DOI:10.19651/j.cnki.emt.2417298

基于元素相乘结构的实时多目标跟踪算法*

周畅达 杨 帆
(河北工业大学电子信息工程学院 天津 300401)

摘 要: 多目标跟踪算法中的 FairMOT 提出了平衡检测和重识别分支的均衡学习策略,有效的平衡了目标检测和重识别两大任务,是目前单阶跟踪段范式算法中最优的算法,但由于 DLA34 骨干网络的特征提取能力有限,面对实际应用场景中复杂的跟踪场景时,往往会因为出现漏检和误跟等现象导致模型的跟踪效果下降。为了有效的提升模型骨干网络的特征提取能力,本文针对此问题设计了基于元素相乘结构的深度聚合骨干网络,提出了 FairMOT-Star 算法。该算法利用了元素相乘结构带来的隐藏维度提升原理,实现了简洁高效的目标特征提取。同时使用 EIoU_Loss 作为检测框回归任务的回归损失函数,更加精准的描述了检测框和真实框之间的位置和形状关系,提升了检测框的预测精度。匹配关联部分使用卡尔曼滤波算法预测目标的运动信息,匈牙利算法完成时序维度上前后帧目标和轨迹的关联匹配。在 MOT16 数据集上进行了实验测试,MOTA 精度达到了 86.0%,模型的权重参数量为 19.59 M,相比于 FairMOT 模型参数量减少 9.7% 的同时,MOTA 精度提升了 3.5%,较好的优化了 FairMOT 算法的计算参数量和跟踪精度。

关键词: 机器视觉;多目标跟踪;特征提取;FairMOT
中图分类号: TP391.4;TN98 **文献标识码:** A **国家标准学科分类代码:** 510.4

Real-time multi-object tracking algorithm based on star operation

Zhou Changda Yang Fan
(School of Electronics Information Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: FairMOT, a multi-object tracking algorithm, proposes a balanced learning strategy between the detection branch and the re-identification branch, effectively balancing the tasks of object detection and re-identification, thereby improving tracking accuracy. However, due to the limited feature extraction capability of its DLA34 backbone network, the model's tracking performance often declines in complex real-world scenarios, leading to missed detections and incorrect tracking. To enhance the backbone network's feature extraction capability, this paper designs a deep aggregated backbone network based on an element-wise multiplication structure and proposes the FairMOT-Star algorithm. This algorithm leverages the principle of hidden dimension enhancement brought by the element-wise multiplication structure to achieve concise and efficient object feature extraction. Additionally, EIoU_Loss is used as the regression loss function for the bounding box regression task, more precisely describing the positional and shape relationships between detection boxes and ground truth boxes, thus improving prediction accuracy. In the matching and association part, the Kalman filter algorithm predicts target motion information, and the Hungarian algorithm associates and matches targets and trajectories across frames in the temporal dimension. Experimental tests on the MOT16 dataset achieved an MOTA accuracy of 86.0%. The model's weight parameters amount to 19.59 M, reducing parameter count by 9.7% compared to the FairMOT model, while increasing MOTA accuracy by 3.5%, effectively optimizing the computational parameters and tracking accuracy of the FairMOT algorithm.

Keywords: machine vision;multi-object tracking;feature extraction;FairMOT

0 引 言

随着人工智能领域的快速发展,在自动驾驶,智能监控等领域对于多目标跟踪(multiple object tracking,MOT)算法的需求也在不断增多。多目标跟踪旨在获取视频流的连续帧内多个目标的位置和身份信息,并给出目标在一组视

收稿日期:2024-11-06
* 基金项目:石家庄市科技合作专项重大项目(SJZZXA23005)资助

频中的连续轨迹。目前主流的多目标跟踪算法通常可根据使用的神经网络结构组合形式划分为两类，一类是以目标检测(object detection)网络为基础，目标重识别网络（re-identification，ReID)为二阶段任务的两阶段范式，该类算法的优点是精度高、组合灵活度高，方法成熟，能够有效的借助目前成熟的目标检测网络进行升级优化，常用的检测网络包括 FasterRCNN[1]、PicoDet[2] 或 YOLOv5[3] 等。该类方法的策略是首先通过目标检测网络对视频中的图像进行逐帧目标检测，并将目标的检测区域进行裁剪保存，随后使用行人重识别网络中对裁剪下来的目标区域进行二次特征提取，并对前后帧的外观特征向量进行匹配关联，最后实现多目标的连续跟踪。DeepSort[4] 是首次将目标重识别网络引入到多目标跟踪的经典算法。但通过介绍也不难看出，两阶段算法需要将数据通过两个神经网络组合才能够实现多目标跟踪，这在实时性要求较高或部署环境算力水平较低的应用场景中很难得到广泛的应用。

