

中文引用格式:郝秦霞,甄浩龙. 基于多模态融合的煤矿输送带运输区矿工不安全行为识别[J]. 中国安全科学学报, 2025, 35(11): 32-41.

英文引用格式:HAO Qinxia, ZHEN Haolong. Unsafe behavior recognition of miners in coal mine belt area based on multimodal feature fusion[J]. China Safety Science Journal, 2025, 35(11): 32-41.

基于多模态融合的煤矿输送带运输区 矿工不安全行为识别*

郝秦霞 副教授, 甄浩龙

(西安科技大学 通信与信息工程学院, 陕西 西安 710054)

中图分类号: X936

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2025.11.0357

资助项目: 陕西省重点研发计划(2024GX-YBXM-526)。

【摘要】 为提高煤矿输送带运输区矿工行为识别的准确性和鲁棒性, 降低不安全行为发生概率, 提出一种基于多模态特征融合的不安全行为识别方法。首先, 结合红绿蓝(RGB)与骨骼2种模态, 通过基于高效注意力机制(EA)改进的SlowFast网络(EA-SlowFast)提取RGB模态特征, 并利用改进后的移动增强型(ME)YOLO(ME-YOLO)检测矿工目标; 然后, 对检测到的矿工送入Lite-HRNet进行姿态估计, 提取骨骼关键点数据, 再通过深度通道卷积注意力时空图卷积网络(DC-STGCN)提取骨骼模态特征, 实现RGB模态与骨骼模态的晚期融合; 最后, 分别选取公开数据集(人类动作数据库51(HMDB51)、中佛罗里达大学101类动作识别数据集(UCF101))与自建煤矿井下不安全行为数据集进行试验验证。结果表明: 相较于单一模态基准模型, EA-SlowFast与DC-STGCN在HMDB51数据集上的识别准确率分别为71.6%和68.3%, 在UCF101动作识别数据集上的识别准确率分别为88.3%和85.4%; 多模态特征融合模型的识别准确率较未改进的融合模型有所提升, 在HMDB51和UCF101数据集上分别达到75.4%和93.5%。在自建数据集上, ME-YOLO的平均检测精度(mAP)@0.5为91.8%, 推理速度为38帧/s, 满足实时性需求。融合模型的识别准确率达到90.6%, 验证了该方法在煤矿输送带运输区环境中的有效性。

【关键词】 多模态; 特征融合; 煤矿输送带运输区; 矿工不安全行为; 目标检测; 红绿蓝(RGB)模态; 骨骼模态

Unsafe behavior recognition of miners in coal mine belt area based on multimodal feature fusion

HAO Qinxia, ZHEN Haolong

(School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China)

Abstract: To enhance the accuracy and robustness of miner behavior recognition in underground belt conveyor areas and reduce the occurrence of unsafe behaviors, a multimodal feature fusion-based recognition method was proposed. The method integrated RGB and skeletal modalities, where an improved Efficient Attention(EA)-SlowFast network was employed to extract RGB features, and an improved Mobile-Enhanced(ME)-YOLO was utilized for miner detection. The detected miners were subsequently processed

by Lite-HRNet for pose estimation to extract skeletal keypoints, which were further encoded using an improved Depthwise Convolution and Channel Prior Attention Spatio-Temporal Graph Convolutional Network (DC-STGCN) to obtain skeletal features. Finally, late fusion of RGB and skeletal features was performed for unsafe behavior recognition. In the experimental design, both public datasets (Human Motion Database 51 (HMDB51), University of Central Florida 101 Action Recognition Dataset (UCF101)) and a self-constructed underground coal mine unsafe behavior dataset were adopted for validation. The results demonstrate that, compared with single-modal baseline models, EA-SlowFast and DC-STGCN achieve recognition accuracies of 71.6% and 68.3% on HMDB51, and 88.3% and 85.4% on UCF101, respectively. The multimodal fusion model outperforms the unimodal baseline, achieving 75.4% and 93.5% on HMDB51 and UCF101, respectively. On the self-constructed dataset, ME--YOLO attains a Mean Average Precision (mAP) @ 0.5 of 91.8% with an inference speed of 38 f/s, satisfying real-time requirements. The fusion model further achieves an accuracy of 90.6%, confirming the effectiveness of the proposed approach in complex underground belt conveyor environments.

Keywords: multimodal; feature fusion; coal mine belt area; unsafe behaviors of miners; object detection; red-green-blue (RGB) modality; skeleton modality

0 引言

在井工煤矿的生产运输体系中,带式输送机作为核心设备,其安全稳定运行对矿井生产至关重要。然而,由于矿工安全意识薄弱、安全培训不足及违规操作习惯难以根除^[1],煤矿输送带区违规操作频发,已成为安全事故的主要诱因,严重威胁煤矿安全生产。统计数据显示,人为因素在矿山事故成因中占比高达 96%,其中,因矿工不安全行为直接导致的事故达 74%^[2]。在煤矿输送带区域中,根据煤矿安全生产网、国家矿山安全监察局、国家矿山安全局及各省级安全监管部门的官方网站资料以及全国范围内因矿工不安全行为导致的事故调查结果,违章穿越、攀爬、倚靠带式输送机等不安全行为尤为突出,极易引发机械伤害、坠落及触电等事故。因此,精准识别矿工不安全行为并及时预警,对减少安全事故、提升煤矿安全管理水平具有重要价值。

基于视频监控的行为识别技术为煤矿井下安全监管提供了新的技术手段^[3]。人体行为识别作为计算机视觉的重要研究方向,主要包括基于轨迹特征^[4]、卷积神经网络^[5]、时空图卷积网络 (Spatial Temporal Graph Convolutional Networks, ST-GCN)^[6]及循环神经网络^[7]方法。随着矿山智能化进程的加速,该技术在井下人员行为识别中得到广泛应用。CAO Xiangang 等^[8]在 ST-GCN 基础上提出了改进的矿工不安全行为识别网络,通过骨架时空图减少环境干扰,重构图结构与优化划分策略,并引入自注意力机制增强全局信息学习,显著提升了不安全行为识别的准确性。李占利等^[9]针对矿井复杂环境下

