



中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 基于可见光-热红外视觉监控的车间工人跌倒检测算法

王鑫, 刘晓楠, 高焕兵, 曾子铭, 张吟龙

引用本文:

王鑫, 刘晓楠, 高焕兵, 曾子铭, 张吟龙. 基于可见光-热红外视觉监控的车间工人跌倒检测算法[J]. 控制与决策, 2024, 39(4): 1142-1150.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1400>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

#### [基于多层次特征的机械臂单阶段抓取位姿检测](#)

Single-stage grasp pose detection of manipulator based on multi-level features

控制与决策. 2021, 36(8): 1815-1824 <https://doi.org/10.13195/j.kzyjc.2019.1840>

#### [复杂背景下全景视频运动小目标检测算法](#)

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249-256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

#### [尺度自适应的多特征融合相关滤波目标跟踪算法](#)

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

控制与决策. 2021, 36(2): 429-435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

#### [高超声速飞行器间歇故障改进自适应容错控制](#)

Improved adaptive fault-tolerant control of intermittent faults in hypersonic flight vehicle

控制与决策. 2021, 36(11): 2627-2636 <https://doi.org/10.13195/j.kzyjc.2020.0483>

#### [基于姿态估计的实时跌倒检测算法](#)

Real-time fall detection algorithm based on pose estimation

控制与决策. 2020, 35(11): 2761-2766 <https://doi.org/10.13195/j.kzyjc.2019.0382>

# 基于可见光-热红外视觉监控的车间工人跌倒检测算法

王鑫<sup>1</sup>, 刘晓楠<sup>1</sup>, 高焕兵<sup>2,3†</sup>, 曾子铭<sup>4</sup>, 张吟龙<sup>5</sup>

(1. 沈阳建筑大学 电气与控制工程学院, 沈阳 110168; 2. 山东建筑大学 电气工程学院, 济南 250101;  
3. 山东建筑大学 山东省智能建筑技术重点实验室, 济南 250101; 4. 深圳职业技术学院 汽车与交通学院,  
广东 深圳 518055; 5. 中国科学院 沈阳自动化研究所, 沈阳 110169)

**摘要:** 为了解决工厂车间视觉监控存在噪声干扰、光线变化、目标遮挡等问题, 提出一种基于多模态视觉监控的工人跌倒检测算法。首先, 采用热像仪和可见光相机获取车间内全天候监控图像, 结合自适应滤波模型对图像进行降噪处理, 以抑制环境噪声对监控图像的干扰; 然后, 构建一种改进的人体姿态特征提取网络, 通过融合串联时间帧合并模块和位姿残差模块, 以简化目标检测的特征图尺度, 实现监控图像中工人区域被部分遮挡时姿态的实时、可靠预测; 最后, 设计人体轴线倾角、人体外接矩形框长宽比以及双膝盖点移动速度作为工人跌倒判别性特征, 进而实现车间内工人的跌倒判别。在自建数据集和公开数据集上对所提出方法进行验证, 实验结果表明, 所提出算法的跌倒检测精度分别为95.6%和96.3%, 与对比算法相比具有更好的准确性和实时性。

**关键词:** 多模态; 跌倒检测; 轻量化卷积; 时间帧合并; 位姿残差融合

中图分类号: TP391.41

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1400

引用格式: 王鑫, 刘晓楠, 高焕兵, 等. 基于可见光-热红外视觉监控的车间工人跌倒检测算法[J]. 控制与决策, 2024, 39(4): 1142-1150.

## Fall detection for industrial workers using RGB and thermal infrared measurements

WANG Xin<sup>1</sup>, LIU Xiao-nan<sup>1</sup>, GAO Huan-bing<sup>2,3†</sup>, ZENG Zi-ming<sup>4</sup>, ZHANG Yin-long<sup>5</sup>

(1. College of Information and Control Engineering, Shenyang Jianzhu University, Shenyang 110168, China; 2. School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China; 3. Key Laboratory of Intelligent Buildings Technology, Shandong Jianzhu University, Jinan 250101, China; 4. School of Automotive and Transportation Engineering, Shenzhen Polytechnic, Shenzhen 518055, China; 5. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China)

**Abstract:** Because of the problems of noise interference, illumination changes, and target occlusion of the visual monitoring in the factory, the existing fall detection algorithms have weak anti-noise ability, high lighting requirements, and poor target detection effect. It is intractable for the state-of-the-art visual surveillance methods to be reliably applied to all-weather factory workshop worker safety monitoring scenarios. Therefore, this paper presents a novel human fall detection method based on multimodal visual monitoring. Firstly, a thermal imager and a visible light camera are used to obtain the monitoring images in the workshop, and an adaptive median filter model is proposed to denoise the images and to suppress the interference of environmental noise on the monitoring images. Secondly, an improved lightweight convolutional neural network is used to extract the worker skeleton and joint sequence, and a time frame merging module and a pose residual fusion module are designed, so that the network can ensure that the worker pose occluded can be detected through the temporal correlation of adjacent frames. Finally, the inclination of the human body axis, the aspect ratio of the human body circumscribed rectangular frame and the moving speed of the double knee points are designed as the discriminative features of the worker fall. The proposed method has been verified on the collected dataset and public dataset. The experimental results show that the fall detection accuracy of the proposed method is 95.6% and 96.3%, respectively. Compared with the traditional methods, it has better accuracy and real-time performance and can be applied to worker fall detection in the factory workshop.

**Keywords:** multimodality; fall detection; lightweight convolution; time-frame merging; pose residual fusion

收稿日期: 2022-08-03; 录用日期: 2022-12-30.

基金项目: 国家自然科学基金项目(62273332); 中国科学院青年创新促进会会员项目(2022201); 山东省智能建筑技术重点实验室开放课题项目(SDIBT202003); 沈阳市科技计划项目(22-322-3-36).

责任编辑: 牛玉刚.

†通讯作者. E-mail: gaohuanbing@126.com.