另一类多目标跟踪算法是由一个统一的共享网络实现特征提取，目标检测和目标重识别任务分别由多个神经网络输出头来负责产出，因为全流程只使用了一个共享的骨干网络进行特征提取，因此称为单阶段多目标跟踪算法。该策略由 Wang 等[5] 在联合检测嵌入模型(jointly learns the detector and embedding model，JDE)中首次提出，相比于两阶段策略能够省去大量的网络计算量，大幅度的提升了多目标跟踪系统的推理速度，因此在实时多目标跟踪领域，该策略成为了主流的选择方案。但由于 JDE 算法的设计局限，在环境复杂，或目标出现频繁遮挡的情况下，该方法容易出现大量目标定位不准确，进而导致目标漏检和跟踪失败问题。为了改善 JDE 算法的跟踪效果，Zhang 等[6] 提出了平衡检测任务和跟踪任务损失的 FairMOT 算法，在 Loss 计算上能均衡了检测头和重识别头的损失分配方案，有效的提升了模型的收敛精度。之后 Zhang 等[7] 又提出了可插拔的 ByteTrack 关联匹配模块，对于匹配失败的检测框进行二次处理，该模块能够有效的提高多目标跟踪系统对目标检测框的利用率，进而有效的提升了 MOT 系统的跟踪精度。为了有效的优化 FairMOT 算法重识别分支的匹配能力，Che 等[8] 提出了圆损失函数对重识别分支进行改进优化，有效的平滑了多目标跟踪过程中目标跳变次数。Hu 等[9] 提出了可变形局部注意力模块和任务感知预测模块，使得样本的中心度测量更加的准确，实现了对模糊样本的有效区分，在维持较好模型速度同时，有效的提升了模型的跟踪精度。近年来，伴随着 Trackformer[10] 结构的兴起，也有部分研究人员利用 Transformer 的 Query-Key 机制来实现对视频流中的多目标跟踪，例如 MeMOT[11]、SMILEtrack[12] 和 OffsetNet[13] 等，但是由于 Transformer 结构的引入，模型的计算量大幅增加，进而导致模型的跟踪速度明显下降，在实时多目标跟踪领域仍然处于探索阶段。

通过以上的总结不难看出，现阶段多目标跟踪算法的改进仍然是向高精度高速度方向发展，对于简洁高效的骨干网络设计仍然较少，模型的高效设计能够更好的提高算法的部署应用范围，因此本文尝试创新性的引入轻量化模型[14] 中的算法思路，来优化模型的参数量和模型精度。为了在有效提升多目标跟踪算法准确性同时，仍能维持模型较好的实时性，本文针对 FairMOT 算法进行优化，提出基于元素乘法结构的多目标跟踪算法 FairMOT-Star。在此之前，通过更换检测模型结构和引入注意力机制能够实现较好的效果提升，但是这些改进都增加了模型的计算参数量，进而影响了跟踪速度。从应用部署的角度来看，更少的参数量和更高的跟踪精度尤为重要，FairMOT-Star 则从以下两点进行优化：1)将由元素相乘结构组成的 StarBlock 首次引入多目标跟踪领域，使用 StarBlock 替换 DLA[15] 骨干网络中的 DLABlock 残差模块，实现在降低模型参数量的同时，提升模型的跟踪精度，简洁高效的完成了目标特征提取，提升模型的实际应用部署能力；2)将回归任务损失函数的 SmoothL1_Loss 改进为 EIoU_Loss[16]，同时考虑重叠损失，中心距离损失，宽高损失 4 个因素来帮助模型快速向目标收敛，提升检测网络的收敛速度和训练精度，帮助模型快速迭代的同时得到更好的跟踪效果。基于以上两点的改进，使得 FairMOT-Star 算法在有效的减少模型计算参数量的同时，提高模型的跟踪效果，对于实际的应用部署场景更加友好。

1　FairMOT-Star 算法原理及改进

FairMOT 算法的主要流程如图 1 所示，由特征提取、目标检测、嵌入特征提取和匹配关联 4 个部分组成。本文针对特征提取和目标检测部分进行优化，提出一种能够实现单镜头下对多目标进行实时跟踪且相对轻量的单阶段范式跟踪算法 FairMOT-Star。算法的数据输入为由摄像头或视频图像中采集的连续视频帧序列，将视频帧序列按照时间顺序依次输入特征提取模块，即可得到每一个视频帧中各个目标的检测信息和外观特征向量信息，随后将两个输出结果输入至关联匹配模块，使用卡尔曼滤波算法[17-18] 和匈牙利匹配算法[19] 实现对前后帧数据中相同目标匹配关联，生成目标的实时跟踪轨迹。本文通过改进 FairMOT 原有的 DLA 骨干网络和输出头的损失函数计算方式，实现适当减少模型参数量的同时，高效的提升该算法在动态视频场景中的实时多目标跟踪能力。

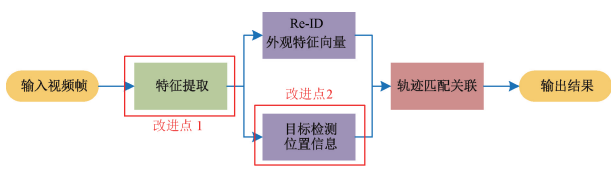

图 1　FairMOT 算法框图

Fig. 1　FairMOT algorithm block diagram

1.1 FairMOT-Star 模型结构

FairMOT-Star 模型由 3 个主要部分构成:改进的 DLA 骨干网络、DLAUP 解码网络和输出头,具体结构如图 2 所示。其中改进的 DLA 骨干网络中引入了基于元素相乘结构设计的基础模块,用以替换原始的基础残差模块。通过使用元素相乘结构旨在提升模型的特征提取能力,同时保持模型结构的简洁和高效性。

对于视频中每一帧图像,首先将图像重新调整为标准的 1 088×608 大小。调整后的图像随后被传入改进的 DLA 骨干网络进行特征提取。该过程首先经过一个 7×7 卷积核、通道数为 16 的卷积层,接着是两个 3×3 卷积核、通道数分别为 16 和 32 的卷积层。最后,特征图会依次通过 4 组由 StarBlock 组成的聚合网络树,实现对输入特征图的特征提取,并对提取到的不同尺寸特征图进行深度的特征聚合,最终得到 4 个尺寸分别为原始输入图像 1/4、1/8、1/16 和 1/32 的多尺度特征图。得到的多尺度特征图随后被输入到 DLAUP 解码器中,通过上采样和可变形卷积操作,实现骨干网络输出多尺度特征图的融合特征结果。最后,经过处理的融合特征图被送入 4 个不同的输出头,分别为中心点预测、检测框回归、中心点偏置回归和外观嵌入特征提取头。在这些输出头中,模型进行联合学习,以实现多目标跟踪中的目标检测、位置回归和特征重识别任务。通过这种多任务联合学习,FairMOT-Star 能够在保持高效率的同时,仍能提升多目标跟踪的精度。