现有模型识别率低、适应性差的问题,提出基于 C3D^[10]的网络,该网络集成多尺度特征融合模块以增强特征表征能力,并结合注意力机制精准聚焦关键区域,从而提升矿井环境的行为识别性能。饶天荣等^[11]提出一种基于交叉注意力机制的多特征融合模型,通过增强特征表示能力,显著提升了不安全行为识别的精度。

然而,以上方法都聚焦于单一模态对井下人员的行为识别,其中,基于红绿蓝 (Red-Green-Blue, RGB) 模态的方法对环境变化如光照、遮挡和复杂背景高度敏感,从而影响识别准确性。基于骨骼模态的方法仅关注人体关键点,忽略了背景和物体信息,在某些场景下会导致识别精度下降。因此,笔者拟融合 RGB 模态与骨骼模态提取的特征,提出一种基于多模态特征融合的方法。该方法结合 RGB 模态与骨骼模态特征,分别利用改进的高效注意力慢通道-快通道视频特征网络 (Efficient Attention-SlowFast) 与深度通道卷积注意力时空图卷积网络 (Depthwise Convolution and Channel Prior Attention Spatio-Temporal Graph Convolutional Network, DC-STGCN) 进行特征提取,并通过晚期融合识别复杂井下环境中矿工不安全行为,旨在提升模型对光照变化、遮挡及复杂背景的适应能力,提升对煤矿输送带运输区域矿工不安全行为的识别精度。

1 矿工不安全行为识别模型

基于多模态的煤矿输送带区矿工不安全行为识别模型如图 1 所示。该模型由 RGB 模态和骨骼模态 2 条特征提取路径组成。针对井下复杂环境对行

为识别的影响,优化各模块。针对 RGB 模态的输入视频数据,采用改进的 SlowFast 网络^[12]强化时空特征的捕捉能力,从而更精准地提取行为信息。对于骨骼模态数据,首先,利用优化后的 YOLOv8m^[13]精确检测 RGB 视频数据中的矿工目标;其次,借助轻量级高分辨率网络 Lite-HRNet^[14]估计检测到的目

标的姿态,提取矿工的 2D 关键点信息并构建骨骼模态数据;然后,将生成的骨骼模态数据输入改进的 ST-GCN 网络,以深度挖掘时空特征并增强对矿工行为模式的理解;最后,通过晚期融合策略整合 RGB 模态与骨骼模态的特征,识别井工煤矿输送带区矿工不安全行为。

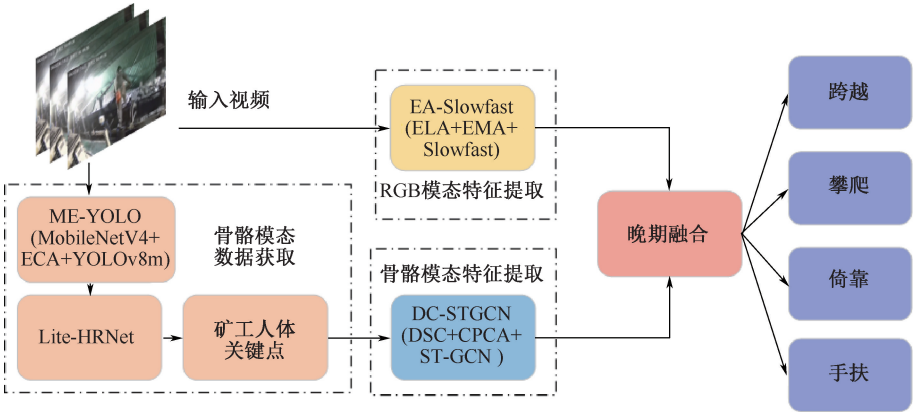


图 1 基于多模态的井工煤矿输送带区矿工不安全行为识别模型

Fig. 1 A multimodal-based model for unsafe behavior recognition of miners in the belt area of underground coal mine

2 骨骼模态数据获取

2.1 基于 YOLOv8m 的人体目标检测

在识别煤矿输送带区矿工不安全行为时,直接检测视频中矿工人体骨骼关键点可能会带来诸多问题,如背景干扰、误检、计算冗余等,而通过目标检测网络先检测矿工,再提取关键点能够显著提升骨骼模态数据的质量,为行为识别提供可靠输入^[15]。针对煤矿井下作业环境中存在的粉尘、遮挡、低光照等特殊因素,YOLOv8m 能够更精确地检测矿工并提取

出更为准确的人体目标信息,为骨骼数据提取提供高质量输入。然而,传统 YOLOv8m 计算量较大,在煤矿输送带区的实时监测任务中,计算量较大的模型可能会引发推理延迟问题,影响系统的实际应用效果。因此,采用 MobileNetV4^[16] 网络替换 YOLOv8m 的 Backbone。MobileNetV4 引入通用倒置瓶颈块(Universal Inverted Bottleneck, UIB)搜索块,通过轻量化设计,在降低模型参数和计算量方面具有显著优势。这一优化方案不仅有效提高了推理速度,同时也使模型更适用于资源受限的井下环境。UIB 模块结构如图 2 所示。

