Computer Science and Application 计算机科学与应用, **2026, 16(7), 226-241**
Published Online July 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.167254

Hans汉斯

基于跨摄像头的车辆在线跟踪关键技术研究

黄懿鑫

温州大学计算机与人工智能学院，浙江 温州

收稿日期：2026年6月25日；录用日期：2026年7月22日；发布日期：2026年7月29日

摘　要

多摄像头多目标车辆跟踪(Multi-Target Multi-Camera Vehicle Tracking, MTMCT)由于车辆外观变化、视角差异以及跨摄像头域分布偏移等问题，仍然面临较大挑战。车辆重识别(Vehicle re-identification, ReID)作为跨摄像头轨迹关联的关键技术，在现有数据集约束下，现有方法主要依赖外观特征建模，导致模型鲁棒性与跨域泛化能力受限。针对上述问题，本文构建了大规模多维标注综合数据集VeCamX，通过融合VeRi-776丰富的语义属性信息与AIC22复杂的时空轨迹特征，实现车辆身份、属性、轨迹及姿态关键点等多层次标注统一，为在线车辆跟踪提供综合性评测基准。在此基础上，提出一种统一优化框架，融合姿态感知模块(Pose-Aware Module, PAM)、元数据辅助模块(Metadata-Assisted Module, MAM)以及搜索剪枝优化策略(Search-and-Pruning, SnP)。其中，PAM用于学习车辆目标的视角不变结构特征；MAM整合车辆类型、品牌、颜色以及时空轨迹等元数据信息；SnP通过构建均衡且高判别性的训练样本集提升模型泛化能力。在VeRi-776、AIC22及VeCamX数据集上的实验结果表明，所提出框架能够稳定提升IDF1 和MOTA 等评价指标，相较现有先进方法表现出更强的鲁棒性与跨场景泛化能力。

关键词

多摄像头多目标跟踪，车辆重识别，VeCamX数据集，姿态感知，元数据辅助

文章引用: 黄懿鑫. 基于跨摄像头的车辆在线跟踪关键技术研究[J]. 计算机科学与应用, 2026, 16(7): 226-241.
DOI: 10.12677/csa.2026.167254

DSESHADE: Research on Key Technologies for Vehicle Online Tracking Based on Cross-Camera Association

Yixin Huang

School of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

Received: June 25, 2026; accepted: July 22, 2026; published: July 29, 2026

Abstract

Multi-Target Multi-Camera vehicle Tracking (MTMCT) remains challenging due to severe appearance variations, viewpoint shifts, and domain discrepancies among cameras. Notably, vehicle re-identification (ReID) plays a crucial role in trajectory association for MTMCT, yet constrained by the limitations of existing datasets, current ReID-based approaches rely mainly on appearance cues, which compromises their robustness and cross-domain generalization. To address these issues, we introduce a large-scale, multi-dimensionally annotated dataset named VeCamX, which integrates the semantic richness of VeRi-776 and the spatio-temporal complexity of AIC22. Specifically, VeCamX unifies annotations for vehicle identities, attributes, trajectories, and poses keypoints, thereby providing a comprehensive benchmark for online vehicle tracking.Building upon VeCamX's multi-dimensional annotations, we propose a unified optimization framework that combines Pose-Aware Module (PAM), Metadata-Assisted Module (MAM), and Search-and-Pruning optimization strategy (SnP). Specifically, PAM captures viewpoint-invariant structural features of vehicle targets; MAM acquires target-specific metadata including vehicle type, brand, color, and spatio-temporal trajectories; and SnP supplies balanced and highly discriminative data inputs for model training, thereby enhancing the model's generalization performance. Experiments on VeRi-776, AIC22, and VeCamX show that the proposed framework consistently improves IDF1 and MOTA over state-of-the-art methods, achieving strong robustness and generalization in complex real-world scenarios.

Keywords

Multi-Target Multi-Camera Vehicle Tracking, ReID, VeCamX Dataset, Pose-Aware, Metadata-Assisted

1. 引言

随着智能交通的发展，多摄像头多目标车辆跟踪(MTMCT)成为智慧交通与自动驾驶中的关键任务 [1]。该任务旨在跨摄像头持续跟踪同一车辆并保持身份一致性 [2]，通常包括单摄像头目标检测与轨迹生成，以及跨摄像头轨迹关联两部分 [3] [4]。其中，车辆重识别(CVehicle ReID)通过外观特征建模，在跨摄像头匹配中起关键作用 [5]。

然而，现有方法仍受数据与建模能力限制。VeRi-776 [6]场景较简单，缺乏复杂交通多样性；AIC22 [7]虽规模更大，但缺少语义属性与结构标注，限制多模态建模能力。同时，不同数据集之间存在标注与分布差异，进一步加剧跨域泛化问题 [8] [9]，影响跨摄像头身份关联稳定性 [4]。

