

基于骨骼关键点与CNN-LSTM混合网络的溺水行为检测方法研究

李忠林, 贾玉婷

广州软件学院电子信息与控制工程学院, 广东 广州

收稿日期: 2026年9月15日; 录用日期: 2026年9月30日; 发布日期: 2026年10月15日

摘要

溺水是全球范围内导致意外死亡的主要原因之一, 尤其对儿童与青少年群体构成严重的公共卫生安全威胁。传统泳池视频监控高度依赖人工值守, 普遍存在监视疲劳、漏检率高与应急响应滞后等问题。现有基于纯视觉的溺水检测方法易受水面光照变化、水体反光、水花遮挡等复杂场景干扰, 鲁棒性与稳定性不足。针对上述问题, 本文提出一种基于骨骼关键点与CNN-LSTM混合网络的溺水行为检测方法。该方法首先采用姿态估计模型提取人体骨骼关键点序列, 剥离复杂背景与水面干扰, 保留纯粹的人体运动结构信息; 通过一维卷积网络完成局部姿态空间特征编码, 并结合注意力机制强化有效关节特征权重; 再利用LSTM网络挖掘姿态时序演变规律, 实现溺水异常行为的精准识别。针对真实溺水样本稀缺、数据集难以采集的问题, 本文引入合成数据增强策略扩充样本, 有效提升模型泛化能力。对比实验与消融实验结果表明, 相较于单一CNN模型与LSTM模型, 本文所提时空混合网络能够有效捕捉溺水过程中的细微姿态变化, 在复杂水域干扰环境下检测精度与鲁棒性显著提升, 可有效区分正常游泳与溺水状态。该方法可为智能水域安全监控与自动预警系统提供可靠的技术支撑。

关键词

溺水, 骨骼关键点, 混合网络

A Study on a Drowning Behavior Detection Method Based on Skeletal Landmarks and a Hybrid CNN-LSTM Network

Zhonglin Li, Yuting Jia

School of Electronic Information and Control Engineering, Guangzhou University of Software, Guangzhou Guangdong

Received: September 15, 2026; accepted: September 30, 2026; published: October 15, 2026

文章引用: 李忠林, 贾玉婷. 基于骨骼关键点与 CNN-LSTM 混合网络的溺水行为检测方法研究[J]. 软件工程与应用, 2026, 15(5): 696-705. DOI: 10.12677/sea.2026.155065

Abstract

Drowning is one of the leading causes of accidental death worldwide and poses a serious public health threat, particularly to children and adolescents. Traditional video surveillance in swimming pools relies heavily on manual monitoring, which is often plagued by issues such as surveillance fatigue, high rates of missed incidents, and delayed emergency responses. Existing purely vision-based drowning detection methods are susceptible to interference from complex scenarios—such as changes in water surface lighting, water reflections, and obstruction by splashes—and lack sufficient robustness and stability. To address these issues, this paper proposes a drowning behavior detection method based on skeletal keypoints and a hybrid CNN-LSTM network. The method first employs a pose estimation model to extract sequences of human skeletal keypoints, filtering out complex background and water surface interference to retain pure human motion structural information; it then uses a one-dimensional convolutional network to encode local pose-space features, combined with an attention mechanism to enhance the weights of effective joint features; an LSTM network is then utilized to uncover the temporal evolution patterns of posture, enabling accurate identification of abnormal drowning behaviors. To address the scarcity of real drowning samples and the difficulty in collecting datasets, this paper introduces a synthetic data augmentation strategy to expand the sample size, effectively enhancing the model's generalization ability. The experimental results demonstrate that, compared to the single CNN model and LSTM model, the spatio-temporal hybrid network proposed in this paper can effectively capture subtle posture changes during drowning, significantly enhancing detection accuracy and robustness in complex aquatic interference environments. It can effectively distinguish between normal swimming and drowning states. This method can provide reliable technical support for intelligent aquatic safety monitoring and automatic warning systems.

