

文章编号: 1001-4632 (2026) 02-0244-12

引用格式: 刘玉鑫. 基于视频分析技术的铁路客站多粒度动态客流评估模型[J]. 中国铁道科学, 2026, 47(2): 244-255.

Citation: LIU Yuxin. Multi-Granularity Yardstick for Dynamic Crowds Model for Railway Passenger Stations Based on Video Analysis Technology [J]. China Railway Science, 2026, 47 (2): 244-255.

基于视频分析技术的铁路客站多粒度 动态客流评估模型

刘玉鑫

(中国铁道科学研究院集团有限公司 电子计算技术研究所, 北京 100081)

摘要: 随着铁路行业快速发展, 客运任务持续增加, 客站面临日益严峻的客流安全问题。为实时监测客流动态、精细化解析客流多粒度特征, 提出 1 种基于视频分析技术的客站多粒度动态客流评估 (MYDC) 模型。首先, 构建铁路客运站客流数据集; 其次, 设计基于 YOLO 和判别式相关滤波 (DCF) 跟踪算法的旅客客流细粒度特征感知网络, 并改进面向客站的自适应人群定位 Transformer (CLTR) 模型, 以捕捉客流整体分布的粗粒度特征; 最后, 基于客流的物理属性及其微观与宏观特征, 构建多注意力时空图卷积网络 (MASTGCN), 挖掘客流的时空动态趋势, 评估站内客流安全风险等级。结果表明: 细粒度特征的提取累计误差为 6.9%, 粗粒度特征的识别精确率为 89.1%, 客流安全评估模型的召回率为 87.5%。该模型可为客流管理提供精准的数据支撑, 具有较强的工程应用价值。

关键词: 客流感知; 视频分析; 铁路客站; 安全评估; Transformer 模型; 时空图卷积网络

中图分类号: U293.1+3

文献标识码: A

doi: 10.3969/j.issn.1001-4632.2026.02.21

近年来, 铁路运输行业经历了飞速发展, 其运输效能和承载能力均实现了跨越式提升, 客运需求持续增长。客流量的急剧增长不仅放大了客站运营管理的复杂性, 还显著增加了安全事故的风险, 因此, 客站客流安全管理问题已成为目前亟待解决的关键议题。

传统客流安全评估多依赖人工经验, 在大客流高峰期难以全面把握车站状态, 易导致客运作业延误与运营效率下降, 亟须引入科学高效的监测技术。目前成熟的车站客流监测技术包括 AFC 检测^[1]、红外检测^[2]、视频识别^[3]、手机信令^[4]等。AFC 通过进出站闸机精准统计人数, 但无法评估客流整体分布与动态趋势; 红外检测需在车站各位置安装传感器, 投资成本高且抗干扰能力弱; 手机信令基于基站数据, 安装成本低、实时性高, 但涉及乘客隐私, 普适性差^[5]。视频识别技术基于车站既有安装的摄像机, 无须 2 次安装, 实时性强,

已逐渐成为主要手段。

传统视频分析如帧差法^[6]、光流法和背景差分法^[7]等虽能通过运动捕捉实现信息表达, 但无法区分旅客与携带物品, 导致直接应用时误差明显。随着人工智能、神经网络等技术的不断发展, 目标检测算法取得了显著进展。其中, 以 R-CNN^[8] 为代表的两阶段目标检测算法, 将兴趣区域检测与分类分步进行, 检测精度高但计算量大且速度慢; 以 YOLO^[9] 为代表的单阶段目标检测算法, 通过 1 个整体网络实现端到端检测, 在视频领域得到广泛应用。在客流细粒度分析方面, LIU 等^[10] 基于深度学习和光流法对上下车乘客进行统计; ZHANG 等^[11] 基于 CNN 对进出站乘客计数; SIPETAS 等^[12] 通过视频数据预测出站人数。然而, 单纯的目标检测无法捕捉运动方向、速度等动态信息, 难以准确反映客流通行情况。视频分析中的目标跟踪算法可准确获取乘客的运动轨迹, 有效

收稿日期: 2025-01-10; 修订日期: 2026-03-19

基金项目: 中国国家铁路集团有限公司科技研究开发计划课题 (N2025X026)

第一作者: 刘玉鑫 (1992—), 男, 山东聊城人, 高级工程师。E-mail: yxin_l@163.com

补充运动信息，如DeepSORT^[13]可在遮挡下实现行人重识别以获取轨迹，但其复杂度高，实际应用中难以满足大规模实时检测需求。在人群密度评估方面，Idress等^[14]采用Composition Loss函数通过头部位置局部最大值构建密度图并预测峰值；Abousamra等^[15]提出TopoCount方法，采用拓扑约束解决定位误差，利用个体间距非线性函数分离目标。然而，这些方法多用于室外通用场景。李瑞等^[16]提出了铁路客站下的人群密度估计方法，但无法实现对人群的定位。因此，人群评估在铁路垂直领域下的研究仍有待深入。在客流安全状态评估方面，豆飞等^[17]分析了城市轨道交通车站设施的类别及乘客聚集程度判断指标，提出1种基于云模型的车站客流控制判别方法；赵保锋等^[18]利用仿真方法分析实现了站台拥堵和预警等级划分；李得伟等^[19]构建了不同维度下城市轨道交通枢纽站台的短时客流密集度指数预测模型；胡国林^[20]则进一步聚焦线网层面，构建了城市轨道交通客流风险评估方法及系统。

综上所述，现有研究多聚焦于城市轨道交通运营系统，依赖日常运营管理经验，且时间分辨率较低，监测周期偏长，导致无法实现车站客流的实时评估；各分析指标间相对独立或片面，缺乏系统的量化评估体系。而铁路客站具有多区域聚集、空间

范围广阔、层次结构复杂等典型特征，其客流安全评估面临更大挑战。因此，亟须结合铁路客站乘客出行特点，引入先进的视频分析技术，实时统计客流指标，综合考虑铁路客流的时空分布、乘客个体和群体行为等多因素，开展定量联合研究。基于此，本文提出1种基于视频分析技术的铁路客站多粒度动态客流评估模型，实现对车站客流安全等级的量化评估，以降低安全事故的发生概率。

1 基于视频分析技术的铁路客站客流安全评估模型

基于视频分析技术的客站多粒度动态客流评估(Multi-granularity Yardstick for Dynamic Crowds, MYDC)模型框架如图1所示。该模型首先获取进站口、出站口、候车厅等区域实时视频，基于YOLO+DCF算法提取客站各通道的细粒度客流通行数据，随后采用自适应人群定位Transformer(Crowd Localization Transformer, CLTR)网络描述客流的粗粒度特征，最终融合细粒度与粗粒度客流数据，结合客站布局设施、旅客流线及列车时刻表等基础信息，通过时空图卷积模型进行综合分析，输出客站客流安全评估指标与风险等级。

