

DOI: 10.19650/j.cnki.cjsi.J2210613

面向人机协作系统的上肢姿态精准识别算法研究*

张堃^{1,2}, 刘志诚¹, 刘纪元¹, 华亮¹, 费敏锐³

(1. 南通大学电气工程学院 南通 266019; 2. 南通市智能计算与智能控制重点实验室 南通 226007;
3. 上海大学机电工程与自动化学院 上海 210053)

摘要:在基于姿态识别协同控制灵巧手机械臂的任务中,会出现身体部位相互遮挡以及非操作人员身体干扰的问题。因此本文提出了一种面向人机协作系统的上肢姿态精准识别算法,能够有效排除遮挡和干扰问题。该算法首先基于 Finger-YOLOv4 算法框选出人体上肢区域;其次通过稀疏性目标提取算法排除非操作人员身体干扰;然后在设计的双特征条件随机场网络中进行深度学习,解决遮挡导致的类内模糊问题,精准定位人体上肢的 48 个关键点坐标;最后,根据关键点坐标进行人体上肢的姿态预测,将人体上肢的姿态与灵巧手机械臂的姿态进行映射,完成人机协作。实验表明,本算法平均检测速度 33 FPS,关键点平均检测精度 75.2%,协同操作完成度 98%。满足实际需求。

关键词: 人机协作; Finger-YOLOv4 算法; 稀疏性目标提取; 双特征条件随机场网络; 灵巧手机械臂

中图分类号: TP273 TH89 **文献标识码:** A **国家标准学科分类代码:** 510. 80

Research on the accurate recognition algorithm of upper limb posture for the human-manipulator cooperation system

Zhang Kun^{1,2}, Liu Zhicheng¹, Liu Jiyuan¹, Hua Liang¹, Fei Minrui³

(1. School of Electrical Engineering, Nantong University, Nantong 226019, China; 2. Nantong Key Laboratory of Intelligent Control and Intelligent Computing, Nantong 226007, China; 3. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 210053, China)

Abstract: In the task of pose recognition based cooperative control of dexterous hand manipulator, the occlusion of body parts and interference of non-operators are always the main factors that affect the control accuracy. To effectively eliminate the aforementioned problems, an accurate upper limb posture recognition algorithm is proposed for human-machine collaboration system. Firstly, a frame selecting scheme is applied to box upper limb region based on Finger-YOLOv4. Then, the sparse target extraction algorithm is applied to exclude body interference of the non-operators. Next, we formulate a deep learning framework DFCRF-Net which aims at accurate positioning of 48 key points' location and solving the problem of intra-class ambiguity. Finally, the upper limb postures is predicted according to the position relationships. The proposed method can accomplish mapping the upper limb posture between humans and manipulators, which could realize the human-machine cooperation of the dexterous hand manipulators. Experiment results demonstrate excellent performance with average detection speed of 33 FPS, average key point detection accuracy of 75.2%, and cooperative operation completion rate of 98%, could meet the practical requirement.

Keywords: man-machine collaboration; finger-YOLOv4; sparse target extraction; DFCRF-Net; dexterous hand manipulators

0 引言

目前面向人机协作的姿态识别算法主要研究近距离无遮挡识别问题^[1-3]。但是在机械臂协同控制任务中,操

作人员的上肢会相互遮挡,导致类内模糊、关键点定位不准确问题;此外来自其他人和物的类间多目标干扰,这也是造成人机协作任务失败的原因之一。

近年来,人机协作技术已成为研究和应用热点,很多研究人员针对人机协作系统中的人体姿态识别技术展开

研究。譬如基于穿戴式传感器获得人体运动参数, Yang等^[4]提出了一种基于惯性和磁测单元的姿态识别传感装置。通过使用磁体的磁偶极子模型和预先估计的相对方位进行姿态识别,但是该方法要求每个人体关节都需要配置一个磁测单元,不仅穿戴不适,而且造假昂贵。为了解决上述问题,研究人员将人体姿态识别技术的研究重心转移到计算机视觉领域,一些基于机器视觉的人体姿态识别算法被提出^[5-7]。Yang等^[8]提出了一个全连接的姿态识别网络,采用端到端的估计框架,避免了繁琐的匹配算法,因此具有一定的可移植性,但是该方法会误检非操作人员动作信息,造成人机协作任务失败。目前开源的姿态识别算法,譬如OpenPose、Mediapipe等能够识别出人体关键点的坐标信息^[9-11]。通过逻辑关系判断也能识别姿态,但实时性及鲁棒性不能满足实际应用需求。

