暴雨灾害
Torrential Rain and Disasters
中国科技核心期刊
中国科技论文统计源期刊

基于深度学习模型Vit-Bi-LSTM的监控相机在夜间强降雨监测的应用初探

王兴 黄浩 周昂 杨正玮 张崔岩

Preliminary exploration of nighttime heavy rainfall detection using surveillance cameras based on deep learning model ViT-Bi-LSTM

WANG Xing，HUANG Hao，ZHOU Ang，YANG Zhengwei，ZHANG Cuiyan

在线阅读 View online: https://doi.org/10.12406/byzh.2025-089

您可能感兴趣的其他文章 Articles you may be interested in

基于深度学习的降水临近预报方法及其在2023年江苏汛期的应用评估

Evaluation of deep learning-based precipitation nowcasting methods during the 2023 flood season in Jiangsu Province

暴雨灾害. 2025, 44(1): 1-8 https://doi.org/10.12406/byzh.2023-206

深度学习方法在强降水临近预报技术中的应用及其改进思考

Thoughts on application and improvement of deep learning in severe precipitation nowcasting technology

暴雨灾害. 2022, 41(5): 506-514 https://doi.org/10.12406/byzh.2022-079

一种基于迁移学习和长短期记忆神经网络的降水预报方法

A precipitation forecast method based on transfer learning and Long Short Term Memory

暴雨灾害. 2024, 43(1): 45-53 https://doi.org/10.12406/byzh.2023-118

武汉和宜昌$PM_{2.5}$和O_3复合污染气象特征分析及其浓度预测模型研究

Study on the characteristics of $PM_{2.5}$ and O_3 compound air pollution and concentration prediction model in Wuhan and Yichang

暴雨灾害. 2025, 44(3): 410-420 https://doi.org/10.12406/byzh.2024-165

区域性暴雨过程评价指标在湖北的本地化修订与应用

Localized revision and application of assessing indices for regional heavy rainfall events in Hubei Province

暴雨灾害. 2020, 39(5): 470-476 https://doi.org/10.3969/j.issn.1004-9045.2020.05.005

2020年7月4—10日长江流域极端强降雨过程MCS活动特征分析

Analysis of MCS activity characteristics during extreme heavy rainfall in the Yangtze River Basin from July 4 to 10, 2020

暴雨灾害. 2020, 39(6): 593-602 https://doi.org/10.3969/j.issn.1004-9045.2020.06.007

关注微信公众号，获得更多资讯信息

王兴,黄浩,周昂,等.2025.基于深度学习模型 Vit-Bi-LSTM 的监控相机在夜间强降雨监测的应用初探[J].暴雨灾害,44(5):592−601.
WANG Xing, HUANG Hao, ZHOU Ang, et al. 2025. Preliminary exploration of nighttime heavy rainfall detection using surveillance cameras based on deep learning model ViT-Bi-LSTM [J]. Torrential Rain and Disasters,44(5):592−601 (in Chinese). doi:10.12406/byzh.2025−089

基于深度学习模型 Vit-Bi-LSTM 的监控相机在夜间强降雨监测的应用初探

王兴[1, 2, 3], 黄浩[2], 周昂[2], 杨正玮[2], 张崔岩[1, 3]

(1. 南京工程学院计算机工程学院, 南京 211167; 2. 南京大学大气科学学院, 南京 210023; 3. 江苏省智能感知技术与装备工程研究中心, 南京 211167)

摘　要: 城市监控相机具有数量多、密度大、传输快的观测优势, 为降雨的高时空分辨率观测提供了新的契机。然而, 现有监控测雨研究聚焦于对白天降雨的监测, 缺少对夜间强降雨的有效测量方法。本文从监控视角出发, 分析了夜间降雨视频的时空特征, 提出了一种融合 Vision Transformer (ViT) 和双向长短时记忆网络 (Bi-LSTM) 的深度学习模型 Vit-Bi-LSTM, 用于夜间降雨强度等级的自动识别。基于 2022—2025 年期间收集的 41 场夜间降雨视频, 构建了一个最大雨强达 80 mm·h^{-1} 的夜间降雨视频数据集, 涵盖 4 500 个视频片段, 总时长达 37.5 h, 利用 Vit-Bi-LSTM 模型对视频数据集开展降雨强度等级识别训练并检验效果。结果表明, 基于降雨视频时空特征的联合建模可有效改善降雨识别精度, 且采用两层 Bi-LSTM 结构的 Vit-Bi-LSTM 模型在自建数据集上取得了最高 85.6% 的精度表现。Vit-Bi-LSTM 对夜间强降雨的实地观测中取得了 76.7% 的降雨等级识别精度, 证明了其在真实环境中的有效性。虽然 Vit-Bi-LSTM 对小于 20 mm·h^{-1} 降雨取得了 80.7%~90.6% 的识别精度, 但随着降雨强度的提升, 特别是降雨强度大于 40 mm·h^{-1} 时, 精度下降至 68.8%~75.0%。

关键词: 强降雨监测; 监控相机; 深度学习; 时空特征

中图法分类号: P412.13　　**文献标志码**: A　　**DOI:** 10.12406/byzh.2025−089

Preliminary exploration of nighttime heavy rainfall detection using surveillance cameras based on deep learning model ViT-Bi-LSTM

WANG Xing[1, 2, 3], HUANG Hao[2], ZHOU Ang[2], YANG Zhengwei[2], ZHANG Cuiyan[1, 3]

(1. *School of Computer Engineering*, *Nanjing Institute of Technology*, *Nanjing* 211167; 2. *School of Atmospheric Science*, *Nanjing University*, *Nanjing* 210023; 3. *Jiangsu Province Engineering Research Center of IntelliSense Technology and System*, *Nanjing* 211167)

Abstract: Urban surveillance cameras, with their large number, high density, and fast transmission capabilities, offer new opportunities for high spatiotemporal resolution rainfall monitoring. However, existing research has mainly focused on daytime rainfall monitoring, with nighttime heavy rainfall monitoring remaining relatively weak. In this study, we analyze the spatiotemporal features of nighttime rainfall videos and propose a deep learning model, Vit-Bi-LSTM, which combines Vision Transformer (ViT) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks, for automatic classification of nighttime heavy rainfall levels. A nighttime rainfall video dataset with a maximum rainfall intensity of 80 mm·h^{-1} was constructed from 4 500 clips totaling 37.5 hours, which included 41 nighttime rainfall videos collected between 2022 and 2025. Utilize the Vit-Bi-LSTM model to conduct rainfall intensity level recognition experiments on video datasets. The results indicate that the joint modeling of spatiotemporal features significantly improves the accuracy of rainfall level classification, with the two-layer Bi-LSTM structure achieving an accuracy of 85.6% on the self-constructed dataset. Additionally, the

收稿日期: 2025−05−12; **定稿日期**: 2025−06−30

资助项目: 国家自然科学基金项目 (42422501, 42405140, 42305006); 中国博士后基金项目 (2024M761383); 南京工程学院校级科研基金项目 (YKJ.202315)

第一作者: 王兴, 主要从事应用气象及计算机视觉方法研究。E-mail: jwangxing0719@163.com

通信作者: 黄浩, 主要从事星-地基多波段雷达反演理论技术及台风降水研究。E-mail: huanghao@nju.edu.cn

© Editorial Office of *Torrential Rain and Disasters*. OA under CC BY−NC−ND 4.0

model achieved an accuracy of 76.7% in field observations of two nighttime heavy rainfall events, demonstrating its effectiveness in real-world environments. Although the model showed high accuracy (80.7%~90.6%) for rainfall intensities below 20 mm·h^{-1}, the accuracy decreased to 68.8%~75.0% for rainfall intensities above 40 mm·h^{-1}.