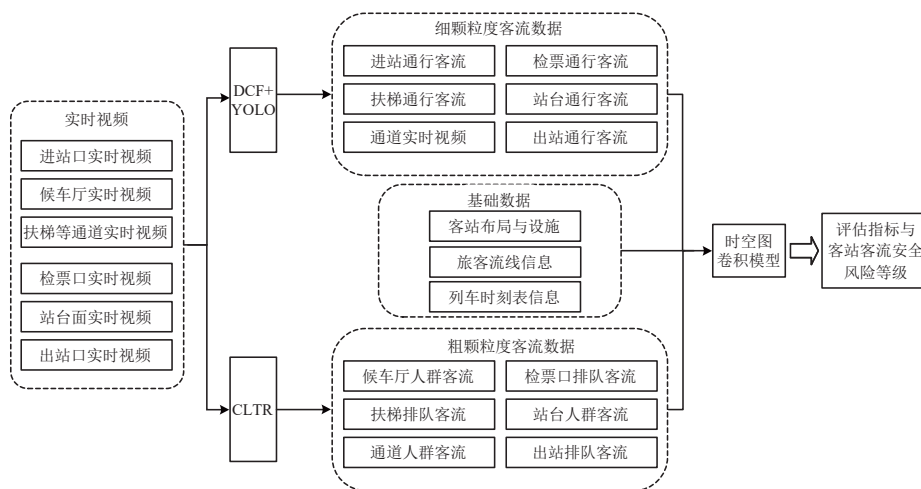


图1 客流评估模型框架

1.1 基于YOLO+DCF的客流细粒度分析

基于YOLO(You Only Look Once)+DCF(Discriminative Correlation Filter)的铁路客站精细化客流分析框架如图2所示，为实现精细化的客流描述，需要实时检测和跟踪站内关键区域的通行旅客，提取旅客在关键卡口的流线轨迹。先将关键监

控点位采集的实时视频数据作为细粒度分析的输入信息，通过YOLO网络^[11]实时推理出视频帧中人体头部的准确位置，再引入头部跟踪模型DCF^[21]，对人体头部数据进行连续实时追踪，并据此确定每位旅客的行进方向与路线。

YOLO系列算法将目标检测的位置框选和类

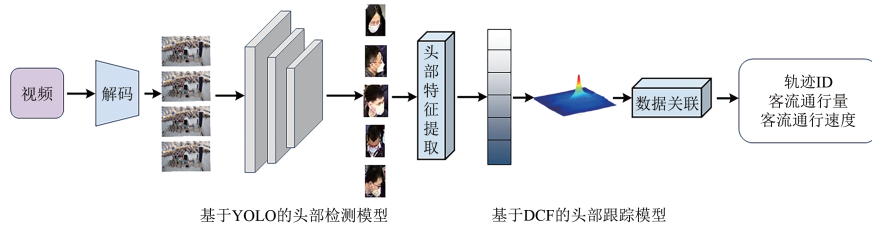


图2 基于YOLO+DCF的细粒度客流分析框架

别检测融合,在保证检测精度的同时显著提升了检测速度,其优良的检测效果在实时目标检测领域得到广泛应用。然而,铁路客站摄像机的部署位置与架设高度往往受限于复杂的现场环境,极易导致视频图像中出现行人遮挡问题,尤其在客流量密集的高峰时段,遮挡现象尤为突出。因此,摒弃依赖于完整人体特征的传统目标检测方法,转而以头部作为主要检测目标,有效解决了因人群密集、相互遮挡而导致的客流精细化监控不足的问题。选取YOLOv5算法对基于旅客头部的目标识别结果进行客流统计。采用DCF算法对提取的头部特征进行轨迹跟踪,将跟踪问题转化为具有循环结构的岭回归问题,通过最小化最小二乘损失函数,利用大量的循环位移样本对滤波器进行在线学习,将学习得到的滤波器应用于目标区域,通过计算两者间的响应估计感兴趣区域的目标状态。具体而言,在视频的某一帧中,DCF基于已知的目标位置,使用目标多通道特征结合损失函数来学习滤波器,该损失函数为

$$L_{DCF} = \arg \min_F \frac{1}{2} \left\| \sum_{d=1}^D \mathbf{X}_{t,d} \otimes F_{t,d} - \mathbf{Y} \right\|_F^2 + \frac{\lambda}{2} \sum_{d=1}^D \left\| F_{t,d} \right\|_F^2 \quad (1)$$

式中: $\mathbf{X}_{t,d}$ 为第 t 帧中跟踪目标中心提取的第 d 个特征通道对应的 $M \times N$ 维矩阵,即向量化训练样本, $\mathbf{X}_{t,d} \in \mathbb{R}^{M \times N}$; $F_{t,d}$ 为待学习优化的对应滤波器, $F_{t,d} \in \mathbb{R}^{M \times N}$; D 为总的特征通道数; \mathbf{Y} 为高斯分布生成的标签, $\mathbf{Y} \in \mathbb{R}^{M \times N}$; \otimes 为卷积运算; λ 为滤波器的正则项参数。

客站细粒度客流通行特征如图3所示。针对进出站闸机组、检票闸机组、站台电梯口等关键区域,在监控视频中心处设置2条横向跨越检测线(a线和b线),并规定每条检测线对应的动作穿越方向。采用YOLO+DCF算法获取每位乘客的行进轨迹,并依据其穿越检测线的先后顺序,判定人员在视频中的位移大小和方向,进而生成任意时间

跨度内旅客通行数量 Q 、通行速度 μ 等细粒度特征。 μ 的计算式为

$$\mu = \frac{Q}{\Delta t} \quad (2)$$

式中: Δt 为统计时间。

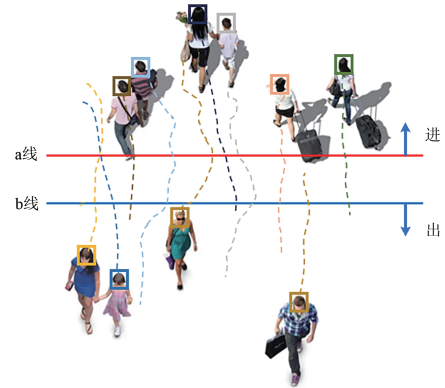


图3 客站细粒度客流通行特征

1.2 基于自适应CLTR的客流粗粒度分析

客流粗粒度分析是1种宏观层面的客流描述,旨在刻画区域内的客流动态分布规律与整体变化趋势。研究基于改进的CLTR^[22]人群动态感知网络,分析候车厅、站台区域、出站厅3个区域的人群分布状态、密度变化、流动模式等特征。

