

基于深度学习的行人重识别技术的研究进展

董亚超 刘宏哲 包 俊

北京联合大学北京市信息服务工程重点实验室 北京 100101

(18810443006@163.com)

摘 要 行人重识别技术广泛应用于辅助驾驶、公共安全、智能医疗、工业检测等多个领域。行人重识别是指在跨区域的摄像头即不同摄像头视角下中匹配同一个目标人物的过程。首先简单介绍使用传统方法的局限性,其次列举一些使用深度学习方法解决行人重识别的数据集,再次详细的总结一下几种典型的解决行人重识别的深度学习方法,如基于表观特征、基于度量学习的方法,对于跨域、泛化能力方面,对于基于特征学习解决遮挡、姿势变换等问题方面,最后展望行人重识别的发展趋势和发展方向。

关键词: 行人重识别;深度学习;跨摄像头;卷积神经网络;计算机视觉

中图法分类号 TP311

Research Progress of Pedestrian Re-identification Technology Based on Deep Learning

DONG Ya-chao, LIU Hong-zhe and BAO Jun

Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China

Abstract Pedestrian re-identification technology is widely used in various fields such as assisted driving, public safety, intelligent medical treatment, and industrial inspection. Pedestrian re-identification refers to the process of matching the same target person in cross-region cameras, that is, under different camera angles. Firstly, this paper briefly introduces the limitations of traditional methods. Secondly, it enumerates some data sets that use deep learning methods for pedestrian re-identification, and again summarizes in detail several typical deep learning methods for pedestrian re-identification, such as methods based on apparent features and metric learning, with regard to cross-domain and generalization capabilities, and for solving problems such as occlusion and posture transformation based on feature learning. Finally, this paper looks forward to the development trend and direction of pedestrian re-identification.

Keywords Person re-identification, Deep learning, Cross-camera, Convolutional neural network, Computer vision

1 引言

行人重识别旨在从不同摄像头视角匹配同一个

目标人物的过程,其在交通、公共安防和视频监控中起着重要作用。该技术与人脸识别类似。一个人的身份可以通过其人脸、身高、衣着和动作等各种信息

基金项目:国家自然科学基金(61871039,61802019,61906017);北京市属高校高水平教师队伍建设支持计划项目(IDHT20170511);北京联合大学领军人才项目(BPHR2019AZ01);北京市教委项目(KM201911417001);国家科技支撑计划项目(2015BAH55F03);智能驾驶大数据协同创新中心(CYXC1902);北京联合大学项目(WZ10201903)

This work was supported by the National Natural Science Foundation of China(61871039,61802019,61906017), Supporting Plan for Cultivating High Level Teachers in Colleges and Universities in Beijing(IDHT20170511), Premium Funding Project for Academic Human Resources Development in Beijing Union University(BPHR2019AZ01), Beijing Municipal Commission of Education Project(KM201911417001), National Key Technology R&D Program(2015BAH55F03), Big Data Collaborative Innovation Center for Intelligent Driving(CYXC1902) and Project of Beijing Union University(WZ10201903).

通信作者:刘宏哲(liuhongzhe@bnu.edu.cn)

来识别,但在一般的监控系统中存在大量的远角监控,摄像头分辨率较低,很难捕捉到清晰的面部信息,因此基于整个行人的重识别技术也非常具有研究价值。在自然场景中,监控视频中存在众多混杂的背景信息,因此需要使用目标检测技术对行人进行定位;然后使用有区分力的特征对行人进行特征表示;最后使用相似性度量,对目标人物和候选人物的特征进行相似性比较,判断这些特征是否属于同一身份。

