Person Re-identification Algorithm Based on Logistic Regression Loss Function

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 feature description of pedestrians is the key to the task of person re-identification. The deep feature extracted from convolutional neural networks proved to have good robustness. Deep feature learning based on convolutional neural network also achieved good performance in person re-identification tasks. However, the feature discriminant is not strong under the constraint of traditional classification loss function, which seriously restricts the person re-identification performance. In order to enhance the discrimination of features, this paper uses the idea of logistic regression to associate the distance between samples and the similarity probability and proposes a logistic regression loss function to constrain the training of the network. The logistic regression loss function proposed in this paper can constrain the network to make the distance between similar samples small, and the distance between different classes of samples large. Secondly, by introducing the positive sample distance threshold, this paper avoids that the network constrains the same sample with some very different samples from being too small, while the distance constraint to other samples fails. Finally, the proposed logic regression loss function is embedded into the Siamese convolutional neural network for end-to-end person re-identification. This paper validates the effectiveness of the algorithm on the mainstream databases of Market1501 and CUHK03.

1. Introduction
Person re-identification [1] is to find the pedestrian image matching the target pedestrian from the massive image database taken by different cameras. With the development of monitoring networks and the rise of public security requirements, person re-ID has become more widely used in pedestrian tracking, security, and monitoring. However, due to differences in shooting angles, lighting, etc., there will be a huge difference in the pictures taken by the same person in different cameras, and different people will become very different due to posture, clothes color, etc. Small, this creates enormous difficulties and challenges for person re-ID tasks.

The key to person re-ID is to compute the similarity of all the pictures in the query sample and the database to be queried. Here, \( g \) is the database to be queried, which is composed of \( N \) types of pedestrian photos, and is denoted as \( \{g_i\}_{i=1}^{N} \). The characteristics of the query sample are represented by \( q \), and its category is determined by equation (1):

\[
i^* = \arg \max_{i=1,2,...,N} d(q, g_i)
\]
Where \( i^* \) is the class of the query sample \( q \) and \( d(p, q) \) is the distance. From equation (1), the person re-identification task consists of a feature description and a distance metric. In the feature description, there are manual design features such as color [2], texture [3], gradient histogram, etc. to characterize the appearance of pedestrians. Zheng [7] combines the classification model with the verification model and uses softmax constraint network training in both the classification model and the verification model. The softmax loss function can constrain the network to classify, but it only has the property of making the data separable [7], and the discriminability is not strong. As shown in Figure 1(b), the query sample and the red dot represent the same person, but in the separable feature space, the distance between the red sample to be queried and the query sample is often greater than the blue sample and the query sample. In the case of the distance between the two, the final network misjudged the blue sample as the same sample of the query sample, which seriously affected the accuracy of person re-identification.

![Discriminative feature learning](image)

In order to enhance the discriminability of features, the distribution of features expected in this paper is shown in Figure 1(c). The distance of similar samples is as small as possible, and the distances of different types of samples are as large as possible. In response to this expectation, this paper proposes a loss function based on logistic regression. The model uses the idea of logistic regression to solve the problem of similarity measure between pedestrians, and links the distance between samples with the similarity probability to constrain the similar samples. The probability of similarity is large and the distance is small, and the distance between the pairs of negative samples is as large as possible, so that the discriminability of the features is enhanced. In this paper, the effectiveness of our method is verified experimentally on the Market1501 and CUHK03 databases.

2. Related work

Our paper is related to two research: 1) Deep metric learning with CNNs, 2) Joint learning with identification model for Re-Id. This section briefly reviews several closely related aspects, CNN-based re-ID methods and face applications.

**Deep metric embedding learning with CNNs**: CNN based methods have obtained good performance in various computer vision tasks, the person reidentification subtask [4][12][15] is no exception to this. Among them, some representative works views the person Re-Id task as a deep metric learning problem and adopt verification loss that include [7] contrastive loss and [8] triplet loss to train the networks.

**Joint learning with identification model for Re-Id**: Combining the two models can make full use the annotated data in terms of image identity and pair-wise similarity. For instance, wang et al.[4] used the contrastive loss and the softmax loss to supervise the network learning. Wen et al.[9] proposed a loss called center loss to obtain discriminative feature of face images, and combined softmax to constrain the network. Sun et al. [14] jointly used the verification and identification losses to training the network.

