

RA Loss: Relation-Aware Loss for Robust Person Re-identification

Kan Wang^{1,2}, Shuping Hu², Jun Cheng¹,
Jun Cheng³, Jianxin Pang^{2,*}, and Huan Tan²

¹ Guangdong Provincial Key Laboratory of Robotics and Intelligent System,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

`kan.wang0128@gmail.com, jun.cheng@siat.ac.cn`

² Guangdong Provincial Key Laboratory of Robot Localization
and Navigation Technology, UBTech Robotics Corp Ltd

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

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

Abstract. Previous relation-based losses in person re-identification (ReID) typically comprise two sequential steps: they firstly sample both positive pair and negative pair and then deploy constraints to simultaneously improve intra-identity compactness and inter-identity separability. However, existing relation-based losses usually place emphasis on exploring the relation between images and therefore consider only several pairs during each optimization. This inevitably leads to different convergence status for pairs of the same kind and brings about the intra-pair variance problem. Accordingly, we propose a novel Relation-Aware (RA) loss to address the intra-pair variance via exploring the informative relation across pairs. In brief, we introduce a macro-constraint and a micro-constraint. The macro-constraint encourages the separation of positive pair and negative pair via pushing far apart the two “centers” of the positive pair and the negative pair. The “center” of each kind of pair are obtained via averaging all the pairs of the same kind. The micro-constraint further enhances the compactness by minimizing the discrepancies among pairs of the same kind. The two constraints work cooperatively to relieve the intra-pair variance and improve the quality of pedestrians’ representation. Results of extensive experiments on three widely used ReID benchmarks, *i.e.*, Market-1501, DukeMTMC-ReID and CUHK03, demonstrate that the RA loss brings improvements over existing relation-based losses.

Keywords: Deep learning · Person re-identification · Metric learning.

1 Introduction

Person re-identification (ReID) intends to retrieve pedestrian images belonging to the same identity from viewpoints across multiple cameras. In recent years,

* Corresponding author

due to the widespread range of potential applications in video surveillance, *e.g.*, multi-camera tracking [60] and forensic search [54, 64], ReID has drawn a lot of attention from both academia and industry [8, 15, 43–45, 47, 57, 70].

The key to robust ReID lies in high-quality representation for pedestrian image. Recently, many loss functions have been proposed to improve the quality of pedestrians representation and achieve superior performance [4, 6, 13, 16, 33, 35, 41, 51, 63]. Among them, a kind of loss functions that explore the relation between pedestrian images emerge and become popular [4, 6, 13, 16, 33, 41, 27]. In common, existing relation-based loss functions typically comprise two steps. First, they sample structures from a batch of pedestrian images. The structure commonly contains two categorizes of pairs, *i.e.*, the positive pair and negative pair that comprise two pedestrian images with identical identity and different identities, respectively. Second, the losses improve both intra-identity compactness and the inter-identity separability via deploying constraints on the two kinds of pairs. For example, as one of the most widely-used relation-based loss in ReID, the triplet loss [16, 31] samples triplet from a batch of pedestrian images. Each triplet comprises one positive pair and one negative pair. Afterwards, the triplet loss enlarges intra-identity similarity while shrinking inter-identity similarity via encouraging the distance between positive pair and negative pair⁴ to be greater than a predefined threshold. Besides, the contrastive loss [6, 13] directly samples positive pairs and negative pairs from a batch of pedestrian images, and then minimizes the embedding distance of positive pairs but demands the distance of negative pairs to be consistently larger than a predefined threshold.

However, although brings performance improvements, during optimization, existing relation-based losses [16, 31] typically increase the distance between individual pairs. As a result, the consideration of only several pairs usually leads to the intra-pair variance problem: pairs of the same kind show significant variation. More specifically, as illustrated in Fig. 1(a,b,c), the positive pairs have significant variance in appearance similarity, due to viewpoint variations and inaccurate detection. Besides, appearance similarity of the negative pairs presented in Fig. 1(d,e,f) also varies, since the two pedestrians in different pairs may wear similar or dissimilar clothes. Accordingly, we propose a novel Relation-Aware (RA) loss which relieves the intra-pair variance via exploring the relation across pairs from both macro- and micro-perspectives.

Though several subsequent methods proposed to utilize information from more pairs [4, 33], the global cues from all pairs are not sufficiently explored. Accordingly, we propose the macro-constraint to improve the separation between positive pairs and negative pairs from a global perspective. In brief, the macro-constraint pushes far apart the two “centers” of the two kinds of pairs. More specifically, at first, we average the distance of all pairs of the same kind as the corresponding “center”. Afterwards, distance between the two “centers” is encouraged to be larger than a per-defined threshold. By using this method, the

⁴ In this paper, the distance **of** one pair denotes the distance between the two pedestrian images contained in this pair. In comparison, the distance **between** two pairs indicates the difference value in the two distances of the two pairs.



Fig. 1. Illustration of the challenges of intra-pair variance in ReID. (a, b, c) Three positive pairs show high, moderate and low similarity in pedestrian appearance, respectively. (d, e, f) Three negative pairs show high, moderate and low similarity in pedestrian appearance, respectively.

macro-constraint is able to explore the useful information from all pairs sampled from a batch; therefore it is more effective in improving the separation between positive pairs and negative pairs and therefore stabilizes the training procedure. Since the introduced constraint optimizes pairs from a global perspective; we therefore name it macro-constraint.

We further propose the micro-constraint to minimize the discrepancies among the positive pairs and negative pairs. As presented in Fig. 1, there exist significant intra-pair variance for both kinds of pairs. Accordingly, in order to enhance the compactness of each kind of pairs, we impose constraint on these “unqualified” pairs, which denotes these pairs that away from the corresponding “center”. More specifically, we drives these “unqualified” pairs within both kinds to be close to the corresponding “center” to a certain extent. By this way, the micro-constraint is able to explicitly alleviate the intra-pair variance; and therefore improves the generalization ability of the pedestrian representation. Besides, different with the macro-constraint which optimizes pairs from a global perspective, this constraint is imposed on individual pairs; therefore we name it micro-constraint.

