A Hybrid Deep Model for Person Re-Identification

Di Wu[1], Si-Jia Zheng[1], Fei Cheng[2], Yang Zhao[2], Chang-An Yuan[3],
Xiao Qin[3], Yong-Li Jiang[4], and De-Shuang Huang[1(✉)]

[1] Institute of Machine Learning and Systems Biology, School of Electronics
and Information Engineering, Tongji University, Shanghai, China
wudi_qingyuan@163.com, zhengsj9878@163.com,
dshuang@tongji.edu.cn
[2] Beijing E-Hualu Info Technology Co., Ltd., Beijing, China
[3] Science Computing and Intelligent Information Processing of Guang Xi Higher
Education Key Laboratory, Guangxi Teachers Education University,
Nanning 530001, Guangxi, China
[4] Ningbo Haisvision Intelligence System Co., Ltd., Ningbo 315000,
Zhejiang, China

Abstract. In this study, we present a hybrid model that combines the advantages of the identification, verification and triplet models for person re-identification. Specifically, the proposed model simultaneously uses Online Instance Matching (OIM), verification and triplet losses to train the carefully designed network. Given a triplet images, the model can output the identities of the three input images and the similarity score as well as make the L_{-2} distance between the mismatched pair larger than the one between the matched pair. Experiments on two benchmark datasets (CUHK01 and Market-1501) show that the proposed method can achieve favorable accuracy while compared with other state of the art methods.

Keywords: Convolutional neural networks · Person re-identification
Identification model · Verification model · Triplet model

1 Introduction

As a basic task of multi-camera surveillance system, person re-identification aims to re-identify a query pedestrian observed from non-overlapping cameras or across different time with a single camera [1]. Person re-identification is an important part of many computer vision tasks, including behavioral understanding, threat detection [2–6] and video surveillance [7–9]. Recently, the task has drawn significant attention in computer vision community. Despite the researchers make great efforts on addressing this issue, it remains a challenging task due to the appearance of the same pedestrian may suffer significant changes under non-overlapping camera views.

Recently, owing to the great success of convolution neural network (CNN) in computer vision community [10–14], many CNN-based methods are introduced to address the person re-identification issue. These CNN-based methods achieve many

promising performances. Unlike the hand-craft methods, CNN-based methods can learn deep features automatically by an end-to-end way [28]. These CNN-based methods for person re-identification can be roughly divided into three categories: identification models, verification models and triplet models. The three categories of models differ in input data form and loss function and have their own advantages and limitations. Our motivation is to combine the advantages of three models to learn more discriminative deep pedestrian descriptors.

The identification model treats the person re-identification issue as a task of multi-class classification. The model can make full use of annotation information of datasets [1]. However, it cannot consider the distance metric between image pairs. Verification model regards the person re-identification problem as a binary classification task, it takes paired images as input and outputs whether the paired images belong to the same person. Thus, verification model considers the relationship between the different images, but it does not make full use of the label information of datasets. As regards to the triplet model [15], it takes triplet unit (x_i, x_j, x_k) as input, where x_i and x_k are the mismatched pair and the x_i and x_j are the matched pair. Given a triplet images, the model tries to make the relative distance between x_i and x_k larger than the one between x_i and x_j. Accordingly, triplet-based model can learn a similarity measurement for the pair images, but it also cannot take advantage of the label information of the datasets and uses weak labels only [1].

We design a CNN-based architecture that combines the three types of popular deep models used for person re-identification, i.e. identification, verification and triplet models. The proposed architecture can jointly output the IDs of the input triplet images and the similarity score as well as force the L2 distance between the mismatched pair larger than the one between the matched pair, thus enhancing the discriminative ability of the learned deep features.

2 Proposed Method

As shown in Fig. 1, the proposed method is a triplet-based CNN model that combines identification, verification and triplet losses.
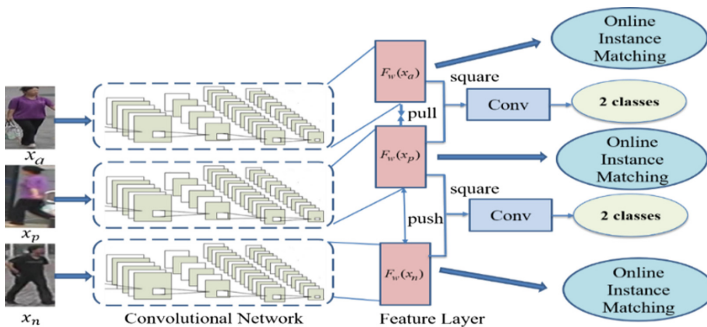

Fig. 1. The proposed model

2.1 Loss Function

Identification Loss. In this work, we utilize the Online Instance Matching [16] to instead of the cross-entropy loss for supervising the identification submodule. For one image M, the similarity score of it belong to the ID j is written as:

$$p_j = \frac{\exp(v_j^T M/\partial)}{\sum_{i=1}^{L} \exp(v_i^T M/\partial) + \sum_{k=1}^{Q} \exp(\mu_k^T M/\partial)} \tag{1}$$

where v_j^T is the transposition of the j-th column of lookup table, μ_k^T represents the transposition of the k-th ID of circular queue. ∂ is the temperature scalar.