3. Person re-identification of loss function based on logistic regression

3.1. Logistic regression loss function

The softmax loss can only constrain the network to correctly classify the samples, and does not form a constraint on the distance, so the feature discriminant is poor. In order to improve the discriminative ability of learning features, this paper propose a loss function based on logistic...
regression model. Logistic regression functions have a wide range of applications in deep learning and neural networks. The expressions of logistic regression functions are: \( \sigma(\theta) = \frac{1}{1 + e^{-\theta}} \). For any input sample \( \theta \), the logistic regression function can map it to a value between \([0,1]\), where the output \( \sigma \) can be used as a probability value. Let \( T = \{ x_i, x_j, s_{ij} \} \) be the training dataset, where \( (x_i, x_j) \) is the input sample pair where \( s_{ij} = 1 \) indicates that the input sample is a similar sample pair, and conversely \( s_{ij} = 0 \), the depth feature of the sample pair is \( \{ f_i, f_j \}_{ij} \). Then define the conditional probability:

\[
p(s_{ij} | F) = \begin{cases} 2\delta(D_{ij}), & s_{ij} = 1 \\ 1 - 2\delta(D_{ij}), & s_{ij} = 0 \end{cases}
\]

(2)

Where \( \delta(D_{ij}) = 1 - \theta(D_{ij}) = \frac{1}{1 + e^{-D_{ij}}} \), \( D_{ij} = \| f_i - f_j \| \geq 0 \) represents the distance between pairs of samples.

The constant 2 in equation (2) is to ensure that the upper bound of the probability value is 1. It can be seen that the smaller the distance \( D_{ij} \) between the samples, the larger the probability value \( p(s_{ij} = 1 \mid F) \) of the sample pair being judged as the same label. Vice versa, where \( p(s_{ij} = 1 \mid F) = 1 \) when \( D_{ij} = 0 \), these properties are in line with the expectations of person re-ID tasks. The likelihood function of the observed sample can be expressed as:

\[
P(S) = \prod_{(i,j) \in T} p(s_{ij})
\]

(3)

Taking a negative logarithm of this formula can be obtained:

\[
L_{\text{verifi}} = -\log \prod_{(i,j) \in T} p(s_{ij})
\]

\[
= - \sum_{(i,j) \in T} (s_{ij} \log 2 - \log (e^{D_{ij}} + 1) + (1 - s_{ij}) \log(e^{D_{ij}} - 1))
\]

(4)

Minimizing this loss function compresses the distance between samples of the same type to near zero \((D_{ij} \rightarrow 0)\) while expanding the distance between different classes of samples to positive infinity \((D_{ij} \rightarrow +\infty)\).

### 3.2. Positive sample distance threshold

In practical applications, because the angle of view and the field of view are different, even the same sample will have a very large distance. If the network is forced to constrain the distance between a part of the original sample with a large difference, it will be Will cause failures in other sample distance constraints. Therefore, this paper introduces a positive sample distance threshold in the loss function proposed above to prevent the network from constraining the distance between positive samples too small. After introducing the positive sample distance threshold, the conditional probability is redefined as:

\[
\hat{p}(s_{ij} \mid F) = \begin{cases} \min(1, 2\delta(D_{ij} - T)), & s_{ij} = 1 \\ 1 - \min(1, 2\delta(D_{ij} - T)), & s_{ij} = 0 \end{cases}
\]

(5)

When the distance \( D_{ij} \) between positive sample pairs is close to \( T \), \( \hat{p}(s_{ij} = 1 \mid F) \) will be close to 1, and the network begins to converge. Therefore, the introduced distance threshold prevents the network from constraining the distance between similar samples too small. Then the corrected loss function is:

\[
L_{\text{verifi}} = - \sum_{(i,j) \in S} \min(0, \log 2 - \log (1 + e^{(D_{ij} - T)})) - \sum_{(i,j) \in D} \log \max(e^{(D_{ij} - T) - 1}, e^{(D_{ij} - T) + 1})
\]

(6)
In this paper, $\epsilon = 0.0001$. In order to simplify the formula, this paper defines a similar sample pair $S$ and a dissimilar sample pair $D$, $(i,j) \in S$ means that the similar sample pair $(x_i, x_j)$ is the same person, $T = S \cup D$. The loss function $L_s$ of the positive sample and the loss function $L_o$ of the negative sample pair. The first term $L_s$ is used to punish a positive sample pair of similar samples for a medium distance exceeding a given threshold $T$. It can constrain the distance between positive sample pairs as much as possible, while avoiding the network to constrain the distance between similar samples. Too small. The second term $L_o$ constrains the distance between dissimilar sample pairs as large as possible.

However, person re-identification has always been a problem of extreme imbalance. The number of positive sample is small, resulting in the network learning being compute by a large number of negative sample pairs, while the positive sample pair is for the loss function. Optimization is easy to ignore. To solve this problem, this paper introduces a weight to balance the proportion of sample pairs in the loss. Then formula (6) is rewritten as:

$$L_{\text{verif}} = L_s + \lambda L_d$$

$$= - \sum_{(i,j) \in S} \min(0, \log 2 - \log (1+e^{(D_s - T)})) - \lambda \sum_{(i,j) \in D} \log \max(\epsilon, e^{(D_o - T)} - 1) + 1$$

(7)

Where $\lambda = \frac{N_s}{N_o}$ is balances the importance of the positive sample loss function and the negative sample loss function, where $N_s$ is the total number of positive sample pairs in a batch and $N_o$ is the total number of negative sample pairs. 1. The model can push the distance between different classes farther ($D_s \rightarrow +\infty$). 2. This model can prevent the network from compressing the distance of similar samples too small. 3. The model expressions in this paper are all exponential, and the curve is smooth for optimization and convergence.

