Reference set based metric learning method for person re-identification against overfitting problem

Chenhao Si*
Courant Institute of Mathematical Sciences, New York University, New York, USA

ABSTRACT

Similarity distance measurement is an important method in data classification, data recognition and other tasks, and has a very wide range of applications in machine learning, computer vision and other fields. However, there are model overfitting problems in complex data classification identification tasks in existing metric learning model. And those problems will negatively infect the accuracy and stability of the metric models. We study on person re-identification (person re-ID) task to design a robust similarity distance metric learning model based on a novel approach of overcoming over-fitting problem. The proposed method sets up a reference set based on training sample. Using the reference set and test images to form similar sample pairs, we can optimize the distribution information and projection feature. Finally, by testing on benchmark dataset, VIPeR, the experimental results validate the effectiveness of the proposed method. It achieves the best identification rates.

Keywords: Metric learning, LDA, distribution distance, reference set

1. INTRODUCTION

Similarity Distance Measurement is a major research area in machine learning and computer vision. In recent years, computer vision tasks have become the hottest topic of technological development in today's society. Traditional Similarity Distance Measurement is mainly studied for the distance metric between image feature vectors to find a discriminate and robust metric model to measure the similarity between two samples. However, traditional similarity distance measurement methods usually design a metric model based on the feature vector of one-shot image, using the label information of the data to learn a metric subspace so that the distance of sample pair with same label is as small as possible and the distance of sample pair with different label is as large as possible. Compared with the tasks of image matching and recognition in traditional computer vision, these problems application scenarios are more complex, and the targets in the images have serious misalignment problems, which leads to extremely unstable image features and brings a great challenge to the similarity distance metric of images.

2. RELATED WORK

Person re-identification is a hot topic and attracts a large number of researchers study on this topic[1-5]. Person re-recognition is very complex and the target's appearance is affected by various factors, including camera angle, gait, background, lighting, etc. Existing models is of poor identification accuracy on test data[4]. Thus, person re-identification task is very challenging. The researchers utilized various methods to improve the identification accuracy of existing methods.

For metric learning based method, researchers target metric learning algorithms by learning a robust and discriminative metric model. Weinberger et al.[6] designed a metric learning algorithm based on the *k*-nearest-neighbor to learn a Mahalanobis distance. This method learns a metric model that draws closer the *k*-nearest neighbor samples with the same label, while expanding the distance of sample pair with different labels. Davis et al.[7] build an optimization model by establishing two Gaussian distributed differential relative entropy functions to learn a novel distance model for pedestrian image similarity measurement. Guillaumin et al.[8] modelled logistic discrimination based metric method such that the similarity of positive pair based on this metric model are smaller than that of negative pair. Zheng et al.[9], on the other hand, established a projection subspace learning method for relative distance comparison and probabilized the model. The model learns a metric subspace by maximizing the probability that the positive samples and the negative samples are correctly recognized. Koestinger et al.[10], on the other hand, use the idea of hypothesis testing to establish a

* tiger59888@163.com

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125066G
 · doi: 10.1117/12.2662666

likelihood ratio model based on the distribution probabilities of positive and negative samples to learn a Mahalanobis distance. Moreover, Liao et al.[11], learn a projection subspace based on a quadratic discriminant analysis based on a sample pair difference vector space. In addition, the KISSME algorithm is used to learn a distance function for the projected samples. Zhao et al.[12] introduced the kernelization approach and proposed a kernelized stochastic KISS algorithm, which enhances the Gaussian distribution characteristics of the training data by kernelizing the original features of training samples. While, these metric learning algorithms have poor generalization ability and low accuracy of testing results.

In recent years, the deep learning methods have been wildly applicated in computer vision tasks and achieved very good results. Li et al.[13] established a person re-recognition network model (FPNN) based on 'Siamese' network structure, which is an early and very representative network for person re-recognition, but its recognition accuracy is very low. Ahmed et al.[14] carried out an improvement on the FPNN person re-recognition network model for discriminative feature extraction. Ustinova et al.[15] built a local feature-based network model using the characteristics of body structure. Liu et al.[16] introduced adversarial generative network (GAN) and attention mechanism to build a network model with the ability to fight against small local interference. However, deep networks need a large amount of data for model training and has limited generalization ability among different surveillance networks, requiring a large amount of data labeled for each surveillance network for model optimization. Deep network-based person re-identification methods have major drawbacks in various detection, recognition and prediction applications.