In conclusion, the macro- and micro-constraints are complementary to each other and work cooperatively to relieve the intra-pair variance and therefore improve the quality of pedestrian representation. From the methodological point of view, the contributions of this work can be summarized as follows:

- First, to the best of our knowledge, this is the first attempt to study the intra-pair variance from a comprehensive perspective for robust ReID.
- Second, we propose a novel and simple-yet-effect loss named RA loss, which relieves the intra-pair variance via exploring the informative relation across pairs from both macro- and micro-perspectives.

- Third, we demonstrate the effectiveness of the proposed RA loss on three popular large-scale ReID benchmarks, *i.e.*, Market-1501[60], DukeMTMC-ReID[64] and CUHK03[22], and the results show that RA loss brings significant improvements over existing relation-based losses.

2 Related Works

2.1 Person Re-identification

Over the past few years, deep learning-based methods [8, 42, 38, 5, 18, 11, 46, 50, 49] have come to dominate in the ReID community. We here categorize existing deep learning-based ReID methods into two groups: the methods based on feature learning and the methods using metric learning, according to the manner adopted to improve the quality of pedestrian representation.

Feature learning-based ReID. Previous methods belonging to this category typically target on learning discriminative holistic representation for pedestrian image [5, 17, 23, 62, 2, 67, 53, 68, 3, 28]. However, the holistic representation-based approaches usually suffer from the overfitting problem [47, 38, 26]. Accordingly, part-level representations [8, 57, 38, 34, 37] have been adopted for robust ReID, as their features contain fine-grained information to distinguish identity. In common, existing part-level representations-based approaches usually use some auxiliary tools, *e.g.*, pose estimators [26, 59, 21], human parsing algorithms [70, 12, 19] or attention modules [45, 68, 58] to infer the body parts' regions, from which the part-level representations are subsequently extracted.

Metric learning-based ReID. According to the manner to employ supervision during the training stage, methods in this category can be further subdivided into two categories, *i.e.*, approaches that optimized with identity-level labels [51, 63, 36, 52] and methods using pair-wise labels [4, 16, 33, 41]. The former methods typically see ReID as an image classification task [63, 36] and employ classification losses to optimize the similarity between representations of pedestrian images and the identity-related weight vectors. Unfortunately, the classification losses may harm the generalization ability of pedestrian representations, as classification losses in ReID usually encourage features to overfit to the identities. However, ReID is actually a zero-shot task: the identities encountered during testing has no overlap with that in the training stage.

In order to enhance the generalization ability of pedestrian representation, subsequent works [16, 33, 41] attempted to explore the relation across pedestrian images and formulate pair-wise labels as supervision during the training stage. In common, existing relation-based losses typically comprise two steps. First, they adopt various strategies to construct loss-specific structures [16, 33, 41, 31]. Second, constraints are designed in order to improve both intra-identity compactness and inter-identity discrepancy. For example, the triplet loss [16] first samples triplet which contains a positive pair and a negative pair, then pushes away the distance of the negative pair to be larger than that of the positive pair. Moreover, quadruplet loss [4] proposed to generalize the triplet loss via

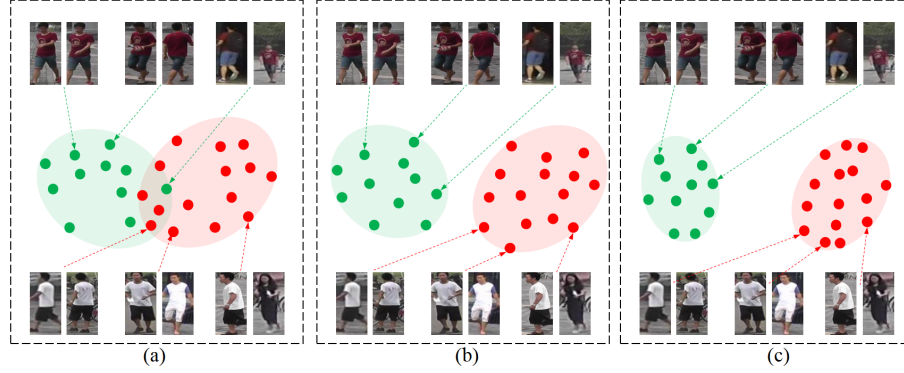


Fig. 2. Distribution of the positive pairs (green dots) and negative pairs (red dots). (a) Significant intra-pair variance for the pairs of the same kind. (b) The macro-constraint improves the separation between positive pairs and negative pairs. (c) The micro-constraint enhances the compactness of each of the two kinds of pairs. Pairs in this and the next figure are demonstrated in the two-dimensional space for better illustration. The vertical coordinates has no indication.

introducing another negative pair for better approximation of the inter-identity distance. However, although brings performance improvements, existing relation-based loss functions consider only several pairs; therefore the relation across pairs are not sufficiently explored.

In comparison, we deploy constraints on all of the positive and negative pairs from both macro- and micro-perspectives; therefore, we make use of the informative relation across pairs in a more comprehensive manner.

2.2 Intra-pair Variance

A few previous methods [35, 55, 20] have attempted to study the intra-pair variance problem. For example, [55] introduced an intra-pair loss which learns a class-independent distance metric by minimizing the intra-pair variance in both positive and negative pairs. Besides, [35] addressed the intra-pair variance via setting up a definite optimization target for all pairs of the same kind. Afterwards, it utilized the re-weighting strategy to highlight the less-optimized pairs in order to benefit the deep feature learning with flexible optimization.

We propose RA loss, which directly pushes apart the positive pair and the negative pair from a comprehensive perspective, and promotes the compactness of each kind of pairs via deploying constraint on each individual pair.

3 Method

3.1 Intra-pair Variance in ReID

Existing loss functions in ReID commonly lay emphasis on exploring the relation between pedestrian images, and only several pairs are taken into consideration

during each optimization. As a result, this type of local constraint for the relation across pairs usually brings the intra-pair variance, *i.e.*, ambiguous convergence status for pairs of the same kind.