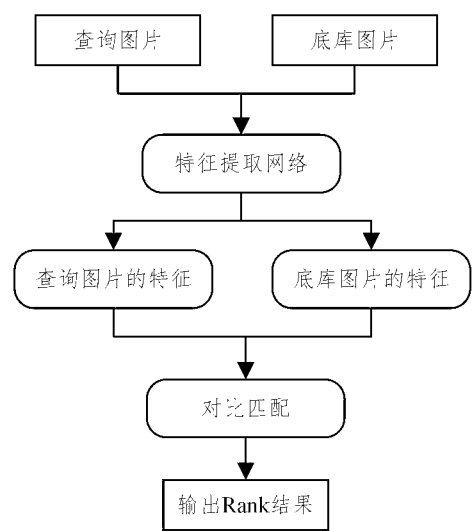


图1 行人重识别基本流程图

Fig. 1 Basic flow chart of pedestrian re-identification

近年来,尤其是随着深度学习在计算机视觉领域取得巨大成就,在深度学习技术出现之前,早期的行人重识别研究主要集中于如何手工制作出更好的

视觉特征和如何学习更好的相似度量。随着深度学习的发展,深度学习技术在行人重识别任务中得到了广泛的应用,图1为行人重识别基本流程图。

2 行人重识别发展的相关工作

行人重识别可以应用到刑事侦查、视频监控、行为理解等多个方面,其在学术界的研究最早可追溯到跨摄像头多目标跟踪问题上。行人重识别任务主要包含特征提取和相似度量两个步骤。传统方法的思路为手工提取一些低级的图像特征,如颜色、局部描述符、人体姿势、HOG (Histogram of Oriented Gradient)、SIFT (Scale Invariant Feature Transform)、LOMO (Local Maximal Occurrence) 等。然后利用 XQDA (Cross-view Quadratic Discriminant Analysis) 或者 KISSME (Keep it Simple and Straight Forward Metric Learning) 来学习最佳的相似度量。然而在数据量较大和场景复杂的情形下,一些传统的方法会很吃力,表现效果差。近几年使用的深度学习方法中,主要侧重于基于表观特征的方法,基于度量学习的方法,行人重识别跨域、泛化能力方面,基于特征学习解决遮挡、姿势变换、光照等问题这四大方面。

3 行人重识别相关数据集介绍及评价指标

3.1 行人重识别常用数据集

表1列出常用的数据集。

表1 Market1501,DukeMTMC-reID 和 CUHK03 数据集具体信息

Table 1 Statistics of Market1501,DukMTMC-reID and CUHK03 datasets

Dataset	Camera ID	Person ID	BBox	Training	Testing	Query
Market1501	6	1 501	32 668	12 936	19 732	3 368
DukeMTMC-reID	8	1 404	36 411	16 522	17 661	2 228
CUHK03	6	1 467	14 097/14 096	7 365/7 368	5 332/5 328	1 400

3.2 行人重识别常用评价指标

首位命中率(Rank-1)、平均准确率均值(mean Average Precision,mAP)和累积匹配特征曲线(Cumulative Matching Curve,CMC)是行人重识别任务中常用的评估标准,其中 CMC 常以 Rank-k 命中率的形式来体现,含义为在底库图像集合中与所查询图像匹配相似度最高的前 k 张图片命中查询图像的概率。mAP 的定义如式(1)所示:

$$mAP=\frac{AP}{Q}$$
(1)

Q 代表查询集的数量,其中 AP 被定义为:

$$AP=\frac{\sum_{k=1}^n p(k) \cdot rel(k)}{R}$$
(2)

其中,k 为预测图片的排名,如果预测排名为 k 的图片与查询图片是同一 ID,rel(k) 为 1,否则为 0,而 p(k) 被定义为:

$$P(k) = \frac{\sum_{i=1}^k rel(i)}{k} \quad (3)$$

CMC 可视为 Ranklist 的可视化, 定义为:

$$CMC(k) = \frac{m(r)}{Q} \quad (4)$$

其中 r 代表 Rank- k , $m(r)$ 是指在前 r 个排名中包含所查询图片的数目, Q 代表查询集的数量。

4 行人重识别中的深度学习方法

4.1 基于表观特征的学习方法

基于表征学习的方法中, Geng 等^[1]用 Classification Loss 和 Verification Loss 两个子网络来训练网络, 用分类网络进行 ID 预测, 用验证网络来判断两张图片是否属于同一个行人。该子网络实质上是一个二分类网络。经过足够数据的训练, 再次输入一张测试图片, 网络将自动提取出一个特征, 用于行人重识别任务。由于仅靠行人 ID 信息学习出的网络表达能力有限, Lin 等^[2]对行人图片进行属性标注, 如性别、衣服、头发、衣着等, 并通过结合预测 ID 信息和属性信息来判定图片是否属于同一个行人, 其性能更加鲁棒。但是表观特征学习容易在数据集上过拟合, 泛化能力不强, 当训练的数据集中的行人数量 ID 达到一定阈值后效果会显得比较乏力。

