

基于姿态估计的护具佩戴检测与动作识别

吴胜昔, 咸博龙, 冒鑫鑫, 顾幸生

能源化工过程智能制造教育部重点实验室, 上海 200237

基金项目: 国家自然科学基金资助项目(61673175)

通信作者: 顾幸生, xsgu@ecust.edu.cn 收稿/录用/修回: 2020-12-14/2021-03-03/2021-09-28

摘要

护具佩戴检测和动作识别是智慧安监系统中的一个重要环节. 针对传统并行识别方法准确率低且实时性不高的问题, 提出了一种基于姿态估计的并行识别方案, 利用一种基于距离和匈牙利算法的数据交互方法, 将人员躯干和护具进行匹配, 快速并行实现了护具佩戴检测与动作识别. 并行识别方案利用深度可分离卷积, 对 Openpose 模型的参数进行压缩, 使 Openpose 模型轻量化, 提高了该模型姿态估计的实时性, 并提出了一种结合姿态信息的四边形检测法, 解决了未佩戴护具(手持安全帽)的误判问题. 在实时检测实验中, 并行识别方案的手持安全帽的误判率下降 29.3%, 系统的整体准确率达到 93.5%; 改进后的 Openpose 模型的速度比原模型每秒提升 12 帧, 为原模型的 2.2 倍. 实验结果表明, 所提的并行识别方案准确度高, 且实时性强, 满足实际护具佩戴检测和动作识别的需求.

关键词

姿态估计
四边形检测法
动作识别
护具佩戴检测
数据交互

中图法分类号: TP273

文献标识码: A

Protective Wearing Detection and Action Recognition Based on Pose Estimation

WU Shengxi, XIAN Bolong, MAO Xinxin, GU Xingsheng

Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, Shanghai 200237, China

Abstract

Protective wearing detection and action recognition are an important part in the intelligent safety supervision system. Aiming at the problem of low accuracy and low real-time performance of traditional parallel recognition methods, We propose a parallel recognition scheme based on pose estimation to realize protective wearing detection and action recognition in parallel quickly, which uses a data interaction method based on distance and Hungarian algorithm to combine the pose with the guardian. Using deep separable convolution to reduce parameters, the parallel recognition scheme improve the Openpose model, which makes the Openpose model lighter and improves the real-time performance of pose estimation. A quadrilateral detection method combined with pose information is proposed to solve the misjudgment problem of not wearing a guardian (holding helmet). In real-time detection experimental, the misjudgment rate of the holding helmet drops by 29.3%, and the overall accuracy rate of the system reaches 93.5%. The speed of improved Openpose model increases 12 frames per second faster than the original model, which is 2.2 times that of the original model. Experimental verification shows that the parallel recognition scheme has high accuracy and strong real-time performance, which meets the needs of actual protective wearing detection and action recognition.

Keywords

pose estimation;
quadrilateral detection;
action recognition;
protective wearing detection;
data interaction

0 引言

随着人工智能、机器视觉技术的提升, 在生产环境的安全防护方面, 研究人员提出了智慧安监的概念. 护具佩戴检测和动作识别的并行实现是实现智慧安监的一个重要

步骤. 文[1]利用数据交互算法, 将姿态估计中的姿态信息引入到护具佩戴检测算法中, 实现两种数据的交互, 从而提出了一种三点定位法, 确定护具检测区域, 有效识别出护具佩戴情况. 姿态估计算法^[2-4], 能够有效地将图像信息中的人体姿态转化为姿态坐标, 人体姿态坐标的特征

丰富,数据量小,便于传输,使动作识别的速度更快,准确性更强;数据交互算法,能够利用一种数据增加其他数据算法的性能,包括准确率和实时性,如匈牙利算法^[5]可以最大程度地将两种集合进行匹配,为护具佩戴检测和动作识别的并行识别提供了新的思路。

护具佩戴检测属于护具检测领域,是利用目标检测的方法,检测人员是否佩戴护具(安全帽、安全背心),并剔除未正确佩戴的护具信息(如手持安全帽),得到人员的护具佩戴情况。传统的目标检测算法需要通过手工设计特征,在实际检测中存在准确率低、模型泛化性差等问题。近年来,国内外学者提出了一系列基于深度学习的目标检测算法,可以用于护具检测系统。2018年 Seong 等^[6]利用安全背心的颜色进行了色彩空间转换和亮度成分评估,提出了一种颜色空间转换的方法,增强安全背心检测的精度。2019年 Fang 等^[7]针对传统检测方法准确率低、鲁棒性差的问题,提出了一种基于改进 YOLOv2 网络的安全帽佩戴情况检测方法,识别速度可以达到每秒 148 帧(YOLOv3 的实时性低于 YOLOv2)。2019年 Dasgupta 等^[8]提出了一种基于卷积神经网络的摩托车安全帽佩戴情况检测结构,通过实验评价,检测结果优于传统的卷积神经网络。2020年 Felix 等^[9]利用交通数据制作摩托车头盔佩戴情况数据集,利用深度学习算法对复杂路况下,摩托车佩戴情况进行检测,并通过实验证明了其方法的有效性。已有的护具佩戴检测方法,在检测过程中没有对未佩戴的护具与佩戴后的护具进行区分,而未佩戴的护具与佩戴后的护具特征相似,属于未正确佩戴护具行为,因此,本文对文[1]的数据交互方法进行改进,在护具佩戴检测算法中引入姿态信息,对这两种特征进行明确区分。

动作识别是深度学习和机器视觉领域的一个分支,是对各种背景下的人员动作(打电话、挥舞等)进行识别。随着人工智能技术的发展,基于机器视觉技术的人员动作监控已成为可能。2019年 Lei 等^[10]提出了一种基于骨架匹配的人体姿态估计方法,通过对姿态信息分类获得动作分类信息,有效地实现了动作分类。2020年 Tufek 等^[11]利用数据平衡和数据扩充方法,在有限的动作数据集上实现了动作分类,并通过实验证明了该方案的有效性。2020年 Tsai 等^[12]使用 3D ConvNet 结构和基于非极大值抑制原理的数据处理方法,实现了基于深度学习算法的多人动作识别系统。针对背景信息影响动作识别精度的问题,2020年周波等^[13]提出一种新的结合目标检测的人体行为识别的算法,在人体动作识别的过程中增加目标检测机制,使神经网络有侧重地学习人体的动作信息,降低了背景信息对动作识别精度的影响。此外,很多学者使用智能传感器进行动作识别,如 Kinect 摄像机,2020年 Wang 等^[14]对基于 Kinect 的动作识别方法进行了综述,并基于综述进行了算法改进,提升了算法的性能。智能传感器简化了动作识别流程,但对设备要求较高,不易于算法的推广。以上方法针对动作识别的准确性出发,但现实需求对动作识别算法的实时性同样有较高的要求^[15]。