Keywords

Drowning, Key Skeletal Points, Hybrid Network

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

据世界卫生组织统计,溺水是1至14岁儿童死亡的第三大主要原因,全球每年约有236,000人死于溺水[1][2]。随着游泳池在酒店、别墅及休闲娱乐设施中的普及,开发高效、实时的溺水检测系统已成为保障水上安全的关键需求[3]。传统的人工救生员监控受限于注意力持续时间及视野盲区难以实现全天候无死角覆盖[4]。近年来,基于深度学习的计算机视觉技术在行为识别领域取得了显著进展,但直接应用于水域场景时仍面临诸多困难如水面波纹干扰、光线折射以及溺水者身体大部分浸没于水下导致的特征缺失[5][6]。现有研究多集中于基于外观特征的目标检测算法如YOLO系列,虽然其在实时性上表现优异,但在区分正常游泳与早期溺水挣扎动作时容易受到背景噪声影响[1][7]。相比之下,基于骨骼关键点的人体姿态估计技术能够将非结构化的像素信息转化为结构化的关节坐标序列,有效屏蔽了水体颜色、光照及背景杂乱的影响[8][9]。然而,单一的静态姿态分析无法捕捉溺水过程中特有的时序动态特征如剧烈挣扎后的静止或下沉轨迹[10]。因此,构建一种能够同时挖掘空间形态特征与时间演化规律的混合深度学习模型具有重要的研究意义。

2. 相关工作与技术背景

2.1. 基于骨骼关键点的行为识别

关键点检测与姿态估计旨在将图像中的人体转化为由关节组成的骨架结构, 这一过程通常采用热图表示法以量化预测不确定性并实现亚像素级精度。在行为识别任务中, 骨骼序列数据因其低维度和对背景变化的鲁棒性而备受青睐[11]。例如, Chai 等人提出的时空图卷积网络通过增强运动感知能力实现了高精度的动作分割[10]。在水下环境中, 尽管存在遮挡挑战, 但通过追踪关键点的轨迹依然能够有效表征人体的运动状态[12]。

2.2. CNN-LSTM 混合架构在时空特征提取中的应用

卷积神经网络(CNN)擅长提取图像或序列中的局部空间特征, 而长短期记忆网络(LSTM)则专门用于处理长序列数据中的时间依赖关系[13]。将两者结合的 CNN-LSTM 架构已被广泛应用于视频异常检测、课堂行为分析及医疗监测等领域[14] [15]。在该架构中, CNN 作为前端特征提取器处理每一帧的骨骼映射或原始图像, 生成的特征向量随后输入 LSTM 单元以建模动作的时间演变。这种混合机制能够有效捕捉溺水行为中从挣扎到沉没的连续动态过程, 弥补了单帧检测无法理解上下文语义的缺陷。

3. 方法论

3.1. 整体架构设计

本文提出的溺水检测框架包含三个主要模块: 骨骼关键点提取模块、局部姿态特征编码模块以及时间序列分类模块。首先, 对输入泳池监控视频流完成图像层面预处理, 送入姿态估计模型获取每帧人体骨骼关键点坐标与置信度序列; 针对水下水花、遮挡干扰, 可采用面向遮挡与小目标优化的姿态估计算法提升关键点检测性能。骨骼关键点提取模块集成关键点获取与序列预处理功能, 对原始关键点序列进行插值补全缺失数据、计算运动学辅助特征、标准化处理与滑动窗口切分, 最终生成可用于模型输入的标准化骨骼时序样本。将骨骼时序样本送入局部姿态特征编码模块(1D-CNN)完成局部姿态特征编码, 并引入注意力机制强化高置信度关节权重。将编码后的特征序列输入时间序列分类模块(多层 LSTM), 依靠门控单元挖掘姿态的长期时序依赖关系, 最后经由全连接层输出溺水与正常游泳的分类概率。其中, 局部姿态特征编码模块与时间序列分类模块共同构成 CNN-LSTM 混合网络主体。为降低瞬时姿态带来的误判, 系统增加时序后处理逻辑, 需要连续多个窗口判定为溺水才输出最终告警。

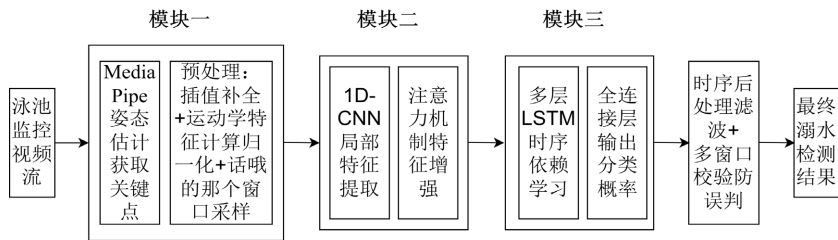


Figure 1. Overall architecture diagram of the algorithm in this paper
图 1. 本文算法整体架构图