CLTR人群动态感知网络结构如图4所示。该网络基于回归思想,将人群定位任务建模为集合预测问题,无须额外的前处理与后处理操作。网络结构主要包含卷积神经网络(CNN)、Transformer编码器与解码器以及 K 近邻匹配算法,可实现密集场景下的人群分布实时检测。其中 K 近邻匹配算法在训练阶段对旅客头部的预测点与真实标注点进行匹配,未成功匹配的预测点被判定为背景点,从而有效利用上下文信息作为补充。

CLTR在大规模通用密集人群数据集上的预训练模型,难以直接适配铁路客站这一垂直场景。客站内人群密度分布不均,为适应不同密集程度下的人群定位与计数任务,提升模型在客站场景的泛化能力, K 近邻匹配算法中的 k 值改进为自适应调整

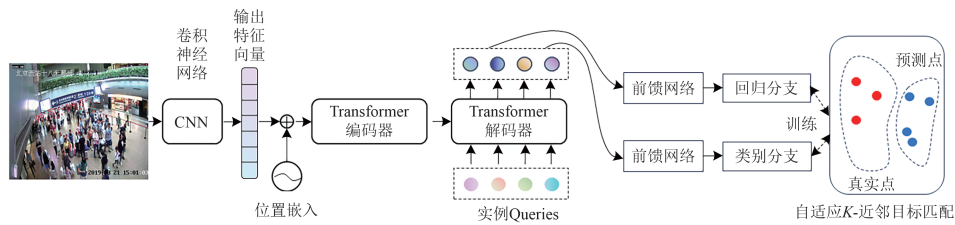


图4 CLTR人群动态感知网络结构

方式，可根据局部人群密度动态选取近邻点数量，具体调整策略如下。

1) 局部密度估计

对每个真实点 y_i ，计算其 m 个最近邻距离的平均值 $d_{i,m}$ (m 为预设的小整数)，以此快速估计局部密度。密度越高， $d_{i,m}$ 取值越小。

2) 全局密度归一化

计算训练集全局平均最近邻距离，并将其作为基准用于标准化不同数据集间的密度范围差异。

3) 自适应 k 值

根据局部距离与全局基准的比例动态调整 k 值，通过剪裁函数控制其范围，防止极端情况下的不合理值。

$$k_i = c \left(\left\lfloor \frac{d_{\text{global}}}{d_{i,m}} k_{\text{base}} \right\rfloor, k_{\text{min}}, k_{\text{max}} \right) \quad (3)$$

式中： k_i 为每个真实点 y_i 对应的 k 值； k_{base} 为基准值，初始取值为 4； d_{global} 为全局平均最近邻距离，即所有训练图像中真实点 $d_{i,m}$ 的平均值； $c(\cdot)$ 为剪裁函数，将 x 限制在 $[a, b]$ 内，防止 k_i 值过小或过大； $\lfloor \cdot \rfloor$ 为向下取整运算，确保 k_i 为整数； k_{min} 和 k_{max} 为 k_i 取值的最小值和最大值， k_{min} 取值为 2， k_{max} 取值为 8。

在训练前预计算 d_{global} 和 $d_{i,m}$ ，避免训练时重复执行复杂运算。由此实现密度自适应， k 值随局部密度动态调整，密集区域采用更大 k 值捕捉广泛上下文，稀疏区域采用较小 k 值抑制噪声。

训练所得自适应 CLTR 模型，可提取区域内排队长度和人群密度等粗粒度客流特征，此类特征是描述区域拥挤程度最直观的指标。当车站内旅客较少时，客流处于自由流状态，其客流密度较低；当车站发生大客流时，客流处于拥堵流状态，客流量和客流密度较大。其中，排队长度指一定统计时间内区域内的聚集人数，可由上述 CLTR 模型直接解析得到，人群密度 K 指统计时间内区域聚集人数 N 与有效聚集面积 S 的比值，其计算式为

$$K = \frac{N}{S} \quad (4)$$

进一步引入排队客流的通行时间以表征客流通行的延误程度，侧面反映客流的通行状态。排队时间 t 的计算式为

$$t = \frac{L}{v} \quad (5)$$

式中： L 为卡口前方的排队长度； v 为当前卡口的实时通行速度。

1.3 客流安全评估

大型铁路客站具有关键卡口多、通行距离长、候车面积广等特点，属于典型复杂系统。一旦站内某节点发生客流风险，易在站内传播并引发连锁反应，对运营安全构成严重威胁，因此，辨识和评估客流风险对于客站安全管理十分重要。铁路客站旅客流线如图 5 所示。根据旅客的走行目的，可将车站设施对旅客的服务流线分为进站流线、出站流线及换乘流线。客流风险关键节点包括：闸机组区域（到达率超过离开率致堆积）、候车厅/站台（流线交织致冲突）、楼扶梯区域（宽窄过渡致通行能力不匹配），均易引发拥堵。

