Hans Journal of Data Mining 数据挖掘, 2023, 13(1), 75-82
Published Online January 2023 in Hans. https://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2023.131008

Hans汉斯

迭代策略下的行人多目标跟踪研究

孟庆权，王永会

沈阳建筑大学计算机科学与工程学院，辽宁 沈阳

收稿日期：2022年12月12日；录用日期：2023年1月12日；发布日期：2023年1月20日

摘　要

多目标跟踪是计算机视觉领域被广泛研究的重要方向，但在实际应用中，目标的快速移动、光照变化、遮挡等问题会导致跟踪性能变差。在本文中提出了一种迭代策略的行人多目标跟踪方法。采用迭代检测方式，可以在两次迭代分别检测出高置信度和低置信度的行人目标，由于前一次迭代检测到的行人预测框将在下一次迭代中以历史特征的形式传递到网络，从而可以避免模型重复检测同一对象，同时提高行人检测的精度。在数据关联阶段，优先对第一次迭代检测结果进行轨迹匹配即高置信度行人检测框，然后是第二次迭代检测结果，这种对检测结果分批次处理可以有效的减少跟踪过程中身份切换问题。在MOT16数据集的实验表明本文方法对行人目标跟踪的可行性和有效性。

关键词

计算机视觉，多目标跟踪，迭代检测，数据关联

Study on Pedestrian Multi-Object Tracking Based on Iterative Strategy

Qingquan Meng, Yonghui Wang

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang Liaoning

Received: Dec. 12th, 2022; accepted: Jan. 12th, 2023; published: Jan. 20th, 2023

Abstract

Multi-object tracking is an important research direction in the field of computer vision, but in practical applications, the fast movement of the object, illumination change, occlusion and other problems will lead to poor tracking performance. In this paper, we propose a pedestrian multi-object tracking method based on iterative strategy. By adopting the iterative detection method, pedestrian targets with high confidence and low confidence can be detected respectively in two iterations. Since the pedestrian prediction box detected in the previous iteration will be transmit-

文章引用: 孟庆权, 王永会. 迭代策略下的行人多目标跟踪研究[J]. 数据挖掘, 2023, 13(1): 75-82.
DOI: 10.12677/hjdm.2023.131008

ted to the network in the form of historical features in the next iteration, the model can avoid repeated detection of the same object and improve the accuracy of pedestrian detection. In the data association stage, the trajectory matching of the first iteration detection results, namely, the high confidence pedestrian detection box, is given priority, followed by the second iteration detection results. This batch processing of the detection results can effectively reduce the problem of identity switching in the tracking process. Experiments on MOT16 data sets show that the proposed method is feasible and effective for pedestrian target tracking.

Keywords

Computer Vision, Multi-Target Tracking, Iterative Detection, Data Association

1. 简介

多目标跟踪是当前计算机视觉领域的研究热点。它的主要任务是在给定视频中同时对多个特定目标进行定位，同时保持目标的 ID 稳定，最后记录它们的轨迹[1]。目前主流的行人多目标跟踪遵循检测跟踪范式，它将 MOT 问题分成两个独立的步骤，即首先通过目标检测算法检测出视频帧中目标对象可能出现的区域，然后通过关联模型将属于同一运动目标的检测框关联到一起，得到目标的关联轨迹，完成目标对象的跟踪。本文沿用检测跟踪范式进行行人多目标的研究。

