DOI:10.11830/ISSN.1000-5013.202105018

结合目标检测与特征匹配的多目标跟踪算法

叶靓玲[1],李伟达[1],郑力新[1],曾远跃[2],黄凯[2]

(1. 华侨大学 工学院,福建 泉州 362021;
2. 福建省特种设备检验研究院 泉州分院,福建 泉州 36021)

摘要: 针对多目标跟踪算法在遮挡频繁的场景下存在目标关联准确性低的问题,提出一种结合检测与特征匹配的多目标跟踪算法. 该算法引入检测精度较高的 YOLOv5 作为多目标跟踪的检测器,能够精准定位目标,有效提高跟踪精度;在面对目标间遮挡时,通过专门设计特征匹配模型提取更为细致的特征,能够有效降低跟踪时目标 ID 的切换次数. 在 MOT16 数据集上对跟踪性能进行评估,结果表明:所提方法可以有效缓解目标遮挡,实现稳定跟踪.

关键词: 多目标跟踪;目标检测;特征匹配;深度学习;YOLOv5

中图分类号: TP 391.41 **文献标志码:** A **文章编号:** 1000-5013(2021)05-0661-09

Multiple Object Tracking Algorithm Based on Detection and Feature Matching

YE Liangling[1], LI Weida[1], ZHENG Lixin[1], ZENG Yuanyue[2], HUANG Kai[2]

(1. College of Engineering, Huaqiao University, Quanzhou 362021, China;
2. Quanzhou Branch, Fujian Special Equipment Inspection and Research Institute, Quanzhou 362021, China)

Abstract: Aiming at the problem that the multiple object tracking algorithm (MOT) had low accuracy of target association in frequent occlusion scenes, an MOT algorithm based on detection and feature matching is proposed in this paper. This algorithm introduces YOLOv5 with high detection accuracy as a detector for MOT, which can accurately locate the target and effectively improve the tracking accuracy. In addition, a feature matching model is specially designed when facing the goals keep out. This can extract more detailed features and effectively reduce the ID switching numbers during tracking. The tracking feature is evaluated on the MOT16 dataset, and the results show that the proposed algorithm can effectively alleviate the occlusion of the target and achieve stable tracking.

Keywords: multiple object tracking; target detection; feature matching; deep learning; YOLOv5

多目标跟踪是计算机视觉和公共安全领域的热点研究问题[1-2],已被广泛应用于各行各业,如自动驾驶、行为分析和智能监控等. 但在复杂场景下,多目标的实际应用仍面临许多挑战,如相似目标的区分、目标遮挡[3]、目标在镜头前的突然产生和消失等. 为了解决这些问题,李星辰等[4]针对目标遮挡,提

收稿日期: 2021-05-14

通信作者: 郑力新(1967-),男,教授,博士,主要从事运动控制、机器视觉、图像处理与模式识别的研究. E-mail:zlx@hqu.edu.cn.

基金项目: 福建省科技计划项目(2020Y0039);福建省泉州市高层次人才创新创业项目(2020C042R)

出目标轨迹修正的相应策略,在一定程度上能够缓解遮挡带来的目标身份切换问题.但这种依靠运动轨迹预测模型进行数据关联,在面对目标过多的复杂场景时,跟踪的准确性仍有待提高;Wang等[5]利用方向梯度直方图(histogram of oriented gradient,HOG)进行在线外观特征判断,只能保证局部的高效特征,对全局的轨迹关联仍有一定影响.针对上述的问题,本文基于深度学习,提出一种结合检测和特征匹配的多目标跟踪算法框架.

1 目标检测和特征匹配的多目标跟踪方法

多目标跟踪算法框架,如图1所示.该框架主要由3个部分组成,即目标检测模型、特征匹配模型和跟踪模型.首先,将图像 I_{t+1} 输入到目标检测模型中,通过YOLOv5检测器获得每一帧中目标边界框的坐标信息 $\boldsymbol{D}_{t+1}$;其次,将 I_{t+1} 输入到预训练好的特征匹配模型中得到128维的特征向量 $\boldsymbol{F}_{t+1}$;最后,将 $\boldsymbol{F}_{t+1}$ 与特征空间 $\boldsymbol{S}_i$ 中已保存的特征向量进行特征匹配(其中 i 代表特征空间中存储的图像帧数).若匹配值越大,代表二者为相同目标的可能性更大;反之,则为相同目标的可能性更小.预测 I_t 帧中所有目标在 $T+1$ 帧中的位置,记为 $\boldsymbol{P}_{t+1}$,并将 $\boldsymbol{P}_{t+1}$ 和目标检测信息 $\boldsymbol{D}_{t+1}$ 与特征匹配的结果进行轨迹关联,从而得到完整的目标跟踪轨迹.