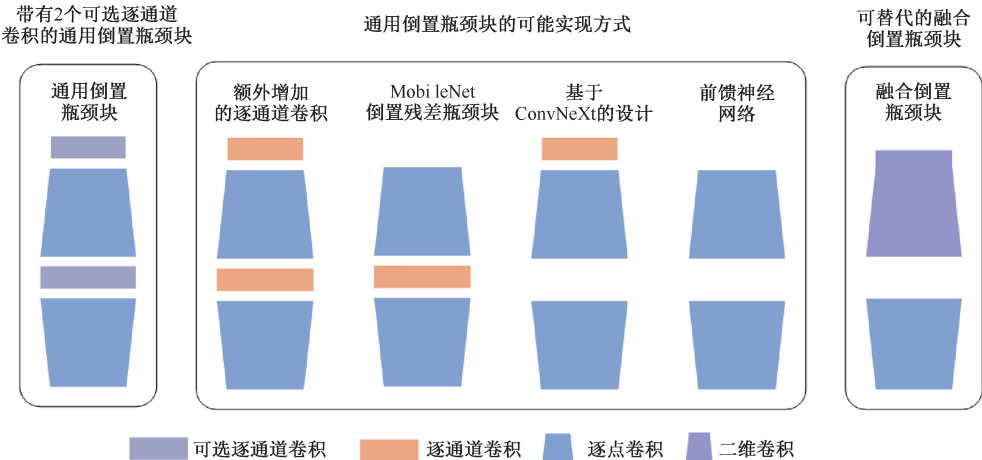


图 2 通用倒置瓶颈模块

Fig. 2 Universal inverted bottleneck blocks

尽管 MobileNetV4 凭借其轻量化设计在计算效率和推理速度上表现优异,但其在特征表达上仍存在不足,尤其在复杂场景中对关键细节的捕捉较为欠缺。因此,在中高层特征提取阶段引入高效通道注意力(Efficient Channel Attention, ECA)^[17]来弥补这一不足,模型如图 3 所示。ECA 模块通过局部跨通道交互实现高效注意力分配,避免了传统全连接层带来的额外计算负担,从而在不增加过多计算负担的同时,增强特征表达能力。通过上述优化,改进后的 YOLOv8m 命名为移动增强型(Mobile-Enhanced, ME)-YOLO。在矿用摄像头等低功耗边缘设备中,部署的目标检测模型需具备轻量化设计、高效特征提取能力、量化与剪枝友好性,并支持多线程视频流处理,以在算力受限的硬件上实现实时检测。ME-YOLO 采用 MobileNetV4+ECA 轻量化骨干网络,优化了模型参数与计算复杂度,同时保持了高效的特征提取能力,满足了矿井复杂环境下的多路视频流实时检测需求。

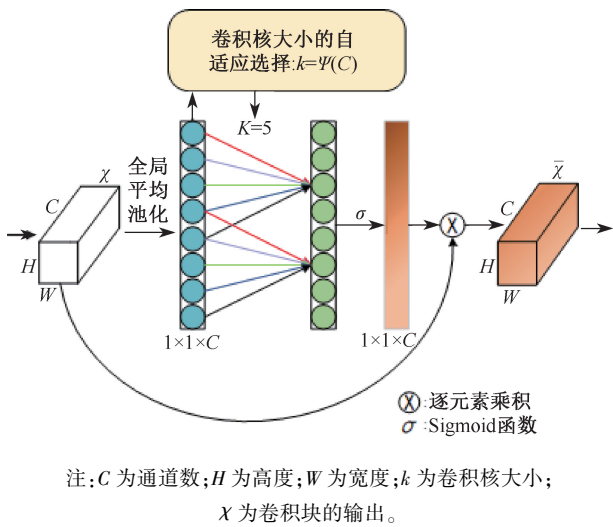


图 3 ECA 模块

Fig. 3 Structure of the ECA module

2.2 基于 Lite-HRNet 的人体骨骼关键点检测

人体骨骼关键点检测是计算机视觉领域的重要研究方向,旨在从静态图像或视频序列中精准定位人体关节关键点,并构建完整的骨架结构。该技术在行为识别、人机交互、安全监测等领域应用广泛。在煤矿井下环境中,通过精准提取姿态信息,对矿工不安全行为识别和安全预警提供了关键数据支持。当前,基于深度学习的人体骨骼关键点检测方法主要分为自上而下和自下而上 2 大策略。自上而下法(如 HRNet、SimpleBaseline)通常采用 2 阶段检测框架,先利用目标检测网络定位人体区域,再在裁剪后

的人体子图上进行关键点检测,虽然能够提供较高的检测精度,但计算成本较高,难以满足实时性要求。相较之下,自下而上方法(如 OpenPose)直接在整个图像中检测、分组与关联所有关键点,具有较高的计算效率,但在人员密集场景中可能存在关键点匹配错误问题。在煤矿井下环境中,由于光照不足、粉尘遮挡以及复杂的背景干扰等因素,为确保安全监测可靠性,不仅需要保持较高的检测精度,同时还必须兼顾计算效率,以适应实时处理要求。

针对煤矿井下环境的特殊需求,采用 Lite-HRNet 检测人体骨骼关键点。作为 HRNet 的轻量化版本,Lite-HRNet 在保持高分辨率特征表达能力的同时,显著降低了计算量,使其更适用于资源受限的环境。相较于传统 HRNet,Lite-HRNet 的核心特点在于引入条件通道加权替换 Shuffle Block 中的 1×1 卷积操作,提高了信息流通效率并降低计算复杂度。使用条件通道加权替换掉卷积后的 Shuffle Block 结构如图 4 所示。此外,该网络采用多尺度特征提取策略,能够在保持全局结构信息的同时,增强对局部关键点的捕捉能力,提升检测精度。在煤矿井下低光照、粉尘干扰等复杂条件下,Lite-HRNet 具备较强的适应性,能够保持较高的关键点检测精度,且其高效结构使其能够在嵌入式设备或边缘计算平台上运行,满足实时监测需求,为煤矿井下安全监测提供高效可靠的技术支持。