为此，本文提出一种面向MTMCT的多模态协同框架，通过引入姿态、语义与时空信息提升跨摄像头关联能力。具体包括：姿态感知模块(PAM)C利用关键点与结构信息增强视角鲁棒性 [10]；元数据辅助模块(CMAM)融合车辆属性及时空信息提升判别能力 [11]；搜索剪枝策略(SnP)通过样本筛选提升数据分布质量 [12]。三者协同实现语义、结构与时空信息融合建模 [13] [14]。

本文贡献如下：(1) 提出多模态协同MTMCT框架，实现跨摄像头稳定身份关联；(2) 在VeRi-776与AIC22上验证方法有效性，达到先进性能(SOTA)。

本文结构如下：第2节介绍相关工作，第3节介绍方法，第4节给出实验，第5节总结全文。

2. 相关工作

2.1. 跨摄像头多目标跟踪(MTMCT)

跨摄像头多目标跟踪(CMTMCT)是智能交通与公共安全领域的重要研究问题 [4]。早期方法多采用离线框架，通过单摄像头轨迹生成与图匹配/网络流模型实现跨摄像头身份关联 [15] [16]，通常具有较高精度但难以满足实时性需求。

近年来，在线MTMCT方法成为主流，强调逐帧关联与实时更新。例如，通过外观特征与IoU进行级联匹配，并结合轨迹恢复策略缓解遮挡导致的断裂问题 [17]。此外，部分方法引入图神经网络或Transformer建模长程时空依赖关系，以提升跨摄像头一致性 [18]。近期也有研究探索轻量化检测驱动框架以兼顾效率与性能 [19]。然而，由于跨视角外观差异显著且车辆外观相似性较高，身份切换与轨迹断裂问题仍然普遍存在。

2.2. 车辆重识别数据集

数据集是推动MTMCT发展的关键因素。VeRi-776是经典车辆ReID数据集，提供车辆身份及颜色、车型、品牌等属性标注，并包含摄像头与时间信息 [20]。AIC22数据集进一步扩展到大规模真实交通场景，提供密集跨摄像头轨迹与时间戳信息，但缺乏细粒度语义属性 [7]。CityFlow-ReID则增强了多摄像头覆盖与复杂场景评估能力 [16]。

总体来看，现有数据集通常在语义属性丰富性与真实场景复杂性之间存在侧重差异，尚未形成统一且全面的评测基准 [5]。

3. 多模态协同优化方法

3.1. VeCamX数据集构建与统计

现有车辆ReID与MTMCT数据集通常针对不同任务需求进行构建。VeRi-776主要提供车辆身份标签与细粒度语义属性信息，而AIC22侧重复杂交通场景下的大规模跨摄像头轨迹关联，两者在标注内容与任务覆盖范围方面存在明显差异。

为支持多模态协同车辆跟踪研究，本文构建统一多模态车辆跟踪数据集VeCamX。该数据集基于VeRi-776与AIC22进行融合扩展，在统一车辆身份体系的基础上，同时整合身份标签、轨迹信息、语义属性、时空元数据以及车辆姿态标注。

在数据构建过程中，首先对不同来源数据进行格式标准化与ID统一映射，建立跨数据源一致的身份编号体系；随后，对车辆类型、颜色及品牌等属性标注进行规范化整理，并保留CameraID、Timestamp及轨迹关联信息，以支持时空建模任务。进一步地，本文设计8关键点车辆姿态标注方案，对车灯、车轮等结构稳定区域进行关键点标注，其中部分样本采用人工精细标注，其余样本利用训练后的关键点检测模型进行自动生成。

表 1给出了VeCamX与现有公开车辆数据集的对比结果。相比VeRi-776与AIC22，VeCamX在统一保留身份监督、轨迹信息以及语义属性的基础上，进一步引入车辆关键点与时空元数据标注，从而为本文提出的PAM、MAM及SnP 等模块提供统一的数据支撑平台。

3.2. 总体框架

本文提出一种面向跨摄像头多目标车辆跟踪(MTMCT)的多模态协同优化框架，整体流程包括

Table 1. Comparison of vehicle reID and cross-camera multi-target tracking datasets
表 1. 车辆重识别与跨摄像头多目标跟踪数据集对比

数据集	车辆数	摄像头数	图像数	轨迹数	主要标注/特征
VeRi-776	776	20	50k+	9k+	细粒度元数据属性(颜色、车型、品牌、车牌)C、Camera ID、时间戳；轨迹规模有限。
AIC22	2000+	46	100k+	20k+	大规模真实交通监控场景；跨摄像头轨迹、Camera ID、帧级时间戳；缺乏元数据标注。
VeCamX (本文)	2700+	66	150k+	25k+	融合VeRi-776语义属性与AIC22跨摄像头轨迹；统一时间戳、场景标签与8关键点标注(20k+人工、130k+自动)，支持多模态学习。