近年来，基于检测的多目标跟踪研究越来越多。Bewley 等[2]提出 Sort 网络，利用双阶段检测器同时结合卡尔曼滤波[3] (Kalman Filter)和匈牙利算法[4] (Hungarian Algorithm)，前者用来预测目标框在当前帧中的位置，后者以预测框和检测框的交并比(Intersection over Union, IOU)为相似性度量来匹配来完成数据关联。Sort 在面对人流较稀疏场景简单有效，在遮挡拥挤的场景下会使目标 ID 的变换次数多，跟踪准确性差。Wojke 等[5]针对 SORT 存在的缺陷提出了 DeepSort 算法，在 Sort 基础上加入了级联匹配和外观模型，级联匹配策略提高了目标匹配准确度，外观模型使用行人重识别(Person Re-identification, ReID)网络[6]提取目标的外观特征作为数据关联的辅助度量，有效的解决了身份切换的问题。但串联两个深度学习模型使得计算量急剧增加。Wang 等[7]提出 JDE 网络，将 ReID 网络与检测网络整合到一个网络中去，在输出目标的同时也输出相应的表观特征，然后结合运动信息进行数据关联。虽然都是采用逐检测跟踪范式，但它们大多数检测器的性能有待提高并且大多数 MOT 方法只保留了满足阈值条件的检测框，并未对检测框进行充分利用。

本文提出了一种迭代策略的多目标跟踪方法，通过迭代检测，分别检测出高置信度和低置信度的目标框，充分考虑所有检测框。并在数据关联中对迭代结果分阶段处理，缓解检测框干扰造成的轨迹漂移而形成的身份切换问题。

2. 提出的方法

本文将多目标跟踪算法分为两个步骤，即第一步是目标迭代检测阶段，通过两次目标检测分别得到高置信度和低置信度目标框；第二步是数据关联阶段，利用轨迹和检测框之间的相似度进行匹配。迭代策略下对检测框分批次处理有效的减少了跟踪过程中身份切换问题同时提高了跟踪精度。图 1 是本文整

体的网络结构。

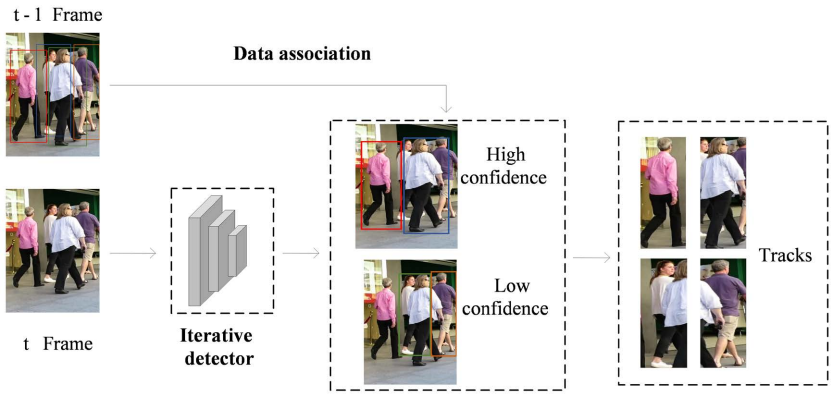

Figure 1. Overall framework
图 1. 整体框架

2.1. 迭代检测

目标检测是多目标跟踪的基础，检测器的性能决定了数据关联的上限。但基于深度学习的检测器更倾向于对相同对象的进行重复检测，然后通过非极大值抑制算法对检测框进行过滤，最后只剩下一个边界框，但这在拥挤环境中通常效果很差，常常出现漏检的情况。针对这个问题提出了迭代策略如图 2 所示，通过两次迭代输出最终的检测结果。由于当前迭代中考虑所有先前迭代检测到的目标框，所以不会两次检测到同一对象。

