Hindawi
Complexity
Volume 2020, Article ID 9860562, 11 pages
https://doi.org/10.1155/2020/9860562

Research Article

Circle-Based Ratio Loss for Person Reidentification

Zhao Yang,[1,2] Jiehao Liu,[1,2] Tie Liu,[1,2] Li Wang,[1,2] and Sai Zhao[1,2]

[1] *School of Electronics and Communication Engineering, Guangzhou University, Guangzhou, China*
[2] *Huangpu Research & Graduate School of Guangzhou University, Guangzhou, China*

Correspondence should be addressed to Zhao Yang; yangdxng100@126.com

Received 27 July 2020; Revised 3 December 2020; Accepted 7 December 2020; Published 29 December 2020

Academic Editor: Atif Khan

Person reidentification (re-id) aims to recognize a specific pedestrian from uncrossed surveillance camera views. Most re-id methods perform the retrieval task by comparing the similarity of pedestrian features extracted from deep learning models. Therefore, learning a discriminative feature is critical for person reidentification. Many works supervise the model learning with one or more loss functions to obtain the discriminability of features. Softmax loss is one of the widely used loss functions in re-id. However, traditional softmax loss inherently focuses on the feature separability and fails to consider the compactness of within-class features. To further improve the accuracy of re-id, many efforts are conducted to shrink within-class discrepancy as well as between-class similarity. In this paper, we propose a circle-based ratio loss for person re-identification. Concretely, we normalize the learned features and classification weights to map these vectors in the hypersphere. Then we take the ratio of the maximal intraclass distance and the minimal interclass distance as an objective loss, so the between-class separability and within-class compactness can be optimized simultaneously during the training stage. Finally, with the joint training of an improved softmax loss and the ratio loss, the deep model could mine discriminative pedestrian information and learn robust features for the re-id task. Comprehensive experiments on three re-id benchmark datasets are carried out to illustrate the effectiveness of the proposed method. Specially, 83.12% mAP on Market-1501, 71.66% mAP on DukeMTMC-reID, and 66.26%/63.24% mAP on CUHK03 labeled/detected are achieved, respectively.

1. Introduction

Person reidentification aims to retrieve the person-of-interest among nonoverlapping camera views according to the given person image. Re-id is an important terminal application technology in the modern intelligent monitoring system, and it becomes gradually significant in the field of public security. However, due to the limitation of work environments and camera devices, the captured images usually have vast differences in illuminations, occlusions, person postures, camera views, etc. These differences would bring about huge variances for different images of a certain pedestrian and degrade the overall re-id performance.

Traditional re-id approaches tackle the aforementioned problems mainly with manual feature representation [1, 2] and metric learning [3, 4] methods. With the rapid development of neural networks and the popularization of large-scale re-id datasets in recent years, the deep learning based methods have been widely applied in person reidentification and obtained remarkable performance. Moreover, the deep learning based approaches can integrate the feature learning and metric learning in an end-to-end framework. Due to various advantages, the deep learning approaches have dominated the research trends of person reidentification.

Person reidentification methods based on deep learning commonly contain two essential parts: network architecture and loss function. Network architecture is generally constructed from convolutional neural networks (CNNs) which are concatenated organically by various network layers, e.g., convolutional layer, pooling layer, and fully connected layer. The designed network architecture can automatically extract pedestrian features from input images. Loss function is used to supervise model training with a predefined constraint objective. According to different constraint objectives, loss functions can be usually divided into two categories: classification loss [5–7] and metric loss [8–10]. In the training stage, classification loss encourages the model to learn the features with label information so the obtained features have well characteristics in the between-class separability. Instead of focusing on the label information exclusively, metric loss takes the feature similarity of different pedestrian images as the constraint objective to guide the model training. In this way, the learned features have distinguishable distribution in the feature space.

Admittedly, the deep model based on convolutional neural networks (CNNs) is able to extract highly abstract pedestrian features, and the large-scale re-id datasets make it possible to tackle re-id tasks with the deep learning methods. Nevertheless, the large-scale datasets with significant changes in illuminations, resolutions, background occlusions, and camera views would bring some great difficulties in the model training, e.g., a huge intraclass gap. Besides, the deep model guided by the traditional classification loss such as softmax loss is hard to fully mine the discriminative pedestrian information. It will make the learned model become susceptible to those adverse variations and cause a lack of generalization ability.

Thus, it becomes critical for re-id task to learn discriminative features which are robust to those adverse variations. To this end, both the within-class similarity and between-class discrepancy of learned features should be as large as possible. One practicable solution is improving or designing loss functions to make it effectively encourage intraclass compactness and interclass separability.

In this paper, we propose a new loss function named as circle-based ratio loss to improve the discriminative ability of learned features. Motivated by Linear Discriminant Analysis (LDA) which seeks for a new subspace where samples have the largest interclass distance and the smallest intraclass distance by optimizing the ratio of these two distances, we take the ratio of the maximal intraclass distance and the minimal interclass distance as a constraint objective in the re-id task. In specific, we first normalize the learned features and classification weights to project these vectors into the hypersphere. After that, we take the distance between a feature and its corresponding classification weight as the intraclass distance and the distance between different classification weights as the interclass distance. Finally, the largest intraclass distance and the smallest interclass distance are selected to formulate the circle-based ratio loss. By minimizing the ratio loss, the between-class similarity and within-class discrepancy could be shrunk simultaneously, and finally the discriminability of learned features would be improved. The diagrammatic explanation of the proposed ratio loss is shown in Figure 1. We use the dots and solid lines in different colors to represent the features and its classification weights of different classes, respectively. Under the supervision of the proposed ratio loss, the variance within a class will decrease and the discrepancy between classes will expand; hence the learned features will be discriminative.

The rest of this paper is organized as follows: Section 2 introduces the works related to our approach. Section 3 gives an elaborate description of our proposed ratio loss. Section 4 provides comprehensive re-id experiments to demonstrate the effectiveness of our method. Section 5 further discusses the effects of the parameters in ratio loss and the relationship between our method and some similar works. The conclusion is drawn in Section 6.