## 0 引言

近年来,随着工业4.0、中国制造2025的提出,制造业逐渐向智能化、规范化方向发展<sup>[1]</sup>。目前,国内外自动化设备因缺少实时的安全感知能力,致使工人在生产过程中时常面临诸多安全隐患。如银川工人手臂卷入机械传动器事件和华中企业员工被压事件<sup>[2]</sup>。上述事件皆因工人误操作被自动化设备卷入或压倒,通常在事故发生的一瞬间工人最直观的表现是跌倒,但是,由于设备自身无法感知危险,而继续运转导致悲剧发生。2021年,中国安全生产协会研究结果显示<sup>[3]</sup>,工人因操作不当而跌倒的事故占工厂安全事故总数的93%。为了进一步提高工厂车间的现代化、规范化,保证工人的生命安全,对工厂内人员的跌倒状态进行实时、准确地判断尤为重要。

国内外学者对复杂场景下的跌倒检测算法进行了初步的探索研究,其研究方法可分为接触式跌倒检测<sup>[4-5]</sup>和非接触式跌倒检测<sup>[6-9]</sup>两类。基于接触式跌倒行为研究,Weng等<sup>[10]</sup>分析了倾度传感器数据中角度变化,并使用Mann-Kendall检验序列突变算法监测佩戴者是否跌倒。Wilk等<sup>[11]</sup>采用非线性加速度时间序列集结合空间轨迹和卷积神经网络(convolution neural network, CNN)的方法,实现了CNN对加速度计数据进行特征提取和分类。Hussain等<sup>[12]</sup>使用了模式识别算法来区别环境变化,进而判断动作类别。虽然基于接触式的跌倒检测算法可利用传感器提取更多的行为特征,但是,所构建的模型会因噪声影响而导致输入信号冗余<sup>[13]</sup>,从而降低识别效率。

非接触式跌倒检测算法主要是基于摄像头获取实时图像信息,并通过提取相应特征来分析人体姿态,完成跌倒检测任务。该类算法具有成本低、检测范围广等优点,广泛应用于人体识别检测的研究。Chen等<sup>[14]</sup>提出了结合RGB图像与光流图像的双流卷积神经网络算法,利用视频相邻帧间的信息差作为网络的输入,实现人体跌倒判别功能。Hoang等<sup>[15]</sup>结合2D-CNN与长短期记忆循环神经网络(long short-term memory, LSTM),以解决跌倒判别分类器训练受限的问题,进而提高了跌倒检测精度。Lu等<sup>[16]</sup>为避免人体特征局限于二维序列,采用了3D-CNN从时间域和空间域提取更充分的行为特征进行分类。基于非接触式的跌倒检测算法准确率一般高于基于接触式的跌倒检测算法,但是,该类算法过度依赖卷积网络的分类效果。而对于2个相似的动作输入,分类器将很难做出准确判断。

相比医疗、家居等结构单一场景,工厂车间环境相对复杂、不确定因素多<sup>[17]</sup>,其主要难点问题如下。

1) 车间内光线明暗差异大。相机在昏暗或夜间场景中成像质量差,无法保证工厂内全天候的跌倒判别。

2) 车间内噪声源复杂。传统检测算法易受到工厂中大型设备、电磁、粉尘等因素干扰<sup>[18]</sup>,造成输入视频散焦或工人位姿部分遮挡等情况,使得特征提取难度加大。

3) 工人动作集包含伪跌倒动作。工人日常动作中存在与跌倒动作相似的情况,传统算法提取的跌倒特征无法对伪跌倒动作做出准确识别<sup>[19]</sup>。

针对上述问题,本文提出一种面向工厂车间的跌倒检测算法,主要内容如下。

1) 融合热成像模态-RGB模态图像实现全天候目标信息采集并降噪处理。为了解决昏暗或无光场景下,可见光相机成像质量差的问题,本文提出一种热成像-RGB多模态融合的方法来获取图像:在光照充足时,仅选用可见光相机采集RGB图像;在环境昏暗时,根据所处环境的光线强度采用不同的加权系数融合热成像模态图像和RGB模态图像;在夜间无光照时,利用热像仪采集人物图像,以弥补无光场景下可见光相机的不足。并通过自适应中值滤波算法降噪处理,得到清晰的人物图像,从而解决工厂车间中昏暗或无光场景的图像采集问题。

2) 改进的人体姿态特征提取网络。将传统VGG网络中的 $7 \times 7$ 大卷积核改进为3个 $3 \times 3$ 的小型卷积核,使得网络在具有相同感受野的同时,减少网络的参数,从而保证实时性。此外,在算法模型中设计相邻时间帧合并模块,以利用视频帧间的时间线索有效聚合缺失帧的关键点,并设计位姿残差融合模块,用于实现跨帧的人体位姿估计和校正。从而保证所提出算法在视频散焦、人物位姿遮挡等情况下检测的准确性、实时性。

3) 基于时域-空间域约束的多判别特征模型。通过分析工人跌倒和伪跌倒动作在时域和空间域的不同特点,提出最小外接矩形宽高比、人体轴线倾角、双膝点下降速度作为跌倒动作的判别性特征,实现工人跌倒检测。从而明确地划分跌倒过程与伪跌倒过程在时间域-空间域上的本质差异。

## 1 算法框架

工人跌倒检测算法框架如图1所示。将热像仪和可见光相机获取的多模态图像经自适应滤波处理,可

有效地滤除图像噪声,进而得到人物轮廓清晰的图像作为模型输入.将原网络中 $7 \times 7$ 的大卷积核替换为3个 $3 \times 3$ 的卷积核,增加网络深度的同时又减少了参数的数量,使得经过改进的VGG网络可实时获取更多准确的特征.将提取到的特征作为OpenPose模型<sup>[20]</sup>的输入,经双分支多级迭代网络处理可为每帧图像中人物生成关键点和骨架图.但是,人物肢体存在部分或全部遮挡等情况,导致检测结果存在误差,故加入

相邻时间帧合并模块,当断帧、跨帧等情况出现时,通过对邻近时间区间3个有效连续帧进行关键点聚合,从而准确定位关键点位置的预测范围,再通过位姿残差融合模块中残差块有效获取当前帧与相邻帧间的位姿残差,进而得到连续、准确的人体姿态预测结果.最终,在时空域分别对所获取的姿态信息进行判别,可准确检测工厂车间工人是否处于跌倒状态.