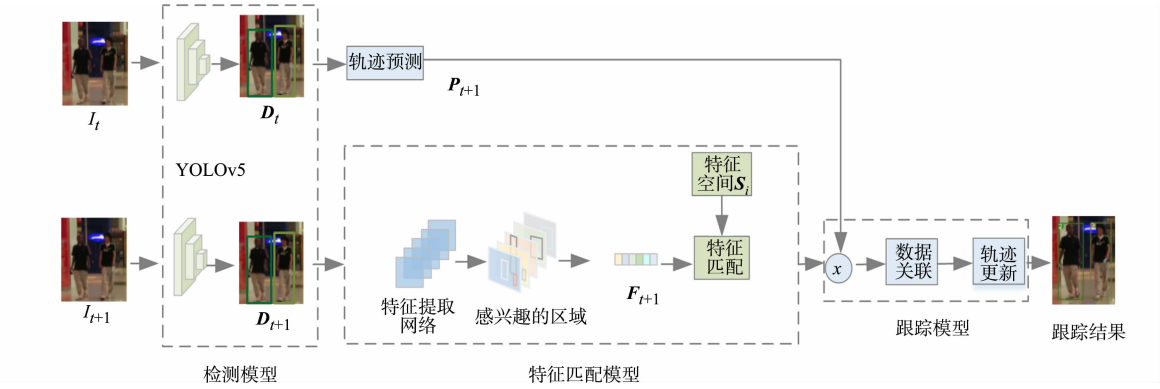


图1 多目标跟踪算法框架图

Fig. 1 Structural framework of multiple object tracking

1.1 目标检测模型

Bewley等[6]提到检测算法的精度能直接影响多目标跟踪的准确性,因此将现阶段检测精度较高的YOLOv5引入多目标跟踪算法中.利用YOLOv5来准确定位并获取目标边界框的坐标信息 $\boldsymbol{D}$,以便作为后续跟踪模型的输入.首先将输入图像按照一定的尺度标准划分成单元格,该单元格负责预测目标边界框的坐标信息.

为了得到更多大小不一的尺度,使检测网络具有更好的鲁棒性,在输入端中使用Mosaic数据增强技术随机将4张图片进行缩放、裁剪和分布.同时,为了减少输入图像的冗余信息,在输入端加入自适应图片缩放技术来自动地添加最少的黑边,达到减少信息的目的.定义原始图像的尺寸为 $M\times N$,预计缩放的尺寸为 $P\times P$,具体的自适应图片缩放的计算式为

$$\left.\begin{aligned} &\alpha=\min\{\alpha_1,\alpha_2\}, \quad \alpha_1=P/M, \quad \alpha_2=P/N,\\ &M'=M\times\alpha, \quad N'=N\times\alpha, \quad H=M'-N',\\ &G=\mathrm{mod}(H,32), \quad X=G/2. \end{aligned}\right\} \tag{1}$$

式(1)中:α 是缩放系数;$M'\times N'$ 是缩放后的图像尺寸;H 是原始图像需要填充的高度;G 是像素个数,通过 H 对32取余得到;X 是图像两端需要填充的数值大小.

为了达到下采样的同时不丢失信息,在图像输入主干网络前,对其进行切片操作.因此在主干部分,采用Focus结构来形成图像特征.具体Focus结构的关键操作,如图2所示.

为了进一步加强检测网络特征融合的能力,在Neck部分[7]加入CSPNet网络设计的CSP2结构[8].CSP2将基础层的特征映射分成两部分,然后利用跨阶层次将二者合并.具体的CSP2结构,如图3所示.其中每个CBL模块分别由卷积、BN(batch normalization)和Leaky ReLu激活函数组成,最后输

出端得到每帧中目标边界框的坐标信息 $\boldsymbol{D}$. 在后续多目标跟踪中，对利用 YOLOv5 得到的目标信息进行跟踪.

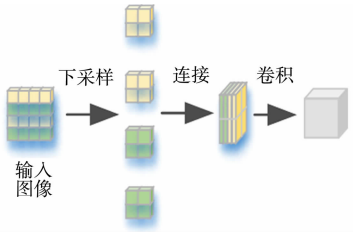


图 2　Foucs 切片操作图

Fig. 2　Foucs slice operation

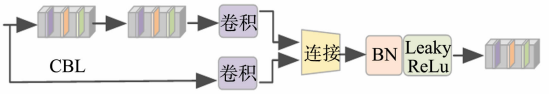


图 3　CSP2 结构

Fig. 3　Structure of CPS2

1.2　特征匹配模型

多目标跟踪中目标特征的提取，可以看作是行人重识别网络(person re-identification，ReID)的具体实现[9]. Wojke 等[10]指出 ReID 能够有效缓解目标遮挡，提高关联的准确性，对关联轨迹可起到有效的辅助作用. 因此，文中利用深度学习强大的表征输出能力，设计一个 ReID 网络作为多目标跟踪的特征匹配模型. 该特征匹配模型以 ResNet50[11]作为主干网络.

特别针对多目标跟踪场景，文中做如下 3 点改进. 1) 为确保网络能捕捉更细致更全面的底层特征，在浅层网络中采用更宽的网络宽度和更大的卷积核(5×5). 虽然采用更大的卷积核会增加少量计算开销，但相应将步长设置为 2，步长较大，后续参与计算的图像会变小，这样可以有效减少模型的计算复杂度. 2) 为有效减少特征的损失，再次采用步长为 2 的卷积操作代替网络中的最大池化操作. 3) 为防止特征在传递过程中的损失，在特征匹配模型中不再使用 ReLU 激活函数，转而使用线性激活函数.

提取目标的特征过程，如图 4 所示. 即先输入一张图像 I_t，经过改进后的特征提取网络，得到输入图像的感兴趣区域，最后经过一个全局平均池化层将提取到的特征映射成一个 128 维的一维向量.