Key words: heavy rainfall monitoring; surveillance cameras; deep learning; spatiotemporal feature

引　言

近年来, 我国城市极端暴雨频发, “城中看海”问题频频出现, 给人民安全、社会发展带来巨大威胁(黄华兵等, 2021; 陈刚等, 2022)。高时空分辨率的降水产品对城市内涝的精细建模、准确分析、科学管理至关重要 (林珲等, 2022; Huang et al., 2023; 盛志军等, 2025), 但现有的降雨观测体系仍难以满足城市内涝等快速响应水文模式对高分辨率降雨产品的需求, 亟需探索新型降水观测手段 (McCabe et al., 2017; 夏军等, 2018; 周月华等, 2019; 徐宗学和叶陈雷, 2021)。智慧城市的发展使得监控相机成为交通、安防、城管等部门的必须。监控相机主要由公安、交通、市政等部门部署, 广泛安装于道路交叉口、建筑外墙、街区主干道及其他公共区域, 具有数量多、密度大、分布广、传输快的观测优势, 为城市降雨的高时空分辨率观测提供新的途径 (王兴等, 2021; Wang et al., 2023a)。此外, 降雨测量任务可在现有的城市监控资源上实现, 无需额外的硬件投入, 布设和维护成本较低。

本文所述的测雨相机为常规市售的监控摄像设备。经调研发现, 目前主流监控相机的图像分辨率普遍高于 1 280×960, 具备较高的图像清晰度, 能够有效记录降雨的发生及降雨强度变化过程。因此, 普通监控相机具备执行降雨测量任务的能力。虽然监控相机主要并非用于气象监测, 但气象部门可以通过政务共享或科研合作等方式获取其视频资料, 从而为城市降雨过程的实时观测与后期分析提供数据支撑。国外已有部分研究尝试利用降水天气过程的监控视频开展定性或定量降雨监测 (Yan et al., 2023; Rajabi et al., 2024; Wang et al., 2025), 目前国内未见此类研究。

视频中雨滴的运动轨迹蕴含雨滴的大小、数量等信息, 是降雨强度反演的关键因子。受监控相机曝光时间较长的影响, 雨滴运动轨迹呈线状 (Li et al., 2021; Byun et al., 2023), 故下文简称雨线。监控测雨的本质是建立“视频雨线-降雨强度”之间的映射关系 (Yan et al., 2023), 现已取得部分成果, 如 Garg 和 Nayar(2007) 开创性地建立降雨的视频成像模型, 在此基础上提出基于相机参数的雨线提取算法, 并依据雨线密度和降雨强度的线性关系实现降雨强度的估测; Allamano 等 (2015) 结合气象学知识与相机的成像原理, 建立雨滴尺寸的推算模型, 突破性地实现雨滴谱的计算; Dong 等 (2017) 采用计算机视觉算法区分视频中聚、散焦区域的雨线, 进一步提升降雨强度计算精度; Jiang 等 (2019) 提出基于稀疏编码的雨线提取算法, 优化已有的降雨反演方案 (Allamano et al., 2015), 实现基于监控相机的降雨反演; Wang 等 (2022) 引入地理学、气象学等先验知识, 采用模糊数学方法建立基于多相机的区域降雨反演模型。近年来, 随着深度学习在计算机视觉领域取得的巨大突破, 部分学者将其应用于监控测雨中, 如 Zen 等 (2019) 以降雨图片为输入, 基于卷积神经网络 (Convolutional Neural Networks, CNN) 建立降雨强度预测模型; Lin 等 (2020) 将 3D-CNN 与卷积长短期记忆网络 (ConvLSTM) 相结合挖掘监控视频中的降雨信息; Saputro 等 (2023) 评估了经典的 VGG16 和 ResNet32 深度学习模型对降雨强度预测的表现。

然而, 现有监控测雨工作研究主要围绕白天降雨开展。由于监控相机在白天通过可见光成像, 而夜间则通过主动发射近红外光的方式采集视频数据, 导致日夜雨滴的成像机理存在本质差异, 降雨的图像/视频特征存在较大区别, 限制了其在夜间的应用效果 (Wang et al., 2023b)。近年来, 夜间降雨测量开始被关注, 如 Lee 等 (2023) 采用最近邻 (KNN) 算法提取夜间监控视频雨线, 然后利用白天降雨反演模型 (Allamano et al., 2015; Jiang et al., 2019) 估算降雨强度; Wang 等(2023b) 建立基于张量的雨线提取算法, 在此基础上采用 1D-CNN 模型估算降雨强度。但上述方法主要针对夜间常规降雨场景设计, 夜间强降雨过程中监控视频内容剧烈变化, 致使上述方法在强降雨监测场景中难以保持有效性。为实现对夜间强降雨的可靠监测, 本文提出了 Vision Transformer (ViT) (Yin et al., 2022) 与双向长短期记忆网络 (Bi-LSTM) (Yang and Wang, 2022) 融合的降雨等级识别深度学习模型 ViT-Bi-LSTM, 基于 2022—2025 年监控视频建立了 41 场夜间强降雨视频数据集, 并在两次降雨场景中评估了模型的有效性, 为改善监控相机对夜间强降雨的测量能力提供新的解决方案。

1 数据与方法

1.1 夜间降雨监控视频数据集构建

为满足深度学习模型的训练与测试需求，本文构建了一个面向夜间强降雨场景的监控视频数据集。数据集的制作过程包括：

(1) 监控视频数据采集。在南京市江宁区布设10 台具有不同拍摄参数的监控相机拍摄夜间降雨过程。共使用 3 种不同类型的监控相机，在本文中，会影响到雨滴图像特征的核心主要参数体现在表 1 所列举的焦距、帧率、曝光时间和分辨率等参数，其中，焦距为所使用相机的关键差别。各相机所处监控背景各异，以确保视频场景的多样性。自 2022 年 3 月—2025 年 3 月，在不同季节中共采集了 41 场夜间降雨事件，涵盖从小雨到暴雨等级的多种降雨强度类型。降雨事件筛选时重点考虑强度等级的代表性，确保不同等级的降雨样本在数据集中分布相对均衡，以增强模型对各类降雨情形的识别能力。将原始监控视频按 30 s 时长划分为片段，构建视频样本集。

表 1 研究使用的监控相机的类型、关键参数及部分场景

Table 1 Key parameters of surveillance cameras used in this study

相机类型	关键参数				数量/台	部分监控场景
	焦距/mm	帧率/fps	曝光时间/s	图像分辨率/pixels		
类型1 (CS-C5S-3C2WFR)	4	25	0.04	1 920×1 080	4	相机安装于楼顶，离地高度约12 m，拍摄角度为俯视，俯视角约为30°~60°。场景以建筑区域为主。
类型2 (CS-C5HC-3G2WFL1)	6	25	0.04	1 920×1 080	3	相机布设于建筑楼顶，离地高度约15 m，拍摄角度接近平视，视野开阔。场景主要覆盖树木、绿化带等。
类型3 (CS-H5-V201-1H3WKFL)	8	25	0.04	1 920×1 080	3	相机安装于建筑外墙，距离地面高度约10 m，拍摄角度亦为平视。场景以城市稀疏建筑区域为主。

