中文引用格式:王海泉,于浩玮,杨岳毅,等. 工业场景下人员行为的多模态信息融合决策策略[J]. 中国安全科学学报,2025,35(8):84-92.

英文引用格式:WANG Haiquan, YU Haowei, YANG Yueyi, et al. Multimodal information fusion decision-making strategy for personnel behavior in industrial scene[J]. China Safety Science Journal, 2025, 35(8): 84-92.

工业场景下人员行为的多模态信息融合决策策略*

王海泉[1]教授, 于浩玮[2], 杨岳毅[1]讲师, 徐晓滨[3]教授,
卜祥洲[4]高级工程师, KURKOVA P[5]教授

(1 中原工学院 智能感知与仪器学院,河南 郑州 450007;2 中原工学院 自动化与电气工程学院,河南 郑州 450007;3 杭州电子科技大学 自动化学院,浙江 杭州 310018;4 河南宏博测控技术有限公司,河南 郑州 450040;5 圣彼得堡国立宇航与仪器制造大学 电子与激光仪器系,俄罗斯 圣彼得堡 14-51)

中图分类号:X928.03 **文献标志码**:A **DOI**: 10.16265/j.cnki.issn1003-3033.2025.08.0084

资助项目:河南重点研发专项项目(251111211600);河南省科技攻关项目(242102320215);河南省高端外国专家引进计划项目(HNGD2024032);中原工学院学科实力提升计划项目(GG202412)。

【摘 要】 为预防工业场景下人员不安全生产行为所导致的安全事故,解决光线不佳、视野受限和遮挡等干扰情况下单一视觉模态动作识别效果不佳的问题,提出一种基于自适应证据推理(S-ER)策略,融合视频信息和惯性测量元件(IMU)信息的人员不安全行为决策方法。首先,构建基于注意力机制的多任务三维卷积模型(M-C3D),分析视频信息,运用融合注意力机制的一维卷积神经网络(1D-CNN)处理IMU信息;其次,运用证据推理(ER)策略实现决策级融合,并通过萤火虫优化算法构建不同环境条件下证据权重和可靠度的优化集合,确保视频和传感器模态信息的权重能够根据环境情况自适应调整;最后,通过德克萨斯大学达拉斯分校的多模态人员行为数据集(UTD-MHAD)和中原工学院的多模态人员行为数据集(ZUT-MHAD)验证模型的有效性。结果表明:在存在干扰的工业场景中,S-ER方法的识别准确率最高可达98.53%,较传统多模态融合方法及单模态识别方法的最高值提升17.52%。

【关键词】 工业场景; 多模态信息; 信息融合; 行为识别; 证据推理(ER)策略

Multimodal information fusion decision-making strategy for personnel behavior in industrial scene

WANG Haiquan[1], **YU Haowei**[2], **YANG Yueyi**[1], **XU Xiaobin**[3], **BU Xiangzhou**[4], **KURKOVA P**[5]

(1 College of Intelligent Sensing and Instrumentation, Zhongyuan University of Technology, Zhengzhou Henan 450007, China; 2 School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou Henan 450007, China; 3 College of Automation, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China; 4 Henan Hongbo Measurment and Control Co., Ltd., Zhengzhou Henan 450040, China; 5 School of Electronic and Laser Instrument, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg 14-51, Russia)

* 文章编号:1003-3033(2025)08-0084-09; 收稿日期:2025-03-01; 修稿日期:2025-05-13

Abstract: In order to reduce the accidents in industrial scenarios which were caused by workers´unsafe operation behaviors, meanwhile improve the performance of visual-based action recognition methods in industrial scene with poor lighting, limited field of view and occlusions, an improved decision-making strategy based on self-adaptive ER (S-ER) was introduced in this paper. This strategy could integrate video information and inertial measurement unit (IMU) information effectively. It firstly analyzed video information and IMU information with attention mechanism-based multi-task convolutional 3D (M-C3D) model as well as one-dimensional convolutional neural network (1D-CNN) fused with attention mechanism, then ER theory was introduced to achieve decision-level fusion, where the set of evidence weights and reliability under different environmental conditions was optimized through the firefly optimization algorithm for improving the recognition accuracy and robustness of the model. The effectiveness of the proposed algorithm was verified on the public dataset Multimodal Human Action Dataset from University of Texas at Dallas (UTD-MHAD) and the self-built dataset Multimodal Human Action Dataset from Zhongyuan University of Technology (ZUT-MHAD). The results show that the identification results of S-ER for workers´unsafe behaviors in complex industrial scenarios can reach up to 98.53%, which is 17.52% higher than the maximum value of traditional multimodal fusion methods and single-modality recognition methods.