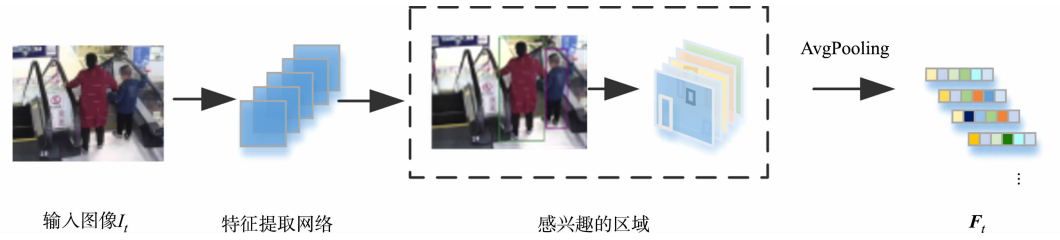


图 4　目标特征的提取过程

Fig. 4　Extracting process of object feature

由于多目标跟踪中目标特征差异性小，为了更好识别目标特征，判断前后两帧中的跟踪目标是否属于同一身份，使用 Triplet 损失[12]进行度量学习，选择 Market 1501 数据集[9]进行训练. Triplet 损失可以使具有相同标签的样本在嵌入空间里尽量接近，不同标签的样本在嵌入空间中尽量远离. 具体地，将 Triplet 损失定义为：输入一个三元组$\langle a,p,n\rangle$，其中 a 是 anchor 锚点[13]，是从训练集中随机选取的一个样本，p 是与 a 同类别的样本，n 是与 a 不同类别的样本. 即将 anchor 作为一个锚点，通过学习后，使得同类样本 p 更加接近 a，而不同类样本 n 远离 a. 在嵌入空间中，这个三元组应该满足

$$L = \max(d(a,p) - d(a,n) + \text{margin}, 0). \tag{2}$$

式(2)中：阈值 margin 是衡量样本相似度的重要指标. 较大的阈值可以增强模型对不同类样本的区分度，较小的阈值则不能有效区分同类样本. 因此，在训练初期先选择一个较小的阈值，接着再针对测试的结果对阈值进行增大或缩小的调整.

1.3　跟踪模型

跟踪模块中进行前后两帧之间多个目标的轨迹关联，主要包括判断新目标的出现、处理旧目标的消失和匹配前后两帧间目标的 ID. 因此可以将前后两帧之间的轨迹关联看作是求最优解的过程. 即将通过目标检测模型得到的目标检测信息 $\boldsymbol{D}$ 看作是一个图解空间，第 $T+1$ 帧中的所有检测框记为 $\boldsymbol{D}_{t+1}$，第 T 帧中的所有预测框记为 $\boldsymbol{P}_{t+1}$. 由于同帧不会存在相同目标，因此同帧中的目标不能进行轨迹关联，故

匹配关联时只需要关注前后两帧间的目标.然而,由于每个目标的匹配地位不相同,需要对每个目标赋予相应的权重,为此引入KM算法[14]来求解前后两帧中目标的最优匹配轨迹.

文中利用KM算法进行轨迹的关联,采用通过目标检测模型得到的检测信息 $\boldsymbol{D}_{t+1}$ 与预测框 $\boldsymbol{P}_{t+1}$ 的交并比(intersection over union,IoU)作为不同目标的占比权重.IoU的计算式为

$$\mathrm{IoU}=\frac{S_{\boldsymbol{D}_{t+1}}\cap S_{\boldsymbol{P}_{t+1}}}{S_{\boldsymbol{D}_{t+1}}\cup S_{\boldsymbol{P}_{t+1}}}. \tag{3}$$

式(3)中:$S_{\boldsymbol{D}_{t+1}}$ 是第 T 帧检测框的面积;$S_{\boldsymbol{P}_{t+1}}$ 是第 $T+1$ 帧预测框的面积.IoU的值越接近1,表明检测框与预测框的关联性越大,意味着二者是相同目标的可能性越大.当IoU大于一定阈值时,认为是相同目标.文中的阈值选为0.4.首先,对每个目标赋值,将 $\boldsymbol{D}_{t+1}$ 中的目标赋值为与其相邻目标的最大权重值;$\boldsymbol{P}_{t+1}$ 中赋值为0.接着利用KM匹配原则对 $\boldsymbol{D}_{t+1}$ 中目标值与其相邻权重值相同的目标进行匹配.

2 实验结果与分析

为了验证文中算法具有良好的跟踪性能,在多目标跟踪数据集MOT16[15]上进行测试评估.实验平台为Linux服务器,python3.8和pytorch编程实现上述算法.在NVIDIA GeForce GTX Titan GPU上进行特征匹配模型的训练和多目标跟踪算法的测试.

文中所提多目标跟踪算法属于SDE框架[16],即目标检测和特征匹配是两个独立的阶段.因此,在实验中,选择阶段式训练.首先,在多目标跟踪的检测模型中引入精度更高的YOLOv5.在特征匹配模型中,用Market1501数据集的行人身份标注训练文中所提的特征匹配模型.模型采用Adam优化器训练60个周期epoch,学习率为设置为0.001,batch_size设置为64.