周继彪^[23]和彭羽飞^[24]等提出城轨交通客流

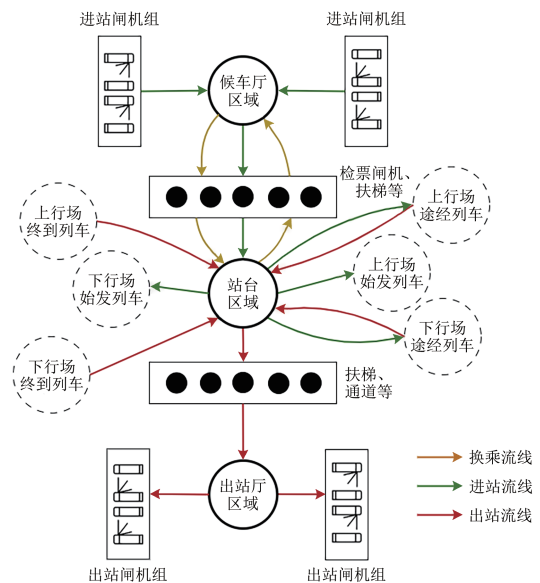


图5 铁路客站旅客流线

的评估因素包含客流拥挤程度、拥挤持续时间等。然而在构建铁路客站客流安全等级时，还需考虑关键区域的客流通行情况。当车站人行设施出现客流拥挤、速度降低、走行受限、短时客流剧增、密度增大且拥挤持续时间较长时，将提升区域客流风险，极易引发安全事故。为此，所提客流安全评估模型综合考虑客流的细粒度与粗粒度特征，将铁路客站客流安全评估转化为图模型^[25]的辨识问题。铁路客站可抽象为有向图 $G(V, E, A)$ ，其中： V 为节点集（ $|V|=N_{\text{key}}$ ），由车站中 N_{key} 个关键区域组成； E 为节点的边集； A 为邻接矩阵， $A \in \mathbb{R}^{N_{\text{key}} \times N_{\text{key}}}$ ，表征节点间的空间连接关系。图 G 中每个节点包含 P 个特征，每个节点在 t 时刻的特征向量为 $\mathbf{x}_{i,t}$ （ $i=1, 2, \dots, N_{\text{key}}$ ）， $\mathbf{x}_{i,t} \in \mathbb{R}^P$ 。定义 t 时刻客站图中全部节点的特征向量 $\mathbf{X}_t = (\mathbf{x}_{1,t}, \mathbf{x}_{2,t}, \dots, \mathbf{x}_{N_{\text{key}},t})^T$ ， $\mathbf{X}_t \in \mathbb{R}^{N_{\text{key}} \times P}$ ，则一定时间段 τ 内车站的安全风险为

$$y = f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\tau) \quad (6)$$

式中： f 为基于图结构的安全风险评估函数。

多注意力时空图卷积网络（Multi-Attention integrated Spatial-Temporal Graph Convolutional Network, MASTGCN）架构如图6所示。该网络包含2个分支：时空融合提取模块和时序特征学习模块。在时空融合提取模块中，多个时空图卷积块（ST blocks）串联构成，通过时空图卷积模块提取不同时刻车站关键区域的客流特征，并建模空间拓扑与时间演化的关联关系，辅以时空注意力机制实现特征的自适应加权与筛选；时序特征学习模块由多个时间卷积块连接而成，以节假日、列车时刻表数据作为输入，引入注意力机制探究不同时间维度因素对客流风险等级评估的影响。为提升训练稳

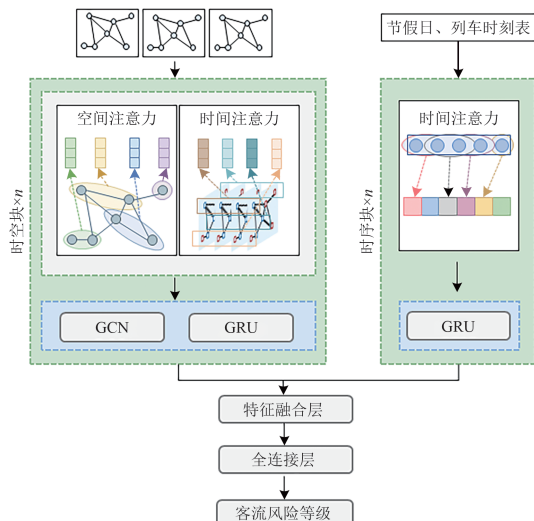


图6 MASTGCN架构

定性并避免梯度消失，2个分支均采用残差学习架构，通过特征融合层和全连接层融合两分支特征，输出车站客流安全风险的最终评估结果。

时空融合提取模块以客站关键区域为节点，输入特征包括排队长度 L 、排队通行时间 t 、通行速度 v 、旅客通行数量 Q 和区域人群密度 K 。这些特征在时空维度上呈现显著异质性：其空间分布随时间动态演变，时间序列特征又随空间位置差异而变化，这种双重变异性充分体现了数据内在的时空耦合特性。

采用图卷积网络（Graph Convolutional Network, GCN）进行空间特征提取，并利用门控循环单元（Gated Recurrent Unit, GRU）从GCN输出的特征矩阵中学习时序依赖关系，从而实现对客流时空特征的全面建模。考虑到不同时间步和区域提取的特征对客流风险评估的贡献度存在差异，因此，引入时间和空间注意力机制进行特征权重分配，能够自适应地评估和捕捉时空中的关键信息。空间注意力矩阵 G 为

$$\begin{cases} G = V_G \cdot \sigma\left(\left(\mathbf{X}_{h,l-1} \mathbf{W}_1\right) \mathbf{W}_2 \left(\mathbf{W}_3 \mathbf{X}_{h,l-1}\right)^T + \mathbf{b}_G\right) \\ G'_{i,j} = \frac{\exp(G_{i,j})}{\sum_{j=1}^{N_{\text{node}}} \exp(G_{i,j})} \end{cases} \quad (7)$$

式中： $\mathbf{X}_{h,l-1}$ 为第 l 层时空块的输入； V_G 、 \mathbf{b}_G 和 \mathbf{W}_i 为可学习的参数， $i=1, 2, 3$ ； $\sigma(\cdot)$ 为激活函数； $G_{i,j}$ 为节点 i 和节点 j 间的关联强度； N_{node} 为节点数量。

时间注意力矩阵 M 为

$$\begin{cases} \mathbf{M} = V_m \cdot \sigma\left(\left(\mathbf{X}_{h,l-1}\right)^T \mathbf{U}_1\right) \mathbf{U}_2 \left(\mathbf{U}_3 \mathbf{X}_{h,l-1}\right) + \mathbf{b}_m \\ M'_{i,j} = \frac{\exp(M_{i,j})}{\sum_{j=1}^{T_{l-1}} \exp(M_{i,j})} \end{cases} \quad (8)$$

式中： V_m 、 \mathbf{b}_m 、 \mathbf{U}_q 为可学习的参数， $q=1, 2, 3$ ； $M_{i,j}$ 为时间 i 和时间 j 间的依赖强度； T_{l-1} 为第 l 层时间维度的长度。

通过图卷积神经网络提取节点集 V 的空间结构特征，GCN中第 l 层至第 $l+1$ 层的传播式为

$$\mathbf{H}_{l+1} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}_l \mathbf{W}_l\right) \quad (9)$$

其中，

$$\tilde{A} = I + A$$

$$\tilde{D} = \text{diag}(\tilde{D}_{i,i}) = \text{diag}\left(\sum_j \tilde{A}_{i,j}\right)$$

式中： I 为单位矩阵； D 为对角阵， W_l 为 l 层的可学习参数； H_l 为第 l 层的输入； H_{l+1} 为第 l 层的输出； \tilde{A} 为自环邻接矩阵； $\tilde{A}_{i,j}$ 为 \tilde{A} 中第 i 行第 j 列元素。

