

弱光照条件下地铁站内乘客移动轨迹跟踪

谷金晶¹, 尹俊凯¹, 冯涛¹, 岳庚¹, 江志彬^{2,3}, 沈立⁴

(1.云南大学信息学院, 云南 昆明 650500; 2.同济大学道路与交通工程教育部重点实验室, 3.同济大学上海市轨道交通结构耐久与系统安全重点实验室, 上海 201804; 4.四川大学计算机学院, 四川 成都 610065)

摘要: 实时精准的乘客移动轨迹跟踪是实现地铁站智能化安全管理的关键技术。然而, 地铁站普遍存在弱光照和乘客频繁遮挡等问题, 导致计算机视觉领域主流的多目标跟踪技术在地铁管理实践中往往难以保证跟踪精度。为此, 本文提出了一种基于弱光图像增强的乘客移动轨迹跟踪方法, 旨在提升弱光照条件下存在遮挡情况下的地铁站内乘客移动轨迹的实时跟踪性能。首先, 引入先进的 Retinexformer 技术, 对弱光照监控视频进行细节恢复和图像质量提升; 随后, 分别采用基于观测中心的多目标跟踪算法 (Observation-Centric SORT, OC-SORT) 和基于稀疏特征的 SparseTrack 算法跟踪存在乘客遮挡情况下的地铁乘客轨迹。实验结果表明: 经图像增强处理后, OC-SORT 模型在弱光环境下的跟踪性能达到 73% 的 HOTA 和 96% 的 MOTA, 而 SparseTrack 模型则达到了 76% 的 HOTA 和 98% 的 MOTA。

关键词: 交通安全; 多目标跟踪; 计算机视觉; 弱光照图像增强

中图分类号: U291.69

文献标志码: A

Tracking of Movement Trajectories of Subway Passengers Under Low-Light Conditions

Gu Jinjing¹, Yin Junkai¹, Feng Tao¹, Yue Gen¹, Jiang Zhibin^{2,3}, Shen Li⁴

(1. School of Information Science and Engineering, Yunnan University, Kunming 650500, China; 2. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China; 3. Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China; 4. School of computing, Sichuan University, Chengdu 610065, China)

Abstract: Real-time and precise passenger trajectory tracking is a key technology for achieving intelligent safety management in subway stations. However, subway stations commonly suffer from low-light conditions and frequent passenger occlusions, which often compromise the tracking accuracy of existing mainstream multi-object tracking (MOT) techniques of computer vision in practical applications of subway management. To address this issue, this paper proposes a passenger trajectory tracking method based on low-light image enhancement, aiming to improve tracking performance in low-light subway station environments and frequent passenger occlusions. First, an advanced Retinexformer technique is introduced to restore details and enhance image quality in low-light surveillance videos. Subsequently, two MOT algorithms, i.e., the observation-centric SORT (OC-SORT) and the feature-sparse SparseTrack, are employed for passenger trajectory tracking. Experimental results demonstrate that after image enhancement, the OC-SORT model achieves 73% HOTA and 96% MOTA in low-light conditions, while the SparseTrack model achieves 76% HOTA and 98% MOTA.

Keywords: Traffic safety; multi-object tracking; computer vision; low-light image enhancement

地铁站作为城市公共交通系统的核心枢纽, 承担着大规模客流输送的重要职能。为保障运营安全并提升应急响应能力, 地铁站普遍部署了视频监控系统。其中, 基于监控视频对乘客移动轨迹进行实时准确的跟踪, 是实现地铁站智慧化管理与服务的关键技术^[1-2]。多目标跟踪 (multiple object tracking, MOT)

收稿日期: 2025-06-09

基金项目: 国家自然科学基金项目(52102382, 52372332); 上海市轨道交通结构耐久与系统安全重点实验室开放基金项目(R202403); 云南省基础研究计划面上项目(202401CF070164)

技术是计算机视觉领域的重要研究方向，其核心任务是从连续视频序列中持续检测并追踪多个目标，同时保持目标身份 ID 的一致性，最终实现时空维度上的完整轨迹重建。传统 MOT 方法主要遵循“检测-关联”范式^[3]，即先通过目标检测算法定位各帧中的物体，再利用时空特征进行跨帧数据关联以构建连续轨迹。目前，基于 YOLOv8^[4]和 SORT 的算法^[5]、DeepSORT^[6]和 ByteTrack^[7]等算法在良好光照条件下展现出优秀的多目标跟踪性能^[9,10]。

然而，地铁站多处于地下或半封闭式环境，其光照条件普遍欠佳，导致监控摄像头捕获的图像往往呈现暗部细节丢失、噪声增强、对比度不足等问题。这些问题不仅降低了视频监控的有效性，还增加了后续事件分析和追踪的难度。如何恢复图像的细节、提高图像的清晰度和可辨识度，成为提高乘客移动轨迹跟踪精度和有效性亟待解决的挑战。

