

文章编号:1007-2780(2022)06-0777-09

基于改进 Fairmot 框架的多目标跟踪

席一帆*, 何立明, 吕 悦
(长安大学 信息工程学院, 陕西 西安 710064)

摘要:针对复杂场景下目标之间遮挡造成跟踪精度降低的问题,提出基于 Fairmot 框架的多目标跟踪改进算法。将主干网特征图通过三重注意力机制进行维度间的信息交互产生注意力掩模,提高对目标的定位能力;行人重识别分支采用 Circle Loss 依据当前状态选择优化程度,提取更为精确的表观特征,区分不同目标对象。实验结果表明,在 MOT15 数据集上跟踪精度提升至 62%,MT (Mostly Tracked) 提升至 358,身份切换降低 68 次,在发生遮挡的场景中拥有更出色的跟踪效果。

关键词:图像处理;深度学习;目标跟踪;注意力机制

中图分类号:TP391 文献标识码:A doi:10.37188/CJLCD.2021-0304

Multi-object tracking based on improved Fairmot framework

XI Yi-fan*, HE Li-ming, LYU Yue

(School of Information Engineering, Changan University, Xi'an 710064, China)

Abstract: Aiming at the problem of reduced tracking accuracy caused by occlusion between targets in complex scenes, an improved multi-object tracking algorithm based on the Fairmot framework is proposed. The feature map of the backbone network is used for information interaction between dimensions through a triplet attention mechanism to generate an attention mask, which improves the positioning ability of the target. Person re-identification branch adopts Circle Loss to select the degree of optimization according to the current state, to extract more accurate appearance features and distinguish different target objects. The experimental results show that the tracking accuracy on the MOT15 data set is increased to 62%, the MT (Mostly Tracked) is increased to 358, and the identity switching is reduced by 68 times. It has a better tracking effect in the scene where occlusion occurs.

Key words: image processing; deep learning; target tracking; attention mechanism

1 引言

多目标跟踪最初源于雷达技术的研究。在军事上,数据关联算法利用目标的位置和运动信息进行轨迹和观测目标的匹配。近年来,随着我国视频监控和无人驾驶行业的飞速发展,基于视

频的多目标跟踪技术显得尤为重要。多目标跟踪根据初始化的方式划分为基于检测跟踪的流程和基于人工初始化的跟踪流程。由于基于人工初始化的跟踪无法处理轨迹的生成和消亡,因此基于检测的跟踪为当前的主流方式。基于检测的跟踪包括目标检测和数据关联,两者功能相

收稿日期:2021-11-27;修订日期:2021-12-27.

*通信联系人, E-mail: 15735169750@163.com

互独立,但却在关系上紧密联系,良好的检测器能为数据关联提供较好的观测结果。

早期的目标检测主要依靠人工设计的特征训练支持向量机进行分类。2005年,Dala^[1]利用HOG特征来训练分类器;2008年,DPM^[2]检测器依据改进HOG特征,采用根滤波器和部件滤波器在多尺度金字塔上滑动检测;2014年,随着深度学习的兴起,传统的目标检测方式被逐渐取代;R-CNN^[3]在传统目标检测方式的基础上,以卷积神经网络作为特征提取器,训练支持向量机;Fast-RCNN^[4]是首个利用全卷积神经网络训练的目标检测器。Faster-RCNN^[5]提出区域建议网络,对任意尺度输入的图像都会生成一组后选框,首次引入锚框机制,速度比Fast-RCNN快一个数量级。YOLO^[6-9]系列框架主要基于锚框(Anchor)机制,将目标的位置和尺寸视为回归问题,该系列框架检测速度快,但锚框机制存在正负样本不均衡,超参数管理复杂等缺点。近些年,基于关键点的目标检测逐渐兴起,Cornernet^[10]通过利用目标的左上角点和右下角点对目标进行定位。Centernet^[11]通过中心点对目标的尺寸、位置和中心点的偏移量进行预测,拥有更高的检测效率。

