

融合多阶语义增强的 JDE 多目标跟踪算法

王俊¹, 王鹏², 李晓艳¹, 王梁³, 孙梦宇¹, 郜辉¹

(1.西安工业大学 电子信息工程学院, 陕西 西安 710021; 2.西安工业大学 发展规划处, 陕西 西安 710021; 3.陕西航天技术应用研究院有限公司, 陕西 西安 710100)

摘要:为了解决联合检测和嵌入(JDE)算法中目标遮挡以及ID信息与位置信息提取不足造成的目标ID切换问题,提出了融合多阶语义增强的JDE多目标跟踪方法。采用SPA特征空间金字塔注意力模块扩大感受野,获得更丰富的语义信息,提高模型对不同尺度目标的检测精度;通过FCN网络使检测头和ID Embedding任务协同学习以缓解两者的过度竞争并增强原始语义信息,有效减少ID切换次数;利用PCCs-Ma运动度量加强卡尔曼滤波的预测和观察之间的联系,提高运动特征相似度判别的可靠性。为了验证算法的有效性,设计了相同实验环境下JDE算法和所提算法的对比实验。实验结果表明,所提算法模型检测平均精度提高了3.94%。在MOT16数据集上,MOTA和IDF1指标均提高了6.9%,改进后的算法ID切换次数明显减少,取得了良好的跟踪效果。

关键词:多目标跟踪;JDE算法;语义信息;SPA;感受野

中图分类号:TP391.9

文献标志码:A

文章编号:1000-2758(2022)04-0944-09

多目标跟踪旨在预测图像序列中多个目标的位置,识别该图像序列中哪些运动物体是同一目标,将其一一匹配并给出各自相应的运动轨迹。在环境感知中多目标跟踪任务是CV(computer vision)中的一项重要研究技术,该任务在智能监控、无人驾驶、无人机巡检等多种军用和民用场景中应用广泛。

近年来,one-shot方法因其速度和准确性均衡而备受关注。2017年Xiao等^[1]首先提出在同一卷积神经网络中处理行人检测和Re-ID任务的端到端框架。Wang等^[2]在one-stage检测器中嵌入表观模型,跟踪准确度达到了62.1%,实现了端对端的联合检测和嵌入(JDE)框架。2020年Zhang等^[3]基于JDE提出FairMOT,采用无锚框的DLA网络进行多层语义特征提取融合,使Re-ID信息同时包含网络中高维和低维语义信息,忽略了2个任务语义信息的差异。为了应对遮挡问题,2021年,Chaabane等^[4]提出DEFT方法,将提取的外观信息用于关联匹配网络,使目标遮挡时具有较强的鲁棒性。同年

Guo等^[5]提出的TADAM网络采用时间感知和干扰注意力,实现多阶语义的融合,并通过记忆聚合模型来增强REID语义信息,使位置预测和嵌入关联之间协同,却忽视了其主干网络提取信息不足的问题。

在上述介绍的以JDE为主流目标跟踪框架的算法中,由于在跟踪过程中检测质量会影响跟踪性能,其中JDE的检测器对深层特征提取不充分,且忽略目标定位信息和ID信息共享嵌入学习的内在差异性,使得在跟踪过程应对不同尺度目标以及遮挡情况的效果不佳,对目标堆叠情况下的目标判别能力不强。由于注意力机制可对目标形成更好的关注以及获得更鲁棒的语义信息,对网络语义信息有极大影响,因此本文借鉴该思想并针对以上不足在原有的Darknet-53特征提取网络^[6]末端加入空间金字塔注意力模块,扩大感受野并弥补CNN对不同尺度目标表征能力不足的问题;在YOLO检测头的分类回归分支和Re-ID特征学习分支应用不同权重的特征相关网络任务,弥补分支任务学习不均衡的矛

收稿日期:2021-11-03

基金项目:国家自然科学基金(62171360)、陕西省科技厅重点研发计划(2022GY-110)与西安工业大学校长基金面上培育项目(XGPY200217)资助

作者简介:王俊(1997—),西安工业大学硕士研究生,主要从事视觉目标跟踪研究。

通信作者:王鹏(1978—),西安工业大学教授,主要从事机器视觉、模式识别及图像处理研究。e-mail: wang_peng@xatu.edu.cn

盾,以应对跟踪过程存在遮挡的问题;在数据关联中将 PCCs 应用到原有度量运动特征相似度公式,使得目标的跟踪轨迹更加具有判别能力。本文针对 JDE 算法的不足提出了融合多阶语义增强的 JDE 多目标跟踪算法,增强了算法的准确度,进一步提升多目标跟踪性能。