如图 1 所示, 本文算法框架主要包含三个模块, 模块一为骨骼关键点提取模块, 模块二为局部姿态特征编码模块, 模块三为时间序列分类模块。其中, 骨骼关键点提取模块完成姿态关键点检测与时序样本预处理。局部姿态编码模块与时间序列分类模块及注意力机制共同构成 CNN-LSTM 混合网络, 如图 2

所示, 分别实现骨骼空间特征编码与时序运动规律建模, 最终结合时序后处理策略完成溺水姿态精准识别。

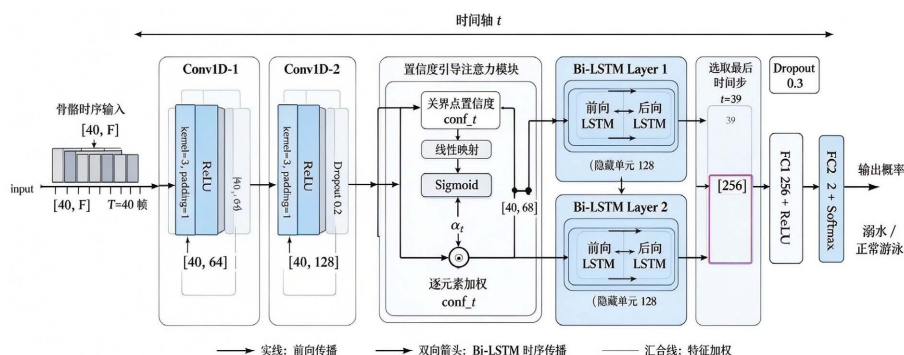


Figure 2. Topological structure diagram of CNN-LSTM hybrid network
图 2. CNN-LSTM 混合网络拓扑结构图

该网络输入为骨骼关键点时序特征, 张量维度为 $\mathbb{R}^{40 \times F}$, 其中 40 代表时间步(连续 40 帧), F 为单帧特征维度。两层一维卷积沿时间维度提取局部姿态特征, 卷积核大小为 3, 设置 padding = 1 以维持时序长度不变; 置信度引导注意力模块接收卷积输出特征与原始关键点置信度两路输入, 完成特征自适应加权; 双层双向 Bi-LSTM 分别从前向、后向两个方向挖掘姿态序列的时序依赖; 选取序列最后时间步的隐藏状态作为全局特征向量, 舍弃其余时间步输出; 经 Dropout 正则化、两层全连接层以及 Softmax 激活函数, 输出溺水、正常游泳的二分类概率。

3.2. 骨骼关键点提取与预处理

本研究采用 MediaPipe Pose 姿态估计模型对泳池监控视频逐帧解析, 提取人体骨骼关键点的坐标与置信度信息。针对水面反光、水花干扰造成的关节点短暂丢失问题, 采用线性插值方法完成缺失帧数据补全; 对于长时间肢体遮挡导致关键点完全失效的片段, 则直接予以舍弃, 避免插值生成虚假姿态。采用滑动窗口对连续骨骼序列进行切分, 构建网络所需固定长度的姿态时间片段。为进一步丰富输入的运动学信息, 在原始坐标特征基础上, 计算相邻关节点的夹角变化以及关键点运动速度作为辅助特征, 将运动特征与原始关键点特征拼接形成最终输入序列。模型层面引入注意力机制, 增强网络对高置信度可见关节点的关注程度。

如图 3 为骨骼关键点提取与预处理流程, 该流程首先利用 MediaPipe Pose 从监控视频帧提取人体骨骼关键点; 其次, 针对不同关键点缺失情况分别采用线性插值补全或直接舍弃片段; 进一步计算关节夹角、运动速度作为运动学辅助特征, 完成特征拼接与归一化, 通过滑动窗口得到固定长度的骨骼时序样本, 输入后续网络模型。