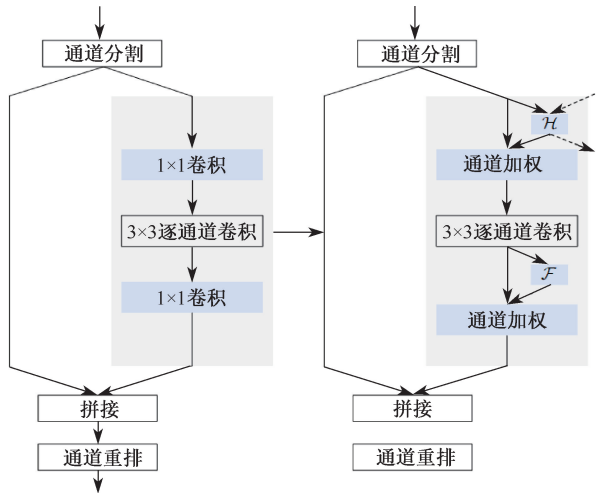


图 4 基于条件通道加权模块的 1×1 卷积替代机制

Fig. 4 Replacement mechanism of 1×1 convolution with conditional channel weighting module

3 多模态特征融合

3.1 RGB 模态特征提取

SlowFast 网络作为 RGB 模态特征提取的关键

架构,凭借其双分支设计在时空信息捕捉方面取得了显著成效。然而,其 Slow 分支在捕捉细粒度空间细节和进行多尺度信息融合时存在一定局限。为解决这一问题,在 Slow 分支的特征提取阶段引入混合注意力模块,结合高效局部注意力(Efficient Local Attention, ELA)^[18]和高效多尺度注意力(Efficient Multi-Scale Attention, EMA)^[19]2种先进的注意力机制,保持较低计算开销的同时强化对局部空间细节的建模能力的同时,提升跨尺度时空特征的交互与融合效果,从而有效增强网络在复杂矿井环境下对不安全行为的识别准确性与鲁棒性。

ELA 通过条带池化提取输入张量的水平和垂直方向特征向量,利用 1D 卷积分别处理这些特征向量以高效编码空间信息,并结合分组归一化(Group Normalization, GN)和非线性激活函数生成位置注意力图,最终通过乘法操作结合 2 个方向的注意力,准确定位感兴趣区域,结构如图 5 所示,该设计避免了通道维度的降维,保持了信息的完整性,同时保持轻量级结构,提升了深度卷积神经网络的空间特征表达能力。

EMA 是一种高效的多尺度注意力机制,旨在提升特征建模过程中全局与局部信息的表达能力,同

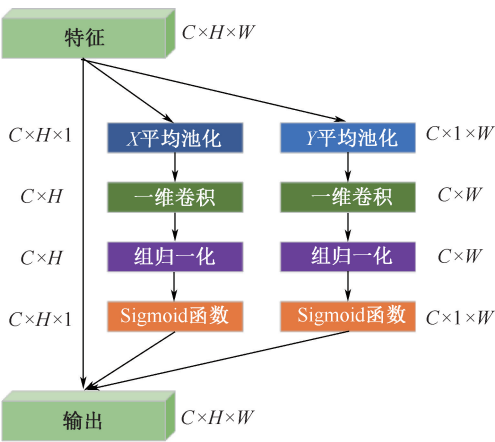
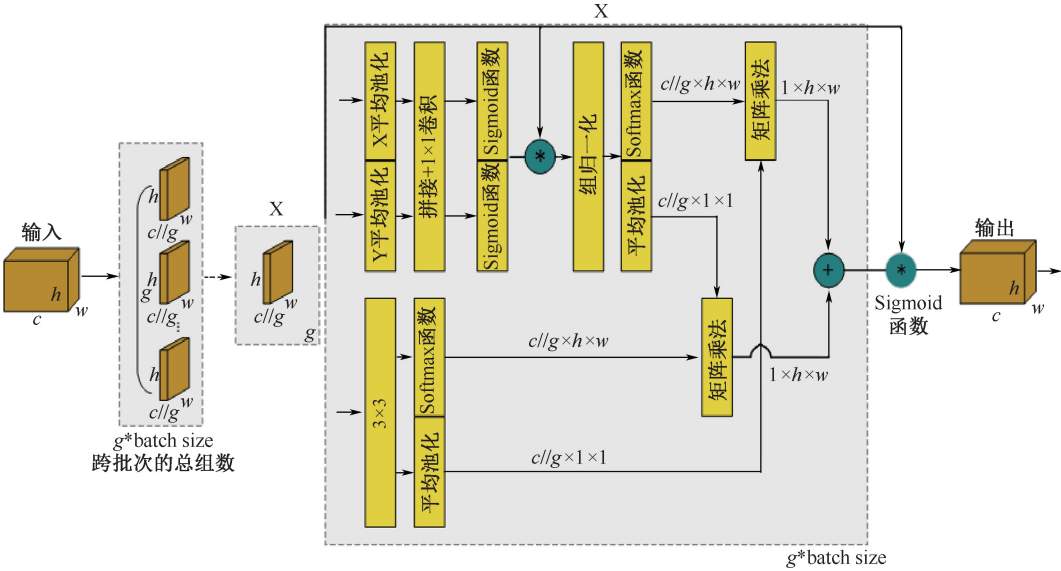


图 5 ELA 结构

Fig. 5 Structure of the ELA module

时保持较高的计算效率。EMA 通过引入跨空间学习机制,实现特征图不同位置间的交互建模,有效捕捉全局上下文依赖关系。同时,EMA 采用多尺度特征聚合策略,在不同尺度上提取并融合特征,使模型能够同时表达全局结构和局部细节。与传统注意力机制相比,EMA 具有更轻量的设计,避免了冗余计算,显著降低了计算复杂度,特别适合煤矿井下等资源受限的场景,模型结构如图 6 所示。