Keywords: industrial scene; multimodal information; information fusion; action recognition; evidence reasoning (ER) theory

0 引　言

在工业领域中,准确判断生产人员的不安全生产行为有极为深远的意义[1]。传统的行为识别和监测大多采用人工巡检或视频巡检的方式,该方式效率低、准确性差。随着人工智能、物联网技术的发展,基于计算机视觉[2-3]和惯性测量元件(Inertial Measurement Unit, IMU)[4]的人员生产行为识别方法得到广泛应用。但基于机器学习和计算机视觉算法的识别方法受光线、遮挡等环境因素影响较大,而基于 IMU 的识别方法又易受传感器位置偏移、信号噪声等影响[5]。因此,结合计算机视觉与 IMU 传感器的优势,利用多模态信息融合技术提升工业场景下人员行为识别的准确性,具有非常重要的现实意义。

目前,基于多模态信息融合的不安全行为识别方法包括决策级融合[6]、特征级融合[7-8]、混合级融合等[9-10]。当不同模型间存在显著的不相关性,尤其是维度不一致时,更合适采用决策级融合,以减少计算量,并降低不同模态间的影响[11]。D-S(Dempster-Shafer)证据理论具有较强的融合多模态信息的能力[12]。ZHANG Guoliang[13]利用 D-S 证据理论融合了视频的时空信息和骨架信息,识别了跳跃、行走和跑步等动作。YANG Jianbo 等[14]基于 D-S 证据理论,提出考虑证据的权重和可靠度的证据推理(Evidence Reasoning, ER)策略,有效解决了 D-S 证据理论存在的与常识判断相悖的问题,并被广泛应用于故障诊断[15]、多属性决策[16]、组件分类[17]、事故分析[18]、健康评估[19]等领域。针对上述方法参数固定、无法动态调整多模态信息融合权重的问题,学者们提出构建最优融合参数集,在融合前依据场景自动选参的方案。赖尉文等[20]提出将各模型的结果作为证据,建立不同证据融合的参数集,同时,以随机森林分类的结果为依据选择参数集。赵蕊蕊等[21]将模型结果作为证据,通过粒子群算法优化模型获取的信息权重,提高了遥感图像蓝藻水华的识别能力。尽管这类方法在一定程度上具备自适应性,但其易受输出结果误差或异常值的干扰,难以保证在模态失效和数据扰动等场景下的稳定性。

鉴于此,笔者拟提出一种面向工业场景的多模态信息自适应融合方法,基于原始数据信息,借助萤火虫算法(Firefly Algorithm, FA)[22]构建由权重和可靠度组成的参数集,通过识别生产环境中光线、遮挡等条件的变化,从参数集中动态选择相应的融合权重,实现多模态信息的综合分析,使推理过程具有更好的灵活性和可解释性,提升复杂工业场景下人员生产行为识别的准确性,以期提升企业安全生产水平和管理质量。

1 基于改进 ER 的多模态信息融合决策策略

1.1 单模态决策模型

在视觉模态层面,鉴于 3D 卷积网络模型[23]能够有效处理视频帧序列的信息,基于 TRAN 等[24]提出的 3D 卷积神经网络(Convolutional 3D, C3D)模型,并结合高效通道注意力机制(Efficient Channel Attention, ECA)[25],构建多任务三维卷积模型(Multi-task Convolutional 3D, M-C3D),其网络结构如图 1 所示。其中,conv1—conv5 分别表示 5 个初始卷积模块,64、128、256、512 分别为卷积模块中滤波器的数量,2 组全连接层为每个任务特定头,处理共享特征并产生特定于每个任务的输出。这种设计能够改善卷积特征图中每个通道维度的自适应加权效果,提高特征识别的精确度和模型的分类性能。