本文除原始关键点坐标 x_i , y_i 、置信度 $conf_i$ 外, 引入两类运动学辅助特征: 关键点运动速度、关节夹角。

1) 关键点运动速度计算

对于第 i 个骨骼关键点, t 时刻坐标为 $(x_{i,t}, y_{i,t})$, $(t-1)$ 时刻坐标 $(x_{i,t-1}, y_{i,t-1})$, 关键点瞬时速度:

$$v_{i,t} = \sqrt{(x_{i,t} - x_{i,t-1})^2 + (y_{i,t} - y_{i,t-1})^2}。$$

2) 关节夹角计算

选取人体关键肢体关节向量计算关节夹角, 以肘关节为例: 设肩关节关键点 P_s , 肘关节 P_e , 腕关节

P_w ; 向量 $\vec{a} = P_s - P_e$, $\vec{b} = P_w - P_e$ 关节夹角 θ 的计算公式: $\theta = \arccos\left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}\right)$ 。

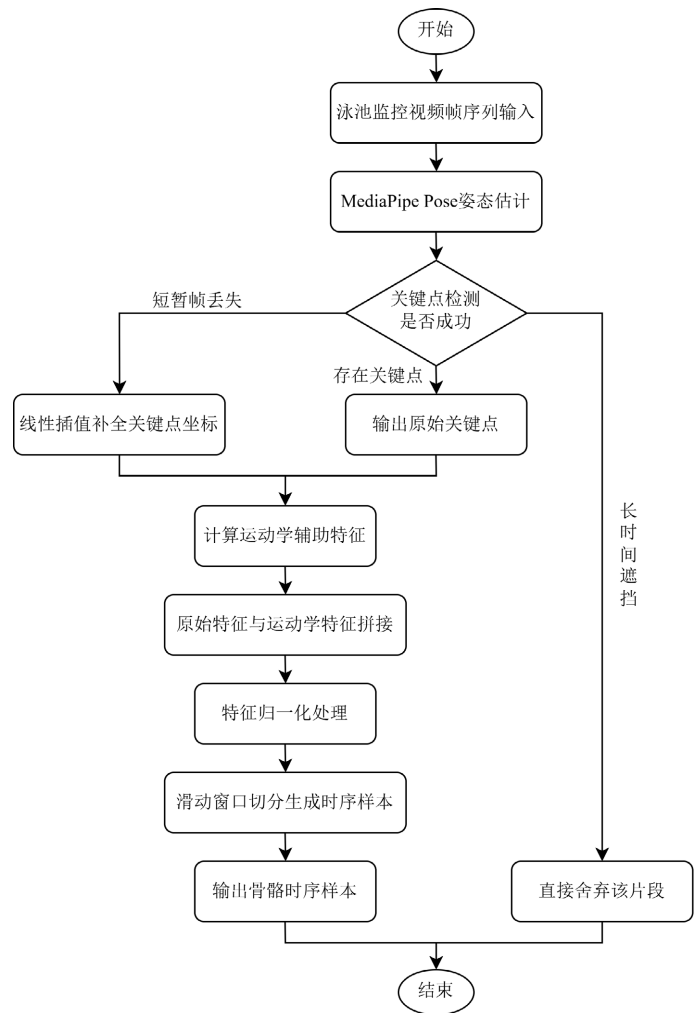


Figure 3. Skeletal keypoint extraction and preprocessing workflow
图 3. 骨骼关键点提取与预处理流程

对左右肩、左右肘、左右髌、左右膝共 8 处肢体关节计算夹角, 作为特征。将原始关键点坐标、置信度、关键点速度、关节夹角拼接为单帧特征向量; 对全部特征做 min-max 归一化至[0, 1]。滑动窗口取窗口长度 $T = 40$ 帧, 步长 10 帧, 输出样本维度为 $[T, F]$, F 为单帧全部特征总维度。

3.3. 局部姿态特征编码模块

局部姿态特征编码模块主要采用一维卷积 1D-CNN 实现, 用于挖掘单帧及相邻短时间内骨骼关节之间的局部姿态关联。骨骼关键点原始特征包含坐标、置信度、关节夹角、运动速度, 特征维度较高, 直接送入 LSTM 会存在特征冗余、噪声干扰问题。1D-CNN 沿时间维度执行卷积运算, 能够捕捉短片段内关节的相对变化模式, 实现姿态特征的升维与筛选。