1 JDE 算法

如图 1 所示,本文使用基准 JDE 架构实现特征提取和检测分支共享特征的同步学习。JDE 架构检测部分以 YOLOv3^[6] 检测算法为基础,将表观模型嵌入检测网络中,共享主干特征提取网络权值。采用 512 个 3×3 卷积学习外观特征,以便模型可以同时输出回归信息、分类及对应的表观特征。基于级联匹配的方式进行跟踪,以卡尔曼滤波^[7] 轨迹预测、运动和外观特征相似度计算以及匈牙利算法^[8] 匹配为主完成跟踪任务。JDE 中 YOLOv3^[6] 的多分支学习方法提高了跟踪效率,但存在检测和 Re-ID^[9] 特征学习不公平造成 ID 切换频繁且准确度降低的问题。本文针对此算法的不足提出了改进的目标跟踪算法,引入了注意力模块、特征相关网络以及 PCCs-Ma 相关度量公式,提高了算法的准确度,有效减少 ID switch 现象。

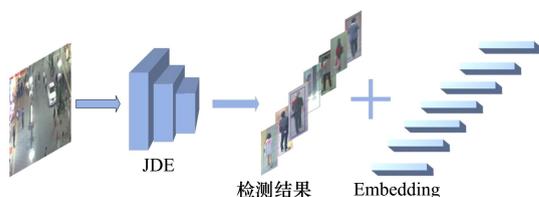


图1 联合检测和嵌入模型

2 JDE 改进算法

本文采用 SPA、FCN 和 PCCs 方法对 JDE 网络改进,故将基于 JDE 网络的改进模型称为 SFP-JDE。

如图 2 所示,第一部分是检测与特征提取,从左至右分别是 Darknet-53 和 SPA^[10] (spatial pyramid attention) 构成的主干特征提取网络、FPN^[11] 和 2 种特征相关网络组成的 Neck 模块、检测器的 Re-ID^[9] 头和 YOLO 头。具体改进为:首先改进主干特征提取网络,将 SPA 模块嵌入 Darknet-53 主干特征提取网络末端。对不同尺度特征融合、重组,提取有效的

多尺度特征,增强对不同大小目标的检测能力;其次,考虑到检测需要嵌入相同类别中具有相似语义的不同对象,Re-ID 倾向于为 2 个对象学习区分语义。为了解决两者任务分支存在的差异冲突,本文在分类回归前和 Re-ID 外观特征提取之前嵌入 FCN 网络,促使各个分支的表观学习,充分实现了模块间的特征信息共享;最后,在线关联的运动亲和力计算中,引入 PCCs 相关系数将运动相关度量改进为 PCCs-Ma 公式,自适应不同轨迹卡尔曼滤波的观察值和预测值的关联程度,提高跟踪运动轨迹的判别能力。

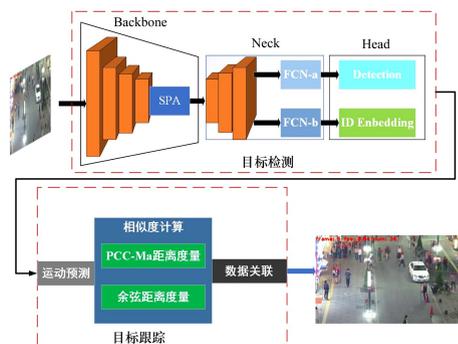


图2 本文算法框架示意图

2.1 SPA 改进主干特征提取网络

注意力模块能够使模型更加关注显著信息,本文采用的 SPA^[10] 模块引入空间金字塔结构等编码和解码操作。考虑全局平均池化会使浅层特征无法充分利用注意力机制,在通道方向引入结构信息,使其同时考虑网络结构正则化和结构信息。

在增强语义信息方面,本文将 SPA^[10] 注意力模块与 SPPNet^[12] 相比。SPPNet 为了得到固定长度的特征向量,通过不同大小的卷积得到全局和局部的语义特征,并融合信息。SPA 则使用更多的结构信息编码特征图,并且在不引入多余参数的情况下,能保留每个通道中的空间语义信息,两者均有扩大感受野的作用。为了证明 SPA 注意模块扩大感受野以增强语义信息的有效性,在 3.3 节进行了实验验证。

由实验结果得,注意力模块放在主干特征提取网络深层时效果最好,且与 SPPNet 实验对比,SPA 既能表示原有特征丰富语义信息,又能扩大感受野,使主干网络继承全局平均聚集的优点,增强 CNN 的表征能力。为了在复杂环境下提高检测、跟踪性能,利用空间金字塔注意力模块,对输入层特征进行多