注: g 为分组。

图 6 EMA 结构

Fig. 6 Structure of the EMA module

将 ELA 和 EMA 结合为高效混合注意力模块,并将其集成至 SlowFast 网络的 Slow 分支,形成 EA-SlowFast 网络。这一结合使 SlowFast 网络有效提升了局部细节建模和多尺度特征融合能力,同时保持轻量化设计,在不增加太多参数数量的同时,提升了模

型效率。通过这一改进,基于 RGB 模态的行为识别可直接通过优化后的 SlowFast 网络实现。

3.2 骨骼模态特征提取

在得到骨骼模态数据后,就要基于骨骼模态的

特征提取。骨骼模态的特征提取主要依赖于图卷积网络。其中,ST-GCN 因其能够同时捕捉骨骼点的空间结构关系(人体关节点连接)和时间维度上的动态变化而被广泛应用。但基于 ST-GCN 的行为识别方法存在以下不足:一是其传统卷积层计算复杂度较高,导致参数量庞大,影响模型的计算效率和部署性能;二是特征提取时缺乏全局依赖建模能力,无法有效关注重要通道和空间信息,导致对关键信息的利用不足。

针对这些问题,提出 2 项改进措施。首先,采用深度可分离卷积(Depthwise Separable Convolution, DSC)替代 ST-GCN 传统的标准卷积,DSC 通过 Depthwise 卷积先在每个通道上独立执行空间卷积,

再通过 pointwise 卷积进行跨通道信息融合,在保证特征提取能力的同时,显著降低了参数量和计算开销,提升推理速度,使模型更加轻量化,适用于实时煤矿井下安全监测任务。其次,在 ST-GCN 的图卷积层后引入通道先验卷积注意力(Channel Prior Convolutional Attention, CPCA)^[20],其结构如图 7 所示。CPCA 通过通道和空间注意力的联合建模,先提取全局通道信息,再利用深度卷积模块生成空间注意力图,最终通过自适应调整不同通道和空间区域的权重,使 ST-GCN 能够更有效地提取关键点序列中的重要特征。相比于传统的注意力机制,CPCA 结合了通道先验信息,使得特征提取更加精准,有助于提升行为识别的准确性。

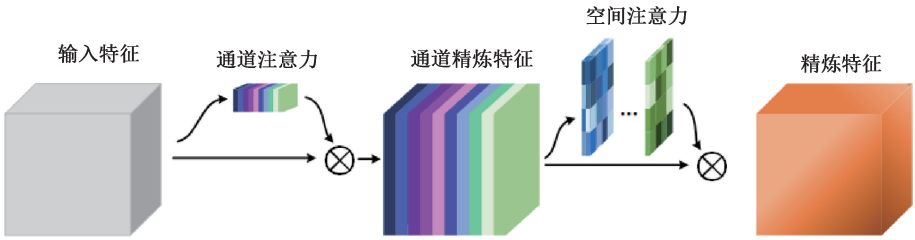


图 7 通道先验卷积注意力

Fig. 7 Channel prior convolutional attention

将上述改进后的网络命名为 DC-STGCN,该网络在骨骼模态特征提取过程中不仅提升了 ST-GCN 的计算效率,同时增强了对关键特征的表达能力。

3.3 特征融合

在利用 EA-SlowFast 提取 RGB 模态特征的同时,运用 ME-YOLO 与 Lite-HRNet 获取人体骨骼关键点数据,并通过 DC-STGCN 提取骨骼模态特征,最终融合 2 种模态。RGB 模态支流凭借其对外观信息和场景背景的丰富感知能力,能够有效捕捉动作的外部表现及其与环境的交互关系,而骨骼模态支流通过人体关键点的结构化表示,专注于行为动作的核心动态特征,从空间和时间维度精准建模人体运动模式,同时具备较强的抗干扰能力。为充分保持各模态特征的完整性与独立性,并挖掘其互补优势,采用特征拼接方式的晚期融合策略,再通过全连接层进行融合特征的建模和分类决策。早期融合对模态对齐要求高,易受噪声干扰,且模态差异大时难以有效整合;中期融合虽然实现模态互补,但计算复杂、设计难度大,容易引入信息冗余;注意力融合虽然能动态调整权重,但模型复杂、计算量大且训练不稳定。相比之下,晚期融合结构简单,能够避免模态间噪声干扰,且对单一模态失效具有较强鲁棒性,更适合矿井复杂环境下的行为识别。

4 矿工不安全行为识别试验与分析

4.1 数据集与试验环境

在公开行为识别数据集(人类动作数据库 51 (Human Motion Database 51, HMDB51),中佛罗里达大学 101 类动作识别数据集(University of Central Florida 101 Action Recognition Dataset, UCF101))与自建井下不安全行为数据集上进行测试验证。UCF101 提供了来自 101 个动作类别的 13 320 段视频。HMDB51 数据库包含有 6 849 段样本,分为 51 类。自建井下不安全行为数据集来自煤矿输送带运输区监控视频,共 900 段,每段视频持续 3~5s,其中 75%作为训练集,25%作为测试集。此外,为使模型充分训练,采用数据增强方法,对样本进行随机裁剪、旋转、放缩等操作。自建数据集通过人工标注,使用专业标注工具逐帧标注视频中的不安全行为,包括不安全行为的起始帧与结束帧;行为类别(如跨越、攀爬等)。标注依据严格参考《煤矿安全规程》及现场安全生产管理规范,确保行为标签与实际生产安全要求一致。