通过对已有方法的研究,针对算法的实时性问题,本

文提出了基于姿态估计的护具佩戴检测与动作识别并行识别方案。并行识别方案使用深度可分离卷积^[16]对 Openpose 模型进行轻量化改进,改善姿态估计的实时性,快速获得多种人体姿态坐标,保证动作特征的丰富性。本方案采用了一种基于距离和匈牙利算法的数据交互方法和基于姿态信息的四边形检测法,利用头肩区域的强鉴别性^[17],对未佩戴护具和佩戴后护具的特征进行明确区分,剔除误判的护具信息,并选取 14 种关键姿态节点,从多方位获得姿态数据,通过深度神经网络(deep neural network, DNN),快速实现动作识别,有效解决了断续的图像序列难以有效识别的问题。

1 并行识别系统框架

基于姿态估计的并行识别系统主要可以分为 4 个部分:姿态估计、动作分类、护具检测和交互。第一部分是姿态估计。在多特征的情况下,深度可分离卷积(depthwise separable convolution, DSC)可以有效压缩网络模型的参数量。本文利用 DSC 对 Openpose 模型的特征提取层进行改进,降低 Openpose 模型的参数量,有效提升姿态估计的实时性。另外,针对 Openpose 模型存在对同一人体躯干重复估计的情况,本文在 Openpose 模型中加入非极大值抑制算法,去除重复的姿态信息,提高姿态估计的精度。第二部分是动作分类。改进的 Openpose 模型将图像特征转化为关节特征,利用算法将关节特征转化为关节坐标,通过 DNN 网络即可对图像序列进行有效的人体动作识别。第三部分是护具检测。Darknet53 模型可以快速准确地提取护具信息的特征,利用 YOLOv3 + Darknet53 模型可以有效地检测出护具信息,包括安全帽和安全背心。第四部分是数据交互。匈牙利算法可以最大程度地将两个集合进行匹配,本文利用人体躯干和护具检测框的距离信息,使用匈牙利算法对人体躯干和护具进行检测,并利用四边形检测法对护具检测框进行检验,剔除未正确佩戴的护具检测框。本文提出的并行识别系统框架如图 1 所示。

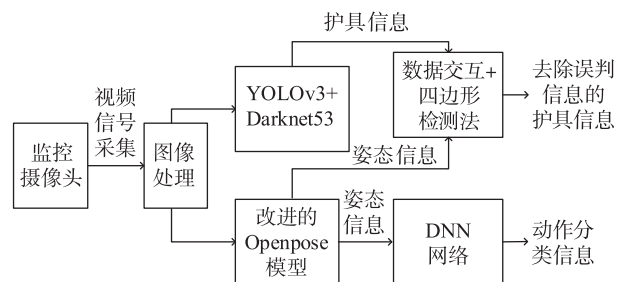


图 1 并行识别系统框架图

Fig.1 Framework of parallel recognition system

考虑到系统的实时性,本并行识别系统选取不同动作切换过程中,较为活跃的 14 个姿态信息,进行 DNN 动作分类。

2 姿态估计模型搭建

本文使用 DSC 对 Openpose 模型的特征提取层进行改

进,降低模型的参数量,增加 Openpose 模型姿态估计的速度,并搭建深度神经网络(DNN),对姿态信息进行动作分类,提出了一种基于改进 Openpose 模型和 DNN 的快速动作分类方案.

2.1 基于 DSC 改进的 Openpose 模型

多人姿态估计包含自顶向下和自底向上两种常用的方法.自顶向下,首先检测出每个人体,再对人体进行关节识别;自底向上,先对所有的关节点进行识别,然后对关节点进行分组、连接,构成人体躯干.在多人场景下,同时对每个人的关节点进行识别速度更快,为保证识别系统具有更高的实时性,本文根据自底向上的思路,使用改进的 Openpose 模型,进行姿态估计.

传统的 Openpose 模型利用 VGGnet (Visual Geometry Group network) 获取图像特征,再利用图像特征进行特征点识别和特征区域识别.特征点识别,利用卷积网络识别出 18 个关节点,输出每个关节点的热图以及 1 个背景热图.特征区域识别,对 19 种特征点信息,两两之间生成骨骼区域的热图,共 38 种(C_{19}^2).可视化后的特征点和特征区域热图如图 2 所示.在图 2 中,图 2(a)为可视化后的脚踝关节点,图 2(b)为可视化后的特征区域(骨骼).

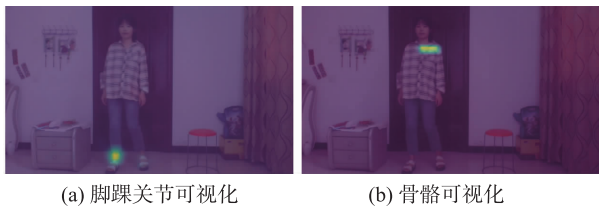


图 2 可视化特征图像
Fig.2 Visual features image

Openpose 模型对一个躯干进行姿态估计时,需要计算 19 种特征点和 38 种特征区域,共 57 种数据特征,网路参数庞大,降低了 Openpose 模型的识别速度.因此,本文使用 DSC 对 Openpose 模型进行改进,降低参数量,可以增加系统的实时性.

DSC 能够有效压缩多特征模型^[16],本文使用 DSC 对 Openpose 的特征提取层进行改进,得到改进的特征提取层 DSCnet (Depthwise Separable Convolution network),使 Openpose 模型拥有更轻量化的网络结构,提升姿态估计模型的实时性. VGGnet、DSCnet 的结构如图 3 和图 4 所示.