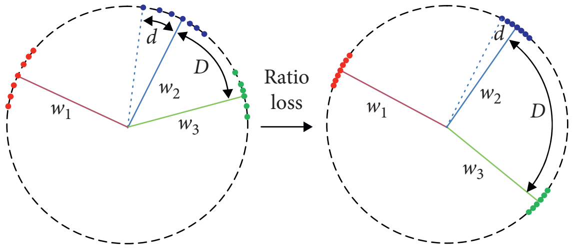

Figure 1: The diagrammatic explanation of the circle-based ratio loss. For simplicity, the feature dimension is set as 2 so the hypersphere can be represented as a circle on the 2D plane. The features and corresponding classification weights of three classes are denoted with the dots and solid lines in three different colors, respectively. d and D represent the maximal intraclass distance and the minimal interclass distance, respectively. By minimizing the ratio of d and D, not only the intraclass distance of the blue class is contracted but also the interclass distance of the blue class and green class is enlarged. The ratio loss can improve the distribution of feature space effectively and help the model learn discriminative features.

2. Related Work

The approaches of addressing person reidentification problems have been widely researched for traditional machine learning methods [1–4, 11], and lots of deep learning frameworks [6–8, 12–15] have been increasingly studied in recent years. Traditional machine learning methods tackle re-id problems mainly from two aspects: manual feature representation and metric learning. The methods of manual feature representation describe the individual image with a feature vector returned from elaborate descriptors. The descriptors will generate specific person features by considering different intrinsic information, e.g., color distribution [1] and texture description [11], and some works would combine multiple features, like LBP and HOG [2] and HSV and SILTP [4]. The metric learning methods seek a well separable metric space for pedestrian features. The most widely used metric learning methods for person reidentification contain KISSME [3], XQDA [4], and so on.

Benefited from the development of the neural network, the deep learning based re-id methods have been widely researched in recent years. They can integrate feature learning and metric learning in an end-to-end framework and achieve remarkable re-id performance. These deep methods commonly contain two essential components: network architecture and loss function. The network architecture of re-id usually comprises a CNN backbone network such as ResNet [16] or GoogleNet [17] and some customized network layers like the pooling layer, batch normalization layer, and L2 normalization layer. The backbone is usually trimmed to extract highly abstract features and some customized network layers are added to meet the requirements of re-id tasks. Besides, various loss functions are used to supervise the model learning during the training process. In certain conditions, loss function has a critical effect on re-id performance. In most existing works of re-id, it can be divided into metric loss [8–10, 13, 15, 18] and classification loss [5–7, 12, 14, 19–21].

The metric loss optimizes the model by considering the similarity of different features. To help the model learn a discriminative feature, the metric loss enlarges the separability of between-class features and promotes the compactness of within-class features. An intuitive metric loss is contrastive loss [9]. Given a pair of images, contrastive loss optimizes the model by reducing the intraclass distance and enlarging the interclass distance which is bigger than a predefined margin. For example, Varior et al. [10] performed a re-id task using contrastive loss in a gated Siamese CNN. Instead of introducing a direct distance constraint between a pair of images, triplet loss [15] constrains a relative relationship between a negative pair and a positive pair. In each iteration, triplet loss makes the distance difference between the negative pair and the positive pair larger than a margin. It is experimentally proved that triplet loss is feasible and effective for re-id tasks. For example, Cheng et al. [18] trained a multichannel parts-based CNN model combined with triplet loss for re-id. Hermans et al. [13] proposed an improved triplet loss by introducing hard sample mining, since they found that those hard triplets contribute more discriminative information in the model optimization. Moreover, Chen et al. [8] proposed the quadruplet loss to enhance the model generalization ability. Admittedly, the metric loss methods obtain outstanding performance in re-id. However, it pays too much attention to the distance information; thus the inherent label information is inevitably less concerned.

Instead of considering the feature similarity of different images, classification loss (ID loss) guides the model to distinguish different individuals according to the label information. A typical ID loss of person reidentification is softmax loss which includes a softmax activation and a cross-entropy loss function. The softmax activation converts an extracted feature into a vector whose elements indicate the possibility that the current sample belongs to a certain class. To learn a correct classification, the cross-entropy loss is used to measure the difference between the estimated probability and the truth label information. So by minimizing softmax loss, the model can progressively learn a correct classification. Zheng et al. [7] applied classification loss to train a network based on ResNet-50 for re-id. Besides, to fully exploit label information, Sun et al. [6] and Wang et al. [20] proposed the PCB and the MGN, respectively, to mine partial information of the pedestrians using classification loss. For better classification effects, many improved versions of softmax loss [5, 14, 19, 21] are proposed. Fan et al. [12] used a modified softmax function and proposed the SphereReID model for person reidentification. Witnessing the excellent performance that metric loss and classification loss have obtained in re-id tasks, some works of literatures [22–25] proposed to train the model by combining metric loss and classification loss and also achieved preferable performance in person reidentification tasks.

By considering that our proposed method in this paper is closely related to loss function, we only give a rough introduction of person reidentification methods based on loss function. It is worth noting that many other inspiring methods have been proposed to address the person reidentification tasks, e.g., pose-guided methods [26], cross-modality based methods [27], and unsupervised learning based methods [28]. One can learn about more detailed information in [29].

3. Methods

In this section, we first review softmax loss which is widely employed in deep learning frameworks of re-id and then introduce its improved version used in our approach. After that, we detail the proposed circle-based ratio loss. Finally, we demonstrate the effectiveness of our method via a toy experiment based on MNIST dataset.

3.1. Normalized Softmax Loss. Softmax loss consists of a softmax activation function and a cross-entropy loss function. The softmax activation function interprets the classification output of the linear layer as the relevant class probability, while the cross-entropy loss function quantifies the distance between calculated classification probability and ground truth label. A typical formulation of the softmax loss function can be expressed as

$$L_s = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{w_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{C} e^{w_j^T f_i + b_j}}, \tag{1}$$

where f_i is the feature extracted from the i-th selected person image in a minibatch of the training set. w_j is the j-th column weight vector of the final linear layer, also called classification weight. b_j is a bias term. y_i is the ground truth label of i-th selected person image. C and n represent the class number of the training set and the sample number in each iteration, respectively. With the optimization under softmax loss, the learned features are equipped with separable characteristics. However, the original softmax loss focuses on the between-class comparison; thus the within-class compactness of learned features is less noticed.