弱光图像增强技术的发展为解决弱光环境下的计算机视觉任务提供了新的解决思路。该技术旨在通过图像亮度增强、对比度调节、细节恢复等手段改善图像质量，为下游的检测、识别与追踪等任务提供更可靠的输入。当前，基于 Retinex 理论的算法如 RetinexNet^[11]、基于分解与恢复的 EnlightenGAN^[11]和 RetinexFormer^[13-15]等深度增强网络在弱光图像增强方面取得了显著成效。这些端到端训练的模型能有效重建图像的纹理结构与空间特征，并提升后续视觉任务在弱光照条件下的鲁棒性和准确性。

基于此，针对地铁站弱光照环境的固有特性，本文提出了一种基于 RetinexFormer 的乘客移动轨迹跟踪方法，如图 1 所示。首先，采用 Retinexformer 算法分离图像的照明与反射率分量，恢复因光照不足损失的全局结构信息，进一步提升局部纹理的可见性。该融合策略有效解决了弱光监控视频中存在的低对比度、噪声干扰和细节模糊问题解决了如何恢复真实弱光照条件下的地铁站监控视频细节这一问题。接着，基于观测中心的多目标跟踪算法(Observation-Centric SORT, OC-SORT)^[14]和基于稀疏特征的 SparseTrack^[16]两种当前性能优越的 MOT 算法，实现乘客轨迹精准追踪。最后，基于真实地铁站监控视频数据，验证所提方法的有效性，并通过跟踪性能评价输出性能较优的跟踪结果。

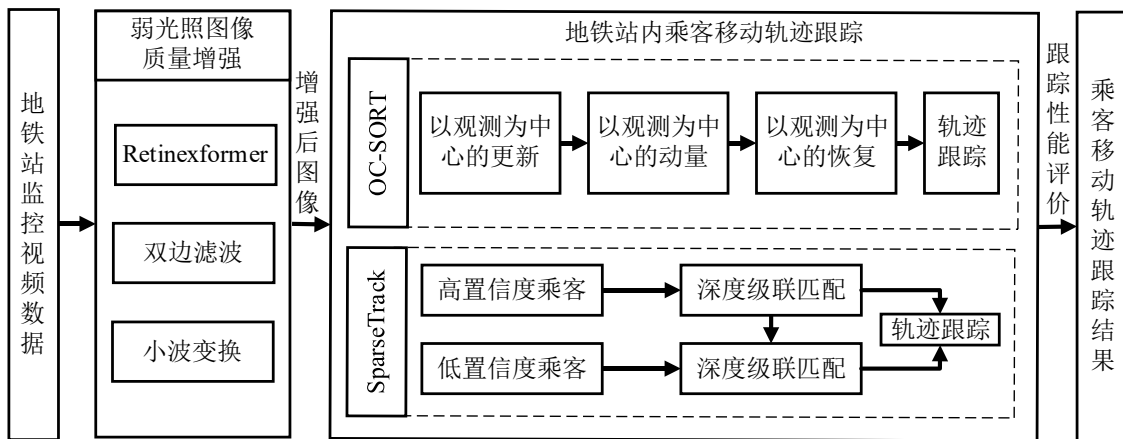


图 1 弱光照条件下地铁站内乘客移动轨迹跟踪

Fig.1 Tracking of Movement Trajectories of Subway Passengers Under Low-Light Conditions

1 弱光照条件下的地铁站监控图像质量增强

1.1 基于 Retinexformer 的弱光照图像质量增强

Retinex 是一种图像增强方法，其核心在于模拟人类视网膜和大脑皮层处理光线的方式，实现在不同光照条件下保持对物体颜色的恒定感知。而基于 Retinex 理论的传统方法假定图像没有噪声，侧重于光照估计，通常会引入严重的局部噪声或颜色失真。此外，传统基于卷积神经网络 (Convolutional Neural

Networks, CNN) 和 Retinex 的方法通常采用多阶段训练, 过程繁琐, 且 CNN 在捕捉长距离依赖方面存在局限性。针对上述问题, Retinexformer 将 Transformer^[20] 和 Retinex 相结合, 简化了训练过程, 通过端到端的方式完成图像增强。Retinexformer 主要由单阶段 Retinex 框架 (One-stage Retinex-based Framework, ORF) 和基于照明指导的 Transformer (Illumination-Guided Transformer, IGT) 构成。

1.1.1 单阶段 Retinex 框架

依据 Retinex 理论将低光照图像 I 表示为反射率图像 R 与照明图 L 的逐点乘积。接着, 采用 ORF 计算得到弱光增强图像, 如式 (1) 所示:

$$(I_{lu}, F_{lu}) = \mathcal{E}(I, L_p), I_{en} = \mathcal{R}(I_{lu}, F_{lu}) \quad (1)$$

其中, L_p 表示输入图像的照明先验, 计算方式为通道均值, \mathcal{E} 表示照明估计器, 输出照增强图像 I_{lu} 和照亮特征 F_{lu} , \mathcal{R} 表示退化恢复器, 将 I_{lu} 和 F_{lu} 送入退化恢复器 \mathcal{R} , 则生成增强图像 I_{en} 。