时空融合模块中时序特征的提取选用GRU，以规避传统RNN易出现的梯度消失与梯度爆炸问题，同时保障模型推理效率。GRU的核心作用是对经图神经网络拓扑聚合后的序列特征进行时间建模，提取其时间特征^[26]。GRU的计算式为

$$H_t = (1 - z_t) \odot H_{t-1} + z_t \odot \tilde{H}_t \quad (10)$$

其中，

$$\tilde{H}_t = \tanh(X_t W_h + (r_t \odot H_{t-1}) U_h + b_h)$$

$$z_t = \sigma(X_t W_z + H_{t-1} U_z + b_z)$$

$$r_t = \sigma(X_t W_r + H_{t-1} U_r + b_r)$$

式中： z_t 为更新门； r_t 为重置门； H_t 为当前时刻空间GCN的输出； \tilde{H}_t 为候选隐状态； H_{t-1} 为上一时刻的隐藏状态； \odot 为Hadamard积运算符。

时序特征提取模块同样采用GRU，以提取节假日和列车时刻表中的时间特征，探究其对客流风险评估的影响。数据编码方式为：若当日为节假日编码为1，否则为0；在列车开检至离站时段内编

码为1，其余时段为0。

MASTGCN模型的输出阶段采用特征融合层对2个关键模块的输出进行整合，将时空特征融合模块提取的时空交互特征与时序特征学习模块捕获的动态演化特征进行张量拼接，经降维操作转换为统一的一维特征表示，再通过全连接层实现特征空间的非线性映射，输出客流风险值。基于此，定义车站客流的安全风险等级： $0 \leq y \leq 0.2$ 为1级； $0.2 < y \leq 0.4$ 为2级； $0.4 < y \leq 0.6$ 为3级； $0.6 < y \leq 0.8$ 为4级； $0.8 < y \leq 1$ 为5级。

2 试验分析

2.1 PFD数据集

构建铁路客运站客流数据集（Passenger Flows Dataset, PFD），该数据集由公开数据集与铁路客运站客流数据集组成，共包含3 000余张图像。其中，公开数据集由JHU-Crowd++的1 000张及NWPU的1 000张构成；铁路客运站客流数据采集自北京西站、北京朝阳站、上海虹桥站等客站，覆盖进站、检票、候车等多个场景，共1 000余张视频帧。所有视频帧均对人员头部信息进行精准标注，用于训练YOLO的头部检测模型和CLTR模型。PFD数据集示例如图7所示。



图7 PFD数据集示例

2.2 评价指标

为证明所提模型在车站客流特征提取和评估中的性能，采用多类指标对各模型进行综合评价。精确率指模型预测为正类的样本中真正例所占比例；召回率指真实正类样本中被模型正确预测为正类的比例；平均精度 A_p 通过计算“精确率-召回率”曲

线下的面积，综合反映模型在该类别上的性能； F_1 分数为精确率和召回率的调和平均数，用于综合评估分类模型的整体性能。精确率 P 、召回率 R 、平均精度 A_p 和 F_1 分数的计算式为

$$P = \frac{T_p}{T_p + F_p} \quad (11)$$

$$R = \frac{T_P}{T_P + F_N} \quad (12)$$

$$A_p = \int_0^1 P(R) dR \quad (13)$$

$$F_1 = \frac{2PR}{P+R} \quad (14)$$

式中： T_P 为正确检测出正样本的数量； F_P 为将负样本预测为正样本的数量； F_N 为将正样本预测为负样本的数量。

帧率（Frames Per Second, FPS）是衡量模型处理速度的核心指标，表示模型每秒可处理的图像帧数，其值越高表明模型实时性越好。平均绝对误差（Mean Absolute Error, MAE）为预测值与真实值间绝对误差的平均值，用于表征预测的整体偏差水平。均方误差（Mean Squared Error, MSE）为预测值与真实值间平方误差的平均值，可放大较大误差的影响，对异常值更为敏感；均方根误差（Root Mean Squared Error, RMSE）为均方误差的平方根，同样适用于回归模型评价，既保留了MSE对大误差的放大特性，又因单位与原始数据一致更便于直观解释。

2.3 细粒度客流监测算法结果

1) 旅客头部检测算法

选取YOLOv5进行旅客头部检测，在PFD数据集上进行模型预训练，硬件为RTX6000，显存为48 G，基于Pytorch深度学习框架实现。为保证对比试验的公平性，训练不同尺寸模型的超参数保持一致，设置迭代次数为200，置信度阈值为0.5。在同一测试数据集下，YOLOv5系列模型性能对比见表1。由表1可知，模型参数越大，检测精度越高，但检测速度会降低。YOLOv5s模型的综合性能最为均衡，因此现场选用其作为基础模型，在保证检测效果的同时保持处理速度，可接入更多摄像头，保证细粒度分析的实时性。

表1 YOLOv5系列模型性能对比

模型名称	参数量/M	平均精度/%	召回率/%	FPS
YOLOv5s	7.3	85.7	86.1	150
YOLOv5m	20.6	87.3	87.6	85
YOLOv5l	46.3	89.4	90.3	45

检测算法面向铁路客站垂直场景，仅针对视频中旅客头部目标进行识别。为验证所选模型的合理性，将YOLOv5s与先进YOLO算法（如YOLOv9s、YOLOv10s）在PFD数据集上进行对比测试，所得结果见表2。由表2可知，在铁路客站专

用数据集上，各模型检测精度差距较小，而YOLOv5s在处理速度上表现更优。考虑到客流安全评估需在候车层（45个摄像头）、站台层（56个摄像头）、出站口（10个摄像头）等关键区域部署设备，且所有摄像头分辨率均为1 920像素×1 080像素，需在保证检测精度的同时满足实时统计分析需求。权衡综合性能与速度，YOLOv5s是实现精度与实时性均衡的最优选择。

表2 不同YOLO算法在客流数据集的测试对比

模型名称	平均精度/%	单图推理时间/ms
YOLOv5s	85.7	1.92
YOLOv9s	87.8	3.54
YOLOv10s	87.6	2.66

2) 人员跟踪性能分析

基于YOLOv5s的人员定位结果，采用DCF算法跟踪目标。为验证DCF在铁路客运站场景下的有效性，选取另外2个主流的跟踪模型DeepSORT与IOU跟踪器进行对比试验，统计结果见表3。由表3可知，YOLOv5s+DeepSORT组合的客流统计精确率最高，达99.6%，但DCF在统计精度与实时性能的权衡上更贴合铁路客站多摄像头、高分辨率的实际应用场景，可在保证较高统计精确率的同时满足现场实时分析需求。

表3 精细化客流统计结果

算法名称	算法统计值/人	客流真实值/人	精确率/%
YOLOv5s+DeepSORT	282	283	99.6
YOLOv5s+IOU	269	283	95.1
YOLOv5s+DCF	281	283	98.3