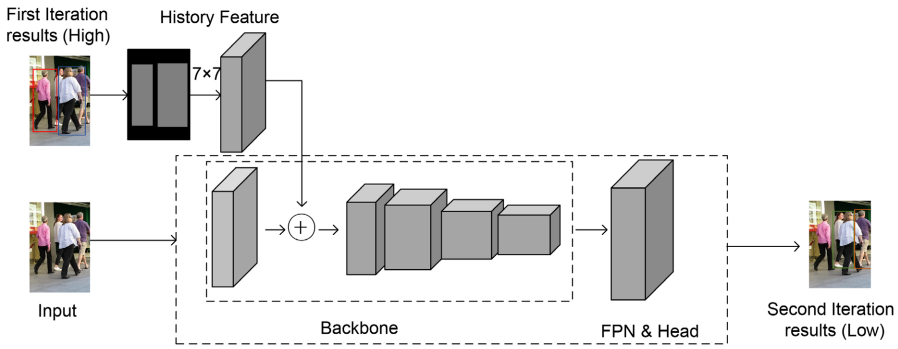

Figure 2. Network structure of iterative detection
图 2. 迭代检测网络结构图

目标检测的一般流程：输入图像 $I \in R^{w\times h\times 3}$ 得到一组边界框 $B=\{(X_k,Y_k,W_k,H_k)\}_{k=1}^{n}$，每个目标框由左上角坐标 (X,Y)，宽度为 W，高度 H 表示。在本文迭代策略中引入了灰度图像作为历史图像 $H \in R^{w\times h\times 1}$ 与输入图像大小相同。初始时所有像素值为 0，当检测器检测出行人位置在灰度图上对应位置像素值加 1，同一位置存在不同检测框时像素值进行累加，如公式(1)所示，

$$H_{(x,y)}=\sum_{k=1}^{n}1,\ X_k\le x\le X_k+W_k, Y_k\le y\le Y_k+H_k \tag{1}$$

其中 n 为先前迭代检测的边界框的数量。

迭代策略下的检测为：在给定图像 I 的情况下，第一次迭代 $t=1$，此时历史 H_1 为空即灰度图像像素值为 0，联合图像 I 和 H_1 并通过阈值 δ 设置可以得出第一组高置信度检测框 D_{high}，其次，D_{high} 被映射到历史 H_2，然后第二次迭代结合图像 I 和 H_2 得出第二组低置信度检测框 D_{low}。

2.2. 数据关联

数据关联是多目标跟踪的核心，它首先计算机轨迹和检测框之间的相似度并根据相似度利用不同的策略进行匹配。位置、运动和外观都是关联的有用线索，在短距离匹配中位置和运动相似性[8]是准确的，外观相似性有助于长距离匹配，在获得相似性度量后，可以通过匈牙利算法或者贪婪赋值算法[9]来进行匹配。

以往的多目标跟踪方法中通常只保留高置信度检测框。在本文迭代检测中充分利用了每个检测框。并按照迭代结果的先后顺序处理不同检测框。我们首先将高置信度检测框与轨迹相关联。有些轨迹无法匹配到高置信度检测框，这通常发生在遮挡、运动模糊或大小变化时。然后，我们将低置信度检测框与这些不匹配的轨迹相关联，以恢复低置信度检测框中的对象，同时可以过滤到检测得到的假阳性目标框即背景。关联的伪代码如下算法所示。

Algorithm 1 Iterative tracking algorithm

Input: A video sequence V; object detector Det; detection score threshold δ

Output: Tracks Γ of the video

Initialization: $\Gamma \leftarrow \emptyset$
for frame f_k in V **do**
  /* First iteration */
  $D_{high} \leftarrow \mathrm{Det}(f_k,\delta)$
  /* predict new locations of tracks */
  for t in Γ **do**
    t $\leftarrow$ KalmanFilter(t)
  end for
  /* Associate high confidence */
  Associate Γ and D_{high} using IOU distance
  $D_{remain} \leftarrow$ remaining object boxes from D_{high}
  $\Gamma_{remain} \leftarrow$ remaining tracks from Γ
  /* Second iteration */
  $D_{Low} \leftarrow \mathrm{Det}(f_k,D_{high})$
  /* Associate low confidence */
  Associate Γ and D_{Low} using IOU distance
  Delete remaining object boxes from D_{Low}
  $\Gamma_{re-remain} \leftarrow$ remaining tracks from Γ_{remain}
  $\Gamma_{lost} \leftarrow \Gamma_{re-remain}$
  /* delete long time unmatched tracks */
  $\Gamma \leftarrow \Gamma \ominus \Gamma_{lost}(\mathrm{time} \geq 30)$
  /* initialize new tracks */
  for d in D_{remain} **do**
    $\Gamma \leftarrow \Gamma \cup \{d\}$
  end for
end for
return Γ