数据关联负责将目标的轨迹与观测目标进行匹配。数据关联算法分为确定性优化算法和概率推断算法。确定性优化算法将其建模成优化问题,通过优化算法解决匹配问题。二分图匹配模型^[12]、动态规划^[13]、最小成本最大流网络模型^[14]、条件随机场^[15]和最大权值独立集模型^[16]属于确定性的优化模型。概率推断模型基于现有的观测状态估计目标状态的概率分布。卡尔曼滤波^[17]、扩展卡尔曼滤波^[18]和粒子滤波^[19]属于概率推断模型。

端到端的多目标跟踪框架近些年飞速发展,将目标检测和数据关联都用神经网络来处理,使得网络的训练效率得到提升。DAN^[20]跨帧提取特征,计算亲和性矩阵,并用交并比信息作为掩模进行匹配。DeepMOT^[21]根据匈牙利算法不可微分的特点,通过MOTA和MOTP的跟踪指标创建损失函数,训练深度匈牙利网络替代数据关联。DMAN^[22]提出空间注意力模块和时间注意力模块,空间注意力模块匹配两幅图像空间相同区域,时间注意力模块对历史轨迹分配不同的权

重,滤除不可靠的轨迹。本文基于检测与数据关联的Fairmot^[23]框架,提出一种改进算法,提高对目标对象的跟踪精度。

2 Fairmot 基本框架

Fairmot框架的目标检测部分包括主干网络、目标检测分支和行人重识别分支。数据关联部分采用DeepSort^[24]框架进行匹配。

2.1 主干网络

Fairmot框架采用改进后深度聚合网络作为特征提取网络,该特征提取网络采用可变形卷积适应不同尺度目标。通过不同层级的跳级连接进行语义信息和空间信息融合,使深度聚合网络以目标尺度、分辨率为关注点。

2.2 目标检测分支

目标中心预测分支负责对目标位置和目标的尺寸进行回归预测,如图1所示。检测分支主要由3部分组成,其中热图分支负责预测目标中心点的位置,框回归分支负责预测目标检测框尺寸,中心点偏移分支负责估计目标中心点偏移量。图1给出了3条分支的仿真图和部分数值仿真结果。目标真实框的位置记为 $p^i = (x_1^i, y_1^i, x_2^i, y_2^i)$,则其目标的中心为 (c_x^i, c_y^i) ,其中 $c_x^i = (x_1^i + x_2^i)/2$ 和 $c_y^i = (y_1^i + y_2^i)/2$,映射到特征图上的坐标为 $(c_x^i, c_y^i) = (\lfloor c_x^i/4 \rfloor, \lfloor c_y^i/4 \rfloor)$,在热图 (x, y) 处的响应为:

$$R_{xy} = \sum_{i=1}^N \exp \frac{-(x - c_x^i)^2 + (y - c_y^i)^2}{2\sigma_c^2}, \quad (1)$$

热图分支中心点的估计的损失函数:

$$L_{\text{heat}} = -\frac{1}{N} \sum_{xy} \begin{cases} (1 - \hat{R}_{xy})^\alpha \log \hat{R}_{xy} & R_{xy} = 1 \\ (1 - R_{xy})^\beta \hat{R}_{xy}^\alpha \log(1 - \hat{R}_{xy}) & \text{otherwise} \end{cases}, \quad (2)$$