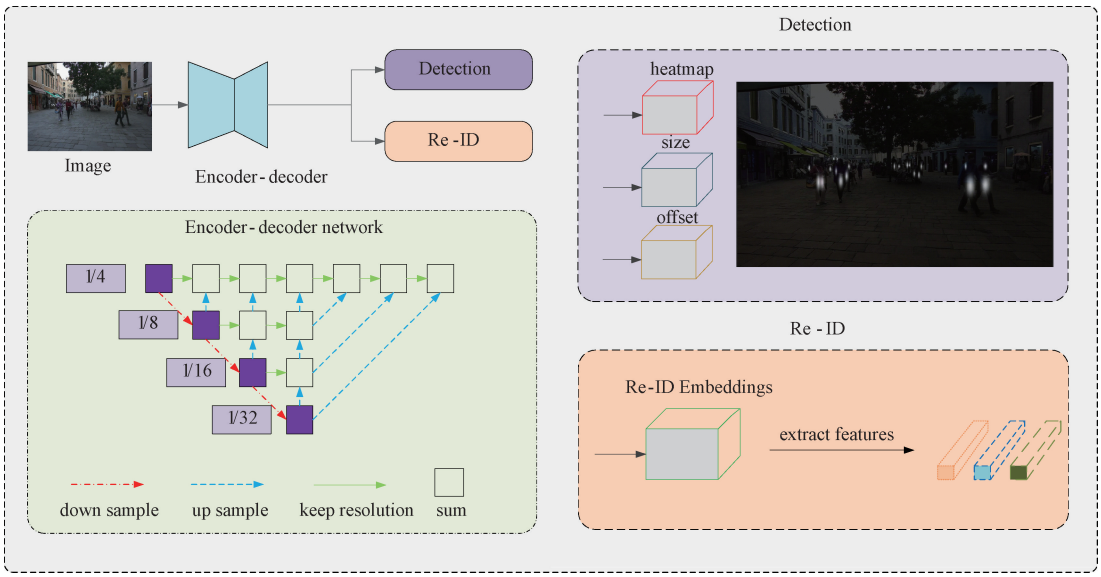

图 2 FairMOT-Star 模型结构

Fig. 2 FairMOT-Star model structure

1.2 改进的深度聚合骨干网络

由于单阶段范式的多目标跟踪算法使用公共的骨干网络进行特征提取,因此骨干网络的特征提取能力对模型最终的多目标跟踪能力和跟踪效果影响很大。FairMOT 使用的 DLA 骨干网络虽然能够通过深度聚合的方式将多个参差模块实现了跳跃链接,但是由于 DLA 默认残差模块的特征提取能力有限,骨干网络的特征提取效率较低。为了有效的提升该骨干网络的特征提取能力,以提升在检测和嵌入特征向量的效果,本文借鉴了 StarNet[20] 中提出的元素相乘结构,针对 DLA 深度聚合骨干网络进行改进,使用重新设计的 Star 模块替换原始的残差模块,如图 3 所示,Star 模块整体可以划分为 5 层,特征图首先经过一个卷积核为 7×7 的深度可分离卷积[21](depthwise deparable convolution,DWConv)在减少计算参数量的同时进行特征提取。深度可分离卷积的输出结果分别通过两个卷积核为 1×1 的卷积以便于增强特征的表达能力,然后对其中的一个输出结果进行 ReLU6 激活函数进行激活输出,该激活函数能够更好的支持模型后续在移动设备上部署。最后,将经过激活输出的特征图与另一个 1×1 卷积的输出结果进行逐元素的相乘。根据 StartNet 中的经验可以得知,该操作具有将输入映射到极高维、非线性特征空间的能力,这种映射方式与传统增加网络宽度的方法不同,而是通过跨通道特征对乘实现了一种类似于多项式核函数的非线性高维映射,这种高效的特征融合方式使得元素相乘结构能够极大的提高模型的表示能力和计算性能。最后将逐元素相乘的结果通过另一个 7×7 大核的深度可分离卷积进行特征提取,得到了原始输入维度相同的残差特征图,最后将此残差特征图与输入特征图进行求和后,经过一个普通的卷积核为 3×3 卷积进行一次卷积计算,得到一个通道数为原始输入通道 2 倍的特征图。该模块由于使用了深度可分离卷积、ReLU6 激活函数和元素相乘结构的特点,实现了对输入特征的高效特征提取。最后经过深度聚合网络 DLANet 的组网方式对该模块进行深度聚合组网,得到一个改进的 DLA 骨干网络,该网络具有更