尺度特征融合、重组,提升主干特征提取网络信息的鲁棒性,故本文在主干特征提取网络末端加入 SPA 模块增强对不同尺度目标检测,使用该模块提取有效特征并提高效率。

设主干特征提取网络 Darknet-53 由 L 层组成,每层输出一个特征图。其中 $l \in [1, L]$ 是层数序列。本文将 SPA 布置在 Darknet-53 的最后一层 ($l = L$), x_l 表示第 l 层的输出。整个模块的具体框架实现如图 3 所示,具体步骤如下:

步骤 1 将 $x_l \in \mathbf{R}^{C \times W \times H}$ 输入 SPA 模块学习注意力权重,并多尺度学习 x_l 中的每个通道的语义信息。空间金字塔结构 $S(x_l)$ 的输出可以表示为

$$S(x_l) = C(R(P(x_l, 4)), R(P(x_l, 2)), R(P(x_l, 1))) \quad (1)$$

式中: $C(\cdot)$ 表示串联运算; $R(\cdot)$ 是指将张量重新调整为向量; $P(\cdot, \cdot)$ 表示自适应平均池化层。

步骤 2 设 $S(x_l) = v, v$ 是 3 个汇集层的输出但非线性表达影响了注意机制的有效性,故采用 2 个全连接的多层感知机层对 v 进行编码,并生成一维注意力特征图。具体见(2)式

$$\tilde{v} = \text{sig}(D_2 p D_1 v) \quad (2)$$

式中: p 为 ReLU 激活函数; D_1 和 D_2 分别表示 2 个全连接层; sig 表示 sigmoid 函数。当忽略 BN 和激活层时,将(1)式代入(2)式中得到 SPA 模块 ξ 为

$$\xi(x_l) = \sigma(F_{fc}(F_{fc}(S(x_l)))) \quad (3)$$

式中: $F_{fc}(\cdot)$ 表示全连接层; $\sigma(\cdot)$ 是 sigmoid 激活函数。

步骤 3 将特征图 x_l 反馈给注意力权重可得 SPA 输出的一维注意力图,由(3)式可得

$$x_l = \xi(x_l) \otimes x_l \quad (4)$$

式中, \otimes 为元素乘法。

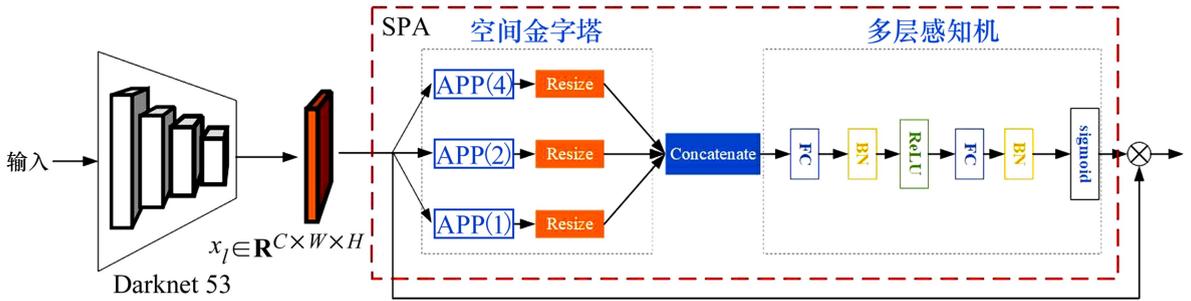


图 3 空间金字塔注意力(SPA)模型

2.2 FCN 改进学习任务

JDE 的检测和外观特征提取 2 项任务存在内在区别,从而导致学习模糊化,造成整体性能降低。为了缓解两者内在矛盾,将 YOLOv3 输出的特征表示

为 $F \in \mathbf{R}^{C \times H \times W}$ 。图 4 为 FCN 网络,分为 3 个模块,图中下标符号 \sim 代表非, k 为 0 或 1。当 $k = 0$ 时, $\sim k$ 为 1,此时的 FCN 网络为嵌入检测分支网络结构的权重分支。同理,当 $k = 1$ 时, $\sim k$ 为 0,FCN 网络作