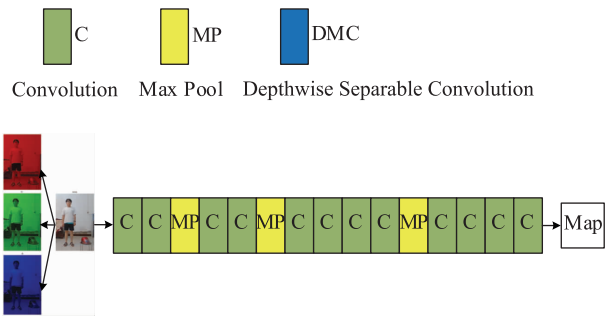


图 3 传统特征提取层: VGGnet
Fig.3 Traditional feature extraction layer: VGGnet

如图 3 所示, VGGnet 的输入图像为三通道的 RGB 图像,图像的宽度为 W ,长度为 L .正常卷积令图像的通道数(深度)为卷积核的第三维度,即正常卷积的卷积核大小为 $M \times N \times 3$,数量为 1,得到的图像特征的大小为 $(W - M + 1) \times (L - N + 1) \times 1$.一个图像特征所需的卷积核参数为 $M \times N \times 3$ 个,姿态估计需要获得 57 个图像特征,所需的卷积核参数为 $M \times N \times 3 \times 57$ 个.

在图 4 中, DSCnet 的输入图像为三通道的 RGB 图像,图像的宽度为 W ,长度为 L .深度可分离卷积的实现分为两步:第 1 步,深度可分离卷积令图像的通道数(深度)为卷积核数量,即深度可分离卷积的卷积核大小为 $M \times N \times 1$,数量为 3,得到的图像特征的大小为 $(W - M + 1) \times (L - N + 1) \times 3$.第 2 步,建立大小为 $1 \times 1 \times 3$ 的特征卷积核(不同图像特征对应不同参数的特征卷积核)对第 1 步获得的图像特征进行卷积,获得最终的图像特征大小为 $(W - M + 1) \times (L - N + 1) \times 1$,与正常卷积的图像特征大小一致.一个图像特征所需的卷积核参数为 $M \times N \times 1 \times 3 + 1 \times 1 \times 3$ 个,姿态估计需要获得 57 个图像特征,因此,建立 57 个大小为 $1 \times 1 \times 3$ 的特征卷积核对第一步中获得的图像特征进行卷积,获得 57 个图像特征,所需的卷积核参数为 $M \times N \times 1 \times 3 + 1 \times 1 \times 3 \times 57$ 个.

由图 3 和图 4 的分析可知,利用 VGGnet 进行 57 个图像特征提取时,卷积层需要的卷积核参数为 $M \times N \times 3 \times 57$ 个;利用 DSCnet 进行 57 个图像特征提取时,卷积层需要的卷积核参数为 $M \times N \times 1 \times 3 + 1 \times 1 \times 3 \times 57$ 个,远远低于 VGGnet 所需要的卷积核参数个数,可以大大提高 Openpose 模型的识别速度.

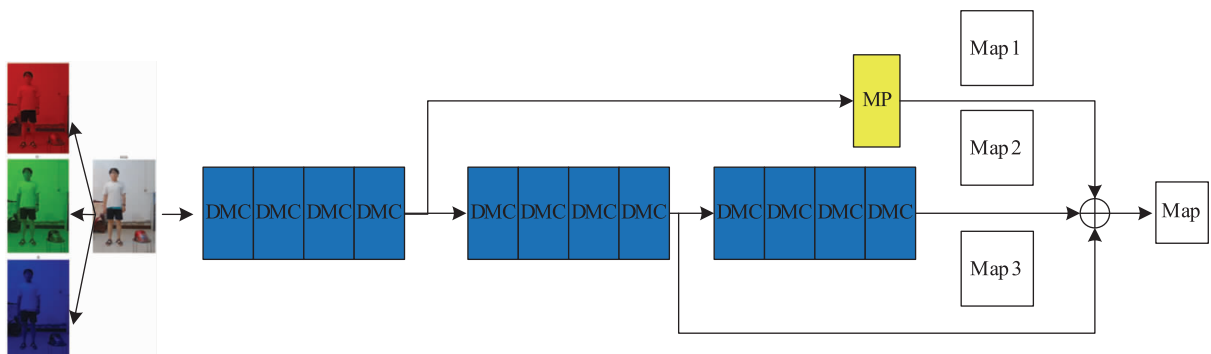


图 4 改进特征提取层: DSCnet
Fig.4 Improve feature extraction layer: DSCnet

2.2 基于姿态信息的 DNN 动作分类模型

基于图像序列的动作识别方法, 难以对断续图像序列有效识别. 本文使用改进的 Openpose 模型将图像中的人体特征转化为人体关节坐标数据, 保证了人体动作特征的丰富性和轻量化(坐标数据格式简单, 数据量小, 便于后续处理和传输), 并在制作数据集时, 从多角度采集数据, 保证了各种动作数据的丰富性. 在数据丰富情况下, 本文

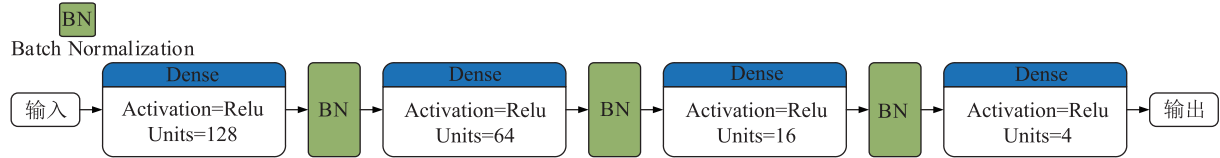


图5 DNN 动作分类网络结构

Fig.5 DNN action classification network structure

基于姿态信息的 Batch Normalization 算法实现可以分为两步: 第一步, 利用式(1)~式(3)将输入的多种姿态特征 P_i 进行正态标准化, 获得标准化后的姿态信息 \hat{P}_i . 第二步, 利用式(4)为 P_i 增加梯度信息, 获得 \tilde{P}_i , 避免 DNN 动作分类网络出现梯度消失问题.

$$\mu_B = \frac{1}{k} \sum_{i=1}^k P_i \quad (1)$$

$$\sigma_B^2 = \frac{1}{k} \sum_{i=1}^k (P_i - \mu_B)^2 \quad (2)$$

$$\hat{P}_i = \frac{P_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (3)$$

在式(1)~式(3)中, k 为姿态特征的种类数, μ_B 为 k 种姿态特征的均值, σ_B^2 为 k 种姿态信息的方差, ε 为方差阈值. 利用式(3), 将姿态信息正态标准化.

$$\tilde{P}_i = \gamma \hat{P}_i + \beta = \text{BN}_{\gamma, \beta}(P_i) \quad (4)$$

在式(4)中, γ 和 β 为梯度函数系数和常数向量. 利用式(4)改变姿态特征变化的梯度, 从而增加网络收敛速度, 避免梯度消失问题.