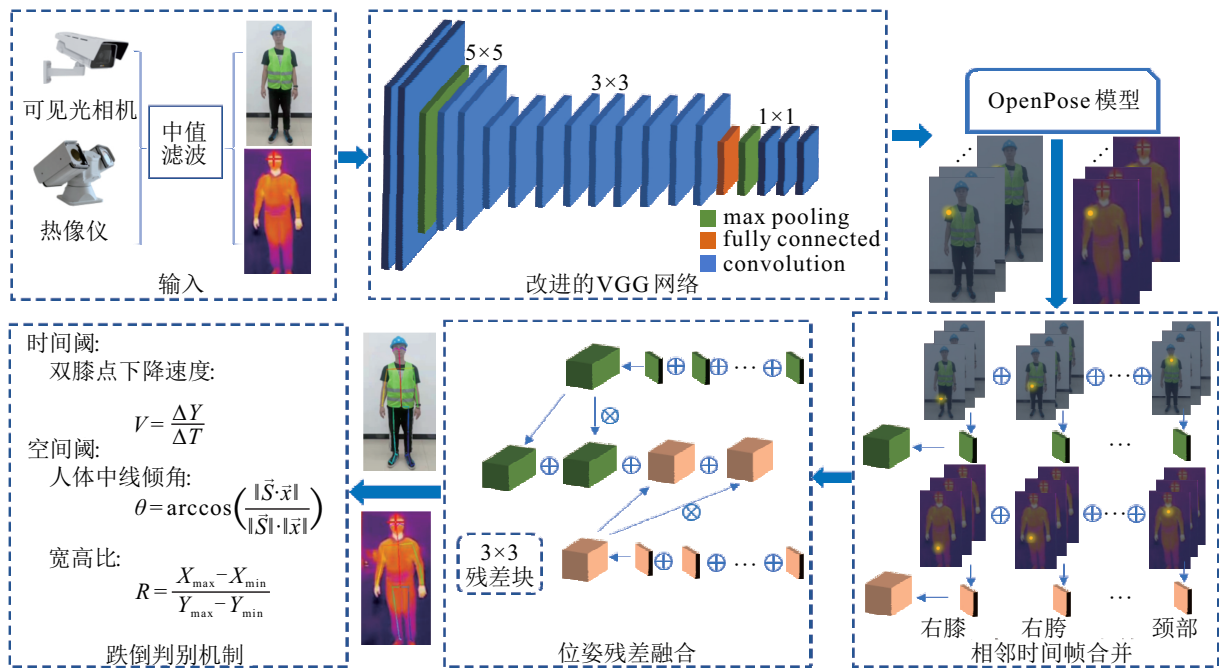


图1 工厂车间在线跌倒检测算法框图

## 2 图像预处理

### 2.1 图像加权融合

面向工厂车间的跌倒检测技术,首先要解决昏暗场景下图像的采集问题<sup>[21]</sup>.本文选用可见光相机和热像仪作为多模态数据采集设备,可见光相机采集的RGB模态图像分辨率高,但是易受云雾、光线等环境因素影响.而通过热像仪采集的热成像模态图像受环境影响较小,但是图像对比度低,分辨细节能力较差.因此,本文提出一种基于环境光强变化的热成像-RGB多模态图像融合方法,用于协同利用多模态图像优势,弥补单模态图像质量易受环境变化影响的不足.

将实时采集的热成像模态图像和RGB模态图像转化为灰度图并用 $X_1$ 、 $X_2$ 表示,大小为 $W \times H$ . $X$ 为 $X_1$ 、 $X_2$ 加权融合后的图像. $X_1$ 、 $X_2$ 两个原始图像加权融合过程如下:

$$X(w, h) = k_1 X_1(w, h) + k_2 X_2(w, h). \quad (1)$$

其中: $w$ 为图像像素的行号, $w = 1, 2, \dots, W$ ;  $h$ 为图像

像素的列号, $h = 1, 2, \dots, H$ ;  $k_1$ 、 $k_2$ 为该时刻图像融合的加权系数,有

$$\begin{cases} k_1 = 1, k_2 = 0, L_t \geq L_h; \\ k_1 = \frac{L_t - L_d}{L_h - L_d}, k_2 = 1 - k_1, L_h > L_t \geq L_d; \\ k_1 = 0, k_2 = 1, L_t < L_d. \end{cases} \quad (2)$$

其中: $L_t$ 为车间环境的实时光照强度大小, $L_h$ 和 $L_l$ 分别为图像融合过程中光照强度的阈值上限和阈值下限.

通过比较图像采集时刻光强与所设阈值的大小,从而确定图像的加权系数,输出该时刻的融合图像.

针对工厂噪声导致获取图像过程中人体关节位置出现跳动等不稳定现象<sup>[22]</sup>,本文提出一种自适应中值滤波算法用于抑制图像中非线性椒盐噪声.在传统中值滤波算法的基础上,改变滤波窗口大小,使得噪声密度低于20%,以增强去噪效果且能够保护图像的细节信息.