数据关联中输入的是视频序列 V，检测器对象 Det 和检测置信度阈值 δ，实验中 δ 设置为 0.6。

输出是视频的轨迹 Γ，每个轨迹包含每个帧中目标的边界框和标识。流程如下：

1) 对于视频的每一帧，使用检测器 Det 第一次迭代检测出高置信度的检测框 D_{high} 即得分大于 δ 的检测框。

2) 利用卡尔曼滤波器来预测 Γ (包括丢失的轨迹)中每个轨迹在当前帧中的新位置。由于轨迹的丢失常常伴随着遮挡，但同时丢失轨迹也在运动，所以为了丢失轨迹的恢复保留其身份，并进行新位置预测。

3) 高置信度检测框 D_{high} 和预测的轨迹框 Γ 之间的 IoU 来计算相似度，采用匈牙利算法来完成基于相似度的匹配，此为第一次关联。

4) 使用检测器 Det 结合前一次迭代的结果第二次迭代检测出低置信度的检测框 D_{low}，将其和剩余轨道 Γ_{remain} 之间执行第二次关联。我们保留不匹配的轨迹 $\Gamma_{re-remain}$，视为丢失的轨迹 Γ_{lost}，只删除所有不匹配的低置信度检测框，因为它们往往是迭代检测出的假阳性框即背景。

5) 对于 Γ_{lost} 中的每个轨道，当它存在超过一定数量的帧(即 30 帧)时，从轨道 Γ 中删除它。

6) 对于第一次关联之后从不匹配的高分检测框 D_{remain} 初始化新轨迹。

3. 实验结果与分析

3.1. 数据集与实验环境

本文在公开的行人检测数据集 CrowdHuman 的训练集[10]训练迭代检测器，用 CrowdHuman 的验证集验证检测器性能，用 MOT16 的训练集[11]作为验证集验证跟踪算法性能。实验环境基于 Ubuntu 18.04 操作系统，Nvidia GeForce RTX 3090 显卡，运行内存 24G，采用 Pytorch1.8.1 深度学习框架，在 Python3.7 的服务器下实现。选择多目标跟踪公开数据 MOT16 测试集测试本文算法，并 MOT Challenge 官网上对结果进行评估，同时与其他算法进行对比，并分析模型性能。

3.2. 评价指标

本文采用一些通用评估指标分别对行人检测和多目标跟踪进行客观评价，以证明算法的准确性。对于行人检测采用平均精度(mAP)，召回率(Recall)，平均丢失率(mMR)作为评估指标。部分评价指标公式如公式(2)，(3)，(4)和(5)所示

$$\mathrm{Recall}(\mathrm{R}) = \frac{\mathrm{T_{rue}P_{ositive}}}{\mathrm{T_{rue}P_{ositive}} + \mathrm{F_{alse}N_{egative}}} \tag{2}$$

$$\mathrm{Precision}(\mathrm{P}) = \frac{\mathrm{T_{rue}P_{ositive}}}{\mathrm{T_{rue}P_{ositive}} + \mathrm{F_{alse}P_{ositive}}} \tag{3}$$

$$\mathrm{mAP} = \int_0^1 \mathrm{PRdR} \tag{4}$$

$$\mathrm{MR} = \frac{\mathrm{F_{alse}N_{egative}}}{\mathrm{T_{rue}P_{ositive}} + \mathrm{F_{alse}N_{egative}}} \tag{5}$$