3. METRIC LEARNING METHOD FOR PERSON RE-IDENTIFICATION

In this problem, the similarity between pedestrian images is measured by learning a Mahalanobis distance model. Given two sample sets: $A=\{x_1,x_2,\cdots,x_i,\cdots,x_N\}$, where $\mathbf{x}_i$ indicates the appearance feature of the sample of the *i*-th pedestrian under camera A; $B=\{z_1,z_2,\cdots,z_j,\cdots,z_N\}$, where $\mathbf{z}_j$ denotes indicates the appearance feature of the sample of the *j*-th pedestrian under camera B. The mathematical expression of the model is as follows

$$d(x_i,z_j)=(x_i-z_j)^T M(x_i-z_j) \tag{1}$$

where $M \geq 0$ is a positive semi-definite matrix. x_i, z_j represent the feature vectors of two samples from different camera view, respectively. Moreover, M is the metric matrix of distance model. In order to learn a discriminative metric subspace, Liao et al.[11] first learn a set of projection features and then learn a Mahalanobis distance by KISSME[5] method.

$$d\left(x_i,z_j\right)=\left(x_i-z_j\right)^T W\left(\Sigma_I^{-1}-\Sigma_E^{-1}\right)W^T\left(x_i-z_j\right) \tag{2}$$

where x_i-z_j represents the difference feature vectors. The model of XQDA11 method is established as follows:

$$\max J(w)=\frac{w^T S_b w}{w^T S_w w} \quad \Leftrightarrow \quad \max J(w)=w^T S_b w;\;\; w^T S_w w=1 \tag{3}$$

According to the related properties of the Rayleigh Entropy, the largest value of the function in equation (3) above is the largest eigenvalue of matrix $S_w^{-1}S_b$ and the solution is corresponding eigenvector w_1. We use the first *r* eigenvectors to obtain a set of projection features by $W=\left(w_1,w_2,\cdots,w_r\right)$.

4. IMPROVEMENT OF METRIC LEARNING ALGORITHM BASED ON SIMILAR SAMPLE CONSTRAINTS

In this task, there is a serious overfitting problem. The projection subspace realizes the separation of positive and negative samples by finding a series of feature transformation patterns. To overcome the over-fitting problem, we use the reference set information to establish a pairwise similarity distance metric learning model against over-fitting problem for complex scenes.

$A' = \{x'_1, x'_2, \cdots, x'_i, \cdots, x'_{N'}\}$ and $B' = \{z'_1, z'_2, \cdots, z'_j, \cdots, z'_{N'}\}$, denote the sample sets under different camera views of the test data, respectively. Then, by using the metric model introduced above, we find the most similar k samples in the reference set for each individual. These k samples are used to calculate the mean difference feature vector to estimate the center of positive pairs' difference vector:

$$v_i = \frac{1}{k}\sum_{i=1}^{N} q_{ij}\alpha_{ij}\left(x'_i - z'_j\right) \tag{4}$$

where $q_{ij} = 0 \ or \ 1$. q_{ij} =1 denotes $\left(x'_i, z'_j\right)$ is similar sample of k nearest neighbors. q_{ij} =0 denotes $\left(x'_i, z'_j\right)$ is dissimilar sample. α_{ij} denotes the sample's weight over $\left(x'_i, z'_j\right)$, defined as follows：

$$\alpha_{ij} = 1 - d(x'_i, z'_j) \Big/ \sum q_{ij} d(x'_i, z'_j) \tag{5}$$

Then, we calculate the center of the training data positive sample difference vector: $u = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - z_i\right)$.

The distribution constraint of positive samples' difference feature vectors between test data and training data is defined as follow,

$$S = \frac{1}{N'}\sum_{i=1}^{N'}\left\| w^T v_i - w^T u \right\|_2^2 = \frac{1}{N'}\sum_{i=1}^{N'} w^T\left(v_i - u\right)\left(v_i - u\right)^T w = \frac{1}{N'} w^T S' w \tag{6}$$

where S denotes the distance between distribution of positive sample pairs in the test data and in the training data. This distance is used to limit the distance between the identified pairs and the positive pairs.