由于工厂车间中图像的采集质量易受灰尘、烟

雾、电流、磁场等环境因素干扰,生成非线性椒盐噪声. 根据该类噪声的极值特性,若像素点 $(w, h)$ 是孤立噪声点,则其与周围像素点的灰度差别很大. 利用这一特点,采用疑似噪声点与其相邻像素点的最大绝对灰度差作为变量,并根据变量的大小判断像素点 $(w, h)$ 是否为噪点,进而求得窗口的噪声密度. 当噪声密度最小时,选用 $3 \times 3$ 滤波窗口,噪声密度最大时,选用 $6 \times 6$ 滤波窗口. 滤波窗口在图像上遍历,噪声密度 $T_i$ 定义如下:

$$T_i = \frac{k_i}{n \times n} \times 100\%. \quad (3)$$

其中: $k_i$ 为当前窗口内噪声点个数, $n \times n$ 为滤波窗口尺寸( $n$ 为整数,3、4、5、6; $n$ 的初始值设为3). 整体流程如下.

step 1: 为保护图像细节和去噪能力,设 $3 \leq n \leq 6$  ( $n$ 越小,去噪能力越弱; $n$ 越大,图像细节越少).

step 2: 当 $T_i < 20\%$ 时,滤波窗口保持不变,在尺寸为 $n \times n$ 的窗口下进行排序取中值操作.

step 3: 当 $T_i \geq 20\%$ 时,增大滤波窗口尺寸,令 $n = n + 1$ ,再进行排序取中值操作.

step 4: 当 $n > 6$ 时,仍然有 $T_i \geq 20\%$ ,保持中心像素原值输出.

## 2.2 改进的人体姿态特征提取网络

OpenPose是一种用于多人实时的人体姿态估计模型<sup>[23]</sup>,拥有较强的准确率和普适性. 传统检测模型中采用VGG作为提取底层特征的主干网络,存在自身模型参数量大、实时性差等缺陷. 鉴于此,本文提出一种改进的人体姿态特征提取网络,利用3个 $3 \times 3$ 的卷积核代替原网络结构中的 $7 \times 7$ 卷积核,以降低OpenPose模型的复杂程度和计算量. 经过裁剪后的轻量化模型可准确识别到人体骨架序列中的19个关节点,每帧图像中人体关节点生成置信图公式如下式所示:

$$C_{n,m}^*(Q) = \exp\left(-\frac{\|Q - X_{n,m}\|_2^2}{\sigma_{n,m}^2}\right). \quad (4)$$

其中: $m$ 为物序号, $n$ 为人体关键点, $Q$ 为置信度图 $C_{n,m}$ 中任意点, $X_{n,m}$ 为人体关键点的真实位置, $\sigma_{n,m}$ 用于描述对应关键点的概率分布. 通过部分亲和场来预测2个关键点间的关联程度,可由积分值 $H$ 表示为

$$H = \int_{t=0}^{t=1} k_b(Q(t)) \cdot \frac{n_2 - n_1}{\|n_2 - n_1\|_2} dt, \quad (5)$$

$$Q(t) = (1-t)n_2 + tn_1, \quad t \in (0, 1). \quad (6)$$

其中: $n_1$ 、 $n_2$ 为相邻关键点, $\|n_2 - n_1\|_2$ 为肢体的长度, $Q(t) \in [n_1, n_2]$ . 若点 $Q$ 在肢体上,则 $k_b(Q(t)) = \mathbf{v}$ ,

式中 $\mathbf{v} = (m_2 - m_1) / \|m_2 - m_1\|_2$ 为单位向量;否则, $k_b(Q(t)) = 0$ .

每个关键点与其相邻关键点间的积分值均会被计算,积分值越大,这对相邻关键点越接近真实骨架连接. 因此,可通过选取 $H$ 的最大值来获得每类肢体的正确连接,再将具有共享相同关键点的肢体连接为人体骨架.

## 2.3 时间帧合并模块与位姿残差融合模块

工厂车间环境复杂,工人在工作过程中身体易被设备、货物遮挡,在作业时仅能够检测到部分身体特征,影响检测结果的准确性,而被遮挡的时间帧 $F(s)$ 与被标记的相邻帧 $F(j)$ 与 $F(l)$ 间具有丰富的时间线索,故处于部分遮挡的人物 $i$ ,其时间帧合并结果 $F_i(j, s, l)$ 为

$$F_i(j, s, l) = \frac{l-s}{l-j} F_i(j) + F_i(s) + \frac{s-j}{l-j} F_i(l), \quad (7)$$

其中 $j, s, l$ 为时间帧索引. 对于每个关节点,可通过规定特定的时间信息来计算并合并为关键点热图,经过重新分组可将其堆叠为特征张量 $\phi_i$ ,表示为

$$\phi_i(j, s, l) = \sum_{n=1}^N \frac{l-s}{l-j} F_i^n(j) \oplus F_i^n(s) \oplus \frac{s-j}{l-l} F_i^n(l). \quad (8)$$

其中: $\oplus$ 为连接帧操作, $N$ 为关节总数量, $n$ 为关节编号. 这样不仅消除了无关关节的干扰,去除算法冗余,并根据时间线索估计了被遮挡时的关节点位置. 随后将得到的不同特征张量输入至残差融合模块,与关键点时间帧合并类似,通过位姿残差融合计算位姿残差 $\psi_i$ ,作为另一种有效的姿态估计线索,实现过程如下:

$$\psi_i(j, s) = F_i(s) - F_i(j), \quad (9)$$

$$\psi_i(s, l) = F_i(l) - F_i(s), \quad (10)$$

$$\psi_i = \frac{l-s}{l-j} \psi_i(j, s) \oplus \frac{s-j}{l-l} \psi_i(s, l). \quad (11)$$

通过连接原始残差 $\psi_i(j, s)$ 、 $\psi_i(s, l)$ 及其对应的权重,并经过 $3 \times 3$ 的残差块处理,可得到最终的位置残差估计. 使得模型可检测出可见光图像和热像仪图像中准确的人体关键点以及骨架位置特征,并将身体部位关键点进行标号,如图2所示. 对应关系如下:0-鼻子,1-颈部,2-右肩,3-右肘,4-右腕,5-左肩,6-左肘,7-左腕,8-重心,9-右胯,10-右膝,11-右脚踝,12-左胯,13-左膝,14-左脚踝,15-右眼,16-左眼,17-右耳,18-左耳. 因此,模型可准确提取不同光照条件下图像中的人体轴线、人体矩形框、左膝盖点、右