单摄像头跟踪、多模态特征增强以及跨摄像头关联三个阶段，如图 1所示。

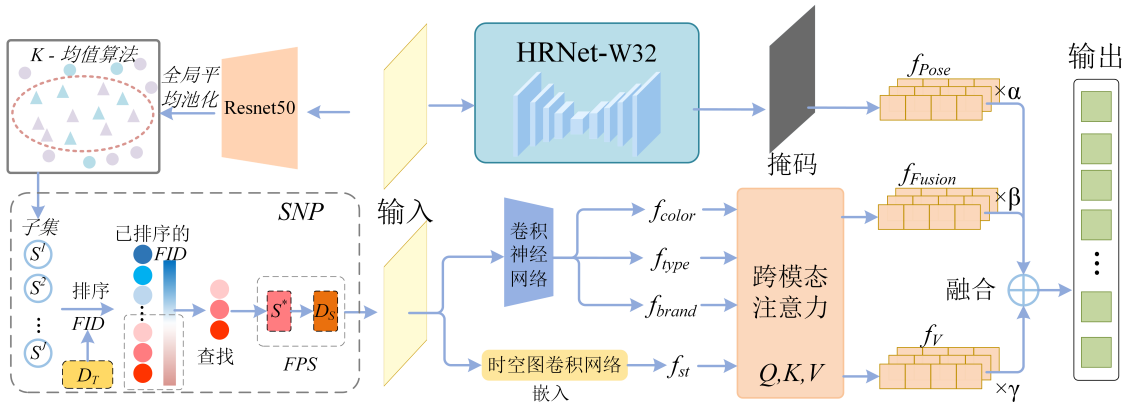

Figure 1. Overall framework of the proposed multimodal collaborative MTMCT method
图 1. 本文提出的多模态协同MTMCT总体框架

首先，各摄像头视频流独立执行目标检测与帧间关联，生成摄像头内部轨迹，为后续跨摄像头身份匹配提供基础输入。随后，在多模态特征增强阶段，分别利用姿态感知模块(Pose-Aware Module, PAM) 与元数据辅助模块(Metadata-Assisted Module, MAM)对车辆结构信息、语义属性以及时空先验进行建模，以增强特征表示的跨视角稳定性与身份判别能力。最后，对轨迹级表示执行全局相似度匹配与跨摄像头关联，将不同摄像头中的局部轨迹整合为统一身份的全局轨迹集合，实现跨摄像头多目标车辆跟踪。

3.3. 单摄像头跟踪模块

本文采用Fast Online MTMCT [2]作为单摄像头跟踪阶段的基础框架，对各摄像头视频流进行独立处理。

给定输入视频序列，首先利用YOLOv7-E6检测器 [21]在每一帧中生成车辆检测结果。每个检

测框定义为：

$$b_t = (x_t, y_t, w_t, h_t, s_t) \tag{1}$$

其中，x_t与y_t表示检测框中心坐标，w_t与h_t分别表示目标框宽度与高度，s_t表示检测置信度。

在目标检测完成后，采用与ByteTrack [22]类似的多阶段关联策略进行帧间目标匹配。该过程结合运动状态预测与检测结果关联，实现目标身份维护与轨迹更新。

对于同一车辆目标，其摄像头内轨迹表示为：

$$\tau_i = \{(b_t, t, c)\} \tag{2}$$

其中，b_t、t与c分别表示目标检测框、时间戳以及摄像头编号。

经过目标检测与单摄像头关联后，最终获得由多个摄像头构成的轨迹集合，并作为后续多模态特征增强与跨摄像头身份关联阶段的基础输入。

3.4. 姿态感知模块(PAM)

跨摄像头车辆跟踪中，同一车辆在不同摄像头下常伴随视角变化、局部遮挡及光照差异，仅依赖视觉外观特征容易导致身份表示不稳定。

为增强特征的跨视角一致性，本文提出姿态感知模块(CPose-Aware Module, PAM)，利用车辆关键点所包含的几何结构信息构建结构约束表示。如图 2所示，PAM主要包括关键点检测与结构引导特征融合两个部分。

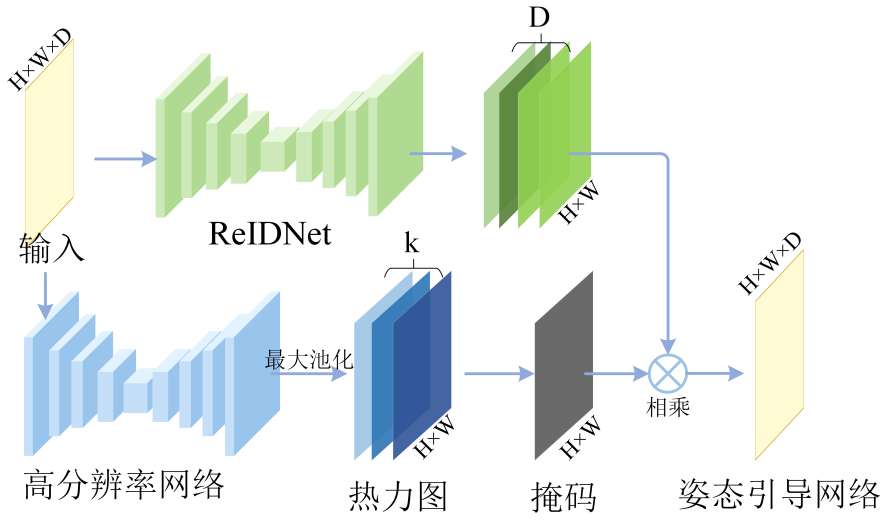

Figure 2. Architecture of the Pose-Aware Module (PAM)