1.1.2 基于照明指导的 Transformer 模块

在低光照图像处理中, 暗光区域通常面临更严重的信息丢失, 这使得恢复这部分内容更加困难。相比之下, 光照条件较好的区域不仅信息更加完整, 还能提供更丰富的上下文信息, 这对增强暗光区域尤为重要。因此, 使用照明特征将不同光照条件下区域的照明信息和交互进行编码, 能够有效指导自注意力机制的计算, 从而帮助网络在处理低光照区域时保持一致性和准确性。为了实现这一点, 我们将输入图像的特征进行重塑, 使其能够适应自注意力机制的处理。首先将特征 $F_m \in \mathbb{R}^{H \times W \times C}$ 重塑为 $HW \times C$ 的 token 序列 X , 然后再分为 k 个注意力头:

$$X = [X_1, X_2, \dots, X_k], \quad X_i \in \mathbb{R}^{HW \times d_k} \quad (2)$$

对于每个注意力头, 分别将其线性投影为查询向量 $Q_i \in \mathbb{R}^{HW \times d_k}$, 键向量 $K_i \in \mathbb{R}^{HW \times d_k}$, 值向量 $V_i \in \mathbb{R}^{HW \times d_k}$, 如下所示:

$$Q_i = X_i W_Q^T, \quad K_i = X_i W_K^T, \quad V_i = X_i W_V^T \quad (3)$$

其中, W_Q^T , W_K^T 和 W_V^T 为全连接层的可学习矩阵, T 表示矩阵转置。

同一图像的暗光区域通常有更严重的信息损失, 恢复起来更加困难, 而光照条件较好的区域可以提供上下文表示, 以帮助增强较暗区域。为了加强照明信息的指导作用, 在计算自注意力时加入了照明特征 Y_i 帮助模型在不同光照条件下进行信息的交互与融合:

$$Attention(Q_i, K_i, V_i, Y_i) = \text{softmax} \left(\frac{K_i^T Q_i + Y_i}{\sqrt{D'}} \right) V_i \quad (4)$$

其中, D' 表示每个 token 的维度。照明特征 Y_i 被引入注意力计算中, 指导模型在较暗区域增强信息, 同时保持光照较好区域的上下文表示。最终, 所有注意力头的输出将被合并为最终的多头自注意力输出:

$$Multi-headAttention(Q_i, K_i, V_i, Y_i) = \text{concat}(head_1, \dots, head_h) W_o \quad (5)$$

其中, W_o 代表输出权重矩阵, 是多头自注意力机制中最后一个步骤中的一个重要组件, 作用是将多个头的输出合并后进行线性变换, 生成最终的自注意力输出。

2 乘客移动轨迹跟踪

2.1 基于 OC-SORT 的地铁站乘客轨迹跟踪

OC-SORT 以 SORT 算法^[2]为基础, 针对遮挡和非线性运动问题, 提出了一种以观测为中心的方法。OC-SORT 提出的改进如下所示:

2.1.1 以观测为中心的更新

跟踪目标被遮挡时, 由于卡尔曼滤波器仅依赖线性预测而无观测更新, 导致状态误差随时间增长, 重关联后卡尔曼滤波器的速度方向估计偏差仍较大, 容易再次丢失目标。为此, OC-SORT 提出了一种以观测为中心的参数更新方法来减少累积误差。一旦目标长时间未被跟踪后再次关联, 会重新检查被追踪目标的丢失时间段, 并用目标丢失时刻的观测状态 z_{t_1} 和重新关联时的观测状态 z_{t_2} 来更新卡尔曼滤波器的参数, 其虚拟轨迹表示为:

$$z_t = \text{Traj}_{\text{virtual}}(z_{t_1}, z_{t_2}, t), t_1 < t < t_2 \quad (6)$$

沿着虚拟轨迹 z_t 执行参数更新:

$$\text{re-update} \begin{cases} K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} \\ \hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (\tilde{z}_t - H_t \hat{X}_{t|t-1}) \\ P_{t|t} = (I - K_t H_t) P_{t|t-1} \end{cases} \quad (7)$$

其中, K_t 为卡尔曼增益, 用于权衡预测状态和观测值的置信度, $P_{t|t-1}$ 为预测误差的协方差矩阵, $P_{t|t}$ 为更新后的误差协方差矩阵, H_t 为观测矩阵, T 表示矩阵转置, R_t 为观测噪声的协方差矩阵, $\hat{X}_{t|t-1}$ 为 t 时刻的预测状态 (基于 $t-1$ 时刻), $\hat{X}_{t|t}$ 为更新后的状态估计, I 为单位矩阵。

2.1.2 以观测为中心的动量

由于 SORT 算法以卡尔曼滤波估计为中心, 而在复杂场景下目标运动状态为非线性, 导致卡尔曼滤波器性能下降, 且使用估计值会导致误差累计。OC-SORT 提出了一种以观测为中心的动量, 通过引入方向一致性动量来降低 SORT 算法对状态噪声的敏感度。关联代价矩阵为:

$$C(\hat{X}, Z) = C_{\text{IoU}}(\hat{X}, Z) + \lambda C_v(T, Z) \quad (8)$$

其中, $\hat{X} \in \mathbb{R}^{N \times 7}$ 为对象状态估计的集合, $Z \in \mathbb{R}^{M \times 5}$ 为新的时间步长上的观测数据集, λ 为权重因子, T 为所有追踪轨迹的集合, C_{IoU} 用于计算错误匹配的 IoU 值, C_v 计算轨迹的前检测值与新检测结果的方向一致性。