3 护具信息与姿态信息交互算法

在人员躯干与护具信息未进行匹配的情况下, 进行数据交互容易出现信息误删问题(护具误删). 基于这一问题, 本文提出一种基于距离和匈牙利算法的数据交互方法, 将目标检测算法获得的护具信息和动作识别中的姿态信息进行匹配, 解决了信息误删问题, 实现了护具佩戴检测和动作识别的数据交互. 利用四边形检测法进行护具佩戴检测算法的误判判断(如手持安全帽), 将未正确佩戴的护具信息进行剔除, 增加了护具佩戴检测算法的准确性.

3.1 基于距离和匈牙利算法的数据交互算法

Openpose 模型在姿态估计过程中, 同一个躯干存在重复估计的问题, 如图 6 所示, 降低了数据交互的准确率. 本文在改进 Openpose 模型的输出阶段, 利用式(5)、式(6)加入非极大值抑制算法(non-maximum suppression, NMS), 对姿态信息进行预处理, 去除重复的姿态信息.

设计 DNN 进行动作分类, 并采用全连接层(Dense)和批量标准化层(batch normalization, BN)交替的网络结构, 如图 5 所示. 针对 DNN 梯度消失问题, 本文利用基于姿态信息的 Batch Normalization 算法, 对中间网络层的输出进行标准化处理后, 降低了不同样本间值域的差异性, 增加姿态特征的变化梯度, 解决了梯度消失的问题.

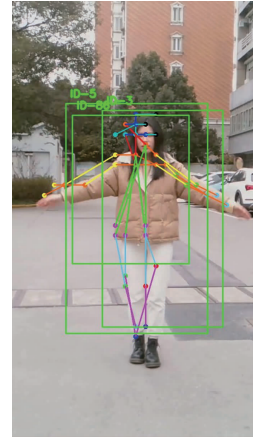


图6 姿态重复信息

Fig.6 Repeat data of pose

$$N_{\text{map}} = \begin{cases} 0, & U_{\text{IoU}}(\mathbf{M}, N_{\text{in}}) \geq N_t \\ N_{\text{score}}, & U_{\text{IoU}}(\mathbf{M}, N_{\text{in}}) < N_t \end{cases} \quad (5)$$

$$N_{\text{ms}} = \text{Max}(N_{\text{map}}, P_{\text{foot}}) \quad (6)$$

在式(5)、式(6)中, N_{in} 为待检测的姿态信息, N_{score} 为待检测姿态信息的原始分数, \mathbf{M} 为得分最高的姿态信息, N_t 为预设阈值, N_{map} 为待检测姿态信息的最后得分, N_{ms} 为非极大值抑制算法的输出值, P_{foot} 为局部最优化窗口. 利用式(5)从 N_{in} 获得每个姿态信息的最后姿态信息的 N_{map} , 再使用 Max 函数, 利用 P_{foot} 进行滚动, 对姿态信息进行去重.

基于距离和匈牙利算法的数据交互方法, 利用式(7)、式(8), 计算护具检测框与躯干的距离, 再利用距离作为匈牙利算法的输入, 进行护具信息和躯干信息的匹配.

$$L_{ij} = L(\mathbf{H}_i, \mathbf{P}_j) = \sqrt{(x_i - x_j)^2 - (y_i - y_j)^2} \\ \mathbf{H}_i \in H_{\text{et}}, \mathbf{P}_j \in P_{\text{osc}} \\ i \in \{1, 2, 3, \dots, n\}, j \in \{1, 2, 3, \dots, m\} \quad (7)$$

$$N_{\text{HP}} = \begin{cases} \infty, & L_{ij} \geq L_i \\ L_{ij}, & L_{ij} < L_i \end{cases} \quad (8)$$

在式(7)中, $\mathbf{H}_i = (x_i, y_i)$ 为第 i 个护具检测框的中心点坐标; H_{et} 为护具检测框中心点坐标集合; n 为护具检测框个数;

$P_j = (x_j, y_j)$ 为第 j 个躯干的 0 号关节点 (Nose) 坐标, 正确佩戴的安全帽; Nose 关节点始终处于安全帽的检测框之内; P_{ose} 是姿态关节坐标集合; m 为躯干数量, L_i 为距离转换阈值, 等于第 i 个护具检测框长和宽的算数平均数. 首先利用式(7), 计算各个护具检测框与躯干的距离 L_{ij} , 然后利用式(8), 将 L_{ij} 中超过阈值 L_i 的值转换为无穷大, 从而得到匈牙利算法的输入量 N_{HP} , 利用匈牙利算法将护具信息与人体躯干进行匹配.

护具信息和躯干数据的交互关系可以分为两种情况, 如图 7 所示, 分别为护具信息与躯干数目相同、护具信息与躯干数目不相同. 护具佩戴检测的目标为判断人员是否正确佩戴护具, 因此, 本文在使用匈牙利算法对护具和躯干进行匹配时, 以躯干为主导, 优先为所有躯干匹配护具检测框.



图 7 数据交互格式
Fig.7 Format of exchange data

式(8)获得的 N_{HP} 为具体数值, 其数值的大小为躯干和护具匹配的具体量纲, 基于此, 本文对传统的匈牙利算法进行改进, 具体步骤如下:

1) 由式(7)中护具检测框集合 H_{et} 和姿态关节坐标集合 P_{ose} 的维度建立行数 m , 列数 n 的匈牙利矩阵 M_{agy} .

