文章编号:1671-1637(2026)06-0137-16

面向夜间道路交通事故的 VSSM-CNN 检测网络构建

杨 洋[1],陈献天[1,2],王健宇*[2],蒲自源[3],赵红专[4],袁振洲[1]

(1. 北京交通大学 交通运输学院,北京 100044; 2. 北京建筑大学 土木与交通工程学院,北京 102616;
3. 东南大学 交通学院,江苏 南京 211189; 4. 桂林电子科技大学 建筑与交通工程学院,广西 桂林 541004)

摘 要:为提升夜间场景下道路交通事故的自动检测性能,基于视觉状态空间模型(VSSM)与卷积神经网络(CNN)构建了 VSSM-CNN 编码器-解码器架构,提出了一种面向该场景的无监督交通事故检测框架;参考特征融合思路,并基于已有的可见光图像作为外观特征,进一步采用循环全对场变换(RAFT)光流估计算法结合卷积长短期记忆网络(ConvLSTM)模块提取并处理视频序列的细粒度光流信息,以表征交通运动状态;将外观与运动两类特征融合后输入至以视觉状态空间模型为主干网络的特征编码器进行全局特征提取,并采用卷积神经网络作为解码器架构,逐层恢复图像以强化局部细节,从而加强特征提取效率与利用效果;采取均方误差、平均绝对误差与结构相似性三重损失函数组合,通过贝叶斯优化确定权重比例,以提升模型在夜间图像重构时对异常结构与噪声的鲁棒性。研究结果表明:所提方法的受试者工作特征曲线下面积(ROC-AUC)与精确率-召回率曲线下面积(PR-AUC)分别为 0.818 与 0.765,较传统生成式网络分别提升了 22.6%和 20.3%,其中 ROC-AUC 在与多种现有方法的比较中取得了最高值;三重损失函数在消融试验中展现出了最强的模型性能提升能力;研究方法取得了最低的模型复杂度与较优的检测速度,充分展现其异常识别的稳定性与部署应用的潜力。研究成果有效提升了夜间场景下的交通事故检测能力,可进一步推广至低光照、低能见度场景下的交通事故检测,并为相关的应用提供可行的技术方案与参考。

关键词:智能交通;夜间交通事故检测模型;无监督学习;视频异常检测;交通安全;视觉状态空间模型
中图分类号:U491.31 **文献标志码**:A **DOI**:10.19818/j.cnki.1671-1637.2026.076

Construction of VSSM-CNN detection network for nighttime road traffic accidents

YANG Yang[1], CHEN Xian-tian[1,2], WANG Jian-yu*[2], PU Zi-yuan[3],
ZHAO Hong-zhuan[4], YUAN Zhen-zhou[1]

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; 2. School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; 3. School of Transportation, Southeast University, Nanjing 211189, Jiangsu, China; 4. School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

出版历程:2025-08-09 收稿,2025-09-17 修回,2025-09-28 录用
基金项目:国家自然科学基金项目(52572336);北京市自然科学基金项目(E2024210149);北京建筑大学培育项目专项资金(X25034);中央高校基本科研业务费专项资金项目(2024JBRC009)
作者简介:杨 洋(1988-),男,河北邢台人,副教授,工学博士,E-mail:bjtuyang@bjtu.edu.cn。
* **通信作者**:王健宇(1991-),男,辽宁沈阳人,副教授,工学博士,E-mail:wangjianyu@bucea.edu.cn。
引用格式:杨 洋,陈献天,王健宇,等.面向夜间道路交通事故的 VSSM-CNN 检测网络构建[J].交通运输工程学报,2026,26(6):137-152.
Citation: YANG Yang, CHEN Xian-tian, WANG Jian-yu, et al. Construction of VSSM-CNN detection network for nighttime road traffic accidents[J]. Journal of Traffic and Transportation Engineering, 2026, 26(6): 137-152.

Abstract: To enhance the automatic detection performance of road traffic accidents in nighttime scenarios, a VSSM-CNN encoder-decoder architecture was constructed based on a visual state space model (VSSM) and a convolutional neural network (CNN), and an unsupervised traffic accident detection framework oriented to this scenario was proposed. By referencing the idea of feature fusion and based on existing visible-light images as appearance features, fine-grained optical flow information of video sequences was further extracted and processed using a recurrent all-pairs field transform (RAFT) optical flow estimation algorithm combined with a convolutional long short-term memory (ConvLSTM) module to represent traffic motion states. The two types of features, appearance and motion, were fused and input into a feature encoder with the VSSM as a backbone network for global feature extraction, and a CNN was adopted as a decoder architecture to recover images layer by layer to enhance local details, so as to strengthen feature extraction efficiency and utilization effects. A triple loss function combination of mean squared error, mean absolute error, and structural similarity was adopted, and weight proportions were determined through Bayesian optimization to improve the robustness of the model to abnormal structures and noise during nighttime image reconstruction. Research results indicate that the area under the receiver operating characteristic curve (ROC-AUC) and area under the precision-recall curve (PR-AUC) of the proposed method are 0.818 and 0.765, respectively, which improve by 22.6% and 20.3%, respectively, compared with traditional generative networks; among them, the ROC-AUC achieves the highest value in the comparison with multiple existing methods; the triple loss function shows the strongest model performance improvement ability in ablation experiments; the research method achieves the lowest model complexity and a favorable detection speed, fully demonstrating its stability in anomaly recognition and potential for deployment and application. The research results effectively improve the traffic accident detection ability in nighttime scenarios, can be further extended to traffic accident detection in low-illumination and low-visibility scenarios, and provide a feasible technical solution and reference for related applications.

Keywords: intelligent transportation; nighttime traffic accident detection model; unsupervised learning; video anomaly detection; traffic safety; visual state space model

Publication history: Received 2025-08-09; Received in revised form 2025-09-17; Accepted 2025-09-28

Funding: National Natural Science Foundation of China (52572336); Natural Science Foundation of Beijing (E2024210149); Cultivation Project Funds for Beijing University of Civil Engineering and Architecture (X25034); Fundamental Research Funds for the Central Universities (2024JBRC009)

* **Corresponding author**: WANG Jian-yu, associate professor, PhD, E-mail: wangjianyu@bucea.edu.cn.

0 引 言

道路交通事故通常会造成不同程度的财产损失与人员伤亡,进而引发如交通拥堵、二次事故等继发交通问题[1-3]。目前,交通事故发生后的处理方法大多通过当事人或目击者进行报警救助之后,相关部门再进行线下的应急预案或线上出警的方式处理[4]。在智能网联、自动驾驶等智能化交通技术迅速发展的背景下[5],这类传统的处理方式实时性差,并且较难准确地获取事故信息。同时,有报告指出在道路交通事故中,未当场死亡但在救护人员到达之前死亡的案例占 27.2%,抢救无效死亡的案例占 52%[6],间接强调了交通事故事后救援效率的重要性与必要性。值得注意的是,夜间环境下的交通事故因环境光照不足、人为驾驶操作失误可能性大及视频监控识别困难等因素,往往具有更高的致死率与延迟救援风险。如何在此类条件下实现事故的快速、准确检测,成为提升应急响应效率的关键问题。

近年来，随着道路交通基础设施的完善以及智能交通系统的快速发展，路侧监控、行车记录仪等覆盖和传播范围逐渐扩大，相较于结构化数据驱动的事故致因分析[7-8]与安全风险识别[9-10]，视频画面等非结构性数据具有更多可挖掘的特征，因此基于各类视角下的事故视频检测方法正逐渐成为研究热点。当前，在视频交通事故检测的有关研究中，多数学者所采用的方式是以各类监督学习为主，并结合交通事故显著的外观特征（如车辆的倾覆、破损，或是行人、非机动车驾驶者的跌倒等有别于正常交通行为的异常情况）和运动特征（如短时间内交通参与者空间位置的快速变化、不同车辆的空间位置交叉、叠加等）实现基于视频的交通事故检测。Batanina 等[11]通过构造三维卷积，捕捉连续帧的时空特征后进行特征抽象，并设立有监督损失进行训练，以此实现事故检测。Fang 等[12]提出了一种基于自监督一致性学习框架的交通事故检测方法，通过考虑时间框架一致性、时间物体位置一致性和道路参与者的时空关系一致性来检测交通事故。Pashaei 等[13]利用卷积神经网络对事故图像进行特征提取，并结合多种极限学习机（Extreme Learning Machine，ELM）的集成分类器，实现交通事故的识别与分类。Ijjina 等[14]提出了一种基于 Mask R-CNN 与质心对象跟踪算法，在车辆与其他车辆出现重叠后，根据速度和轨迹异常来确定是否发生事故。尽管现有方法已经被证实在交通事故视频检测任务上是有效的，但在识别夜间场景下的交通事故时仍有较大的改进空间，同时针对夜间交通事故检测的研究正逐渐被重视[1]。现有方法大多在日间条件下进行训练与验证，在应用到夜间环境，尤其是面向复杂交通场景进行检测时，多数方法的性能会显著下降[15]。此外，受制于低照度、模糊、强光干扰等环境因素，交通状态特征的提取会受到不同程度的干扰；同时，夜间交通事故视频资源较为稀缺，相关数据集覆盖不足，进一步制约了方法的研究与推广[16-17]。

针对上述此类的问题，有部分学者认为利用无监督学习的方式能够有效弥补因数据量缺失造成的训练困难问题[18-19]。此外，受制于眩光和噪声等因素的影响，夜间场景下的外观特征难以完整反映交通运行状态，而运动信息在一定程度上能够更直观地表征交通参与者的动态行为，因此引入不同模态的数据进行特征融合能够有效提升夜间场景的检测任务精度。Singh 等[20]使用去噪自编码器提取深度表征，并结合车辆轨迹交点、光流特征等进行分析，能够有效地识别交通事故，降低误报率。Zhou 等[21]利用卷积神经网络提取事故外观特征，以及光流信息表征的运动特征，实现了在拥堵场景下的交通事故检测。Chen 等[22]提出了一种基于 3D 目标检测和自适应空间划分的交通事故定量推断方法，通过结合 RGB 图像和点云数据，实现了高精度的 3D 边界盒回归和分割。张奇等[23]提出了一种融合递进式域适应和交叉注意力机制的交通事故检测方法，通过对事故的外观特征和运动特征进行提取并融合实现面向中国场景的交通事故检测。因此，采用无监督学习的方法，并利用不同类型的特征融合策略，加强检测任务下信息的提取，是实现复杂环境下交通事故检测的解决思路之一。