4.2 基于度量学习的方法

基于度量学习的方法旨在对两张图片进行相似性度量时使两张图片更加容易匹配到或区分开。在行人重识别问题上, 具体为同一行人的不同图片相似度大于不同行人的不同图片。最后网络的损失函数使得相同行人图片(正样本对)的距离尽可能小, 不同行人图片(负样本对)的距离尽可能大。常用的度量学习损失方法有对比损失^[3]、三元组损失^[4]、四元组损失^[5]、难样本采样三元组损失(Triplet Hard Loss with Batch Hard Mining, Triplet Hard loss)^[13]、边界挖掘损失(Margin Sample Mining Loss, MSML)^[14]。与这些不同的是 Sun 等提出的 Circle Loss^[6], 通过让每个相似性得分以不同的步调学习, 收敛目标更加明确, 其针对传统损失函数优化缺乏灵活性、收敛状态不明确等问题的角度考虑提出了具有统一损失函数的 circleloss。从统一的相似度配对优化角度出发, 其为两种基本学习范式(使用类别标签和使用样本对标签的学习)提出了一种统

一的损失函数; 其次具有灵活的优化方式和明确的收敛状态, 有利于提高特征的鉴别能力。

4.3 行人重识别跨域、泛化能力

为增强行人重识别跨域、泛化能力和解决行人重识别中的姿态多变性问题, Qian 等^[7]使用 GAN 网络生成了 8 个姿势的图像样本, 用来代表所有摄像头视角的行人姿势。文献[15]使用条件 GANs 到输出图像的映射。文献[16]提出了新的 GAN 网络 SPGAN, 其可以将生成的图像结合行人重识别 loss 来解决区域偏差问题。文献[17]提出了一个新的数据集 SyRI, 通过更改 HDR 等参数, 可以生成 100 多个环境中的行人样本数据。文献[18]提出区域转换网络, 在保留原始样本身份的同时, 结合多类 GAN 损失来生成不可见区域的图像样本。Zheng 等^[8]则是通过 GAN 生成换衣的数据, 更换行人衣服的颜色。文献[19]结合了人的精细和粗略姿态信息来学习判别嵌入。文献[20]基于数据生成器研究了视角对行人重识别的影响, 说明了摄像头视角的变化对行人重识别的性能影响很大。

4.4 基于特征学习解决遮挡、姿势变换、光照等问题

图 2 所示为行人重识别在遮挡、姿势变换等场景下的一些特殊情况。为提取更细节的局部特征信息, 以进行遮挡、姿势变换等复杂情况下的行人重识别, Sun 等^[9]提出了分块卷积(PCB)。其将特征在垂直方向上均匀切分为 6 块并分别预测 ID, 通过这种方法可以提取更细节更具有鲁棒性的特征, 但是分完块后会出现图像不能对齐的情况。Wei 等^[10]利用提取的人体关键点把图片分为头部、上身和下身 3 个部分提取特征。Zheng 等^[11]则是先用姿态估计模型估计出行人关键点, 然后用仿射变换使得相同的关键点对齐。Wang 等^[12]提出了多粒度网络, 其将特征分成 3 个分支, 并使用多个分类损失和三元组损失联合学习, 最后整合所有特征。文献[21]首次将语义分割应用到行人重识别中, 并将分体部位进行分块、对齐, 以像素级的角度提取更高级的局部特征。文献[22]同样使用语义分割方法来进行块对齐, 但借鉴了 DensePose, 将身体分为 24 个部分, 并映射到空间中, 不仅实现了各个部位的语义对应, 甚至还实现了像素级别的对应。