国内学者也在人体姿态识别领域做了深层次的研究。蔡轶珩等^[12]在处理3D人体姿态估计的问题中提出一种基于自适应融合特征提取和误分类处理机制的改进算法,能有效去除部位误分点,并显著改善了3D人体姿态估计。本课题组多年来在姿态识别领域做了深入研究,冯文字等^[13]提出了一种CT-OpenPose模型检测人体姿势,使用带自适应软阈值残差网络、跨层连接机制和权值修剪的方法对传统模型在底层特征提取方式、底层特

征处理流程、模型训练和压缩方面进行改进,提高了关键部位被衣物遮挡时检测的准确性。此外朱洪堃等^[14]在传统OpenPose模型基础上,使用添加二阶项融合的残差网络提取底层特征,再通过L1范数权值对训练后的模型进行权值修剪,提出了改进型OpenPose模型。

本文提出了一种面向人机协作系统的上肢姿态精准识别算法,能够无接触地实时捕捉人体上肢姿态信息,有效排除类内遮挡干扰与类间多目标干扰,满足用于人机协作的姿态识别算法的实际需求。

1 上肢姿态精准识别算法

上肢姿态精准识别算法检测流程为上肢区域识别、上肢干扰排除、上肢关键点检测以及上肢姿态推理。

1.1 上肢区域识别

本文采用Finger-YOLOv4进行上肢区域识别。Finger-YOLOv4是在YOLOv4的基础上改进,相较原算法在识别人体上肢的任务中实时性和鲁棒性都大幅提高。该算法结构如图1所示,图中左侧虚线框区域为改进型MobileNetv3,较YOLOv4改进如下:

- 1) 采用改进型MobileNetv3替换YOLOv4的主干网络CSPDarkNet53;
- 2) 将YOLOv4的Hand结构简化为2个分支。

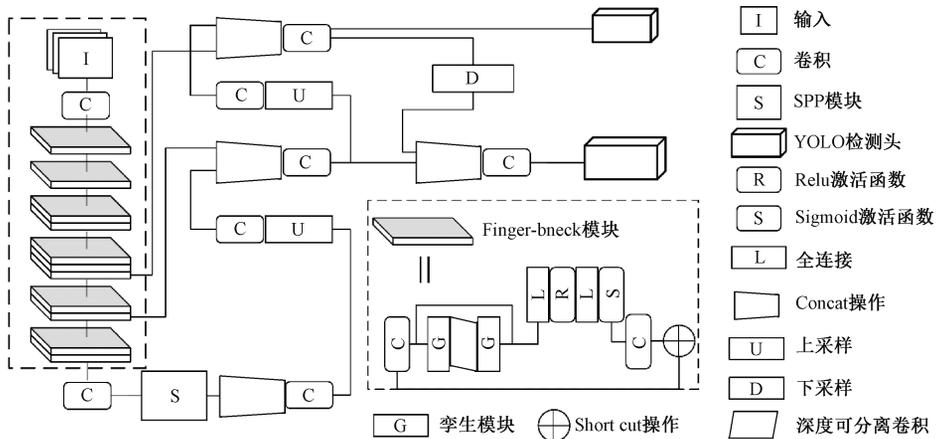


图1 Finger-YOLOv4 算法结构

Fig. 1 Finger-YOLOv4 algorithm architecture

MobileNetv3的改进方式是结合通道注意力机制^[15]、孪生模块^[16]、深度可分离卷积^[17]搭建Finger-bneck网络模块,用于替换MobileNetv3中的bneck。