As illustrated in Fig. 2(a), three positive pairs inherently show high, moderate and low similarity in pedestrian appearance, respectively. Moreover, existing relation-based loss functions introduce no constraint to encourage the distance of the three pairs to be consistent. Therefore, after optimized by existing relation-based loss functions, the three positive pairs may get into different convergence status: for example, the distance of them are 0.9, 0.7 and 0.3, respectively. Similarly, the distance of three negative pairs in Fig. 2(a) also obviously different and show significant intra-pair variance.

3.2 The Macro-constraint

The macro-constraint targets on improving the separation between positive pairs and negative pairs from a global perspective. To this end, the macro-constraint comprises two sequential steps to explore the informative cues from all the positive and negative pairs that sampled from a batch.

First, the macro-constraint respectively computes the “center” for both the category of positive pairs and negative pairs via averaging the distances of all the pairs within each of the two categories, as follows:

$$C_{pos} = \frac{1}{|\mathcal{P}|} \sum_{i=1}^{|\mathcal{P}|} \mathcal{D}(\mathbf{f}_i^{p_1}, \mathbf{f}_i^{p_2}), \quad (1)$$

$$C_{neg} = \frac{1}{|\mathcal{N}|} \sum_{i=1}^{|\mathcal{N}|} \mathcal{D}(\mathbf{f}_i^{n_1}, \mathbf{f}_i^{n_2}). \quad (2)$$

Here \mathcal{P} and \mathcal{N} represent the set of positive pairs and negative pairs that are sampled from the same batch, respectively. Besides, $|\mathcal{P}|$ and $|\mathcal{N}|$ denotes the number of elements in \mathcal{P} and \mathcal{N} , respectively. At last, $\mathcal{D}(\mathbf{f}_i^{p_1}, \mathbf{f}_i^{p_2})$ and $\mathcal{D}(\mathbf{f}_i^{n_1}, \mathbf{f}_i^{n_2})$ indicates the cosine distance between two pedestrian representations for the i -th positive pair $(\mathbf{f}_i^{p_1}, \mathbf{f}_i^{p_2})$ in \mathcal{P} and the i -th negative pair $(\mathbf{f}_i^{n_1}, \mathbf{f}_i^{n_2})$ in \mathcal{N} , respectively.

Second, in order to improve the separation between positive pairs and negative pairs, the macro-constraint encourages the distance between the two “centers” to be larger than a per-defined margin, as follows:

$$\mathcal{L}_{macro} = [C_{pos} - C_{neg} + \alpha]_+, \quad (3)$$

where α represents the margin of the proposed macro-constraint. Besides, $[\cdot]_+ = \max(\cdot, 0)$ denotes the hinge loss.

During each optimization, previous methods [16, 33, 41] commonly utilize the information from a part of pairs; therefore, they improve the separation between only several positive pairs and negative pairs. Compared with them, the proposed macro-constraint shows its superiority in taking full advantage of the comprehensive cues from all the pairs and improves the separation between positive pairs and negative pairs from a global perspective.

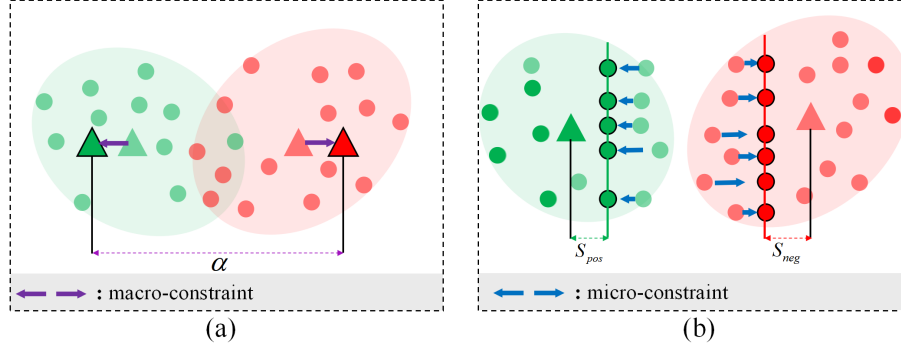


Fig. 3. (a) The macro-constraint pushes away the two “centers” of positive pairs and negative pairs. Rectangles without and with black border represent the “centers” before and after macro-constraint. (b) The micro-constraint encourages pairs within each category to be close to the corresponding decision boundary, which is denoted by the green/red vertical lines. Dots without and with black border represent the pairs before and after micro-constraint.

3.3 The Micro-constraint

We further propose the micro-constraint, which is complementary to the macro-constraint, to explicitly address the intra-pair variance. We can observe that, as presented in Fig. 1, due to the significant variance in appearance similarity, the distance of pairs within the same category varies. Accordingly, we propose the micro-constraint to explicitly minimize the discrepancies among pairs for both the category of positive pair and negative pair. As illustrated in Fig. 3(b), the micro-constraint is imposed on a part of individual pairs. More specifically, the micro-constraint consists two sequential steps.

First, we identify the pairs on which the micro-constraint should be deployed. We name these pairs as “unqualified” pairs. Intuitively, in order to address the intra-pair variance, it seems reasonable to encourage all pairs of the same kind to be close to corresponding “center”. However, the above constraint is too rigid and unreasonable. This is because the distance of different pairs are inherently inconsistent, as illustrated in Fig. 1. Besides, it is also counterintuitive to further increase distance of these positive pairs whose distance is already smaller than its corresponding “center” (C_{pos}), and decrease distance of these negative pairs whose distance is already larger than its corresponding “center” (C_{neg}).

Based on the above analysis, as presented in Fig. 3(b), the sets of “unqualified” pairs for the positive pairs and negative pairs are defined respectively as:

$$\hat{\mathcal{P}} = \{(\mathbf{f}_i^{p1}, \mathbf{f}_i^{p2}) \in \mathcal{P} : \mathcal{D}(\mathbf{f}_i^{p1}, \mathbf{f}_i^{p2}) > (C_{pos} + \beta \cdot S_{pos})\}, \quad (4)$$

$$\hat{\mathcal{N}} = \{(\mathbf{f}_i^{n1}, \mathbf{f}_i^{n2}) \in \mathcal{N} : \mathcal{D}(\mathbf{f}_i^{n1}, \mathbf{f}_i^{n2}) < (C_{neg} + \beta \cdot S_{neg})\}. \quad (5)$$

Here S_{pos} and S_{neg} denotes the standard deviation for the category of positive pair and negative pair, respectively. β is a hyper-parameter which determines

the decision boundaries of the micro-constraint. Please refer to Fig. 3(b) for a clear illustration of the decision boundaries.