膝盖点等信息,为后续生成跌倒判别特征提供基础数据。

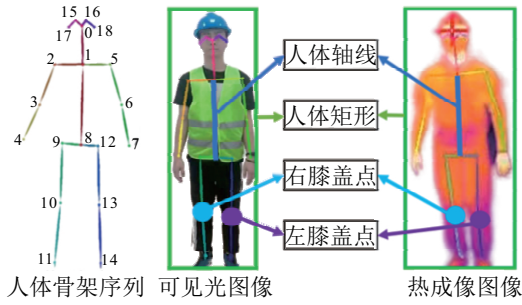


图2 人体骨架和跌倒判别特征示意图

### 3 跌倒判别机制

车间内工人的行为与自然场景下的普通行为不同,工人行为与其工作性质、车间环境等关系更为紧密,常包含一些伪跌倒动作,仅通过深度学习的方法很难区分跌倒与伪跌倒。为此,本文提出一种时域-空域约束条件用于辅助实现伪跌倒动作的准确判别。

#### 3.1 时域约束

伪跌倒动作可理解为主动式跌倒动作<sup>[24]</sup>,其膝盖点的下降速度总保持在一定安全范围内。然而,工人发生意外跌倒时处于被动状态,其膝盖点的下降速度会在短时间内发生突变,进而超出安全速度范围。因此,本文记录了不同动作发生过程中膝盖点的运动速度,并汇总测试数据如图3所示。

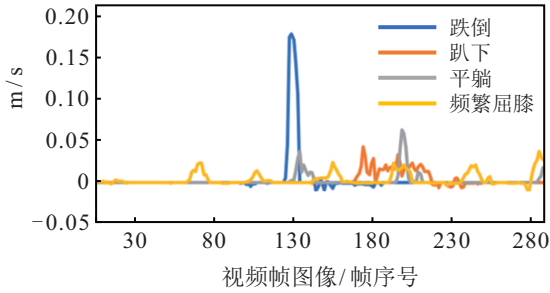


图3 不同动作发生过程中速度变化

由图3可见,在动作的发生过程中,趴下、平躺以及频繁屈膝等动作的膝盖点移动速度均在 $0.078\text{ m/s}$ 以下,而跌倒动作发生时膝盖点的瞬时速度达到了 $0.17\text{ m/s}$ ,因此,本文最终选定 $\alpha = 0.1\text{ m/s}$ 作为判别跌倒和其他伪跌倒动作的阈值。判别过程中,通过计算相邻帧间的膝盖点移动速度 $V$ ,并与阈值比较来实现判别,其中膝盖点的移动速度计算公式为

$$V_{lk} = \frac{Y_{13} - Y'_{13}}{T - T'}, \quad (12)$$

$$V_{rk} = \frac{Y_{10} - Y'_{10}}{T - T'}, \quad (13)$$

$$\Phi = \left( \left( \frac{Y_{10} - Y'_{10}}{T - T'} \right) > \alpha \right) \cup \left( \left( \frac{Y_{13} - Y'_{13}}{T - T'} \right) > \alpha \right). \quad (14)$$

由人体骨架提取可得到关键点坐标,其中人体左膝盖点坐标为 $(X_{13}, Y_{13})$ ,右膝盖点坐标为 $(X_{10}, Y_{10})$ ,故式中: $Y_{10}$ 为当前帧右膝盖点纵坐标; $Y'_{10}$ 为上一帧右膝盖点纵坐标; $Y_{13}$ 为当前帧左膝盖点纵坐标; $Y'_{13}$ 为上一帧左膝盖点纵坐标; $T$ 为当前时刻; $T'$ 为上一帧时刻;当 $\Phi$ 为真时,表示至少有一膝盖点下降速度发生了突变,若连续3帧均发生突变,则判定为跌倒动作,反之,则为非跌倒动作。

#### 3.2 空间域约束

仅通过计算膝盖点下降速度的方法区分伪跌倒动作,易产生误判,因为膝盖点移动速度的突然改变可由人体主动发起,如下蹲等动作。仅依靠时域的约束条件无法准确判断人物状态,故提出空间域约束条件共同作为人体跌倒的判别依据。工人在工厂车间作业时,站立、下蹲等动作的人体外接矩形宽高比和身体轴线倾角与跌倒状态有较大差异,如图4所示。因此,根据身体处于不同状态时满足的物理属性,本文提出人体轴线倾角 $\theta$ 以及人体外接矩形框宽高比 $R$ 作为空间约束条件。

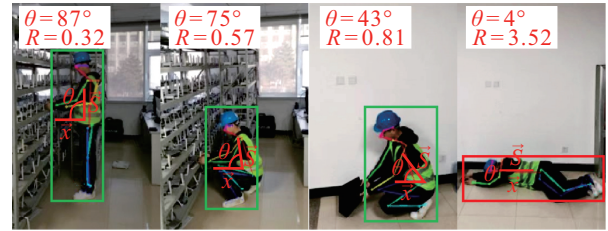


图4 不同空间域约束条件示意图

为了更准确地标注人体外接矩形框,本文选用关键点坐标的最大值 $X_{\max}$ 、 $Y_{\max}$ 和最小值 $X_{\min}$ 、 $Y_{\min}$ 构建人体外接矩形框,得到外接矩形框的宽高比 $R$ 为

$$R = \frac{X_{\max} - X_{\min}}{Y_{\max} - Y_{\min}}. \quad (15)$$

通过选取颈部关键点 $(X_1, Y_1)$ 、中心点 $(X_8, Y_8)$ 生成代表人体轴线位置和方向的向量 $\vec{S}$ ,如下式所示:

$$\vec{S} = (X_8 - X_1, Y_8 - Y_1). \quad (16)$$

人体轴线与水平地面的单位向量 $\vec{x}$ 所构成的夹角可表示为

$$\theta = \arccos \left( \frac{\|\vec{S} \cdot \vec{x}\|}{\|\vec{S}\| \cdot \|\vec{x}\|} \right). \quad (17)$$

当 $R$ 小于阈值 $\beta = 0.8$ ,且人体倾角满足 $\gamma = 63^\circ \leq \theta \leq \delta = 117^\circ$ 时,判定该动作为非跌倒,不满足此条件则判定为跌倒。

## 4 实验与分析

### 4.1 实验环境

实验环境为CPU intel i7-9 700(8核心8线程),内存容量32GB,显卡为Nvidia GTX 3070,使用Cuda、

Cudnn加速,操作系统为Ubuntu 16.04.