式中 N 为视频帧中目标对象的个数, σ_c 为标准差, $\hat{R}$ 为目标中心点的估计值, α, β 为预定义参数。

视频帧经过主干网络会产生下采样4倍的特征图,当目标的中心点预测结果映射回原图时会产生4个像素的误差,因此通过中心点预测偏移

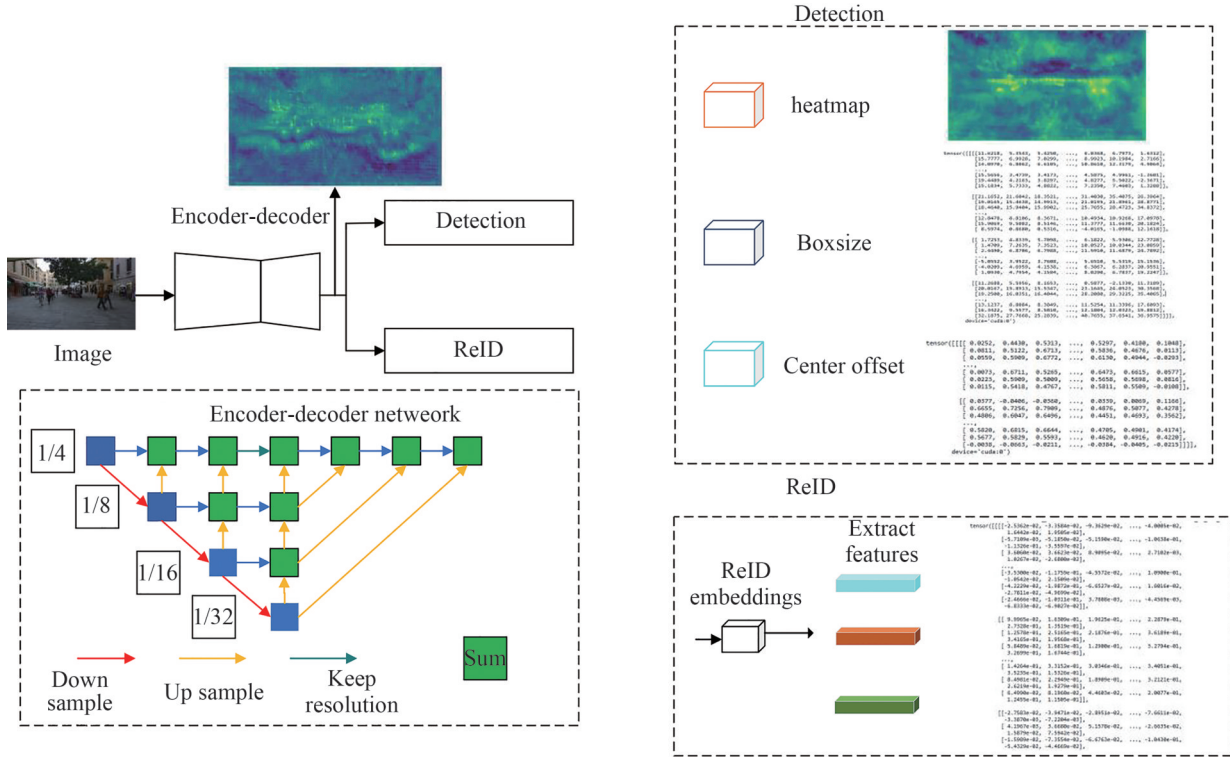


图1 Fairmote 框架

Fig. 1 Fairmote framework

量弥补误差。视频中目标的尺寸为 $s^i = (x_2^i - x_1^i, y_2^i - y_1^i)$, 真实标签中心点的偏移量为 $o^i = (c_x^i/4, c_y^i/4) - (\lfloor c_x^i/4 \rfloor, \lfloor c_y^i/4 \rfloor)$, 则预测的目标尺寸和中心点偏移量表示为 $\hat{s}^i$ 和 $\hat{o}^i$, 框回归和中心点偏移分支的损失函数为:

$$L_{\text{box}} = \sum_{i=1}^N \|o^i - \hat{o}^i\|_1 + \|s^i - \hat{s}^i\|_1. \quad (3)$$

2.3 行人重识别分支

行人重识别分支通过提取的特征向量区分视频中的不同目标。利用128个通道的 1×1 的卷积核在特征图上提取目标的表观特征, 在目标中心 (x, y) 处提取的表观特征为 $E_{x,y} \in R^{128}$ 。通过分类损失学习表观特征, 假设视频帧的真实目标框 $b^i = (x_1^i, y_1^i, x_2^i, y_2^i)$, 目标在热图分支预测的中心点位置为 $(\tilde{c}_x^i, \tilde{c}_y^i)$, 在特征图上提取表观特征向量 $E_{\tilde{c}_x^i, \tilde{c}_y^i}$, 且将该表观特征映射成 $[0, 1]$ 的概率分布向量 $P = \{p(k), k \in [1, K]\}$, 将真实标签进行独热编码 $L^i(k)$, 训练时的损失函数为:

$$L_{\text{identity}} = - \sum_{i=1}^N \sum_{k=1}^K L^i(k) \log p(k), \quad (4)$$