To tackle the mentioned defect, sorts of improvements [5, 14, 19, 21] are conducted on softmax loss. One simple but effective improvement is normalizing both classification weights and features to map these vectors into the hypersphere. In this way, the learned features are more angularly separable in the feature space. Softmax loss with the normalization is benefit to the feature learning [21], and it can be expressed as follows:

$$\begin{aligned} L_{NSL} &= -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{\left\|w_{y_i}\right\|\left\|f_i\right\|\cos\left(\theta_{y_i}\right)}}{\sum_{j=1}^{C} e^{\left\|w_j\right\|\left\|f_i\right\|\cos\left(\theta_j\right)}} \\ &= -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{s\cdot\cos\left(\theta_{y_i}\right)}}{\sum_{j=1}^{C} e^{s\cdot\cos\left(\theta_j\right)}}, \end{aligned} \tag{2}$$

where s denotes a scaling factor. It is noteworthy that a margin term is often added to obtain a more powerful constraint on the interclass and intraclass distance in many research works. In this paper, we use the improved softmax loss expressed in (2) as the classification loss and name it as normalized softmax loss [21, 30] to distinguish from the original softmax loss function.

3.2. Circle-Based Ratio Loss. As classification weights and features are both normalized, the magnitude variations are eliminated and the learned features are angularly dependent in the hypersphere. Thus the similarity between features can be directly measured with their cosine distances. In the task of person reidentification, the extracted features should have enough discrimination. It means that the between-class discrepancy should be as large as possible while the within-class compactness should be as tight as possible. Inspired by LDA, we formulate our loss function with a ratio of the maximal intraclass distance and the minimal interclass distance. Since the features and classification weights have been normalized, the learned features and classification weights spread out on a circle. Therefore, we name the proposed loss as circle-based ratio loss, and its mathematical expression is

$$L_{\text{Ratio}} = \frac{1}{C}\sum_{j=1}^{C}\frac{\max_{y_i=j}\left(f_i, w_j\right)}{\min_{k\neq j}\left(w_k, w_j\right)+\varepsilon}, \tag{3}$$

where ε is a moderating factor. Considering that the most classification weights cannot obtain a satisfactory distribution at the initial training stage, which may cause a disturbance for the ratio loss, we introduce ε factor to help the model learn smoothly.

We design the ratio loss for two main reasons. One is that the distance between feature vectors and classification weight vectors can be effectively measured in the hypersphere. The other one is that the maximal intraclass distance will gradually decrease and the minimal interclass distance will progressively enlarge by minimizing the ratio loss. Under the supervision of the ratio loss, the learned features have a well distribution in the embedding space, which can help improve the re-id accuracy.

3.3. Joint Training. The normalized softmax can learn angularly separable features in the hypersphere. However, the within-class restraint will gradually become slack along with the increase of interclass distance. Hence, the learned features are not sufficiently discriminative. Thus we propose a joint training of the normalized softmax loss and the ratio loss for the re-id task to maintain continual constraining force on the between-class discrepancy and within-class compactness. Therefore, the final loss function is formulated as follows:

$$\begin{aligned} L &= L_{NSL} + \lambda L_{\text{Ratio}} \\ &= -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{s\cdot\cos\left(\theta_{y_i}\right)}}{\sum_{j=1}^{C}e^{s\cdot\cos\left(\theta_j\right)}} + \lambda\frac{1}{C}\sum_{j=1}^{C}\frac{\max_{y_i=j}\left(f_i, w_j\right)}{\min_{k\neq j}\left(w_k, w_j\right)+\varepsilon}, \end{aligned} \tag{4}$$

where λ is a balance parameter to adjust the weight of the ratio loss. As normalized softmax loss restricts the features and classification weights to the hypersphere, the ratio loss can effectively optimize the between-class and within-class characteristics of features. The final loss can be easily optimized by SGD or Adam in the Pytorch framework [31].

3.4. A Toy Example Based on MNIST. To verify the feasibility and effectiveness of our proposed method, we conduct a toy experiment based on MNIST dataset [32] with a designed 8 layers CNN network by adopting the same experimental setting as [30]. We set the feature dimension as 2 so the learned features can be visualized on the 2-D plane, and 2000 training samples of each class are used to train the model. The visualizations of the original softmax, normalized softmax [30] are used to compare the effects of our proposed method (normalized softmax with ratio loss) in Figure 2.

From the experimental results, we could roughly make some conclusions as follows. (1) The original softmax focuses on separating the samples of different classes instead of learning discriminative features directly. So the learned features are able to reach preferable separability in the feature space but cause large within-class sparsity. (2) The normalized softmax removes variations in radial directions to optimize the model by normalizing the classification weights and features simultaneously. As a result, the learned features are angularly separable on a sphere and exhibit tighter within-class compactness. (3) On the basis of the normalized softmax, the proposed ratio loss could further improve the discriminability of the features by constraining the relation of intraclass and interclass distances. It can achieve a tighter within-class compactness as well as more obvious separability than the other two loss functions. These observations verify the effectiveness of our method and provide an experimental support for its application on person reidentification tasks.

4. Experiments

In this section, we give the experimental details of the proposed ratio loss for person reidentification and compare the experimental results on re-id datasets, e.g., Market-1501 [33], DukeMTMC-reID [34], and CUHK03 [35] with some state-of-the-art works. All involved experiments are conducted in the Pytorch framework.

4.1. Dataset Descriptions. Market-1501 is a large-scale person reidentification dataset which is collected in Tsinghua University. In the Market-1501 dataset, 1501 pedestrians are captured by six cameras (five 1280 × 1080 HD, one 720 × 576 SD), and 32668 bounding boxes of these 1501 pedestrians are detected by Deformable Part Model (DPM). Market-1501 is composed of a training set and a testing set. The training set contains 751 identities with 12936 training pedestrian images. The testing set includes 750 identities with 19732 gallery pedestrian images and 3368 query pedestrian images.