(2) 降雨等级标签获取。根据降雨强度将其划分为不同等级。具体而言，以 10 mm·h^{-1} 为间隔定义降雨等级(如：Level_0 表示无雨，Level_1 表示降雨强度为 0~10 mm·h^{-1}，Level_2 表示 10~20 mm·h^{-1}，依此类推)，在 0~80 mm·h^{-1} 的范围内共划分为 9 个等级。因视频样本量受限，大于等于 80 mm·h^{-1} 降雨视频未被考虑在内。降雨强度参考数据来源于国际上广泛使用的二维视频雨滴谱仪 (Two-Dimensional Video Disdrometer, 2-DVD) 的同步观测结果 (Kruger and Krajewski, 2002)。2-DVD 与各监控摄像机之间的最大水平距离不超过 200 m，垂直距离低于 6 m，确保了视频样本与对应标签之间降雨场景的一致性。2-DVD 的降雨强度计算频率设定为 1 min 一次，对于同一分钟内采集的多个视频片段，统一赋予该时间段对应的降雨等级标签。最终，每个降雨等级包含 500 个视频片段，构成模型训练数据集，并按照 7∶2∶1 的比例划分为训练集、验证集和测试集。

1.2 夜间降雨视频特征

在夜间监控视频中，降雨通常以雨线的形式表现，其时空特征在不同降雨强度下存在显著差异。从时间维度看，降雨过程中雨线随时间动态变化，在常规降雨条件下，这种变化较为平缓，周期性稳定，有利于通过时序分析提取连续性特征 (Garg and Nayar, 2007)；而在强降雨时，雨线频繁涌现，帧间变化剧烈，伴随大量重叠与遮挡，使得其时序结构更加复杂。参考《降水量等级》规定 (国家气象中心，2012)，本文将雨强<20 mm·h^{-1} 定义为一般性降雨，将雨强≥20 mm·h^{-1} 定义为强降雨。从空间维度看，如图 1 所示，一般性降雨下的雨线数量适中、分布稀疏，与背景对比明显，便于从单帧图像中提取其数量、形状和像素尺度等空间特征，进而用于估算降雨强度与雨量；但在强降雨条件下，雨线空间上呈现密集重叠状态，图像内容剧烈变化，背景被遮挡严重，使得雨线轮廓难以分辨，空间结构信息严重丢失，极大干扰了现有基于图像的降雨识别与测量方法有效性。因此，如何突破强降雨条件下雨线特征提取困难这一瓶颈，已成为提升监控相机在实际降雨测量中应用价值的关键问题。

鉴于本文使用深度学习模型开展研究，高质量、标注准确的降雨视频数据集是确保模型性能和实用性的基础保障。由于全尺度、多监控场景的强降雨视频收集周期较长，而且强降雨过程中通常伴随大风等因素干扰，降雨强度绝对真值获取存在较大困难。为规避上述问题，本文将降雨强度划分为不同的等级，通过降雨等级的识别替代了降雨强度的直接测量，确保夜间降雨的监测准确性与可靠性。综

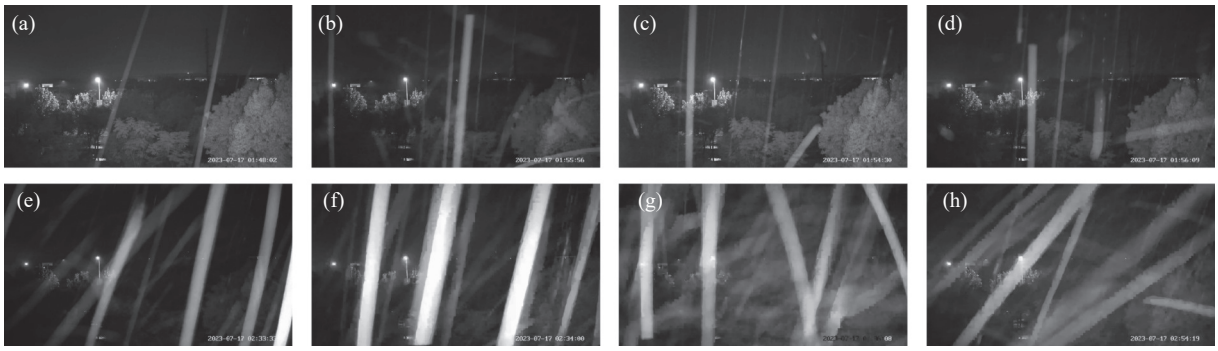


图 1 夜间不同降雨强度 (a: 2 mm·h^{-1}; b: 8 mm·h^{-1}; c: 12 mm·h^{-1}; d: 15 mm·h^{-1}; e: 43 mm·h^{-1}; f: 51 mm·h^{-1}; g: 69 mm·h^{-1}; h: 75 mm·h^{-1}) 视觉特征差异的表现

Fig. 1 Variation in visual features under different nighttime rainfall intensities (a: 2 mm·h^{-1}, b: 8 mm·h^{-1}, c: 12 mm·h^{-1}, d: 15 mm·h^{-1}, e: 43 mm·h^{-1}, f: 51 mm·h^{-1}, g: 69 mm·h^{-1}, h: 75 mm·h^{-1})

上所述, 本文将夜间降雨测量任务转化为对降雨监控视频的分类任务。

1.3 ViT-Bi-LSTM 降雨等级识别模型

针对夜间强降雨视频时空维度特征, 本文将 ViT 与 Bi-LSTM 相结合, 提出名为 Vit-Bi-LSTM 的视频降雨测量方法。如图 2 所示, 该模型使用 ViT 在图像空间特征提取方面的优势 (Yuan et al., 2021) 与 Bi-LSTM 对视频时间序列建模的能力 (Mughees et al., 2021), 充分挖掘视频序列中的时空信息, 实现夜间强降雨等级的判别。