交通事故作为有别于正常交通状态的事件，其本质上仍属于异常检测的范畴。夜间场景下的交通事故样本数量过少，因此采用无监督学习的方式能够较为有效地缓解这一难题导致的检测效果不佳的问题。基于无监督学习的异常检测相关方法可分为基于重构误差[24-26]和基于预测误差[27-29]的无监督异常检测方法，其编码器和解码器多采用两类主干模型架构：基于卷积神经网络（Convolutional Neural Network，CNN）的采样模块与基于 Transformer 的采样模块。CNN 具有较强的特征挖掘能力，在同种检测任务下对不同类别场景进行识别时具有较好的兼容能力，但也导致 CNN 在训练过程会将一些异常事件视为正常存在的误差；此外，在捕捉长序列视频帧之间的时序特征时会有概率地忽略部分特征，导致特征信息流失，特别是其时间信息会被过滤从而导致精度下降[30]。基于 Transformer 的采样模块可以有效提取长序列视频帧之间的特征，但不同的交通事故视频样本存在发生场景、拍摄角度不统一等问题，加之 Transformer 模型在训练和推理过程中需要大量的计算资源等问题，导致其在训练过程中效率下降[31]，因此并不完全适用于交通事故检测任务。

考虑目前研究存在的问题，本研究参考交通异常行为检测的思路，设计了一种基于无监督学习范式的夜间交通事故检测网络。以自编码器结构为基础，在基于可见光图像检测的基础上融合包含车辆运动信息的光流特征，并采用特征融合的策略对夜间场景的交通状态进行学习；利用对事故场景的重构异常分数差异进行判别，实现该场景下的事故检测，以弥补夜间场景的交通事故数据较为稀缺的不足。同时，考虑到车辆运动具有细微特征与强时序

依赖性，在光流提取网络中采用循环全对场变换(Recurrent All-pairs Field Transforms，RAFT)[32]光流估计算法，并嵌入卷积长短期记忆网络(Convolutional Long Short-term Memory，ConvLSTM)模块以显式建模帧间时序关联并增强对车辆急停与碰撞等异常交通状态的敏感性，提取更为丰富的运动特征。此外，为应对 CNN 或 Transformer 等架构带来的问题，基于 Jiao 等[33]提出的在长序列建模性能优异且计算复杂度较低的 VMamba 模型的核心架构视觉状态空间模型(Visual State Space Model，VSSM)构建编码器，并结合由均方误差(Mean Squared Error，MSE)、平均绝对误差(Mean Absolute Error，MAE)和结构相似性指数(Structural Similarity Index，SSIM)构成的组合损失函数进行训练，以保证在噪声干扰严重的夜间场景中保持较强的鲁棒性与重建性能。最后采用 CNN 作为解码器，加强编码后的局部特征约束能力，最终形成以 VSSM-CNN 为架构的自编码器模型，实现夜间场景下事故图像的重构并检测事故的发生与否。

1 研究方法

1.1 模型构建

为了使模型能够充分学习正常状态下的交通图像，并尽可能地实现图像重建，本研究参考经典的生成模型，基于视觉状态空间模型构建编码器-解码器模型，如图 1 所示。该模型的整体训练流程可概述为如下过程。

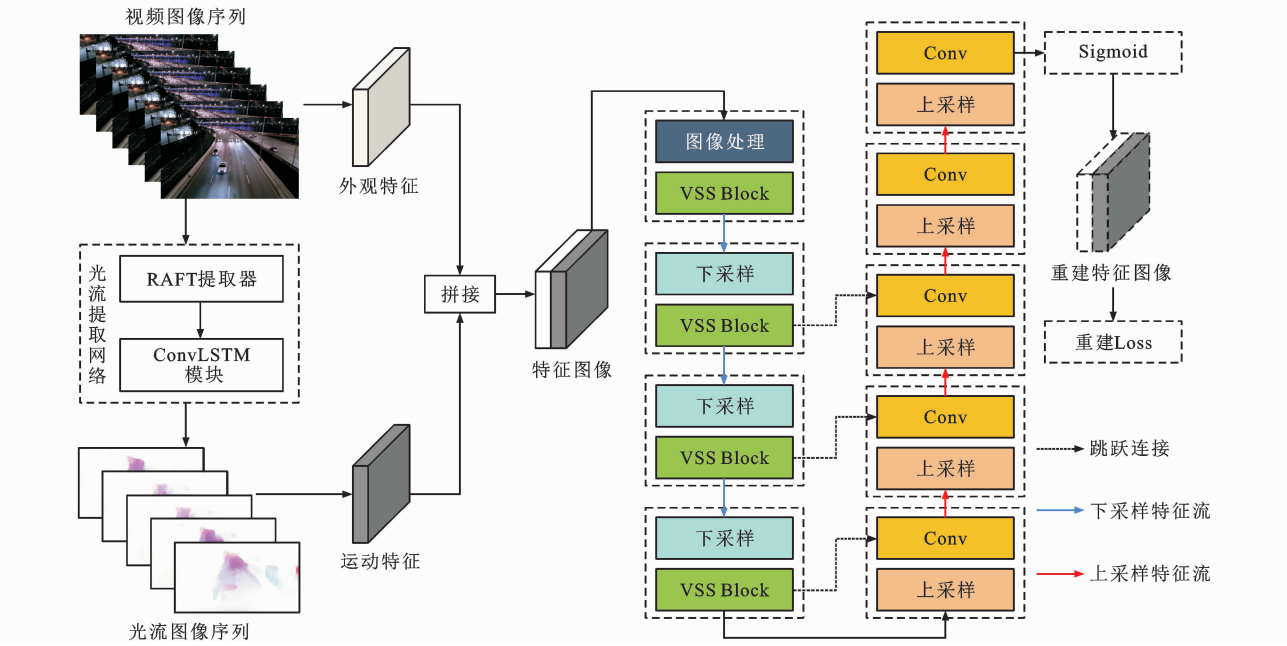


图 1 夜间环境交通事故检测网络

Fig. 1 Traffic accident detection network in nighttime environments

步骤 1：输入帧数为 T，高、宽分别为 H、W 的可见光视频片段 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T)$，$\boldsymbol{x}_T\in\mathbb{R}^{H\times W\times 3}$。

步骤 2：提取某一时刻下连续 3 帧视频帧，并利用 RAFT 算法对夜间场景下的正常交通视频图像序列估计光流，得到正常交通状态下运动特征的运动分量 $\boldsymbol{m}_t$。将光流序列 $\boldsymbol{m}_t$ 输入至 ConvLSTM 模块，捕捉时序依赖关系，得到时序增强的运动特征 $\boldsymbol{m}_t'$。

步骤 3：由于光流帧已经具备了视频序列中的时序信息，因此只需提取一张可见光的图像作为其外观特征提取的输入，并采用沿通道拼接的方式将外观特征与运动特征进行融合得到 z_t。沿通道拼接作为早期融合的方式能够最大程度保留原始模态信息，相较于自注意力融合或相乘融合具有较低的复杂度。

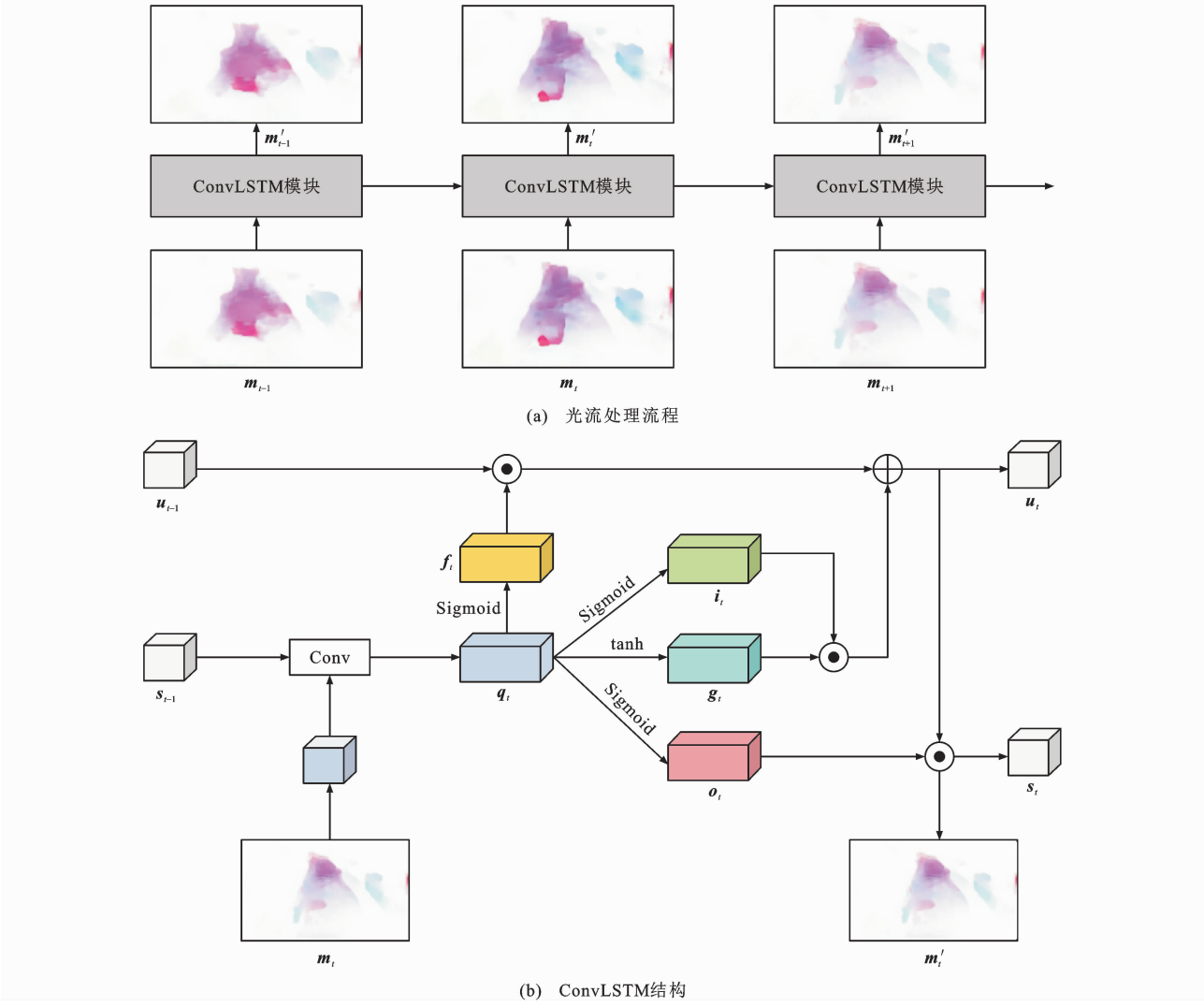
步骤 4：通过编码器进行序列建模，记录输入特征图的全局信息，得到下采样后的特征信息；将编码器输出的特征信息传入解码器，采用卷积操作逐层恢复特征信息，实现特征图的局部细化。此外，在特征的采样过程中，采用跳跃连接直接传递对应层之间的特征信息，能够起到防止梯度消失和信息遗漏的作用。