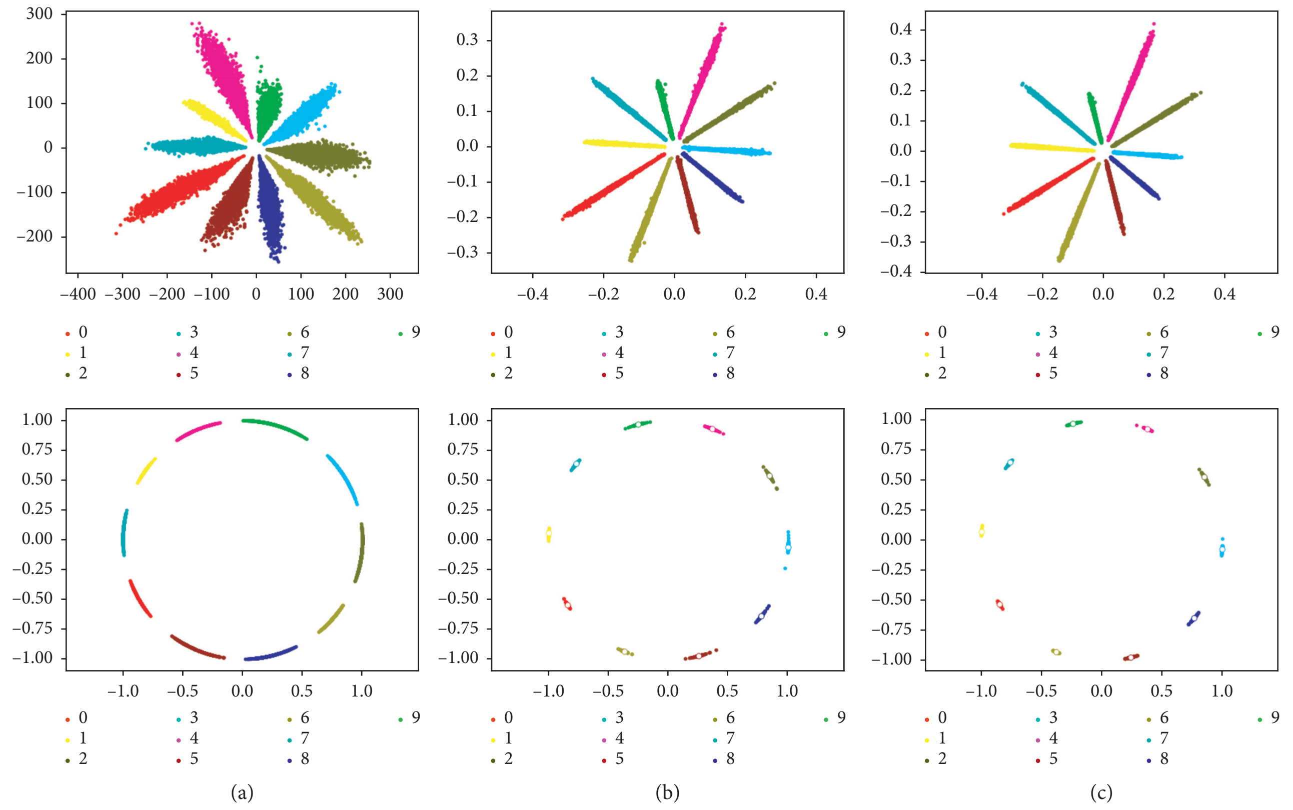

Figure 2: MNIST experiment results with the original softmax, normalized softmax, and normalized softmax with ratio loss, respectively. For the intuitive demonstration, we use a subset of MNIST for the experiments and 2000 training samples of each class are used to train the model. By setting the output dimension of the last feature layer as 2, the learned features can be visualized in 2D space, where the x-axis and y-axis correspond to the two dimensions of the learned features. In the figure, the first row gives the distributions of the original features in 2D space and the second row gives the corresponding normalized features. Best viewed in color. (a) Original softmax. (b) Normalized softmax. (c) Normalized softmax with ratio loss.

DukeMTMC-reID is a subset of the multitarget multicamera tracking dataset [36] which is collected outdoors in Duke University campus using 8 synchronized cameras. By selecting and cropping pedestrian regions from the videos of the tracking dataset, DukeMTMC-reID has 36411 pedestrian images of 1404 identities. The organization format of DukeMTMC-reID is the same as that of Market-1501. Concretely, 702 pedestrians constitute the training set with 16522 training images, and the remaining 702 pedestrians constitute the testing set with 2228 query images and 17661 gallery images.

CUHK03 re-id dataset is collected with 5 pairs of cameras in CUHK campus and contains 14096 pedestrian images of 1467 identities. The dataset provides a detected version in which the pedestrians are algorithmically detected and a labeled version where the pedestrians are manually labeled. It is worth noting that the original dataset is designed for a single-shot situation. Therefore, Zhong et al. [37] reorganized the CUHK03 dataset according to the format of Market-1501. In the new training/testing protocol, 767 pedestrians are used for training and the remaining 700 pedestrians constitute the testing set. In our experiments, we use the new training/testing protocol of CUHK03 to evaluate our method comprehensively.

4.2. Implementation Details

4.2.1. Preprocessing. First, all input training images are resized to 288 × 144 before they are randomly cropped to 256 × 128. Then, each input image would be flipped horizontally with a probability of 0.5. This operation is beneficial to the generalization ability of the model. Moreover, we use the random erasing trick [38] with a probability of 0.5 for each input image. It means that a small random rectangle region of a pedestrian image may be erased with zero value in the training procedure. This operation can enhance the robustness of the model by making a small area of input images invisible to the network.

4.2.2. Network Architecture. We construct a network architecture based on ResNet-50 in which the parameters have been pretrained in the ImageNet dataset as [30]. We remove the last fully connected layer of the original ResNet-50, and the remainder makes up a backbone which can automatically extract pedestrian features from input images. Besides, we change the last stride of ResNet-50 from 2 to 1 to retain more fine-grained pedestrian information with tiny extra computation cost.