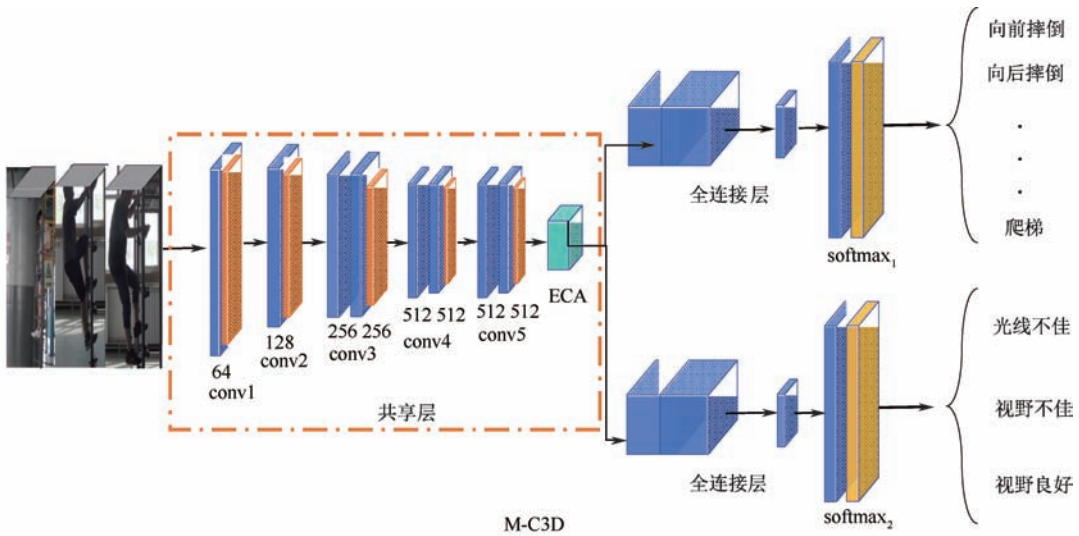

图 1 M-C3D 网络结构

Fig. 1 M-C3D network structure

选择一维卷积神经网络(1D Convolutional Neural Network, 1D-CNN)处理一维的 IMU 信息,利用通道注意力(Channel Attention Module, CAM)调整不同通道内的权重,并借助激励注意力(Excitation Attention Module, EAM)[26]精确捕捉与任务相关的关键时刻,提升学习效率,实现对关键时序片段和敏感传感维度的双重关注。

IMU 模态模型结构如图 2 所示。该模型包括 5 层一维卷积层,每一层卷积层均依次连接 CAM 模块和 EAM 模块。在第 5 层卷积层后,模型通过全局平均池化层、softmax 层完成最终的分类任务。

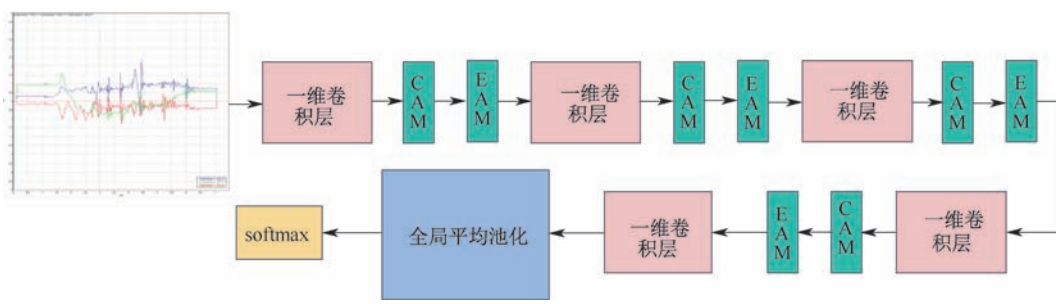

图 2 IMU 模态模型结构

Fig. 2 IMU modal model structure

1.2 多模态信息决策级融合策略

为有效融合单一模态决策信息,提升工业场景下不安全生产行为识别的准确性,提出基于 ER 规则的自适应 ER(Self-adaptive ER, S-ER)方法,通过引入环境感知机制、参数集动态调节策略,关联环境和模态的权重,以解决多模态信息融合中模态失效和环境适应能力差的问题,多模态信息自适应决策级融合策略如图 3 所示。

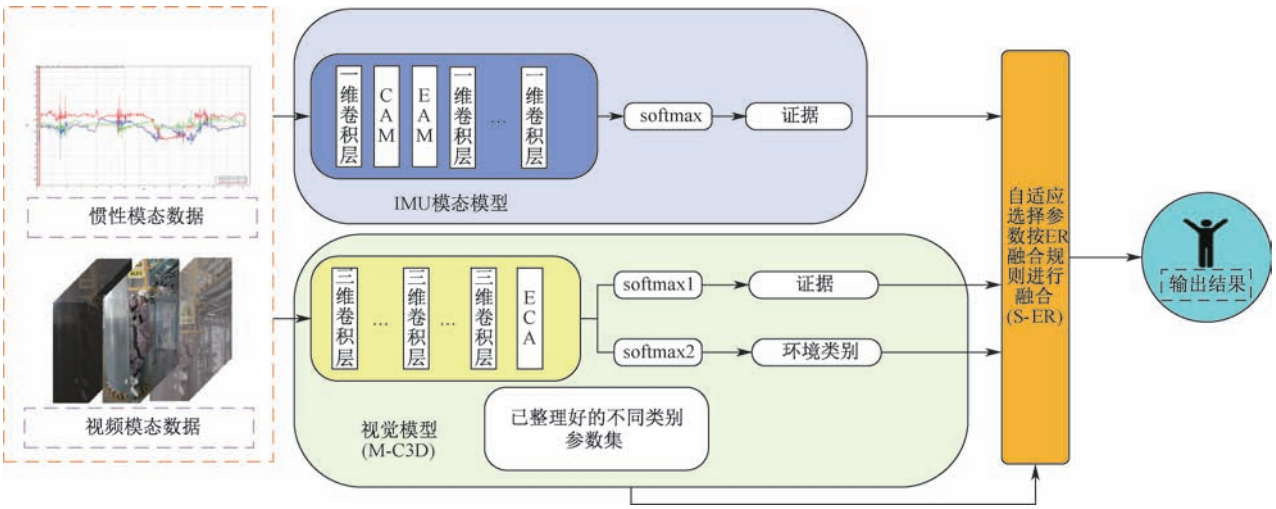

图 3　多模态信息自适应决策级融合架构

Fig. 3　Adaptive decision-level fusion architecture based on multimodal information