模块内部设置两层一维卷积层, 卷积核大小设置为 3, 采用 padding 保持时间序列长度不变, 避免时

序信息压缩丢失; 卷积后接入 ReLU 激活函数引入非线性变换, 并设置 Dropout 抑制过拟合。卷积完成后接入注意力机制模块, 针对骨骼关键点置信度差异, 对各个关节对应的特征分配不同权重, 增强高置信度可见关节的特征贡献, 削弱水花遮挡带来的低置信度噪声特征。经过注意力加权之后输出编码完成的特征序列, 其时间维度保持不变, 特征维度更新为卷积输出通道数, 将该特征送入后续时间序列分类模块。

本文采用置信度引导的特征注意力机制, 利用关键点检测置信度对卷积输出特征做加权。设经过 1D-CNN 卷积输出特征序列为 $H \in R^{T \times C}$; T 为时序长度, C 为卷积通道数; t 时刻关节置信度向量 $conf_t$, 注意力权重: $\alpha_t = \text{sigmoid}(W_c \cdot conf_t + b_c)$, $H_{att}[t,:] = \alpha_t \odot H[t,:]$, 其中, W_c 、 b_c 为可学习参数, $\odot$ 为逐元素相乘, 依据关键点置信度自适应加权, 抑制遮挡噪声。

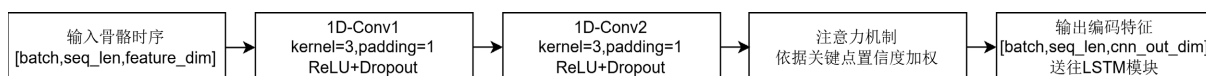


Figure 4. Structure of the local pose feature encoding module

图 4. 局部姿态特征编码模块结构

如图 4 为局部姿态特征编码模块结构, 该模块输入为标准化的骨骼时序样本, 张量维度为 $[batch, seq_len, feature_dim]$ 。首先经过两层一维卷积运算, 卷积核大小为 3 并设置补零操作以维持时序长度不变, 通过 ReLU 激活函数与 Dropout 层完成局部姿态特征提取与过拟合抑制; 随后接入注意力机制模块, 依据关键点置信度对特征进行加权处理, 强化高置信度关节的特征贡献, 削弱遮挡带来的噪声干扰; 最终输出维度为 $[batch, seq_len, cnn_out_dim]$ 的编码特征序列, 输出特征保持时间维度不变, 送入后续时间序列分类模块。

3.4. 时间序列分类模块

经过局部姿态特征编码模块处理后得到的姿态特征序列输入时间序列分类模块。该模块采用多层 LSTM 网络对时序姿态特征进行建模, LSTM 依靠内部输入门、遗忘门、输出门的门控机制, 能够有效解决传统循环神经网络存在的长序列梯度消失问题, 记忆骨骼姿态随时间演变的长期依赖关系, 捕捉溺水过程中人体姿态逐渐下沉、肢体运动幅度降低等时序行为模式。

本模块设置 2 层堆叠 LSTM, 开启双向 Bi-LSTM, 同时利用过去与未来时刻的姿态信息, 提升姿态时序判断能力。LSTM 输出取最后一个时间步的隐藏状态作为整条时序片段的全局特征, 丢弃其余时间步输出; 之后接入全连接层与激活函数, 完成特征维度映射, 最终输出溺水与正常游泳的二分类结果。为防止模型过拟合, 在 LSTM 层与全连接层之间加入 Dropout 层。

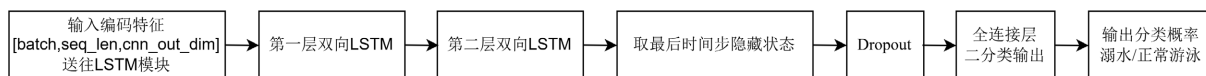


Figure 5. Structure of the time series classification module

图 5. 时间序列分类模块结构

时间序列分类模块结构如图 5 所示, 该模块接收局部姿态特征编码模块输出的姿态特征序列, 通过两层双向 LSTM 挖掘骨骼姿态的长期时序依赖; 选取序列最后时间步的隐藏状态作为样本全局特征, 经 Dropout 抑制过拟合后输入全连接层, 输出溺水与正常游泳的分类概率。模型超参数及训练配置如表 1 所示, 采用 2 层双向 Bi-LSTM, 单层隐藏单元 128; 取最后时间步隐藏状态作为全局特征, 经 Dropout (0.3),