2.1.3 以观测为中心的恢复

以观测为中心的恢复 (Observation-centric Recovery, OCR) 是一种观测数据中心的恢复方法, 用于解决目标跟踪中的累积误差问题。当目标轨迹出现丢失或错误匹配等情况时, 传统的 SORT 算法可能无法正确恢复目标的轨迹。OCR 会在常规关联阶段之后, 重新尝试将未匹配轨迹的最后一次观测与之后的未匹配观测进行关联, 从而缓解目标静止或短暂被遮挡而跟踪失败的情况, 提高乘客运动轨迹跟踪的连续性。

2.2 基于 SparseTrack 的地铁站乘客轨迹跟踪

在复杂场景中, MOT 面临密集人群导致的遮挡目标精确关联难的挑战。当前主流的“检测-跟踪”方

法 ByteTrack 通过单独关联低置信度目标以缓解遮挡问题，但在严重拥挤或大面积遮挡情形下，仍然容易出现误匹配现象。SparseTrack 通过伪深度估计将密集场景中的目标集进行稀疏子集划分，并通过深度级联匹配进行数据关联，从而有效降低遮挡目标间位置相似性带来的关联混淆。

2.2.1 伪深度估计

伪深度估计的目的是从二维图像中推断出目标之间的相对远近关系，无需使用真实的深度传感器。该方法建立在以下两个基本假设之上：摄像头位置高于地面，且场景中所有目标处于平坦地面上。设图像高度为 H ，目标检测框下边缘中心点在图像坐标系中的纵坐标为 y_p ，伪深度估计值为 $L_p = H - y_p$ 。虽然伪深度值不能表示目标在三维空间中的真实深度，但能充分测量同一地面上不同物体之间的相对深度关系，为密集人群分割提供了基础。

2.2.2 深度级联匹配

伪深度值的引入能够将原本密集的目标集划分为多个非重叠的深度层，每层包含伪深度值相近的目标。基于此，SparseTrack 设计了深度级联匹配算法，核心步骤如下：

步骤 1：深度划分。给定所有目标的伪深度值，设其最大最小值为：

$$L_{\min} = \min_{d \in D} L_p(d), \quad L_{\max} = \max_{d \in D} L_p(d) \quad (9)$$

其中， $L_p(d)$ 表示目标 d 的伪深度值，由检测框下边缘纵坐标计算， D 表示当前帧的所有检测目标集合。

接着将该区间平均划分为 k 个深度区间：

$$I_i = [L_{\min} + i \cdot \Delta, L_{\min} + (i+1) \cdot \Delta), \quad \Delta = \frac{L_{\max} - L_{\min}}{k} \quad (10)$$

其中， I_i 将第 i 个深度区间， Δ 表示每个深度区间的宽度。

步骤 2：构造层级子集。将检测集合 D 和轨迹集合 T 分别按深度区间划分为 k 个子集：

$$D = \bigcup_{i=0}^{k-1} D_i, \quad T = \bigcup_{i=0}^{k-1} T_i, \quad D_i = \{d \in D \mid L_p(d) \in I_i\} \quad (11)$$

其中， D_i 属于第 i 个深度区间的检测目标子集， T_i 属于第 i 个深度区间的轨迹子集。

步骤 3：级联匹配。对每一层 $i \in [0, k-1]$ ，首先利用 IoU 构造轨迹子集 T_i 与检测子集 D_i 的匹配代价矩阵：

$$C_i(j, l) = 1 - \text{IoU}(T_i^j, D_i^l) \quad (12)$$

其中， $C_i(j, l)$ 表示第 i 层的代价矩阵，表示轨迹 T_i^j 与检测 D_i^l 的匹配代价，接着使用匈牙利算法求最小代价匹配，而未匹配的轨迹与检测子集，将参与下一层匹配。

3 实验结果展示及案例分析

3.1 数据集及评价指标

1) 地铁站视频监控数据采集

本研究采集真实的地铁站的监控视频数据，视频帧率为 25 帧/秒，分辨率为 1280*720。视频视角固

定，场景光线较暗，不同时刻的图像内均包括上下楼梯的乘客、乘坐扶梯的乘客、站台上移动的乘客和静止的候车乘客。

2) 评价指标

平均精度 AP (Average Precision) 用于衡量检测器在单帧中的目标定位精度，反映检测阶段性能。多目标跟踪准确率 MOTA (Multiple Object Tracking Accuracy) 用于综合衡量跟踪器在检测精度 (漏检、误检) 和关联一致性上的整体表现。高阶跟踪准确率 HOTA (Higher Order Tracking Accuracy) 用于同时评估检测精度和关联一致性，避免 MOTA 的偏检测倾向。多目标跟踪精确率 MOTP (Multiple Object Tracking Precision) 用于衡量跟踪器定位精度 (如边界框重合度)，忽略 ID 一致性。多目标检测准确率 MODA (Multiple Object Detection Accuracy)，评估检测性能，无 ID 切换惩罚。身份 F1 分数 IDF1 (Identity F1 Score) 用于评估 ID 保持能力，计算匹配轨迹与真实轨迹的 F1 分数。

