lin, Y., Dong, X., Yan, Y., Ouyang, W., Yang, Y.: Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning. In: CVPR, pp. 5177–5186 (2018)
38. Ye, M., Lan, X., Leng, Q., Shen, J.: Cross-modality person re-identification via modality-aware collaborative ensemble learning. TIP **29**, 9387–9399 (2020)
39. Ye, M., Li, J., Ma, A.J., Zheng, L., Yuen, P.C.: Dynamic graph co-matching for unsupervised video-based person re-identification. TIP **28**(6), 2976–2990 (2019)
40. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint [arXiv:1411.7923](https://arxiv.org/abs/1411.7923) (2014)
41. Yu, B., Liu, T., Gong, M., Ding, C., Tao, D.: Correcting the triplet selection bias for triplet loss. In: ECCV, pp. 71–87 (2018)
42. Yu, S., et al.: Multiple domain experts collaborative learning: multi-source domain generalization for person re-identification. arXiv preprint [arXiv:2105.12355](https://arxiv.org/abs/2105.12355) (2021)
43. Zhang, Z., Lan, C., Zeng, W., Chen, Z.: Densely semantically aligned person re-identification. In: CVPR, pp. 667–676 (2019)
44. Zhao, X., Qi, H., Luo, R., Davis, L.: A weakly supervised adaptive triplet loss for deep metric learning. In: ICCV W (2019)
45. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: ICCV, pp. 1116–1124 (2015)
46. Zheng, M., Karanam, S., Wu, Z., Radke, R.J.: Re-identification with consistent attentive siamese networks. In: CVPR, pp. 5735–5744 (2019)
47. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: ICCV, pp. 3754–3762 (2017)
48. Zhong, Y., Wang, X., Zhang, S.: Robust partial matching for person search in the wild. In: CVPR, pp. 6827–6835 (2020)
49. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR, pp. 1318–1327 (2017)
50. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: AAAI, pp. 13001–13008 (2020)
51. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification. In: ICCV, pp. 3701–3711 (2019)
52. Zhou, Z., Li, Y., Gao, J., Xing, J., Li, L., Hu, W.: Anchor-free one-stage online multi-object tracking. In: PRCV, pp. 55–68 (2020)