加轻量且高效的特点。能够为后续多目标跟踪的目标检测和 ReID 联合学习过程在使用更少的模型参数量条件下获得更高的模型精度。

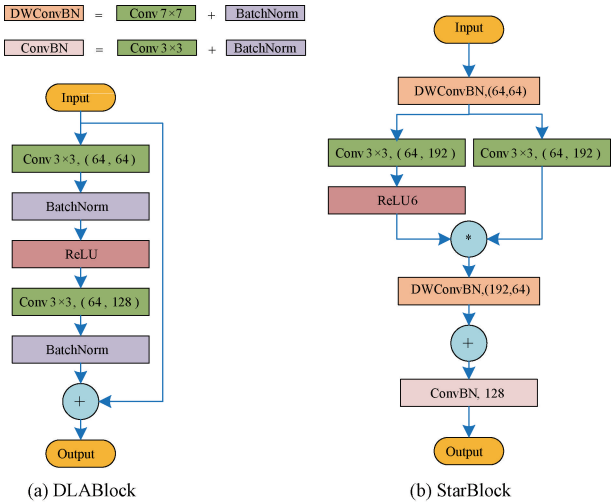

图 3　改进的 StarBlock 结构图

Fig. 3　Improved StarBlock structure diagram

1.3　改进的检测框回归损失策略

根据图 2 中的模型结构，输入图片经过编解码网络后得到了一个多尺度融合的特征图输出。随后该融合特征图会分别送入两个任务分支输出中，检测任务分支包括中心点、检测框回归和中心点偏置回归 3 个输出头，重识别分支则为嵌入式特征提取输出头。检测分支中的中心点输出头输出的是在目标 1/4 原始尺寸的下的热力图，在反向传播计算 Loss 时采用交叉熵损失计算该热力图与目标真实标签的损失；检测框输出头的输出是由多个 4 点坐标组成的检测框坐标，在 FairMOT 算法中，该输出头反向传播 Loss 采用 SmoothL1_Loss 计算；中心点偏置输出头输出的是由多个 2 点偏置组成的二维列表，其中的每一组偏置分别代表其对应的中心点在还原至原始尺寸时横纵方向上的偏移量，使用该方法能够更加准确的预测目标的中心点坐标，该输出头在反向传播损失计算时仍采用 SmoothL1_Loss 计算。本文针对 FairMOT 检测分支中的检测框损失函数进行改进，由于检测框的 4 个点并非相互独立的，使用 SmoothL1_Loss 并未考虑到检测框中的 4 个点在几何意义上的关联性，因此使用该 4 点组成的检测框和目标真实框的交并比(intersection over union，IoU)来衡量检测框损失更能符合实际应用场景中的需求，因为对于不同的检测框可能具有相同的 SmoothL1_Loss，但他们的 IoU 差异却非常大。Unitbox[22] 中首次提出了 IoU_loss，同时将 4 个点构成的目标边界看做一个整体进行回归，该 Loss 首先通过式(1)计算出检测框和真实框的交并比，随后计算采用式(2)作为该回归任务的损失函数，进而有效的提高了检测框的回归精度。

$$R_{IoU}=\frac{B\cap B_{gt}}{B\cup B_{gt}} \tag{1}$$

$$L_{IoU}=1-R_{IoU} \tag{2}$$

但 IoU_Loss 存在着存在着以下几个问题：1)在检测框和真实框无交集时，IoU 值始终为 0，此时无法反馈出两个框的位置关系，在网络进行反向传播时梯度不可导，模型无法对此情况进行有效收敛，因此对于检测框与真实框无交集的检测框，模型无法通过优化使检测框朝向真实框收敛。2)当检测框完全包含在真实框内时，此时无论检测框在真实框内部的哪个位置，两者的 IoU 值均一致，此时位置较为居中的检测框无法获得更高的 IoU 值，不利于模型朝向更好的方向收敛。3)只考虑了检测框的面积关系，未考虑检测框和真实框的形状，在交并比相同的情况下，与真实框具有相同长宽比的检测框无法得到更高的 IoU 值，同样不利于模型朝向更好的方向收敛。

不同 IoU_Loss 对比如图 4 所示。

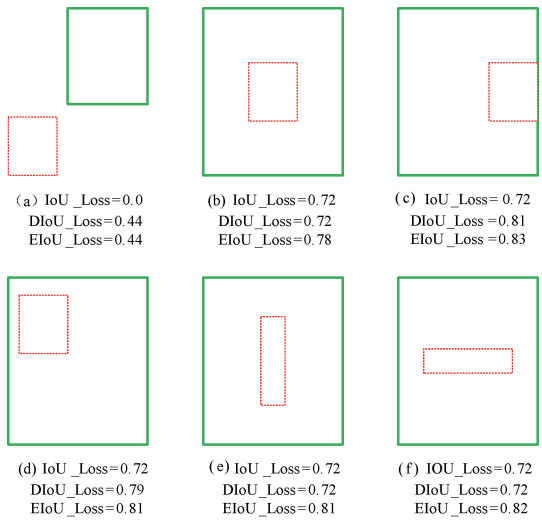

图 4　不同 IoU_Loss 对比图

Fig. 4　Different IoU_Loss comparison diagram

为了提高 IoU_Loss 的收敛效率，研究人员相继提出了考虑检测框和真实框最小包闭区域的 GIoU_Loss[23]，通过引入最小包闭框的概念，有效的解决了检测框和真实框无交集使 IoU=0 的问题；考虑了检测框和真实框之间距离的 DIoU_Loss[24]，以解决检测框完全在真实框内部，检测框位置居中时无法获得更高 IoU 值的问题；考虑了检测框长宽比例的 CIou_Loss，以解决检测框和真实框横纵比不一致问题。本文采用更为全面的 EIoU_Loss，其计算公式如式(3)所示，原理是在 CIoU 的基础上，使用宽度和高度的差异值，代替 CIoU 中的宽高比例差异，能够更好的加快模型的收敛速度。

$$L_{EIoU}=L_{IoU}+L_{dis}+L_{asp}=1-R_{IoU}+\frac{\rho^2(b,b_{gt})}{c^2}+\frac{\rho^2(w,w_{gt})}{c_w^{\ 2}}+\frac{\rho^2(h,h_{gt})}{c_h^{\ 2}} \tag{3}$$

其中,ρ 代表两点之间的欧式距离,c 代表的是包围检测框和真实框最小矩形的对角线长度,c_w 是包围框的宽度,c_h 是包围框的高度。使用 EIoU 作为检测框收敛的损失函数,能够更加全面的考虑到检测框和真实框之间的几何关系,符合实际的应用场景,实现有效的加快模型的收敛速度。