Verification Loss. For the verification subnetwork, we treat person re-identification as a task of binary classification issue. Similar to identification model, we also adopt cross-entropy loss as identification loss function, which is:

$$q = soft \max(\text{conv}),$$
$$L_2 = \sum_{i=1}^{2} -n_i \log(q_i) \tag{2}$$

in which n_i is the labels of paired image, when the pair is the same pedestrian, $n_1= 1$ and $n_2= 0$; otherwise, $n_1= 0$ and $n_2= 1$.

Triplet Loss. The triplet subnetwork is adopted to make the Euclidean distance between the positive pairs smaller than that between negative pairs. For one triplet unit R_i, the triplet loss function can be written as:

$$L_3 = [thre + d(F_w(R_i^o), F_w(R_i^+)) - d(F_w(R_i^o), F_w(R_i^-))]_+ \tag{3}$$

where *thre* is a threshold value and is a positive number, $[x]+ = \max(0, x)$, and $d\,()$ is Euclidean distance.

Hybrid Loss. The deep architecture is jointly supervised by three OIM losses, two cross-entropy losses and one triplet loss. During the training phase, the hybrid loss function can be written as:

$$L = \alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3 \tag{4}$$

in which α_1, α_2 and α_3 are the balance parameters.

2.2 Training Strategies

We utilize the pytorch framework to implement our network. In this work, we use the Adaptive Moment Estimation (Adam) as the optimizer of the deep model. We use two types of training strategies proposed by [17]. Specially, for large-scale dataset like Market-1501, we use the designed model directly to transfer on its training set. As for

the small datasets (e.g. CUHK01), we first train the model on the large-scale person re-identification dataset (e.g. Market-1501), then fine-tune the model on the small dataset.

3   Experiments

We first resize the training images into 256*128, then the mean image is subtracted by those resized training images. For our hybrid model, it is crucial for it to organize the mini-batch that can satisfy the training purpose of both identification, verification and triplet subnetworks. In this study, we use the protocol proposed by, we sample Q identities randomly, and then all the images R in the training set are selected. Finally, we use QR images to constitute one mini-batch. Among these QR images, we choose the hardest positive and negative sample for each anchor to form the triplet units. And we randomly selected 100 paired images for verification training. As to the identification subnetwork, we use all QR images in the mini-batch for training.

(1) **Evaluation on CUHK01**

For this dataset, 485 pedestrians are randomly selected to form the training set. The remainder 486 identities are selected for test. From Table 1, we can observe that the proposed method beats the most of compared approaches, which demonstrates the effectiveness of the proposed method.

Table 1. Fig. 1. Table 1. Results (Rank1, Rank5 and Rank10 matching accuracy in %) on the CUHK01 dataset. '-' means no reported results is available.

Method	Rank-1	Rank-5	Rank-10
Deep Ranking [18]	50.41	75.93	84.07
GOG [19]	57.80	79.10	86.20
DGD [20]	66.60	-	-
Quadruplet [21]	62.55	83.44	89.71
Deep Transfer [17]	**77.00**	-	-
CNN-FRW-IC [22]	70.50	90.00	94.80
Our	76.84	**95.60**	**96.12**

(2) **Evaluation on Market-1501**

We compare the proposed model with eight state-of-the-art methods on Market-1501 dataset. We report the performances of mean average precision (mAP), Rank-1 and Rank-5. Both the results are based on single-query evaluation. From Table 2, it can be observed that the accuracies of mAP, Rank-1 and Rank-5 of the proposed model achieve 77.43%, 91.60% and 98.33%, respectively, and our method beats all the other competing methods, which further proves the effectiveness of the proposed method.

Table 2. Results (mAP, Rank-1 and Rank-5 matching accuracy in %) on the Market-1501 dataset in the single-shot. '-' means no reported results is available.