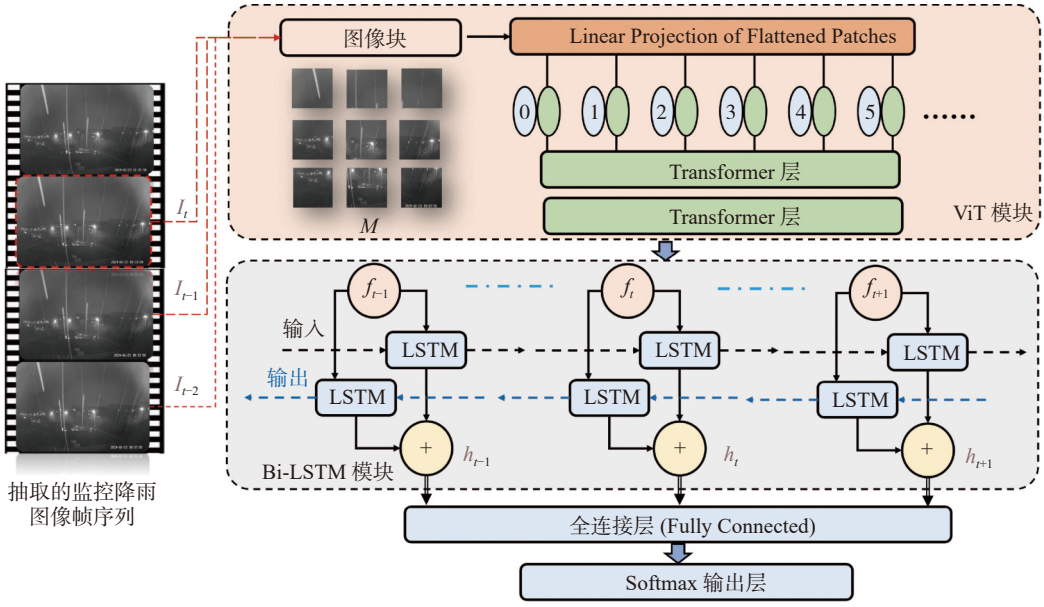


图 2 Vit-Bi-LSTM 网络结构图

Fig. 2 Framework of the Vit-Bi-LSTM Network

1.3.1 基于 ViT 的空间特征提取模块

监控视频经预处理为统一长度的视频片段后, 按固定间隔抽取N帧图像序列。设第 I 帧图像为 $I \in R^{H\times W\times 3}$, R为视频多维向量, H与W分别为图像高、宽, 3 表示 RGB 三通道。ViT 将图像划分为$P\times P$尺寸的图像块 (patch), 共形成 $M = H\times W/P^2$个图像块。每个图像块通过线性投影映射至D维特征向量, 再拼接位置编码, 作为 Transformer 编码器输入。Transformer 利用多头自注意力机制 (Multi-Head Self Attention, MHSA) 对图像全局特征进行建模, 自注意力机制 (Attention) 的计算公式如下

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)V \tag{1}$$

式中, Q, K, V分别为查询、键、值矩阵, d_{k}为键向量维度, softmax 为神经网络中特殊层常用称呼, 用于类别的区分, T 表示转置计算。

经过 2 层 Transformer 编码器, 提取每帧图像的空间特征$f_i \in R^D$。为适配时序建模, 将 ViT 去除原有分类头, 输出中间层 768 维特征向量。

1.3.2 基于 Bi-LSTM 的时间特征提取模块

连续多帧监控图像中的强降雨过程具有复杂的时间依赖性，仅使用单帧图像难以准确刻画降雨演变特征。本文引入 Bi-LSTM 对时序图像特征进行建模，将t帧序列 (对应 I 桢图像在网格里的序列) 所对应的 ViT 特征拼接形成时序特征矩阵

$$F = [f_1, f_2, \cdots, f_t]^{\mathrm{T}} \in R^{N\times D} \tag{2}$$

式中，N为帧数，D为 Vit 输出特征的维度 (768)。

将F输入 Bi-LSTM 网络，分别为正向 LSTM 处理时间顺序的特征序列、反向 LSTM 处理时间逆序的特征序列。正反向隐藏状态分别记作

$$\overrightarrow{h_t} = LSTM_{\mathrm{fwd}}(f_t);\ \overleftarrow{h_t} = LSTM_{\mathrm{bwd}}(f_t) \tag{3}$$

通过拼接正反向状态，获得每帧的时序特征表示为：$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \in R^{2H}$，$H$为单向 LSTM 的隐藏层维度。Bi-LSTM 能够同时捕捉图像序列中的前向与后向动态信息，从而更充分地反映降雨演变的过程特征。相比于单向 LSTM，其在处理降雨强度变化的上下文关系上具有更强的表达能力，尤其适用于突发性强、变化剧烈的夜间强降雨过程。鉴于降雨过程存在显著的时序变化，为增强时间建模能力，本文采用堆叠式 Bi-LSTM 结构进行时序建模。该模块支持多层堆叠形式部署，其中 Bi-LSTM 的最优层数设置基于实验结果确定，详见 2.1 节。

1.3.3 降雨等级分类模块

将 Bi-LSTM 输出特征h_N输入全连接层，映射至降雨等级类别空间，分类器输出为

$$\hat{y} = \mathrm{Softmax}(Wh_N + b) \tag{4}$$

式中W为权重矩阵，b为偏置向量，$\hat{y}_i$为预测概率，Softmax 函数定义如下

$$\mathrm{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \tag{5}$$

式中C为降雨等级类别数。采用交叉熵损失函数衡量预测结果与真实标签的差异

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{6}$$

式中y_i 为真实标签。

1.4 深度学习模型训练

首先，对原始监控视频数据进行预处理。每段视频片段均采用固定间隔的抽帧方式进行均匀采样，采样频率为 3 帧/秒，每个视频片段 (30 s) 共获得 90 帧图像。为统一输入尺寸，所有图像帧在输入模型前被重采样至 224×224 的分辨率。在 ViT 模块中，每帧图像进一步被划分为大小为 8×8 的图像块，总计生成 784 个图像块。在时间序列建模阶段，本文采用每 10 帧图像构成一个序列样本的策略，作为 Bi-LSTM 模块的输入。考虑到 Bi-LSTM 层数对时间特征提取能力的影响，本文对不同层数设置下的深度学习模型性能进行了对比分析 (见 2.2 节)。

在模型训练方面，采用交叉熵损失函数 (Cross-Entropy Loss) 评估模型预测与真实标签之间的差异，使用 Adam 优化器进行参数更新，初始学习率设为 1×10^{-4}。为提高训练过程的自适应性，采用 ReduceLROnPlateau 策略，在验证集性能连续若干轮未提升时动态降低学习率。此外，为评估批次大小对模型性能的影响，训练批次大小 (Batch Size) 分别设定为 32、64 和 128。最大训练轮数为 200，并引入早停机制 (Early Stopping)，当模型在验证集上在连续若干轮内无性能提升时，自动提前终止训练过程，以降低过拟合风险。训练过程中保存在验证集上表现最优的模型参数，并将其用于后续性能评估与分析。

2 模型训练结果与分析

2.1 自建数据集训练

为探讨 Bi-LSTM 网络层数对模型性能的影响，分别构建了包含 1 至 5 层 Bi-LSTM 的 Vit-Bi-LSTM 模型，并设置了一种不含 Bi-LSTM 模块 (0 层) 的 ViT 基线模型作为对照。为确保对比的合理性，所有实验均采用相同的数据预处理流程、损失函数、优化器参数和训练轮数。考虑到随着 Bi-LSTM 层数的增加，模型参数量呈指数增长易产生过拟合问题，采用层次递减的隐藏单元配置，即 1 层：128，2 层：128—64，3 层：128—64—32，4 层：128—96—64—32，5 层：128—96—64—32—16，模型训练结果如图 3 所示。

由图 3 可知，ViT 基线模型 (0 层 Bi-LSTM) 虽可从单帧图像中提取出丰富的空间特征，但是缺乏对降雨时序演化信息的利用，对降雨等级识别的平均精度仅为 75.2% (图 3a)。进一步对比图 3b–f 可见，当 Bi-LSTM 层数设为 2 层时，Vit-Bi-LSTM 模型在测试集上的识别精度最高，达 85.6%；当 Bi-LSTM 层数为 1 层和 3 层时，识别精度分别为 83.8% 和 83.6%，略低于 2 层配置；当 Bi-LSTM 层数增加至 4 层和 5 层时，识别精度下降至 80.6% 和 76.6%，显示出明显的性能退化。这一现象可能与网络深度增加导致模型参数量显著提升、进而引发过拟合有关。引入 Bi-LSTM 结构后，有助于建模降雨过程中的时间依赖性，从而提升模型性能 (图 3b–f)。该结果