1.4 匹配关联

对于算法的匹配关联部分,仍保持原始算法的匹配策略,通过检测分支和 ReID 分支输出的目标检测信息和外观嵌入特征进行关联匹配,首先对第 1 帧中的检测结果进行初始化,由于此时轨迹池为空,因此会直接保存其结果在轨迹池中,并标记该轨迹为待确认跟踪状态,从第 2 帧开始获取的目标检测信息和嵌入特征向量,首先使用检测信息和嵌入特征信息与现有轨迹的卡尔曼滤波算法预测结果进行关联匹配,该过程会使用余弦距离计算当前帧中目标与预测轨迹之间的距离信息,形成代价矩阵,随后使用匈牙利算法通过代价矩阵中的排序依次为当前帧中的目标分配轨迹。经过匈牙利算法分配后,对于一次匹配成功的,可以直接加入跟踪轨迹池更新轨迹信息,对于一次匹配后轨迹池中仍未分配到检测结果的跟踪轨迹,此时采用 IoU 对剩余的检测信息和剩余跟踪轨迹进行二次分配,对于二次分配仍未成功匹配的轨迹,则将该轨迹标记为失配状态,对于连续失配 30 帧以上的轨迹,将会从轨迹池中删除,此时认为该轨迹目标已从画面中消失。对于两次分配后仍然未匹配成功的检测目标,视为新目标,加入轨迹池并标记问为待确认跟踪状态,当新目标连续跟踪指定阈值后,则将该轨迹状态从待确认跟踪状态更新到已跟踪状态,形成新的跟踪轨迹。对于轨迹的嵌入向量信息采用式(4)进行更新:

$$f_t = \gamma f_{t-1} + (1-\gamma) f \tag{4}$$

其中,$\tilde{f}$ 表示目标中一个目标的嵌入向量,f_t 表示 t 时刻的嵌入向量表示,γ 是平滑动量项。当完成匹配之后,最后会输出轨迹池中非失配状态的轨迹作为跟踪结果,每帧图像中的不同目标会分配一个唯一的 ID 作为身份标识。

如图 5 所示,为算法整体流程图,输入视频数据即可得到目标的跟踪轨迹信息。

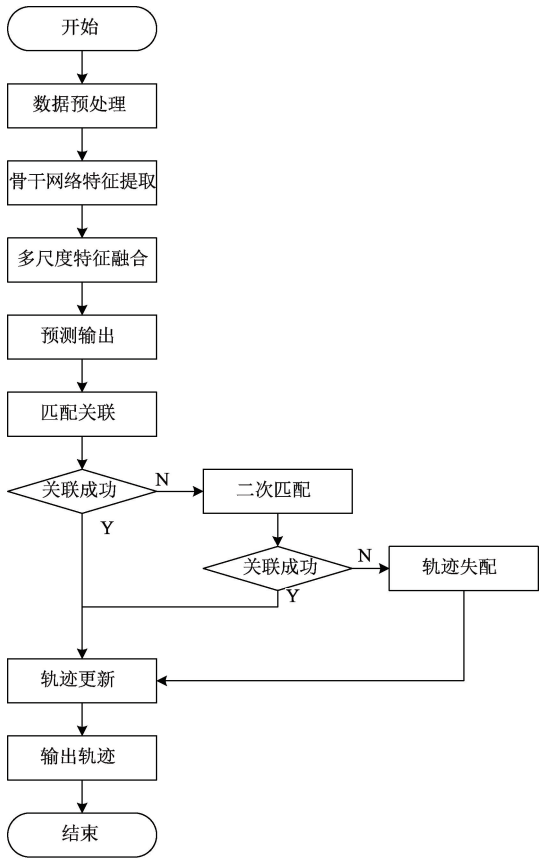

图 5 算法整体流程图

Fig. 5 Overall algorithm flow diagram

2 实验及实验数据分析

2.1 数据集与评价指标

本文使用 MOT17、Caltech、CUHKSYSU、PRW、Cityscapes 和 ETHZ 六个数据集组成的 MIX 数据集作为训练集,使用单一小数据集的优点是能够针对该特定场景进行模型优化,但是会大幅度的限制模型的通用场景能力,因此本文使用 MIX 的多场景混合数据集作为训练集,使用公开的 MOT16 训练集作为本文的测试集。本文使用的训练集图片数量为 53 694 张,已标注目标轨迹 14 455 个;验证集图片数量 5 316 张,已标注目标轨迹 517 个。

本文使用与人类感知比较一致的 MOTA 作为评测指标来评估整个 MOT 系统的运行效果,其中的关键性能指标如下:

1) MT(mostly tracked):预测轨迹与 GT(ground truth)轨迹 80%以上的时间内都匹配成功的轨迹数量,此处不考虑轨迹 ID 发生转变的情况,只考虑与 GT 匹配成功的部分。

2)ML(mostly lost):预测轨迹与 GT 轨迹 20%以上的时间都未匹配成功的轨迹数量,考虑范围与 MT 相同。

3)FP(false positive):模型将负样本预测为正样本的数量,在多目标跟踪系统中 FP 代表每一帧中 FP 的总和。

4)FN(false negative):模型将正样本预测为负样本的数量,在多目标跟踪系统中 FN 代表每一帧 FN 的总和。

5)IDF1:被检测和跟踪目标中获取正确 ID 的目标占总目标的比例,用于考察跟踪连续性和 ReID 的准确性。

6)IDs(id switch):对于相同目标发生 ID 跳变的次数,理想的情况下对于相同的目标应该分配固定的身份 ID,当跟踪算法能力不足无法满足实际跟踪需求时会导致对同一目标的跟踪轨迹碎片化,进而导致跟踪目标的 ID 发生跳变。IDs 是衡量跟踪算法跟踪效果的重要性能指标之一。