图 2. 姿态感知模块(PAM)整体结构

3.4.1. 关键点检测

本文采用HRNet-W32 [23]作为车辆关键点检测网络，对车灯、车轮及车顶等具有较强结构稳定性的区域进行预测。网络输出对应关键点热力图，并利用VeCamX数据集标注的关键点信息生成二维高斯热力图作为监督信号，训练阶段采用均方误差损失进行优化。

上述关键点区域在不同摄像头视角之间通常保持相对稳定的空间关系，可为跨视角车辆表征提供可靠的几何先验。

3.4.2. 结构引导特征融合

在获得关键点预测结果后，首先对全部关键点热力图进行聚合生成统一结构掩码M，用于突出车辆稳定结构区域并抑制背景干扰。

随后，将结构掩码引入视觉特征表示中。设视觉主干网络提取的中间特征为F_v，则结构增强特征定义为：

$$f_{\text{pose}} = F_v \otimes M \tag{3}$$

其中，$\otimes$表示逐元素乘法操作。

通过结构引导融合，模型能够更加关注跨视角条件下保持稳定的几何区域，从而减弱背景噪声、局部遮挡及视角变化对身份表示学习的影响。最终生成的结构增强特征f_{pose}将作为后续多模态融合的重要输入。

3.5. 元数据辅助模块(MAM)

跨摄像头车辆跟踪中，仅依赖视觉外观特征容易受到视角变化、遮挡以及光照差异影响，而车辆语义属性与时空信息通常具有较强稳定性，可作为身份判别的重要补充信息。

为提升跨摄像头关联的鲁棒性与判别能力，本文提出元数据辅助模块(Metadata-Assisted Module, MAM)，通过融合车辆语义属性与时空先验构建多模态增强表示。如图 3所示，MAM主要由语义属性建模、时空嵌入以及跨模态注意力融合三个部分组成。

3.5.1. 语义属性与时空建模

本文利用车辆类型、颜色及品牌等语义属性补充视觉身份信息。给定输入图像I，采用轻量化ResNet-18子网络分别对不同属性进行编码，生成对应语义嵌入：

$$\{f_{\text{type}}, f_{\text{color}}, f_{\text{brand}}\} \tag{4}$$

上述属性特征能够从语义层面增强车辆身份表示，并降低视角变化带来的影响。

除语义属性外，CameraID与Timestamp同样包含重要的跨摄像头关联先验。为此，本文构建可学习嵌入表对摄像头信息与时间信息进行编码，并通过特征映射获得统一时空表示f_{st}，用于刻

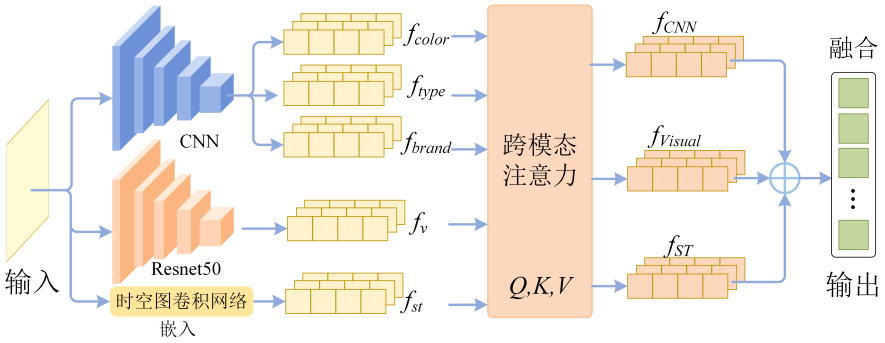

Figure 3. Architecture of the Metadata-Assisted Module (MAM)
图 3. 元数据辅助模块(MAM)整体结构

画摄像头拓扑关系与时间连续性规律。

3.5.2. 跨模态注意力融合

为实现视觉特征与元数据特征之间的有效交互，本文引入跨模态多头注意力机制。

其中，视觉外观特征f_v作为查询向量，语义属性特征与时空嵌入作为元数据Token参与注意力计算：

$$f_{\text{att}} = \text{Attention}\left(f_v, [f_{\text{type}}, f_{\text{color}}, f_{\text{brand}}, f_{\text{st}}]\right) \tag{5}$$