To make the model more suitable for re-id tasks and facilitate the optimization of our proposed loss, we add several network layers behind the backbone. Concretely, we use a global average pooling (GAP) layer to aggregate the convolutional maps via an average operation. Then a batch normalization (BN) layer is attached to the GAP to shrink the internal covariate shift. Subsequently, a fully connected (FC) layer followed by another BN layer is used to compress the feature dimension into 1024. After that, the learned features and classification weights are both normalized in an L2 normalization layer. Finally, another fully connected layer is used as the classification layer in which the normalized softmax loss and proposed ratio loss can be calculated. After the training phase, this FC layer will be removed and the rest of the networks become a feature extractor used in the evaluation phase. The entire network architecture used in our re-id experiments is shown in Figure 3. We name the normalized softmax loss as ID loss for the sake of brevity.

4.2.3. Experiment Settings. All experiments are implemented in the Pytorch framework with an NVIDIA GTX 1080 Ti GPU. We use a balanced sampling strategy [12] during the training process. This strategy fixes the pedestrian number P and the image number K of per pedestrian in each sampling. By comparing with a random sampling strategy, the balanced sampling strategy can improve the re-id performance as well as accelerating the training process. In our experiments, we set P and K as 16 and 4, respectively, so the size of a minibatch in each iteration is 64.

We choose Adam optimizer to upgrade the parameters of the network. Besides, a warm-up strategy is adopted to initialize the learning rate at the beginning of training. In specific, the value of the learning rate will linearly increase from 10^{-5} to 10^{-3} during the first 20 epochs. After the warm-up stage, the learning rate remains unchanged until the 90^{th} epoch. Then we decay the learning rate by 0.1 at 90^{th} and 130^{th}, respectively, to fine-tune the parameters. It has been experimentally proved that the warm-up strategy can help the network achieve a better initial state for re-id problems [12]. The total number of training epochs is 150, and the learning rate curve is plotted in Figure 4. Moreover, we also use an online hard example mining (OHEM) scheme in our proposed method. In specific, we sort the training samples in descending order according to the value of the normalized softmax loss during each iteration, and the last 20% samples will be discarded. The OHEM scheme can effectively alleviate the model overfitting caused by overwhelming easy samples. Thus the robust and generalization ability of the learned model can be enhanced. We set the parameters λ and ε in ratio loss as 1 and 0.5, respectively, in our experiments. The scale coefficient s in the normalized softmax is set as 14.

4.2.4. Evaluation Metrics. In the evaluation phase, we remove the last FC layer from the training network to obtain the feature extractor for the person reidentification task. The testing images are resized to 288×144 before they are fed to the feature extractor. In specific, we extract the features of both the original input image and its horizontal flipping version, respectively. Then the final embedding is obtained by averaging these two features. The similarity between pedestrian images can be easily measured via their cosine distance of the features in the hypersphere.

We use two evaluation metrics including cumulative match characteristic (CMC) and mean average precision (mAP) to evaluate the performance of our proposed method. The re-id task is taken as a ranking problem in the CMC evaluation metric and a retrieval problem in the mAP evaluation metric. We report the cumulative match characteristic at Rank-1 in our results. The single-query/multi-shot mode is used for all experiments.

4.3. Experimental Results. The experimental results are given in the following tables. To be fair, we only make a comparison with some state-of-the-art methods based on deep learning, e.g., Deep-Person [39] and PCB [6]. Besides, the model trained by the normalized softmax loss without OHEM is regarded as the baseline of our method.

The experimental results on Market-1501 and DukeMTMC-reID are listed, respectively, in Table 1. We can find that the mAP increases on Market-1501 and DukeMTMC-reID respectively when the ratio loss or OHEM scheme is applied to the baseline model. Based on them, our proposed methods finally brings +1.47% and +0.56% increments than the baseline method in mAP on the two datasets respectively. Besides, our approach outperforms the most compared state-of-the-art methods on both mAP and Rank-1, like GSRW and PCB.

The experimental results on CUHK03 dataset under the new training/testing protocol are listed in Table 2. We observe that the proposed ratio loss and the OHEM scheme can improve the model performance dramatically. Finally, compared with the baseline method, our proposed method brings +6.38% improvements on mAP in the labeled version and +4.22% on mAP in the detected version. Moreover, we find that the performance of our method surpasses the listed state-of-the-art works by a large margin.

By analyzing the experimental results, we observe that the proposed ratio loss can further improve the re-id performance, which demonstrates the effectiveness of the ratio loss. Meanwhile, our method outperforms most listed state-of-the-art works on three re-id datasets and shows promising competitiveness.

5. Discussion

In this section, we first discuss the influences of two parameters λ and ε in the ratio loss by fixing one parameter and varying the other. Then we compare our

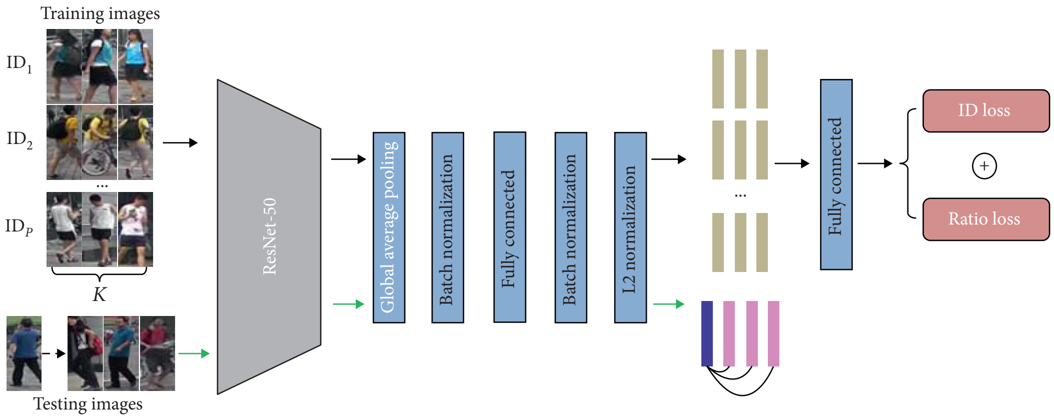

Figure 3: The network architecture for our person reidentification experiments is comprised of the backbone, global average pooling layer, batch normalization layer, fully connected layer, and L2 normalization layer. In the training procedure, the training images are organized as $P * K$ format in which P and K denote the number of identities and the sample number for each identity, respectively. Then the model learns the pedestrian features under the supervision of the ID loss and the ratio loss. In the testing phase, the last fully connected layer is removed and the remaining networks make up the feature extractor. The testing images are fed to the feature extractor to obtain pedestrian features, and the re-id task is conducted by comparing the similarity between extracted features.