为了验证所提出方法的泛化性和准确性,本文分别在公开数据集和自建工人动作数据集进行实验. 公开数据集包括Le2i跌倒数据集和URFD数据集,数据集中包含的常见动作有行走、蹲起、弯腰、坐下、躺下、跌倒等,共1200组视频片段. 其中:训练集1000组,测试集200组.

自建样本数据模拟车间生产全过程主要包括货物搬运、设备维修、产品制造、安防巡逻4类工人参与的日常动作,共计1000组动作序列,每组序列中各动作按一定顺序多次出现且动作的方向、速度、位置以及环境的光线强度均不同. 其中:训练集800组,测试集200组.

#### 4.2 自建样本数据集测试

为了体现所提出算法的优越性能,本文选取2种具有代表性的跌倒检测算法,分别为基于YOLOv4对老年人进行跌倒检测算法(YOLOv4+LSTM)<sup>[25]</sup>和基于RGB-D视频的跌倒检测算法(深度图像特征+SVM)<sup>[26]</sup>与所提出算法进行对比实验.

为了验证所提出算法在检测精度上的优越性,使用自建样本数据集进行训练和验证,所有样本图像参与一次训练过程作为一轮训练周期,其训练损失曲线如图5(a)所示,其中:横坐标为训练周期,纵坐标为损失值. 实验结果显示,3种检测算法的损失值随着训练次数的增加总体呈现下降趋势,基于YOLOv4+LSTM检测算法和基于深度图像特征+SVM检测算法的损失值分别为0.9和0.7,所提出算法损失值呈现的下降趋势更加明显,最终稳定在0.24左右,明显优于其他2种算法. 这是由于在模型训练过程中,神经网络训练效果受网络层数和训练次数影响. 在参数合理的情况下,层数和神经元数越多,正确率相对提高,但是易出现过拟合,损失值下降缓慢,在到达某个程度后不再下降. 而本文采用的改进轻量化卷积神经网络在保持模型性能的前提下降低模型大小,提升模型速度,同时可迅速降低损失值,以达到较好的训练效果.

在准确率方面,实验结果如图5(b)所示,横坐标为训练周期,纵坐标为准确度. 随着训练次数的增加,3种检测算法的准确率不断提高,但是,其他2种算法准确率提高幅度有限,且没有趋于平稳,最终准确率在60%~70%间波动. 所提出算法的准确率随着训练次数的增加不断提高,并稳定在95%左右,明显优于另外2种算法. 这是由于自建样本训练集中夹杂工人的伪跌倒动作,如躺下这类由人体主动发出的

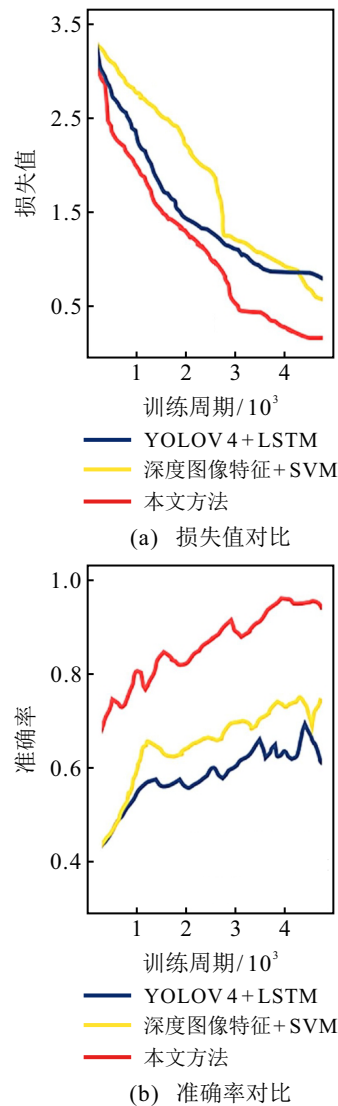


图5 自建样本训练结果

伪跌倒动作,在动作完成后单帧的图像特征与跌倒特征无差异,致使传统YOLOv4+LSTM检测算法频繁出现误判,导致准确率低下. 而所提出检测算法不仅增加了空间域上对工人特殊动作的判别,还增加了时间域上对邻近3个有效连续帧双膝点下降速度的分析,进而判别此时动作是否由人体主动发出,使得检测准确率大幅提升,并增强了模型的泛化能力.

为了验证所提出算法在光线变化、非线性噪声以及人体姿态部分遮挡等干扰因素相叠加的情况下仍然具有较高的检测精度,在自建数据集中随机抽取明亮、昏暗以及无光场景下各900张样本图片,其中夹杂货物遮挡人物的样本图片和噪声密度大于15%的样本图片,共计300组. 同时,为了对比分析所提出算法与其他检测算法的检测性能,选取在明亮、昏暗以及无光场景下各3张样本图片,均为具有噪声、人物遮挡等检测难度较大的图片,检测结果如图6所示. 通过对比分析,所提出方法在明亮场景、昏暗场景以及无光场景下均具有准确的检测结果,原因在于可

自动调节热像仪与可见光相机对场景图片的融合权重比例系数,结合所提出滤波算法自适应调整去噪窗口大小,得到清晰的全天候实时图像. YOLOv4 + LSTM检测算法和深度图像特征 + SVM检测算法虽然可对明亮且清晰的图片进行准确识别,但是,在光线相对昏暗或噪声干扰严重的场景下,准确率明显降低. 特别是在无光或噪声密度大于35%的场景下,将很难提取图像细节特征,无法实现有效判别.