步骤 5：将重构的特征图与输入特征图进行对比，得到重构损失，并根据其大小与阈值判断该序列是否存在交通事故。

1.1.1　光流提取网络

道路交通事故通常具有较为明显的异于正常交通行为的运动状态，例如车辆或行人位置在短时间内出现的不规律位移，包括行驶轨迹偏离与停止等，因此运动信息也可以作为判断事故发生与否的一个参考标准。目前，卷积神经网络正逐步成为光流估计的主要框架，近年来更是不乏许多性能优异的光流估计算法。其中，Teed 等[32]提出的 RAFT 光流提取网络成为许多研究领域的主流框架。该方法使用特征编码层逐像素提取特征，为所有元素建立关联信息，并通过循环迭代的方式不断更新光流场，在保持计算效率的同时，将运动场估计精度提升至亚像素级，进一步提高了光流估计的精度和鲁棒性。对于该场景下的学习任务，相较于 FlowNet[34]，RAFT 算法能够较好地捕捉夜间低光照条件下的细微运动信息。此外，RAFT 算法通过迭代更新和全对场变换，比 PWC-Net[35]网络更适应不同分辨率的图像，因此本研究选择 RAFT 算法作为光流提取的骨干网络。

观察交通运行状态下的视频序列，可以直观地发现其具有较为明显的规律性和时序依赖性：在正常情况下，车辆和行人的速度变化相对平稳，轨迹与道路空间边界保持一致，表现出明显的时序连续性和平滑性；而在事故发生时，这些规律往往会被打破，表现为车辆或行人突然急停、偏离既定轨迹或发生碰撞等运动突变。因此，为了更好地刻画这种由“规律性”向“异常性”的转变细节，本研究将 RAFT 算法所提取的光流场输入至 ConvLSTM 模块，用于在时间维度上显式建模帧间依赖关系，使模型能够捕捉并利用交通运行状态的内在规律，其处理过程如图 2(a)所示，图 2(b)则展示了 ConvLSTM 模块的结构。具体处理过程可概述为如下步骤。



(a)　光流处理流程

(b)　ConvLSTM结构

图 2　ConvLSTM 光流处理模块

Fig. 2　ConvLSTM optical flow processing module

步骤 1：对输入的连续 T 帧，高、宽分别为 H、W 可见光视频片段 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T)$，$\boldsymbol{x}_T\in\mathbb{R}^{H\times W\times 3}$ 采用 RAFT 算法从相邻帧估计逐帧光流，得到表征运动信息的光流帧，并组成光流帧序列 $\boldsymbol{M}=(\boldsymbol{m}_1,\boldsymbol{m}_2,\cdots,\boldsymbol{m}_{T-3})$，其最后一帧光流帧的输出过程为

$$\boldsymbol{m}_{T-3}=\mathrm{RAFT}(\boldsymbol{x}_{T-2},\boldsymbol{x}_{T-1},\boldsymbol{x}_T)\quad \boldsymbol{m}_{T-3}\in\mathbb{R}^{H\times W\times 2} \tag{1}$$

步骤 2：将 RAFT 算法提取的光流帧序列 $\boldsymbol{M}$ 输入至 ConvLSTM 模块，得到时序增强后的光流帧，并组成新序列 $\boldsymbol{M}'=(\boldsymbol{m}'_1,\boldsymbol{m}'_2,\cdots,\boldsymbol{m}'_{T-3})$。在某一时刻 t，ConvLSTM 首先将上一时刻的隐藏状态 $\boldsymbol{s}_{t-1}$ 以及当前时刻的光流帧 $\boldsymbol{m}_t$ 进行卷积操作，并沿通道维度进行拼接，得到中间特征张量 $\boldsymbol{q}_t$，即

$$\boldsymbol{q}_t=\mathrm{Conv}(\boldsymbol{m}_t)+\mathrm{Conv}(\boldsymbol{s}_{t-1}) \tag{2}$$

步骤 3：将中间特征张量经过 4 次非线性变化，分别得到输入门信息 $\boldsymbol{i}_t$、遗忘门信息 $\boldsymbol{f}_t$、输出门信息 $\boldsymbol{o}_t$、记忆门信息 $\boldsymbol{g}_t$。其中，$\boldsymbol{i}_t$、$\boldsymbol{f}_t$、$\boldsymbol{o}_t$ 遵循形式相同、参数不同的激活函数 Sigmoid(·)，$\boldsymbol{g}_t$ 采用激活函数 tanh(·)，计算过程为

$$\begin{cases}\boldsymbol{i}_t=\mathrm{Sigmoid}(\boldsymbol{q}_t)\\ \boldsymbol{f}_t=\mathrm{Sigmoid}(\boldsymbol{q}_t)\\ \boldsymbol{o}_t=\mathrm{Sigmoid}(\boldsymbol{q}_t)\\ \boldsymbol{g}_t=\tanh(\boldsymbol{q}_t)\end{cases} \tag{3}$$

步骤 4：利用上一时刻的记忆状态 $\boldsymbol{u}_{t-1}$，经过各门控操作得到该时刻的输出，分别为当前时刻的隐藏状态 $\boldsymbol{s}_t$ 和记忆状态 $\boldsymbol{u}_t$，即

$$\begin{cases}\boldsymbol{u}_t=\boldsymbol{f}_t\odot\boldsymbol{u}_{t-1}+\boldsymbol{i}_t\odot\boldsymbol{g}_t\\ \boldsymbol{s}_t=\boldsymbol{o}_t\odot\boldsymbol{u}_t\end{cases} \tag{4}$$

通过 ConvLSTM 模块处理后的光流特征不仅保留了原始光流的空间信息，还融入了时间序列的长期依赖型，此类方式能够加强光流特征图对交通运动状态的描述，更易于辨识正常与非正常的交通状态光流。

1.1.2 编码器

在以无监督学习为基础的异常行为检测任务中，上下文的全局信息以及局部细节信息同样重要，尤其是以特征融合为主要方式的多模态数据融合策略，充分提取与利用不同模态的信息能够有效提升检测精度[36]。尽管在编码器阶段同时使用 CNN＋Transformer 作为骨干网络可以有效提升精度，但仍较难平衡因 Transformer 而导致的计算效率问题。因此，本研究在编码器阶段采用具有线性计算复杂度的视觉状态空间架构模型 VSSM，借助其长序列建模优势记录正常交通状态中的全局信息。

图像处理模块是利用 Patch Embedding 操作将输入的特征图切割成均等大小且互不重叠的块，随后传入编码器的特征提取模块 VSS Block，总共经过 4 次下采样过程(1→1/4→1/8→1/16→1/32)，最终得到编码器的输出，并保存第 2、3、4 次下采样的输出作为跳跃连接的输入。在 VSS Block 中，其核心计算单元为 2D 选择性扫描(2D Selective-scan，SS2D)，其具体的信息记录模式可概述为如下过程：首先，对输入的特征图在 4 个方向上进行扫描，确保从不同方向获取特征图的信息，保持空间尺度上的关联性，使其能够捕捉该特征图中更多的上下文信息；随后不同方向上的特征会被分别传输至选择性扫描机制模块进行信息整合，以过滤掉无关信息；最后经过处理后的特征图会被重新组合为原始输入的大小，此输出保留了丰富的上下文信息，并且为每一部分的信息保留了不同的权重。在编码器部分，针对前文处理过后的运动特征，将其与外观特征拼接得到编码器的输入序列 $\boldsymbol{z}$，经过编码器实现完整下采样操作后得到输出序列 $\boldsymbol{z}'$。

编码器中 VSS Block 的核心来源是状态空间模型，其被视为一个线性时不变系统，将其应用于深度学习领域时需要对其数据来源进行处理，即根据零阶保持原则对连续输入的数据进行离散化操作，随后再应用于状态空间模型。因此，该模块的计算过程如式(5)所示，处理流程示意如图 3 所示。在实际训练前，参数 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$ 在初始化阶段随机，训练过程中可学习并自动更新直至模型收敛，即

$$\begin{cases}\bar{\boldsymbol{A}}=\mathrm{e}^{\Delta\boldsymbol{A}}\\ \bar{\boldsymbol{B}}=(\mathrm{e}^{\Delta\boldsymbol{A}}-\boldsymbol{I})\boldsymbol{A}^{-1}\boldsymbol{B}\\ \boldsymbol{h}_t=\bar{\boldsymbol{A}}\boldsymbol{h}_{t-1}+\bar{\boldsymbol{B}}\boldsymbol{z}_t\\ \boldsymbol{y}_t=\boldsymbol{C}\boldsymbol{h}_t+\boldsymbol{D}\boldsymbol{z}_t\end{cases} \tag{5}$$

式中：Δ 为时间尺度参数；$\boldsymbol{I}$ 为单位矩阵；$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别为状态转移矩阵和控制输入矩阵；$\bar{\boldsymbol{A}}$ 和 $\bar{\boldsymbol{B}}$ 分别为 $\boldsymbol{A}$、$\boldsymbol{B}$ 离散转换后的参数；$\boldsymbol{h}_t$ 为系统在时间 t 上的潜在状态；$\boldsymbol{z}_t$ 为特征融合后的输入向量；$\boldsymbol{y}_t$ 为输出向量；$\boldsymbol{C}$ 和 $\boldsymbol{D}$ 分别为投影参数和跳跃连接参数。

