Person re-identification based on linear classification margin

Bing Chen*, Yufei Zha, Wu Min, Zhou Yuan
Aeronautics Engineering College, Air Force Engineering University, Xi’an 710038, China
*Corresponding author e-mail: deeplearningcb@163.com

Abstract. The core of the person re-identification task is to find a discriminative feature map or a good measure to measure the similarity between two pedestrians. Under the traditional softmax loss function constraint, the characteristics of the FC layer output of the convolutional neural network are only separable and the discriminative ability is insufficient. Our proposes a person re-ID method based on decision margin. By adding the traditional softmax loss function to the angle decision margin, the characteristics of strong discriminative power in network learning are enhanced. This paper provides a clear geometric interpretation by normalizing the last fully connected layer parameters and pedestrian characteristics of the network and projecting the embedded features into the geometric space for person re-identification. Combining two different classification margins, a linear classification margin is proposed and good performance is achieved. This paper validates the effectiveness of the algorithm on the mainstream databases of Market1501 and DukeMTMC-reID. The results show that our method can get more discriminative features, and its experimental performance is significantly improved compared to the existing advanced algorithms.

1. Introduction
The core of the person re-ID[1] task is to find a discriminative feature map. Recently, convolutional neural networks [2] have developed rapidly in computer vision task. The re-ID algorithm [3-6] based on CNN has largely enhanced the performance of person re-ID. However, existing methods [7-13] based on CNN are not able to improve the discriminability of pedestrian characteristics.

The most commonly used loss function in a classification network is softmax, which can easily optimize network parameters. The method [7] solves the person re-identification as a classification problem, which uses the softmax loss function to constrain the network training, and uses a single-channel CNN to obtain features. However, the softmax loss function does not explicitly constrain the distance between samples, so the resulting features are not discriminative. In order to enhance the discriminability of features, the article [8] introduces the multiplicative angular margin to constrain the angle between the feature and the weight to one of the original n points, but the constraint is too strict, which makes the network difficult to converge. In [7], the weights and features are normalized, and the angle is constrained to be smaller than the traditional classification loss function. This paper aims to reduce the intersection of the inter-class sample distance distribution by constraining the distance between similar samples and the distance between different types of samples.
In the face recognition task, adding the classification margin \cite{8}\cite{9} to the classification loss function can effectively enhance the discriminability of the feature. This paper adds the classification margin to the traditional softmax loss function to enhance the discriminability of pedestrian characteristics. These classification-based classification loss functions are first applied to the field of person re-identification. Secondly, the pedestrian characteristics and network weights are normalized, and a linear classification margin is proposed by combining the multiplicative margin \cite{8} and the additive margin \cite{9}. The addition of a linear classification margin to the classification loss function enhances the constraints on the model and effectively improves the discriminability of pedestrian characteristics.

2. Related work
This paper is closely related to two mainstream studies. 1. Improved depth feature learning under softmax loss function constraints. 2. Person re-identification based on convolutional neural network. This section will review the work closely related to the above two aspects.

**Improved depth feature learning under softmax loss function constraints.** In the field of face recognition, many methods for improving the traditional softmax loss function have appeared in recent years. The article \cite{8} proposes a large margin softmax loss function, which introduces the angle between the feature and the weight by the angle margin of the multiplicative into one of the original data points. The article \cite{9} has further improved the large margin softmax. It preserves the multiplicative margin while normalizing the weight vector, mapping the pedestrian's features to the three-dimensional spherical surface for analysis, providing a clear physical explanation for the theory of large margin softmax. The article \cite{7} proposed a L2 softmax loss function, which normalizes features and weights to L2, and uses cosine distance to measure similarity.

**Person re-identification based on CNN.** The method based on CNN has achieved outstanding performance in the person re-identification task. Some of the representative work can be subdivided into: feature learning and metric learning based on CNN. In feature learning \cite{9}, the literature uses a framework of joint learning to unify embedding of individual pictures and embedding of intersecting images. The literature \cite{13} proposes to first align the pedestrian image, and then divide the pedestrian image into multiple parts to learn the fine-grained features of the pedestrian. In terms of distance measurement, the literature \cite{5} proposes to use a siamese model to learn the similarity of sample pairs.