Define $C = B$ as a reference set, where B is the data collected under camera B in the training data. Then, the similarity distance between each individual x'_i in the test data and each sample in the reference set C is calculated. Samples from Reference set C and test instances from A' will form negative sample pairs. We define the distance between the similar sample pairs in the reference set and the positive sample pairs as follows:

$$D = \left\| w^T\left(\frac{1}{N_{A'}N_C}\sum_{i=1}^{N_{A'}}\sum_{j=1}^{N_C} p_{ij}\left(x'_i - z_j\right)\right)\right\|_2^2 \tag{7}$$

where $\frac{1}{N_{A'}N_C}\sum_{i=1}^{N_{A'}}\sum_{j=1}^{N_C} p_{ij}\left(x'_i - z_j\right)$ is the similar sample pair difference vector distribution center of test individuals in the reference set. p_{ij} denotes the matching relationship between x_s and z_t. $x'_i \in A'$, $z_j \in C$. p_{ij} =1 denotes $\left(x'_i, z_j\right)$ is a sample that is similar to less than (negative sample), and p_{st} =0 denotes $\left(x_s, z_t\right)$ is a pair of similar samples of the k' nearest neighbors. We use the metric model in Section 3 to calculate the similarity between samples of different camera views. The samples with the highest similarity are treated as similar samples.

$$\begin{aligned} D &= \left\| w^T\left(\frac{1}{N_{A'}N_C}\sum_{i=1}^{N_{A'}}\sum_{j=1}^{N_C} p_{ij}\left(x'_i - z_j\right)\right) - 0\right\|_2^2 = w^T\left(\frac{1}{N_{A'}N_C}\sum_{i=1}^{N_{A'}}\sum_{j=1}^{N_C} p_{ij}\left(x'_i - z_j\right)\right)\left(\frac{1}{N_{A'}N_C}\sum_{i=1}^{N_{A'}}\sum_{j=1}^{N_C} p_{ij}\left(x'_i - z_j\right)\right)^T w \\ &= w^T D' w \end{aligned} \tag{8}$$

where D' indicates the distance from the center of similar sample pairs in the reference set to the center of the theoretical distribution of positive samples. The distance in equation (6) is introduced as a constraint to the learning of the projection subspace in the second part, so as to improve the generalization ability of the metric model to the test data. The mathematical expression of improved metric learning model is as introduced in equation (9):

$$\max J(w) = \frac{w^T S_b w}{w^T \left(S_w + D + S\right) w} \tag{9}$$

The improved metric learning model is transformed to solve the following problem,

$$\left(S_w + D + S\right)^{-1}\left(S_b\right) w = \lambda w \tag{10}$$

The improved metric model is obtained by solving the eigenvalue problem in equation (10). It is worth noting that during the computation, we train the metric model separately for each test individual. Because the test individuals are individual-specific, the characteristics of each test individual vary.

5. EXPERIMENT

To validate the performance of our model, the identification rates on VIPeR[3] dataset is summarized. For the evaluation of the accuracy of algorithms for this task, the widely used cumulative accuracy curve (CMC)[5] is used to evaluate the performance of our model.

The identification results are compared to some first-class methods to further prove the improvement effect of our novel model. The results are displayed in Figure 1 and Table 1. The identification results of rank-1, 5, 10, and 20 are listed in Table 1. According to the results in this table, it can be clearly seen that the algorithm in this paper has achieved the best recognition accuracy on all indicators. Especially in the recognition accuracy of rank-1, compared to the best comparison model, OSNet, rank-1 accuracy is improved by 3.4%. The recognition accuracy of rank-1, rank-5, rank-10 and rank-20 is improved to 71.4%, 93.6%, 98.2% and 99.8%, respectively. Compared to the baseline comparison algorithm XQDA, the recognition accuracy of the rank-1 of our model is reised by 30.8%. The recognition accuracy improvement effect is very significant.