通过跨模态注意力学习，模型能够依据视觉上下文动态选择更具判别价值的元数据线索，从而强化重要语义与时空信息。

最终，注意力增强后的输出生成元数据表示：

$$f_{\text{meta}} = \text{Concat}(f_{\text{att}}) \tag{6}$$

所得特征f_{meta}综合编码视觉外观、语义属性以及时空先验信息，为后续跨摄像头身份关联提供更稳定的辅助表示。

3.6. 多模态协同增强与轨迹表示融合

在完成姿态感知模块(PAM)与元数据辅助模块(MAM)建模后，本文进一步对视觉外观特征、结构特征以及元数据特征进行统一融合，构建用于跨摄像头关联的轨迹级表示。

给定车辆目标的视觉外观特征f_v、结构增强特征f_{pose}以及元数据增强特征f_{meta}，统一帧级表示定义为：

$$g = \text{Concat}(f_v, f_{\text{pose}}, f_{\text{meta}}) \tag{7}$$

其中，Concat(·)表示特征拼接操作。

由于跨摄像头身份关联建立在轨迹层面，因此需要进一步将帧级特征聚合为轨迹级表示。对

于轨迹τ_i中的全部帧特征，采用时间池化操作生成统一轨迹表示：

$$f_i = \text{Pool}\left(\{g^t | (b_t, t, c) \in \tau_i\}\right) \tag{8}$$

其中，Pool(·)表示时间维度聚合操作，可采用平均池化或最大池化方式。

最终生成的轨迹表示f_i综合编码视觉外观、车辆结构及元数据信息，并作为后续跨摄像头身份匹配与全局轨迹关联的输入特征。

3.7. Search-and-Pruning优化策略

多源数据集融合能够提供更加丰富的语义、结构与时空监督信息，但同时也容易带来数据分布不均衡与跨域偏差问题。

本文构建的VeCamX数据集融合了VeRi-776细粒度属性信息与AIC22大规模轨迹信息，不同数据源在采集环境、视角分布及样本密度等方面存在明显差异。直接进行数据拼接容易产生冗余采样、类别失衡以及跨域伪差异问题，从而影响模型泛化能力 [5]。

为此，本文提出Search-and-Pruning(SnP)优化策略，通过目标导向子集搜索与预算约束剪枝两个阶段完成训练数据重组与筛选。如图 4所示，SnP整体流程主要包括目标域适配搜索与多维冗余抑制两个阶段。

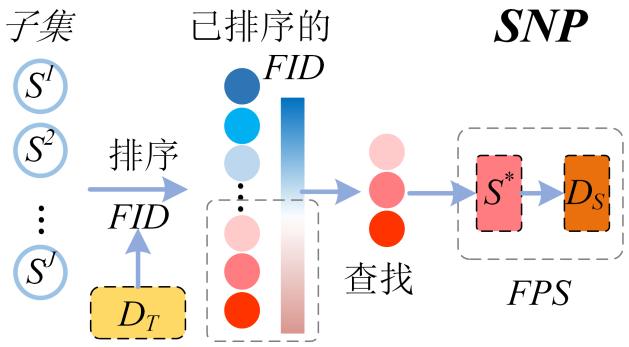

Figure 4. Overview of the Search-and-Pruning optimization strategy

图 4. Search-and-Pruning优化策略整体流程

3.7.1. 目标导向子集搜索

首先，将多个数据源统一组织为训练集合$S = \{S_1, S_2, \ldots, S_n\}$。随后，利用预训练ResNet-50提取样本特征，并通过K-means聚类获得候选子集。

为了评估候选子集与目标域之间的分布一致性，本文采用Fréchet Inception Distance (FID)作为筛选指标：

$$\text{FID} = \|\mu_j - \mu_T\|^2 + \text{Tr}\left(\Sigma_j + \Sigma_T - 2(\Sigma_j\Sigma_T)^{1/2}\right) \tag{9}$$

其中，(μ_j, Σ_j)与(μ_T, Σ_T)分别表示候选子集与目标域特征分布的均值及协方差矩阵。

完成FID排序后，进一步采用最远点采样(FPS)策略执行重采样，优先保留与已选样本差异较大的候选数据，从而提升训练集的分布覆盖能力与样本多样性。

3.7.2. 预算约束剪枝

在目标导向搜索基础上，SnP进一步执行预算约束剪枝，以降低训练数据中的空间与时间冗余。

在特征空间层面，采用迭代式FPS持续进行样本筛选，在固定预算条件下保留具有较强外观多样性的训练样本；在时间维度层面，通过帧率下采样与相邻冗余帧抑制减少长轨迹中的高度重复观测。

通过双维度剪枝，最终训练集能够在保持多模态信息完整性的同时，获得更优的数据均衡性与表示紧凑性，为后续跨摄像头身份学习提供更加稳定的数据基础。

3.8. 多分支联合损失函数设计

为实现视觉外观、车辆结构以及元数据信息的协同优化，本文采用多分支联合损失函数对主ReID分支、姿态感知模块(PAM)以及元数据辅助模块(MAM)

进行统一监督。

总体损失函数定义为：

$$L_{\text{total}} = \alpha L_{\text{ReID}} + \beta L_{\text{PAM}} + \gamma L_{\text{MAM}} \tag{10}$$