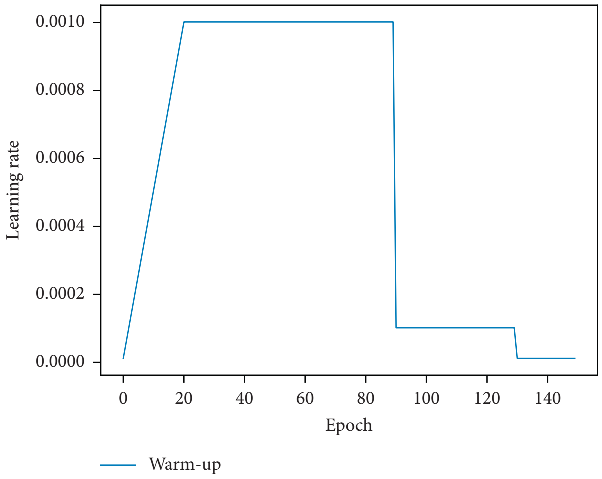

Figure 4: The learning rate curve with the warm-up strategy. During the first 20 epochs, the learning rate linearly increases from a small value of 10^{-5} to 10^{-3}, and it remains unchanged in the latter 70 epochs. Then the learning rate is decayed by 0.1 at 90th and 130th, respectively, to fine-tune the network parameters.

method with two similar works including LMCL [21] and ArcFace [5].

5.1. Parameter Analysis. The parameter λ is used for adjusting the weight of the ratio loss in the joint training. In order to observe the influence of λ on the re-id performance, we set ε as 0.5 and vary λ from {0.1, 0.2, 0.5, 1.0, 1.5, 2.0} on Market-1501, DukeMTMC-reID, and CUHK03, respectively. The results are given in Figure 5. From the results on Market-1501, we find that the mAP increases slightly as λ grows and achieves a peak when λ is 1.0, and then it reduces gradually with a larger λ. The similar mAP tendency can be observed on DukeMTMC-reID dataset. We also find that the mAP is greatly influenced by λ in CUHK03 dataset. For example, the mAP rises from 63.82% to 66.87% as λ increases from 0.1 to 1.5 in the labeled version.

The parameter ε could prevent the disturbance of ratio loss in the initial training stage. Similarly, we fix λ to 1 and change the value of ε from 0.0 to 0.5 with the step of 0.1. The results are shown in Figure 6. We find that ε has less influence on Market-1501 and DukeMTMC-reID. Specifically, the fluctuation of mAP is limited in 0.7% (0.43% for Market-1501 and 0.63% for DukeMTMC-reID). However, the fluctuation of the mAP reaches 1.67% for the labeled version and 1.29% for the detected version in CUHK03. We think the main reason for this phenomenon is the number difference of samples in different datasets. In large-scale

Table 1: The experiment results and comparisons with some state-of-the-art works for Market-1501 and DukeMTMC-reID datasets on mAP and Rank-1.

Methods	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
IDE [7]	46.00	72.54	—	—
SVDNet [40]	62.1	82.3	56.8	76.7
AACN [41]	66.87	85.90	59.25	76.84
TriNet [13]	69.14	84.92	—	—
DPFL [42]	72.6	88.6	60.6	79.2
GLAD [43]	73.9	89.9	—	—
HA-CNN [44]	75.7	91.2	63.8	80.5
DuATM [45]	76.62	91.42	64.58	81.82
Deep-person [39]	79.58	92.31	64.80	80.90
PCB [6]	77.4	92.3	66.1	81.8
PCB + RPP [6]	81.6	**93.8**	69.2	83.3-
GSRW [46]	82.5	92.7	66.4	80.7
Baseline	81.65	92.31	71.10	83.44
Baseline + ratio loss	82.14	92.43	71.25	84.16
Baseline + OHEM	82.44	92.73	71.61	84.16
Baseline + ratio loss + OHEM (ours)	**83.12**	92.64	71.66	84.34

The bold values indicate the best results of all the methods on each metric. They are beneficial to compare between our proposed method and the other methods.

Table 2: The experiment results and comparisons with some state-of-the-art works for CUHK03 labeled version and detected version on mAP and Rank-1.

Methods	Labeled		Detected	
	mAP	Rank-1	mAP	Rank-1
IDE [7]	21.0	22.2	19.7	21.3
SVDNet [40]	37.83	40.93	37.3	41.5
DPFL [42]	40.5	43.0	37.0	40.7
HA-CNN [44]	41.0	44.4	38.6	41.7
PAN [47]	35.0	36.9	34.0	36.3
PAN + Re-rank [47]	45.8	43.9	43.8	41.9
MLFN [48]	49.2	54.7	47.8	52.8
PCB [6]	—	—	53.2	59.7
PCB + RPP [6]	—	—	56.7	62.8
Baseline	59.88	61.14	59.02	60.57
Baseline + ratio loss	63.47	64.21	61.30	62.86
Baseline + OHEM	63.34	64.43	60.36	61.43
Baseline + ratio loss + OHEM (ours)	**66.26**	**68.57**	**63.24**	**65.07**

The bold values indicate the best results of all the methods on each metric. They can clearly demonstrate that our proposed method achieves the best performance compared with the other methods.

datasets such as Market-1501 and DukeMTMC-reID, massive samples bring a relatively small disturbance in the ratio loss, which is beneficial to the stabilized learning of the model. On the contrary, the model training in CUHK03 may risk a fluctuation in the ratio loss so the value of mAP is relatively insensitive to ε.