3) 实际场景下的精细化客流数据

选取南昌东站作为试点站，以检票口和出站口客流通行量为核验对象。基于视频分析技术实时计算关键区域的通行人数，并与该站官方购票数据进行对比验证，结果如图8所示。经统计并计算可得视频分析结果与购票真实数据的累计误差为6.9%，验证了细粒度客流表征在实际客站场景下的统计准确性与可靠性。

2.4 粗粒度客流监测算法结果

本试验在PFD数据集上进行训练和测试，具体划分为训练集1 930张，验证集300张，测试集900张。采用ResNet50作为主干网络，Transformer架构的编码器与解码器层数均设置为6层，实例查询集的数量设定为500。数据集增强策略包含随机

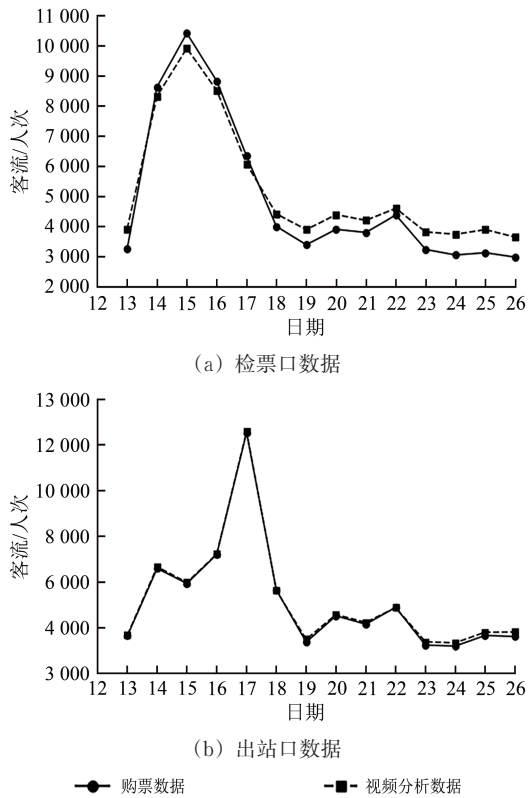


图 8 算法结果对比

裁剪、随机缩放和水平翻转（翻转概率为 0.5），输入图像尺寸统一为 256 像素 × 256 像素，经无重叠裁剪后，不足部分用 0 填充。设置学习率为 0.000 1、批处理大小为 16，参数 k 在 2~8 之间自适应选择。人群定位与计数性能对比见表 4。由表 4 可知：相较于基础 CLTR 模型，改进 CLTR 模型在各项指标上均取得显著优势，在计数任务中，其精确率、召回率及 F_1 分数分别提升 4.2%，5.3% 和 4.8%；在计数任务中，其 MAE 和 MSE 分别下降 10.4% 和 17.8%。

表 4 人群定位与计数性能对比

模型名称	精确率/%	召回率/%	F_1 分数/%	MAE	MSE
基础 CLTR 模型	84.9	82.2	83.5	68.3	280.5
改进 CLTR 模型	89.1	87.5	88.3	61.2	230.6

基于自适应的 CLTR 模型在不同场景下的粗粒度客流可视化结果如图 9 所示。该图展示了检票口、站台区、候车厅 3 个不同场景下提取粗粒度特征的人群数量，图中：黑色点表示解析出的旅客位置；点的数量表示区域中旅客的总数量。

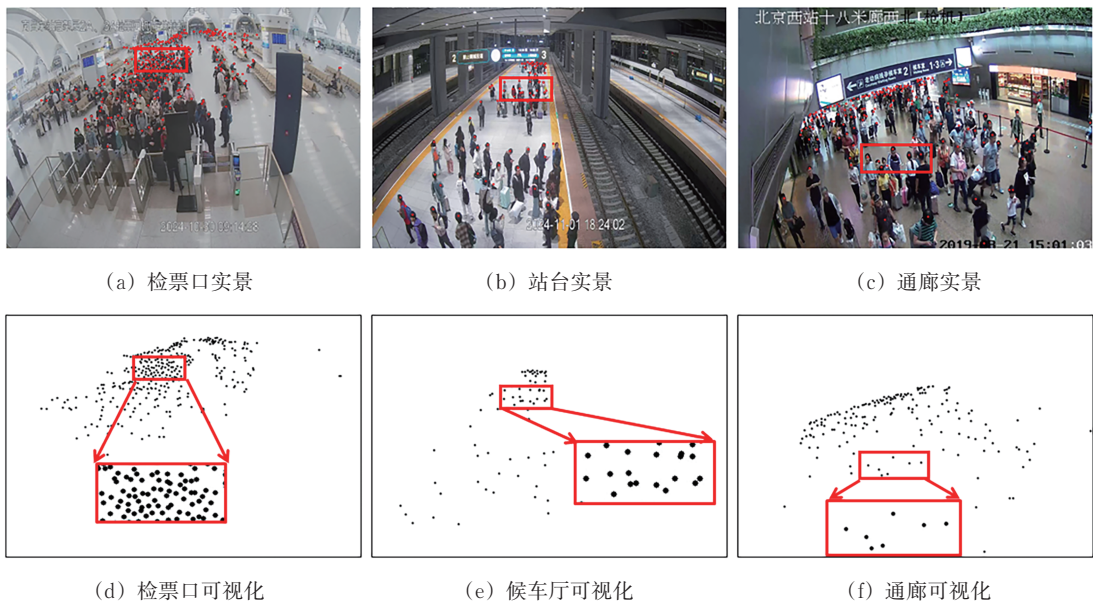


图 9 不同场景下的粗粒度客流实景及可视化结果

2.5 客站客流安全评估

通过多位专家打分的方式，将客流安全等级进行分级。基于粗细粒度客流特征数据，对南昌东站客流风险数据集进行扩充，得到 28 934 条高质量样本，数据项包含排队长度、排队通行时间、通行速度、旅客通行数量、区域人群密度、节假日属性及列车时刻表等。

模型训练以估计值与真实值之间的均方误差 (MSE) 作为损失函数，设训练阶段批处理大小为 64，学习率为 0.000 1。为验证所提模型的有效性，选取多种经典模型作为基准进行对比，包括历史平均 (HA) 模型、自回归移动平均 (ARIMA) 模型、长短时记忆 (LSTM) 网络、门控循环单元 (GRU)、时空图卷积网络 (STGCN) 以及基于多级注意力