1.2.1　ER 原理

ER 方法是一种用于处理多个相互独立证据的规则,能综合考虑证据的可靠性和权重。假设在某个分类任务中,有 L 条独立的证据 e_i ($i=1,2,\cdots,L$),并且辨识框架 Θ 由 N 个类别 θ_N 组成,即 $\Theta=\{\theta_1,\theta_2,\cdots,\theta_N\}$,则第 j 条证据的置信分布表示为:

$$e_j=\{(\theta,\Omega_{\theta,j})\mid \forall\theta\subseteq\Theta,\sum_{\theta\subseteq\Theta}\Omega_{\theta,j}=1\} \quad (1)$$

式中:$\Omega_{\theta,j}$ 为支持类别 θ 和第 j 条证据的信度;θ 为类别,可取幂集 $P(\Theta)$ 中除空集外的任一元素。

ER 中每条证据 e_j 的特性都由可靠性因子和重要性权重来衡量。可靠性因子是证据的固有特性,反映证据 e_j 本身的可信度或质量;重要性权重是带有决策者主观判断的可信度。对于多个信息源提供的相互独立的证据,通过 ER 规则,融合得到类别 θ 的信度。

1.2.2　S-ER 策略

为提升 ER 策略在应对不同状态动作识别任务时的灵活性,引入外部环境识别结果,融入 ER 策略构建基于 S-ER 规则的多模态决策级信息融合策略,如图 4 所示。

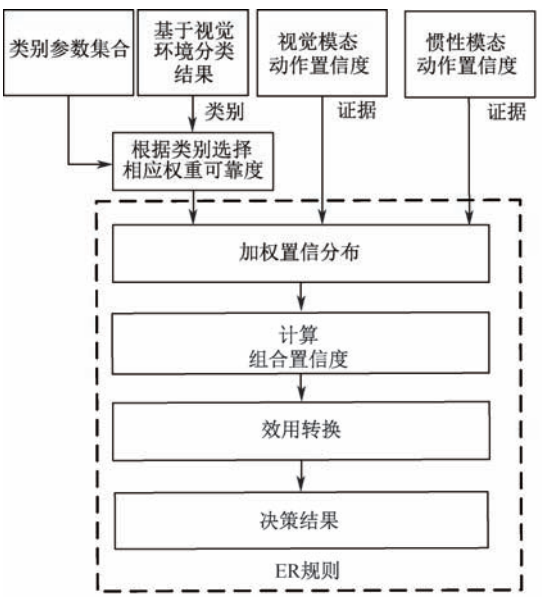

图 4　基于 S-ER 规则的多模态决策级信息融合策略

Fig. 4　Multimodal decision-level fusion strategy based on S-ER

该策略的输入包括惯性模态和视觉模态判断的结果 $e_{1\alpha}$、$e_{2\alpha}$,环境分类结果 Q 和类别参数集 ε。根据客观环境的识别结果 Q 选择类别参数集 ε 中的参数,将得到的可靠性因子和重要性权重分别赋值 $e_{1\alpha}$ 和 $e_{2\alpha}$,并通过 ER 规则进行融合,输出对动作的识别结果。融合过程可表示为:

$$h=f([Z_{11}\cdot Z_{21}],[Z_{12}\cdot Z_{22}],\cdots,[Z_{1\alpha}\cdot Z_{2\alpha}],\tau) \quad (2)$$

式中:h 为最终分类结果;$f([\cdot])$ 为 ER 策略下每个模态模型的融合过程;α 为需要判断的动作类型;$Z_{1\alpha}$ 和 $Z_{2\alpha}$ 分别为 IMU 模态模型和视频模态模型的分类结果,同时也是 ER 策略的证据;τ 为 S-ER 规则的自适应融合过程,可表示为:

$$\tau=g([Z_{11}\cdot Z_{21}],[Z_{12}\cdot Z_{22}],\cdots,[Z_{1\alpha}\cdot Z_{2\alpha}],\varepsilon) \quad (3)$$

式中 $g([\,.\,])$ 为 S-ER 规则下每个模态模型的融合过程。根据环境识别结果,模型实时地从预先构建的参数集中自适应地选取合适的参数组合,实现融合策略的动态调节。

整个动态信息融合过程中,参数集的选择是最核

心的一环,是每种环境类别 Q 对应的一组最优权重与可靠度,其最优值采用萤火虫算法(Firefly Algorithm, FA)寻优确定。参数集构建流程如图 5 所示。