其中，α、β与γ分别表示各损失项权重系数。

主视觉分支采用ReID监督损失：

$$L_{\text{ReID}} = L_{\text{triplet}} + L_{\text{ce}} \tag{11}$$

其中，L_{triplet}用于增强类间可分离性与类内紧致性，L_{ce}用于强化身份分类能力。

对于姿态感知模块，采用均方误差损失监督车辆关键点热力图预测：

$$L_{\text{PAM}} = L_{\text{MSE}} \tag{12}$$

该损失用于约束预测结果与真实关键点热力图之间的差异。

元数据辅助模块采用多属性联合监督策略：

$$L_{\text{MAM}} = L_{\text{type}} + L_{\text{color}} + L_{\text{brand}} \tag{13}$$

其中，L_{type}、L_{color}以及L_{brand}分别对应车辆类型、颜色与品牌属性监督损失。

最终，通过联合优化不同分支目标，实现视觉外观、结构信息与元数据表示之间的跨模态协同学习。

4. 实验

4.1. 数据集与评价指标

为了验证所提出MTMCT框架的有效性，本文在VeRi-776 [5]、AIC22 CityFlow-ReID [7] [16]以及本文构建的VeCamX数据集上进行实验评估。为了全面评估所提出方法的性能，本文采用IDF1、IDP、IDR以及MOTA等标准指标。

IDP、IDR及IDF1定义如下：

$$IDP = \frac{IDTP}{IDTP + IDFP} \tag{14}$$

$$IDR = \frac{IDTP}{IDTP + IDFN} \tag{15}$$

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{16}$$

其中$IDTP$、$IDFP$与$IDFN$分别表示身份级真阳性、假阳性与假阴性。

此外，整体跟踪性能采用MOTA评价，其定义为：

$$MOTA = 1 - \frac{FN + FP + ID_{sw}}{GT} \tag{17}$$

其中FN、FP、ID_{sw}和GT分别表示漏检、误检、身份切换次数以及真实目标数量。

4.2. 对比实验

4.2.1. VeCamX数据集对比结果

表 2给出了VeCamX数据集上的对比结果可以看出，传统关联方法在身份一致性方面存在明显局限，相比之下，本文方法取得77.7% IDF1与74.5% MOTA，均优于现有方法，验证了PAM、MAM与SnP在多模态车辆跟踪场景中的有效性。

4.2.2. AIC22数据集对比结果

从表 3可以看出，MCVT在MOTA指标上表现较强，但IDF1相对偏低，说明其跨摄像头身份保持能力有限。本文方法在IDF1 上取得最佳结果，同时保持较高MOTA表现，表明所提出多模态优化策略能够有效提升复杂交通场景下的身份一致性。

Table 2. Comparison experiments on VeCamX dataset
表 2. VeCamX数据集上的对比实验结果

Method	Year	IDF1 (%)	MOTA (%)
BUPT	2020	70.2	67.0
Hsu et al. (TCLM+ReID)	2020	75.2	68.7
ANU	2021	74.1	71.5
MCVT (CVPRW 2021)	2021	77.0	70.4
Chung et al. (TransReID-V2)	2022	73.5	66.9
Nguyen et al. (Graph-based online)	2023	72.0	65.5
Huang et al. (Transformer-CLM)	2023	74.3	67.4
Ours	2025	**77.7**	**74.5**

Table 3. Comparison experiments on AIC22 dataset
表 3. AIC22数据集上的对比实验结果

Method	Year	IDF1 (%)	MOTA (%)
Hsu et al. (TCLM+ReID)	2020	74.9	71.0
MCVT (CVPRW 2021)	2021	72.5	76.5
Chung et al. (TransReID-V2)	2022	72.6	69.0
Nguyen et al. (Graph-based online)	2023	75.0	70.5
Huang et al. (Transformer-CLM)	2023	73.7	69.8
Ours	2025	**76.8**	73.2

4.2.3. VeRi-776数据集对比结果

Table 4. Comparison experiments on VeRi-776 dataset
表 4. VeRi-776数据集上的对比实验结果

Method	Year	IDF1 (%)	MOTA (%)
BUPT	2020	74.5	72.8
Hsu et al. (TCLM+ReID)	2020	75.8	73.4
ANU	2021	75.2	73.0
MCVT (CVPRW 2021)	2021	76.5	73.2
Chung et al. (TransReID-V2)	2022	76.1	72.9
Nguyen et al. (Graph-based online)	2023	77.0	73.6
Huang et al. (Transformer-CLM)	2023	77.2	73.8
Ours	2025	**78.3**	**75.0**

从表 4可以看出，随着方法从传统关联模型逐渐发展至Transformer与图网络结构，整体性能持续提升。本文方法在VeRi-776上取得最佳综合结果，说明所提出模块具有较好的跨数据集泛化能力。