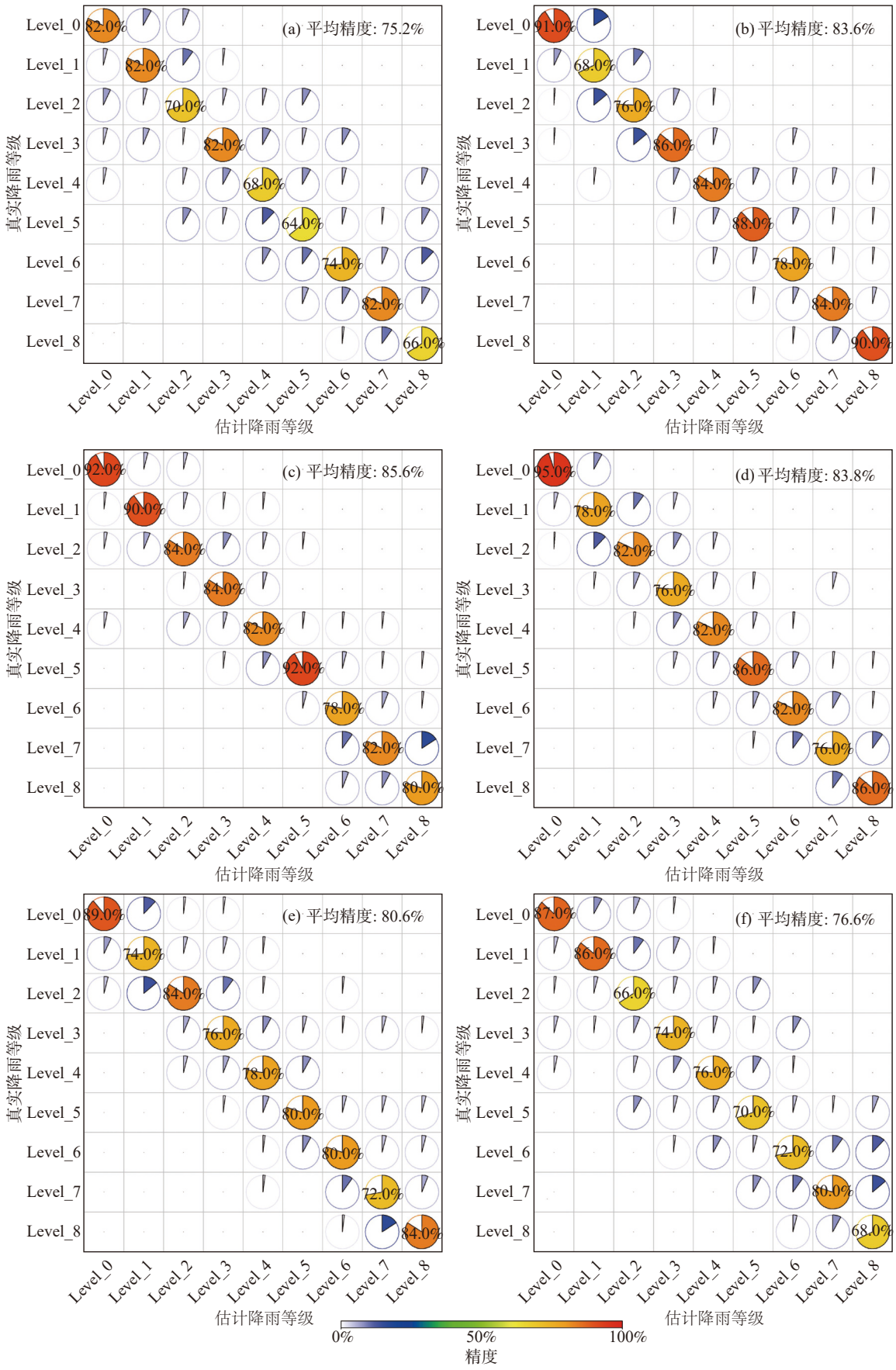


图 3 Bi-LSTM 0 层 (a)、1 层 (b)、2 层 (c)、3 层 (d)、4 层 (e)、5 层 (f) 对降雨等级的识别精度 (对角线位置数字表示正确预测的占比)

Fig. 3 Recognition accuracy of (a) 0, (b) 1, (c) 2, (d) 3, (e) 4, (f) 5 layer of Bi-LSTM on rainfall intensity classification (the number on diagonal line indicates the proportion of correct predictions)

说明，引入时序建模结构对于刻画降雨等级变化中的动态特征具有积极作用，可增强模型对不同强度降雨的判别能力。综合考虑性能表现与模型复杂度，后续实地降雨观测检验采用 2 层 Bi-LSTM 结构作为最终模型配置。

2.2 实地降雨观测检验评估

选取江苏省南京市江宁区作为实地观测区域，开展真实降雨条件下的检验评估。如图 4 所示，共部署 4 台监控摄像机 (表 1 所列举的类型 1 和类型 2 各一台，类型 3 两台)，从不同角度同步采集降雨视频数据。需要说明的是，检验所使用的 4 台相机与前期数据收集阶段部署的 10 台监控相机并非同一组设备。此设计旨在充分考虑不同监控场景对模型性能的影响，并避免训练数据与检验数据之间的交叉泄露，从而确保模型评估结果的客观性和在实际复杂环境下的泛化能力。同时，借助安装在观测区的 2-DVD 获取对应时段的高精度降雨等级数据。

使用 2023 年 6 月 18 日 01:20—06:00 (北京时，下同) 和 2023 年 7 月 17 日 03:00—05:00 江苏南京的两场夜间强降雨作为观测对象。图 5 给出了 2-DVD 测量的两场降雨的强度和过程信息，包括降雨强度以及按照 1.3 节中描述的方法划分的降雨等级。

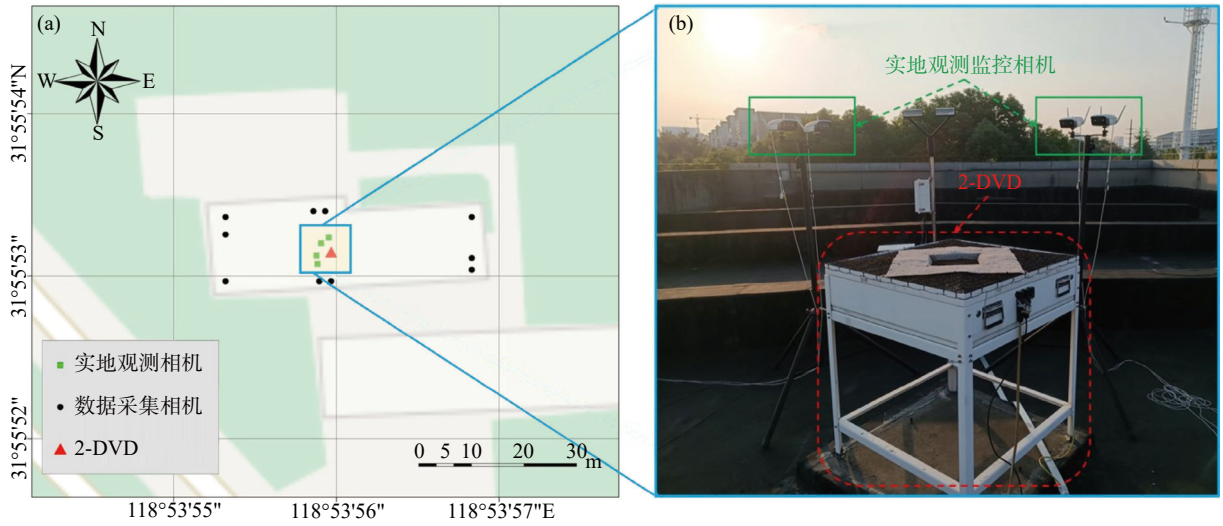


图 4 监控相机分布 (a) 与区域环境 (b)