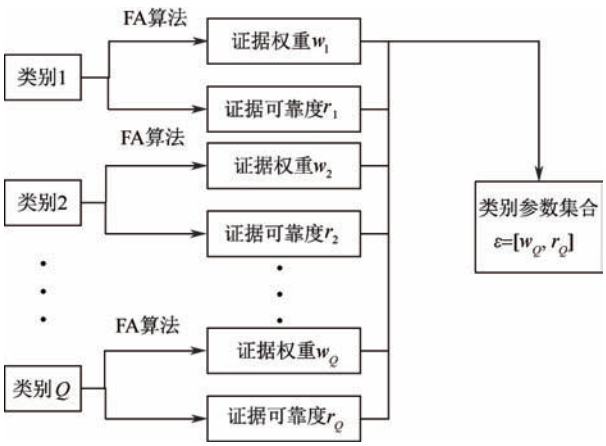

图 5 参数集构建流程

Fig. 5 Construction flow of parameter set

FA 模拟萤火虫利用自身发出的光亮吸引其他个体的特性,通过不断向光亮度更高的个体移动,从而找到最优解。其中,萤火虫表示最优解,光亮度大小表示每个解的适应度值,即通过 ER 规则融合后的准确率。

初始萤火虫 i 的位置 X_i 为:

$$X_i = (x_{i1}, x_{i2}, \cdots, x_{iM}) \tag{4}$$

式中 $x_{i1}, x_{i2}, \cdots, x_{iM}$ 为萤火虫 i 在维度 M 下的坐标位置。

每只萤火虫间的相互吸引力 ρ_{ij} 为:

$$\rho_{ij} = \rho_0 \exp(-\gamma \xi_{ij}) \tag{5}$$

式中:ρ_0 为最大吸引度($\xi_{ij}=0$),一般取值为 1;γ 为光吸收系数;ξ_{ij} 为萤火虫 i 和 j 间的笛卡尔距离。

当发现更亮的萤火虫时,其他萤火虫位置移动过程可表示为:

$$X_i^{t+1} = X_i^t + \rho_{ij}(X_j^t - X_i^t) + \sigma\varepsilon \tag{6}$$

式中:t 为迭代次数;ε 为满足高斯分布的随机数;σ 为随机项系数,$\sigma\varepsilon$ 用以扰动处于最优位置的萤火虫;$X_j^t - X_i^t$ 为不同萤火虫之间的距离,计算如下:

$$X_j^t - X_i^t = (x_{j1}^t - x_{i1}^t, x_{j2}^t - x_{i2}^t, \cdots, x_{jM}^t - x_{iM}^t) \tag{7}$$

视觉模态模型与 IMU 模态模型通过 FA 算法寻找到最佳权重 w_Q 后,由 ER 算法融合求得最终结果。

综上所述,S-ER 规则模型的自适应信息融合流程可表述为:

1) 将图像信息和惯性信息作为输入,图像信息由 M-C3D 模型分类,得到 Z_{2p} 和 Q_P;惯性信息经过惯性模型输出得到 Z_{1p}。其中,Q_P 作为预设类别。

2) 模型训练时,利用 FA 算法通过准确率反推不同的类别 Q 下的证据权重和证据可靠度,统计不同类别下每组权重和可靠度的集合,得到参数集 $W(\cdot)$ 与 $R(\cdot)$。

2 人员行为决策策略验证

2.1 试验数据集

德克萨斯大学达拉斯分校的多模态人员行为数据集(Multimodal Human Action Dataset from University of Texas at Dallas, UTD-MHAD)涵盖了视频、深度视频、骨架和惯性 4 种模态信息,共包含 8 名被试的 27 种不同动作,其中包含 21 种手部动作和 6 种腿部动作。每名被试需重复执行每种动作 5 次。为方便对比该数据集的使用方法[27-31],数据集同样划分为特定对象(Subject Specific, SS)和通用对象(Subject Generic, SG)2 类。其中,SS 将 1、3、5 和 7 号被试的数据作为训练集,其余被试的数据作为测试集;而 SG 将所有被试第 2—第 4 次采集的数据作为训练集,第 1 次采集的数据作为测试集。中原工学院的多模态人员行为数据集(Multimodal Human Action Dataset from Zhongyuan University of Technology, ZUT-MHAD)是针对工业场景下人员不安全生产行为的自建多模态数据集,包含视频和惯性传感器 2 种模态的数据。惯性传感器数据来源于固定在生产人员身体的 2 个无线惯性传感器节点,由 3 轴加速度计和 3 轴陀螺仪组成,采样率为 128 Hz。视频信号则来自能够覆盖生产环境的红绿蓝三原色(Red Green Blue, RGB)摄像头,分辨率为 1 080 p,帧率为 25 帧/s。

邀请 15 名被试参与构建数据集,年龄 22 ~ 50 岁(平均值为 27.1 岁,标准差为 3.87 岁),身高 160 ~ 190 cm(平均值为 173.5 cm,标准差为 8.14 cm),体质量 50 ~ 90 kg(平均值为 68.64 kg,标准差为 10.10 kg)。

ZUT-MHAD 数据采集场景和动作类别如图 6 和图 7 所示。每个动作执行 4 次,每次持续 2 ~ 3 s。测试时,SS 将第 1—10 名被试的数据作为训练集,第 11—15 名被试的数据作为测试集;SG 将 15 名被试在每种场景下的前 3 次动作作为训练集,其余动作数据作为测试集。