4.3. 消融实验

为了验证所提出三个核心模块PAM、MAM与SnP的有效性，本文在VeRi-776、AIC22及VeCamX三个数据集上开展消融实验。实验以Fast Online MTMCT为Baseline，通过逐步引入不同模块分析其独立贡献与协同作用。

4.3.1. VeRi-776数据集消融结果

Table 5. Ablation study on VeRi-776 dataset
表 5. VeRi-776数据集上的消融实验结果

Method	IDF1	IDP	IDR	MOTA
Baseline	77.31	77.66	76.89	74.32
+SnP	77.47	77.79	77.06	74.45
+MAM	77.68	78.08	77.22	74.63
+PAM	78.15	78.52	77.73	74.89
All	**78.33**	**78.71**	**77.95**	**75.03**

从表 5可以看出，三个模块均能够稳定提升性能。其中PAM带来的增益最明显，说明结构姿态信息有助于缓解视角变化导致的身份混淆；MAM进一步增强语义判别能力；SnP通过数据优化改善训练样本质量。联合使用全部模块时获得最佳结果。

4.3.2. AIC22数据集消融结果

Table 6. Ablation study on AIC22 dataset
表 6. AIC22数据集上的消融实验结果

Method	IDF1	IDP	IDR	MOTA
Baseline	75.86	76.32	75.34	72.10
+SnP	75.97	76.41	75.46	72.32
+MAM	76.03	76.48	75.58	72.51
+PAM	76.57	77.02	76.04	72.81
All	**76.82**	**77.26**	**76.41**	**73.20**

表 6表明，在真实复杂交通场景中，PAM与MAM均能够有效增强跨摄像头身份保持能力，而SnP有助于提升训练稳定性。整体模型在IDF1与MOTA指标上均取得最优性能。

4.3.3. VeCamX数据集消融结果

从表 7可以看出MAM在VeCamX上的增益最为明显，说明丰富的语义属性与时空信息能够有效提升多模态特征表达能力；PAM通过关键点结构建模增强跨视角鲁棒性；SnP则有效降低数据

Table 7. Ablation study on VeCamX dataset
表 7. VeCamX数据集上的消融实验结果

Method	IDF1	IDP	IDR	MOTA
Baseline	76.25	76.63	75.82	73.21
+SnP	76.77	77.15	76.39	73.73
+MAM	77.55	77.95	77.10	74.22
+PAM	77.25	77.60	77.15	73.91
All	**77.75**	**78.10**	**77.45**	**74.50**

冗余带来的训练偏差。综合三个数据集结果可以发现，三个模块均具有稳定正向贡献，且联合优化能够获得最优性能，验证了本文提出方法在不同车辆MTMCT场景中的有效性与泛化能力。

5. 结论

本文提出了一种统一的多目标多摄像头车辆跟踪(MTMCT)框架，并构建了新的数据集VeCamX，用于推动该领域的性能提升。通过引入姿态感知模块(PAM)以增强视角不变性，引入元数据辅助模块(MAM)实现语义与时空信息的融合，以及提出搜索与剪枝策略(S nP) 以优化训练数据质量，本文有效缓解了外观变化与数据分布不均衡带来的影响。

在VeRi-776、AIC22 以及VeCamX 三个基准数据集上的实验结果表明，所提出方法在IDF1和MOTA 指标上均取得了持续且稳定的性能提升，验证了各模块设计的有效性与整体框架的优越性。

参考文献

[1] Veres, M. and Moussa, M. (2020) Deep Learning for Intelligent Transportation Systems: A Survey of Emerging Trends. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 3152-3168. https://doi.org/10.1109/tits.2019.2929020

[2] Shim, K., Ko, K., Hwang, J., Jang, H. and Kim, C. (2023) Fast Online Multi-Target Multi-Camera Tracking for Vehicles. *Applied Intelligence*, **53**, 28994-29004. https://doi.org/10.1007/s10489-023-05081-7

[3] Yang, H., Cai, J., Liu, C., Ke, R. and Wang, Y. (2023) Cooperative Multi-Camera Vehicle Tracking and Traffic Surveillance with Edge Artificial Intelligence and Representation Learning. *Transportation Research Part C: Emerging Technologies*, **148**, Article ID: 103982. https://doi.org/10.1016/j.trc.2022.103982

[4] Fei, L. and Han, B. (2023) Multi-Object Multi-Camera Tracking Based on Deep Learning for Intelligent Transportation: A Review. *Sensors*, **23**, Article 3852. https://doi.org/10.3390/s23083852

[5] Wang, H., Hou, J. and Chen, N. (2019) A Survey of Vehicle Re-Identification Based on Deep Learning. *IEEE Access*, **7**, 172443-172469. https://doi.org/10.1109/access.2019.2956172

[6] Lou, Y., Bai, Y., Liu, J., Wang, S. and Duan, L. (2019) VERI-Wild: A Large Dataset and a New Method for Vehicle Re-Identification in the Wild. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 3230-3238. https://doi.org/10.1109/cvpr.2019.00335