7)MOTA:多目标跟踪算法的整体评价指标,有 MOT 系统中各项重要指标计算得出的综合性能指标,具体计算公式如下:

$$MOTA = 1 - \frac{FN + FP + IDs}{GT} \tag{5}$$

2.2　实验环境

本实验的硬件条件为 Intel(R) Xeon(R) Gold 6271C CPU,629 G 内存,2.60 GHz 主频,8 张 NVIDIA Tesla V100-SXM2 显卡,单卡显存 32 G。软件环境使用 registry. baidubce. com/paddlepaddle/paddle: 3.0.0b1-gpu-cuda11.8-cudnn8.9-trt8.5 虚拟 Docker 环境,在 Ubuntu 20.04.6 LTS 操作系统中,以 PaddlePaddle2.6.1 作为开发框架,以 PaddleDetection 作为开发套件,Python3.10 环境中运行。

2.3　骨干网络预训练

为了初步验证 Star 模块的特征提取有效性,本文首先选取了另一个流行的轻量化骨干网络 FasterNet[25] 的 FasterBlock 和 DLA 骨干网络默认的基础残差 DLABlock 与 StarBlock 分别作用于 DLA 骨干网络的聚合树中,在 ImageNet 1k 数据集上做了简单的对比实验,综合参考了原始算法的超参数信息,并适当进行学习率搜索实验,最后在批大小为 64,初始学习率为 0.1,在第 3、6、9 epoch 中下降 10%学习率策略下,分别训练了 10 个 epoch,对比 3 个模块得到的训练结果如表 1 所示。

表 1　基础模块对比

Table 1　Comparison of basic modules

基础模块	Acc/%	Params/M
StarBlock	55.27	13.65
FasterBlock	47.76	11.25
DLABlock	52.70	15.75

通过对比实验可以看出 StarBlock 不同于其他模块的特点是其应用于深度聚合网络结构下能够实现提升骨干网络特征提取精度同时,仍能节省一部分网络的计算参数量。证明了其简洁高效的特征提取能力。因此本文基于上述实验继续在 ImageNet 1k 数据集上做了 120 epoch 的全量实验,修改学习率的下降策略为第 30、60 和 90 epoch 时下降 10%,最后得到的分类预训练模型精度为 75.2%,相比于 DLA34 的原始精度 74.62%提升了 0.79%,有效的证明了元素相乘模块在骨干网络中的有效性。因此本文将此网络替换至 FairMOT 中,使用 ImageNet 1k 的全量实验训练权重作为 FairMOT-Star 算法的预训练权重开展后续实验。

2.4　实验结果分析

对于 FairMOT-Star 的微调训练策略上,首先使用图像缩放、随机仿射变换和反转等数据增强操作,以减少模型在训练过程中的过拟合风险,随后将图片统一调整到指定分辨率,默认调整分辨率为 1 088×608。采用 Piecewise Decay 分段衰减优化器,以 2×10^{-3} 为初始学习率,并在 20 个 epoch 之后下降 10%策略训练至 30 epoch。每卡 batch size 设置为 6,全局 batch size 为 48。

图 6 所示为 FairMOT-Star 以上述训练策略训练过程中的主要损失函数曲线,横轴为模型的迭代 iter 数,纵轴为不同损失函数的损失值。通过观察图中数据可以看出,伴随着模型迭代次数的增加,模型的整体混合损失、中心点损失、检测分支损失和外观嵌入向量损失均在不断下降,并在模型训练后期趋于平稳。说明模型在不断的训练中得到了有效的收敛。

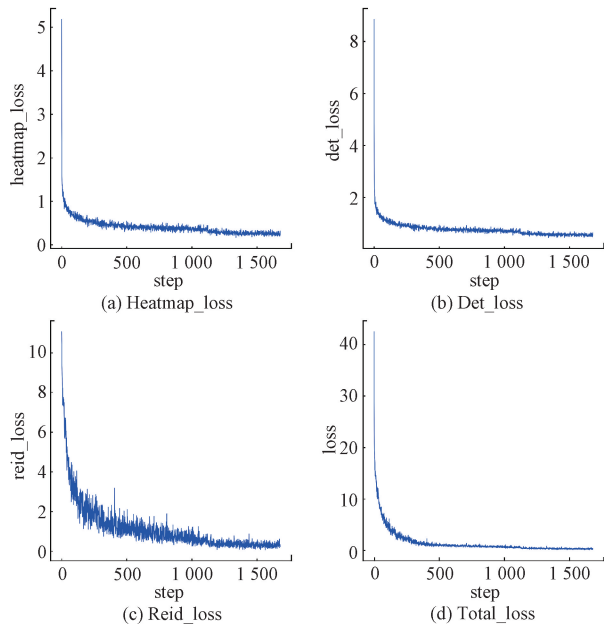

图 6　训练 Loss 收敛曲线

Fig. 6　Training Loss convergence curve

表 2 所示为改进后的算法与 DeepSort、JDE 等常见的主流多目标跟踪模型在相同设备环境上训练后使用 MOT16 测试集评估的对比结果,其中两阶段模型使用 YOLOv3 作为目标检测模型,因此算法参数量以 YOLOv3 参数量作为参考。通过实验结果可以看出,本文提出的 FairMOT-Star 算法在 MOTA 指标上达到了 86.0%,同时模型的参数量只有 19.59 M,达到了更加高效跟踪的效果。与原 FairMOT 算法相比,模型的参数量下降了 2.11 M,同时 MOTA 精度提升了 3.5%。与两阶段范式的模型相比,由于 FairMOT-Star 使用的是更加轻量高效的骨干网络,因此在模型的参数量大幅度减少的同时,取得了更高的跟踪精度。与现有的单阶段跟踪算法相比,无论是 MOTA 指标还是 IDF1 指标,FairMOT-Star 算法都达到了领先效果,实现了高效稳定的多目标跟踪性能。