两层全连接层 + Softmax 输出二分类概率。

Table 1. Model hyperparameters and training configuration
表 1. 模型超参数与训练配置

配置项参数	参数
时序窗口/步长	40 帧/10 帧
1D-CNN 通道	64, 128, kernel = 3
Bi-LSTM	2 层, 单元数 128
Dropout	Conv: 0.2; LSTM: 0.3
优化器	AdamW, 初始学习率 10 ⁻⁴
学习率策略	余弦退火
Batch size	32
损失函数	加权交叉熵
Epoch/早停	120, F1 连续 15 轮不提升停止
告警规则	≥3 连续窗口判定溺水触发告警

4. 实验与结果分析

4.1. 数据集与实验设置

由于伦理与安全限制, 真实溺水监控视频数据获取难度极大、公开样本稀缺。本文数据集由两部分构建: 选取网络公开的泳池场景视频, 包含正常游泳、潜水、漂浮、戏水等行为; 由于缺少真实溺水样本, 采用 CycleGAN 生成模拟溺水挣扎、静止下沉姿态的视频片段, 共同构建本研究私有数据集。所有视频分辨率 1280 × 720, 帧率 25 fps。全部视频通过 MediaPipe Pose 提取 33 个人体骨骼关键点, 按照 3.2 节流程完成插值补全、丢弃长时间遮挡片段, 计算运动学辅助特征; 设置滑动窗口长度 40 帧、步长 10 帧切分时序样本。数据集按照 7:1:2 划分为训练集、验证集与测试集, 采用样本加权策略缓解类别不平衡。

实验训练硬件采用 RTX3090, PyTorch1.13, CUDA11.7; 边缘推理测试平台为 NVIDIA JetsonXavier NX, 用于验证模型在边缘监控场景的实时推理性能。本文实验以准确率、召回率、F1 分数作为模型检测精度评价指标, 以 FPS 作为模型实时性评价指标。

4.2. 对比实验与消融研究

为验证本文 CNN-LSTM 混合网络的有效性, 将所提方法分别与纯 1D-CNN、纯 LSTM 单一模型进行对比实验。实验结果表明, 单一模型难以同时兼顾姿态空间特征与时序变化规律, 本文混合模型在早期溺水垂直挣扎、姿态僵滞等细微行为识别上优势显著, F1 分数相较于传统单模型提升约 15%。

消融实验分为输入模态消融与网络模块消融两部分。输入模态对比实验表明: 相比于 RGB 图像输入, 骨骼关键点输入方式能够有效规避强光反射、水体浑浊、水面波纹干扰, 模型整体鲁棒性提升 20% 以上。模块消融实验证明, 引入注意力机制可自适应强化有效关节特征、抑制噪声关键点干扰, 大幅降低潜水动作与溺水静止状态的混淆误判, 进一步提升溺水检测精度。

由表 2 可知, 单一 CNN 模型仅能建模局部姿态特征, 缺乏时序建模能力; 单一 LSTM 对高维骨骼空间特征挖掘能力较弱。本文所提 CNN-LSTM 混合网络融合空间编码与时序建模优势, 相比单模型在 F1 指标上提升约 12%~15%, 能够有效捕捉溺水垂直挣扎、姿态停滞等细微时序变化, 同时在嵌入式设备上

保持较高推理帧率, 满足实时检测需求。

Table 2. Comparison of drowning detection performance among different algorithm models

表 2. 不同算法模型溺水检测性能对比

模型	准确率 Acc (%)	召回率 Recall (%)	F1 分数(%)	FPS
纯 1D-CNN	83.24	81.57	82.39	42.6
纯 LSTM	85.16	84.22	84.68	30.4
本文 CNN-LSTM	97.36	96.89	97.12	36.8