考虑到视觉模态(25 帧/s)与 IMU 模态(128 Hz)帧率差异较大,记录视频图像与 IMU 信号

统一的起止时间和时间戳,确保不同模态数据在时间轴上的一致性。起始时刻一致,总采集时长相同,从而实现视觉与惯性模态在采集阶段的时间同步。若动作视频的总帧数为 T,每 16 帧视频图像被视为一个有效片段,输入至 M-C3D 模型,则 IMU 数据的滑动窗口长度 l 为:

$$l = \frac{T}{16 \times k} \tag{8}$$

式中 k 为 IMU 数据的总行数。

(a) 光线良好无遮挡场景　(b) 光线不佳有遮挡场景　(c) 光线不佳无遮挡场景

图 6　构建 ZUT-MHAD 数据集的不同场景

Fig. 6　Different environments for constructing ZUT-MHAD dataset

(a) 向前摔倒　(b) 向后摔倒　(c) 爬梯子　(d) 电焊　(e) 清理　(f) 爬行

(g) 行走　(h) 上楼梯　(i) 下楼梯　(j) 拧阀门　(k) 寻求救援　(l) 坐

图 7　不同动作类别

Fig. 7　Different actions

2.2　试验结果

2.2.1　UTD-MHAD 数据集验证结果

UTD-MHAD 数据集与文献[27-31]的方法对比结果见表 1。模态信息 D 为带深度信息的图片,S 为骨架信息,V 为 RGB 图片信息,I 为惯性信息。由表 1 可知:S-ER 在 UTD-MHAD 数据集上应用时,在 SS 基准上,相较于文献[28]和文献[31],最高的准确率分别提升 2.6% 和 8.6%。在 SG 基准上,相较于其他文献,准确率最高能提升 2.3%。对于模型的识别时间,S-ER 方法从多模态数据输入到模型识别输出结果的平均时间为 0.413 s,对比文献[28]的平均计算时间 0.526 s,其运算速度提升明显,验证了 S-ER 方法良好的信息融合和识别能力。

2.2.2　ZUT-MHAD 数据集动作识别结果

针对 ZUT-MHAD 数据集,在工业场景下,光线良好无遮挡、光线不佳无遮挡、光线不佳有遮挡情况时的数据量为 1∶1∶1,动作识别结果如图 8 所示。将改进前后的 ER 规则应用于 SG 基准,混淆矩阵如图 9 所示。显然,S-ER 方法识别 12 个动作的效果较好,均高于传统单模态识别方法和 DS/ER 规则融合方法的结果,其中改进 S-ER 方法相比视觉或惯性传感器的单模态识别方法的识别精度在 SS 基准

表 1 UTD-MHAD 数据集对比结果

Table 1 Comparative results on UTD-MHAD dataset

数据集		模态信息	准确率/%	方法
UTD-MHAD	SS	D,S	90.0	多模态协同自注意力动作识别网络(Multimodal Cooperative Self-attention Network for Action Recognition,MGAT)[31]
		D,S,V	96.0	焦点通道知识蒸馏的多模态行为识别方法(Focal Channel Knowledge Distillation for Multi-Modality Action Recognition,FCKD)[29]
	SG	V,I	98.60	S-ER
		D,I	97.20	常州大学综合多模态人员识别方法(Changzhou University: Comprehensive Multi-modal Human Action, CZU-MHAD)[27]
		V,S,I	97.90	堆叠密集流差分图像(Stacked Dense Flow Difference Image,SDFD)[30]
		V,I	98.75	新型监督式自适应多模态融合方法(New Supervised Adaptive Multi-modal Fusion Method, AMFM)[28]
		V,I	99.50	S-ER

下最高能提升 17.52%,相比其他未改进融合方法,识别精度最高能提升 5.35%;而在 SG 基准下,S-ER 方法相比其他单模态方法,准确率最高能提升 15.85%,相比未改进融合方法,准确率最高提升 4.62%。而与改进前 ER 算法相比,从多模态数据输入到模型识别输出结果的平均时间仅增加 0.099 s,而平均准确率提升了 2%。

当工业场景环境变化时,视觉模型的准确率也发生变化。针对不同的环境类别,采用 FA 优化 ER 规则库的最佳参数组合,ZUT-MHAD 数据集在 SG 基准下的识别效果见表 2。

表 2 ZUT-MHAD 数据集 SG 基准下 S-ER 识别效果

Table 2 S-ER recognition performance under SG benchmark in ZUT-MHAD dataset

环境类别	M-C3D 准确率/%	1D-CNN 准确率/%	M-C3D 权重	1D-CNN 权重
光线良好	90.00	82.60	0.85	0.55
光线不佳	83.51	82.60	0.78	0.64
有遮挡时	77.80	82.90	0.55	0.67