Second, we encourage the pairs in $\hat{\mathcal{P}}$ and $\hat{\mathcal{N}}$ to be close to the corresponding decision boundary for better intra-pair compactness. More specifically, as illustrated in Fig. 3(b), the micro-constraint deploys constraint on the pairs in $\hat{\mathcal{P}}$ and $\hat{\mathcal{N}}$, as follows:

$$\mathcal{L}_{micro}^{pos} = \frac{1}{|\hat{\mathcal{P}}|} \sum_{i=1}^{|\hat{\mathcal{P}}|} \{ \mathcal{D}(\mathbf{f}_i^{p_1}, \mathbf{f}_i^{p_2}) - (C_{pos} + \beta \cdot S_{pos}) \}, \quad (6)$$

$$\mathcal{L}_{micro}^{neg} = \frac{1}{|\hat{\mathcal{N}}|} \sum_{i=1}^{|\hat{\mathcal{N}}|} \{ (C_{neg} + \beta \cdot S_{neg}) - \mathcal{D}(\mathbf{f}_i^{n_1}, \mathbf{f}_i^{n_2}) \}. \quad (7)$$

Based the above, the micro-constraint is summarized as follows:

$$\mathcal{L}_{micro} = \mathcal{L}_{micro}^{pos} + \mathcal{L}_{micro}^{neg}. \quad (8)$$

Finally, the proposed RA loss which comprises both macro-constraint and micro-constraint is formulated as follows:

$$\mathcal{L}_{RA} = \mathcal{L}_{macro} + \lambda_1 \mathcal{L}_{micro}. \quad (9)$$

Here λ_1 , whose value is set to 1 for the sake of simplicity, indicates the hyper-parameter which balances the macro-constraint and micro-constraint.

3.4 Person ReID via RA Loss

During the training stage, the proposed RA loss is deployed on the ReID feature extractor (*i.e.*, the PCB model [38]) and optimized simultaneously with the popular Cross-Entropy (CE) loss as well as the triplet loss. Therefore, the overall objective function can be written as:

$$\mathcal{L} = \mathcal{L}_{ID} + \lambda_2 \mathcal{L}_{TP} + \lambda_3 \mathcal{L}_{RA}. \quad (10)$$

Here \mathcal{L}_{ID} and \mathcal{L}_{TP} represents the cross-entropy loss and triplet loss, respectively. λ_2 and λ_3 are represents the weights of loss functions. For the sake of simplicity, they are consistently set to 1.

During the testing stage, the cosine metric is adopted in order to measure the similarity between two pedestrian representations \mathbf{f}_1 and \mathbf{f}_2 :

$$\rho = \frac{\mathbf{f}_1^\top \mathbf{f}_2}{\|\mathbf{f}_1\| \|\mathbf{f}_2\|}. \quad (11)$$

Here $\|\cdot\|$ indicates the L2 norm of \cdot .

4 Experiments

4.1 Datasets and Evaluation Protocols

In order to verify the effectiveness of RA loss, we perform extensive experiments on three large-scale ReID benchmarks, *i.e.*, Market-1501 [60], DukeMTMC-ReID [64] and CUHK03 [22]. The official protocol for each database is respectively followed. The widely utilized Rank-1 accuracy and mean Average Precision (mAP) are consistently adopted as metrics for performance evaluation.

Market-1501 includes images of 1,501 identities. A total of 12,936 images of 751 identities are utilized as the training set, while images of the other 750 identities are used for testing. The testing set is further split into a gallery set containing 19,732 images and a query set including the other 3,368 images.

DukeMTMC-reID comprises pedestrian images belonging to 1,404 identities. This dataset is divided into a training set containing 16,522 images belonging to 702 identities, and a testing set comprising images of the remaining 702 identities. Similar to that of Market-1501, the testing set is further subdivided into a gallery set of 17,661 images and a query set of 2,268 images.

CUHK03 contains 14,097 images of 1,467 identities in total. This dataset provides both hand-labeled and DPM-detected bounding boxes. We follow the train/test protocol proposed in [65] to split this dataset into a training set of 767 identities and a testing set of the remaining 700 identities.

4.2 Implementation Details

We use Pytorch to implement the proposed RA loss and adopt PCB [38], one of the most popular part-level representation-based approach in ReID, to extract the pedestrian representations. Besides, in this paper, we denote the PCB model which optimized with the CE loss and triplet loss as the baseline model. We implement the PCB model and the triplet loss following [47]. For the proposed RA loss, we empirically set the hyper-parameters α (in Eq. 3) and β (in Eq. 6 and Eq. 7) to 0.5 and 1, respectively.

During training, all images are resized to 384×128 pixels. The training sets of all three datasets are augmented by means of offline translation [22], online horizontal flipping, and random erasing [66] with ratio 0.5. We construct a batch as follows: each batch comprises 6 identities and each identity has 8 random sampled images; therefore the size of a batch is 48. The stochastic gradient descent optimizer with a weight decay of 5×10^{-4} and a momentum [39] value of 0.9 is utilized for model optimization. The PCB model is fine-tuned from the IDE model [61] and trained in an end-to-end manner for 70 epochs, with the learning rate is initially set to 0.01 and then multiplied by 0.1 every 20 epochs.

4.3 Ablation Study

In this subsection, we first validate the effectiveness of the macro- and micro-constraints. Afterwards, we evaluate the values of two important hyper-parameters and conclude this subsection by evaluating the universality of RA Loss.