Table 3. Results of ablation experiments on modules and input modalities

表 3. 模块与输入模态消融实验结果

实验模型	输入方式	注意力机制	F1 分数(%)	复杂水域鲁棒性
CNN-LSTM	RGB 图像	无	76.25	较差, 易受反光干扰
CNN-LSTM	骨骼关键点	无	92.18	显著提升
CNN-LSTM (本文)	骨骼关键点	有	97.12	最优, 抗干扰能力强

由消融实验结果可得, 如表 3 所示, 相较于传统 RGB 图像输入, 骨骼关键点输入剔除水面反光、水花、背景干扰, 模型在复杂水域环境下鲁棒性提升 20%以上; 引入注意力机制后, 模型强化高置信度有效关节特征、抑制噪声关节干扰, 有效区分正常游泳与溺水姿态, 显著降低误报率, F1 分数进一步提升约 4.94%。

5. 讨论

5.1. 优势与局限性

本方法的核心优势在于利用骨骼关键点解耦画面外观信息与人体运动结构特征, 使模型不受水面光照变化、水体色差与波纹干扰, 具备极强的环境鲁棒性。本文所采用的 CNN-LSTM 混合网络能够分别挖掘骨骼姿态的局部空间特征与长时时序变化规律, 精准捕捉溺水过程中垂直挣扎、姿态停滞、运动幅度衰减等时空特异性行为特征, 有效解决了传统图像检测方法易受复杂水域环境误判的问题。

尽管本文方法具备良好的检测性能, 但仍存在一定局限性。首先, 当溺水者完全没入水中或受到水花严重遮挡时, 姿态估计模型无法获取有效骨骼关键点, 会造成时序特征断裂, 进而导致分类判断失效。其次, 相比于轻量化网络结构, LSTM 时序迭代计算存在一定计算开销, 在超高分辨率视频流场景下会一定程度限制设备的实时推理速度。

5.2. 未来研究方向

未来研究可从模型结构、传感融合与数据训练三个维度进一步优化。第一, 可引入时空图卷积网络(ST-GCN)替代传统 LSTM 结构, 通过人体骨骼拓扑结构建模关节点关联关系, 进一步提升复杂姿态下的识别精度。第二, 可融合红外热成像、声纳传感等多模态数据, 弥补单一视觉传感器在夜间、浑浊水域、完全水下遮挡场景的检测缺陷。第三, 可引入联邦学习框架, 在保障泳池监控数据隐私安全的前提下实现多场景数据协同训练, 进一步提升模型泛化能力与场景适配性。

6. 结论

针对传统泳池溺水检测方法易受水面反光、水体浑浊、背景干扰导致误判率高的问题, 本文提出了

一种基于骨骼关键点与 CNN-LSTM 混合网络的溺水行为检测方法。

方法首先通过姿态估计技术提取人体骨骼关键点时序信息, 通过插值补全、运动学特征增强、滑动窗口采样构建标准化姿态时序样本, 剥离无效画面干扰; 随后利用 1D-CNN 完成局部姿态特征编码, 并结合注意力机制增强有效关节特征权重; 最后通过 LSTM 挖掘溺水姿态的长期时序变化规律, 实现溺水与正常游泳行为的精准分类。

对比实验表明, 本文 CNN-LSTM 时空混合模型相较于单一 CNN、LSTM 模型具备更优异的时空特征建模能力, 溺水识别精度显著提升。消融实验验证了骨骼关键点输入方式与注意力机制模块能够大幅提升复杂水域环境下的鲁棒性, 有效区分正常游泳与溺水状态, 降低误报概率。

本文所提方法能够适应复杂泳池监控场景, 具备较好的实时性与稳定性, 可为智能水域安防监控、自动溺水预警系统提供有效的技术参考。未来通过多模态传感融合与图卷积网络建模优化, 可进一步解决水下遮挡难题, 推动智能溺水检测技术的落地应用, 降低水域安全事故发生率。