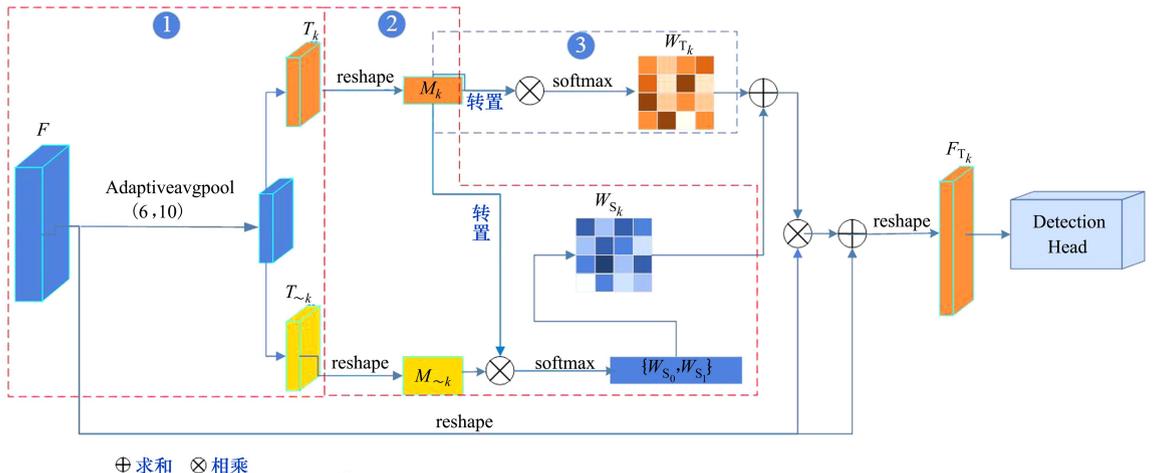


图 4 特征相关网络结构

为嵌入外观特征分支网络结构的权重分支。本文将 $k = 0/1$ 的 FCN 网络分别嵌入检测分支和 Re-ID 特征提取之前,实现任务的协同学习。

模块1 首先采用 Adaptiveavgpool 获得特征信息 $F' \in \mathbf{R}^{C \times H' \times W'}$;其次将 F' 作为卷积层的输入得到 2 个学习任务的特征映射 T_0 和 T_1 ;最后将 T_0 和 T_1 重塑为 $\{M_0, M_1\} \in \mathbf{R}^{C \times N'}$,其中 $N' = H' \times W'$ 。将特征图通道分为 0/1。

模块2 特征层互相关权重计算公式为

$$\omega_s^{ij} = \frac{\exp(M_{0/1}^i \cdot M_{1/0}^j)}{\sum_{j=0}^c \exp(M_{0/1}^i \cdot M_{1/0}^j)} \quad (5)$$

式中, ω_s^{ij} 代表任务 0/1 的第 i 个通道对任务 1/0 的第 j 个通道的影响,且 $\{W_{s_0}, W_{s_1}\} \in \mathbf{R}^{C \times C}$ 。

模块3 将 2 个任务与自身的转置矩阵相乘, softmax 层计算 W_{T_0} 和 W_{T_1} 且 $\{W_{T_0}, W_{T_1}\} \in \mathbf{R}^{C \times C}$, 计算公式如下

$$\omega_{T_k}^{ij} = \frac{\exp(M_k^i \cdot M_k^j)}{\sum_{j=0}^c \exp(M_k^i \cdot M_k^j)}, k \in \{0, 1\} \quad (6)$$

式中: $\omega_{T_k}^{ij}$ 为第 i, j 通道在通道注意力特征图中两者的关系。 W_{T_0} 和 W_{T_1} 表示任务的自相关权重图。

最终通过 λ 将模块 2 和 3 的相关权重融合,得到 $\{W_0, W_1\} \in \mathbf{R}^{C \times C}$,其中 λ 为训练参数。

$$W_{0/1} = \lambda W_{T_0/T_1} + (1 - \lambda) W_{S_0/S_1} \quad (7)$$

根据(5)式在 M_0 和 M_1 的转置之间进行矩阵运算,学习 2 个任务的共性并遵循 softmax 层输出互相关权重;其次,将 M_0 和 M_1 分别代入(6)式得到 2 个学习任务的自相关特征;最后,将(5)~(6)式代入(7)式计算融合特征相关值。为了增强每个任务的原始语义信息,应用残差连接将增强后的特征与原始特征 F 融合。在 FCN 网络引入 ELU 激活函数^[13]如图 5 所示。