Fig. 4 The (a) distribution of surveillance cameras and (b) overview of the experimental site

图 5 显示所选取的两场真实夜间降雨过程覆盖了从小雨到强降雨的多个降雨等级，等级跨度大，过程演变明显，有利于全面评估模型对不同降雨强度的区分能力。此外，两场降雨持续时间较长，均发生于夜间，期间光照条件随时间推移发生显著变化，能够较为真实地反映实际监控场景中降雨视频的时空特征变化，适用于检验模型在低照度与动态光照条件下的鲁棒性与适应性表现。实验中，多台监控相机的多角度布设提供了丰富的空间视角，有助于从不同方位评估模型对降雨图像特征的提取能力。值得注意的是，两场降雨部分时段的降雨强度超过 80 mm·h^{-1}，而该强度级别在训练集中未曾出现。为更准确标注该类极端强度的降雨，本文引入 Level_9 等级，用于区分超出原有最大等级 (Level_8) 范围的降雨事件。图 6 给出了 Vit-Bi-LSTM 模型在这两场夜间降雨过程中的等级识别性能表现，图中所示数值为基于 4 台监控相机估计结果的平均值。

从两场实地降雨观测结果 (图 6) 来看，Vit-Bi-LSTM 模型在 Level_0 至 Level_8 各等级的平均识别精度分别为 90.6%、86.3%、80.7%、75.0%、75.9%、68.8%、73.2%、69.2% 和 70.0%，整体呈现出随降雨等级上升而识别精度逐渐下降的趋势。具体而言，模型在 Level_2 及以下等级降雨场景中表现出较高的识别准确率 (均高于 80.7%)，表明其在弱降雨条件下具备良好的判别能力。相较之下，模型在强降雨 (雨强 ≥20 mm·h^{-1}) 场景中的识别性能则出现不同程度的下降。例如，6 月 18 日降雨过程中 Level_7 与 Level_8 (雨强 ≥60 mm·h^{-1}) 等级的识别精度均低于 70.0% (图 6a)；而在 7 月 17 日，Level_5 (雨强 ≥40 mm·h^{-1}) 及以上等级的识别精度亦明显下降，最低降至 68.8% (图 6b)。原因在于，不同降雨强度下视频图像的视觉特征复杂度存在显著差异。低强度降雨条件下，图像特征较为清晰、背景干扰较少，有利于模型提取关键空间信息并建模其时序演化过程，因而识别精度较高。相比之下，强降雨场景中往往伴随雨线密集、亮度剧烈波动等复杂视觉干扰，易引发图

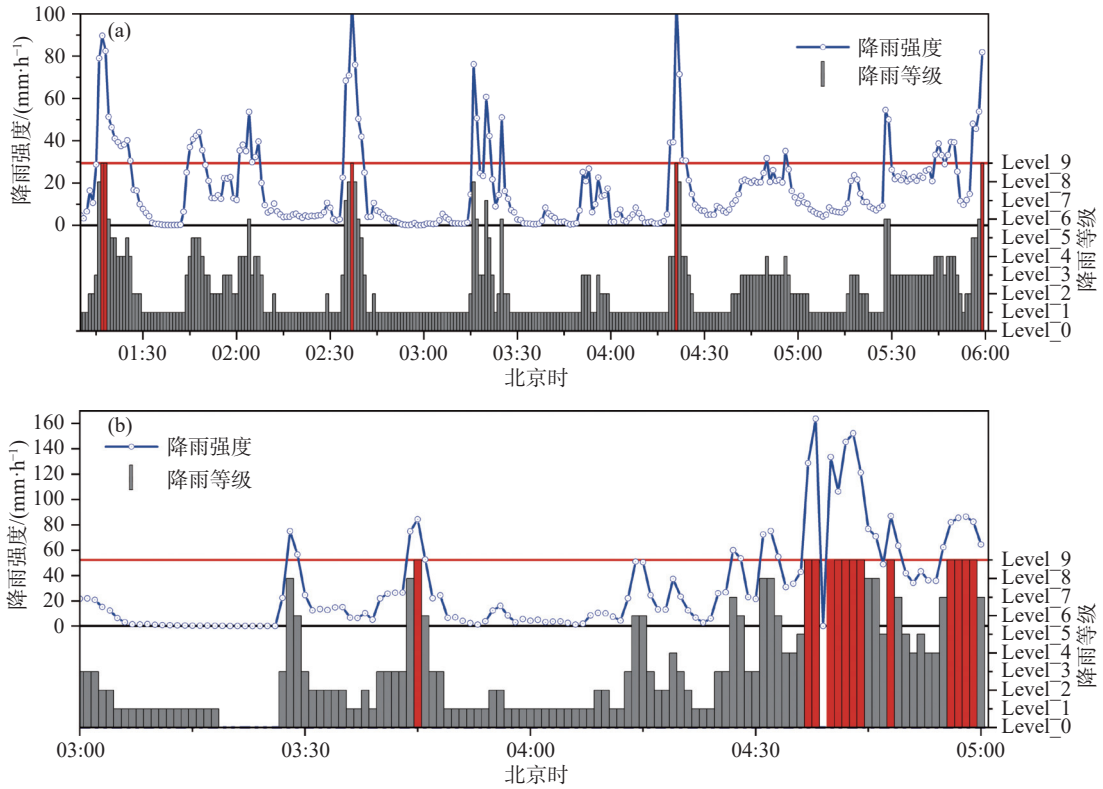


图 5 二维雨滴谱仪 (2DVD) 观测的 2023 年 6 月 18 日 (a) 和 2023 年 7 月 17 日 (b) 两场夜间降雨强度和等级 (红色柱状图表示不参与训练的降雨等级, 红色直线表示 Level_9 参考线)

Fig. 5 Rainfall process information of two nighttime heavy rainfall events on (a) June 18, 2023 and (b) July 17, 2023, observed by 2-DVD (The red bar chart denotes the predicted rainfall intensity levels that did not participate in the training, the red dashed line serves as the reference threshold for Level_9)