1.1.3 解码器

在编码器部分，利用 VSS Block 能够有效过滤无关信息，但可能会不可避免地忽略部分细节信息。因此，在解码器阶段采用转置卷积对编码器阶段所提取到的深层特征图进行复原，并约束细节信息防止其丢失。解码器与编码器属于非对称式结构，在解码器的前 3 层采用逐层跳跃连接以保证模型对图

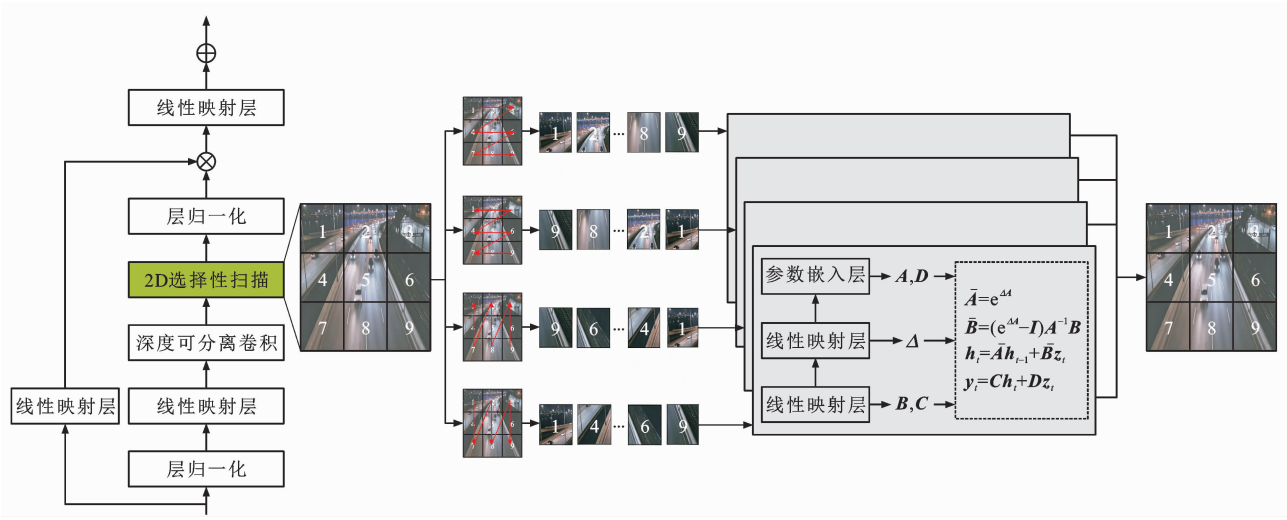


图 3　VSS Block 与 SS2D 模块

Fig. 3　VSS block and SS2D module

像细节的恢复能力，从最深层特征图恢复到原始图像需经历 5 次上采样(1/32→1/16→1/8→1/4→1/2→1)后得到重建图。在本研究中，下采样操作后得到的输出即作为解码器的输入张量，应用转置卷积恢复并逐层输出的过程为上采样，具体计算如式(6)、(7)所示

$$\boldsymbol{K}_l=\begin{cases}\mathrm{ReLU}[\mathrm{Conv}(\boldsymbol{Z}_{l-1})+\boldsymbol{S}_l] & l=1,2,3\\ \mathrm{ReLU}[\mathrm{Conv}(\boldsymbol{Z}_{l-1})] & l=4,5\end{cases} \tag{6}$$

$$\boldsymbol{Y}=\mathrm{Sigmoid}(\boldsymbol{K}_5) \tag{7}$$

式中：$\boldsymbol{K}_l$ 为第 l 层上采样后的特征图；$\boldsymbol{Z}_{l-1}$ 为 $l-1$ 层的输入特征图；$\boldsymbol{S}_l$ 为来自编码器对应尺度的跳跃连接特征，当 $l\leqslant 3$ 时与该层的特征图相加融合；ReLU(·)与 Sigmoid(·)均为非线性运算；$\boldsymbol{Y}$ 为最终的重建图像。

1.2　损失函数

在异常检测模型的训练过程中，损失函数多采用均方误差 L_{MSE}，其表示 2 个图像中对应像素点之间的误差，取值越接近 0 则表明误差越低，即

$$L_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{x}_i)^2 \tag{8}$$

式中：N 为特征图中的像素数量；x_i 为真实图像的第 i 个像素值；$\hat{x}_i$ 为重构图像的第 i 个像素值。

在夜间交通场景中，交通运行状态较为复杂，尤其在干扰正常交通行为出现时，这些行为常被视为噪声信号，在重构过程中可能会出现遗漏，进而导致均方误差较大。为应对上述问题，受文献[37]、[38]启发，本研究引入了平均绝对误差和结构相似性指数，以增强模型对噪声的鲁棒性和对结构信息的保持能力。对误差的惩罚较为宽容，适合噪声较多的场景。在图像重构任务中，平均绝对误差损失能够有效地抑制异常值的影响，避免模型过度拟合噪声，从而提高模型的泛化能力。结构相似性指数则用于衡量图像在亮度、对比度和结构上的相似性。在夜间环境下，由于车灯、路灯等光源的存在，图像的亮度和对比度可能存在差异，结构相拟性指数损失能够有效地捕捉这些结构上的变化，保持图像的结构信息，从而提高重构图像的质量。平均绝对误差损失 L_{MAE} 和结构相似性指数损失 L_{SSIM} 的计算公式分别为

$$L_{\mathrm{MAE}}=\frac{1}{N}\sum_{i=1}^{N}|x_i-\hat{x}_i| \tag{9}$$

$$L_{\mathrm{SSIM}}=1-\frac{(2\mu_x\mu_{\hat{x}}+C_1)(2\sigma_{x\hat{x}}+C_2)}{(\mu_x^2+\mu_{\hat{x}}^2+C_1)(\sigma_x^2+\sigma_{\hat{x}}^2+C_2)} \tag{10}$$

$$\begin{cases}\mu_x=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}x_i\\ \mu_{\hat{x}}=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\hat{x}_i\\ \sigma_x^2=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(x_i-\mu_x)^2\\ \sigma_{\hat{x}}^2=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(\hat{x}_i-\mu_{\hat{x}})^2\\ \sigma_{x\hat{x}}=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(x_i-\mu_x)(\hat{x}_i-\mu_{\hat{x}})\end{cases}$$

式中：μ_x、$\mu_{\hat{x}}$ 分别为真实图像和重构图像的像素均值；σ_x^2、$\sigma_{\hat{x}}^2$ 分别为真实图像和重构图像的像素方差；$\sigma_{x\hat{x}}$ 为真实图像和重构图像的协方差；C_1 和 C_2 为常数，依据文献与实际试验情况分别取 0.065 和 0.585，用于避免除零错误。

因此，训练过程的总损失函数 L_{total} 为

$$L_{\text{total}} = w_1 L_{\text{MAE}} + w_2 L_{\text{SSIM}} + w_3 L_{\text{MSE}} \quad (11)$$

式中：w_1、w_2、w_3 分别为 L_{MAE}、L_{SSIM}、L_{MSE} 损失对应的权重参数，取值分别为 0.840 4、1.244 7、4.694 5。

在本次训练过程中，权重参数由贝叶斯超参数优化规则确定。为了节约计算资源并提高效率，本研究采用“代理度量＋贝叶斯优化”的两阶段策略来确定总损失函数中的权重组合。参考 Klein 等[39]和 Wu 等[40]的研究思路，短周期验证损失与长周期验证损失在超参数排序层面具有较高的相关系数，证明其在早期淘汰劣势解的可行性。在本研究中，损失函数的权重参数确定过程如下：

步骤 1：在搜索空间内随机采样 n 组超参数组合 $\boldsymbol{w}^{(i)} = (w_1^{(i)}, w_2^{(i)}, w_3^{(i)})$，使用短周期训练得到对应的验证损失。由超参数组合及其对应的损失值构成初始观测集，表示为

$$\mathcal{D}_0 = \{(\boldsymbol{w}^{(i)}, f^{(i)}) \mid i = 1, 2, \cdots, n\} \quad (12)$$

式中：$\mathcal{D}_0$ 为初始观测样本集合；$\boldsymbol{w}^{(i)}$ 为第 i 组待评估的超参数组合；$f^{(i)}$ 为第 i 组参数对应的验证损失值。

步骤 2：构建代理模型，拟合高斯过程，表示为

$$f(\boldsymbol{w}) \sim \mathcal{G}[\zeta(\boldsymbol{w}), k(\boldsymbol{w}, \boldsymbol{w}')] \quad (13)$$

式中：$\mathcal{G}(\cdot)$ 为高斯过程先验模型拟合；$f(\boldsymbol{w})$ 为定义在超参数空间上的损失函数；$\boldsymbol{w}$ 与 $\boldsymbol{w}'$ 分别为超参数空间中任意 2 个样本点；$\zeta(\boldsymbol{w})$ 为高斯过程的先验均值函数，通常设置为 0；$k(\boldsymbol{w}, \boldsymbol{w}')$ 为协方差函数，用以度量 2 组超参数间的相似度。

步骤 3：采集函数选点并用于期望提升，即

$$E(\boldsymbol{w}) = \mathbb{E}[\max\{0, f_{\min} - f(\boldsymbol{w})\}] \quad (14)$$

式中：$E(\boldsymbol{w})$ 为在 $\boldsymbol{w}$ 处的期望提升；$\mathbb{E}[\cdot]$ 为数学期望算子，用于决定下一个待评估的超参数；$f_{\min}$ 为当前已观测到的最小验证损失。

步骤 4：利用采集函数选出使期望提升最大的下一组超参数，表示为 $\boldsymbol{w}_{\text{next}}$。在新的参数组合中重复短周期训练并得到新的损失，表示为 f_{next}。更新观测集为 $\mathcal{D}_{t+1}$，并重复步骤 2～4，直到完成迭代，将验证损失最小的超参数作为最终权重。新观测集 $\mathcal{D}_{t+1}$ 为

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(\boldsymbol{w}_{\text{next}}, f_{\text{next}})\} \quad (15)$$