通道注意力机制主要用于通道之间重要性调节。具体操作方式为首先对输入特征图进行轻量化操作,使用池化层获取通道的全局特征值;然后利用全连接层对全局特征值进行调节排序,学习各通道的权重;最后对原特征进行加权处理,能够有效减少无用计算量,提高检测速度的同时且不影响检测精度。

孪生模块是用来降低神经网络大小和计算资源占用的网络模块,相比于卷积操作能够使用更少的计算量获取同等的特征,定义整个卷积操作的计算量为 F ,计算方法如式(1)所示。

$$F = n \times h' \times w' \times c \times k \times k \quad (1)$$

式中: F 为卷积操作的计算量, n 是卷积操作的输出维度, c 是卷积操作的通道数, h' 是输出的高, w' 是输出的宽, k 是卷积操作中卷积核的高和宽。

孪生模块相对卷积操作进行了改进,第1步使用较卷积操作更少的卷积核生成特征图,第2步对第1步生成的每一张特征图进行深度卷积操作,第3步将特征图拼接到一起。定义原卷积核个数为 n ,第1步使用的卷积核个数为 m ,第2步中每张特征图生成 s 张新的特征图,保证 $m \times s = n$ 即保证孪生操作和普通卷积输出的特征相同;定义整个孪生操作的计算量为 F' ,计算方法如式(2)所示。

$$F' = (m \times h' \times w' \times c \times k \times k) + [(s-1) \times m \times h' \times w' \times d \times d] \quad (2)$$

式中: F' 为孪生操作的计算量, m 是孪生操作的输出维度, c 是通道数, h' 是输出的高, w' 是输出的宽, k 是卷积操作中卷积核的高和宽, d 是孪生操作中卷积核的高和宽;由式(1)和(2)可以看出输出特征相同情况下,卷积操作的计算量是孪生模块的 s 倍。

深度可分离卷积是将一个完整的卷积分两部分进行,一部分是逐点卷积,另外一部分是逐深度卷积,然后重新堆叠到一起,使用较小的卷积核来构建有效的卷积操作,其计算方法如式(3)所示。

$$F_d = n \times h' \times w' + c \times k \times k \quad (3)$$

式中: F_d 为深度可分离卷积所需参数量, n 是卷积操作的输出维度, c 是卷积操作的通道数, h' 是输出的高, w' 是输出的宽, k 是卷积核的高和宽。对比式(1)和(3),所有参数一致的情况下,卷积操作的计算量明显高于深度可分离卷积的计算量。

将 MobileNetv3 中的 bneck 替换为上文设计的 Finger-bneck 模块,进行一次卷积后以 Finger-bneck 作为基础模块,对其进行特征堆叠。将改进型 MobileNetv3 第7层、第9层 Finger-bneck 模块,分别与原 YOLOv4 颈部网络中前两个上采样后的卷积进行 Concat 连接。堆叠 11 层 Finger-bneck 后进行 1 次卷积,将特征与 SPP 附加模块提取的特征进行级联。

1.2 上肢区域干扰排除

本文采用稀疏性目标提取算法(sparse target extraction algorithm, STEA)排除非目标人员上肢区域。在图像数据中,运动频率高的像素区域组成的矩阵具有稀疏性,运动频率低的像素区域组成的矩阵具有低秩性。基于该理论本文提出 STEA 算法,将采集到的图像信息分成 H 、 S 、 V 三个通道分别处理。每个通道的数据可以分解为分别具有低秩性和稀疏性两部分。由于目标人员上肢相对非目标人员运动频率高,故其像素值组成的矩阵具有稀疏性,由此锁定操作人员上肢区域,具体操作步骤如下:

第1步分别构建 H 、 S 、 V 三通道的观测矩阵,定义观测矩阵为 D ,帧数为 f ,则 H 、 S 、 V 通道的观测矩阵分别为 $D^H = \{h_1, h_2, \dots, h_f\}$ 、 $D^S = \{s_1, s_2, \dots, s_f\}$ 、 $D^V = \{v_1, v_2, \dots, v_f\}$ 。