式中 K 为目标类别个数。总体的损失函数:

$$L_{\text{detection}} = L_{\text{heat}} + L_{\text{box}}, \quad (5)$$

$$L_{\text{total}} = \frac{1}{2} \left(\frac{1}{e^{w_1}} L_{\text{detection}} + \frac{1}{e^{w_2}} L_{\text{identity}} \right), \quad (6)$$

式中 w_1 和 w_2 是学习参数。

2.4 数据关联

数据关联部分采用DeepSort框架。如图2所示, DeepSort首先通过级联匹配得到最初的匹配轨迹集合、未匹配的检测集合和未匹配的轨迹集合, 然后将级联匹配结果中的未匹配轨迹集合和未匹配检测集合进行IOU匹配得到最终的匹配结果。

匹配的轨迹集合作为观测结果进行卡尔曼滤波更新, 经过IOU匹配得到的未匹配检测框集合。如果连续3帧都匹配上轨迹, 则认为是新的轨迹, 然后进行卡尔曼滤波更新。最终的未匹配集合依据状态来判断该轨迹是否消亡。卡尔曼滤波更新得到的轨迹若为确认态则送入级联匹配, 否则送入IOU匹配。图2右下角为部分视频的两次匹配结果。

函数优化缺乏灵活性,采用Circle Loss根据当前的状态选择优化程度,使其提取更为精确的身份嵌入向量。

3.1 三重注意力机制

针对深度聚合网络后端高维信息缺乏维度间信息交互的问题,通过三重注意力机制^[25](图3)进行维度间信息交互。该机制能分别从 (C, H) , (C, W) , (H, W) 维度捕捉信息产生注意力掩模。其中的Z-Pool模块通过最大池化和平均池化将特征图的第0维度的通道数降至2,使特征图保持丰富语义信息的同时,进一步简化计算量。其公式如式(16)所示:

$$\text{Z-Pool} = [\text{MaxPool}_{0d}(x), \text{AvgPool}_{0d}(x)]. \quad (16)$$

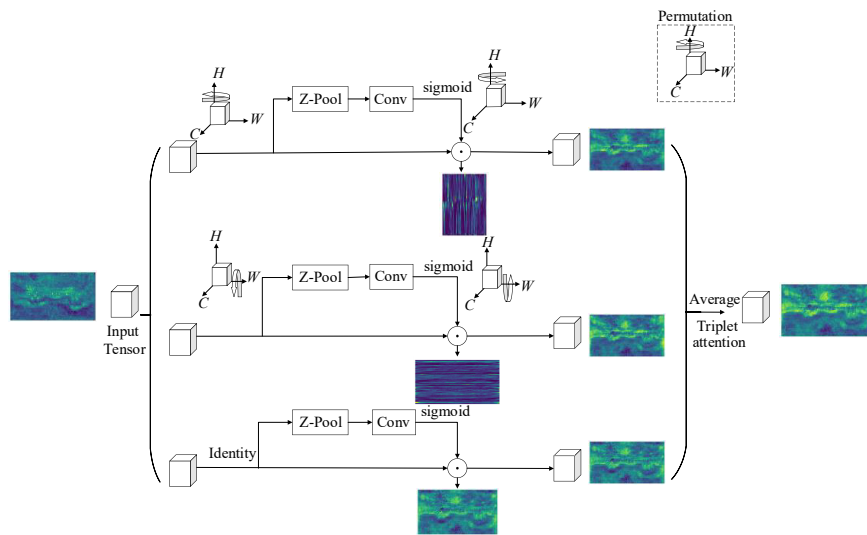


图3 三重注意力机制

Fig. 3 Triplet attention mechanism

3.2 Circle Loss

深度学习的目标是将类内特征的相似度最大化,类间特征的相似度最小化,因此Circle Loss^[26]概括出一个统一的损失函数表达式:

$$L_{\text{uni}} = \log \left[1 + \sum_{i=1}^K \sum_{j=1}^L \exp(r(s_n^i - s_o^j + m)) \right] = \log \left[1 + \sum_{j=1}^L \exp(r(s_n^j + m)) \sum_{i=1}^K \exp(r(-s_p^i)) \right], \quad (17)$$