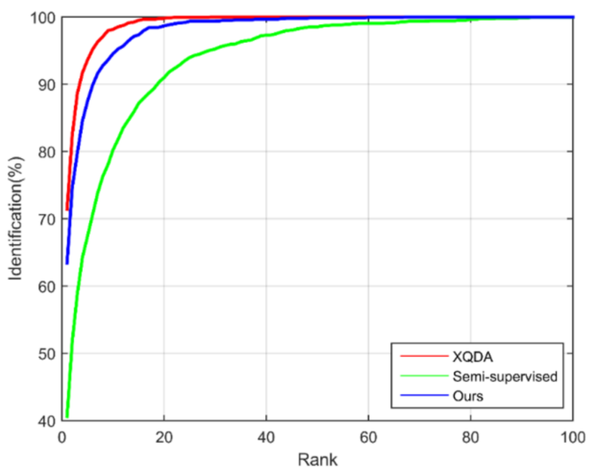

Figure 1. Comparison of identification accuracy of different methods on VIPeR dataset.

Table 1. Comparison results on VIPeR dataset.

Methods	r=1	r=5	r=10	r=20	Methods	r=1	r=5	r=10	r=20
KISSME[10]	18.7	47.2	61.7	75.6	Improved deep[14]	34.8	63.5	75	80
SCML[17]	40.6	-	-	-	SCIR[22]	35.8	-	-	-
XQDA-LOMO[11]	40.6	68.3	80.5	91.8	TCP[23]	47.8	74.7	84.8	91.1
NFST-LOMO[18]	42.3	71.5	82.9	92.1	MCK-CCA[24]	47.2	-	87.3	94.7
LSSCDL-LOMO[19]	42.7	-	84.3	91.9	KEPLER[25]	42.4	-	82.4	90.7

Methods	r=1	r=5	r=10	r=20	Methods	r=1	r=5	r=10	r=20
MLAPG-LOMO[20]	40.7	69.9	82.3	92.4	HRNet[26]	48.7	73.4	81.7	-
IPMLLSL[21]	46.5	69.3	80.7	86.5	HydraPlus-Net[27]	56.6	-	-	-
Semi-supervised	63.4	87.5	94.6	98.7	OSNet[28]	68.0	-	-	-
Ours-LOMO	**71.4**	**93.6**	**98.2**	**99.8**					

6. CONCLUSION

Overfitting is a common phenomenon that affects the accuracy of algorithms in machine learning tasks. We take person re-identification problem as the object and studies similarity measurement model against the overfitting problem. Person re-identification technology is based on the similarity measurement of pedestrian image pairs, which is more complicated than the traditional data point classification problem and cannot be constrained by the direct sample population distribution distance. In this paper, similar samples of individuals are tested to form pseudo positive pairs. The optimization of the metric model is guided by constraining the difference between the test individual's positive pairs and the theoretical center of probability distribution. Finally, through extensive comparative experiments and parameter analysis, the effectiveness of our model is verified.

REFERENCE

[1] Liu, X., Tao, D., Song, M., Zhang L., Bu, J. and Chen, C., "Learning to track multiple targets," IEEE Transactions on Neural Networks & Learning Systems 26(5), 1060 (2015).

[2] Gray, D. and Tao, H., "Viewpoint invariant pedestrian recognition with an ensemble of localized features," European Conference on Computer Vision, 262-275 (2008).

[3] Du, Y., Ai, H. and Lao, S., "Evaluation of color spaces for person re-identification," IEEE Conference on Pattern Recognition, Tsukuba, 1371-1374 (2012).

[4] Prosser, B., Zheng, W., Gong, S. and Xiang, T., "Person re-identification by support vector ranking," British Machine Vision Conference, 231-242 (2010).

[5] Hirzer, M., Roth, P. M., Kostinger, M. and Bischof, H., "Relaxed pairwise learned metric for person re-identification," European Conference on Computer Vision, 780-793 (2012).

[6] Weinberger, K. Q. and Saul, L. K., "Distance metric learning for large margin nearest neighbor classification," The Journal of Machine Learning Research 10(1), 207-244 (2009).

[7] Davis, J. V., Kulis, B., Jain, P., Sra, S. and Dhillon, I. S., "Information-theoretic metric learning," International Conference on Machine Learning, 209-216 (2007).

[8] Guillaumin, M., Verbeek, J. and Schmid, C., "Is that you? Metric learning approaches for face identification," International Conference on Computer Vision, 498-505 (2009).