2.1. Traditional softmax classification loss function
The softmax loss is often used to constrain the training of classification networks. The expression is as follows:

\[
L = -\sum_{i=1}^{n} \log \frac{e^{\|f(x_i)\| \cos(\theta_i) + b_j}}{\sum_{j=1}^{n'} e^{\|f(x_i)\| \cos(\theta_i) + b_j}}
\]

(1)

\[ n \] is the size of the batch, and \[ n' \] is categories. \[ f(x_i) \in \mathbb{R}^d \] is the depth embedding of the \( i \)-th, with a category of \( y_i \) and a dimension of \( d \). \( W_j \in \mathbb{R}^d \) is the \( j \)-th row parameter of \( W \), and \( b \in \mathbb{R}^n \) is the bias of this layer. A classification model training network is commonly used in person re-ID task. However, the traditional softmax loss function only constrains the network for correct classification, and does not explicitly constrain the distance between samples, which makes the feature discriminative. In the case of two classifications, \( n' = 2 \), where \( p_i \) is the probability that the sample predicted by the classifier belongs to the first category. \( p_2 \) means the sample belongs to the second category, then the decision boundary is the point of \( p_1 = p_2 \), and the traditional softmax loss function constraint: When \[ W_1 \|f(x_i)\| \cos(\theta_i) + b_j = W_2 \|f(x_i)\| \cos(\theta_i) + b_j \] samples are the first category, and vice versa.
2.2. Multiplicative classification margin

In Sphereface [8], in order to visually discuss the properties of high-dimensional eigenvectors, which project features onto three-dimensional geometric spheres, a loss called Sphere loss is proposed. In order to simplify the calculation, the bias term is set to 0 and the weight W is normalized to L2. This makes it possible to rely only on the angle between features and weights when predicting sample categories. In order to optimize the discriminability of the features, it introduces an angular margin so that the spacing between samples of the same type can be compressed less. Thus, a decision margin is generated at the time of discrimination. Its formula is as follows:

\[
L = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{e^{(\cos(\theta_{j,i})}}{\sum_{j \neq y_i} e^{(\cos(\theta_{j,i}))}}
\]

(2)

Where \( \theta_{j,i} \in [0, \frac{\pi}{m}] \). To remove this limit and make it an optimizable function, it introduces a monotonically increasing function

\[
\rho(\theta) = \begin{cases} 
-1 & \cos(m\theta) - 2k, \theta_{j,i} \in \left[\frac{k\pi}{n}, \frac{(k+1)\pi}{n}\right] \& k \in [0, m-1]
\end{cases}
\]

which is strictly equal to \( \cos(m\theta) \) in \( [0, \frac{\pi}{m}] \). Then its loss function expression becomes:

\[
L = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{e^{(\rho(\theta_{j,i})}}{\sum_{j \neq y_i} e^{(\cos(\theta_{j,i}))}}
\]

(3)

In the case of the two classifications, before the addition of the angular margin, for the sample feature \( f(x) \), when it belongs to the first class, the condition for correctly determining it as the first class is

\[
\cos(\theta_{j,i}) > \cos(\theta_{j})
\]

After adding the angle margin, when it belongs to the first class, it needs \( \rho(\theta_{j,i}) > \rho(\theta_{j}) \) to be correctly predicted as the first class, and when it belongs to the second class, it is correctly predicted as the second class when it meets \( \cos(\theta_{j,i}) > \cos(\theta_{j}) \) \& \( \theta_{j,i} \in [0, \frac{\pi}{2}] \), in this case, the sample needs to be discriminated as the angle between the first class of the required sample and the weight of the corresponding category is smaller, which is equivalent to strengthening the constraint on the network, that is, the output of the network is close to the expected output. An angular margin of angle \( \frac{\pi}{m} \) \( \theta^j \) is the angle between \( W_1 \) and \( W_2 \) is generated.