式中 $\{s_p^i\}$ ($i=1, 2, \dots, K$)为类内相似性得分, $\{s_p^j\}$ ($j=1, 2, \dots, L$)为类间相似性得分, r 为缩放

第一条分支将输入的特征图 $(C \times H \times W)$ 以 H 为轴进行逆时针旋转 90° 得到 $(W \times H \times C)$ 的特征图,首先通过Z-Pool单元得到 $(2 \times H \times C)$,再利用 $k \times k$ 的标准卷积层、批归一化层和Sigmoid激活函数层产生 $(1 \times H \times C)$ 的注意力掩模,然后通过残差连接与 $(W \times H \times C)$ 的特征图元素相乘得到通道维度与空间高维度的注意力热图,再将特征图进行顺时针旋转 90° 得到 $(C \times H \times W)$,第二条分支与其类似。第三条分支只需捕捉空间维度的信息,无需旋转,得到空间注意力效果图。最后通过将3条分支的注意力热图进行平均得到最终的注意力效果图。图3分别给出了不同维度注意力掩模作用后的注意效果图,该效果图是将四维张量在第1维度压缩可视化得到的,展示了不同维度信息交互的过程与结果。

系数, m 是相似性分类间隔。如果 w_j 为第 j 个非目标类别的权重向量,样本表示为 x ,将 $s_n^j = w_j^T x / (\|w_j\| \|x\|)$, $s_p = w_y^T x / (\|w_y\| \|x\|)$ 代入式(17)则式(17)退化成AM-Softmax损失:

$$L_{\text{am}} = \log \left[1 + \sum_{j=1}^{N-1} \exp(r(s_n^j + m)) \exp(-rs_p) \right] = -\log \frac{\exp(r(s_p - m))}{\exp(r(s_p - m)) + \sum_{j=1}^{N-1} \exp(rs_n^j)}, \quad (18)$$

该损失函数对 s_n 和 s_p 优化梯度相等,反向传播时

的惩罚项是一样的,因此不利于寻找最优点,优化方式缺乏灵活性。Circle Loss 提供一个能够灵活优化目标的损失函数,其公式如式(19)所示:

$$L_{\text{circle}} = \log \left[1 + \sum_{i=1}^K \sum_{j=1}^{L_i} \left(r \left(a_n^j s_n^j - a_p^i s_p^i \right) \right) \right] = \log \left[\sum_{j=1}^{L_i} \exp \left(r a_n^j s_n^j \right) \sum_{i=1}^K \exp \left(-r a_p^i s_p^i \right) \right], \quad (19)$$

式中 a_n^j 、 a_p^i 为非负权值。此时优化目标为 $(a_n^j s_n^j - a_p^i s_p^i)$, 若相似性得分距最优点较远,则通过较大的梯度进行梯度下降。自适应更新的权重如式(20)所示:

$$\begin{cases} a_p^i = [O_p - s_p^i]_+ \\ a_n^j = [s_n^j - O_n]_+ \end{cases}, \quad (20)$$

式中 O_n 和 O_p 为 s_n^j 、 s_p^i 的最优点。

4 实验结果与分析

4.1 实验环境与数据集

实验运行环境为 Ubuntu 16.04 操作系统, GPU 型号为:2块 NVIDIA GeForce GTX 1080TI (11G 显存), 基于 Pytorch 1.3 深度学习框架。采用的数据集为 MOT 数据集, MOT 数据集分为 MOT15、MOT16 和 MOT17, 该数据集包含了静止或者移动拍摄、低中高角度拍摄以及黑夜等复杂的环境。实验首先在 CrowdedHuman 数据集进行预训练, 然后通过 MOT16 的训练集进行训练, 在 MOT15 的训练集进行消融实验。对比实验在 MOT15 的训练集训练, 通过 MOT15 测试集测