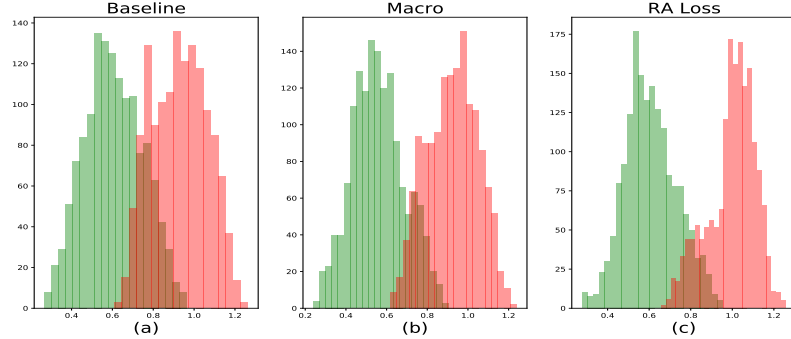


Fig. 4. Distribution of the positive pairs (green) and negative pairs (red) for: (a) baseline, (b) macro-constraint (c) RA loss. Values of the horizontal and vertical coordinates denotes the distance and number of pairs within each histogram, respectively.

Table 1. Ablation study on each key component of RA loss

Dataset	Components		Market-1501		DukeMTMC		CUHK03-D		CUHK03-L	
Metric	\mathcal{L}_{macro}	\mathcal{L}_{micro}	Rank-1 mAP		Rank-1 mAP		Rank-1 mAP		Rank-1 mAP	
Baseline	-	-	94.2	84.4	88.2	77.4	70.9	66.7	75.6	71.3
Macro	✓	-	94.9	85.7	89.0	78.8	74.1	69.3	77.2	73.8
Micro	-	✓	95.2	86.2	89.8	79.2	74.9	69.8	78.6	74.0
RA Loss	✓	✓	95.7	86.6	90.6	79.8	76.6	70.6	79.6	74.5

Effectiveness of the Macro-constraint. We equip the baseline model with the macro-constraint only to demonstrate its effectiveness. The experimental results presented in Table 1 show that the macro-constraint consistently promote the performance of baseline model on all benchmarks. For example, performance improvements of 0.7% and 1.3% can be observed on Market-1501 in terms of Rank-1 accuracy and mAP, respectively. These experimental results firmly justify the effectiveness of macro-constraint.

Effectiveness of the Micro-constraint. We then verify the effectiveness of micro-constraint in this experiment via equipping the baseline model with the micro-constraint only. The experimental results tabulated in Table 1 clearly demonstrate that micro-constraint achieves significantly superior results than those of the baseline model. For example, it improves the Rank-1 accuracy and mAP on DukeMTMC-ReID from 88.2% and 77.4% to 89.8% and 79.2%, respectively. These experimental results validate the effectiveness of micro-constraint.

Effectiveness of the RA loss. We finally equip the baseline model with both the macro- and micro-constraints. The results can be found in the row “RA Loss”

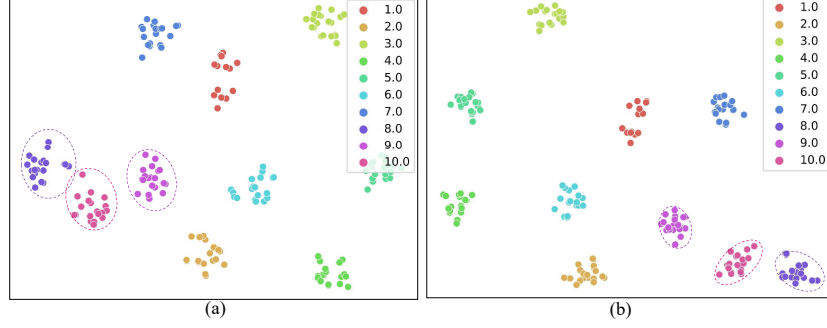


Fig. 5. Visualization using t-SNE [25] for the feature embeddings produced by (a) the baseline model and (b) the baseline model with the proposed RA loss. We sample 20 pedestrian images for each of 10 random identities, which are denoted using different colors, from the testing set of Market-1501.

of Table 1. After assessing the results, we can observe that the combination of the two components creates a considerable performance boost relative to the use of one component. Finally, the combination of the two constraints improves the performance of baseline by 1.5%, 2.4%, 5.7%, and 4.0% in terms of Rank-1 accuracy, as well as by 2.2%, 2.4%, 3.9%, and 3.2% in terms of mAP on each dataset, respectively. These results show that the macro- and micro-constraints are complementary and convincingly demonstrate the effectiveness of RA loss.

Visualization of the Distribution of Pairs. To support the above quantitative experimental results, we compare the distributions of the positive pairs and negative pairs produced by three representative models: (a) baseline, (b) baseline with macro-constraint, and (c) baseline with RA loss, in Fig. 4.

As can be seen in Fig. 4(a), the distribution of positive pairs significantly overlaps to that of the negative pairs. Moreover, we note that the compactness of both kinds of pairs is unsatisfactory. By contrast, Fig. 4(b) illustrates that the macro-constraint enlarges the separation between positive pairs and negative pairs. Moreover, as shown in Fig. 4(c), the micro-constraint further improves the compactness of both kinds of pairs. These above qualitative visualization firmly justify the effectiveness of the two components in RA loss.

Visualization of the Pedestrian Representations. In addition, in order to further justify the effectiveness of the RA loss, we visualize the pedestrian representations produced by (a) baseline and (b) baseline with RA loss.

After assessing the visualizations presented in Fig. 5, we can draw the conclusion that RA loss effectively improve the intra-identity compactness of feature embeddings as well as enhance the separability between different identities. These visualization results qualitatively justify the capability of RA loss in improving the robustness of pedestrian representations.

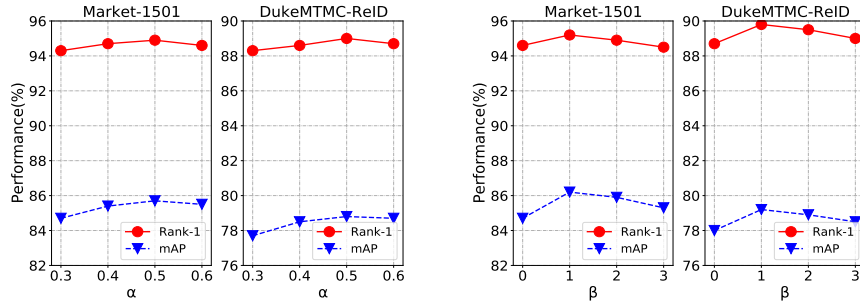


Fig. 6. Evaluation on the value of hyper-parameters α and β .