3.2 图像质量增强

Retinexformer 图像增强的代码环境为 python 3.8.18, pytorch 1.11.0, torchvision 0.12.0, cuda 11.3。使用公开数据集 FiveK (MIT-Adobe FiveK) 的预训练模型进行图像增强，本文采集的视频数据时长 8 分 31 秒，帧率为 25 帧/秒，总共 12781 帧图像。使用 Retinexformer 模型进行图像质量增强后的效果如图 2 所示。



图 2 图像增强效果对比示例

Fig.2 Image Enhancement Results

由图 2 可见，图片增强后的光照强度明显比增强前要高，更多暗部的细节 (如红框所示) 能够更清晰的被展示出来。图像增强后能够很好地恢复出弱光照图像中的细节，帮助 MOT 模型在能够在图像增强后的视频中提取到更有用的信息。

3.3 跟踪模型训练

训练 OC-SORT 和 SparseTrack 模型所用 GPU 为 NVIDIA RTX4090，训练环境为 python 3.8.20, pytorch 1.11.0, torchvision 0.12.0, cuda 11.3。将图像质量增强前后的图像分别送入 OC-SORT 和 SparseTrack 模型进行跟踪模型训练，模型训练迭代 300 轮次，结果如表 1 所示。

表 1 模型训练效果对比

Tab. 1 Model Training Performance Comparison

评价指标	跟踪模型	无图像增强	基于双边滤波	基于小波变换	基于 Retinexformer
AP[IoU=0.50:0.95 area= all]	OC-SORT	0.806	0.809	0.808	0.809
	SparseTrack	0.85973	0.83410	0.83734	0.86332

AP[IoU=0.50 area=all]	OC-SORT	0.990	0.990	0.990	0.990
	SparseTrack	0.99007	0.99007	0.99008	0.99009
AP[IoU=0.750 area=all]	OC-SORT	0.943	0.953	0.953	0.953
	SparseTrack	<u>0.97725</u>	<u>0.96592</u>	<u>0.96600</u>	<u>0.97745</u>
AP[IoU=0.50:0.95 area= small]	OC-SORT	0.593	0.594	0.589	0.595
	SparseTrack	0.65852	0.65250	0.66301	0.64565
AP[IoU=0.50:0.95 area= medium]	OC-SORT	0.790	0.791	0.790	0.792
	SparseTrack	0.85008	0.82170	0.82285	0.85242
AP[IoU=0.50:0.95 area= large]	OC-SORT	0.839	0.838	0.840	0.842
	SparseTrack	0.87748	0.86306	0.86515	0.88076

如表 1 所示，AP 是模型在不同类别下的检测平均准确度，数值范围通常在 0 到 1 之间，值越高表示性能越好。最优和次优的结果分别用粗体和加下划线标出。IoU=0.50:0.95 | area= all 表示表示在目标检测任务中，评估模型性能时使用的交并比（Intersection over Union, IoU）阈值范围为 0.50 到 0.95，并且考虑所有大小的目标。相对于基于传统双边滤波和小波变换的图像质量增强方法，基于 Retinexformer 的 SparseTrack 模型在 IoU 阈值为 0.50 时，所有目标大小范围内的 AP 最高并达到了 0.99009。基于 Retinexformer 模型增强后的视频送入跟踪模型后，AP 值均普遍得到提升。因此，弱光照环境下的图像增强能够对 MOT 模型训练阶段起到正向作用，有利于模型更好地理解图像中的细节。此外，对比两种 MOT 跟踪算法发现，无论是在哪种参数限制下，在相同的训练条件下，SparseTrack 模型的训练效果都明显优于 OC-SORT 模型。

3.4 跟踪模型测试

完成模型训练之后，在[IoU=0.50 | area=all]的设置下进行模型测试，结果如表 2 所示。

表 2 模型跟踪性能对比
Tab. 2 Model Prediction Performance Comparison

评价指标	跟踪模型	无图像增强	基于双边滤波	基于小波变换	基于 Retinexformer
MOTA	OC-SORT	0.96213	0.95082	0.95807	0.96326
	SparseTrack	0.98416	0.97215	0.98143	0.98589
HOTA	OC-SORT	0.72042	0.69692	0.71745	0.73387
	SparseTrack	0.75228	0.72452	0.74553	0.75657
MOTP	OC-SORT	0.89761	0.86936	0.88343	0.89791
	SparseTrack	0.90744	0.88480	0.90079	0.90822
MODA	OC-SORT	0.96605	0.95474	0.96189	0.96686
	SparseTrack	0.98838	0.97680	0.98582	0.98811
IDF1	OC-SORT	0.74629	0.75352	0.76130	0.76259
	SparseTrack	0.76519	0.75643	0.76491	0.76950