4.2 基于 ME-YOLO 的矿工目标检测

4.2.1 参数设置

矿工目标检测模型训练参数设置如下:算法采

用随机梯度下降优化器,初始学习率为 0.01,权值衰减为 0.000 5,动量值为 0.93,批次大小为 16,训练轮数为 200。

4.2.2 目标检测模型对比试验

为衡量 ME-YOLO 网络模型的检测性能,选择在行业中应用更为广泛且检测效率高的 YOLOv3, YOLOv5s, YOLOv7, YOLOv8m 等 4 种检测模型与本文的 ME-YOLO 模型在自建煤矿输送带运输区监控视频数据集上进行比较和测试。其中,采用平均检测精度(Mean Average Precision, mAP)@0.5 网络的参数量和浮点数运算次数以及帧率(Frames Per Second, FPS)作为评价标准。结果见表 1, ME-YOLO 模型的 mAP@0.5 最高,并在参数量、浮点数运算数和 FPS 方面相比除 YOLOv5s 以外的模型具有显著优势。

表 1 不同目标检测模型的试验结果

Table 1 Experimental results of different object detection models				
模型	mAP@0.5/ %	参数量/ 10 ⁶	浮点运算 数/10 ⁹	FPS/ (帧·s ⁻¹)
YOLOv3	89.8	61.5	154.7	21
YOLOv5s	85.6	7.2	16.5	40
YOLOv7	91.1	36.5	103.3	27
YOLOv8m	90.9	25.9	78.9	35
ME-YOLO	91.8	16.7	47.2	38

4.2.3 目标检测模型消融试验

消融试验结果见表 2。从表 2 看出, YOLOv8m 中加入 ECA 后平均检测精度提升了 1.4%,但降低了其他 3 个指标的效果,将未改进的 MobileNetV4 直接替换 YOLOv8m 的 backbone 后,参数量、浮点运算数、FPS 效果得到提升,而 mAP 下降 1.8%。将加入 ECA 的 MobileNetV4 替换 YOLOv8m 的 backbone 后, mAP 仅次于 YOLOv8m+ECA,但其他 3 个指标效果具有显著优势,说明 ME-YOLO 模型在有较高精度的同时,满足模型轻量化和实时监测需求,适合部署于算力受限的边缘设备。

表 2 消融试验结果

Table 2 Results of ablation experiment				
模型	mAP@ 0.5/%	参数量/ 10 ⁶	浮点运 算数/10 ⁹	FPS/ (帧·s ⁻¹)
YOLOv8m	90.9	25.9	78.9	35
YOLOv8m+ECA	92.3	27.2	81.7	33
YOLOv8m+ MobileNetV4	89.1	15.4	46.4	39
ME-YOLO	91.8	16.7	47.2	38

4.3 煤矿输送带运输区不安全行为识别

4.3.1 试验方案

在公开数据集 UCF-101、HMDB51 以及自建的

井下不安全行为数据集上验证文中的煤矿输送带运输区矿工不安全行为识别模型,并将其与基于单一骨骼模态的 ST-GCN、InfoGCN^[21]以及基于单一 RGB 特征的 C²LSTM^[22]、SlowFast、SIN^[23]进行对比分析。训练前,通过预训练的 ME-YOLO 和 Lite-HRNet 模型从视频数据中提取人体骨骼关键点数据作为识别模型的输入。训练过程中,采用随机梯度下降作为优化器,初始学习率设定为 0.1,并使用余弦退火策略动态调整学习率,权值衰减系数为 0.000 1,动量值设为 0.9,批处理大小为 8,总训练轮数为 160。

4.3.2 公开数据集试验结果分析

表 3 为不同行为识别模型在公开数据集上的对比试验结果,从表 3 中看出,模型在 UCF101 数据集上的表现都强于 HMDB51,在 HMDB51 和 UCF101 数据集上, SIN 和 C²LSTM 分别取得最高识别准确率,为 74.6%和 92.8%。文中模型在所有模型中取得了最优的识别准确率,在 HMDB51 上较 SIN 提高 1.2%,在 UCF101 上较 C²LSTM 提高 1.5%。这一结果表明:该模型能够有效整合 RGB 特征和骨骼模态特征,提升识别性能。

表 3 不同行为识别模型在公开数据集上的试验结果

Table 3 Experimental results of different action recognition models on public datasets				
模型	识别准确率/%		参数量/ 10 ⁶	浮点运算 数/10 ⁹
	HMDB51	UCF101		
ST-GCN	64.5	82.5	3.1	5.4
InfoGCN	67.8	84.9	12.6	15.7
C ² LSTM	61.3	92.8	49.2	74.8
SIN	74.6	91.6	53.2	86.3
Slowfast	68.9	86.8	34.6	65.7
文中模型	75.8	94.3	45.7	71.4

4.3.3 行为识别模型消融试验

基于公共数据集 HMDB51 和 UCF101 进行消融试验,结果见表 4。可以看出,在基于 RGB 特征的 SlowFast 网络中,引入 ELA 和 ECA 注意力机制后,模型在 HMDB51 数据集上的识别准确率分别提升 0.4%和 0.6%,在 UCF101 数据集上分别提升了 0.8%和 0.3%。进一步将二者融合为一个模块后,模型在 HMDB51 和 UCF101 上的准确率分别提升 2.7%和 1.5%,表明该模块能够有效增强 RGB 特征的表达能力。在基于骨骼特征的 ST-GCN 网络中,加入 DSC 后,模型准确率分别下降 0.8%和 1.3%。这可能是由于标准卷积能够直接建模通道间及空间信息的交互,而 DSC 采用分步处理方式,导致部分信息损失。然而,在 ST-GCN 网络中引入通道注意