2.1 实验数据集

MOT16(multiple object tracking 16)是在2016年被Milan[15]提出,主要用于衡量多目标跟踪算法性能的,也是多目标跟踪领域中最具有挑战的数据集之一,包括静态摄像机和动态摄像机拍摄的7个不同场景,共11 235张图像.标注的主要目标为行人和车辆.图5为MOT16部分场景.从图5可知:每个场景都拥有丰富的画面信息,包含多个行人目标,目标间存在严重遮挡、光照变化和复杂天气等挑战.因此,利用MOT16数据集对文中提出的算法进行验证,可以进一步说明在复杂场景下该算法有较好的泛化能力和鲁棒性.

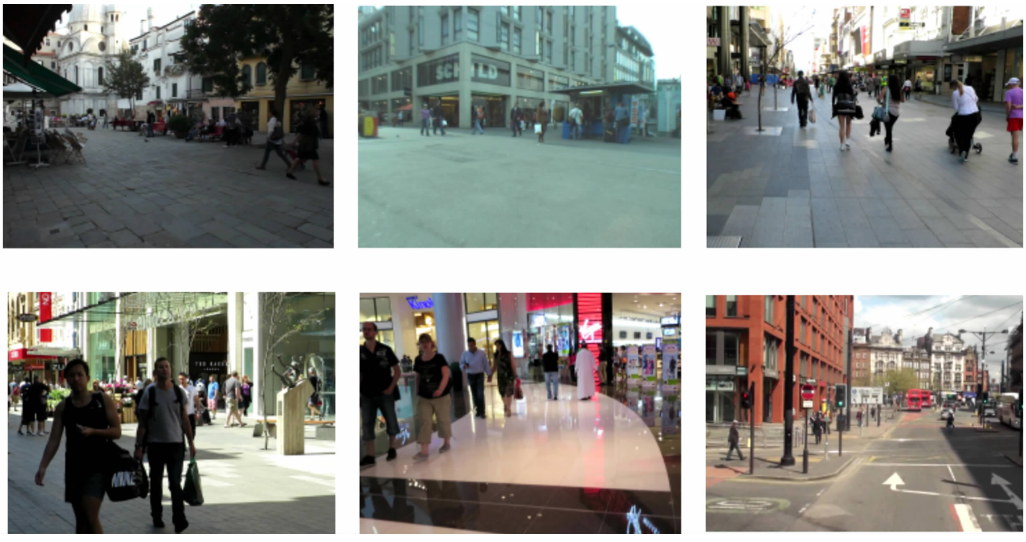


图5 MOT16部分场景

Fig. 5 Some scenes in MOT16

2.2 检测模型对跟踪器性能的影响

采用多目标跟踪领域流行的评估标准对算法进行评估.对跟踪影响最大的四个指标分别为:多目标跟踪准确度(multiple object tracking accuracy,MOTA)、多目标跟踪精度(multiple object tracking pre-

cision，MOTP）、识别 F1 分数（identification F-score，IDF1）和目标身份切换次数（identity switches，IDs）. 为了验证 YOLOv5 检测算法对跟踪器的有效性，文中挑选了 3 种当下检测效果较好的检测器（FasterR-CNN[17]，YOLOv3[7]，YOLOv5），并组合两种不同的特征匹配模型（WRN[18]，ResNet50[11]）在 MOT16-05 序列中对其进行评估，结果分别如表 1，2 所示. 每一组评估中只设置一个变量，以保证算法的可信度.

表 1　不同检测算法组合相同的 WRN 特征匹配模型

Tab. 1　Different detection algorithms are combined with same WRN feature matching model

算法	MOTA ↑	MOTP ↑	IDs ↓	IDF1 ↑
FasterR-CNN＋WRN	26.9	67.5	65	34
YOLOv3＋WRN	43.5	70.1	58	35.4
YOLOv5s＋WRN	47.5	74.1	62	59.5
YOLOv5x＋WRN	46.8	75.3	60	58.8

表 2　不同检测算法组合相同的 ResNet50 特征匹配模型

Tab. 2　Different detection algorithms are combined with same ResNet50 feature matching model

算法	MOTA ↑	MOTP ↑	IDs ↓	IDF1 ↑
FasterR-CNN＋ResNet50	29.8	67.7	66	38.2
YOLOv3＋ResNet50	44.3	70.1	56	39.4
YOLOv5s＋ResNet50	47.9	74.8	49	41.9
YOLOv5x＋ResNet50	48.2	75.7	50	58.2

MOTA 是评估跟踪准确性的重要指标，而 MOTP 是衡量检测器的定位精度. 从表 1 可知：引入 YOLOv5 检测算法（YOLOv5s＋WRN 和 YOLOv5x＋WRN）时，MOTA，MOTP 和 IDF1 的得分更高. 这说明，检测器性能的好坏在一定程度上能够影响跟踪的鲁棒性；采用 YOLOv5 检测器进行跟踪时产生的目标身份切换次数（IDs）会略高于 YOLOv3，但并不能单独利用 IDs 的得分来评估跟踪性能的优异. 有时可能出现目标身份切换次数较少，但产生较多的轨迹片段，跟踪的稳定性也会受到影响. 因此，需要结合 IDF1，IDs 和 MOTA 的得分来评估跟踪的鲁棒性才更具有说服力.

从表 2 可知：用与表 1 相同的 3 个检测器搭配 ResNet50[11] 为主干的特征匹配模型进行跟踪，结果表明使用 YOLOv5 的 MOTA 得分更高. 这说明，将 YOLOv5 作为多目标跟踪的检测器引入到 MOT 中，能够有效地提高 MOT 的跟踪精度.