2) 利用式(8), 计算 $m \times n$ 种 P_j 与 H_i 的 N_{HP} , 并对每个躯干的 n 个 N_{HP} 进行升序排序, 然后将排序的名次填入匈牙利矩阵 M_{agy} .

3) 先将矩阵 M_{agy} 的每一行减去该行的最小值, 获得矩阵 M_{agyF} , 再将矩阵 M_{agyF} 的每一列减去该列的最小值, 获得矩阵 M_{agyS} .

4) 将矩阵 M_{agyS} 中, 所有不同行且不同列的 0 元素取出, 读出 0 元素对应坐标值 (j, i) , 将第 j 个躯干与第 i 个护具检测框进行匹配.

3.2 四边形检测法

针对未佩戴护具的误判问题(如手持安全帽), 本文提出了结合姿态信息的四边形检测法, 示意图如图 8 所示. 利用基于距离和匈牙利算法的数据交互方法将躯干与护具检测框进行匹配, 通过计算节点坐标与安全帽的位置关系, 将误判的护具信息剔除. 假设, 护具检测框 box 的左上角坐标 $Q_i(x_{min}, y_{min})$ 和右下角坐标 $Q_j(x_{max}, y_{max})$. 利用

公式(9)通过 Q_i 和 Q_j 建立第一级数值四边形 $Q_{uaF}(Q_{F11}, Q_{F12}, Q_{F21}, Q_{F22})$.

$$Q_{uaF}(Q_{F11}, Q_{F12}, Q_{F21}, Q_{F22}) = \begin{bmatrix} Q_{F11} & Q_{F12} \\ Q_{F21} & Q_{F22} \end{bmatrix} = \begin{bmatrix} (x_{min}, y_{min}) & (x_{max}, y_{min}) \\ (x_{min}, y_{max}) & (x_{max}, y_{max}) \end{bmatrix} \quad (9)$$

如图 8 所示, box1、box2、box3 均为护具检测框, box1 和 box3 为包含了 0 号关节点 (Nose) 的护具误判信息, 为了将 box1 和 box3 进行剔除, 必须对数值四边形进行压缩. 设置压缩比例 θ , 压缩比例 θ 由式(10)获得. 获得压缩比例 θ 后, 利用式(11), 获得第二级数值四边形 $Q_{uaS}(Q_{S11}, Q_{S12}, Q_{S21}, Q_{S22})$.

$$\theta = \begin{cases} 1 - \frac{n_h}{10}, & n_h \leq 5 \\ 0.5, & n_h > 5 \end{cases} \quad (10)$$

在式(10)中, n_h 为每个躯干匹配的护具检测框数量, 压缩比例最多为 0.5, 保证算法的准确性.

$$Q_{uaS}(Q_{S11}, Q_{S12}, Q_{S21}, Q_{S22}) = Q_{uaF}(Q_{F11}, Q_{F12}, Q_{F21}, Q_{F22}) \cdot \theta - \frac{\theta}{2} \cdot \begin{bmatrix} (\Delta x, \Delta y) & (\Delta x, \Delta y) \\ (\Delta x, \Delta y) & (\Delta x, \Delta y) \end{bmatrix} \quad (11)$$

$$\Delta x = x_{max} - x_{min}, \Delta y = y_{max} - y_{min}$$

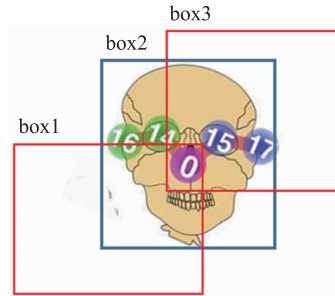


图 8 四边形检测法示意图
Fig.8 Schematic diagram of quadrilateral detection

判断 0 号关节点坐标 $P_j = (x_j, y_j)$ 和第二级数值四边形 Q_{uaS} 的位置关系, 若 $Q_{S11} < P_j < Q_{S22}$, 则 0 号关节点在 Q_{uaS} 之内, 保留此护具检测框; 否则, 将此护具检测框去除.

4 实验与分析

本文实验中, 使用准确率 (Ar, Accuracy rate), 帧速 (FPS, Frames Per Second), 漏检率 (Mr, Miss rate), 误检率 (FFr, False Fositive rate) 来测试模型的性能. 实测实验中, 使用单 GeForce2080 进行算法的性能验证.

$$Ar = \frac{TP}{TP + FP} \quad (12)$$

$$FFr = \frac{FP}{RP + FP} \quad (13)$$

$$Mr = \frac{MP}{RP + MP} \quad (14)$$

在式(12)~式(14)中, FP 为误判帧, RP 为正确识别帧, MP 为漏判帧. 在利用公式计算 Ar、FFr、Mr 时, 考虑到动作的延续性, 本实验以每秒 2 帧的格式提取和计算 FP、RP、MP 的值.

4.1 数据集

本文的数据集包括护具佩戴检测数据集(protective appliance wearing dataset, PAWD)、动作识别数据集(skeleton Openpose dataset, SOD)两部分. 在开源的安全帽检测数据集(safety helmet wearing dataset, SHWD)的基础上, 加入了 hat 和 vest 两类数据进行扩充, 构成了最终的护具佩戴检测数据集 PAWD. SOD 由 NTU RGB + D 动作识别数据集^[18]和 Openpose 数据集组成, 其中 Openpose 数据集为本文自制数据集.

4.1.1 PAWD 数据集

SHWD 数据集中只包括 helmet、person 两类标签, 无法对 helmet 和 hat 进行区分, 且无法识别安全背心, 因此本文对 SHWD 数据集进行扩充, 得到 PAWD 数据集. 经过对 SHWD 数据集的扩充, 本文的 PAWD 数据集共 7 462 张图片, 标签设置为 hat、helmet、person 和 vest 四类. 从数据集中, 选取 80% 的数据作为训练集, 20% 的数据作为测试集, 然后再从训练集中选取 10% 的数据作为验证集, 最后进行交叉验证测试, 结果如表 1 所示.