力机制 CPCA 后,模型在 HMDB51 和 UCF101 数据集上的准确率分别提升了 4.6% 和 3.2%,这表明 CPCA 能更有效地提取关键点序列中的重要特征。当 DSC 和 CPCA 共同作用时,相较于仅引入 CPCA,识别准确率略微下降 0.8% 和 0.3%,但仍显著高于原始 ST-GCN,验证了 CPCA 在骨骼特征提取中的优势。最终将 SlowFast 和 ST-GCN 的识别结果进行晚期融合后,模型在 HMDB51 和 UCF101 数据集上的识别准确率分别达到 73.2% 和 90.9%,均高于任何单一模态特征的改进版本,进一步证明了多模态特征融合的有效性。此外,与原始融合模型相比,EA-SlowFast 和 DC-STGCN 结合后,识别准确率分别提升 2.6% 和 3.4%,表明煤矿输送带区矿工不安全行为识别的模型在特征提取能力和识别精度方面均具有显著优势。

表 4 消融试验结果

Table 4 Results of ablation experiment %		
模型	识别准确率	
	HMDB51	UCF101
Slowfast	68.9	86.8
ST-GCN	64.5	82.5
Slowfast+ST-GCN	73.2	90.9
Slowfast+ELA	69.3	87.6
Slowfast+ECA	69.5	87.1
Slowfast+ELA+ECA	71.6	88.3
ST-GCN+DSC	63.7	81.2
ST-GCN+CPCA	69.1	85.7
ST-GCN+DSC+CPCA	68.3	85.4
本文模型	75.8	94.3

4.3.4 自建井下不安全行为数据集试验结果及分析

在自建井下不安全行为数据集上开展试验验证,不同行为识别模型在测试集上的识别准确率见表 5,可以看出,在煤矿输送带区的行为识别任务中,文中提出的模型展现出较高的识别准确率,明显优于仅基于单一骨骼模态或 RGB 模态的行为识别模型。表明所提模型在复杂井下环境中,尤其是在矿工不安全行为识别方面具有更强的适用性。

为全面评估识别模型在区分不同动作方面的表现,图 8 展示了混淆矩阵,矩阵中呈现了 4 种不安全行为的识别结果:跨越输送带、手扶输送带、攀爬输送带和倚靠输送带。这 4 种动作的平均识别准确率为 90.6%。其中,跨越输送带和倚靠输送带的识别准确率较高,因为这 2 个动作相对简单,容易识别。而手扶输送带的识别准确率较低,可能是由于该动作较为细微,容易受到光照、遮挡等环境因素的影响。

表 5 自建井下不安全行为数据集识别结果

Table 5 Recognition results on the self-constructed underground unsafe behavior dataset %	
模型	识别准确率
ST-GCN	73.9
InfoGCN	75.4
C ² LSTM	83.2
SIN	86.7
Slowfast	81.5
文中模型	90.6

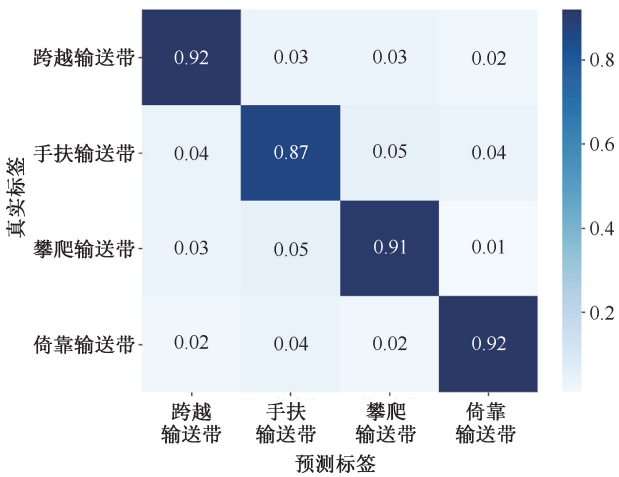


图 8 不安全行为识别混淆矩阵

Fig. 8 Confusion matrix for unsafe behavior recognition

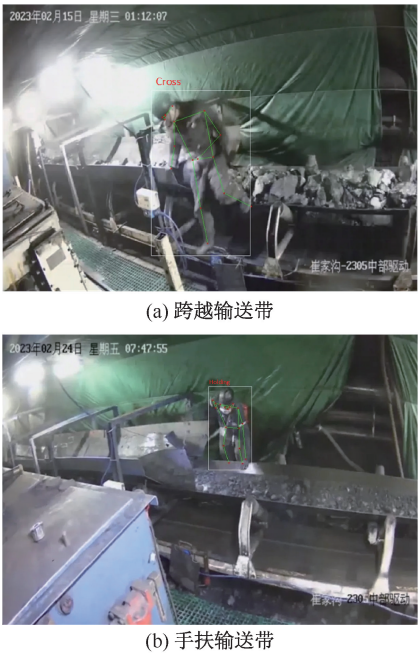


图 9 多模态特征融合的不安全行为识别结果

Fig. 9 Recognition results of unsafe behavior based on multimodal feature fusion

响。本文模型在井工煤矿输送带区对矿工不安全行为的部分识别结果如图9所示,可以看出,文中模型能够对煤矿输送带区矿工的不安全行为进行有效识别与区分。

5 结 论

1) 将 ME-YOLO 模型与 YOLOv3, YOLOv5s, YOLOv7, YOLOv8m 在自建井下不安全行为数据集上进行对比,试验结果表明:ME-YOLO 的 mAP@0.5 最高,达到 91.8%。

2) 在公开数据集 HMDB51 以及 UCF101 上的

结果表明:在基于单一模态的行为识别模型中,改进后的 EA-SlowFast 与 DC-STGCN 的识别准确率均比未改进的初始方法高,在 HMDB51 上分别高出 2.7% 和 3.8%;在 UCF101 上分别高出 1.5% 和 2.9%。对比未改进的融合模型,改进后的行为识别模型在 2 个数据集上的准确率分别提高到了 75.8% 和 94.3%。

3) 在自建井下不安全行为数据集上的试验结果表明:在煤矿输送带区环境中,文中的提出的行为识别模型表现出最高的识别准确率,达到 90.6%,能够精准识别煤矿输送带区矿工不安全行为。