表 2 不同算法在 MOT16 测试集上的结果对比

Table 2 Comparison of results of different algorithms on the MOT16 test set

算法	MOTA/%	IDF1/%	MT	ML	IDs	Params/M
DeepSort	72.2	59.5	322	30	1 087	61.53
BoT-SORT[26]	71.9	65.5	307	29	1 353	61.53
OC-Sort[27]	70.6	65.5	221	43	876	61.53
JDE	73.1	68.9	324	28	1 312	73.08
FairMOT	82.5	81.8	317	28	501	21.70
FairMOT-Star	**86.0**	**84.4**	396	13	**496**	**19.59**

图 7 所示为改进前后对目标轨迹发生重叠时效果对比。在 a 时刻均正确跟踪目标,改进前的模型在 b 时刻就已经发生了轨迹断裂,进而导致在 c 时刻目标重叠时导致目标的 ID 发生了跳变。而改进后的模型在 c 时刻虽然发生了目标重叠而导致的轨迹丢失,在 d 时刻发生的轨迹跳变,当 e 时刻两个目标再次分开后,改进后的模型仍然维持着两个目标原始 ID,并未发生 ID 跳变现象。

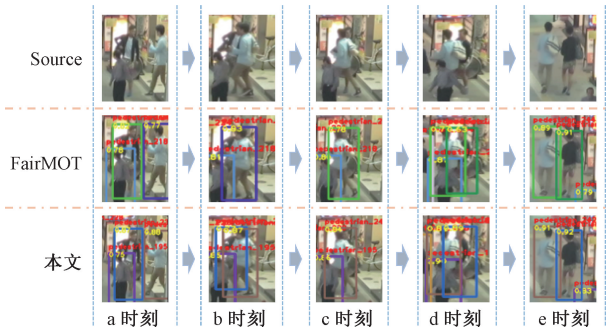

图 7 跟踪效果对比图

Fig. 7 Tracking effect comparison

如图 8 所示为改进前后对小目标和遮挡目标跟踪效果的对比,图 8(a)和(b)对照组展示的是改进后的模型凭借更有效的特征提取能力,能够更加精确的发现图像中的小目标,实现对小目标进行跟踪,有效的避免了目标漏检现象。图 8(c)和(d)对照组展示的是改进后的模型能够有效的解决目标被严重遮挡而导致的轨迹断裂问题。通过两幅图片的对比可以看出,改进后的模型拥有更好的鲁棒性和,更加细粒度的检测能力,能够实现更加稳定和精确的多目标跟踪。

为了验证该模型在实际应用场景的跟踪效果,本文在实际的校园场景中进行了一个跟踪测试,具体跟踪效果如图 9 所示,根据图中信息可以看出,该模型在实际场景中跟踪效果良好,能够很好的检测视频中的目标,同时应对目标的遮挡情况不会发生 ID 跳变问题,是一个稳定性和鲁棒性较好的跟踪算法,具有实际应用意义。

为了验证 StarBlock 和 EIoU_Loss 两处改进相对于原

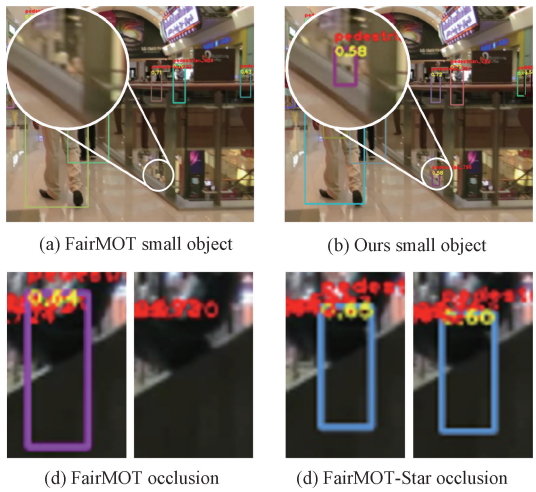

图 8 小目标和遮挡情况对比图

Fig. 8 Comparison of small targets and occlusion situations

图 9 实际场景跟踪效果图

Fig. 9 Actual scene tracking renderings

算法的有效性,本文在使用的 MOT16 测试集上进行了相关消融实验,具体结果如表 3 所示。

表 3 消融实验

Table 3 Ablation experiment

StarBlock	EIoU_Loss	DIoU_Loss	MOTA/%	IDF1/%	IDs
			82.5	81.8	501
√			**86.7**	**84.6**	507
	√		82.6	81.3	502
		√	81.1	80.8	520
√		√	85.9	84.2	501
√	√		86.0	84.4	**496**

更换 FairMOT 中聚合树残差模块为 StartBlock,MOTA 指标提升了 4.2%,但同时 IDs 指标上涨了 6 次;将损失函数更换为 EIoU_Loss 后,MOTA 指标提升了 0.1%,IDs 指标上涨了 1 次,将损失函数更换为 DIoU_Loss,MOTA 指标下降了 1.4%,同时 IDs 指标上涨了 19 次,且 DIoU_Loss 和 StartBlock 组合后对 MOTA 指标提升有限的同时,在 IDs 指标并无优化。消融实验结果表明,

StartBlock 和 EIoU_Loss 两者均能够带来 MOTA 指标的提升，但各自独立使用时会导致目标 ID 跳变次数的增加，对两处同时优化后的目标 ID 跳变次数得到了一定的减少，同时相比于原始模型仍然有 3.5%的精度提升，因此得到一个跟踪精度和稳定性均更好的多目标跟踪算法。