2.3. Additive classification margin

In fact, the multiplicative classification margin[9] makes the constraint of the loss function too strict, so it becomes difficult to converge during network training. The article normalizes sample features and network weights. The constraints of the network are enhanced by adding an additive classification margin to the traditional softmax loss function. Its formula is as follows:

\[
L = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{e^{(\cos(m\theta_{j,i})}}{\sum_{j \neq y_i} e^{(\cos(\theta_{j,i}))}} + \sum_{j \neq y_i} e^{\cos(\theta_{j,i})}
\]

\[\text{s.t.} \quad W_j = \left[ \begin{array}{c} W_j \\ x_i \end{array} \right], \quad x_i = \left[ \begin{array}{c} x_i \\ 1 \end{array} \right], \quad \cos(\theta) = W_j^T x_i
\]

(4)

Similar to the multiplicative margin, the margin of \( \cos(m\theta) \leq \cos(\theta) \) \& \( 0 \leq \theta \leq \frac{\pi}{n} \) enhances the constraint on the classification loss, and a 2m classification margin is generated.
2.4. Linear margin

Both of these margins have shortcomings. This paper proposes a linear classification margin for the shortcomings of the two classification margins, which can simultaneously limit the distance within the class and the distance between the classes. Its formula is as follows:

\[
L = -\frac{1}{n} \sum_{i=1}^{n} \log \left( e^{\cos(\theta_i s m)} + \sum_{j \neq i} e^{\cos(\theta_j)} \right)
\]

s.t. \( W_j = \frac{W_j}{\|W_j\|}, x_i = \frac{x_i}{\|x_i\|}, \cos \theta_j = W_j^T x_j. \)

The distance between the classes is compressed by linear spacing, so that the angle between the feature and the weight of the sample is compressed to \( N \). This can more reasonably compress the angle between the samples within the class, which helps to learn more discriminative features.

3. Person re-identification based on margin softmax

Unlike the framework used in the previous chapters, the framework used in this chapter is the PCB structure in [13]. The difference from the previous frame is that the training input image is 256*128 instead of 244*244. This paper is consistent with the sample input training phase. The difference is that the feature normalization and weight normalization layers are added to the network. The loss function part uses the softmax loss function based on the linear classification margin proposed above. The network structure diagram of the re-ID algorithm in our paper is:

![Network Structure Diagram](image)

Figure 1 Algorithm network structure diagram

After entering the network, the pedestrian image gets a three-dimensional tensor. The resulting three-dimensional tensor level is then divided into six regions. Then, it is averaged to obtain 6 column vectors, and then the 6 column vectors are respectively reduced in dimension, and then the 6 features of the reduced dimension are input into the classifier for classification. In the training phase, the loss of this paper is used as the sum of the six-part classification loss function. In the test phase, the reduced-dimensional six-part feature \( h_i \) is normalized and the combined set \( H \) is used as the final feature description. The feature \( G \) before dimensionality reduction can also be used.

4. Experimental results and analysis

4.1. Experiment setup

The experimental platform in this paper is configured as an i7-5830k CPU and a 2*GTX1080ti GPU machine, which is implemented under the Pytorch deep learning framework. In order to reflect the contribution of the classification margin proposed in this paper to the experimental performance, this paper uses the same network framework as PCB [13], with ResNet50 as the backbone network framework. After pool5, a FC layer with a dimension of 256 dimensions is added.

4.2. Comparison of various classification margins

The nature of the additive margin and the multiplicative margin are all added to the classification margin when classifying, which is beneficial to the model to learn more discriminative embedding. Several margins are different in expression and geometric sense, and there are also differences in experimental performance. As can be seen from the data in the figure, the model with the classification margin has improved performance compared to the model with no margin. Different margin have different performance enhancements, and it can be seen that the linear spacing in this paper is the
The biggest improvement in performance. The model in this paper is 3.5% and 2.3% higher than the uninterrupted models mAP and rank-1, respectively. Compared with the additive margin model, the models were increased by 1.3% and 0.8%, respectively, and the models with multiplicative margin were increased by 1.5% and 1%, respectively. It can be seen that the model of this paper can effectively enhance the discriminability of pedestrian characteristics.

### Table 1 Performance of various margin models

| Model                  | Market1501 |
|------------------------|------------|
|                        | rank-1    | rank-5  | rank-10 | mAP  |
| PCB [13]               | 92.3      | 97.2    | 98.2    | 77.4 |
| PCB [13]+ Additive margin | 93.8      | 97.6    | 98.1    | 79.6 |
| PCB [13]+ Multiplicative margin | 93.6      | 96.8    | 98.5    | 78.8 |
| PCB [13]+ Linear margin | 94.6      | 97.9    | 98.7    | 80.9 |

#### 4.3. Comparison with the other method

This paper compares the performance of this algorithm with the current latest mainstream algorithms on the two mainstream databases, Market1501 and DukeMTMC-reID, including SOMAnet [14], ReRank [12], PANL [15], Trihard [11], PCB [13], PDC [16] and other advanced algorithms.