参 考 文 献

- [1] 王露露,尘兴邦,袁嘉淙,等. 职业心理学视域下矿工不安全行为研究综述[J]. 中国安全科学学报, 2023, 33(1): 48-55.
WANG Lulu, CHEN Xingbang, YUAN Jiacong, et al. Review of research on miners' unsafe behaviors in perspective of occupational psychology[J]. China Safety Science Journal, 2023, 33(1): 48-55.
- [2] 王海军,齐庆杰,梁运涛,等. 我国煤矿重特重大事故统计分析及对策建议[J]. 中国安全科学学报, 2024, 34(9): 9-18.
WANG Haijun, QI Qingjie, LIANG Yuntao, et al. Statistical analysis and countermeasures of major accidents in coal mines in China[J]. China Safety Science Journal, 2024, 34(9): 9-18.
- [3] 沈铭华,马昆,杨洋,等. AI 智能视频识别技术在煤矿智慧矿山中的应用[J]. 煤炭工程, 2023, 55(4): 92-97.
SHEN Minghua, MA Kun, YANG Yang, et al. Application of ai identification technology in intelligent coal mine[J]. Coal Engineering, 2023, 55(4): 92-97.
- [4] XU Leiyang, WANG Qiang, LIN Xiaotian, et al. Skeleton-based tai chi action segmentation using trajectory primitives and content[J]. Neural Computing and Applications, 2023, 35(13): 9 549-9 566.
- [5] WU Yangkai, QIU Luhua, WANG Jinming, et al. The use of convolutional neural networks for abnormal behavior recognition in crowd scenes [J]. Information Processing & Management, 2025, 62(1): DOI: 10.1016/j. ipm. 2024. 103880.
- [6] HUANG Zengxi, QIN Yusong, LIN Xiaobing, et al. Motion-driven spatial and temporal adaptive high-resolution graph convolutional networks for skeleton-based action recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(4): 1 868-1 883.
- [7] HE Junyan, WU Xiao, CHENG Zhiqi, et al. DB-LSTM: densely-connected Bi-directional LSTM for human action recognition[J]. Neurocomputing, 2021, 444: 319-331.
- [8] CAO Xiangang, ZHANG Chiyu, WANG Peng, et al. Unsafe mining behavior identification method based on an improved st-gen[J]. Sustainability, 2023, 15(2): DOI: 10.3390/su15021041.
- [9] 李占利,权锦成,靳红梅. 基于 3D-Attention 与多尺度的矿井人员行为识别算法[J]. 国外电子测量技术, 2023, 42(7): 95-104.
LI Zhanli, QUAN Jincheng, JIN Hongmei. Mine personnel behavior recognition algorithm based on 3D-Attention and multi-scale[J]. Foreign Electronic Measurement Technology, 2023, 42(7): 95-104.
- [10] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 4 489-4 497.
- [11] 饶天荣,潘涛,徐会军. 基于交叉注意力机制的煤矿井下不安全行为识别[J]. 工矿自动化, 2022, 48(10): 48-54.
RAO Tianrong, PAN Tao, XU Huijun. Unsafe action recognition in underground coal mine based on cross-attention

mechanism[J]. Journal of Mine Automation, 2022, 48(10): 48-54.

[12] FEICHTENHOFER C, FAN Haoqi, MALIK J, et al. Slowfast networks for video recognition[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6 202-6 211.

[13] CAO Zhong, CHEN Kaihong, CHEN Junzhuo, et al. CACS-YOLO: a lightweight model for insulator defect detection based on improved YOLOv8m[J]. IEEE Transactions on Instrumentation and Measurement, 2024; DOI: 10.1109/TIM.2024.3453332.

[14] YU Changqian, XIAO Bin, GAO Changxin, et al. Lite-HRNet: a lightweight high-resolution network[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10440-10450.

[15] 王宇,于春华,陈晓青,等. 基于多模态特征融合的井下人员不安全行为识别[J]. 工矿自动化, 2023, 49(11): 138-144.

WANG Yu, YU Chunhua, CHEN Xiaoqing, et al. Recognition of unsafe behaviors of underground personnel based on multi modal feature fusion[J]. Journal of Mine Automation, 2023, 49(11): 138-144.

[16] ZHOU Huanxin, LI Wenhao, WEI Shuiyi, et al. Steel surface defect detection method based on YOLOv11-MobileNetv4[J]. International Core Journal of Engineering, 2025, 11(2): 10-16.

[17] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11 534-11 542.

[18] HAO Shengnan, LI Xinlei, PENG Wei, et al. YOLO-CXR: a novel detection network for locating multiple small lesions in chest X-ray images[J]. IEEE Access, 2024, 12; DOI:10.1109/ACCESS.2024.3482102.

[19] OUYANG Daliang, HE Su, ZHANG Guozhong, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[20] HUANG Hejun, CHEN Zuguo, ZOU Ying, et al. Channel prior convolutional attention for medical image segmentation[J]. Computers in Biology and Medicine, 2024, 178; DOI: 10.1016/j.combiomed.2024.108784.

[21] CHI Seunggeun, CHI HyungGun, HUANG Qixing, et al. InfoGCN++: learning representation by predicting the future for online skeleton-based action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024; DOI: 10.1109/TPAMI.2024.3466212.

[22] MAJD M, SAFABAKHSH R. Correlational convolutional LSTM for human action recognition[J]. Neurocomputing, 2020, 396: 224-229.

[23] SHIQING Hongya, ZHANG Hongbo, LI Zhe, et al. Shuffle-invariant network for action recognition in videos[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2022, 18(3): 1-18.

作者简介： 郝秦霞 （1980—）,女,陕西西安人,博士,副教授,主要从事物联网工程应用,煤矿安全方面的研究。E-mail:haoqinxia@xust.exu.cn。