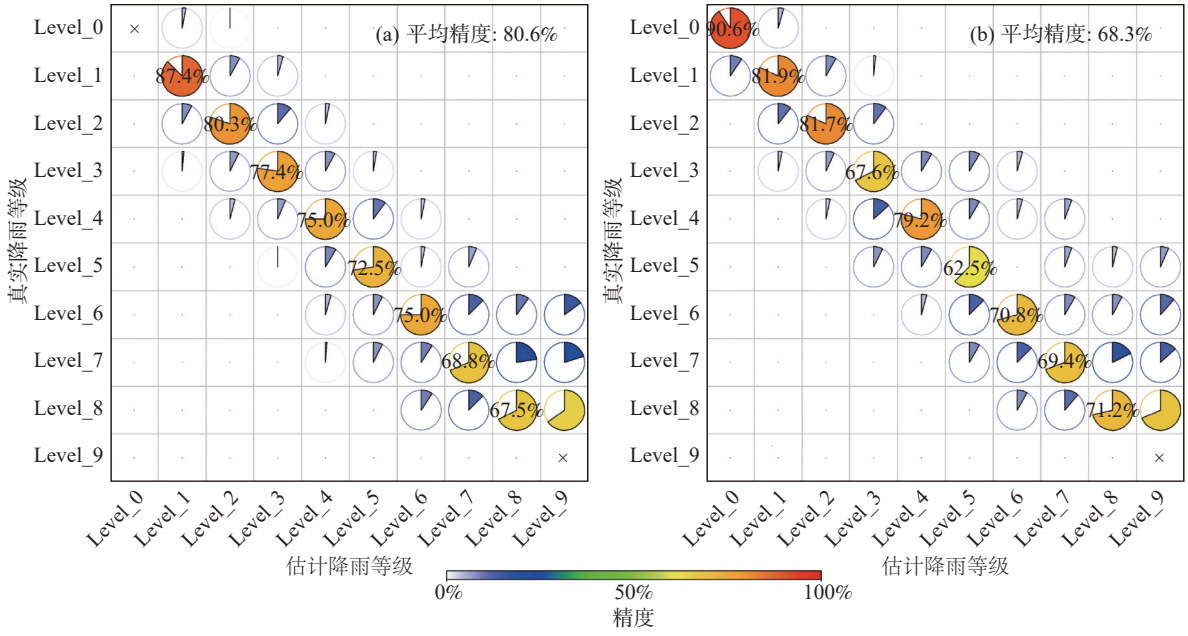


图 6 同图 3, 但为 2023 年 6 月 18 日 (a) 和 2023 年 7 月 17 日 (b) 两场夜间强降雨实地观测精度

Fig. 6 Same as Fig. 3, but for two nighttime heavy rainfall events on (a) June 18 and (b) July 17, 2023

像突变 (图 1), 从而增加了模型在空间特征提取与时序建模方面的难度, 进而导致识别性能下降。由此可见, Vit-Bi-LSTM 模型对强降雨的识别性能仍存需进一步提升。

对比两场降雨事件, Vit-Bi-LSTM 模型在 6 月 18 日的整体识别精度为 80.6% (图 6a), 而在 7 月 17 日降至 68.3% (图 6b), 平均精度为 76.7%。分析认为, 该差异主要源于 7 月 17 日降雨过程中 Level_5

及以上的强降雨 (即降雨强度大于 40 mm·h^{-1}) 持续时间较长。如图 6b 所示，6 月 18 日降雨中高于 Level_4 的时长为 53 min，占比 18.9%，而 7 月 17 日降雨则为 31.6%。这种差异导致了上述 Vit-Bi-LSTM 在高降雨等级场景中出现的精度衰减问题更加显著，从而影响了在两场降雨整体识别性能的差异。

此外，由图 5 可以看出，在两场降雨过程的部分时段降雨强度超过 80 mm·h^{-1}，共持续 18 min，已超出 Level_8 的最大降雨强度范围 (即 Level_9)。然而，尽管 Vit-Bi-LSTM 未在训练集中学习到这种极端降雨场景，模型仍将其归类为最接近的 Level_8 等级：6 月 18 日降雨强度超过 80 mm·h^{-1} 的部分被判定为 Level_8 的比例为 65%，而 7 月 17 日则为 68.3%。这一定程度上表明模型较好地掌握了强降雨的时空特征，并具备了一定的泛化能力。未来将进一步收集超过 80 mm·h^{-1} 极端降雨视频数据，扩展降雨等级的划分，以提升模型的适应性和准确性，增强监控相机对强降雨的测量能力。

3 结论与讨论

针对监控相机在夜间强降雨场景下测量能力较弱的问题，本文从降雨视频的时空维度出发，提出将 ViT 与 Bi-LSTM 相结合的深度学习模型 Vit-Bi-LSTM，用于实现夜间降雨等级的自动识别。通过收集 2022—2025 年间的 41 场夜间降雨视频，构建了最大降雨强度为 80 mm·h^{-1} 的夜间降雨视频数据集，用于模型的训练与评估。得出主要结论如下：

(1) 在自建多等级降雨视频数据集上的实验结果表明，与仅依赖单帧图像特征的 Vit 模型相比，Vit-Bi-LSTM 通过降雨视频时空特征联合建模策略可有效提升降雨判别能力。其中，两层 Bi-LSTM 结构的 Vit-Bi-LSTM 模型在测试集上取得了最优表现，对降雨等级的识别精度达到 85.6%。

(2) 在两场典型夜间强降雨观测实验中，Vit-Bi-LSTM 模型的平均识别精度为 76.7%，证明其在复杂环境下具有可行性。然而，随着降雨等级上升，尤其在 Level_5 及以上 (>40 mm·h^{-1}) 强降雨场景中，识别精度下降至 68.8%~75.0%，显示出模型在极端降雨条件下仍存在性能瓶颈。值得注意的是，模型对训练集中未出现的>80 mm·h^{-1} 降雨仍能合理识别，体现出一定的泛化能力。

由于现有方法主要聚焦于常规降雨条件下雨强的定量估计 (Lee et al., 2023; Wang et al., 2023b)，且缺乏对强降雨条件下性能的系统性评估，因此与本文方法难以直接进行比较。本文方法专为夜间强降雨监测场景设计，具备明确的应用针对性，能够更有效应对复杂图像条件下的测量挑战。因此，本文方法可作为对现有监控相机测雨方法提供有益的补充，特别是在夜间强降雨条件下展现出更高的适应性与实用价值。

未来，研究将持续收集不同监控场景下、涵盖多种降雨强度的夜间降雨视频数据，以扩展数据集的多样性和代表性，提升模型对复杂场景的适应能力。随着数据集的不断丰富，模型将采用更细化的降雨强度数值代替降雨等级为输出，从而进一步提高降雨观测结果的实用价值。气象部门与相关研究团队通过与城市安防、交通、水利等管理部门合作，获取多源监控视频数据，基于本文方法构建强降雨过程的动态监测，为气象及防灾部门提供更加精准的降雨信息，助力城市极端降雨监测与内涝预警工作。

参考文献 (References)：

陈刚,赵坤,吕迎辉,等.2022.河南 “21·7” 特大暴雨过程微物理特征变化分析[J].中国科学:地球科学,65(10):1861−1878. Chen G, Zhao K, Lyu Y H, et al. 2022. Variability of microphysical characteristics in the “21·7” Henan extremely heavy rainfall event [J]. Science China Earth Sciences,65(10):1861−1878 (in Chinese). doi:10.1007/s11430−022−9972−9

国家气象中心.2012.降水量等级:GB/T 28592—2012[S].北京:中国标准出版社. National Meteorological Centre,2012. Standard of precipitation: GB/T 28592-2012 [S]. Beijing:China Standard Press (in Chinese).