3.3. Person re-identification framework based on logistic regression loss

The logistic regression loss is embedded in the training of model. In this paper, a two-way twinning network is adopted. The model consists of two parts: (1) A verification model based on logistic regression loss function. (2) The classification model. The expression for the softmax loss function is:

$$L_{\text{Identif}} = -\sum_{i=1}^{m} \log \frac{e^{W_i f(x_i) + b_i}}{\sum_{j=1}^{n} e^{W_j f(x_j) + b_j}}$$

(8)

In the classification network, $m$ is the numbers of input sample in one batch. $f(x_i) \in R^d$ is the depth embedding of the $i$-th sample, its class is $y_i$, and its dimension is $d$. $W \in R^{d \times n}$ is the parameter of the last FC layer, $b \in R^d$ is the bias of this layer, and $W_j \in R^{d \times i}$ is the $j$-th row parameter of $W$. The loss of the whole model is the weighted sum of the loss of the classification models and the verification models. The expression is:

$$L = L_{\text{Identif}} + \lambda_2 L_{\text{verif}}$$

(9)

The parameter $\lambda_2$ is used to balance the weights of the two partial loss, and the parameters of the network are constrained by the two partial loss functions in the optimization.

4. Experiments

The algorithm of this paper is use in Matlab 2016a. The MatconvNet deep learning toolbox is used to verify the effect of the method in the two mainstream person re-identification databases of Market1501 and CUHK03.

4.1 Implementation details

This paper uses the ResNet50 network to verify the method, using the pre-training model obtained on the ImageNet database. In the network training, the last FC layer of the pre-training model is removed, and the value of the pool5 layer is extracted as the feature of the sample. Then using these
features for similarity metrics, all input pictures are cropped to 224*224 before entering the network to fit ResNet50. In this paper, the gradient descent method is used to optimize the model end-to-end.

4.2 Comparison with other loss

In order to compare the performance of the logistic regression loss function proposed in this paper with other mainstream loss functions, this paper uses the RestNet50 network to conduct experiments on the Market1501 database. This paper compares the loss functions commonly used in the verification model with Softmax loss [7], Contrastive loss, and Triplet loss. (V) indicates that the network only has the verification model part. It can be seen from Table 1 that under the constraints of the model, 72.52% of the rank-1 accuracy rate and 54.9% of the mAP were obtained. Compared with the performance of the other three loss functions, the performance of the logistic regression loss function can constrain the network to learn more discriminative features.

Table 1 Comparison with other loss functions

| Method             | Market1501 |
|--------------------|------------|
|                    | rank-1 | mAP  |
| Contrastive loss   | 0.7128  | 0.5073|
| Triplet loss       | 0.6736  | 0.4537|
| Softmax loss[7]    | 0.6458  | 0.4494|
| Our(V, T=0.1)      | 0.7252  | 0.549 |

4.3 Influence of distance threshold T on experimental performance

In order to illustrate the effect of the positive distance threshold T introduced in this paper on the result. Figure 3 shows the experimental results on the Market1501 database.

![Fig 2. Comparison of the results under different distance thresholds](image)

When T=0, the mAP value and rank1 obtained by the model were 49.95% and 65.16%, respectively. In reality, the illumination and angle between the positive sample images are different, so there must be a certain distance between the image features, and strictly constraining the Euclidean distance to zero will reduce the robustness of the model. When T=0.1, mAP and rank1 at this time are 54.9% and 72.52%, respectively, which have reached the highest value. The performance of mAP and rank1 is 4.95% and 7.36%, respectively, compared with the case without distance threshold. The introduction of the positive sample distance T has a large impact on the performance improvement.
4.4 Comparison with the other method

| Algorithm                  | Market1501 |   | CUHK03 |   |
|----------------------------|------------|---|--------|---|
|                            | rank-1     | mAP | rank-1 | mAP|
| LSTM Siamese[9]            | 0.616      | 0.353 | 0.504 | 0.275 |
| Gate Reid[10]              | 0.659      | 0.396 | 0.547 | 0.314 |
| SOMAnet[11]                | 0.7387     | 0.4789 | -     | 0.689 |
| SGLE(R)[12]                | 0.723      | 0.4678 | 0.732 | -    |
| ReRank[13]                 | 0.771      | 0.636 | 0.64  | 0.693 |
| SVD[13]                    | 0.823      | 0.621 | 0.818 | 0.848 |
| In Defense[15]             | 0.8492     | 0.6419 | -     | -    |
| Baseline[7]                | 0.7951     | 0.5987 | 0.834 | 0.864 |
| Our (V)                    | 0.7252     | 0.549 | 0.7486 | 0.7983 |
| Our (I+V)                  | 0.8165     | 0.6527 | 0.8573 | 0.8771 |