[9] Zheng, W. S., Gong, S. and Xiang, T., "Person re-identification by probabilistic relative distance comparison," Conference Computer Vision. Pattern Recognition, 649-656 (2011).

[10] Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P. and Bischof, H., "Large scale metric learning from equivalence constraints," IEEE Conference on Computer Vision and Pattern Recognition, 2288-2295 (2012).

[11] Liao, S., Hu, Y., Zhu, X. and Li, S. Z., "Person re-identification by local maximal occurrence representation and metric learning," IEEE Conference on Computer Vision and Pattern Recognition, 2197-2206 (2015).

[12] Zhao, C., Chen, Y., Wang, X., Wong, W. K., Miao, D. and Lei, J., "Kernelized random KISS metric learning for person re-identification," Neurocomputing 275, 403-417 (2018).

[13] Li, W., Zhao, R., Xiao, T. and Wang, X., "Deepreid: Deep filter pairing neural network for person re-identification," IEEE Conference on Computer Vision and Pattern Recognition, 152-159 (2014).

[14] Ahmed, E., Jones, M. and Marks, T. K., "An improved deep learning architecture for person re-identification," IEEE Conference on Computer Vision and Pattern Recognition, 3908-3916 (2015).

[15] Ustinova, E., Ganin, Y. and Lempitsky, V., "Multi-region bilinear convolutional neural networks for person re-identification," International Conference on Advanced Video and Signal Based Surveillance, 1-6 (2017).

[16] Liu, A., Liu, X., Fan, J., et al., Perceptual-sensitive GAN for generating adversarial patches," Proceedings of the AAAI Conference on Artificial Intelligence, 1028-1035 (2019).

[17] Feng, Y., Yuan, Y. and Lu, X., "Person reidentification via unsupervised cross-view metric learning," IEEE Transactions on Cybernetics, (2019).

[18] Zhang, L., Xiang, T. and Gong, S., "Learning a discriminative null space for person re-identification," IEEE Conference on

Computer Vision and Pattern Recognition, 1239-1248 (2007).
[19] Zhang, Y., Li, B., Lu, H., Irie, A. and Ruan, X., "Sample-specific svm learning for person re-identification," IEEE Conference on Computer Vision and Pattern Recognition, 1278-1287 (2016).
[20] Liao, S. and Li, S. Z., "Efficient psd constrained asymmetric metric learning for person re-identification," IEEE Conference on Computer Vision and Pattern Recognition, 3685-3693 (2015).
[21] Zhao, Z., Zhao, B. and Su, F., "Person re-identification via integrating patch-based metric learning and local salience learning," Pattern Recognition 75, 90-98 (2017).
[22] Lin, T. Y., Roychowdhury, A. and Maji, S., "Bilinear convolutional neural networks for fine-grained visual recognition," IEEE Transactions on Pattern Analysis & Machine Intelligence 40(6), 1309-1322 (2017).
[23] Cheng, D., Gong, Y., Zhou, S., Wang, J. and Zheng, N., "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," IEEE Conference on Computer Vision and Pattern Recognition, 1335-1344 (2016).
[24] Lisanti, G., Karaman, S. and Masi, I., "Multichannel-kernel canonical correlation analysis for cross-view person reidentification," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 13(2), 13 (2017).
[25] Zhao, Z., Zhao, B. and Su, F., "Person re-identification via integrating patch-based metric learning and local salience learning," Pattern Recognition 75, 90-98 (2017).
[26] Zhang, G., Ge, Y., Dong, Z., Wang, H., Zheng, Y. and Chen, S., "Deep high-resolution representation learning for cross-resolution person re-identification," in arXiv preprint arXiv: 2105.11722, (2021).
[27] Liu, X., Zhao, H., Tian, M., Sheng, L., Shao, J., Yi, S., Yan, J. and Wang, X., "Hydraplus-net: Attentive deep features for pedestrian analysis," IEEE International Conference on Computer Vision, 350-359 (2017).
[28] Zhou, K., Yang, Y., Cavallaro, A. and Xiang, T., "Learning generalisable omni-scale representations for person re-identification," IEEE Transactions on Pattern Analysis and Machine Intelligence 44(9), 5056-5069 (2021).