2 算例搭建

2.1 数据来源

目前，已有许多公开的数据集[21,41]被用于交通事故视频检测任务，如 CADP（Car Accident Detection and Prediction）、HWID12（Highway Incidents Detection Dataset）等，这些数据集包含了丰富的交通事故与非交通事故的视频片段。基于上述文献中总结的数据，本研究在视频网络上额外收集了部分夜间场景下的交通事故视频，以及在该场景下自采集的交通视频数据构成了本研究所有用于训练和测试的数据集。经过视频处理、筛选并按照无监督网络的学习需求，共保留了 1 674 段非事故视频，总计 41 850 帧；保留 342 段事故视频，总计8 550 帧。在试验中，训练阶段所用数据均为公共开源数据集且不含事故数据，共计 874 段非事故视频进行训练；而测试阶段数据集采用部分公开数据集与自采集数据集结合，剩余 800 段非事故视频用于测试并加入事故视频数据。

此外，为减少不同来源数据在分辨率与输入结构上的异质性带来的影响，在数据预处理阶段，将所有视频按其原始帧率切分为静态帧序列，随后对每一帧进行归一化操作，并统一缩放分辨率至 224 像素×224 像素，以兼顾计算效率与图像质量。各部分数据情况如表 1 所示。

表 1 数据集信息

Table 1 Dataset information

数据集	片段数	帧数
非事故（公开数据集 HWID12）	1 074	26 850
非事故（自采集数据集）	600	15 000
事故（公开数据集 CADP＋网络视频）	342	8 550

2.2 试验环境与设置

本研究所使用的计算机环境为 Ubuntu 22.04 操作系统，采用英特尔（R）酷睿（TM）i7-11800H@ 2.30 GHz CPU，NVIDIA GeForce RTX 3060 Laptop GPU。数据处理以及相关代码编写均基于 Python3.10.16 以及 Pytorch 2.6.0 深度学习框架实现。

在训练期间，采取 Adam 优化器，将初始学习率设置为 0.000 1，设置其在 10 个 epoch 内验证集准确率没有提升后下降为原来的一半。训练轮次设置为 200 轮，采取训练早停机制以防止过拟合，当损失函数在连续 5 个 epoch 内没有下降超过 0.1 时，触发早停，迭代结束。针对输入的图像，在训练初始阶段实施归一标准化，选取训练批量为 8。

2.3 评价指标

基于无监督学习范式的异常检测本质上是一个二分类问题，即对正常样本与异常样本进行区分。同时，考虑到本研究任务中正负样本数量差

距较大，构成了较为经典的数据不平衡场景。为此，试验选取对二分类模型判别能力具有较强表征性的受试者工作特征曲线下面积（Area Under the Receiver Operating Characteristic Curve，ROC-AUC）作为主要评价指标，用 S_{ROC} 表示，计算公式为

$$S_{\mathrm{ROC}}=\frac{\sum_{i\in\mathcal{P}}r_i-M_{\mathrm{pos}}(M_{\mathrm{pos}}+1)/2}{M_{\mathrm{pos}}M_{\mathrm{neg}}} \tag{16}$$

式中：$\mathcal{P}$ 为所有事故样本的索引集合；M_{pos} 为正常交通场景样本数量；M_{neg} 为训练阶段或者测试阶段中事故样本的数量；r_i 为第 i 条事故样本在全样本中模型打分降序排列中的排位。

此外，针对样本不平衡的问题，已有研究通常引入精确率-召回率曲线下面积（Area Under the Precision-recall Curve，PR-AUC）以评估模型在不均衡数据上的鲁棒性，用 S_{PR} 表示。因此本研究同时引入 PR-AUC 作为评估指标之一，计算公式为

$$S_{\mathrm{PR}}=\sum_{k=1}^{K-1}\frac{(R_{k+1}-R_k)(P_k+P_{k+1})}{2} \tag{17}$$

式中：K 为不重复阈值候选，且不多于样本总数；P_k 为在第 k 个阈值下模型预测为事故的准确率；R_k 为在第 k 个阈值下事故样本的检出率，即被成功检测为事故的样本占全部事故样本的比例。

两类指标中，S_{ROC} 能够较为全面刻画二元分类器在所有阈值下的表现，而 S_{PR} 能作为其在少数异常样本上的识别能力的体现。

此外，在交通事故检测任务中，受不平衡数据特性干扰，若是直接选择精确率 P_{pre} 等指标进行评价可能会造成结果失真，因此本研究选择召回率 R_{rec} 以及 F1 分数（F_1）作为核心指标，以平衡这一缺陷。计算公式为

$$\begin{cases}R_{\mathrm{rec}}=\dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}\\ P_{\mathrm{pre}}=\dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}}\\ F_1=\dfrac{2P_{\mathrm{pre}}R_{\mathrm{rec}}}{P_{\mathrm{pre}}+R_{\mathrm{rec}}}\end{cases} \tag{18}$$

式中：N_{TP} 为事故样本被正确识别为事故的数量；N_{FP} 为正常交通场景样本被错误识别为事故的数量；N_{FN} 为事故样本被错误识别为正常交通场景的数量。

进一步地，考虑到交通事故检测的实际应用中需要尽可能地检测出全部事故事件，而召回率指标又常被用于衡量模型对异常事件的覆盖能力，因此额外引入 F2 分数（F_2）作为评价指标之一，即

$$F_2=\frac{(1+2^2)P_{\mathrm{pre}}R_{\mathrm{rec}}}{2^2P_{\mathrm{pre}}R_{\mathrm{rec}}} \tag{19}$$

F2 分数是在 F1 的基础上对召回率赋予更高权重的调和平均，其在兼顾精确率的同时更侧重召回率的作用。

3　结果与讨论

3.1　光流估计

为比较深度光流估计算法 RAFT 与传统光流提取算法全变分 L1 范数（Total Variation L1 Regularization，TV-L1）光流估计在夜间交通场景下的表现效果，并进一步验证本研究引入时序建模模块 ConvLSTM 后的改进效果，本研究选取了自采集数据中一个典型的夜间城市道路交通场景[图 4(a)]，分别在 3 种方法下进行光流估计可视化比较（图 4）。图 4(b)～(f)展示了 TV-L1 算法在该场景下的光流提取效果，从图中可以看出，其提取的光流整体较为轻微且模糊，颜色分布较浅，像素块边界明显，较难

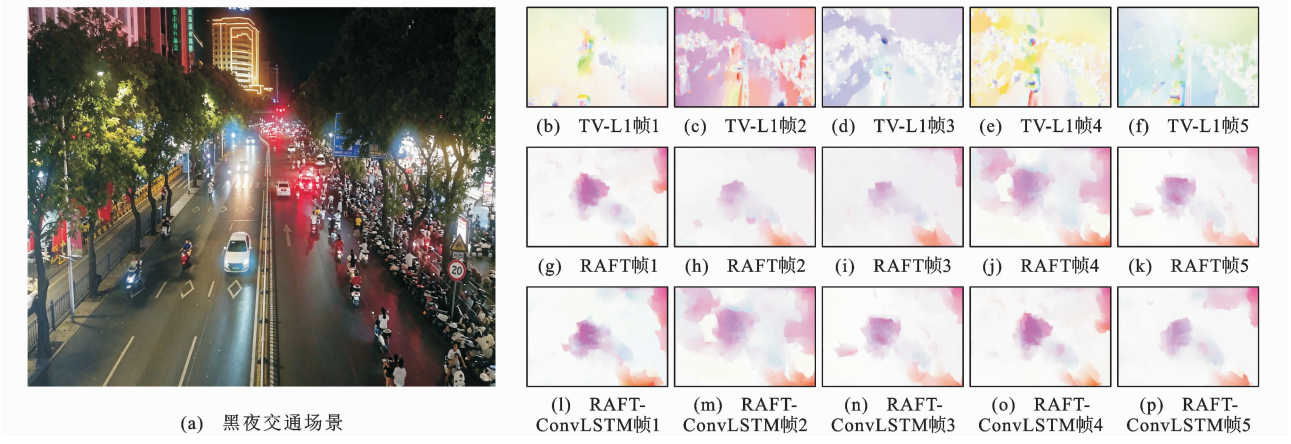


图 4　不同光流提取方法在自采集数据上的提取效果对比

Fig. 4　Comparison of extraction effects of different optical flow estimation methods on self-collected data

准确反映运动主体的细节与方向特征。由图 4(g)～(k)可以看出，相较于 TV-L1，RAFT 提取的光流能够较为准确地捕捉到较细粒度的运动信息，且能够反映运动目标的运动趋势与方向。然而，部分区域仍存在边缘模糊现象，同时道路两侧背景依然呈现出不同程度的误判。而从图 4(l)～(p)可见，在经过 ConvLSTM 处理后，通过建模帧间的时序关联，静止的背景部分被逐渐抑制，每一帧中运动主体的光流信息得到进一步加强和稳定，表明 RAFT 算法结合 ConvLSTM 模块具备一定的可行性。

3.2 消融试验

为了验证本研究所提出的方法对检测效果以及模型性能的影响，通过消融试验验证各部分对总体检测效果的贡献，并验证各模块的必要性以及不同模块对本研究所提模型性能的影响。

(1)基础模型：以经典的 UNet 作为基础模型，通过通道拼接直接传入视频帧序列进行训练。

(2)拓展模型 1：在本研究模型的基础上，将损失函数模块删除，替换为单个损失函数。

(3)拓展模型 2：在本研究模型的基础上，将 VSSM 模块删除，编码器替换为卷积结构。

(4)拓展模型 3：在本研究模型的基础上，删除运动信息模块，使用可见光视频帧序列作为输入。

(5)拓展模型 4：在本研究模型的基础上，删除光流提取网络的 ConvLSTM 模块，将 RAFT 算法所提取的光流帧用于模型输入。

表 2 展示了本次消融试验的结果。可以看出，与基础模型相比，本研究所提出的模型有效提升了在夜间场景下的交通事故检测性能。具体而言，在嵌入运动信息、综合损失函数并将 VSSM 作为编码器的主干架构后，本研究模型相较于基础模型在 ROC-AUC 与 PR-AUC 值上分别提升了 22.6%和 20.3%，表明本文构建的 3 个模块起到的综合性能提升最强。此外，通过与拓展模型 1～3 比较可以发现，损失函数模块对整体模型的性能贡献最大，在删除损失函数模块后 ROC-AUC 值与 PR-AUC 值分别下降了 8.8%和 7.0%，VSSM 模块次之，融合运动信息后对模型整体性能提升最小，证明本文所构建的组合损失及其权重优化机制具有一定的优越性。值得注意的是，在删除光流提取网络中的 ConvLSTM 模块后(即拓展模型 4)，其性能略低于本研究模型，进一步证明融合 ConvLSTM 模块对光流帧的时序建模具有一定的有效性。