如图 5 所示 ELU 让输入的负值能返回一些信息,更大程度上保留有效信息,并让整体输出值的均值维持在 0 附近,收敛速度加快,模型的泛化能力变强。在 SFP-JDE 中使用 ELU 激活函数能够提升网络特征学习能力,从而利于提高 2 个任务学习的公平性并有效减少 ID 切换的次数。经实验测试,FCN 网络使用 ELU 激活函数能够提升整体 SFP-JDE 网络架构对于目标的检测能力,从而更有利于对重叠目标的检测与跟踪,能有效降低 IDS、增大 MOTA。ELU 激活函数的数学形式如(8)式所示。

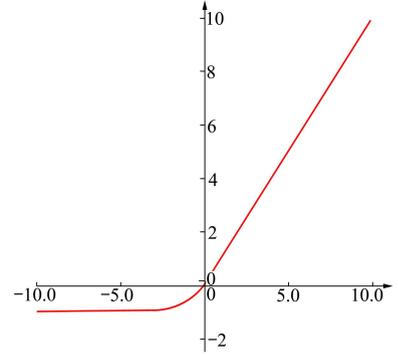


图5 ELU 激活函数

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{if } x \leq 0 \end{cases} \quad (8)$$

2.3 PCCs-Ma 改进关联度量

在线关联中卡尔曼滤波器^[7]预测轨迹状态,计算目标轨迹的外观特征和运动特征的相似度,作为匈牙利算法^[9]的相似代价矩阵来解决的轨迹匹配问题。设运动信息的距离为 d_1 , 马氏距离为 $d^{(1)}(i, j)$, 用 $\rho_{d_{y_i}}$ 表示目标轨迹的观察值和预测值向量的相关程度,外观特征向量之间的距离为 d_2 。传统方法采用马氏距离计算运动信息,尽管马氏距离可以考虑到各种变量之间的联系,但同时也放大了极小权值变量的作用。本文考虑到不同轨迹信息的有效性,采用度量 2 个变量之间相关程度的 PCCs 融合马氏距离,以减少信息冗余。

本文为了进一步测量矢量的距离并包含原有运动信息,组合 2 个度量表征运动信息实现数据关联。马氏距离 d_1 的基本计算公式如(9)式所示。

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \quad (9)$$

式中, d_j 为物体的 Bbox 信息。 $\rho_{d_{y_i}}$ 的定义为

$$\rho_{d_{y_i}} = \frac{\text{Cov}(d_j, y_i)}{\sqrt{D(d_j)} \sqrt{D(y_i)}} = \frac{\mathbf{E}((d_j - \mathbf{E}d_j)(y_i - \mathbf{E}y_i))}{\sqrt{D(d_j)} \sqrt{D(y_i)}} \quad (10)$$

式中, y_i 为目标跟踪的 Bbox 信息。

由于目标的预测值和观察值存在不相关的情况,故由(9)式和(10)式得

$$d_1 = \begin{cases} d^{(1)}(i, j) \times \rho_{d_{y_i}}, & \rho_{d_{y_i}} \geq 0 \\ d^{(1)}(i, j) \times C, & \rho_{d_{y_i}} < 0 \end{cases} \quad (11)$$

式中, C 的值趋于无穷大。

3 算法训练与实验结果分析

实验平台主要由硬件和软件两部分平台构成,其中硬件平台主要配置包括:单卡 NVIDIA RTX2060。软件环境为所搭建的深度学习平台,包括:Ubuntu16.04 操作系统, Cuda10.2, Cudnn7.6, pytorch1.7.1-gpu, OpenCV 4.5.1.48, cython-bbox0.1.3, scikit-learn0.24.1, python3.6 等,针对本文中研究的行人多目标跟踪,训练基准 JDE 和 SFP-JDE 模型。

3.1 算法训练

1) 数据集的构成

本文选用 CUHK-SYSU^[14]、PRW^[15]、MOT17^[16] 3 个多目标公共数据集,总共 22 222 张图片,均为同一个类别标注。数据集中的目标类别仅包含行人一类,同时已剔除训练集中与测试集重复部分。通过

CUHK-SYSU^[14]、PRW^[15]数据集的测试集测试模型的检测准确率,通过 MOT 基准数据集中的 MOT15^[17]、MOT16^[16]、MOT20^[18]评估 SFP-JDE 多目标跟踪算法的性能。

2) 模型训练

由于 JDE 模型是在 8 张 Nvidia Titan Xp 显卡、批量大小为 32 的环境下训练。为了避免不公平的对比,将模型训练分为 JDE 模型训练和 SFP-JDE 模型训练。以本实验硬件环境训练基准 JDE 模型并加载预训练模型,训练 JDE 模型未使用预训练权重。根据硬件配置本实验 batch-size 为 2,初始学习率为 0.01,采用等间隔随机调整学习率,动量值 0.9,衰减因子为 10^{-4} ,使用 SGD 进行 30 个 epochs 的训练,图片尺寸被调整到 864×480 后再输入到网络中,且其余超参数保持不变。2 个模型的训练损失变化如图 6 所示。