平均丢失率 mMR，是在(False Positive Per Image) FPPI@0.01-1 下漏检数的平均值越低越好。

对于多目标跟踪采用多目标跟踪领域通用的评估指标进行评估，评估指标如下：

1) MOTA (Multi-object Tracking Accuracy)代表多目标跟踪的准确度，综合考虑了误检、漏检和身份切换 3 种因素，衡量模型在目标检测和轨迹关联的整体性能，如公式(6)所示。

$$\mathrm{MOTA} = 1 - \frac{N_{FN} + N_{FP} + N_{IDs}}{N_{GT}} \tag{6}$$

其中 N_{GT} 表示真实边界框数量； N_{FN} 表示整个视频的漏检数； N_{FP} 表示整个视频的误检数； N_{IDs} 表示总的行人 ID 切换次数；

2) MOTP (Multi-object Tracking Precision)代表多目标跟踪精度，计算目标检测框与真实框在所有帧之间的平均度量距离。如公式(7)所示。

$$\mathrm{MOTP}=\frac{\sum_{i,t} d_t^i}{\sum_t C_t} \tag{7}$$

其中 t 表示当前帧为第 t 帧，$t\in[1,N]$，d_t^i 表示第 t 帧中第 i 个预测框与真实框之间的重叠率，即 IoU (Intersection-Over-Union)距离；c 表示目标成功匹配数量。

3) IDF1 (Identification F1 Score)代表多目标跟踪器 ID 维持能力。

4) IDs (ID Switch)代表整个跟踪过程行人身份(ID)切换的数目。

5) MT (Mostly Tracked)代表至少在 80%匹配成功的跟踪轨迹。

6) ML (Mostly Lost)代表在小于 20%的时间成功匹配的跟踪轨迹即大多数丢失目标百分比。

3.3. 结果分析

本文使用 Faster R-CNN [12]作为行人检测框架，使用 ResNet50 [13]作为骨干网络，添加 FPN 模块，并添加 BN 层改善 FPN 模块的特征融合效果，训练使用 Adam 优化器，epoch 设置为 24，学习率设置为 0.0005，在第 16 个 epoch 和第 22 个 epoch 时学习率分别乘以 0.1。在训练过程中，在标注框随机选取一部分标注框并生成历史特征，同时该部分标注框将不参与训练，将历史特征添加到骨干网略中去训练网络使得网络预测到的结果更加接近剩余的标注框，从而迫使网络学习利用历史特征。引进迭代策略改进后的模型与基准模型在 CrowdHuman 验证集上的结果对比，如表 1 所示。

Table 1. Comparison results of iterative detection and benchmark model on CrowdHuman verification set
表 1. 迭代检测与基准模型在 CrowdHuman 验证集上的对比结果

检测模型	Recall ↑	mAP ↑	mMR ↓
基准模型	89.2	84.8	50.5
迭代检测	94.5	88.1	49.1

对于多目标跟踪，本文选择 MOT16 数据集进行实验，与几种先进多目标跟踪算法进行对比，结果如表 2 所示，本文在迭代策略下充分考虑到所有检测框，并且对检测框的分批次处理，因此在 MOTA、MOTP 等指标上都有所提升。跟踪效果如图 3 所示。在 MOT16-03 中 61 号和 MOT16-06 中 71 号行人在面对遮挡情况，依然可以准确追踪。

Table 2. Comparison results between the proposed algorithm and other algorithms on MOT16 data sets
表 2. 本文算法与其他算法在 MOT16 数据集上的对比结果

方法	MOTA ↑	MOTP ↑	IDF1 ↑	MT/% ↑	ML/% ↓	IDs ↓
SORT [2]	59.8	79.6	53.8	25.4	22.7	1423
DeepSORT [4]	61.4	79.1	62.2	32.8	18.2	**781**
JDE [5]	64.4	-	55.8	**35.4**	20.0	1544
TubeTK [14]	64.9	59.4	59.4	33.5	19.4	1117
CNNMTT [15]	65.2	-	62.2	32.4	21.3	946
Ours	**65.4**	**80.8**	**65.8**	35.3	**17.8**	982

Figure 3. Multi-object tracking effect. (a) MOT16-03; (b) MOT16-06
图 3. 多目标跟踪效果。(a) MOT16-03；(b) MOT16-06