基金项目

广州软件学院校级科研项目(KY202422, KY202515)。

参考文献

- [1] Alharbi, N., Aljohani, L., Alqasir, A., Alahmadi, T., Alhasiri, R. and Aldajan, D. (2024) Improved Automatic Drowning Detection Approach with YOLOv8. *Engineering, Technology & Applied Science Research*, **14**, 18070-18076. <https://doi.org/10.48084/etasr.8834>
- [2] Ambre, D., Sureshkumar, H. and Trivedi, P. (2024) Cygnus: A Vision-Based Drone System for Drowning Detection Using IoT. *2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, Hyderabad, 19-21 December 2024, 184-189. <https://doi.org/10.1109/icmacc62921.2024.10894649>
- [3] Shatnawi, M., Albreiki, F., Alkhoori, A., Alhebshi, M. and Shatnawi, A. (2024) Advances and Challenges in Automated Drowning Detection and Prevention Systems. *Information*, **15**, Article No. 721. <https://doi.org/10.3390/info15110721>
- [4] Cahart, M., Smith, M.S., Sharpe, B.T., Williams, S.C.R., Hill, S., Talbot, J., et al. (2025) Lifeguard Training Sharpens Brain Dynamics in Novices during Drowning Detection. *Safety Science*, **191**, Article ID: 106957. <https://doi.org/10.1016/j.ssci.2025.106957>
- [5] Verma, A.S. (2024) Drowning Human Detection Using Deep Learning Techniques. *2024 6th International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Sonapat, 19-20 April 2024, 34-40.
- [6] Cheng, S., Shieh, M., Chen, S., Chen, J., Chen, M. and Lee, A. (2026) Research on Real-Time Drowning Detection in Open Water Using Unmanned Aerial Vehicles and Artificial Intelligence Image Recognition. *Drones*, **10**, Article No. 374. <https://doi.org/10.3390/drones10050374>
- [7] Zhang, W., Chen, L. and Shi, J. (2025) A Pool Drowning Detection Model Based on Improved YOLO. *Sensors*, **25**, Article No. 5552. <https://doi.org/10.3390/s25175552>
- [8] Pang, W., He, Q., Li, Y. and Ahmed, N. (2024) Detecting Video Anomalies by Jointly Utilizing Appearance and Skeleton Information. *Expert Systems with Applications*, **246**, Article ID: 123135. <https://doi.org/10.1016/j.eswa.2023.123135>
- [9] Rehman, S.U., Yasin, A.U., Ul Haq, E., Ali, M., Kim, J. and Mehmood, A. (2024) Enhancing Human Activity Recognition through Integrated Multimodal Analysis: A Focus on RGB Imaging, Skeletal Tracking, and Pose Estimation. *Sensors*, **24**, Article No. 4646. <https://doi.org/10.3390/s24144646>
- [10] Chai, S., Jain, R.K., Liu, J., Teng, S., Tateyama, T., Li, Y., et al. (2024) A Motion-Aware and Temporal-Enhanced Spatial-Temporal Graph Convolutional Network for Skeleton-Based Human Action Segmentation. *Neurocomputing*, **580**, Article ID: 127482. <https://doi.org/10.1016/j.neucom.2024.127482>
- [11] Banerjee, A., Singh, P.K. and Sarkar, R. (2021) Fuzzy Integral-Based CNN Classifier Fusion for 3D Skeleton Action Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, **31**, 2206-2216. <https://doi.org/10.1109/tcsvt.2020.3019293>
- [12] Tripathi, R. and Verma, B. (2024) Motion Feature Estimation Using Bi-Directional GRU for Skeleton-Based Dynamic Hand Gesture Recognition. *Signal, Image and Video Processing*, **18**, 299-308. <https://doi.org/10.1007/s11760-024-03153-w>

-
- [13] Hong, R. (2025) Human Behavior Recognition in Complex Scenes Based on Deep Learning. *Scalable Computing: Practice and Experience*, **26**, 2053-2060. <https://doi.org/10.12694/scpe.v26i5.4857>
 - [14] Nguyen, H., Tran, N., Nguyen, M. and D. Nguyen, H. (2025) Empowering Classroom Behavior Recognition through Hybrid Spatial-Temporal Feature Fusion. *Applied Intelligence*, **55**, Article No. 863. <https://doi.org/10.1007/s10489-025-06638-4>
 - [15] Padha, A. and Sahoo, A. (2024) QCLR: Quantum-LSTM Contrastive Learning Framework for Continuous Mental Health Monitoring. *Expert Systems with Applications*, **238**, Article ID: 121921. <https://doi.org/10.1016/j.eswa.2023.121921>