试。实验超参数设置如表 1 所示, 实验评价指标如表 2 所示。

表 1 实验超参数设置

Tab. 1 Experimental hyperparameter settings

实验超参数	未改进 ReID Loss 的实验	改进 ReID Loss 的实验
优化器	Adam	Adam
动量参数	0.9	0.9
Batchsize	6	6
初始学习率	1e-4	1e-4
训练轮数	45	70

表 2 多目标跟踪评价指标

Tab. 2 Multi-target tracking evaluation index

指标	含义
MOTA	跟踪精度, 衡量目标轨迹的保持程度
MOTP	衡量检测器的定位性能
IDF1	衡量跟踪器身份维持能力
MT	整个视频中超过 80% 的时间被正确跟踪的轨迹个数
ML	整个视频中不超过 20% 时间被正确跟踪的轨迹个数
IDS	目标身份的切换次数
FM	轨迹碎片的个数

4.2 消融实验

对 Fairmot 模型、采用注意力机制的 Fairmot (Fairmot+A) 模型和采用 Circle Loss 和注意力机制的 Fairmot (Fairmot+A+CL) 模型进行实验, 结果如表 3 所示。

表 3 3 种模型的消融实验在 MOT15 训练集上的测试结果

Tab. 3 Ablation experiments of the three models tested on the MOT15 training set

模型	MOTA ↑ / %	IDF1 ↑ / %	MOTP ↑ / %	MT ↑	ML ↓	IDS ↓	FM ↓	Recall ↑ / %	Prec ↑ / %
Fairmot	59.6	67.7	77.1	171	172	255	1 439	70.7	70.7
Fairmot+A	60.7	66.5	76.8	173	177	209	1 436	71	87.2
Fairmot+A+CL	62.9	64.3	77.5	255	123	232	1 173	78	84.3

采用三重注意力机制后, MOTA 得到了 1.1% 的提升, 且身份切换次数明显降低。注意力机制能够提供更可靠的目标检测, 从而提升跟踪精度。Fairmot+A+CL 在采用 Circle Loss 后相比原模型在 MOTA 上提升 3.3%, 且在 MOTP、

MT、ML、FM 等指标上明显优于原模型。但 Fairmot+A+CL 模型与 Fairmot+A 模型相比, IDS 指标上升许多, 可能采用 Circle Loss 之后对多任务学习目标检测分支的性能产生影响, 产生漏检或虚检现象, 使改进后的模型身份切换指标

上升。

图4展示了3种模型的跟踪能力对比。对于图4(a)蓝色箭头所指的女士,Fairmot模型上只在第一个视频帧中检测到该女士,在后续的视频帧中出现部分遮挡未检测出该行人。待遮挡结束时,行人身份发生切换。Fairmot+A模型在前两幅视频帧中跟踪到该女士,采用注意力机制能够明显提高其跟踪精度,但在遮挡结束时,行人的身份发生切换。Fairmot+A+CL模型在全程视频帧中均跟踪到该女士。可见,引入Circle Loss之后,增强了行人重识别分支的特征提取能力,使其能够提取更精确的表观特征。

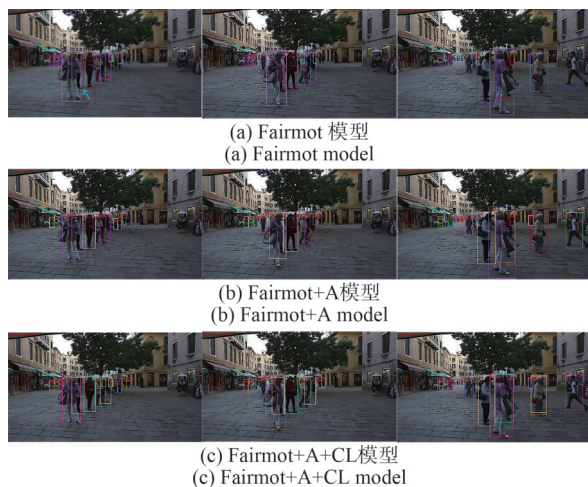


图4 3种模型在MOT15训练集上的测试结果

Fig.4 Test results of the three models on the MOT15 training set

4.3 对比实验

如表4所示,改进后的模型在MOTA、IDF1和MT上要明显优于其他4种模型。与原模型相比,MOTA提升1.4%,MT得到稍许提升。引入注意力机制和Circle Loss之后,提高了对目标的定位能力和跟踪能力,使得提取的表观特征更具区分性。