5.2. Comparison with Similar Works. In recent years, many excellent works have been proposed to enhance the discriminability of learned features, for example, LMCL [21] and ArcFace [5] loss functions. Both of them learn discriminative features by introducing a margin in the cosine space and the angular space, respectively. However, the value of the margin needs to be selected scrupulously because an inappropriate value would cause optimization difficulty. In our proposed ratio loss, the ratio formulation can effectively encourage the between-class separability and within-class compactness simultaneously without an extra margin.

For detailed comparisons, we conduct reidentification experiments on Market-1501, DukeMTMC-reID, and CUHK03 datasets with our method and the two loss functions. In the experiments, all the previous experimental settings are kept unchanged except for the loss function. For LMCL and ArcFace, we vary the margin parameter m from {0.01, 0.1, 0.3, 0.5, 1.0} to seek the best results as in [30], and the comparative results are recorded in Table 3. From the results, we can find

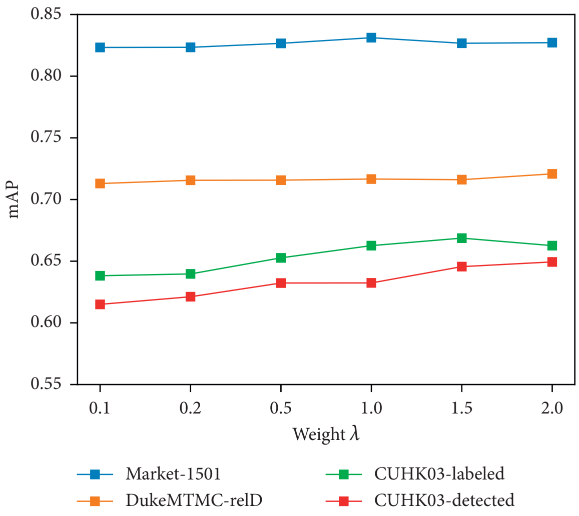

FIGURE 5: The sensitivity of the mAP to λ when ε is set as 0.5. The mAP of Market-1501 and DukeMTMC-reID is less sensitive to λ. Yet the mAP of CUHK03 overall shows a rising trend with the increase of λ.

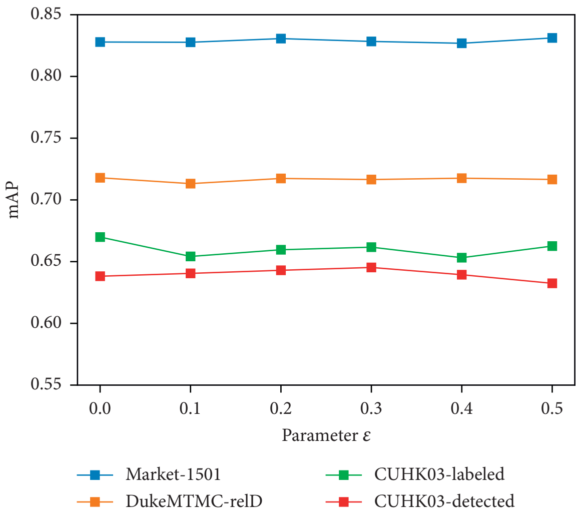

FIGURE 6: The sensitivity of the mAP to ε when λ is set to 1. The mAP of Market-1501 and DukeMTMC-reID is relatively stable to the change of ε. The mAP of CUHK03 fluctuates along with different ε.

TABLE 3: The comparisons of our proposed method with LMCL and ArcFace on Market-1501, DukeMTMC-reID, and CUHK03 datasets.

Methods	m	Market-1501		DukeMTMC-reID		CUHK03 labeled		CUHK03 detected	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
LMCL [21]	0.01	82.49	92.70	71.87	84.29	63.12	63.71	61.42	63.36
	0.1	82.93	**93.11**	**72.32**	**84.92**	63.55	66.50	62.54	63.64
	0.3	82.05	92.49	71.43	83.93	63.32	64.64	60.71	62.00
	0.5	81.70	93.02	70.68	84.07	62.15	64.21	58.49	60.43
	1.0	81.15	92.25	70.51	84.11	61.86	63.50	58.46	59.79
ArcFace [5]	0.01	82.31	92.99	71.19	84.02	63.02	64.21	60.92	63.00
	0.1	82.70	92.96	71.28	83.93	64.33	66.50	62.54	64.93
	0.3	82.52	92.99	71.09	84.78	65.20	67.57	62.65	63.43
	0.5	81.28	92.34	70.54	83.89	64.62	65.86	61.83	63.07
	1.0	80.86	91.42	69.60	83.08	62.76	64.14	60.72	61.50
Ours	—	**83.12**	92.64	71.66	84.34	**66.26**	**68.57**	**63.24**	**65.07**

The bold values indicate the best results of all the methods on each metric. They are beneficial to compare between our proposed method and the other methods.

that our method has a higher mAP value than LMCL and ArcFace on Market-1501 and achieves comparable performance with them on DukeMTMC-reID. Moreover, it outperforms LMCL and ArcFace completely on CUHK03 dataset even if they are with the best margin parameters.

6. Conclusions

In this paper, we proposed a circle-based ratio loss to learn discriminative features for person reidentification. To enhance feature discriminability, we first use the normalized softmax to regulate the magnitudes of feature vectors and classification weight vectors. In this way, the network will concentrate on the angle relationship between features and classification weights, and their distance can be effectively measured in the hypersphere. Then we take the ratio of the maximal intraclass distance and the minimal interclass distance as the objective loss, so that the intraclass compactness and interclass separability can be optimized at the same time. With the joint training of the normalized softmax and proposed ratio loss, the model could learn discriminative pedestrian features for person reidentification tasks. Extensive experiments on Market-1501, DukeMTMC-reID, and CUHK03 are conducted to demonstrate the effectiveness of our proposed re-id method.

Data Availability

The underlying data related to our submission are publicly available for research.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by National Natural Science Foundation of China (nos. 61501177 and 61902084), Guangzhou University's Training Program for Excellent New-Recruited Doctors (no. YB201712), and the Featured Inovation Project of Guangdong Education Department under Grant 2018KTSCX174.