表 1 护具佩戴检测数据集测试结果
Tab.1 Test results of protective wear detect dataset

数据集	Epoch: 10 mAP	Epoch: 20 mAP	Best mAP
SHWD	0.734	0.815	0.821
PAWD	0.692	0.791	0.791

在目标检测问题中, mAP(mean Average Precision) 是衡量检测结果的一个重要指标. 表 1 中记录了两种数据集在 YOLOv3 + Darknet 53 网络下训练 10 次、20 次达到的平均精度. 在第 13 次, SHWD 数据集的 mAP 达到的最大值; 在第 17 次, PAWD 数据集的 mAP 达到的最大值. 由于标签类别数的增加, PAWD 数据集的 mAP 值稍低于 SHWD 数据集, 且 mAP 值的最大值晚于 SHWD 数据集出现, 符合训练规律, 数据集扩充成功.

4.1.2 SOD 数据集

本文实验中选取关节点标号对应表中的 14 种关节点作为动作识别的输入, 关节点标号对应表如表 2 所示; 并利用 NTU RGB + D 动作识别数据集中的 Skeleton 数据和自制 Openpose 数据制作本文实验中的 SOD 数据集.

1) Openpose 数据

Openpose 数据为本文的实测数据, 通过改进的 Openpose 模型进行姿态估计, 得到各个动作对应的关节点数据, 制作数据集. 步骤如下: (1)选取 6 位体型相差较大的同学依次做各种动作, 对每种连续动作的各种情况进行采集, 如图 9 所示, 将各种打电话动作均进行采集. (2)设置姿态估计网络, 以每秒固定帧数的方式, 输出指定的 14 个关节点信息. (3)建立数据集白板, 将输出的关节点信息导入.



图 9 动作实测数据

Fig.9 Measured data of action

2) Skeleton 数据

在动作识别的研究范围内, COCO、NTU RGB + D 两个数据集的认可度较高. Skeleton 数据是 NTU RGB + D 数据集中的三维骨骼数据, 由 Kinect 传感器采集获得, 含深度信息. 本文实验使用 NTU RGB + D 数据集中等量 Skeleton 数据和自制 Openpose 数据构造数据集, 进行动作分类网络的训练. Skeleton 数据与 Openpose 数据的格式如图 10 所示. 由于 Skeleton 数据与 Openpose 数据格式不同, 因此需要对 Skeleton 数据进行格式转换.



(a) Skeleton数据格式

(b) Openpose数据格式

图 10 动作数据格式图

Fig.10 Format of action data

表 2 为 Skeleton 数据和 Openpose 数据的关节点标号对应表. 本文根据关节点标号对应表, 提取 Skeleton 数据中 Openpose 数据对应的 14 个关节点数据, 再利用去除深度值的方法, 将三维 Skeleton 数据转化为二维 Openpose 数据.

3) 动作识别数据集测试

将数据集分为三类进行测试, 包括 Skeleton Dataset、Openpose Dataset、Skeleton + Openpose Dataset. 从各自数据集中, 选取 10% 的数据做验证集, 90% 的数据为训练集,

进行交叉训练测试, 获得动作识别的 Ar (Accuracy rate) 如表 3 所示.

表 2 关节点标号对应表
Tab.2 Correspondence table of joint label

关节名	Openpose	Skeleton
Nose	0	4
Neck	1	21
RShoulder	2	9
Relbow	3	10
RWrist	4	12
LShoulder	5	5
LElbow	6	6
LWrist	7	8
RHip	8	17
Rknee	9	18
Rankle	10	19
LHip	11	13
LKnee	12	14
LAnkle	13	15

表 3 动作识别数据集测试结果

Tab.3 Test results of action recognition dataset

数据集	Skeleton	Openpose	Skeleton + Openpose
Ar	0.94	0.938	0.91

由测试可知, Skeleton 数据集和 Openpose 数据集的准确率较高, 但 SOD 数据集的鲁棒性更高. 在人员数据可测的情况下, 可选择 Openpose 数据集进行动作识别, 在人员信息未知的情况下, 可选择 SOD 数据集进行动作识别.

4.2 实验结果与分析

4.2.1 Openpose 改进验证实验

本文针对于 Openpose 模型的实时性不高的问题, 利用 DCS 压缩模型参数的特性, 对 Openpose 模型的特征提取层进行改进, 并设计实验进行验证. 本实验使用准确率 Ar 和帧速 FPS 来评价系统的性能, 分为单人识别和多人识别两类, 分别在楼梯间、停车场和户外道路三种环境下进行实验, 实验结果如图 11 和表 4 所示. 图 11 左侧 3 幅图为楼梯间实验, 中间 3 幅图为停车场实验, 右侧 3 幅图为户外道路实验.



图 11 Openpose 改进验证实验结果图

Fig.11 Result of Openpose improved verification experiment

表 4 Openpose 改进验证对比实验结果

Tab.4 Result of Openpose improved verification comparison experiment

网络类别	Ar	FPS	模型大小
DSCnet	0.947	22	7.6 MB
VGGnet	0.955	10	204.4 MB

由表 4 中的测试结果可知, VGGnet 的模型大小是 DSCnet 的 26.9 倍, 每秒识别动作的帧数比 DSCnet 少 12 帧. 原因在于, DSCnet 中, 利用深度可分离卷积对网络的参数进行轻量化处理, 大大提高了网络的处理速度. 通过实验得到结论: 使用 DSCnet 作为特征提取层, 进行基于

Openpose 模型的动作识别, 大大提升动作识别的实时性.

4.2.2 安全帽检测对比实验

本文中, 利用数据交互方法和四边形检测法, 将姿态信息加入到护具检测算法中, 剔除误判信息, 提升护具检测算法的准确率. 为验证基于姿态信息的护具检测算法的有效性, 本文设计安全帽检测对比实验, 使用准确率 Ar、漏检率 Mr 和误判率 FFr 对算法的进行验证. 实验分为二个实验组: YOLOv3^[20] 和 YOLOv3 + PoseData (基于姿态信息的安全帽检测算法). 实验分别在楼梯间、停车场和户外道路三种环境下进行, 实验结果如图 12 和表 5 所示.