基于稀疏性矩阵分解理论将观测矩阵分解为前景矩阵和背景矩阵^[18],定义具有稀疏性的前景矩阵为 A ,具有低秩性的背景矩阵为 B 。

第2步分别对 H 、 S 、 V 三通道的的前景矩阵 A 的 $1 \sim f$ 帧进行去向量化,得到 H 、 S 、 V 三通道的的前景图像序列帧分别为 $(ag_1^H, ag_2^H, \dots, ag_f^H)$ 、 $(ag_1^S, ag_2^S, \dots, ag_f^S)$ 、 $(ag_1^V, ag_2^V, \dots, ag_f^V)$ 。

第3步按帧组合 H 、 S 、 V 三通道的的前景部分得到的前景图像,第1帧前景图像为 $agimg_1 = \{ag_1^H, ag_1^S, ag_1^V\}$,以此类推第 f 帧前景图像为 $agimg_f = \{ag_f^H, ag_f^S, ag_f^V\}$ 。对前 f 帧前景图像进行求和处理,得到前景图像矩阵 $Agimg$,如式(4)所示。

$$Agimg = abs(agimg_f) \quad (4)$$

式中: $Agimg$ 表示对前 f 帧前景图像进行组合, $agimg_f$ 表示第 f 帧前景图像矩阵, $abs()$ 表示求和。

将 Finger-YOLOv4 框选的若干人体上肢区域中前景图像占比最大的一块区域定为操作人员上肢,进而锁定目标人员上肢区域。

1.3 上肢关键点检测

本文采用双特征条件随机场网络(dual features CRF-Net, DF-CRF-Net)定位人体上肢关键点,其网络结构如图2所示,使用不同特征提取器:图卷积神经网络(graph convolutional networks, GCN)与卷积神经网络(convolutional neural networks, CNN)共同提取目标人员上肢区域的关键点特征^[19-20]。通过条件随机场(conditional random fields, CRF)进行推理,解决遮挡导致的类内模糊问题,进一步优化关键点位置信息,提高人体姿态估计在遮挡下的鲁棒性。

使用卷积神经网络提取特征是基于残差网络提取的,定义使用图卷积神经网络提取的特征为向量 α^c ,残差网络在全连接层 FC 后提取的特征为向量 α^c ,图卷积神经网络中以及卷积神经网络中待推理的特征分别为向量 β^c 以及 β^c ,利用权重矩阵 W^{GC} 调节注意门(attention gate, AG),调节两种潜在特征 β^c 和 β^c 进入条件随机场的流量,权重矩阵 W^{GC} 计算方法如式(5)所示。

$$\begin{cases} W^{GC} = \sigma(-(\beta_{i-1}^c \otimes \beta_{i-1}^c)), & i \neq 0 \\ \beta_i^c = \alpha^c & i = 0 \\ \beta_i^c = \alpha^c, & \end{cases} \quad (5)$$

式中: W^{GC} 表示注意门控制两种潜在特征的信息流的权重矩阵, \otimes 表示元素相乘, σ 表示 sigmoid 函数归一化, β_i^c 为图卷积中待推理的特征, β_i^c 为卷积中待推理的特征, α^c 为卷积提取的特征, α^c 为图卷积提取的特征, β^c 和 β^c 的更新方法如式(6)、(7)所示。

$$\begin{cases} \beta_i^c = Conv^{GC} * \beta_{i-1}^c, & i \neq 0 \\ \beta_i^c = \alpha^c, & i = 0 \end{cases} \quad (6)$$