References

[1] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1622–1634, 2013.

[2] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3594–3601, New York, OR, USA, June 2013.

[3] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 2288–2295, Providence, RI, USA, June 2012.

[4] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 2197–2206, Boston, MA, USA, June 2015.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 4685–4694, Long Beach, CA, USA, July 2019.

[6] 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 *Proceedings of the European Conference On Computer Vision*, pp. 501–518, Munich, Germany, September 2018.

[7] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: past, present, and future," 2016, http://arxiv.org/abs/1610.02984.

[8] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: a deep quadruplet network for person re-identification," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 1320–1329, Honolulu, HI, USA, July 2017.

[9] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 1735–1742, New York, NY, USA, June 2006.

[10] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *Proceedings of the European Conference On Computer Vision*, pp. 791–808, Amsterdam, Netherlands, October 2016.

[11] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proceedings of the European Conference On Computer Vision*, pp. 22–275, Marseille, France, October 2008.

[12] X. Fan, W. Jiang, H. Luo, and M. Fei, "SphereReID: deep hypersphere manifold embedding for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 51–58, 2019.

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

[14] W. Liu, Y. Wen, Z. Yu et al., "SphereFace: deep hypersphere embedding for face recognition,," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6738–6746, Honolulu, HI, USA, July 2017.

[15] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, Boston, MA, USA, June 2015.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[17] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.

[18] 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 *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 1335–1344, Las Vegas, NV, USA, June 2016.

[19] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proceedings of the*

International Conference On Machine Learning, pp. 507–516, New York, NY, USA, June 2016.

[20] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proceedings of the ACM International Conference On Multimedia*, pp. 274–282, Amsterdam, Netherlands, June 2018.

[21] H. Wang, Y. Wang, Z. Zhou et al., "CosFace: large margin cosine loss for deep face recognition," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 5265–5274, Salt Lake City, UT, USA, June 2018.

[22] D. Wu, S.-J. Zheng, W.-Z. Bao, X.-P. Zhang, C.-A. Yuan, and D.-S. Huang, "A novel deep model with multi-loss and efficient training for person re-identification," *Neurocomputing*, vol. 324, pp. 69–75, 2019.

[23] Y. Yang, X. Liu, Q. Ye, and D. Tao, "Ensemble learning-based person re-identification with multiple feature representations," *Complexity*, vol. 2018, Article ID 5940181, 12 pages, 2018.

[24] F. Zheng, C. Deng, X. Sun et al., "Pyramidal person re-identification via multi-loss dynamic training," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 8514–8522, Long Beach, CA, USA, June 2019.

[25] W. Zhong, L. Jiang, T. Zhang, J. Ji, and H. Xiong, "Combining multilevel feature extraction and multi-loss learning for person re-identification," *Neurocomputing*, vol. 334, pp. 68–78, 2019.

[26] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *Proceedings of the IEEE International Conference On Computer Vision*, pp. 542–551, Seoul, Korea, November 2019.

[27] Y. Zhu, Z. Yang, L. Wang, S. Zhao, X. Hu, and D. Tao, "Hetero-center loss for cross-modality person re-identification," *Neurocomputing*, vol. 386, pp. 97–109, 2020.

[28] Y. Fu, Y. Wei, G. Wang et al., "Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification,," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6111–6120, Seoul, Korea, October 2019.

[29] H. Luo, W. Jiang, Y. Gu et al., "A strong baseline and batch normalization neck for deep person re-identification," *IEEE Transactions on Multimedia*, vol. 22, 2020.

[30] Z. Yang, T. Liu, J. Liu, L. Wang, and S. Zhao, "A novel soft margin loss function for deep discriminative embedding learning," *IEEE Access*, vol. 8, pp. 202785–202794, 2020.

[31] "Pytorch," https://pytorch.org/.

[32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[33] L. Zheng, L. Shen, L. Tian et al., "Scalable person re-identification: a benchmark,," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, Santiago, Chile, December 2015.

[34] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proceedings of the IEEE International Conference On Computer Vision*, pp. 3774–3782, Venice, Italy, October 2017.

[35] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 152–159, Columbus, OH, USA, June 2014.

[36] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proceedings of the European Conference On Computer Vision Workshop*, pp. 17–35, Las Vegas, NV, USA, July 2016.

[37] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 3652–3661, Honolulu, HI, USA, July 2017.

[38] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," 2017, http://arxiv.org/abs/1708.04896.

[39] X. Bai, M. Yang, T. Huang et al., "Deep-Person: learning discriminative deep features for person re-identification," *Patten Recognition*, vol. 98, 2020.

[40] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proceedings of the IEEE International Conference On Computer Vision*, pp. 3820–3828, Venice, Italy, October 2017.

[41] J. Xu, R. Zhao, F. Zhu, H. Wang, and W. Ouyang, "Attention-aware compositional network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2119–2128, Salt Lake City, UT, USA, June 2018.

[42] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," in *Proceedings of the IEEE International Conference On Computer Vision Workshop*, pp. 2590–2600, Honolulu, HI, USA, July 2017.

[43] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian, "GLAD: global-local-alignment descriptor for pedestrian retrieval," in *Proceedings of the ACM International Conference On Multimedia*, pp. 420–428, New York, NY, USA, October 2017.

[44] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2285–2294, Salt Lake City, UT, USA, June 2018.

[45] J. Si, H. Zhang, C. Li et al., "Dual attention marching network for context-aware feature sequence based person re-identification," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 5363–5372, Salt Lake City, UT, USA, June 2018.

[46] Y. Shen, H. Li, T. Xiao et al., "Deep group-shuffling random walk for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2265–2274, Salt Lake City, UT, USA, June 2018.

[47] Z. Zheng, L. Zheng, and Y. Yang, "Pedestrian alignment network for large-scale person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3037–3045, 2019.

[48] X. Chang, T. M. Hospedales, and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, pp. 2109–2118, Salt Lake City, UT, USA, June 2018.