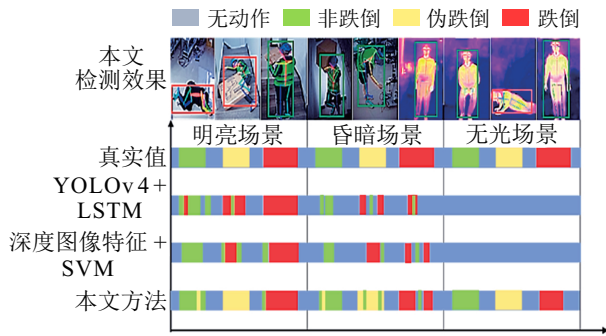


图6 不同光强下跌倒判别效果自建样本训练结果

由于其他算法缺少对昏暗场景下图像信息的提取能力,在检测过程中会存在大量误判现象,这将极大地影响其准确率和召回率,具体结果如表1所示. 由表1可见,所提出算法在自建数据集的准确率达到95.6%,召回率达到了93.7%,远优于传统算法的性能. 同时,所提出算法的误检率和漏检率分别为4.6%和3.7%,较其他检测方法有明显的降低. 原因在于在本文利用视频帧间的时间线索有效聚合缺失帧的关键点,并设计了位姿残差融合模块,有效降低了误检、漏检的可能性. 在处理相同清晰度视

频帧时,所提出算法处理每帧的平均耗时降至52 ms,其他方法处理每帧的平均耗时分别达到了141 ms和287 ms,原因在于改进了原始模型的底层特征提取网络,加入了相邻时间帧合并模块和位姿残差融合模块,有效提高了算法的准确性,并利用多个小型卷积核减小了计算成本,降低了模型复杂度,保证了实时性. 所提出方法在自建数据集的跌倒检测效果如图7所示.

### 4.3 公开数据集测试

为了更好地验证融合卷积神经网络和手工判别性特征对人体姿态识别的有效性,本文在公开数据集上对上述3种算法进行实验对比,检测性能比较结果如表2所示. 由表2可见,所提出算法在召回率、误检率、漏检率、耗时等方面的优越性能与自建样本数据上的测试结果类似,并均优于基于YOLOv4 + LSTM的跌倒检测算法和基于深度图像特征 + SVM的跌倒检测算法. 而所提出算法准确率略低于YOLOv4 + LSTM算法的检测准确率,原因在于所提出算法相对于YOLOv4 + LSTM采用轻量化模型,在牺牲少量精度的前提下,显著地降低了算法的复杂度,平均每帧的耗时为55 ms,所提出方法可实现实时地在线检测要求. 基于深度图像特征 + SVM算法由于需要计算大量数据导致检测耗时过长,实时性较差. 此外,人体跌倒后其背景减除算法易导致目标丢失,降低了检测准确率. 所提出方法在公开数据集上的检测效果如图8所示,可准确地检测出实验者的跌倒状态.

表1 自建数据集检测性能比较

方法	准确率/%	召回率/%	误检率/%	漏检率/%	平均耗时/ms
YOLOv4 + LSTM <sup>[24]</sup>	68.8	32.1	22.4	17.1	141
深度图像特征 + SVM <sup>[25]</sup>	74.4	51.6	15.3	24.9	287
本文方法	95.6	93.7	4.6	3.7	52

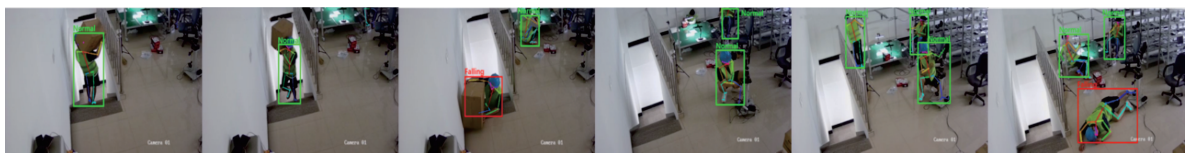


图7 自建数据集跌倒检测效果

表2 公开数据集检测性能比较

方法	准确率/%	召回率/%	误检率/%	漏检率/%	平均耗时/ms
YOLOv4 + LSTM <sup>[24]</sup>	97.1	85.4	5.65	9.8	132
深度图像特征 + SVM <sup>[25]</sup>	89.3	86.2	9.26	11.2	356
本文方法	96.3	93.2	5.43	4.2	55

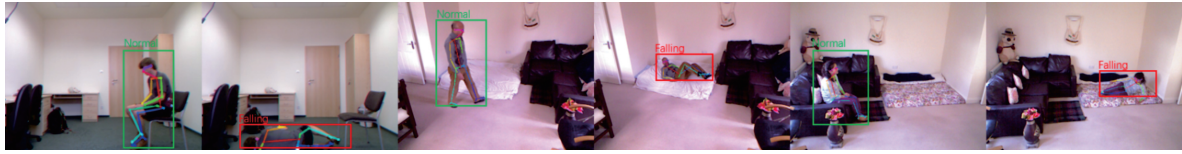


图8 公开数据集跌倒检测效果

## 5 结论

本文提出了一种面向工厂车间的工人跌倒检测算法。首先,该算法设计了热成像-RGB多模态融合的方式获取图像信息,并通过自适应中值滤波器对图像进行预处理,再采用改进的轻量级卷积神经网络模型来提取监控工人的骨架信息,并设计了相邻时间帧合并模块和位姿残差融合模块,保证了人体在被部分遮挡时的检测准确性;然后,在人体跌倒检测算法中融入时域-空域判别特征,使得神经网络能够有侧重地学习人体的动作信息,并设计了合适的检测阈值进而准确地判别跌倒动作;最后,在自建数据集以及公开数据集上进行了实验,准确率分别为95.6%和96.3%,同时,平均检测耗时较其他方法明显缩短,验证了所提出方法的准确性和实时性。今后的工作中将充分考虑在噪声种类多或干扰严重条件下,实现动作特征的分离和提取,在计算资源有限的情况下进一步提高算法效率。