Table 2. Evaluation on the universality of RA loss

Dataset	Market-1501		DukeMTMC	
Metric	Rank-1	mAP	Rank-1	mAP
CE + Contrastive	93.1	82.9	86.8	75.4
CE + Contrastive + RA	95.2	84.5	88.7	77.4
CE + Quadruplet	94.4	84.8	88.5	77.5
CE + Quadruplet + RA	95.6	86.7	90.6	79.7

Evaluation on the Value of Hyper-Parameters α and β . In this experiment, we evaluate the performance of RA loss at different values of α and β . To facilitate clean comparison, we equip the baseline model with only the proposed macro-constraint and micro-constraint for two hyper-parameters, respectively.

We can make three observations from Fig. 6. First, the performance becomes better when α increases to a certain extent while it drops when α further increases. This is because a moderately large α is helpful for the macro-constraint to separate the positive pairs and negative pairs. Second, the performance tends better when β increases from 0 to 1. This is because the micro-constraint demands all pairs of the same kind to be consistent when β is set to 0; therefore the micro-constraint becomes too strict. Third, the performance drops when β further increases; this is because a large β lowers the effect of micro-constraint.

Evaluation on the Universality of RA Loss. We finally validate the universality of RA loss. For a clean comparison, we simply replace the triplet loss in baseline with two other popular relation-based losses, *i.e.*, contrastive loss [6, 13] and quadruplet loss [4], and then add the RA loss.

It can be seen from the experimental results in Table 2 that the RA loss consistently improve the performance of both losses. For example, it promotes the performance of two losses by 2.1% and 1.2% in terms of Rank-1 accuracy, as well as by 1.6% and 1.9% in terms of mAP on Market-1501. These results firmly verify the universality of RA loss in improving existing relation-based losses.

4.4 Comparisons with State-of-the-Art Methods

Table 3. Performance comparisons on Market-1501, DukeMTMC-ReID and CUHK03. **blue** and **Red** indicates the best results obtained by HF- and PF-based methods, respectively. Results of our methods are marked in **bold**. “-” represents these results are not available, “RR” denotes the re-ranking operation in [65].

Methods	Market-1501		DukeMTMC		CUHK03-D		CUHK03-L	
	Rank-1 mAP		Rank-1 mAP		Rank-1 mAP		Rank-1 mAP	
HF-based	PSE [30]	87.7 69.0	79.8 62.0		87.7 69.0		- -	
	DuATM [32]	91.4 76.6	81.8 64.6		- -		- -	
	SFT [24]	93.4 82.7	86.9 73.2		- -		71.7 60.8	
	BDB [7]	94.2 84.3	86.8 72.1		72.8 69.3		73.6 71.7	
	Circle loss [35]	94.2 84.9	- -		- -		- -	
	IANet [17]	94.4 83.1	87.1 73.4		- -		- -	
	Res50+NFormer[44]	94.7 87.7	87.4 74.9		- -		- -	
	ViT + DCAL [69]	94.7 87.5	89.0 80.1		- -		- -	
	CFPR [56]	94.8 87.7	87.4 74.9		- -		- -	
	OSNet [67]	94.8 84.9	88.6 73.5		72.3 67.8		- -	
	TransReID [15]	94.9 88.1	90.2 81.3		- -		- -	
	BDB-Cut [7]	95.3 86.7	89.0 76.0		76.4 73.5		79.4 76.7	
	AAFormer [71]	95.6 87.7	90.1 80.0		77.6 74.8		79.9 77.8	
PF-based	HA-CNN [23]	91.2 75.7	80.5 63.8		41.7 38.6		44.4 41.0	
	PCB [38]	92.3 77.4	81.7 66.1		61.3 54.2		- -	
	PCB+RPP [38]	93.8 81.6	83.3 69.2		63.7 57.5		- -	
	HPM [10]	94.2 82.7	86.6 74.3		63.9 57.5		- -	
	Auto-ReID [29]	94.5 85.1	88.7 78.4		73.3 69.3		77.9 73.0	
	BIN [40]	94.8 87.2	89.4 79.6		72.6 69.8		74.3 72.4	
	BAT-net [9]	95.1 81.4	87.7 77.3		76.2 73.2		78.6 76.1	
	MHN-6[2]	95.1 85.0	- -		71.7 65.4		77.2 72.4	
	DSLNet [48]	95.1 87.3	90.4 78.5		76.3 72.4		- -	
	AdaMine [1]	95.2 85.9	89.9 79.0		- -		- -	
	CDPM [45]	95.2 86.0	88.2 77.5		71.9 67.0		75.8 71.1	
	MuDeep [28]	95.3 84.7	88.2 75.6		71.9 67.2		75.6 70.5	
	FPR [14]	95.4 86.6	88.6 78.4		76.1 72.3		- -	
	MGN [43]	95.7 86.9	88.7 78.4		66.8 66.0		68.0 67.4	
	DSA-reID [57]	95.7 87.6	86.2 74.3		78.2 73.1		78.9 75.2	
	RA Loss	95.7 86.6	90.6 79.8		76.6 70.6		79.6 74.5	
	RA Loss + RR	96.3 94.1	93.1 90.9		85.1 85.2		87.9 88.2	

We here compare the performance of RA loss with that of state-of-the-art methods on three ReID benchmarks: namely, Market-1501 [60], DukeMTMC-ReID [64] and CUHK03 [22]. Moreover, to facilitate fair comparison, we divide existing approaches into two categories: holistic feature-based (HF) methods and part feature-based (PF) methods, according to the properties of the pedestrian representation. The comparisons are tabulated in Table 3.