如表 2 所示，SparseTrack 模型的跟踪性能综合比 OC-SORT 模型更优秀。经图像增强处理后，OC-SORT 模型在弱光环境下的跟踪性能达到 73% 的 HOTA 和 96% 的 MOTA，而 SparseTrack 模型则达到了 76% 的 HOTA 和 98% 的 MOTA。基于 Retinexformer 的 SparseTrack 模型的所有指标值都明显优于基于 Retinexformer 的 OC-SORT 模型，且图像增强也起到了显著效果，说明图像增强后的检测和跟踪效果比增强前更加优秀，对于多目标跟踪效果起到了不错的性能提升。

3.5 跟踪结果

1) 基于 Retinexformer 的 OC-SORT 模型跟踪结果

基于 Retinexformer 的 OC-SORT 模型对连续视频帧进行跟踪，效果如图 3 所示。



图 3 基于 Retinexformer 的 OC-SORT 模型跟踪示意图

Fig.3 Tracking Example of the OC-SORT Model Based on Retinexformer

如图 3 所示，基于 Retinexformer 的 OC-SORT 模型能准确跟踪视频中乘客的移动轨迹。此外，以典型乘客的跟踪结果为例，对比图像增强前后 OC-SORT 模型的跟踪结果。如图 4 所示。

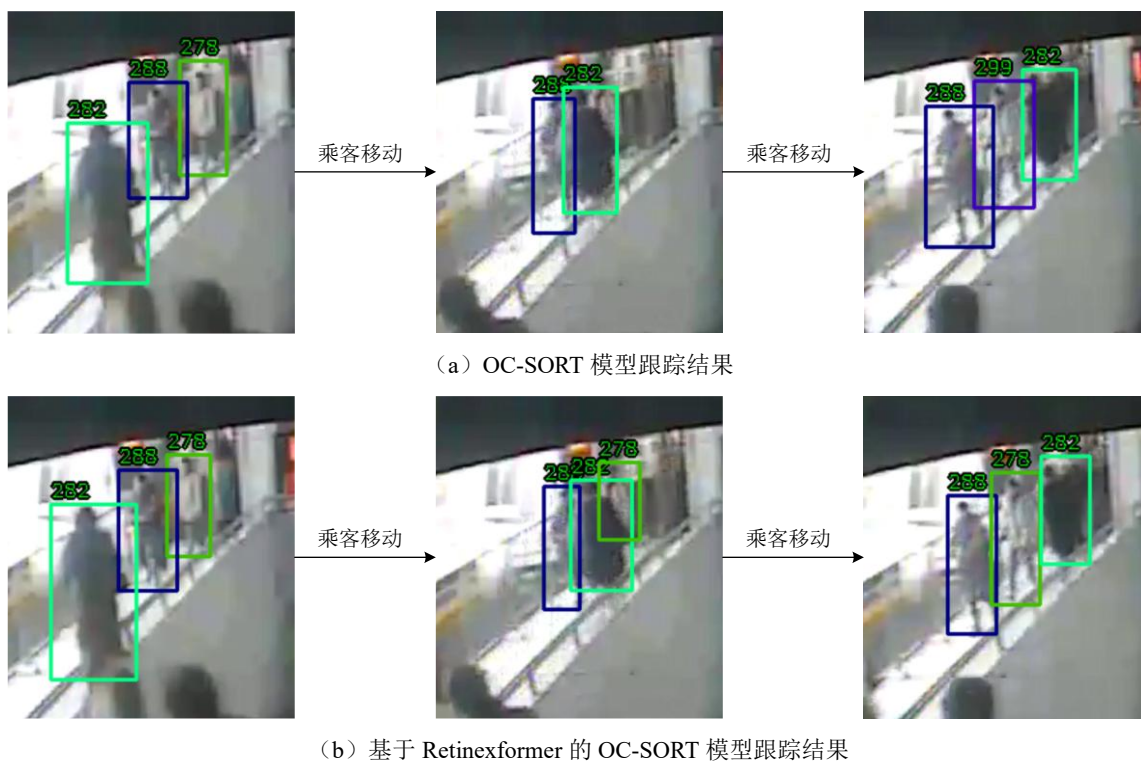


图 4 图像质量增强前后 OC-SORT 模型的跟踪效果示例

Fig.4 Tracking Effect of the OC-SORT Model With/Without Image Quality Enhancement

由图 4 (a) 所示，使用 OC-SORT 模型对未图像增强前的视频进行乘客个体轨迹跟踪，ID 为 278

的乘客移动至与他人重叠时，模型无法识别此乘客，后续未能保持该乘客的 ID 继续跟踪。由图 4 (b) 可知，在图像增强后，OC-SORT 模型能够捕捉到更加深度的信息。针对同一乘客，在图像增强前，模型丢失了乘客的轨迹，轨迹跟踪发生了乘客身份 ID 漂移（ID 由 278 变为了 299）。而图像增强之后，模型能够准确跟踪这一乘客，过程中不中断且没有发生 ID 漂移的情况，跟踪性能得到了进一步的提升。