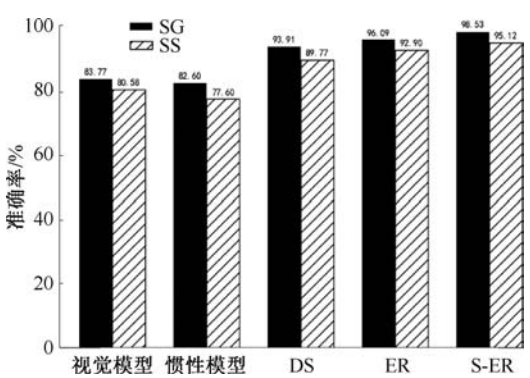

图 8 在 ZUT-MHAD 数据集上的对比试验

Fig. 8 Comparative experiments on ZUT-MHAD dataset

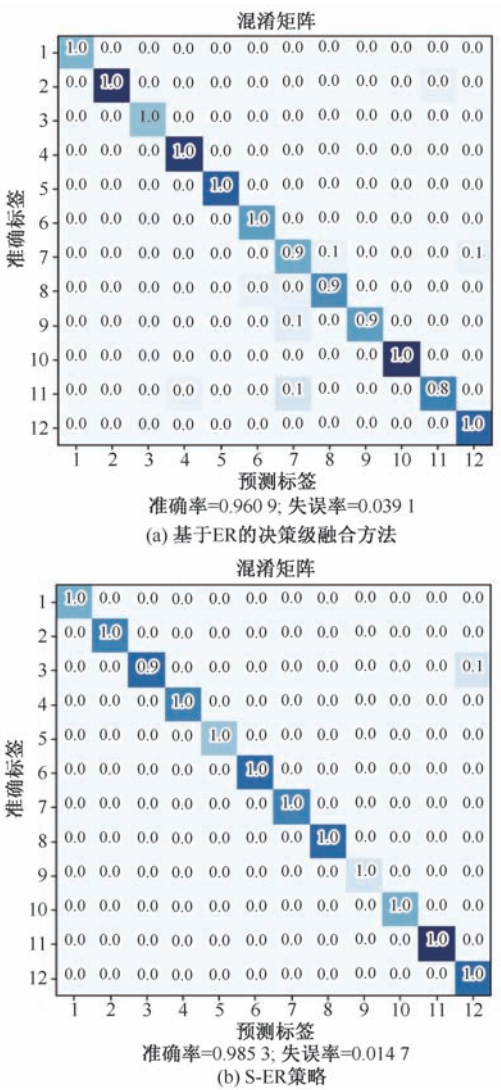

图 9 ZUT-MHAD 中 SG 基准下试验结果

Fig. 9 Experimental results on SG benchmark in ZUT-MHAD

3　结　论

1) S-ER 信息融合方法在有干扰的工业场景中,人员行为识别准确率最高能达到 98.53%,相比传统多模态融合方法与单模态识别方法的识别准确率,最高能提升 17.52%。

2) 下一步将开展算法实际部署中资源占用与实时性能的测试与分析,在带有视觉检测功能的可移动机器人平台上,优化算法的实际部署效果。

参 考 文 献

[1] 纪执安,周云奕,张玉媛,等. 基于改进 YOLOv5 的工业现场不安全行为检测[J]. 中国安全科学学报,2024,34(7):38-43.
JI Zhian, ZHOU Yunyi, ZHANG Yuyuan, et al. Industrial site unsafe behavior detection based on improved YOLOv5[J]. China Safety Science Journal, 2024, 34(7): 38-43.

[2] ZHAO Jiajun, LIU Zhiqiang, XIE Sijia, et al. Research on the application of body posture action feature extraction and recognition comparison[J]. IET Image Processing, 2023, 17(1): 104-117.

[3] 刘耀,焦双健. ST-GCN 在建筑工人不安全动作识别中的应用[J]. 中国安全科学学报,2022,32(4):30-35.
LIU Yao, JIAO Shuangjian. Application of ST-GCN in unsafe action identification of construction workers[J]. China Safety Science Journal, 2022, 32(4): 30-35.

[4] QIU Sen, ZHAO Hongkai, JIANG Nan, et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges[J]. Information Fusion, 2022, 80: 241-265.

[5] LIU Liujun, YANG Jiewen, LIN Ye, et al. 3D human pose estimation with single image and inertial measurement unit (IMU) sequence[J]. Pattern Recognition, 2024: DOI: 10.1016/j.patcog.2023.110175.

[6] 孙晴,杨超宇. 基于多模态的井下登高作业专人扶梯检测方法[J]. 工矿自动化,2024,50(5):142-150.
SUN Qing, YANG Chaoyu. A multi-modal detection method for holding ladders in underground climbing operations[J]. Journal of Mine Automation, 2024, 50(5): 142-150.

[7] DAI Zhuangzhuang, PARK J, KASZOWSKA A, et al. Detecting worker attention lapses in human-robot interaction: an eye tracking and multimodal sensing study[C]. International Conference on Automation and Computing (ICAC), 2023: 1-6.

[8] 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.

[9] 年立辉. 基于多模态融合构建的建设工程工人不安全行为识别模型[J]. 佳木斯大学学报:自然科学版,2024,42(9):118-120, 148.
NIAN Lihui. A model for identifying unsafe behaviors of construction workers based on multimodal fusion[J]. Journal of Jiamusi University, 2024, 42(9): 118-121, 148.