的递归神经网络 (GeoMAN)。不同模型的性能对比见表6。

表6 不同模型的性能对比

模型名称	RMSE	MAE	R/%
HA	40.09	26.35	72.5
ARIMA	39.21	24.32	70.2
LSTM	34.56	21.63	81.1
GRU	33.83	20.17	80.3
STGCN	25.46	15.78	82.6
GeoMAN	21.73	15.04	83.8
MASTGCN	18.14	11.21	87.5

传统时间序列分析模型 (如 HA 和 ARIMA) 的预测误差显著偏高、召回率较低, 反映出这类方法在刻画客流数据非线性特征与复杂时空关联方面存在局限性; 考虑时间相关性与空间相关性的 STGCN, GeoMAN 及 MASTGCN 模型, 相较 LSTM, GRU 等仅聚焦时序特征的传统深度学习

模型表现更优; GeoMAN 的性能优于 STGCN, 验证了多级注意力机制在捕捉客流时空动态变化中的有效性; MASTGCN 在数据集上的各项评估指标均取得最优表现, 该网络在融合多注意力机制的基础上, 额外融入时序特征信息, 实现了预测误差的进一步降低, 充分证明其在客流预测任务中的优越性与创新性。

通过所提 MYDC 模型对南昌东站的客流进行评估, 整体客流风险等级如图 10 所示。由图 10 可知, 客流风险等级的变化趋势与客流规模呈显著正相关, 风险等级的 2 个极大值点出现在 15:30 及之后, 此时间段前后站内客流显著增长, 进站与出站客流叠加, 形成短时大客流; 客流叠加引发旅客排队进出站频次增加, 进一步放大了客流安全的不稳定因素, 待后续旅客逐步疏散、站内通行恢复稳定后, 风险等级随之回落, 整体态势维持在动态较低水平。

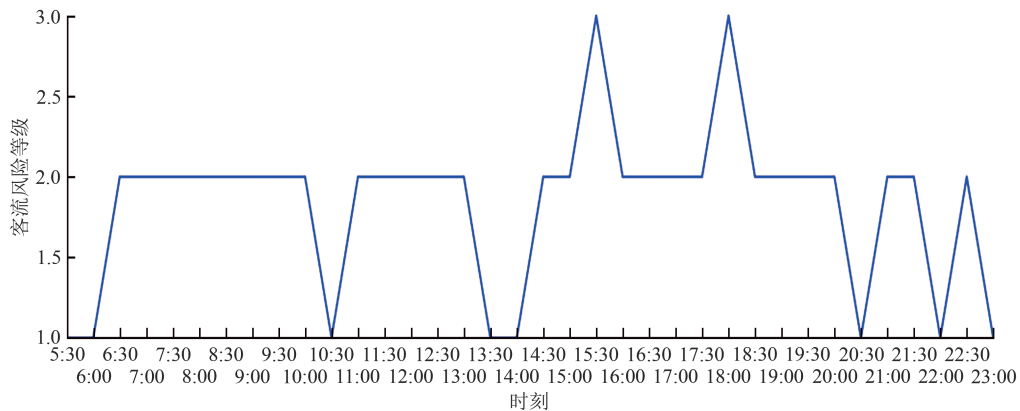


图10 南昌东站整体客流风险等级

3 结 论

本文紧密结合铁路客站内客流安全管控的实际需求, 为实现车站客流安全风险的及时量化与高效管控, 提出1种基于视频分析技术的客站多粒度动态客流评估 (MYDC) 模型。构建了铁路客运站客流数据集 (PFD), 并引入视频分析技术, 设计了粗、细双维度客流特征提取方案。其中, 细粒度特征提取采用 YOLO 与 DCF 技术, 实现了客流速度、流量等核心参数的实时精准计算; 粗粒度特征提取基于改进的 CLTR 模型, 有效完成了区域排队状态、人群空间分布等宏观特征的准确分析。在此基础上, 综合考虑车站区域功能差异与列车时刻表约束, 构建基于时空图卷积的客站客流安全评估

模型 MASTGCN。南昌东站的实际应用表明, 所提 MYDC 模型能够精准刻画现场实时客流的时空演化特征, 全面、客观地评估客流安全风险状态。

未来研究可从3个方面进一步深化与拓展: 一是扩大数据集覆盖范围, 纳入不同季节、节假日高峰及异常客流 (如突发聚集、紧急疏散) 等场景数据, 提升模型的泛化能力; 二是融合多源数据 (如旅客出行画像、设备运行状态等), 优化 MASTGCN 模型结构, 进一步提升客流安全风险评估的精准度与提前预警能力; 三是探索模型的轻量化部署方案, 适配车站边缘计算设备, 实现客流安全评估与管控指令的实时响应, 推动技术成果的规模化落地应用。