2) 基于 Retinexformer 的 SparseTrack 模型跟踪结果

基于 Retinexformer 的 SparseTrack 模型的跟踪示例，如图 5 所示。

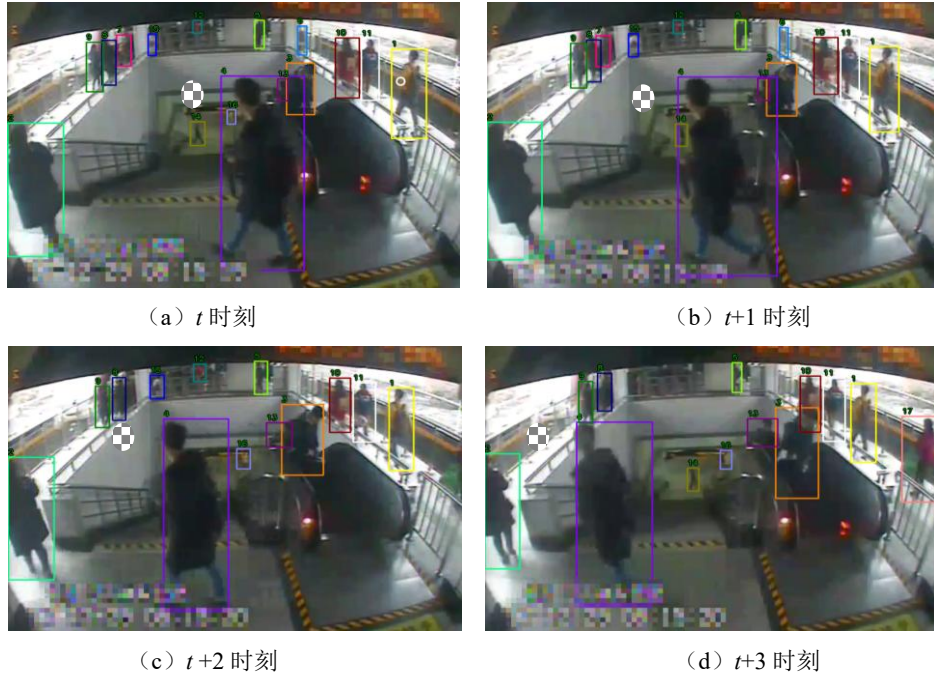


图 5 基于 Retinexformer 的 SparseTrack 模型跟踪示意图

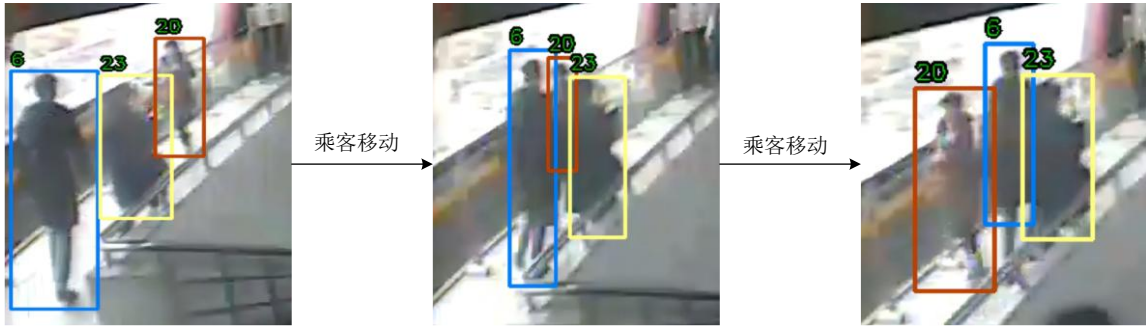
Fig.5 Tracking Example of the SparseTrack Model Based on Retinexformer

如图 5 中，远处微小、不清晰等的所有乘客都能够被成功识别并跟踪，能够在弱光照条件下地铁乘客轨迹跟踪。对比图像增强前后 SparseTrack 模型的跟踪结果。如图 6 所示。

由图 6 (a) 所示，未进行图像质量增强时，SparseTrack 模型没有检测出 ID 为 20 的乘客。由图 4 (b) 可见，基于 Retinexformer 的 SparseTrack 模型能够避免乘客移动至与他人重叠时，OC-SORT 模型无法识别此乘客，后续未能保持 ID 继续跟踪的问题。



(a) SparseTrack 模型跟踪结果



(b) 基于 Retinexformer 的 SparseTrack 模型跟踪结果

图 6 图像质量增强前后 SparseTrack 模型的跟踪效果示例

Fig.6 Tracking Effect of the SparseTrack Model with/without image quality enhancement

4 结论

本文采用弱光图像增强方法对乘客轨迹跟踪进行了研究，提出了一种基于 Retinexformer 技术与 OC-SORT 和 SparseTrack 算法相结合的多目标轨迹跟踪方法。最终得出以下结论：

1) 经过 Retinexformer 技术对弱光图像进行增强处理后，实验结果表明，OC-SORT 模型在弱光环境下的跟踪性能有所提升，具体表现为 HOTA 提高了 1.3%，MOTA 提高了 0.1%；而 SparseTrack 模型的跟踪性能也得到提升，HOTA 提高了 0.4%，MOTA 提高了 0.1%。这表明，图像质量的增强能够有效提升多目标跟踪算法的性能。