In order to prove the advanced nature of the method, our compares with the current advanced person re-identification algorithm in the two mainstream databases of Market1501 and CUHK03. (V) represents only the verification model, (I+V) Representing the combination of the verification model and the classification model, red represents the best performing algorithm, followed by blue and green again.

In the field of person re-ID, many mainstream algorithms combine the similarity verification model with the category classification model to significantly improve experimental performance. In order to test the fairness of the comparison, this paper compares the results of this model with the classification model and other advanced mainstream algorithms in the comparative experiment. The weight of the two-way classification loss function and the verification loss function is 1:1:0.001(λ = 0.001). In the Market1501 database, the algorithm achieved 81.6% rank-1 accuracy and 65.2% mAP, of which mAP exceeded all comparison algorithms and rank-1 ranked second. The rank-1 and mAP algorithms of this paper have a significant improvement compared with the Baseline [7] algorithm, because the logistic regression loss function proposed in this paper can compress the distance between similar samples very small, and will be between different types of samples. The distance is far away, which improves the discriminability of the features. In the CUHK03 data, the rank-1 and mAP of this paper are 85.73% and 87.39%, respectively, which is 2.33% and 1.31% higher than the Baseline [7] algorithm. Compared with advanced algorithms such as SGLE [12], ReRank [13], SVD [13] and In defense [15], the performance of this paper has improved. In addition, when the network has the constraints of the logistic loss constraint and the classification loss, its performance is much higher than that under the constraint of the logistic regression loss.

5. Conclusion

This paper proposes a logistic regression loss function based on the idea of binary logistic regression, and uses the distance threshold to correct the proposed loss function to prevent the network from constraining the distance of similar samples too small. It is then embedded into the deep network to work with the traditional softmax classification loss function to constrain the network to obtain discriminative features. In addition, this paper proposes that the loss function is not only applicable to person re-identification tasks, but also to other distance measurement problems. It is verified on the two mainstream databases, Market1501 and CUHK03.

References
[1] Zheng L, Yang Y, Hauptmann A G. Person Re-identification: Past, Present and Future. arXiv preprint arXiv:1610.02984, 2016.
[2] Xu Yuhua, Tian Zunhua, Zhang Yuqiang, et al. Body contour tracking adaptively integrates color and depth information. Journal of Automation, 2014, 40(8):1623-1634.
[3] Hu Min, Jiang He, Wang Xiaohua, et al. Expression level classification method based on geometric and texture features. Journal of Electronics, 2017, 45(1):164-172.

[4] Bazzani L, Cristani M, Perina A, et al. Multiple-Shot Person Re-identification by HPE Signature // Proceedings of 2010 International Conference on Pattern Recognition. Istanbul, 2010:1413-1416.

[5] Pedagadi S, Orwell J, Velastin S, et al. Local Fisher Discriminant Analysis for Pedestrian Re-identification // Proceedings of 2013 Computer Vision and Pattern Recognition. Portland, 2013:3318-3325.

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

[7] Zheng Z, Zheng L, Yang Y. A Discriminatively Learned CNN Embedding for Person Re-identification. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 14(1):13.

[8] Wen Y, Zhang K, Li Z, et al. A Discriminative Feature Learning Approach for Deep Face Recognition // Proceedings of 2016 European Conference on Computer Vision. Amsterdam, 2016:499-515.

[9] Varior R R, Shuai B, Lu J, et al. A Siamese Long Short-Term Memory Architecture for Human Re-identification // Proceedings of 2016 European Conference on Computer Vision. Amsterdam, 2016:135-153.

[10] Rama Varior R, Haloi M, Wang G. Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification // Proceedings of 2016 European Conference on Computer Vision. Amsterdam, 2016: 791-808.

[11] Barbosa I B, Cristani M, Caputo B, et al. Looking beyond appearances: Synthetic training data for deep CNNs in re-identification. Computer Vision and Image Understanding, 2017, 8(1):1379-1494.

[12] Cheng D, Gong Y, Li Z, et al. Deep Feature Learning via Structured Graph Laplacian Embedding for Person Re-Identification. arXiv preprint arXiv:1707.07791, 2017.

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

[14] Sun Y, Zheng L, Deng W, et al. SVDNet for Pedestrian Retrieval // Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, 2017: 3820 - 3828.

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