图 12 安全帽检测对比实验结果图

Fig.12 Result of helmet detect comparison experiment

图 12 上面 3 幅图为三种环境下, YOLOv3 测试实验. 图 12 下面 3 幅图为三种环境下, YOLOv3 + PoseData 测试实验.

表 5 安全帽检测对比实验结果

Tab.5 Result of helmet detect comparison experiment

网络类别	Ar (HaH)	FFr (HaH)	Ar (Worn)	Mr (Worn)	Ar
YOLOv3	0.675	0.325	0.952	0.048	0.867
YOLOv3 + PoseData	0.968	0.032	0.945	0.065	0.959

由表 5 中的实验结果可知, 利用数据交互算法将误判的护具检测框进行剔除, 可以提高护具检测的准确性. 由结果可知, YOLOv3 + PoseData 获得的手持安全帽误判率比

文[19]中的安全帽检测方法降低 29.3%, 整体准确率增加 9.2%. 因此, 由实验结果得到结论: 结合了姿态信息的 YOLOv3 + PoseData 安全帽检测模型, 大幅降低了手持安全帽的误判率, 提升了护具佩戴检测系统的整体准确率, 增加了系统的实用性.

4.2.3 相关研究对比实验

利用基于距离和匈牙利算法的数据交互方法, 本文并行完成了动作识别和护具佩戴检测, 提出一种快速的并行识别方案. 本文设计并行识别实验, 使用准确率 Ar、漏检率 Mr 和 FPS 对并行识别方案进行验证. 实验分为单人并行识别和多人并行识别两类, 分别在楼梯间、停车场和户外道路三种环境下进行实验, 实验结果如图 13 和表 6 所示.



图 13 人员行为识别实验结果图

Fig.13 Result of human behavior recognition experiment

图 13 左侧 3 幅图为楼梯间实验, 中间 3 幅图为停车场实验, 右侧 3 幅图为户外道路实验。

表 6 并行识别实验结果

Tab.6 Result of parallel recognition experiment

Ar(Action)	Ar(Helmet)	Ar(Vest)	Ar	FPS
0.942	0.953	0.910	0.935	9

在并行识别实验中, 并行识别模型的整体正确率 Ar 为动作识别准确率、安全帽检测准确率和安全背心准确率的算术平均值。由表 6 中的实验结果可知, 模型的整体正确率为 93.5%。动作识别准确率和安全帽检测准确率与单独测试获得结果一致, 并行识别实验结果合理。考虑到行为的延续性, 本文并行识别模型 FPS 达到每秒 9 帧, 基本满足识别要求。因此, 并行实验结果得到以下结论: 本文提出的 D-ODY (DSC-Openpose + DNN + YOLOv3, D-ODY) 并行识别方案, 在满足系统实时性的要求下, 能够有效识别出动作信息、安全帽和安全背心佩戴情况。

D-ODY 并行识别方案在 NTU RGB + D 标准数据集上的动作识别准确率, PAWD 数据集上的护具佩戴检测的准确率整体优于其他文献。比文 [9] 的护具佩戴检测准确率高 5.2%, 比文 [12] 的动作识别准确率高 3.4%, 比文 [20] 的护具佩戴检测准确率高 9.2%, 比文 [21] 的动作识别准确率高 7.3%。

参考文献

- [1] 王雨生, 顾玉宛, 封晓晨, 等. 基于姿态估计的安全帽佩戴检测方法研究[J]. 计算机应用研究, 2021, 38(3): 937-940, 945. Wang Y S, Gu Y W, Feng X C, et al. Research on detection method of helmet wearing based on attitude estimation[J]. Application Research of Computers, 2021, 38(3): 937-940, 945.
- [2] Munea T L, Jembre Y Z, et al. The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation [J]. IEEE Access, 2020, 8: 133330-133348.
- [3] Liang G, Zhong X, Ran L, et al. An adaptive viewpoint transformation network for 3D human pose estimation[J]. IEEE Access, 2020, 8: 143076-143084.
- [4] Xu J, Yu Z, Ni B, et al. Deep kinematics analysis for monocular 3D human pose estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2020: 896-905.
- [5] Pang C, Shan G L, Ma W N, et al. Sensor radiation interception risk control in target tracking[J]. Defence Technology, 2020, 16(3): 695-704.
- [6] Seong H, Son H. A comparative study of machine learning classification for color-based safety vest detection on construction-site images[J]. KSCE Journal of Civil Engineering, 2018(22): 4254-4262.
- [7] Fang M, Sun T T, Shao Z. Fast helmet-wearing-condition detection based on improved YOLOv2[J]. Optics and Precision Engineering, 2019, 27(5): 1196-1205.
- [8] Dasgupta M, Bandyopadhyay O, Chatterji S. Automated helmet detection for multiple motorcycle riders using CNN[C]//IEEE Conference on Information and Communication Technology. Piscataway, USA: IEEE, 2019. DOI: 10.1109/CICT48419.2019.9066191.
- [9] Siebert F W, Lin H H. Detecting motorcycle helmet use with deep learning[J]. Accident Analysis the Prevention, 2020, 134. DOI: 10.1016/j.app.2019.105319.
- [10] Lei F, An Z H, Wang X L. Pose estimation of complex human motion[C]//International Conference on Video and Image Processing. Piscataway, USA: IEEE, 2019: 153-156.
- [11] Tufek N, Yalcin M, Altintas M, et al. Human action recognition using deep learning methods on limited sensory data[J]. IEEE Sensors Journal, 2020, 20(6): 3101-3112.
- [12] Tsai J K, Hsu C C, Wang W Y, et al. Deep learning-based real-time multiple-person action recognition system[J]. Sensors, 2020, 20(17): 4758.
- [13] 周波, 李俊峰. 结合目标检测的人体行为识别[J]. 自动化学报, 2020, 46(9): 1961-1970. Zhou B, Li J F. Human action recognition combined with object detection[J]. Acta Automatica Sinica, 2020, 46(9): 1961-1970.