4. 结论

本文针对行人多目标跟踪过程中因为遮挡导致的行人 ID 频繁切换，跟踪效果差的问题，提出一种迭代策略的多目标跟踪算法。该算法选用 TBD 跟踪范式，通过迭代检测，分别检测出高置信度和低置信度的目标框，充分考虑所有检测框。并在数据关联中对迭代结果分阶段处理。其准确的检测性能和关联低置信度检测框的帮助，对遮挡非常鲁棒。在未来，我们将继续探索更加可靠的相似性度量和匹配策略，以提升多目标跟踪任务的整体性能。

参考文献

[1] Luo, W.H., Xing, J.L., Milan, A., *et al.* (2021) Multiple Object Tracking: A Literature Review. *Artificial Intelligence*, **293**, Article ID: 103448. https://doi.org/10.1016/j.artint.2020.103448

[2] Bewley, A., Ge, Z.Y., Ott, L., Ramos, F. and Upcroft, B. (2016) Simple Online and Realtime Tracking. *Proceedings of* 2016 *IEEE International Conference on Image Processing* (*ICIP*). Phoenix, 25-28 September 2016, 3464-3468. https://doi.org/10.1109/ICIP.2016.7533003

[3] Kalman, R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Journal of Fluids Engineering*, **82,** 35-45. https://doi.org/10.1115/1.3662552

[4] Kuhn, H.W. (1955) The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, **2**, 83-97. https://doi.org/10.1002/nav.3800020109

[5] Wojke, N., Bewley, A. and Paulus, D. (2017) Simple Online and Realtime Tracking with a Deep Association Metric. *Proceedings of* 2017 *IEEE International Conference on Image Processing* (*ICIP*). Beijing, 17-20 September 2017, 3645-3649. https://doi.org/10.1109/ICIP.2017.8296962

[6] 罗浩, 姜伟, 范星, 张思朋. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.

[7] Wang, Z.D., Zheng, L., Liu, Y.X., Li, Y.L. and Wang, S.J. (2020) Towards Real-Time Multi-Object TrackingIn: Vedaldi, A., Bischof, H., Brox, T. and Frahm, JM., Eds., *Computer Vision—ECCV* 2020. *ECCV* 2020. *Lecture Notes in Computer Science*, Vol. 12356, Springer, Cham, 107-122. https://doi.org/10.1007/978-3-030-58621-8_7

[8] 花景培, 陈昌红, 干宗良, 刘峰. 基于运动和外形度量的多目标行人跟踪[J]. 南京邮电大学学报, 2016, 36(71): 1673-5439.

[9] Zhou, X.Y., Koltun, V. and Krähenbühl, P. (2020) Tracking Objects as Points. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV* 2020. *ECCV* 2020. *Lecture Notes in Computer Science*, Vol. 12349, Springer, Cham, 474-490. https://doi.org/10.1007/978-3-030-58548-8_28

[10] Shao, S., Zhao, Z.J., Li, B.X., *et al.* (2018) Crowdhuman: A Benchmark for Detecting Human in a Crowd. ArXiv Preprint ArXiv: 1805.00123.

[11] Milan, A., Leal-Taixe, L., Reid, I., Roth, S. and Schindler, K. (2016) MOT16: A Benchmark for Multi-Object Tracking.

ArXiv: 1603.00831.

[12] Ren, S.Q., He, KM., Girshick, R.B. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *NIPS*, 91-99.

[13] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[14] Pang, B., Li, Y.Z., Zhang, Y.F., Li, M.C. and Lu, C.W. (2020) TubeTK: Adopting Tubes to Track Multi-Object in a One-Step Training Model. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 6307-6317. https://doi.org/10.1109/CVPR42600.2020.00634

[15] Mahmoudi, N., Ahadi, S.M. and Rahmati, M. (2019) Multi-Target Tracking Using CNN-Based Features: CNNMTT. *Multimedia Tools and Applications*, **78**, 7077-7096. https://doi.org/10.1007/s11042-018-6467-6