综合表 1，2 可知：通过比较 WRN 和 ResNet50 两个特征匹配模型在不同检测器上（FasterR-CNN，YOLOv3，YOLOv5s 和 YOLOv5x）的影响，使用 ResNet50 的特征匹配模型在 MOTA，MOTP，IDs 等 3 个指标上的得分相对更高. 特别是当检测器选择 YOLOv5 时，使用 ResNet50（表 2 中 YOLOv5s＋ResNet50 和 YOLOv5x＋ResNet50）相比较于使用 WRN（表 1 中 YOLOv5s＋WRN 和 YOLOv5x＋WRN）的特征匹配模型，IDs 减少约 20％和 17％. 这说明，当目标面对遮挡时，基于 ResNet50 的特征匹配模型能够正确关联到相应目标的可能性更高，跟踪稳定性相对更好. 造成两个特征匹配模型有如此差异的原因，是因为 ResNet50 的网络层数更深，能够提取更细致更全面的目标外观特征，故当目标存在遮挡时，也能够有效减少目标间的身份切换问题. 因此，将 ResNet50 作为特征匹配模型的主干网络能够在目标的数据关联上起到一定的辅助作用.

2.3　特征匹配模型的性能评估

特征匹配模型是基于 ReID，通过分类网络来具体实现的. 因此可以利用 ReID 中的性能指标 mAP（mean average precision）和 rank＝1 作为评估特征匹配模型的性能. 利用 WRN 和 ResNet50 探究特征匹配模型对跟踪性能的影响，可以看出，在检测器相同的情况下，ResNet50 在跟踪性能上取得的效果更好. 为了进一步充分利用 ResNet50 强大的特征提取能力，文中提出改进的特征匹配模型. 即在浅层网络中采用更宽的网络宽度和更大的卷积核（5×5），并将步长设置为 2 来减少卷积操作带来的计算开销. 最后，在特征匹配模型中再次使用步长为 2 的卷积代替最大池化操作，利用线性激活函数来防止特征在传递过程中的损失.

改进的特征匹配模型在 Market1501 上训练 60 个周期 epoch 后得到的函数损失，如图 6 所示. 图 6

中:L 代表的是 Triplet 损失函数在 Market1501 数据集上训练时的收敛过程;e_{top1} 代表的是 top1 的错误率. 从图 6 可知:Triplet 损失函数在训练开始后逐渐收敛.

进一步选用当下流行且性能优越的 ReID 算法(SPReID[19],BFE[20],Mancs[21])与文中改进的特征匹配模型进行评估,结果如图 7 所示. 图 7 中:A 为精度. 从图 7 可知:文中通过 ResNet50 改进的特征匹配模型在 mAP 和 rank1 上的得分明显优于其他算法. 这说明文中所提的特征匹配模型能够有效提取到目标的细致特征.

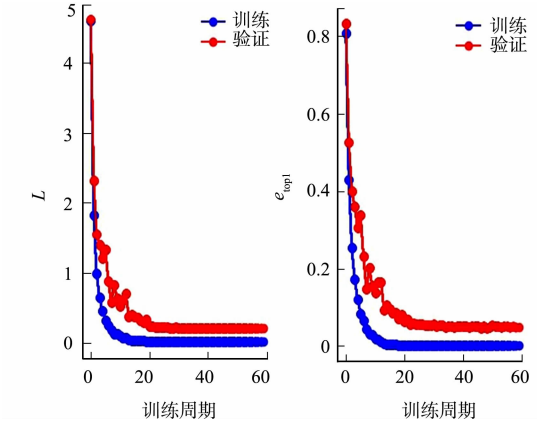


图 6 特征匹配模型训练的损失函数图

Fig. 6 Loss function of feature matching model training

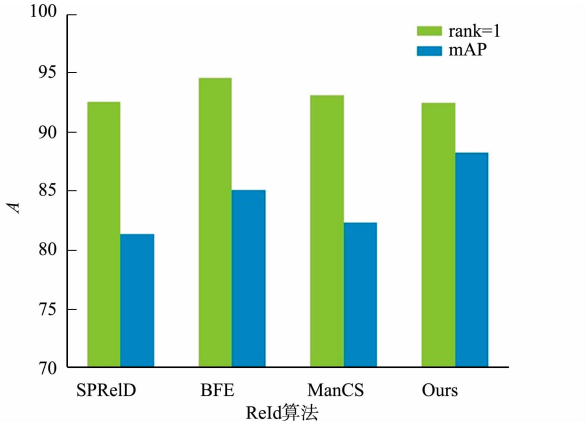


图 7 特征匹配模型与性能优异的 ReID 算法的比较

Fig. 7 Comparison of feature matching model with ReID algorithm in excellent performance

为了进一步验证引入的 YOLOv5 检测模型与改进的特征匹配模型对提高跟踪器的准确性具有一定的效果. 文中所提算法在检测模型部分选择 YOLOv5s 模型,在特征匹配模型部分,以 ResNet50 为主干网络进行网络结构的调整并训练后得到特征匹配模型. 将文中所提算法在 MOT16-05 序列上与上节中表现优异的组合相对比,得到的评估结果,如表 3 所示.