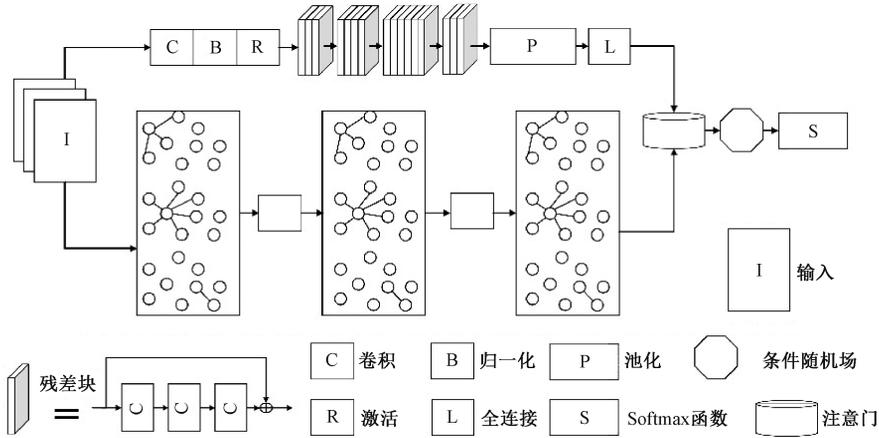


图2 双特征条件随机场网络结构

Fig. 2 DFCRF-Net architecture

$$\begin{cases} \beta_i^C = \beta_{i-1}^C \oplus (W^{GC} \otimes \beta_{i-1}^C), & i \neq 0 \\ \beta_i^C = \alpha^C, & i = 0 \end{cases} \quad (7)$$

式中: β_i^C 为图卷积中待推理的特征, β_i^C 为卷积中待推理的特征, α^C 为卷积提取的特征, α^C 为图卷积提取的特征, W^{GC} 表示注意力门控制两种潜在特征的信息流的权重矩阵, $Conv^{GC}$ 表示 1×1 的卷积核, $*$ 表示卷积操作, \otimes 表示元素相乘, \oplus 表示元素相加。

β_i^C 和 β_i^G 不断更新, 将该特征输入条件随机场 (conditional random field, CRF) 进行推理^[21], 当能量损失值最小时得到最优融合特征, 能量函数如式(8)所示。

$$E(\beta_i^C, \beta_i^G, X^F, I) = \sum_{i=0}^n \phi(\beta_i^C, \alpha^C) + \quad (8)$$

$$\sum_{i=0}^n \phi(\beta_i^C, \alpha^C) + \sum_{i=1}^n \varphi(\beta_i^C + \beta_i^G + W^{GC})$$

式中: E 表示条件随机场的能量函数的能量损失值, β_i^C 为图卷积中待推理的特征, β_i^G 为卷积中待推理的特征, X^F 为融合待推理卷积特征和图特征的最终可观测特征, I 为输入的原始图片, α^C 为卷积提取的特征, α^G 为图卷积提取的特征。最终将最优的融合特征输入 Softmax 分类层对人体上肢区域的关键点特征进行精准分类。

1.4 上肢姿态推理

上肢的姿态是基于获取的上肢关键点的 3 维坐标点的逻辑关系进行判断。具体分为 4 个部分: 手指关节姿态推理、手腕关节姿态判断、手肘关节姿态推理以及肩关节姿态推理。

以指关节为例, 其余关节原理相似。手指运动是各个指关节围绕着与它相连关节点的运动, 可由相邻关节间的夹角表示, 即关节的角度值。每个手指的第一关节 B 、第二关节 C 、第三关节 D 都有一个向掌心 O 旋转的自由度。如图 3 所示, 本文以右手食指为例进行基于空间坐标的角度计算。

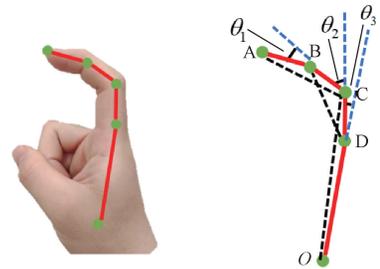


图3 食指姿态信息推理示意图

Fig. 3 Inference diagram of index finger posture information

根据上文已知 A 、 B 、 C 、 D 以及 O 的 3 维坐标, 来计算图 3 中 3 个三角形外角 $\angle \theta_1$ 、 $\angle \theta_2$ 和 $\angle \theta_3$ 的角度, 进而推理出该手指的运动状态。

1.5 协作方法

灵巧手机械臂是自主设计的, 通过 3D 打印机械结构, 其自由度上肢姿态识别算法能够捕捉的人体上肢的自由度是一一对应的。左肢各关节夹角示意如图 4 所示, 右肢同理。