如图5所示,改进模型在目标检测和跟踪上明显优于其他4种模型。对远处的小目标,改进模型跟踪效果最佳。在第195帧中,改进模型能准确检测出坐在左侧的行人,而原模型却未检测出,表明三重注意力机制和Circle Loss增强了对目标的定位能力和表观特征表达能力,产生了较好的跟踪效果(图6)。

表4 5种模型在MOT15测试集上的对比实验

Tab.4 Comparative experiments of five models on the MOT15 test set

模型	MOTA ↑ / %	IDF1 ↑ / %	MT ↑	ML ↓	IDS ↓
DMT ^[27]	44.5	49.2	250	159	684
MDP-Subcnn ^[28]	47.5	55.7	216	134	628
TRID ^[29]	55.7	61	293	105	351
Fairmot	60.6	64.7	343	79	591
Fairmot+A+CL	62.0	65.1	358	88	523

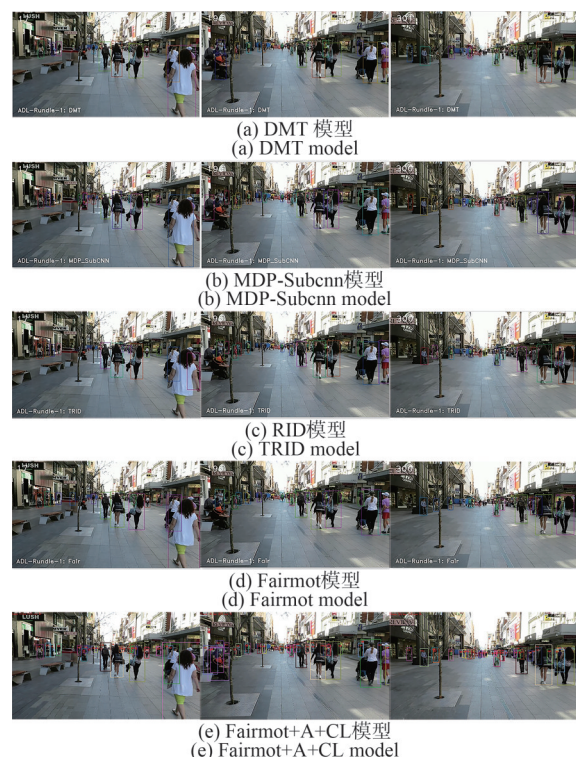


图5 5种模型在MOT15测试集上的对比效果图

Fig.5 Comparison of the five models on the MOT15 test set

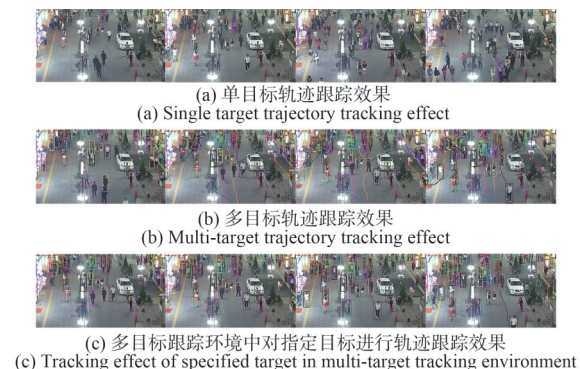


图6 轨迹跟踪功能展示

Fig.6 Display of trajectory tracking function

5 结 论

本文对 Fairmot 框架提出两种改进措施,首先利用三重注意力机制提高对高维信息的维度交互能力,产生精确定位;然后通过 Circle Loss 损失函数优化行人重识别分支,使其根据当前距