表 2 消融试验结果

Table 2 Results of the ablation study

模型	运动信息	VSSM	损失函数	ConvLSTM_module	ROC-AUC	PR-AUC
基础模型	×	×	×	×	0.592	0.562
拓展模型 1	√	√	×	√	0.730	0.695
拓展模型 2	√	×	√	√	0.759	0.769
拓展模型 3	×	√	√	√	0.806	0.750
拓展模型 4	√	√	√	×	0.808	0.752
本研究模型	√	√	√	√	0.818	0.765

为更直观地展现重构异常得分与交通事故图像之间的关联，本研究采用“视频帧-异常得分”曲线图进行可视化。在试验过程中，所有帧均根据其是否包含事故场景被标注为“0”(非事故)或“1”(事故)。图 5 中展示了含有事故的视频帧序列索引，每条曲线表示模型在测试阶段对随机选取的 100 个片段逐帧计算得到的异常得分，得分越高表示该帧出现异常的可能性越大。由图 5 可知，异常得分的波动趋势与事故帧的实际位置具有较高的一致性，异常分数折线图右侧的帧图为折线图中蓝色高亮区域所对应异常得分较高的帧图。

通过计算原始图像与重构图像之间每个像素点的误差，并以热力图的形式可视化误差分布，可以间接评估模型在事故区域的异常检测能力。通常情况下，事故发生区域会表现为高重构误差，即热力图中的高响应区域，从而突出模型在异常检测方面的表现。

图 6 展示了本次试验中 2 种不同的模型(即基础模型与本研究模型)在多个夜间交通事故场景中的重构误差热力图分布情况。从图 6 可以看出：对于部分夜间场景下的交通事故，基础模型未能有效识别出事故区域的异常特征，表现为事故区域在热力图中仅呈现低误差分数的蓝色区域。相比之下，本研究模型能够在不同场景中较为有效地定位到事故发生区域，热力图中对应位置显著呈现高误差分值的红色区域，表明本研究所提出的模型在夜间环境下具备较强的异常检测与事故感知能力。

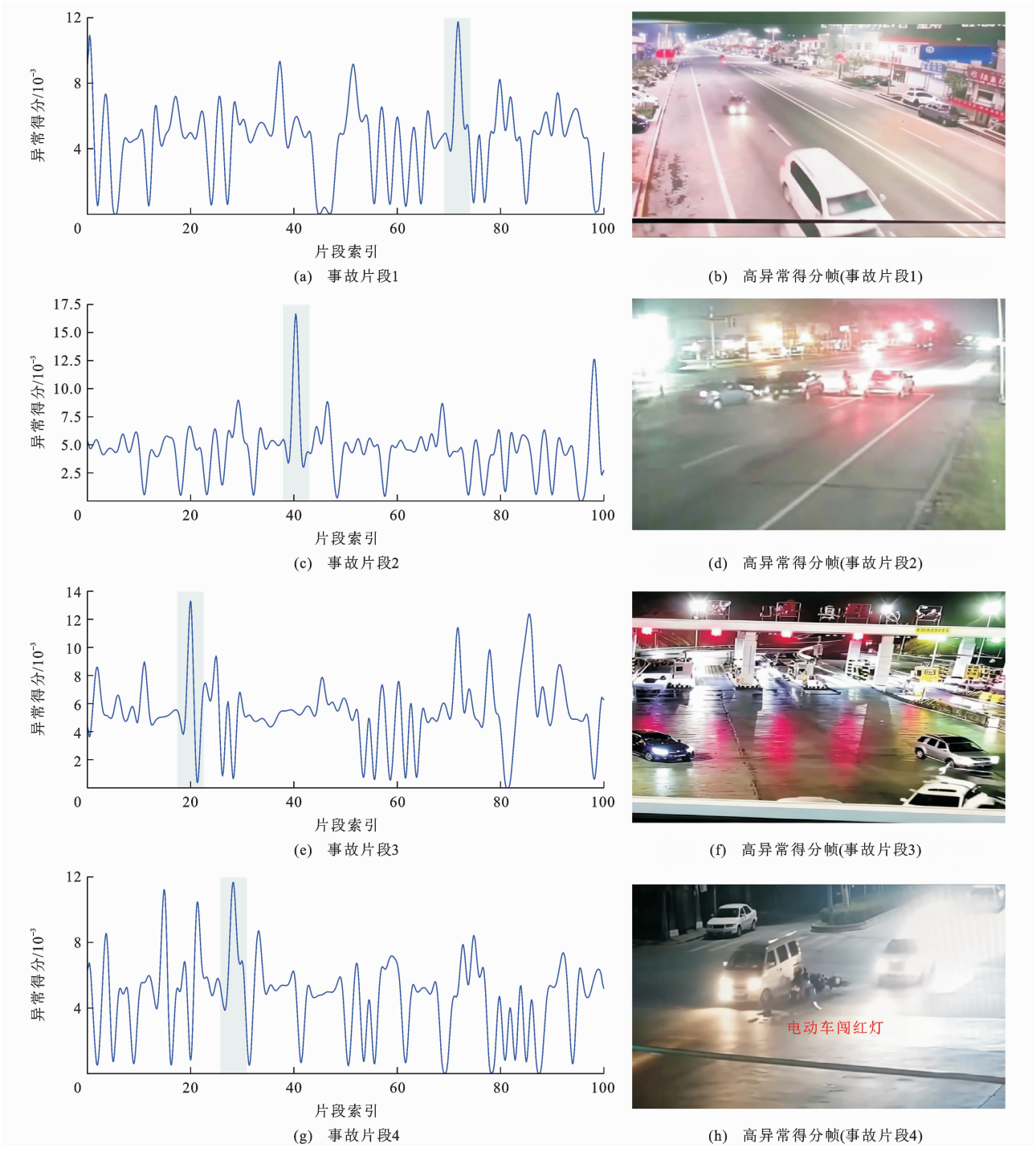


图 5　事故检测异常分数曲线

Fig. 5　Anomaly score curves for accident detection

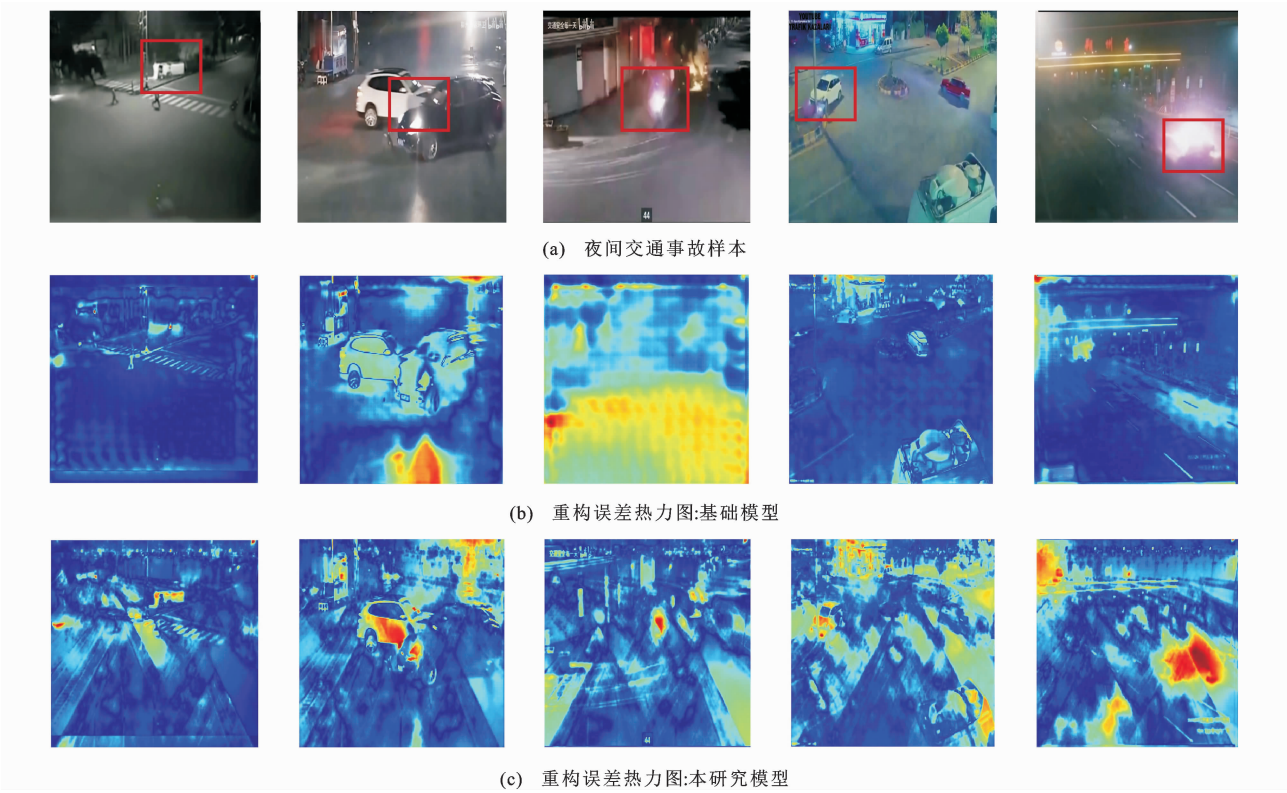
3.3　对比试验

为了验证本研究方法的优越性，选择在异常检测或事故检测相关研究中的模型进行性能以及计算效率方面的比较，包括如下模型。

(1)C3D 模型[42]：将多个三维卷积层进行堆叠而形成的三维卷积神经网络，通过三维卷积对事故和非事故视频进行时空特征的提取，再利用全连接层进行事故的分类检测。