从表 3 可知:文中所提算法在检测器上选择了精度更高的 YOLOv5,相比较于 FasterR-CNN+WRN,多目标跟踪的准确性 MOTA 提高了约 20%,IDs 减少了近 24%. 即引入 YOLOv5 能够提升一定的跟踪精度. 更重要的是,面对频繁遮挡的场景,利用文中改进的特征匹配模型可以提取更为鲁棒的特征,有效减少目标间的身份切换次数和提高 IDF1 的得分.

虽然文中所提算法在 MOT 中引入特征匹配模型来关联目标轨迹,但并没有引入额外的计算开销. 针对 ResNet50 改进的特征匹配模型,文中算法在模型参数量 Params 和浮点运算次数(floating point of operations,FLOPs)上都优于其余算法. 其中,FLOPs 可以衡量多目标跟踪算法复杂度. 计算目标检测模型和特征匹配模型复杂度时,输入图像的大小分别为 680 px×680 px 和 224 px×224 px. 可见,文中所提算法在没有引入额外开销的同时还能有效缓解遮挡时的身份切换问题并实现稳定跟踪. 这主要得益于文中检测算法引入参数量较小,但检测精度相对较高的 YOLOv5s. 在特征匹配模型中虽然加大了卷积核来提取更细致的特征,但同时可对网络结构采用步长为 2 的设置,来减少参与计算图像的大小. 与 YOLOv5x+ResNet50 相比,文中所提算法虽然会牺牲小部分 MOTA 和 MOTP 的精度,但其参数量和复杂度远比 YOLOv5x+ResNet50 小得多,并且在跟踪稳定性 IDs 和 IDF1 两个指标上也优于 YOLOv5x+ResNet50. 综合来看,文中所提算法在减少推理时间的同时,还能实现稳定的跟踪.

表 3 总体跟踪性能评估(MOT16-05)

Tab. 3 Overall tracking performance evaluation(MOT16-05)

算法	MOTA ↑	MOTP ↑	IDs ↓	IDF1 ↑	Params(Million) ↓	FLOPs ↓ (G)
FasterR-CNN+WRN	26.9	67.5	65	34	—	—
YOLOv3+WRN	43.5	70.1	58	35.4	117.8	552.5
YOLOv5x+ResNet50	48.2	75.7	50	58.2	111.2	282.4
文中所提算法	47.8	74.8	49	59.2	30.9	34.5

2.4 与当下同类跟踪算法进行评估

为了进一步体现文中所提算法具有良好的跟踪性能,利用 MOT16 数据集中的 7 个测试序列进行

全面评估，并将其与当下 MHT-bLSTM[22]，CDA_DDALv2[23]，MTDF[24]，AM_ADM[25]和 OVBT[26]等同类型算法进行对比，结果如表 4 所示.

从表 4 可知：文中所提算法的 MOTP 在所有方法中得分最高，说明文中引入的 YOLOv5 检测器对定位目标的位置起到一定效果；其次，文中所提算法在 IDs 和 IDF1 两个指标上的表现性能也最好，这表明所提出的特征匹配模型能够有效减少目标间身份切换的次数，有利于稳定跟踪. MHT-bLSTM[22]，CDA_DDALv2[23]算法的思想与文中所提算法的思想相似，都对目标的特征进行了建模. 由此可知，在所有算法中，这三者算法的 IDs 和 IDF1 性能表现最好，特别是文中所提算法，在三者中指标得分最高. 这说明，在 MOT 上引入目标特征的建模中，文中所提出特征匹配算法的性能最优，对关联轨迹和维持跟踪的稳定性有一定效果.

表 4　不同方法在 MOT16 测试集上的结果对比

Tab. 4　Comparison of different methods on MOT16 test set

算法	MOTA ↑	MOTP ↑	IDs ↓	IDF1 ↑	算法	MOTA ↑	MOTP ↑	IDs ↓	IDF1 ↑
MHT-bLSTM	42.1	75.9	753	47.8	AM_ADM	40.1	75.4	789	43.8
MTDF	45.7	72.6	1987	40.1	CDA_DDALv2	43.9	74.7	676	45.1
OVBT	38.4	75.4	1321	37.8	文中所提算法	42.7	76.1	655	47.8

文中所提算法在 MOTA 上得分不如 MTDF[24]，这是因为 MOT 是一项复杂的任务，特别是针对 SDE 框架下设计的跟踪算法[16]，往往会受到检测模型、时空信息等因素的影响. MTDF 加入了时空信息，可以在目标相互接近时消除模糊的轨迹关联，因此跟踪的准确性会略高. 但该算法在 IDs 和 IDF1 这两个指标上表现不佳，导致跟踪的稳定性不如文中所提算法. 它在面对遮挡时依然能够正确关联大部分轨迹，维持稳定鲁棒的跟踪. 文中重点关注的是检测器和特征匹配模型对跟踪性能的影响，忽略了时空信息对轨迹关联的影响，后续将进一步加入时空信息来提高多目标跟踪的准确性.

2.5　行人多目标跟踪

为了进一步验证文中所提算法在复杂场景下的行人跟踪效果，利用文中所提算法对 MOT16 测试集中的序列进行跟踪，对得到的跟踪结果随机截取图像帧，如图 8 所示. 从图 8 可知：文中所提算法对 MOT16 测试集场景中出现的目标都成功关联到身份 ID，做到了有效的跟踪.