### 参考文献(References)

- [1] 彭开香, 张丽敏. 基于核典型相关性-熵成分分析的工业过程质量监测方法[J]. 控制与决策, 2021, 36(12): 2999-3006.  
(Peng K X, Zhang L M. A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis[J]. Control and Decision, 2021, 36(12): 2999-3006.)
- [2] 张权. 工厂企业范围内的安全生产管理及安全事故预防策略[J]. 中国化工贸易, 2019, 11(5): 26-27.  
(Zhang Q. Safety production management and safety accident prevention strategy within the factory and enterprise[J]. China Chemical Trade, 2019, 11(5): 26-27.)
- [3] Wu P S, Lin C Y, Cheng T Y, et al. Analysis and design of industrial safety automatic identification system based on Tiny-YOLOv3[C]. IEEE International Future Energy Electronics Conference. Taipei, 2022: 1-6.
- [4] Saadeh W, Butt S A, Altaf M A B. A patient-specific single sensor IoT-based wearable fall prediction and detection system[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(5): 995-1003.
- [5] Clemente J, Li F Y, Valero M, et al. Smart seismic sensing for indoor fall detection, location, and notification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(2): 524-532.
- [6] Sadreazami H, Bolic M, Rajan S. Contactless fall detection using time-frequency analysis and convolutional neural networks[J]. IEEE Transactions on Industrial Informatics, 2021, 17(10): 6842-6851.
- [7] Wang H, Zhang D Q, Wang Y S, et al. RT-fall: A real-time and contactless fall detection system with commodity WiFi devices[J]. IEEE Transactions on Mobile Computing, 2017, 16(2): 511-526.
- [8] Vishnu C, Datla R, Roy D, et al. Human fall detection in surveillance videos using fall motion vector modeling[J]. IEEE Sensors Journal, 2021, 21(15): 17162-17170.
- [9] 张晓平, 纪佳慧, 王力, 等. 基于视频的人体异常行为识别与检测方法综述[J]. 控制与决策, 2022, 37(1): 14-27.  
(Zhang X P, Ji J H, Wang L, et al. Overview of video based human abnormal behavior recognition and detection methods[J]. Control and Decision, 2022, 37(1): 14-27.)
- [10] Weng W X, Lo S C. Fall detection based on tilt angle and acceleration variations[C]. IEEE Trustcom/BigDataSE/ISPA. Tianjin, 2017: 1712-1717.
- [11] Wilk B, Augustyn M, Wilk G. Algorithm for human fall detection based on acceleration measurement[C]. Signal Processing: Algorithms, Architectures, Arrangements, and Applications. Poznan, 2020: 13-17.
- [12] Hussain F, Hussain F, Ehatisham-ul-Haq M, et al. Activity-aware fall detection and recognition based on wearable sensors[J]. IEEE Sensors Journal, 2019, 19(12): 4528-4536.
- [13] Singh V, Dev R, Dhar N K, et al. Adaptive type-2 fuzzy approach for filtering salt and pepper noise in grayscale images[J]. IEEE Transactions on Fuzzy Systems, 2018, 26(5): 3170-3176.
- [14] Chen E Q, Bai X, Gao L, et al. A spatiotemporal heterogeneous two-stream network for action recognition[J]. IEEE Access, 2019, 7: 57267-57275.
- [15] Hoang V D, Hoang D H, Hieu C L. Action recognition based on sequential 2D-CNN for surveillance systems[C]. IECON the 44th Annual Conference of the IEEE Industrial Electronics Society. Washington, 2018: 3225-3230.

- [16] Lu N, Wu Y D, Feng L, et al. Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 23(1): 314-323.
- [17] Giorgini M, Barbieri F, Aleotti J. Ground segmentation from large-scale terrestrial laser scanner data of industrial environments[J]. IEEE Robotics and Automation Letters, 2017, 2(4): 1948-1955.
- [18] Villani V, Sabattini L, Barańska P, et al. The inclusive system: A general framework for adaptive industrial automation[J]. IEEE Transactions on Automation Science and Engineering, 2021, 18(4): 1969-1982.
- [19] Vishnu C, Datla R, Roy D, et al. Human fall detection in surveillance videos using fall motion vector modeling[J]. IEEE Sensors Journal, 2021, 21(15): 17162-17170.
- [20] Qiao S, Wang Y L, Li J. Real-time human gesture grading based on OpenPose[C]. The 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Shanghai, 2018: 1-6.
- [21] Feng Q, Gao C Q, Wang L, et al. Spatio-temporal fall event detection in complex scenes using attention guided LSTM[J]. Pattern Recognition Letters, 2020, 130: 242-249.
- [22] 石跃祥, 许湘麒. 基于改进 DenseNet 网络的人体姿态估计[J]. 控制与决策, 2021, 36(5): 1206-1212.  
(Shi Y X, Xu X Q. Improved DenseNet network for human pose estimation[J]. Control and Decision, 2021, 36(5): 1206-1212.)
- [23] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1302-1310.
- [24] Chaccour K, Darazi R, El Hassani A H, et al. From fall detection to fall prevention: A generic classification of fall-related systems[J]. IEEE Sensors Journal, 2017, 17(3): 812-822.
- [25] Chutimawattanukul P, Samanpiboon P. Fall detection for the elderly using YOLOv4 and LSTM[C]. The 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Prachuap Khiri Khan, 2022: 1-5.
- [26] Panahi L, Ghods V. Human fall detection using machine vision techniques on RGB-D images[J]. Biomedical Signal Processing and Control, 2018, 44: 146-153.

### 作者简介

王鑫(1979—), 男, 副教授, 博士, 从事无线通信、视觉惯性导航和计算机视觉等研究, E-mail: wangx7988@sjzu.edu.cn;

刘晓楠(1997—), 女, 硕士生, 从事计算机视觉的研究, E-mail: 1217577407@qq.com;

高焕兵(1979—), 男, 副教授, 博士, 从事智能控制与机器人系统等研究, E-mail: gaohuanbing@126.com;

曾子铭(1983—), 男, 讲师, 博士, 从事计算机视觉、人工智能、无人驾驶等研究, E-mail: zzm1983@szpt.edu.cn;

张吟龙(1988—), 男, 副研究员, 博士, 从事工业视觉、多源信息融合等研究, E-mail: zhangyinlong@sia.cn.