(2)Conv-AE[43]：一种基于二维卷积自编码器的重构模型，通过对正常交通视频帧进行建模，使其在异常帧重构时产生较大的重构误差，从而实现对交通事故的检测。

(3)3DConv-AE[44]：在 Conv-AE 基础上引入三维卷积结构，参考三维卷积网络的时空特征建模思想，使模型能够同时捕捉视频序列中的空间与时间信息，通过时空重构误差实现异常检测。



(a) 夜间交通事故样本

(b) 重构误差热力图:基础模型

(c) 重构误差热力图:本研究模型

图 6 夜间交通事故样本及其重构误差分布热力图

Fig. 6 Accident samples under nighttime conditions and their heatmaps of reconstruction error distributions

(4)MemAE[45]:在自编码器结构中引入记忆模块,利用显著性记忆单元存储正常模式的关键特征,从而增强模型对异常信息的识别能力,避免因自编码器泛化能力过强而重构异常帧。

(5)MNAD[46]:提出基于时序建模的多任务异常检测框架,分别采用预测分支与重构分支学习正常模式的动态演化规律,通过预测误差与重构误差联合判断异常事件。

(6)MONAD[47]:集成多目标深度学习模块与统计异常检测方法,前者用于学习视频中的高维行为表示,后者提供理论支撑的异常评分机制,实现对交通监控视频中异常事件的检测。

在所引入的模型当中,C3D 模型是基于监督学习构建,剩余 5 个模型是基于无监督学习构建。因此为了公平比较模型性能,将前文所介绍的用于训练的非事故数据集和公开事故数据集进行标注,并用于 C3D 模型的训练,其余模型均使用未标注的非事故数据集进行训练。在测试阶段,统一使用本研究构建的测试数据集进行模型测试。表 3 定量地展示了各个模型的性能评价指标。

表 3 各模型性能对比

Table 3 Comparison of the performance of various models

模型	召回率	F1 分数	F2 分数	ROC-AUC	PR-AUC	运算量/GFLOPs	推理速度/(样本·s^{-1})
C3D	0.546	0.636	0.578	0.757	0.722	468.43	9.18
Conv-AE	0.520	0.670	0.571	0.702	0.713	136.72	7.87
3DConv-AE	0.894	0.494	0.675	0.591	0.505	102.96	8.54
MemAE	0.574	0.715	0.623	0.778	0.771	635.83	7.79
MNAD	0.543	0.668	0.587	0.713	0.711	188.12	7.04
MONAD	0.554	0.704	0.606	0.809	0.782	1 271.66	6.82
本研究模型	0.593	0.668	0.621	0.818	0.765	5.19	7.45

通过表 3 可以看出,不同模型在各项评价指标上的表现差异较为显著。3DConv-AE 虽然在召回率上达到了 0.894 的最高水平,但其 F1 分数仅为 0.494,表现出明显不协调。这说明其高召回率并非真实有效,而是伴随着大量误报,导致整体检测性能下降。相比之下,其余模型在 F1 与 F2 分数上相对

更为均衡，展现出较好的检测效果。

值得注意的是，本研究所构建的模型在召回率取得了除 3DConv-AE 以外最优的性能，同时在 F2 分数上取得了仅次于 MemAE 模型的效果，其综合性能较优，避免了 3DConv-AE 单纯追求召回率而牺牲精度的问题。除此之外，本研究所构建的模型在测试数据集上的计算复杂度仅为 5.19 GFLOPs，显著低于其他对比模型，展现出更高的计算效率；在推理速度上保持在 7.45 样本 · s^{-1} 的水平，能够兼顾实时性与准确性。因此，该模型在夜间复杂环境下表现出了更优的综合检测性能与应用潜力。

图 7 展示了每个模型的 ROC-AUC 曲线与 PR-AUC 曲线。结合表 6 并由图 7(a)可知，本研究模型取得了最佳的 ROC-AUC 值，说明本研究模型在夜间场景下进行交通事故检测任务的优越性。3DConv-AE 在本次试验中的性能明显低于其余模型，可能的原因在于其卷积参数量较大，在数据规模较小且不平衡的情况下，无标注的无监督学习相较于监督学习(以 C3D 为例)容易导致训练不充分，此外相较于空间建模(以 Conv-AE 为例)以及额外的显示引导使模型关注异常区域(以 MemAE 与 MNAD 等为例)，较为复杂的模型容易学习到过多无用特征导致训练效果下降。此外，无监督学习中 Conv-AE 与 MNAD 取得了相当的效果，MONAD 取得了除本研究模型外的最佳效果。值得注意的是，C3D 模型作为在本试验中唯一一个监督学习模型，其性能要优于 Conv-AE 和 MNAD，说明在数据相当的情况下，将正反例数据同时训练并加以标注更有利于模型训练。

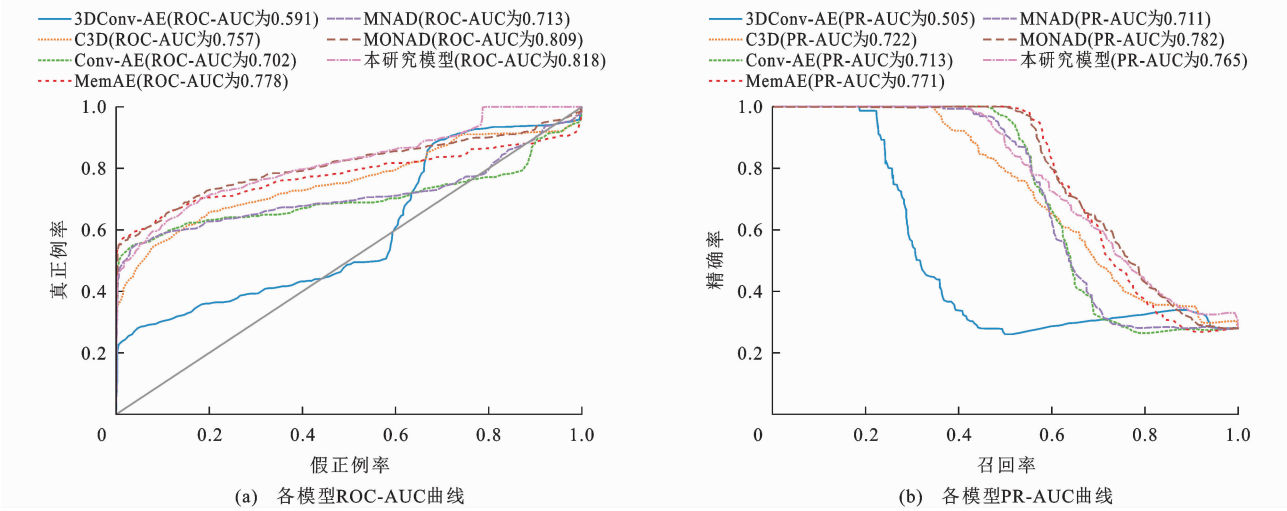


图 7　各类模型的 ROC-AUC 与 PR-AUC 曲线

Fig. 7　ROC-AUC and PR-AUC curves for various models

由图 7(b)可知，在对不平衡数据的鲁棒性方面，本研究模型虽并未取得最高值，但与 MemAE(PR-AUC 为 0.771)和 MONAD(PR-AUC 为 0.782)相差不大，且更优于其余模型，结合 ROC-AUC 值来看其综合稳定性更优。特别是在中高召回率区域，本研究方法的精确率下降趋势相对平缓，说明其在检测多数异常事件的同时，仍能有效控制误报率。此外，相比于传统重构模型如 Conv-AE(PR-AUC 为 0.713)和 3DConv-AE(PR-AUC 为 0.505)，本研究方法在保持高召回的基础上显著提升了精确率，进一步验证了其对异常模式识别的敏感性和判别能力。

4　结　语

(1)本研究提出了一种用于夜间场景下道路交通事故检测网络，通过显示引入光流信息表征交通运动状态，并结合基于视觉状态空间模型的自编码器架构，实现该场景下外观与运动的多特征融合建模思路。试验表明，所提的模型架构能有效提升夜间场景下的交通事故识别，能为相关部门提供一定的技术参考。

(2)在面对如夜间交通事故等复杂环境与不平衡样本的情况下，相较于标签引导或单一特征挖掘的监督学习方法，本研究所提出的基于特征融合与记忆引导的无监督 VSSM-CNN 自编码器表明，无监督学习方法在数据稀缺与复杂条件下具有较强的稳定性与泛化能力。

(3)在光流估计算法的基础上进一步引入时序建模，有效捕捉了细粒度的运动信息，并逐帧抑制静止物体的运动信息表达，有效提升了光流估计的准确性。结合三重损失函数与贝叶斯优化机

制，能够有效增强模型在事故区域的可分性与鲁棒性，验证了本研究方法在夜间场景下的适用性与稳定性。

(4)未来的研究应将重心聚焦于夜间叠加恶劣天气等复杂场景的事故检测，考虑融合更多场景的事故数据进行训练，并引入图像增强、消除无关背景等方法进一步提升特征提取能力，从而提升事故检测精度。

参 考 文 献：

References：

[1] 王 晨，周 威，严隽逸，等. 一种用于道路交通事故自动检测的改进双流网络[J]. 中国公路学报，2023，36(5)：185-196.
WANG Chen，ZHOU Wei，YAN Jun-yi，et al. Improved two-stream network for vision-based traffic accident detection[J]. China Journal of Highway and Transport，2023，36(5)：185-196.

[2] 杨 洋，王文慧，吴先宇，等. 高速公路非常规交通事故研究综述[J]. 应用基础与工程科学学报，2024，32(3)：601-626.
YANG Yang，WANG Wen-hui，WU Xian-yu，et al. Review of the research toward freeway unconventional traffic accidents[J]. Journal of Basic Science and Engineering，2024，32(3)：601-626.

[3] WANG W C，YANG Y，YANG X B，et al. A negative binomial Lindley approach considering spatiotemporal effects for modeling traffic crash frequency with excess zeros[J]. Accident Analysis & Prevention，2024，207：107741.

[4] 黄思德，黄荣军，张岩坤，等. 基于监控视频的交通事故自动识别算法[J]. 公路交通科技，2024，41(1)：169-176.
HUANG Si-de，HUANG Rong-jun，ZHANG Yan-kun，et al. An algorithm for automatic traffic incident detection based on surveillance video[J]. Journal of Highway and Transportation Research and Development，2024，41(1)：169-176.