参 考 文 献

- [1] 熊豪. 基于AFC数据的城市轨道交通客流特征及加权网络抗毁性研究[D]. 西安: 长安大学, 2020.
(XIONG Hao. Research on Urban Rail Transit Passenger Flow Characteristics and Weighted Network Invulnerability Based on AFC Data [D]. Xi'an: Chang'an University, 2020. in Chinese)
- [2] 唐锟, 韦先霜, 王双保. 基于红外流量传感器的客流监测系统的设计[C]//第九届全国光电技术学术交流会议论文集(下册). 北京: 中国宇航学会光电技术专业委员会, 2010: 346-349.
(TANG Kun, WEI Xianshuang, WANG Shuangbao. Design of Supervisory Control System for Passenger Flow Volume Based on Infrared Flow Sensor [C]// The 9th National Conference on Photoelectronic Technology (Volume 2). Beijing: The Chinese Society of Astronautics Technical Committee on Optoelectronic Technology, 2010: 346-349. in Chinese)
- [3] KHAN S D. Congestion Detection in Pedestrian Crowds Using Oscillation in Motion Trajectories [J]. *Engineering Applications of Artificial Intelligence*, 2019, 85: 429-443.
- [4] FUKUZAKI Y, MOCHIZUKI M, MURAO K, et al. A Pedestrian Flow Analysis System Using Wi-Fi Packet Sensors to a Real Environment [C]// Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. Seattle: ACM, 2014: 721-730.
- [5] 郁琦晖, 杨海红, 李傲松. 基于多源监测数据融合的城市轨道交通客流监测系统研究[J]. *交通与港航*, 2021, 8(3): 68-72.
(YU Qihui, YANG Haihong, LI Aosong. Research on Urban Rail Transit Passenger Flow Monitoring System Based on Multi-Source Monitoring Data Fusion [J]. *Communication & Shipping*, 2021, 8(3): 68-72. in Chinese)
- [6] HARIYONO J, HOANG V D, JO K H. Motion Segmentation Using Optical Flow for Pedestrian Detection from Moving Vehicle [C]// *Computational Collective Intelligence. Technologies and Applications*. Cham: Springer, 2014: 204-213.
- [7] 王爱丽, 董宝田, 王泽胜, 等. 融合光流速度场与背景差分的自适应背景更新方法[J]. *中国铁道科学*, 2014, 35(6): 131-137.
(WANG Aili, DONG Baotian, WANG Zesheng, et al. Adaptive Background Updating Based on Optical Flow and Background Difference [J]. *China Railway Science*, 2014, 35(6): 131-137. in Chinese)
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2014: 580-587.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 779-788.
- [10] LIU Q L, GUO Q, WANG W, et al. An Automatic Detection Algorithm of Metro Passenger Boarding and Alighting Based on Deep Learning and Optical Flow [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-13.
- [11] ZHANG J, LIU J Z, WANG Z Z. Convolutional Neural Network for Crowd Counting on Metro Platforms [J]. *Symmetry*, 2021, 13(4): 703.
- [12] SIPETAS C, KEKLIKOGLOU A, GONZALES E J. Estimation of Left behind Subway Passengers through Archived Data and Video Image Processing [J]. *Transportation Research Part C: Emerging Technologies*, 2020, 118: 102727.
- [13] 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). New York: IEEE, 2017: 3645-3649.
- [14] IDREES H, TAYYAB M, ATHREY K, et al. Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds [C]// *Computer Vision ECCV 2018*. Cham: Springer, 2018: 544-559.
- [15] ABOUSAMRA S, HOAI M, SAMARAS D, et al. Localization in the Crowd with Topological Constraints [J]. *Proceedings of AAAI Conference on Artificial Intelligence*, 2021, 35(2): 872-881.
- [16] 李瑞, 李平, 代明睿, 等. 基于铁路客运站视频多视角特征图投影融合的人群密度估计模型[J]. *中国铁道科学*, 2022, 43(3): 182-192.
(LI Rui, LI Ping, DAI Mingrui, et al. Crowd Counting Estimation Model Based on Multi-View Projection Fusion of Railway Passenger Station Video [J]. *China Railway Science*, 2022, 43(3): 182-192. in Chinese)

- [17] 豆飞,潘晓军,秦勇,等.基于云模型的城市轨道交通车站客流控制触发判别方法[J].东南大学学报(自然科学版),2016,46(6):1318-1322.
(DOU Fei, PAN Xiaojun, QIN Yong, et al. Identification Method for Passenger Inflow Control in Urban Rail Transit Station Based on Cloud Model [J]. Journal of Southeast University (Natural Science Edition), 2016, 46 (6): 1318-1322. in Chinese)
- [18] 赵保锋,邹晓磊,屈晓宜.基于仿真的城市轨道交通站台客流滞留分级预警方法[J].城市轨道交通研究,2017,20(9):107-110,115.
(ZHAO Baofeng, ZOU Xiaolei, QU Xiaoyi, et al. Classification of Early Warning for Passenger Flow Retention at Urban Rail Transit Station Based on Simulation [J]. Urban Mass Transit, 2017, 20 (9): 107-110, 115. in Chinese)
- [19] 李得伟,尹浩东.基于物联网设备检测数据的城市轨道交通车站站台多维度实时客流密集度指数预测方法研究[J].铁道学报,2014,36(3):9-13.
(LI Dewei, YIN Haodong. Real-Time Forecast of Passenger Crowd Index on Urban Rail Transit Station Platform Based on Test Data of Internet of Things [J]. Journal of the China Railway Society, 2014, 36 (3): 9-13. in Chinese)
- [20] 胡国林.城市轨道交通线网客流风险评估方法及系统研究[D].北京:北京交通大学,2022.
(HU Guolin. Research on Passenger Flow Risk Assessment Method and System of Urban Rail Transit Network [D]. Beijing: Beijing Jiaotong University, 2022. in Chinese)
- [21] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 583-596.
- [22] LIANG D, XU W, BAI X. An End-to-End Transformer Model for Crowd Localization [C]// Computer Vision-ECCV 2022. Cham: Springer, 2022: 38-54.
- [23] 周继彪,赵鹏飞,董升,等.基于蚁群算法的地铁车站行人拥挤等级划分方法[J].城市交通,2019,17(4):105-113.
(ZHOU Jibiao, ZHAO Pengfei, DONG Sheng, et al. Pedestrian Congestion Levels at Subway Stations with Ant Colony Algorithm [J]. Urban Transport of China, 2019, 17 (4): 105-113. in Chinese)
- [24] 彭羽飞.城市轨道交通客流实时感知估计研究[D].北京:北京交通大学,2023.
(PENG Yufei. Research on Real-time Perception Estimation of Urban Rail Transit Passenger Flow [D]. Beijing: Beijing Jiaotong University, 2023. in Chinese)
- [25] GUO S, LIN Y, FENG N, et al. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33 (1): 922-929.
- [26] CHUNG J, GÜLÇEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [J]. ArXiv e-Prints, 2014: arXiv:1412.3555 [cs. CV].

Multi-Granularity Yardstick for Dynamic Crowds Model for Railway Passenger Stations Based on Video Analysis Technology

LIU Yuxin

(Institute of Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

Abstract: With the rapid development of the railway industry and the continuous increase of passenger transport tasks, railway passenger stations are facing increasingly severe passenger flow safety issues. To realize real-time monitoring of passenger flow dynamics and finely analyze the multi-granularity characteristics of passenger flow, a Multi-granularity Yardstick for Dynamic Crowds (MYDC) model for railway passenger stations based on video analysis technology is proposed. Firstly, a passenger flow dataset for railway passenger stations is constructed. Secondly, a fine-grained feature perception network for passenger flow is designed based on

YOLO and Discriminative Correlation Filter (DCF) tracking algorithm, and the adaptive crowd localization Transformer (CLTR) model for railway passenger stations is improved to capture the coarse-grained features of the overall passenger flow distribution. Finally, based on the physical attributes of passenger flow as well as its micro and macro characteristics, a Multi-Attention Spatio-Temporal Graph Convolutional Network (MASTGCN) is constructed to mine the spatio-temporal dynamic trends of passenger flow and assess the safety risk level of passenger flow in the station. The results show that the cumulative error of fine-grained feature extraction is 6.9%, the recognition accuracy of coarse-grained features is 89.1%, and the recall rate of the passenger flow safety assessment model is 87.5%. The proposed model can provide accurate data support for passenger flow management and has strong engineering application value.

Key words: Crowd perception; Video analysis; Railway passenger station; Security assessment; Transformer model; Spatio-Temporal Graph Convolutional Network (STGCN)

(责任编辑 杨婧婕)