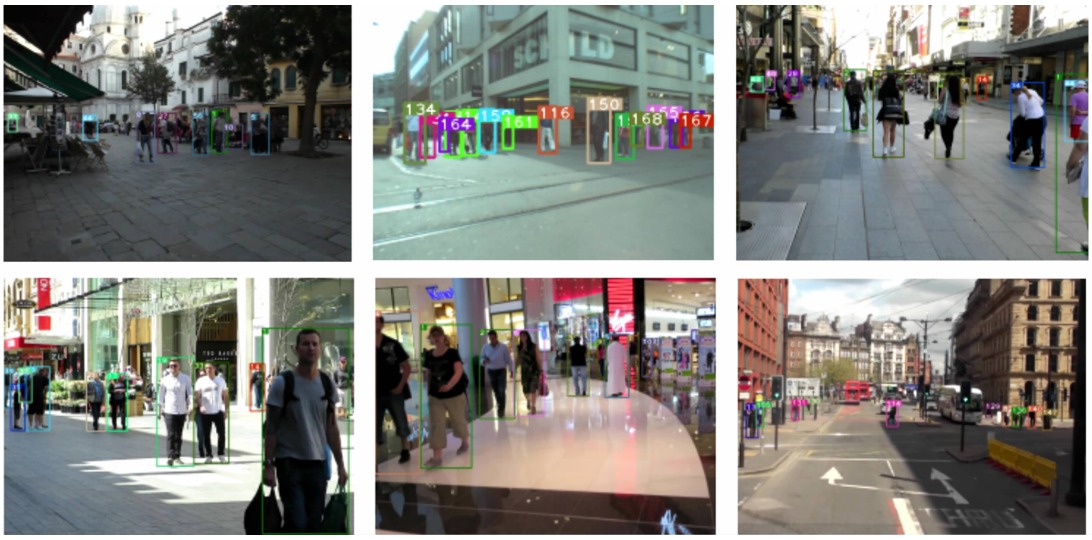
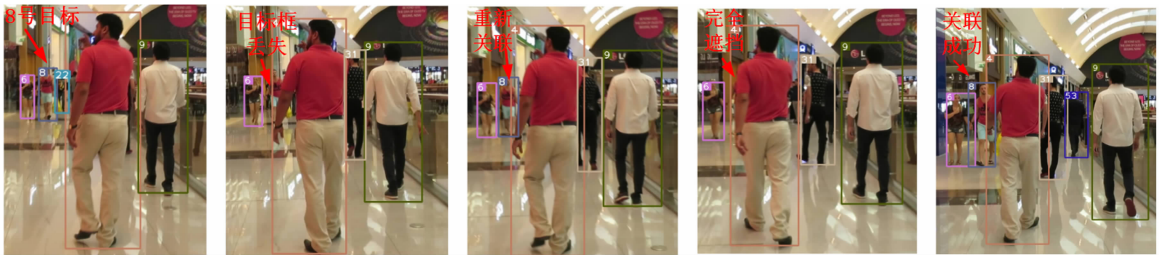


图 8　行人多目标跟踪结果图

Fig. 8　Pedestrian multiple object tracking results

为了进一步说明文中所提算法在遮挡频繁场景下能够有效识别目标的产生和消失，选择 MOT16 中移动相机拍摄的视频序列进行重点说明，如图 9 所示. 在图 9(a)中，利用文中方法有效跟踪到当前场景下出现的目标，并赋予目标相应的 ID 编号，重点分析编号为 8 的目标；在图 9(b)中，8 号的目标框发生丢失，经过一段时间后，8 号目标的 ID 在图 9(c)中被重新正确关联上；在图 9(d)中，8 号目标被完全遮挡，消失在镜头中，经过若干帧后，当 8 号目标重新出现在镜头下时，利用特征匹配模型辅助关联，成功将其外观特征与特征空间中的特征进行匹配计算；在图 9(f)中，利用相同目标的匹配关联值最大的

特点,将8号目标重新关联,得到8号目标完整的行人轨迹.这说明文中结合YOLOv5检测与特征匹配模型的多目标跟踪算法,经过数次遮挡,依然能够稳定维持目标编号,实现稳定的跟踪.



(a) 关注目标 (b) 目标丢失 (c) 重新关联 (d) 完全遮挡 (f) 关联成功

图9 移动场景下的行人多目标跟踪

Fig. 9 Pedestrian multiple object tracking in moving scenarios

3 结论

为了解决多目标跟踪在目标检测精度低和多目标遮挡时存在轨迹匹配难的问题,通过一系列实验探究表明YOLOv5检测器对提高跟踪的准确性有一定效果,因此将YOLOv5引入到多目标跟踪中作为跟踪的检测器.为了提取更全面更鲁棒的特征,提出改进的特征匹配模型,来解决目标间由于遮挡导致的身份切换问题.在MOT16数据集上的评估也表明,文中所提算法在处理遮挡能力和关联轨迹方面都有优异表现,并且能够在维持稳定跟踪的前提下减少相应的推理时间,这为多目标跟踪在实际设备中的应用提供了更大的可能性.

然而,该算法也存在一些问题,如文中所提算法属于SDE算法框架[16],即检测和跟踪分成两个阶段训练,这会对实时性产生一定的影响.下一阶段的研究目标是联合检测和跟踪,并进行多任务训练得到端到端的多目标跟踪网络,从而进一步提高多目标跟踪的实时推理速度.