[10] 陈佳菁. 基于 WiFi 和视频的多模态人体动作识别研究与实现[D]. 北京: 北京邮电大学,2024.
CHEN Jiajing. Research and implementation of multimodal human activity recognition based on WiFi and vision[D]. Beijing: Beijing University of Posts and Telecommunications, 2024.

[11] 何俊,张彩庆,李小珍,等. 面向深度学习的多模态融合技术研究综述[J]. 计算机工程,2020,46(5):1-11.
HE Jun, ZHANG Caiqing, LI Xiaozhen, et al. Survey of research on multimodal fusion technology for deep learning[J]. Computer Engineering, 2020,46(5):1-11.

[12] SHAFER G. Dempster-shafer theory[J]. Encyclopedia of Artificial Intelligence,1992,1:330-331.

[13] ZHANG Guoliang, JIA Songmin, LI Xiuzhi, et al. Weighted score-level feature fusion based on Dempster-Shafer evidence theory for action recognition[J]. Journal of Electronic Imaging, 2018, 27(1): DOI: 10.1117/1.jei.27.1.013021.

[14] YANG Jianbo, XU Dongling. Evidential reasoning rule for evidence combination[J]. Artificial Intelligence, 2013, 205: 1-29.

[15] XU Xiaobin, HUANG Weidong, ZHANG Xuelin, et al. An evidential reasoning-based information fusion method for fault diagnosis of ship rudder[J]. Ocean Engineering, 2025, 318: DOI: 10.1016/j.oceaneng.2024.120082.

[16] FAN Xuecheng, XU Zeshui. Double-level multi-attribute group decision-making method based on intuitionistic fuzzy theory and evidence reasoning[J]. Cognitive Computation, 2023, 15: 838-855.

[17] NABOT A. A correction: utilizing evidential reasoning (ER) approach for software components selection[J]. SN Computer Science, 2025, 295(6): DOI: 10.1007/s42979-025-03878-6.

[18] XIE Jiming, ZHANG Yan, LI Ke, et al. Inspiration in human reasoning logic: automating the inference and analysis of traffic accident information via macro-micro integration[J]. Traffic Injury Prevention, 2025: DOI: 10.1080/15389588.2025.2486572.

[19] 王树畅. 基于证据推理规则的锂离子电池健康状态评估方法研究[D]. 哈尔滨: 哈尔滨师范大学,2024.
WANG Shuchang. Research on health state assessment method of lithium-ion batteries based on the evidence reasoning rule[D]. Harbin: Harbin Normal University, 2024.

[20] 赖尉文,贺维. 一种基于自适应证据推理规则的集成学习方法[J]. 计算机应用研究,2023,40(8):2 281-2 285, 2 297.
LAI Weiwen, HE Wei. Ensemble learning method based on adaptive evidential reasoning rule[J]. Application Research of Computers, 2023, 40(8): 2 281-2 285, 2 297.

[21] 赵蕊蕊,孙建彬,游雅倩,等. 动态 ER Rule 分类器构建与应用[J]. 系统工程理论与实践,2022,42(8): 2 258-2 276.
ZHAO Ruirui, SUN Jianbin, YOU Yaqian, et al, Construction and application of dynamic classifier based on evidential reasoning rule[J]. Systems Engineering-Theory & Practice, 2022, 42(8): 2 258-2 276.

[22] YANG Xinshe. Firefly algorithm, stochastic test functions and design optimisation[J]. International Journal of Bio-inspired Computation, 2010, 2(2): 78-84.

[23] JI Shuiwang, XU Wei, YANG Ming, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(1): 221-231.

[24] 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.

[25] 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.

[26] WANG Huan, LIU Zhiliang, PENG Dandan, et al. Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2019, 16(9): 5 735-5 745.

[27] CHEN Chen, JAFARI R, KEHTARNAVAZ N. UTD-MHAD: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor[C]. 2015 IEEE International Conference on Image Processing (ICIP), 2015: 168-172.

[28] LIN Fang, WANG Zhelong, ZHAO Hongyu, et al. Adaptive multi-modal fusion framework for activity monitoring of people with mobility disability[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 4 314-4 324.

[29] GAN Lipeng, CAO Runze, LI Ning, et al. Focal channel knowledge distillation for multi-modality action recognition[J]. IEEE Access, 2023,11: 78 285-78 298.

[30] IMRAN J, RAMAN B. Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(1): 189-208.

[31] ZHONG Zhuokun, HOU Zhenjie, LIANG Jiuzhen, et al. Multimodal cooperative self-attention network for action recognition[J]. IET Image Processing, 2023,17(6): 1 775-1 783.

作者简介： 王海泉 (1981—),男,河南郑州人,博士,教授,硕士生导师,主要从事多模态数据挖掘技术、故障诊断等方面的研究。E-mail: wanghq@zut.edu.cn。