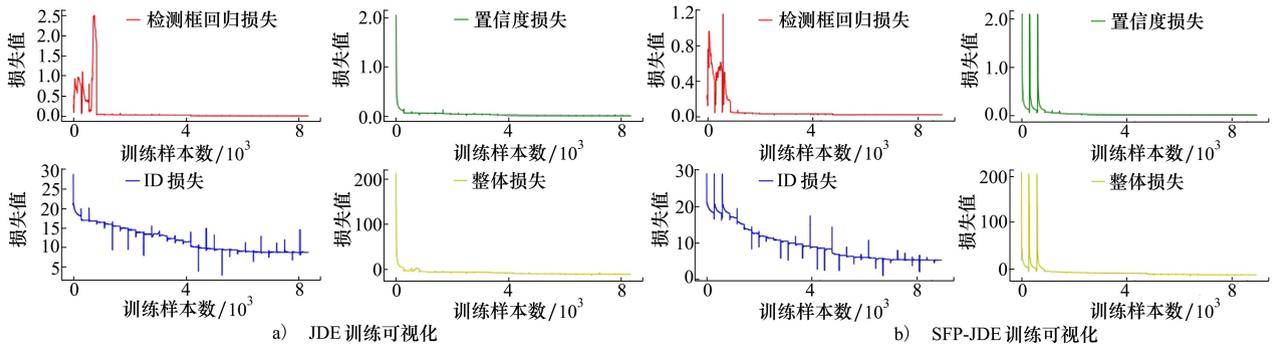


图 6 30 批次训练损失变化

本文 SFP-JDE 与基准 JDE 模型的检测性能相比,在测试过程中 SFP-JDE 的各个损失收敛速度更快,对类内目标准确分布且无漏检现象。在 CUHK-SYSU、PRW 数据集的测试上定量评估模型的检测平均准确率,基准模型针对行人的检测平均准确率为 82.37%,而本文的检测器模型可达 86.31%,相比提升约 3.94%。

3.2 定性分析

跟踪算法的非极大值最大重叠率为 0.4,置信度阈值和 IOU 阈值最大余弦距离为 0.5。以典型 MOT16 中的一个视频场景对本文算法定性分析,并和基准 JDE 算法对比目标跟踪效果。可视化结果分别如图 7~8 所示,图片右上角为跟踪状态,右下

角为目标的 ID 号。

从图 7 中可以看到,51 和 91 号目标在被遮挡后产生了 ID switch 现象,转换为一个新的轨迹标志号,而如图 8 所示,采用了 SPA 注意力模块、FCN 网络和皮尔逊相关改进关联度量的改进算法能够在目标被遮挡后仍然保持原有的标志号,这使得目标跟踪的准确度进一步提高,并且有助于保持目标轨迹的完整性。由可视化结果可见,本文算法对于复杂场景的行人多目标跟踪有较高的位置信息准确度、ID 信息准确度,能够更有效地避免 ID switch 现象,并有助于生成目标运动的全局轨迹,取得了良好的效果。



图 7 基准 JDE 跟踪可视化结果



图 8 SFP-JDE 跟踪可视化结果

3.3 定量分析

在 MOT challenge 数据集上评估本文算法,其中 MOTA 为多目标跟踪准确度,IDs 是在视频流跟踪过程中总共出现的 ID 切换次数,IDP 为目标 ID 准确性,IDR 为目标 ID 召回率,IDF1 表征跟踪器的好坏。SPA 模块的嵌入使得网络层数变多、网络更深,从而在构建特征金字塔时,使用更加鲁棒的信息,获得强语义信息以提高检测效果。本文首先探讨在网络的不同深度嵌入 SPA 模块对模型的影响。其次,为证明 SPA 的语义增强有效性,本文在最优位置使用同样具有扩大感受作用的 SPPNet 与 SPA 模块进行对比验证实验。在 CUHK-SYSU、PRW 测试集上通过检测平均准确率 AP 评价模型。JDE 是一种 one-shot 方法, MOTA 等指标与架构中检测到

关联所有部分都有关系。通过表 1 实验结果对比可得深层嵌入 SPA 最优,由相同深度的 SPPNet 实验结果可得 SPA 对模型提升比 SPPNet 高,故可得 SPA 模块嵌入到深层网络的检测效果最优。