参考文献:

[1] XIANG Jun, XU Guohan, MA Chao, *et al*. End-to-end learning deep CRF models for multi-object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(1): 275-288. DOI: 10.1109/TCSVT.2020.2975842.

[2] RISTANI E, TOMASI C. Features for multi-target multi-camera tracking and re-identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Press, 2018: 6036-6046. DOI: 10.1109/CVPR.2018.00632.

[3] 顾培婷,黄德天,黄炜钦,等.抗遮挡的相关滤波目标跟踪算法[J].华侨大学学报(自然科学版),2018,39(4):611-617. DOI:10.11830/ISSN.1000-5013.201608031.

[4] 李星辰,柳晓鸣,成晓男.融合YOLO检测的多目标跟踪算法[J].计算机工程与科学,2020,42(4):665-672. DOI:10.3969/j.issn.1007-130X.2020.04.013.

[5] WANG Jiahui, GUO Yulan, TANG Xing, *et al*. Semi-online multiple object tracking using graphical tracklet association[J]. IEEE Signal Processing Letters, 2018, 25(11): 1725-1729. DOI: 10.1109/LSP.2018.2872403.

[6] BEWLEY A, GE Zongyuan, OTT L, *et al*. Simple online and realtime tracking[C]// IEEE International Conference on Image Processing. Phoenix: IEEE Press, 2016: 3464-3468. DOI: 10.1109/ICIP.2016.7533003.

[7] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. [2018-04-08]. https://arxiv.org/abs/1804.02767.

[8] WANG C Y, LIAO H Y M, YEH I H, *et al*. CSPNet: A new backbone that can enhance learning capability of CNN [C]// IEEE Conf-erence on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE Press, 2020: 1571-1580. DOI: 10.1109/CVPRW50498.2020.00203.

[9] ZHENG Liang, ZHANG Hengheng, SUN Shaoyan, *et al*. Person re-identification in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Press, 2017: 3346-3355. DOI: 10.1109/CVPR.2017.357.

[10] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]// IEEE International Conference on Image Processing. Beijing: IEEE Press, 2017: 3645-3649. DOI: 10.1109/ICIP.2017.8296962.

[11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2016: 770-778. DOI: 10.1109/CVPR.2016.90.

[12] WEINBERGER K, SAUL L. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10: 207-244. DOI: 10.5555/1577069.1577078.

[13] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE Press, 2015: 815-823. DOI: 10.1109/CVPR.2015.7298682.

[14] KUHN H W. The hungarian method for the assignment problem[M]. JÜNGER M, LIEBLING T M, NADDEF D, *et al*. 50 Years of Integer Programming 1958－2008: From the Early Years to the State-of-the-Art. Heidelberg: Springer Berlin Heidelberg, 2010: 29-47.

[15] MILAN A, LEAL-TAIXE L, REID I, *et al*. MOT16: A benchmark for multi-object tracking[EB/OL]. (2016-03-02)[2016-05-03]. https://arxiv.org/abs/1603.00831v2.

[16] WANG Zhongdao, ZHENG Liang, LIU Yixuan, *et al*. Towards real-time multi-object tracking[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 107-122. DOI: 10.1007/978-3-030-58621-8_7.

[17] REN Shaoqing, HE Kaiming, 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, 2015, 39(6): 1137-1149. DOI: 10.1109/TPAMI.2016.2577031.

[18] WOJKE N, BEWLEY A. Deep cosine metric learning for person re-identification[C]// IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE Press, 2018: 748-756. DOI: 10.1109/WACV.2018.00087.

[19] KALAYEH M M, BASARAN E, GÖKMEN M, *et al*. Human semantic parsing for person re-identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Press, 2018: 1062-1071. DOI: 10.1109/CVPR.2018.00117.

[20] DAI Zuozhuo, CHEN Mingqiang, GU Xiaodong, *et al*. Batch dropblock network for person re-identification and beyond[C]// IEEE International Conference on Computer Vision. Seoul: IEEE Press, 2019: 3690-3700. DOI: 10.1109/ICCV.2019.00379.

[21] WANG Cheng, ZHANG Qian, HUANG Chang, *et al*. Mancs: A multi-task attentional network with curriculum sampling for person re-identification[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 384-400. DOI: 10.1007/978-3-030-01225-0_23.

[22] KIM C, LI F, REHG J M. Multi-object tracking with neural gating using bilinear LSTM[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 208-224. DOI: 10.1007/978-3-030-01237-3_13.

[23] BAE S, YOON K. Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 595-610. DOI: 10.1109/TPAMI.2017.2691769.

[24] FU Zeyu, ANGELINI F, CHAMBERS J, *et al*. Multi-level cooperative fusion of GM-PHD filters for online multiple human tracking[J]. IEEE Transactions on Multimedia, 2019, 21(9): 2277-2291. DOI: 10.1109/TMM.2019.2902480.

[25] LEE S, KIM M, BAE S. Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures[J]. IEEE Access, 2018, 6: 67316-67328. DOI: 10.1109/ACCESS.2018.2879535.

[26] BAN Y, BA S, ALAMEDA-PINEDA X, *et al*. Tracking multiple persons based on a variational Bayesian model [C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 52-67. DOI: 10.1007/978-3-319-48881-3_5.

（责任编辑： 黄仲一　　**英文审校：** 吴逢铁）