Method	mAP	Rank-1	Rank-5
Deep Transfer [17]	65.50	83.70	-
CAN [23]	35.90	60.30	-
CNN Embedding [24]	59.87	79.51	90.91
CNN + DCGAN [25]	56.23	78.06	-
IDE(R) + ML(re-ra) [26]	63.63	77.11	-
TriNet [27]	69.14	84.92	94.21
Our	**77.43**	**91.60**	**98.33**

4   Conclusions

In this paper, we design a hybrid deep model for the person re-identification. The proposed model combines the identification, verification and triplet losses to handle the intra/inter class distances. Through the hybrid strategy, the model can learn a similarity measurement and the discriminative features at the same time. The proposed model outperforms most of the state-of-the-art on the CUHK01 and Market-1501 datasets.

References

1. Zheng, L., et al.: Person Re-identification: Past, Present and Future. Computer Vision and Pattern Recognition (2016). arXiv: 1610.02984
2. Huang, D.S.: A constructive approach for finding arbitrary roots of polynomials by neural networks. IEEE Trans. Neural. Netw. **15**(2), 477–491 (2004)
3. Huang, D.S., et al.: A case study for constrained learning neural root finders. Appl. Math. Comput. **165**, 699–718 (2005)
4. Huang, D.S., et al.: Zeroing polynomials using modified constrained neural network approach. IEEE Trans. Neural Netw. **16**, 721–732 (2005)
5. Huang, D.S., et al.: A new partitioning neural network model for recursively finding arbitrary roots of higher order arbitrary polynomials. Appl. Math. Comput. **162**, 1183–1200 (2005)
6. Huang, D.S., et al.: Dilation method for finding close roots of polynomials based on constrained learning neural networks ☆. Phys. Lett. A **309**, 443–451 (2003)
7. Huang, D.S.: The local minima-free condition of feedforward neural networks for outer-supervised learning. IEEE Trans. Syst. Man Cybern. Part B Cybern. **28**, 477 (1998). A Publication of the IEEE Systems Man & Cybernetics Society

8. Huang, D.S.: The united adaptive learning algorithm for the link weights and shape parameter in RBFN for pattern recognition. Int. J. Pattern Recognit. Artif. Intell. **11**, 873–888 (1997)

9. Huang, D.S., Ma, S.D.: Linear and nonlinear feedforward neural network classifiers: a comprehensive understanding: journal of intelligent systems. J. Intell. Syst. **9**, 1–38 (1999)

10. Huang, D.S., Du, J.X.: A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. IEEE Trans. Neural Netw. **19**, 2099 (2008)

11. Wang, X.F., Huang, D.S.: A novel density-based clustering framework by using level set method. IEEE Trans. Knowl. Data Eng. **21**, 1515–1531 (2009)

12. Shang, L., et al.: Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network. Neurocomputing **69**, 1782–1786 (2006)

13. Zhao, Z.Q., et al.: Human face recognition based on multi-features using neural networks committee. Pattern Recogn. Lett. **25**, 1351–1358 (2004)

14. Huang, D.S., et al.: A neural root finder of polynomials based on root moments. Neural Comput. **16**, 1721–1762 (2004)

15. Ding, S., et al.: Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn. **48**, 2993–3003 (2015)

16. Xiao, T., et al.: Joint detection and identification feature learning for person search. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3376–3385. IEEE (2017)

17. Geng, M., et al.: Deep transfer learning for person re-identification (2016)

18. Chen, S.Z., et al.: Deep ranking for person re-identification via joint representation learning. IEEE Trans. Image Process. **25**, 2353–2367 (2016)

19. Matsukawa, T., et al.: Hierarchical Gaussian Descriptor for Person Re-identification. In: Proceedings of the Computer Vision and Pattern Recognition, pp. 1363–1372 (2016)

20. Xiao, T., et al.: Learning deep feature representations with domain guided dropout for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1249–1258 (2016)

21. Chen, W., et al.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: Proceedings of the CVPR, pp. 1320–1329 (2017)

22. Jin, H., et al.: Deep person re-identification with improved embedding and efficient training. In IEEE International Joint Conference on Biometrics, pp. 261–267 (2017)

23. Liu, H., et al.: End-to-End comparative attention networks for person re-identification. IEEE Trans. Image Process. **26**(7), 3492–3506 (2017). A Publication of the IEEE Signal Processing Society

24. Zheng, Z., et al.: A discriminatively learned CNN embedding for person re-identification. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **14**(1), 13 (2016)

25. Zheng, Z., et al.: Unlabeled samples generated by GAN improve the person re-identification baseline in vitro, 3774–3782 (2017). arXiv preprint arXiv:1701.07717

26. Zhong, Z., et al.: Re-ranking person re-identification with k-reciprocal encoding. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3652–3661 (2017)

27. Hermans, A., et al.: In Defense of the Triplet Loss for Person Re-Identification (2017). arXiv preprint, arXiv:1703.07737

28. Huang, D.S.: Systematic Theory of Neural Networks for Pattern Recognition, Publishing House of Electronic Industry of China, May 1996