[7] Li, F., Wang, Z., Nie, D., Zhang, S., Jiang, X., Zhao, X., *et al.* (2022) Multi-Camera Vehicle Tracking System for AI City Challenge 2022. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, 19-20 June 2022, 3264-3272. https://doi.org/10.1109/cvprw56347.2022.00369

[8] Zhang, F., Zhang, L., Zhang, H. and Ma, Y. (2023) Image-To-Image Domain Adaptation for Vehicle Re-Identification. *Multimedia Tools and Applications*, **82**, 40559-40584. https://doi.org/10.1007/s11042-023-14839-7

[9] Wei, R., Gu, J., He, S. and Jiang, W. (2023) Transformer-Based Domain-Specific Representation for Unsupervised Domain Adaptive Vehicle Re-Identification. *IEEE Transactions on Intelligent Transportation Systems*, **24**, 2935-2946. https://doi.org/10.1109/tits.2022.3225025

[10] Li, J., Zhang, S., Tian, Q., Wang, M. and Gao, W. (2022) Pose-Guided Representation Learning for Person Re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 622-635. https://doi.org/10.1109/tpami.2019.2929036

[11] Hsu, H., Cai, J., Wang, Y., Hwang, J. and Kim, K. (2021) Multi-Target Multi-Camera Tracking of Vehicles Using Metadata-Aided Re-Id and Trajectory-Based Camera Link Model. *IEEE Transactions on Image Processing*, **30**, 5198-5210. https://doi.org/10.1109/tip.2021.3078124

[12] Jain, E., Nandy, T., Aggarwal, G., Tendulkar, A., Iyer, R. and De, A. (2023) Efficient Data Subset Selection to Generalize Training across Models: Transductive and Inductive Networks. *Advances in Neural Information Processing Systems 36*, New Orleans, 10-16 December 2023, 4716-4740. https://doi.org/10.52202/075280-0210

[13] Xue, C., Deng, Z., Wang, S., Hu, E., Zhang, Y., Yang, W., *et al.* (2024) GLSFF: Global-Local Specific Feature Fusion for Cross-Modality Pedestrian Re-Identification. *Computer Communications*, **215**, 157-168. https://doi.org/10.1016/j.comcom.2023.12.035

[14] Zhai, Y., Zeng, Y., Huang, Z., Qin, Z., Jin, X. and Cao, D. (2024) Multi-Prompts Learning with Cross-Modal Alignment for Attribute-Based Person Re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 6979-6987. https://doi.org/10.1609/aaai.v38i7.28524

[15] Nguyen, T.T., Nguyen, H.H., Sartipi, M. and Fisichella, M. (2024) Multi-Vehicle Multi-Camera Tracking with Graph-Based Tracklet Features. *IEEE Transactions on Multimedia*, **26**, 972-983. https://doi.org/10.1109/tmm.2023.3274369

[16] Chen, Y.C., Jing, L.L., Vahdani, E., Zhang, L., He, M.Y. and Tian, Y.L. (2019) Multi-Camera Vehicle Tracking and Re-Identification on AI City Challenge 2019. *CVPR Workshops 2019*, Long Besch, 16-20 June 2019, 324-332.

[17] Yoon, Y., Song, Y., Yoon, K. and Jeon, M. (2018) Online Multi-Object Tracking Using Selective Deep Appearance Matching. *2018 IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia)*, JeJu, 24-26 June 2018, 206-212. https://doi.org/10.1109/icce-asia.2018.8552105

[18] Nguyen, T.T., Nguyen, H.H., Sartipi, M. and Fisichella, M. (2024) Lammon: Language Model Combined Graph Neural Network for Multi-Target Multi-Camera Tracking in Online Scenarios. *Machine Learning*, **113**, 6811-6837. https://doi.org/10.1007/s10994-024-06592-1

[19] Nguyen, T.T., Nguyen, H.H., Sartipi, M. and Fisichella, M. (2023) Real-Time Multi-Vehicle Multi-Camera Tracking with Graph-Based Tracklet Features. *Transportation Research Record: Journal of the Transportation Research Board*, **2678**, 296-308. https://doi.org/10.1177/03611981231170591

[20] Ayala-Acevedo, A., Devgun, A., Zahir, S. and Askary, S. (2019) Vehicle Re-Identification: Pushing the Limits of Re-Identification. *CVPR Workshops 2019*, Long Beach, 16-20 June 2019, 291-296.

[21] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-Of-Freebies Sets New State-Of-The-Art for Real-Time Object Detectors. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/cvpr52729.2023.00721

[22] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., *et al.* (2022) ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M. and Hassner, T., Eds., *Computer Vision—ECCV 2022*, Springer, 1-21. https://doi.org/10.1007/978-3-031-20047-2_1

[23] Sun, K., Xiao, B., Liu, D. and Wang, J. (2019) Deep High-Resolution Representation Learning for Human Pose Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Kushtia, 23-24 October 2025, 1-6. https://doi.org/10.1109/cvpr.2019.00584