黄华兵,王先伟,柳林.2021.城市暴雨内涝综述: 特征,机理,数据与方法[J].地理科学进展,40(6):1048−1059. Huang H B, Wang X W, Liu L. 2021. A review on urban pluvial floods: Characteristics, mechanisms, data, and research methods [J]. Progress in Geography,40(6):1048−1059 (in Chinese). doi:10.18306/dlkxjz.2021.06.014

林珲,吴贤宇,潘家祎,等.2022.中国城市洪涝实时预报研究:现状与挑战[J].测绘学报,51(7):1306−1316. Lin H, Wu X Y, Pan J Y, et al. 2022. Real-time forecast of urban flood in China: past, present and future [J]. Acta Geodaetica et Cartographica Sinica,51(7):1306−1316 (in Chinese). doi:10.11947/j.AGCS.2022.20220144

盛志军,周雨,孟明华,等.2025.基于随机森林算法的城市内涝风险预估方法及应用-以南昌市为例[J].暴雨灾害,44(1):93−101. Sheng Z J, Zhou Y, Meng M H, et al. 2025. Method and application of urban waterlogging risk prediction based on the random forest algorithm-taking Nanchang as an example [J]. Torrential Rain and Disasters, 44(1):93−101 (in Chinese). doi:10.12406/byzh.2023−215

王兴,王美珍,刘学军.2021.视频节点协同的雨量反演精度控制模型[J].光学精密工程,29(11):10. Wang X, Wang M Z, Liu X J. 2021. Precision control model of rainfall inversion based on visual sensor nodes collaboration [J]. Optics and Precision Engineering,29(11):10 (in Chinese). doi:10.37188/OPE.20212911.2714

夏军,张印,梁昌梅,等.2018.城市雨洪模型研究综述[J].武汉大学学报

(工学版),51(2):95−105. Xia J, Zhang Y, Liang C M, et al. 2018. Review on urban storm water models [J]. Engineering journal of wuhan university,51(2):95−105 (in Chinese). doi:10.14188/j.1671−8844.2018−02−001

徐宗学,叶陈雷.2021.城市暴雨洪涝模拟:原理,模型与展望[J].水利学报,52(4):381−392. Xu Z X, Ye C L. 2021. Simulation of urban flooding/waterlogging processes: Principle, models and prospects [J]. Journal of Hydraulic Engineering,52(4):381−392 (in Chinese). doi:10.13243/j.cnki.slxb.20200515

周月华,彭涛,史瑞琴.2019.我国暴雨洪涝灾害风险评估研究进展[J]. 暴雨灾害,38(5):494-501. Zhou Y H, Peng T, Shi R Q. 2019. Research progress on risk assessment of heavy rainfall and flood disasters in China [J]. Torrential Rain and Disasters, 38(5):494-501 (in Chinese). doi:10.3969/j.issn.1004−9045.2019.05.011

Allamano P, Croci A, Laio F. 2015. Toward the camera rain gauge [J]. Water Resources Research,51(3):1744−1757. doi:10.1002/2014WR016298

Byun J, Jun C, Kim J, et al. 2023. Deep learning-based rainfall prediction using cloud image analysis [J]. IEEE Transactions on Geoscience and Remote Sensing, 61:1-11. doi:10.1109/TGRS.2023.3263872

Dong R, Liao J, Li B, et al. 2017. Measurements of rainfall rates from videos [C]//10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, pp. 1-9. doi:10.1109/CISP−BMEI.2017.8302066

Garg K, Nayar S K. 2007. Vision and rain [J]. International Journal of Computer Vision,75(1):3−27. doi:10.1007/s11263−006−0028−6

Huang H, Tao R, Zhao K, et al. 2023. Potential of snowfall nowcasting using polarimetric radar data and its link to ice microphysics: Study of two snowstorms in East China [J]. Journal of Geophysical Research: Atmospheres,128(15):e2022JD037654. doi:10.1029/2022JD037654

Jiang S, Babovic V, Zheng Y, et al. 2019. Advancing opportunistic sensing in hydrology: A novel approach to measuring rainfall with ordinary surveillance cameras [J]. Water Resources Research,55(4):3004−3027. doi:10.1029/2018WR024480

Kruger A, Krajewski W F. 2002. Two-dimensional video disdrometer: A description [J]. Journal of Atmospheric and Oceanic Technology,19(5):602−617. doi:10.1175/1520−0426(2002)019<0602:TDVDAD>2.0.CO;2

Lee J, Byun J, Baik J, et al. 2023. Estimation of raindrop size distribution and rain rate with infrared surveillance camera in dark conditions [J]. Atmospheric Measurement Techniques,16(3):707-725. doi:10.5194/amt−16−707−2023

Li M, Cao X, Zhao Q, et al. 2021. Online rain/snow removal from surveillance videos [J]. IEEE Transactions on Image Processing,30:2029−2044. doi:10.1109/TIP.2021.3050313

Lin C W, Lin M, Yang S. 2020. SOPNet method for the Fine-Grained measurement and prediction of precipitation intensity using outdoor surveillance cameras [J]. IEEE Access,8:188813−188824. doi:10.1109/ACCESS.2020.3032430

McCabe M F, Rodell M, Alsdorf D E, et al. 2017. The future of Earth observation in hydrology [J]. Hydrology and earth system sciences,21(7):3879−3914. doi:10.5194/hess−21−3879−2017

Mughees N, Mohsin S A, Mughees A, et al. 2021. Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting [J]. Expert Systems with Applications,175:114844. doi:10.1016/j.eswa.2021.114844

Rajabi F, Faraji N, Hashemi M. 2024. An efficient video-based rainfall intensity estimation employing different recurrent neural network models [J]. Earth Science Informatics,1-14. doi:10.1007/s12145−024−01290−x

Saputro A H, Bustamam A, Sopaheluwakan A, et al. 2023. Rainfall estimation using convolutional neural networks in various environmental backgrounds [C]//2023 International Conference on Information Technology Research and Innovation (ICITRI). IEEE, pp. 70-74. doi:10.1109/ICITRI59340.2023.10249806

Wang X, Shi S, Zhu L, et al. 2023a. Traditional and novel methods of rainfall observation and measurement: a review [J]. Journal of Hydrometeorology,24(12):2153−2176. doi:10.1175/JHM−D−22−0122.1

Wang X, Wang M, Liu X, et al. 2022. A novel quality control model of rainfall estimation with videos–A survey based on multi-surveillance cameras [J]. Journal of Hydrology,605:127312. doi:10.1016/j.jhydrol.2021.127312

Wang X, Wang M, Liu X, et al. 2023b. Near-infrared surveillance video-based rain gauge [J]. Journal of Hydrology,618:129173. doi:10.1016/j.jhydrol.2023.129173

Wang X, Zhou A, Zhao K, et al. 2025. Simulation of regional rainfall observation using urban surveillance camera networks [J]. Environmental Research Letters,20:5. doi:10.1088/1748−9326/adc825

Yan K, Chen H, Hu L, et al. 2023. A review of video - based rainfall measurement methods [J]. Wiley Interdisciplinary Reviews:Water,10(6):e1678. doi:10.1002/wat2.1678

Yang M, Wang J. 2022. Adaptability of financial time series prediction based on BiLSTM [J]. Procedia Computer Science,199:18−25. doi:10.1016/j.procs.2022.01.003

Yin H, Vahdat A, Alvarez J M, et al. 2022. A-vit: Adaptive tokens for efficient vision transformer [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.10809-10818. doi:10.48550/arXiv.2112.07658

Yuan L, Chen Y, Wang T, et al. 2021. Tokens-to-token vit: Training vision transformers from scratch on imagenet [C]//In Proceedings of the IEEE/CVF international conference on computer vision, pp.558-567. doi:10.1109/ICCV48922.2021.00060

Zen R, Arsa D M S, Zhang R, et al. 2019. Rainfall estimation from traffic cameras [C]//International Conference on Database and Expert Systems Applications. Springer, 18-32. doi:10.1007/978−3−030−27615−7_2

(责任编辑　杨　浩)