Table 2 reveals the experimental results on the Market1501 database. The comparison algorithms in the table are currently advanced algorithms. PCB is an excellent algorithm that takes the performance of person re-ID to another level in a simple way. Therefore, this paper uses it as a benchmark algorithm to improve. It can be seen that the use of this model on the PCB frame can enhance the performance of person re-ID. Compared with the PCB algorithm, the algorithm is improved by 3.5% on mAP and 2.2% on rank-1. There are also different levels of improvement in other performance indicators. Compared with advanced algorithms such as SOMAnet [14], ReRank [12], PANL [15], Trihard [11], PCB [13] and PDC [16], the performance of this paper is significantly superior to them.

### Table 2 Comparison with the other method on the Market1501 Dataset

| Method             | Market1501 |
|--------------------|------------|
|                    | rank-1    | rank-5  | rank-10 | mAP  |
| SOMAnet[14]        | 73.87     | 88.03   | 92.22   | 47.89|
| ReRank(R)[12]      | 77.1      | -       | -       | 63.6 |
| PANL(R)[15]        | 82.81     | 93.53   | -       | 63.35|
| Trihard [11]       | 84.92     | 94.21   | -       | 69.14|
| PDC [16]           | 84.4      | 92.7    | 94.9    | 63.4 |
| SVDNet[7]          | 82.3      | 92.3    | 95.2    | 62.1 |
| PCB[13]            | 92.3      | 97.2    | 98.2    | 77.4 |
| PCB+Our model(R)   | **94.5**  | **97.9**| **98.7**| **80.9**|

### 5. Conclusion

Inspired by the related methods face recognition task, this paper proposes a person re-ID method based on linear classification margin when studying the problem of person re-ID. The feature of strong discriminative power in network learning is enhanced by adding a classification margin to the traditional softmax loss function. In addition, theoretical analysis of different kinds of margins is carried out to find out the deficiencies, and a linear classification margin is proposed for these deficiencies. It is verified on the two mainstream databases, Market1501 and DukeMTMC-reID.

### References

[1] Zheng Z, Zheng L, Yang Y. A Discriminatively Learned CNN Embedding for Person Re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 14(1):13.
[2] Ranjan R, Castillo C D, Chellappa R. L2-constrained Softmax Loss for Discriminative Face Verification[J]. 2017.

[3] W. Liu, Y. Wen, Z. Yu, et al. Sphereface: Deep hypersphere embedding for face recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6738 – 6746.

[4] Deng J, Guo J, Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J]. arXiv preprint arXiv:1801.07698, 2018.

[5] Yi D, Lei Z, Liao S, et al. Deep Metric Learning for Person Re-identification[C]. International Conference on Pattern Recognition, 2014:34-39.

[6] Liu J, Zha Z J, Tian Q I, et al. Multi-Scale Triplet CNN for Person Re-Identification[C]. ACM on Multimedia Conference, 2016:192-196.

[7] Sun Y, Zheng L, Deng W, et al. SVDNet for Pedestrian Retrieval[C]. IEEE International Conference on Computer Vision, 2017: 3820 - 3828.

[8] Zhang X, Luo H, Fan X, et al. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification[J]. arXiv preprint arXiv:1711.08184, 2017.

[9] Li W, Zhao R, Xiao T, et al. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014:152-159.

[10] Liu Y, Li H, Wang X. Learning Deep Features via Congenous Cosine Loss for Person Recognition[J]. 2017.

[11] Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-Identification. arXiv preprint arXiv:1703.07737, 2017.

[12] Zhong Z, Zheng L, Cao D, et al. Re-ranking Person Re-identification with k-reciprocal Encoding[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:3652-3661

[13] Sun Y, Zheng L, Yang Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling [C]. European Conference on Computer Vision, 2018:501-518.

[14] Barbosa I B, Cristani M, Caputo B, et al. Looking beyond appearances: Synthetic training data for deep CNNs in re-identification[J]. Computer Vision and Image Understanding.

[15] Zheng Z, Zheng L, Yang Y. Pedestrian Alignment Network for Large-scale Person Re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018.

[16] Su C, Li J, Zhang S, et al. Pose-Driven Deep Convolutional Model for Person Re-identification[C]. IEEE International Conference on Computer Vision, 2017:3980-3989.