First, the performance of RA loss is comparable with that of the state-of-the-art methods on **Market-1501**. For example, when compared TransReID [15] and AAFormer [71], RA loss outperforms them by 0.8% (95.7% - 94.9%) and 0.1% (95.7% - 95.6%) for Rank-1 accuracy, respectively. Moreover, compared with the two methods, RA loss shows superiority in efficiency by using a much simpler feature extractor (*i.e.*, PCB). Besides, RA loss also surpasses Circle Loss [35], one of the most recent loss for ReID, by a significant margin in terms of both Rank-1 accuracy (95.7% vs 94.2%) and mAP (86.6% vs 84.9%). At last, we note that the Re-ranking [65] further promotes the performance of RA loss to 96.3% and 94.1% for Rank-1 accuracy and mAP, respectively. These results firmly validate the effectiveness of RA loss. Second, the RA loss consistently beats all PF-based methods on both Rank-1 accuracy and mAP on **DukeMTMC-ReID**. For example, RA loss suppresses DSLNet [48], one of the most recent methods, by 0.2% (90.6% - 90.4%) for Rank-1 accuracy and 1.3% (79.8% - 78.5%) for mAP. Moreover, the Rank-1 accuracy obtained by the RA loss is also superior to that scored by all HF-based methods, particularly, including two recent transformer-based methods, *i.e.*, TransReID [15] and AAFormer [71]. These comparisons clearly justify the overall effectiveness of RA loss. Third, the performance scored by RA loss on **CUHK03** are comparable to that obtained by other PF-based methods: in particular, RA loss achieves the best Rank-1 accuracy of 79.6% on the CUHK03-Labeled dataset, beating DSA-reID [57], the second best PF-based method, by 0.7% (79.6% - 78.9%). These above comparisons convincingly demonstrate the effectiveness of RA loss.

5 Conclusion

In this work, we propose a novel and simple-yet-effective loss function, named RA loss, which addresses the intra-pair variance for robust ReID. The proposed RA loss constructs a pair of constraints to explore the relation across pairs. First, the macro-constraint improves the discrepancy between positive pairs and negative pairs. Second, the micro-constraint promotes the compactness for both the positive pairs and negative pairs. The macro- and micro-constraints are complementary to each other and work collaboratively to address the intra-pair variance; it therefore improves the quality of pedestrian representation. We conduct extensive experiments on three popular large-scale ReID benchmarks, thereby demonstrating the effectiveness of the proposed RA loss.

Acknowledgements This research is supported by the National Natural Science Foundation of China (U2013601), and Key-Area Research and Development Program of Guangdong Province, China (2019B010154003), and the Program of Guangdong Provincial Key Laboratory of Robot Localization and Navigation Technology (2020B121202011), and the Natural Science Foundation of China (U21A20487), and Shenzhen Technology Project (JCYJ20180507182610734, KCXFZ20201221173411032, Y795001001), and CAS Key Technology Talent Program, and Guangdong Technology Project (No. 2016B010125003).

References

1. Carvalho, M., Cadene, R., Picard, D., Soulier, L., Thome, N., Cord, M.: Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 35–44 (2018)
2. Chen, B., Deng, W., Hu, J.: Mixed high-order attention network for person re-identification. In: ICCV. pp. 371–381 (2019)
3. Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., Ren, Z., Wang, Z.: Abd-net: Attentive but diverse person re-identification. In: ICCV. pp. 8350–8360 (2019)
4. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: CVPR. pp. 403–412 (2017)
5. Chen, X., Fu, C., Zhao, Y., Zheng, F., Song, J., Ji, R., Yang, Y.: Saliency-guided cascaded suppression network for person re-identification. In: CVPR. pp. 3300–3310 (2020)
6. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: CVPR. vol. 1, pp. 539–546 (2005)
7. 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)
8. Ding, C., Wang, K., Wang, P., Tao, D.: Multi-task learning with coarse priors for robust part-aware person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* (2020)
9. Fang, P., Zhou, J., Roy, S.K., Petersson, L., Harandi, M.: Bilinear attention networks for person retrieval. In: ICCV. pp. 8029–8038 (2019)
10. Fu, Y., Wei, Y., Zhou, Y., Shi, H., Huang, G., Wang, X., Yao, Z., Huang, T.: Horizontal pyramid matching for person re-identification. In: AAAI. pp. 8295–8302 (2019)
11. Gu, X., Ma, B., Chang, H., Shan, S., Chen, X.: Temporal knowledge propagation for image-to-video person re-identification. In: ICCV. pp. 9647–9656 (2019)
12. Guo, J., Yuan, Y., Huang, L., Zhang, C., Yao, J.G., Han, K.: Beyond human parts: Dual part-aligned representations for person re-identification. In: ICCV. pp. 3641–3650 (2019)
13. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: CVPR. vol. 2, pp. 1735–1742 (2006)
14. He, L., Wang, Y., Liu, W., Zhao, H., Sun, Z., Feng, J.: Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In: ICCV. pp. 8449–8458 (2019)
15. He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W.: Transreid: Transformer-based object re-identification. In: ICCV. pp. 15013–15022 (2021)
16. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017)
17. 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)
18. Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., Chen, X.: Vrsrc: Occlusion-free video person re-identification. In: CVPR. pp. 7183–7192 (2019)
19. Kalayeh, M.M., Basaran, E., Gökmen, M., Kamasak, M.E., Shah, M.: Human semantic parsing for person re-identification. In: CVPR. pp. 1062–1071 (2018)
20. Kim, S., Kim, D., Cho, M., Kwak, S.: Proxy anchor loss for deep metric learning. In: CVPR. pp. 3238–3247 (2020)