最优点的距离选择优化目标和程度,提取更精确的表观特征。实验结果表明,本文所提模型明显优于其他模型,在 MOT15 测试集上的跟踪精度为 62%,IDF1 提升至 65.1%,身份切换降低 68 次。但是对于长时间遮挡的目标,本文方法会发生身份切换,产生较多的轨迹碎片,未来将着重研究长时遮挡问题以及模型压缩问题。

参 考 文 献:

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego, CA, USA: IEEE, 2005: 886-893.
- [2] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model [C]//2008 *IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN [C]//2015 *IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, 2015: 1440-1448.
- [5] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [6] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 6517-6525.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv*:1804.02767, 2018.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. *arXiv*: 2004.10934, 2020.
- [10] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 734-750.
- [11] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points [J]. *arXiv*: 1904.07850, 2019.
- [12] XING J L, AI H Z, LAO S H. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses [C]//2019 *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE, 2009: 1200-1207.
- [13] WOLF J K, VITERBI A M, DIXON G S. Finding the best set of K paths through a trellis with application to multitarget tracking [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1989, 25(2): 287-296.
- [14] CHOI W, SAVARESE S. A unified framework for multi-target tracking and collective activity recognition [C]//*Proceedings of the 12th European Conference on Computer Vision*. Florence, Italy: Springer, 2012: 215-230.
- [15] YANG B, HUANG C, NEVATIA R. Learning affinities and dependencies for multi-target tracking using a CRF model [C]//*The 2011 IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs, CO, USA: IEEE, 2011: 1233-1240.
- [16] BRENDDEL W, AMER M, TODOROVIC S. Multiobject tracking as maximum weight independent set [C]//*Proceedings of 2011 Conference on Computer Vision and Pattern Recognition*. Colorado Springs, CO, USA: IEEE,

- 2011: 1273-1280.
- [17] REID D. An algorithm for tracking multiple targets [J]. *IEEE Transactions on Automatics Control*, 1979, 24(6): 843-854.
- [18] MITZEL D, LEIBE B. Real-time multi-person tracking with detector assisted structure propagation [C]//2011 *IEEE International Conference on Computer Vision Workshops*. Barcelona, Spain: IEEE, 2011: 974-981.
- [19] HU W M, LI X, LUO W H, *et al.* Single and multiple object tracking using Log-Euclidean Riemannian subspace and block-division appearance model [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(12): 2420-2440.
- [20] SUN S, AKHTAR N, SONG H S, *et al.* Deep affinity network for multiple object tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(1): 104-119.
- [21] XU Y H, BAN Y, ALAMEDA-PINEDA X, *et al.* DeepMOT: a differentiable framework for training multiple object trackers [J]. *arXiv*: 1906.06618, 2019.
- [22] ZHU J, YANG H, LIU N, *et al.* Online multi-object tracking with dual matching attention networks [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 379-396.
- [23] ZHANG Y F, WANG C Y, WANG X G, *et al.* FairMOT: on the fairness of detection and re-identification in multiple object tracking [J]. *International Journal of Computer Vision*, 2021, 129(11): 3069-3087.
- [24] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C]//2017 *IEEE International Conference on Image Processing (ICIP)*. Beijing, China: IEEE, 2017: 3645-3649.
- [25] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al.* Rotate to attend: convolutional triplet attention module [C]//*Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE, 2021: 3138-3147.
- [26] SUN Y F, CHENG C M, ZHANG Y H, *et al.* Circle loss: a unified perspective of pair similarity optimization [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 6397-6406.
- [27] KIM H U, KIM C S. CDT: cooperative detection and tracking for tracing multiple objects in video sequences [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, 2016: 851-867.
- [28] XIANG Y, ALAHI A, SAVARESE S. Learning to track: online multi-object tracking by decision making [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015: 4705-4713.
- [29] MANEN S, GYGLI M, DAI D X, *et al.* PathTrack: fast trajectory annotation with path supervision [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE, 2017: 290-299.

作者简介:



席一帆(1995—),男,山西运城人,硕士研究生,2018年于太原科技大学获得学士学位,主要从事目标检测和多目标跟踪研究。E-mail: 15735169750@163.com