表 1 不同深度嵌入 SPA 以及 SPP 对比验证实验

浅层嵌入 SPA	中层嵌入 SPA	深层嵌入 SPA	深层嵌入 SPP	AP
✓				0.655 1
	✓			0.844 0
		✓		0.877 8
			✓	0.848 5

本文主要研究多目标准确度及当存在目标遮挡

情形时的标签切换问题,在 MOT 基准数据集的视频序列上做跟踪实验。表 2 展示了在 MOT 数据集上

SFP-JDE 与基准 JDE 算法的量化指标对比。

表 2 多目标跟踪实验结果对比

数据集	算法	MOTA/%	IDF1/%	IDP/%	IDR/%	IDs
MOT15	JDE(864×480)	48.3	57.5	56.5	58.5	652
	SFP-JDE	48.1	60.9	57.5	65.3	626
MOT16	JDE(864×480)	54.6	55.8	69.5	46.6	1 342
	SFP-JDE	61.5	62.7	73.2	54.9	1 279
MOT20	JDE(864×480)	16.7	17.9	54.2	10.7	9 877
	SFP-JDE	18.8	19.6	58.7	11.7	7 885

为进一步验证本文算法的有效性,本文将 JDE 的改进算法与其他算法在 MOT16 基准数据集序列上进行对比分析。其中,TADAM 为 JDE 改进算法;Deep SORT 为非 JDE 算法。为保证相对公平的条件对比,本文将在相同实验环境下训练 TADAM^[5]模型。Deep SORT^[19]算法为非 JDE 算法,使用原作者提供的 POI 检测器的检测文件做定量评估。上述算法定量评估测试结果如表 3 所示。

表 3 不同跟踪算法实验对比

算法	MOTA/%	IDF1/%
Deep SORT ^[19]	51.6	60.2
TADAM ^[5]	57.1	60.3
基准 JDE 算法	54.6	55.8
本文算法	61.5	62.7

由实验结果可以看出,本文的引入 SPA 模块、FCN 网络以及改进运动度量方程的改进算法的 MOTA 高于基准 JDE 算法,ID 准确率与 ID 召回率指标均有明显提升,ID switch 现象大幅减少,轨迹 ID 稳定性明显提高。具体分析如下:

1) 由表 2 分析可得,在最初的 MOT 挑战数据集 MOT15 上,相较于原算法,本文算法 IDF1 指标有 3.4%的提升,ID 召回率明显提高。在 MOT16 数据集上显著超过基准 JDE 算法,MOTA 和 IDF1 指标提高了 6.9%。MOT20 与上述数据集相比,更具挑战性,其数据集在 3 个非常拥挤的场景中拍摄,数据呈现高度、亮度多样性,故整体算法指标过低,但与基

准 JDE 算法相比,本文算法 MOTA 指标提高 2.1%,IDF1 指标提高 1.7%。

2) 由表 3 可知,本文算法相较于 Deep SORT^[17]算法,MOTA 指标提升将近 10%,IDF1 指标提高 2%。与 TADAM 算法相比,MOTA 提升 4.4%,IDF1 提升 2.4%。

综上所述,本文算法相较原算法跟踪能力明显提高,判断目标轨迹是否是同一个目标的能力变强。

4 结 论

为了解决 JDE 将检测和嵌入共同学习造成在目标短时遮挡以及 2 个学习任务提取信息不足造成的 ID 切换问题,本文提出了 SPA 注意力模块、FCN 网络以及利用相关度改进运动度量的多目标跟踪算法。注意力模块兼顾网络架构信息和正则化能够增强基础特征网络,获得更多的语义,从而提取深层有效特征;FCN 有利于保留原信息,获得目标的检测信息与 Re-ID 外观特征不同的语义信息增强;PCCs-Ma 改进运动特征相关度量,有利于强化跟踪过程中卡尔曼滤波的预测值与观察值之间的联系,提升了在短时遮挡场景下持续追踪目标运动轨迹的关联能力。实验结果表明,本文算法在行人目标短时遮挡复杂场景下能有效提升目标跟踪性能,使得目标跟踪定位更加准确。后续工作将会在模型中加入长短时间记忆感知的目标注意,以形成对目标更好的关注,进一步提升跟踪性能。

参考文献:

- [1] XIAO Tong, LI Shuang, WANG Bochao, et al. Joint detection and identification feature learning for person search[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3415-3424
- [2] WANG Zhongdao, ZHENG Liang, LIU Yixuan, et al. Towards real-time multi-object tracking[C]//Computer Vision-European Conference on Computer Vision, 2020: 107-122
- [3] ZHANG Yifu, WANG Chunyu, WANG Xinggang, 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
- [4] CHAABANE M, ZHANG P, BEVERIDGE J R, et al. DEFT: detection embeddings for tracking[EB/OL]. (2021-02-03) [2021-11-01]. <https://arxiv.org/abs/2102.02267>
- [5] GUO Song, WANG Jingya, WANG Xinchao, et al. Online multiple object tracking with cross-task synergy[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 8132-8141
- [6] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. (2018-04-08) [2021-11-01]. <https://arxiv.org/abs/1804.02767>
- [7] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82: 35-45
- [8] KUHN H W. The hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly, 1955, 2(1/2): 83-97
- [9] ZHANG Xuan, LUO Hao, FAN Xing, et al. AlignedReID: surpassing human-level performance in person re-identification[J/OL]. (2017-11-22) [2021-11-01]. <https://arxiv.org/abs/1711.08184>
- [10] GUO Jingda, MA Xu, SANSOM A, et al. Spanet: spatial pyramid attention network for enhanced image recognition[C]//2020 IEEE International Conference on Multimedia and Expo, London, 2020: 1-6
- [11] LIN T Y, DOLLAR P, GIRSHICK R. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 936-944
- [12] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916
- [13] DJORK-ARNÉ C, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs)[C]//International Conference on Learning Representations, San Juan, Puerto Rico, 2016: 1-14
- [14] XIAO T, LI S, WANG B, et al. Joint detection and identification feature learning for person search[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3415-3424
- [15] ZHENG L, ZHANG H, SUN S, et al. Person re-identification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1367-1376
- [16] MILAN A, LEAL-TAIXE L, REID I, et al. Mot16: a benchmark for multi-object tracking[J/OL]. (2016-03-02) [2021-11-01]. <https://arxiv.org/abs/1603.00831>
- [17] LEAL-TAIXE L, MILAN A, REID I, et al. MOTChallenge 2015: towards a benchmark for multi-target tracking[EB/OL]. (2015-04-08) [2021-11-01]. <https://arxiv.org/abs/1504.01942>
- [18] DENDORFER P, REZATOFIGHI H, MILAN A, et al. MOT20: a benchmark for multi object tracking in crowded scenes[EB/OL]. (2020-03-19) [2021-11-01]. <https://arxiv.org/abs/2003.09003>
- [19] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing, 2017: 3645-3649

JDE multi-object tracking algorithm integrating multi-level semantic enhancement

WANG Jun¹, WANG Peng², LI Xiaoyan¹, WANG Liang³,
SUN Mengyu¹, GAO Hui¹

(1.School of Electronics and Information Engineering, Xi'an Technological University, Xi'an 710021, China;
2.Development Planning Service, Xi'an Technological University, Xi'an 710021, China;
3.Shaanxi Academy of Aerospace Technology Application Co., Ltd, Xi'an 710100, China)

Abstract: In order to solve the problem of target ID switching caused by target occlusion and insufficient ID information and location information extraction in JDE (joint detection and embedding) algorithm, an improved multi-target tracking algorithm based on JDE is proposed in this paper. Firstly, the SPA feature space pyramid attention module is used to expand the receptive field and obtain more abundant semantic information to improve the detection accuracy of the model for different scale targets. Secondly, the FCN network makes the header and ID Embedding task collaborative learning to alleviate the excessive competition and enhance the original semantic information, effectively reducing the number of ID switching. Finally, PCCs-Ma motion measurement can strengthen the connection between Kalman filtering prediction and observation, and improve the reliability of similarity discrimination of motion characteristics. In order to verify the effectiveness of the algorithm, the JDE algorithm and the proposed algorithm are compared in the same experimental environment. The experimental results show that the average accuracy of model detection is improved by 3.94 %. On the MOT16 dataset, the MOTA and IDF1 indexes are increased by 6.9 %, and the number of ID switching of the improved algorithm is significantly reduced, and good tracking results are achieved.

Keywords: multi-object tracking; JDE algorithm; semantic information; SPA; receptive field

引用格式: 王俊, 王鹏, 李晓艳, 等. 融合多阶语义增强的JDE多目标跟踪算法[J]. 西北工业大学学报, 2022, 40(4): 944-952
WANG Jun, WANG Peng, LI Xiaoyan, et al. JDE multi-object tracking algorithm integrating multi-level semantic enhancement[J]. *Journal of Northwestern Polytechnical University*, 2022, 40(4): 944-952 (in Chinese)