21. Li, J., Zhang, S., Tian, Q., Wang, M., Gao, W.: Pose-guided representation learning for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* (2019)
22. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: *CVPR*. pp. 152–159 (2014)
23. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: *CVPR*. pp. 2285–2294 (2018)
24. Luo, C., Chen, Y., Wang, N., Zhang, Z.: Spectral feature transformation for person re-identification. In: *ICCV*. pp. 4975–4984 (2019)
25. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. *JMLR* pp. 2579–2605 (2008)
26. Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y.: Pose-guided feature alignment for occluded person re-identification. In: *ICCV*. pp. 542–551 (2019)
27. Nguyen, B., De Baets, B.: Kernel distance metric learning using pairwise constraints for person re-identification. *IEEE Trans. Image Process.* **28**(2), 589–600 (2019)
28. Qian, X., Fu, Y., Xiang, T., Jiang, Y.G., Xue, X.: Leader-based multi-scale attention deep architecture for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 371–385 (2020)
29. 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)
30. Saquib Sarfraz, M., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: *CVPR*. pp. 420–429 (2018)
31. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: *CVPR*. pp. 815–823 (2015)
32. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: *CVPR*. pp. 5363–5372 (2018)
33. Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective. *NIPS* **29** (2016)
34. Suh, Y., Wang, J., Tang, S., Mei, T., Lee, K.M.: Part-aligned bilinear representations for person re-identification. In: *ECCV*. pp. 402–419 (2018)
35. Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z., Wei, Y.: Circle loss: A unified perspective of pair similarity optimization. In: *CVPR*. pp. 6398–6407 (2020)
36. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval. In: *ICCV*. pp. 3800–3808 (2017)
37. Sun, Y., Zheng, L., Li, Y., Yang, Y., Tian, Q., Wang, S.: Learning part-based convolutional features for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* (2019)
38. 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)
39. Sutskever, I., Martens, J., Dahl, G.E., Hinton, G.E.: On the importance of initialization and momentum in deep learning. *ICML* pp. 1139–1147 (2013)
40. Tang, Z., Huang, J.: Branch interaction network for person re-identification. In: *ACCV* (2020)
41. Tang, Z., Huang, J.: Harmonious multi-branch network for person re-identification with harder triplet loss. *ACM Trans. Multimedia Comput. Commun. Appl.* **18**(4), 1–21 (2022)

42. Tao, D., Guo, Y., Yu, B., Pang, J., Yu, Z.: Deep multi-view feature learning for person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology* **28**(10), 2657–2666 (2017)
43. 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)
44. Wang, H., Shen, J., Yongtuo, L., Gao, Y., Gavves, E.: Nformer: Robust person re-identification with neighbor transformer. In: *CVPR*. pp. 7297–7307 (2022)
45. Wang, K., Ding, C., Maybank, S.J., Tao, D.: Cdpm: Convolutional deformable part models for semantically aligned person re-identification. *IEEE Trans. Image Process.* **29**, 3416–3428 (2020)
46. Wang, K., Ding, C., Pang, J., Xu, X.: Context sensing attention network for video-based person re-identification. *arXiv preprint arXiv:2207.02631* (2022)
47. Wang, K., Wang, P., Ding, C., Tao, D.: Batch coherence-driven network for part-aware person re-identification. *IEEE Trans. Image Process.* **30**, 3405–3418 (2021)
48. Wang, L., Fan, B., Guo, Z., Zhao, Y., Zhang, R., Li, R., Gong, W.: Dense-scale feature learning in person re-identification. In: *ACCV* (2020)
49. Wang, P., Ding, C., Shao, Z., Hong, Z., Zhang, S., Tao, D.: Quality-aware part models for occluded person re-identification. *IEEE Transactions on Multimedia* (2022)
50. Wang, P., Ding, C., Tan, W., Gong, M., Jia, K., Tao, D.: Uncertainty-aware clustering for unsupervised domain adaptive object re-identification. *IEEE Transactions on Multimedia* (2022)
51. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: *ECCV*. pp. 499–515. Springer (2016)
52. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A comprehensive study on center loss for deep face recognition. *IJCV* **127**(6), 668–683 (2019)
53. Xia, B.N., Gong, Y., Zhang, Y., Poellabauer, C.: Second-order non-local attention networks for person re-identification. In: *ICCV*. pp. 3759–3768 (2019)
54. Yao, H., Zhang, S., Hong, R., Zhang, Y., Xu, C., Tian, Q.: Deep representation learning with part loss for person re-identification. *IEEE Trans. Image Process.* **28**(6), 2860–2871 (2019)
55. Yu, B., Tao, D.: Deep metric learning with triplet margin loss. In: *ICCV*. pp. 6490–6499 (2019)
56. Zhang, A., Gao, Y., Niu, Y., Liu, W., Zhou, Y.: Coarse-to-fine person re-identification with auxiliary-domain classification and second-order information bottleneck. In: *CVPR*. pp. 598–607 (2021)
57. Zhang, Z., Lan, C., Zeng, W., Chen, Z.: Densely semantically aligned person re-identification. In: *CVPR*. pp. 667–676 (2019)
58. Zhao, L., Li, X., Zhuang, Y., Wang, J.: Deeply-learned part-aligned representations for person re-identification. In: *ICCV*. pp. 3219–3228 (2017)
59. Zheng, L., Huang, Y., Lu, H., Yang, Y.: Pose invariant embedding for deep person re-identification. *IEEE Trans. Image Process.* **28**(9), 4500–4509 (2019)
60. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: *ICCV*. pp. 1116–1124 (2015)
61. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., Tian, Q.: Person re-identification in the wild. In: *ICCV*. pp. 1367–1376 (2017)
62. Zheng, M., Karanam, S., Wu, Z., Radke, R.J.: Re-identification with consistent attentive siamese networks. In: *CVPR*. pp. 5735–5744 (2019)

63. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned cnn embedding for person reidentification. *ACM Trans. Multimedia Comput. Commun. Appl.* **14**(1), 1–20 (2017)
64. 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)
65. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: *CVPR*. pp. 1318–1327 (2017)
66. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. pp. 13001–13008 (2020)
67. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification. In: *ICCV*. pp. 3701–3711 (2019)
68. Zhou, S., Wang, F., Huang, Z., Wang, J.: Discriminative feature learning with consistent attention regularization for person re-identification. In: *ICCV*. pp. 8039–8048 (2019)
69. Zhu, H., Ke, W., Li, D., Liu, J., Tian, L., Shan, Y.: Dual cross-attention learning for fine-grained visual categorization and object re-identification. In: *CVPR*. pp. 4692–4702 (2022)
70. Zhu, K., Guo, H., Liu, Z., Tang, M., Wang, J.: Identity-guided human semantic parsing for person re-identification. In: *ECCV* (2020)
71. Zhu, K., Guo, H., Zhang, S., Wang, Y., Huang, G., Qiao, H., Liu, J., Wang, J., Tang, M.: Aaformer: Auto-aligned transformer for person re-identification. *arXiv preprint arXiv:2104.00921* (2021)