[5] 郭延永，刘 佩，袁 泉，等. 网联自动驾驶车辆道路交通安全研究综述[J]. 交通运输工程学报，2023，23(5)：19-38.
GUO Yan-yong，LIU Pei，YUAN Quan，et al. Review on research of road traffic safety of connected and automated vehicles[J]. Journal of Traffic and Transportation Engineering，2023，23(5)：19-38.

[6] 公安部办公厅，国家卫生健康委办公厅. 关于健全完善道路交通事故警医联动救援救治长效机制的通知：公交管〔2020〕161号[EB/OL]. (2020-07-02)[2025-08-04]. http：//www.nhc.gov.cn/vzvgj/s3594q/202007/c41e38d7db744ad9890ecb5d710ab59b.shtml.
General Office of the Ministry of Public Security，General Office of the National Health Commission. Notice on improving the long-term mechanism of police-medical joint rescue and treatment for road traffic accidents：Gong Jiao Guan〔2020〕No. 161[EB/OL]. (2020-07-02)[2025-08-04]. http：//www.nhc.gov.cn/vzvgj/s3594q/202007/c41e38d7db744ad9890ecb5d710ab59b.shtml.

[7] 阎 莹，王玉莹，周 旋，等. 基于混合模型的自动驾驶车辆事故严重程度影响因素分析[J]. 交通运输工程学报，2025，25(1)：184-196.
YAN Ying，WANG Yu-ying，ZHOU Xuan，et al. Analysis of autonomous vehicle accident severity factors based on hybrid model[J]. Journal of Traffic and Transportation Engineering，2025，25(1)：184-196.

[8] 杨 洋，贺 昆，王云鹏，等. 面向动态交通流的高速公路事故风险模型空间移植研究[J]. 交通运输系统工程与信息，2023，23(3)：174-186.
YANG Yang，HE Kun，WANG Yun-peng，et al. Spatial transplantation for modeling of freeway traffic crash risk based on dynamic traffic flow[J]. Journal of Transportation Systems Engineering and Information Technology，2023，23(3)：174-186.

[9] 牛世峰，母俊杰，濮泽昱，等. 基于错时序因果变量和分目标训练的车辆事故风险分类分层评估方法[J]. 交通运输工程学报，2025，25(6)：293-306.
NIU Shi-feng，MU Jun-jie，PU Ze-yu，et al. Hierarchical assessment method for vehicle accident risk classification based on temporal misalignment causal variables and sub-objective training[J]. Journal of Traffic and Transportation Engineering，2025，25(6)：293-306.

[10] 杨 洋，袁振洲，王印海，等. 基于 WOMDI-Apriori 算法的高速公路交通事故风险识别[J]. 交通工程，2021，21(6)：1-10，16.
YANG Yang，YUAN Zhen-zhou，WANG Yin-hai，et al. Freeway crash risk identification based on a new improved method of WOMDI-Apriori algorithm[J]. Journal of Transportation Engineering，2021，21(6)：1-10，16.

[11] BATANINA E，BEKKOUCH I E I，YOUSSRY Y，et al. Domain adaptation for car accident detection in videos[C]// IEEE. 2019 Ninth International Conference on Image Processing Theory，Tools and Applications (IPTA). New York：IEEE，2019：1-6.

[12] FANG J W，QIAO J H，BAI J，et al. Traffic accident detection via self-supervised consistency learning in driving scenarios[J]. IEEE Transactions on Intelligent Transportation Systems，2022，23(7)：9601-9614.

[13] PASHAEI A，GHATEE M，SAJEDI H. Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images[J]. Journal of Real-time Image Processing，2020，17(4)：1051-1066.

[14] IJJINA E P，CHAND D，GUPTA S，et al. Computer vision-based accident detection in traffic surveillance[C]// IEEE. 2019 10th International Conference on Computing，Communication and Networking Technologies (ICCCNT). New York：IEEE，2019：1-6.

[15] JIANG Y，WANG Y H，ZHAO M H，et al. Nighttime traffic object detection via adaptively integrating event and frame domains[J]. Fundamental Research，2025，5(4)：1633-1644.

[16] YUE Y G, ZHANG S, WU Z Y, et al. Research on nighttime road visibility monitoring based on video images[J]. Traffic Injury Prevention, 2026, 27(3): 287-295.

[17] HOSSAIN A, SUN X D, SHAHRIER M, et al. Exploring nighttime pedestrian crash patterns at intersection and segments: Findings from the machine learning algorithm[J]. Journal of Safety Research, 2023, 87: 382-394.

[18] REN J S, WANG W, WANG J W, et al. An unsupervised feature learning approach to improve automatic incident detection [C] // IEEE. 2012 15th International IEEE Conference on Intelligent Transportation Systems. New York: IEEE, 2012: 172-177.

[19] YAO Y, XU M Z, WANG Y C, et al. Unsupervised traffic accident detection in first-person videos [C] // IEEE. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). New York: IEEE, 2019: 273-280.

[20] SINGH D, MOHAN C K. Deep spatio-temporal representation for detection of road accidents using stacked autoencoder[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(3): 879-887.

[21] ZHOU W, WEN L H, ZHAN Y F, et al. An appearance-motion network for vision-based crash detection: Improving the accuracy in congested traffic[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(12): 13742-13755.

[22] CHEN P, ZHANG W W, XIAO Z Y, et al. Traffic accident detection based on deformable frustum proposal and adaptive space segmentation[J]. Computer Modeling in Engineering & Sciences, 2022, 130(1): 97-109.

[23] 张　奇，周　威，胡伟超，等. 融合递进式域适应和交叉注意力的事故检测方法[J]. 计算机工程与应用，2025，61(6)：349-360.
ZHANG Qi, ZHOU Wei, HU Wei-chao, et al. Accident detection method integrating progressive domain adaptation and cross-attention[J]. Computer Engineering and Applications, 2025, 61(6): 349-360.

[24] TRAN T M, BUI D C, NGUYEN T V, et al. Transformer-based spatio-temporal unsupervised traffic anomaly detection in aerial videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 8292-8309.

[25] LIANG R Q, LI Y M, YI Y X, et al. A memory-augmented multi-task collaborative framework for unsupervised traffic anomaly detection in driving videos[J]. Pattern Recognition, 2025, 168: 111789.

[26] LUO W X, LIU W, LIAN D Z, et al. Video anomaly detection with sparse coding inspired deep neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(3): 1070-1084.

[27] ZHANG Y, NIE X S, HE R D, et al. Normality learning in multispace for video anomaly detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(9): 3694-3706.

[28] QIE K, WANG J Y, LI Z H, et al. Recognition of occluded pedestrians from the driver's perspective for extending sight distance and ensuring driving safety at signal-free intersections[J]. Digital Transportation and Safety, 2024, 3(2): 65-74.

[29] 石洋宇，谢承杰，郑棣文，等. 基于 Mamba-CNN 的多尺度异常行为检测方法[J/OL]. 北京航空航天大学学报，(2024-11-15). https://doi.org/10.13700/j.bh.1001-5965.2024.0416.
SHI Yang-yu, XIE Cheng-jie, ZHENG Di-wen, et al. Multi-scale anomaly behavior detection method based on Mamba-CNN[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, (2024-11-15). https://doi.org/10.13700/j.bh.1001-5965.2024.0416.

[30] PENG Z L, GUO Z H, HUANG W, et al. Conformer: Local features coupling global representations for recognition and detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 9454-9468.

[31] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C] // IEEE. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2021: 16514-16524.

[32] TEED Z, DENG J. RAFT: Recurrent all-pairs field transforms for optical flow[C] // ECCV. Computer Vision-ECCV 2020. Berlin: Springer International Publishing, 2020: 402-419.

[33] JIAO J B, LIU Y, LIU Y F, et al. VMamba: Visual state space model[C] // NeurIPS. Advances in Neural Information Processing Systems 37. Vancouver: Curran Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 103031-103063.

[34] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning optical flow with convolutional networks [C] // IEEE. 2015 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2015: 2758-2766.

[35] SUN D Q, YANG X D, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume[C] // IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2018: 8934-8943.

[36] CHEN H J, WANG Z, QIN H D, et al. CFDHI-net: Correlation-driven feature decoupling and hierarchical integration network for RGB-thermal semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(10): 17173-17184.

[37] FAN J, CHEN M, GU Z Y, et al. SSIM over MSE: A new perspective for video anomaly detection[J]. Neural Networks, 2025, 185: 107115.

[38] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C] // IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 1125-1134.

[39] KLEIN A, FALKNER S, BARTELS S, et al. Fast Bayesian optimization of machine learning hyperparameters on large datasets[C] // PMLR. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Fort

Lauderdale: PMLR, 2017: 528-536.

[40] WU J, TOSCANO-PALMERIN S, FRAZIER P I, et al. Practical multi-fidelity Bayesian optimization for hyperparameter tuning[C]// PMLR. Proceedings of the 35th Uncertainty in Artificial Intelligence Conference. Fort Lauderdale: PMLR, 2020: 788-798.

[41] SHAH A P, LAMARE J B, NGUYEN-ANH T, et al. CADP: A novel dataset for CCTV traffic camera based accident analysis[C]// IEEE. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). New York: IEEE, 2018: 1-9.

[42] HUANG X H, HE P, RANGARAJAN A, et al. Intelligent intersection: Two-stream convolutional networks for real-time near-accident detection in traffic video [J]. ACM Transactions on Spatial Algorithms and Systems, 2020, 6(2): 1-28.

[43] HASAN M, CHOI J, NEUMANN J, et al. Learning temporal regularity in video sequences [C] // IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 733-742.

[44] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks[C]// IEEE. 2015 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2015: 4489-4497.

[45] GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]// IEEE. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 1705-1714.

[46] PARK H, NOH J, HAM B. Learning memory-guided normality for anomaly detection[C]// IEEE. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 14372-14381.

[47] DOSHI K, YILMAZ Y. Online anomaly detection in surveillance videos with asymptotic bounds on false alarm rate[C]// Springer International Publishing. Pattern Recognition-ICPR International Workshops and Challenges. Berlin: Springer International Publishing, 2021: 292-304.