图 2 行人重识别在特殊场景下的情况

Fig. 2 Some special cases in pedestrian re-identification

5 挑战与未来

现有的行人重识别算法在几个常见的行人重识别数据集集中的 mAP 和 Rank-1 命中率已经很高了,目前的挑战主要在于复杂场景多变,同一种算法在风格差异较大的数据集中不能有很好的表现。目前的主要研究趋势在于跨域无监督方面的行人重识别,以及 3D 行人重识别。3D 能表达更多的信息但是需要提取更多的信息进行匹配。

结束语 行人重识别是计算机视觉领域的一个热门且具有极大实际应用价值的研究方向。本文首先介绍了传统方法和深度学习的一些主流数据集,然后针对基于表观特征的方法,基于度量学习的方法,行人重识别跨域、泛化能力方面,基于特征学习解决遮挡、姿势变换、光照等问题的 4 类方法对现有的研究成果进行了综述。最后,针对现有研究中存在的缺陷进行论述,指出了该领域仍待解决的问题,并深入探讨了未来发展的方向。

参 考 文 献

[1] GENG M Y, WANG Y W, XIANG T, et al. Deep transfer learning for person re-identification[J]. arXiv:1611.05244, 2016.

[2] LIN Y T, ZHENG L, ZHENG Z D, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151-161.

[3] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification[C]// European Conference on Computer

Vision, Springer, 2016: 791-808.

[4] CHENG D, GONG Y H, ZHOU S P, et al. Person re-identification by multichannel parts-based cnn with improved triplet loss function[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1335-1344.

[5] CHEN W H, CHEN X T, ZHANG J G, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 403-412.

[6] SUN Y F, CHENG C M, ZHANG Y H, et al. Circle Loss: A Unified Perspective of Pair Similarity Optimization[J]. arXiv:2002.10857, 2020.

[7] QIAN X L, FU Y W, WANG W X, et al. Pose-Normalized Image Generation for Person Re-identification [C]// ECCV, 2018: 661-678.

[8] ZHENG Z D, YANG X D, YU Z D, et al. Joint discriminative and generative learning for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 2138-2147.

[9] SUN Y F, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) [C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.

[10] WEI L H, ZHANG S L, YAO H T, et al. Glad: Global-local-alignment descriptor for pedestrian retrieval[C]// Proceedings of the 25th ACM international conference on Multimedia, ACM, 2017: 420-428.

[11] ZHENG L, HUANG Y J, LU H C, et al. Pose invariant

- embedding for deep person reidentification[J]. arXiv: 1701.07732, 2017.
- [12] WANG G S, YUAN Y F, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//Proceedings of the 26th ACM International Conference on Multimedia, 2018:274-282.
- [13] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person reidentification[J]. arXiv:1703.07737, 2017.
- [14] XIAO Q, LUO H, ZHANG C. Margin Sample Mining Loss: A Deep Learning Based Method for Person Re-identification[J]. arXiv:1710.00478, 2017.
- [15] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//CVPR, 2017:5967-5976.
- [16] DENG W, ZHENG L, KANG G, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification [C] // CVPR, 2018:994-1003.
- [17] BAK S, CARR P, LALONDE J F. Domain Adaptation through Synthesis for Unsupervised Person Re-identification[DB/OL]. <https://arxiv.org/pdf/1804.10094.pdf>, accessed April 2018.
- [18] PENG P, XIANG T, WANG Y, et al. Unsupervised cross-dataset transfer learning for person re-identification[C]//CVPR, 2016:1306-1315.
- [19] SAQUIB S M, SCHUMANN A, EBERLE A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:420-429.
- [20] SUN X, ZHENG L. Dissecting Person Re-identification from the Viewpoint of Viewpoint[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:608-617.
- [21] KALAYEH M M, BASARAN E, GÖKMEN M, et al. Human semantic parsing for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:1062-1071.
- [22] ZHANG Z, LAN C, ZENG W, et al. Densely Semantically Aligned Person Re-Identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:667-676.



DONG Ya-chao, born in 1995, male, master student of Beijing Union University. His main research interests include computer vision, person re-identification and so on.



LIU Hong-zhe, born in 1971, female, professor. Her main research directions include semantic computing, digital image processing and artificial intelligence.