3　结　　论

本文提出了基于元素相乘结构的多目标跟踪算法 FairMOT-Star，在 FairMOT 算法上做出了以下两点优化：首先使用由元素相乘结构重新设计的 StarBlock 作为深度聚合网络 DLA 聚合树中的基本模块，实现了对输入目标和嵌入特征更加精准的提取，节省了 9.7%模型参数量的同时，提升了模型跟踪精度。然后在检测分支的检测框回归任务中，将 Smooth L1_Loss 改进为更能提醒模型损失效果的 EIoU_Loss，使得模型训练过程中的能够有效的提升目标的检测框精度，同时凭借 EioU_Loss 的精准描述特性，也加快了模型训练时的收敛速度。通过实验结果表明，该算法在节约模型参数量的同时，算法的多目标跟踪精度得到了有效提升，相比于原始的 FairMOT 算法及相关多目标跟踪算法在参数节约和精度提升的平衡上做的更好。但由于为对模型匹配关联部分进行更多的优化，导致算法在复杂场景的使用中，仍然具有 ID 切换频繁，IDF1 指标不高等问题，未来计划从匹配关联的角度，进一步优化算法的匹配关联能力，提高算法跟踪稳定性。

参考文献

[1] REN SH 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, 2016, 39(6): 1137-1149.

[2] YU G H, CHANG Q Y, LYU W Y, et al. PP-PicoDet: A better real-time object detector on mobile devices[J]. ArXiv preprint arXiv:2111.00902, 2021.

[3] 强栋,王占刚.基于改进 YOLOv5 的复杂场景多目标检测[J].电子测量技术, 2022, 45(23): 82-90.
QIANG D, WANG ZH G. Improved YOLOv5 complex scene multi-target detection[J]. Electronic Measurement Technology, 2022, 45(23): 82-90.

[4] 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). IEEE, 2017: 3645-3649.

[5] WANG ZH D, ZHENG L, LIU Y X, et al. Towards real-time multi-object tracking [C]. European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 107-122.

[6] 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(1): 3069-3087.

[7] ZHANG Y F, SUN P Z, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box[C]. European conference on Computer Vision, Cham: Springer Nature Switzerland, 2022: 1-21.

[8] CHE J, HE Y, WU J M. Pedestrian multiple-object tracking based on FairMOT and circle loss [J]. Scientific Reports, 2023, 13(1): 4525.

[9] HU W M, WANG S, ZHOU Z, et al. One-stage anchor-free online multiple target tracking with deformable local attention and task-aware prediction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 11446-11463.

[10] MEINHARDT T, KIRILLOV A, LEAL-TAIXE L, et al. Trackformer: Multi-object tracking with transformers [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8844-8854.

[11] CAI J R, XU M Z, LI W, et al. Memot: Multi-object tracking with memory[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8090-8100.

[12] WANG Y H, HSIEH J W, CHEN P Y, et al. Smiletrack: Similarity learning for occlusion-aware multiple object tracking[C]. AAAI Conference on Artificial Intelligence, 2024, 38(6): 5740-5748.

[13] ZHANG W, LI J M, XIA M, et al. OffsetNet: Towards efficient multiple object tracking, detection, and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(2): 949-960.

[14] 冉险生，贺帅，苏山杰，等. 融合特征增强与 DeepSort 的疲劳驾驶检测跟踪算法[J]. 国外电子测量技术, 2023, 42(8):54-62.
RAN X SH, HE SH, SU SH J, et al. Fatigue driving detection tracking algorithm incorporating feature enhancement and DeepSort [J]. Foreign Electronic Measurement Technology, 2023,42(8):54-62.

[15] SU Y F, LIU W Q, YUAN Z M, et al. DLA-Net: Learning dual local attention features for semantic segmentation of large-scale building facade point clouds[J]. Pattern Recognition, 2022, 123(1): 108372.

[16] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506(1):

146-157.

[17] 白晓娟,道伟,关露,等. 室内复杂环境处理及泰勒公式改进定位算法研究[J]. 电子测量技术,2021,44(3):55-59.
BAI X J, DAO W, GUAN L, et al. Research on indoor complex environment processing and improved positioning algorithm of Taylor formula[J]. Electronic Measurement Technology, 2021, 44(3): 55-59.

[18] 杨艳华,吕童,柴利. 基于ESKF-MPC的四旋翼无人机轨迹跟踪控制[J]. 电子测量与仪器学报,2022,36(7):24-32.
YANG Y H, LYU T, CHAI L. Path tracking control for a quadrotor UAV based on ESKF-MPC[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(7):24-32.

[19] 杨闻宇. 基于深度学习的多目标跟踪算法研究[D]. 绵阳:西南科技大学,2023.
YANG W Y. Research on multi-object tracking algorithm based ondeep learning [D]. Mianyang: Southwest University of Science, 2023.

[20] MA X, DAI X Y, BAI Y, et al. Rewrite the stars[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694-5703.

[21] SINHA D, EL-SHARKAWY M. Thin mobilenet: An enhanced mobilenet architecture[C]. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, 2019: 0280-0285.

[22] YU J H, JIANG Y, WANG Z Y, et al. Unitbox: An advanced object detection network[C]. 24th ACM International Conference on Multimedia, 2016: 516-520.

[23] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

[24] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[25] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.

[26] AHARON N, ORFAIG R, BOBROVSKY B Z. BoT-SORT: Robust associations multi-pedestrian tracking[J]. ArXiv preprint arXiv:2206.14651, 2022.

[27] CAO J K, PANG J M, WENG X S, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9686-9696.

作者简介

周畅达,硕士研究生,主要研究方向为计算机视觉、视频多目标跟踪。

E-mail:changda1650@163.com

杨帆(通信作者),博士研究生,教授,博士生导师,主要研究方向为电子电路与计算机视觉。

E-mail:yangfan@hebut.edu.cn