(下转第 739 页)

表 7 D-OPY 与相关文献对比

Tab.7 Comparison of D-OPY and related literature

方法	Ar(Action)	Ar(Helmet)
D-ODY	0.942	0.953
RetinaNet ^[9]		0.901
3DConvNet + NMS ^[12]	0.908	
DeepSORT + YOLOv3 ^[19]		0.867
PoT2I + CNN ^[20]	0.869	

5 结论

本文提出了一种 D-ODY 并行识别方案, 快速并行实现了护具佩戴检测和动作识别。D-ODY 使用 DSC 对 Openpose 模型的特征提取层进行改进, 使姿态估计的 FPS 达到每秒 22 帧, 较未改进的 Openpose 模型提高 12 帧。本文在 D-ODY 中, 使用基于距离和匈牙利算法的数据交互算法, 以及四边形检测法, 将姿态信息加入到安全帽检测算法中, 使手持安全帽的误判率下降 29.3%。并以实时性为前提, 并行完成动作识别和护具佩戴检测, 整体准确率达到 93.5%, 解决了护具佩戴检测和动作识别并行识别实时性低的问题。但 D-ODY 并行识别方案依然处于实验室研究阶段, 系统的鲁棒性还有待提升, 在今后的研究中, 会继续改进, 提高方案的实用性。

- [10] Fasahat M, Manthouri M. State of charge estimation of lithium-ion batteries using hybrid autoencoder and long short term memory neural networks[J]. *Journal of Power Sources*, 2020, 469. DOI: 10.1016/j.jpowsour.2020.228375.
- [11] Shateri N, Shi Z H, Auger D J, et al. Lithium-sulfur cell state of charge estimation using a classification technique[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(1): 212–224.
- [12] Huang G B, Zhu Q Y, Siew C. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1/2/3): 489–501.
- [13] Huang G, Huang G B, Song S J, et al. Trends in extreme learning machines: A review[J]. *Neural Networks*, 2015, 61: 32–48.
- [14] 宋绍剑, 王志浩, 林小峰. 基于极限学习机的磷酸铁锂电池 SOC 估算研究[J]. *电源技术*, 2018, 42(6): 806–808, 881.
Song S J, Wang Z H, Lin X F. Research on SOC estimation of LiFePO₄ batteries based on ELM[J]. *Chinese Journal of Power Sources*, 2018, 42(6): 806–808, 881.
- [15] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 513–529.
- [16] 吴忠强, 尚梦瑶, 申丹丹, 等. 基于 BSA-RELM 的纯电动汽车锂离子电池 SOC 估计[J]. *计量学报*, 2019, 40(4): 693–699.
Wu Z Q, Shang M Y, Shen D D, et al. Estimation of SOC of li-ion battery in pure electric vehicle by BSA-RELM[J]. *Acta Metrologica Sinica*, 2019, 40(4): 693–699.
- [17] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer[J]. *Advances in Engineering Software*, 2014, 69: 46–61.
- [18] Wang M J, Chen H L, Li H Z, et al. Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction[J]. *Engineering Applications of Artificial Intelligence*, 2017, 63: 54–68.
- [19] Heidari A A, Abbaspour R A, Chen H L. Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training[J]. *Applied Soft Computing*, 2019, 81. DOI: 10.1016/j.asoc.2019.105521.
- [20] Cai Z N, Gu J H, Luo J, et al. Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy[J]. *Expert Systems with Applications*, 2019, 138. DOI: 10.1016/j.eswa.2019.07.031.
- [21] 方一鸣, 赵晓东, 张攀, 等. 基于改进灰狼算法和多核极限学习机的铁水硅含量预测建模[J]. *控制理论与应用*, 2020, 37(7): 1644–1654.
Fang Y M, Zhao X D, Zhang P, et al. Prediction modeling of silicon content in liquid iron based on multiple kernel extreme learning machine and improved grey wolf optimizer[J]. *Control Theory & Applications*, 2020, 37(7): 1644–1654.
- [22] Mohamed F A, Koivo H N. System modelling and online optimal management of MicroGrid using mesh adaptive direct search[J]. *International Journal of Electrical Power & Energy Systems*, 2010, 32(5): 398–407.

作者简介

赵超(1976), 男, 博士, 副教授. 研究领域为群智能算法, 电动汽车电池管理系统技术开发, 新能源系统调度优化.
王延峰(1995), 男, 硕士生. 研究领域为动力电池状态估计.
林立(1985), 男, 博士, 讲师. 研究领域为新能源系统设计, 燃料电池系统仿真与系统集成.

(上接第 730 页)

- [14] Wang L, Huynh D Q, Koniusz P. A comparative review of recent kinect-based action recognition algorithms[J]. *IEEE Transactions on Image Processing*, 2020, 29: 15–28.
- [15] Zhao X, Huang Y, Yang J, et al. Discriminative pose analysis for human action recognition[C]//6th World Forum on Internet of Things (WF-IoT). Piscataway, USA: IEEE, 2020. DOI: 10.1109/WF-IoT48130.2020.9221390.
- [16] Zhang T, Zhang X, Shi J, et al. Depthwise separable convolution neural network for high-speed SAR ship detection[J]. *Remote Sensor*, 2019, 11(21): 2483–2520.
- [17] 姬东飞, 丁学明. 基于特定区域去相关的行人检测算法[J]. *信息与控制*, 2020, 49(4): 404–413.
Ji D F, Ding X M. Specific region decorrelation feature channel on pedestrian detection[J]. *Information and Control*, 2020, 49(4): 404–413.
- [18] Shahroudy A, Liu J, Ng T, et al. NTU RGB + D: A large scale dataset for 3D human activity analysis[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2016: 1010–1019.
- [19] 何超. 基于改进 YOLOv3 的安全帽检测系统研究[D]. 武汉: 华中科技大学, 2019.
He C. Research on safety helmet detection system based on improved YOLOv3[D]. Wuhan: Huazhong University of Science and Technology, 2019.
- [20] Thien H T, Hua C H, Ngo T T, et al. Image representation of pose-transition feature for 3D skeleton-based action recognition[J]. *Information Sciences*, 2020, 513: 112–126.

作者简介

吴胜昔(1976–), 女, 博士, 副教授. 研究领域为计算机视觉在工业领域的应用.
戚博龙(1994–), 男, 硕士生. 研究领域为机器视觉, 行为识别.
冒鑫鑫(1998–), 男, 硕士生. 研究领域为机器视觉, 行为识别.