2) 在弱光环境下，OC-SORT 模型的跟踪性能分别达到了 73% 的 HOTA 和 96% 的 MOTA，SparseTrack 模型则达到了 76% 的 HOTA 和 98% 的 MOTA。两种算法均能够有效应对弱光照及遮挡条件下的地铁站乘客轨迹跟踪问题，验证了其在实际应用中的可行性与鲁棒性。

3) 根据目前实验结果，将进一步优化 OC-SORT 和 SparseTrack 模型，特别是提高其对目标遮挡、运动模糊以及快速运动目标的鲁棒性，从而提升整体跟踪性能。

参考文献

- [1] 徐玉萍, 梅哲源, 胡永葳. 高铁开通对城市绿色创新“量质齐升”影响研究[J]. 华东交通大学学报, 2025, 42(3): 87-95.
XU Y P, MEI Z Y, HU Y W. Research on the Impact of High-Speed Rail Opening on the “Quantitative and Qualitative Improvement” of Urban Green Innovation [J]. Journal of East China Jiaotong University, 2025, 42(3): 87-95.
- [2] 张凌, 彭思捷, 冯青松, 等. 地铁车辆段试车线减振改造后振动特性研究[J]. 华东交通大学学报, 2025, 42(3): 12-21.
ZHANG L, PENG S J, FENG Q S, et al. Research on Vibration Characteristics and Vibration Control Effect after the Retrofit of the Test Track in the Metro Depot [J]. Journal of East China Jiaotong University, 2025, 42(3): 12-21.
- [3] LUO W, XING J, MILAN A, et al. Multiple object tracking: A literature review[J]. Artificial intelligence, 2021, 293: 103448.
- [4] XIAO X, FENG X. Multi-object pedestrian tracking using improved YOLOv8 and OC-SORT[J]. Sensors, 2023, 23(20): 8439.
- [5] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE international conference on image processing (ICIP). IEEE, 2016: 3464-3468.
- [6] 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.
- [7] ZHANG Y, SUN P, 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] ZHENG Y, QI H, LI L, et al. Motion-guided and occlusion-aware multi-object tracking with hierarchical matching[J]. Pattern Recognition, 2024, 151(6): 110369.
- [9] YI A, ANANTRASIRICHAH N. A comprehensive study of object tracking in low-light environments[J]. Sensors, 2024, 24(13): 4359.

- [10] WEI C, WANG W, YANG W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.
- [11] JIANG Y, GONG X, LIU D, et al. Enlightengan: Deep light enhancement without paired supervision[J]. IEEE transactions on image processing, 2021, 30(1): 2340-2349.
- [12] 王登飞, 苏宏升, 陈光武, 等. 基于聚焦线性注意力 Retinexformer 的 TEDS 图像实时暗光增强方法研究[J]. 铁道科学与工程学报, 2024, 21(11): 4840-4850.
WANG D F, SU H S, CHEN G W, et al. Real time low light enhancement method of TEDS images based on focused linear attention Retinexformer[J]. Journal of Railway Science and Engineering, 2024, 21(11): 4840-4850.
- [13] 黄丹丹, 张钰晨, 陈广秋, 等. 基于 Transformer 的轻量级单目标跟踪算法[J]. 华中科技大学学报(自然科学版), 2025, 53(03): 41-47+158.
HUANG D D, ZHANG Y C, CHEN G Q, et al. Lightweight single-object tracking algorithm based on Transformer[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2025, 53(03): 41-47+158.
- [14] CAO J, PANG J, WENG X, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 9686-9696.
- [15] CAI Y, BIAN H, LIN J, et al. Retinexformer: One-stage retinex-based transformer for low-light image enhancement[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2023: 12504-12513.
- [16] LIU Z, WANG X, WANG C, et al. Sparsetrack: Multi-object tracking by performing scene decomposition based on pseudo-depth[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(5): 4870-4882
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 5998-6008.
- [18] LIU Z, WANG X, WANG C, et al. Sparsetrack: Multi-object tracking by performing scene decomposition based on pseudo-depth[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025.
- [19] GUO K, HHUANG S, SONG X, et al. SmartTrack: Sparse multiple objects association with selective re-identification tracking[J]. Computers and Electrical Engineering, 2025, 123: 110116.
- [20] 彭洋, 吴文欢, 张溟坤. 基于多尺度 Transformer 特征的道路场景语义分割网络[J]. 华东交通大学学报, 2025, 42(2): 110-118.
PENG Y, WU W H, ZHANG H K. Road Scene Semantic Segmentation Network Based on Multi-Scale Transformer Features[J]. Journal of East China Jiaotong University, 2025, 42(2): 110-118.



第一作者：谷晶晶（1990—），女，讲师，博士，硕士生导师，研究方向为计算机视觉，交通运输规划与管理。



通信作者：江志彬（1980—），男，1980年5月生，研究员，博士，博士生导师，研究方向为轨道交通运输组织优化。