由计算平台将上肢姿态识别算法推理出的各关节姿态信息转换为舵机和电机驱动信号, 通过 USB 转 TTL 模块将总线舵机的驱动信号传至通信板, 最后由通信板向总线舵机的控制器发送指令驱动舵机运动到相应位置; 通过 USB 转 CAN 模块将无刷电机的驱动信号传至双编码 CAN 通信无刷电机, 驱动电机转动到对应位置, 进而使得灵巧手机械臂完成协同动作。

2 实验分析

2.1 实验平台及环境

实验所用深度学习平台配置两块 NVIDIA RTX-3090 24 GB 显卡。其他信息如下: Ubuntu20.04、Torch1.7.0、CUDA11.1、Python3.6。

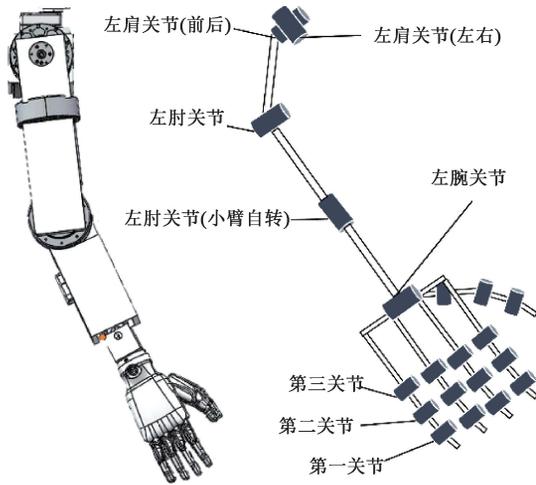


图 4 灵巧手机械臂各关节自由度示意图

Fig. 4 Freedom diagram of each joint of dexterous hand manipulators

2.2 数据集

1) 上肢目标检测数据集

上肢目标检测数据集为自制的上肢区域 (upper limb-region, ULB-R) 数据集,在复杂环境下采集不同类型的人体上肢图片共计 1 800 张,并制作对应标签文件。采用随机裁剪、镜像、翻转等方式对数据集进行增广^[22],最终得到 ULB-R 数据集,共计 12 000 张。其中 8 400 张作为训练集,3 600 张作为测试集。

2) 上肢关键点检测数据集

上肢关键点检测数据集分为两部分,第 1 部分选择开源数据集 COCO-WholeBody 数据集^[23]进行训练和评价,其中 70% 作为训练集,30% 作为测试集。第 2 部分是自制的上肢关键点 (upper limb-point, ULB-P) 数据集,ULB-P 数据集用于极端环境下关键点检测的实验,在身体部位相互遮挡、非操作人员身体干扰等极端环境采集数据,并自制数据集 2 000 张,通过数据增广的方式。最终得到 ULB-Point 数据集,共计 10 000 张,其中 7 000 张作为训练集,3 000 张作为测试集。

2.3 验证上肢区域识别算法改进有效性

如表 1 所示,测试了 YOLOv3、YOLOv4、MobileNetv3-YOLOv4、Finger-YOLOv4 这 4 种目标检测算法在 ULB-R 数据集的性能。其中 MobileNetv3-YOLOv4 是采用 MobileNetv3 替换 YOLOv4 的主干网络。通过对比可知采用 MobileNetv3 替换 YOLOv4 的主干网络 CSPDarknet53 模型检测速度明显高于 YOLOv3 和 YOLOv4,但检测精度有所下降,本文提出的 Finger-YOLOv4 在检测精度、检测速度以及模型体积各方面都优于 YOLOv3 和 YOLOv4,检测精度和模型体积优于 MobileNetv3-YOLOv4,检测速度与 MobileNetv3-YOLOv4 基本一致,综上所述,本文提出的 Finger-YOLOv4 对上肢区域的检测效果最优。

表 1 目标检测算法对比实验

Table 1 Comparison experiment of the target detection algorithm

主干模型	输入尺寸	mAP /%	帧率 /FPS	模型体积 /MB
YOLOv3	416×416	89.4	62	234
YOLOv4	416×416	91.67	66	245
MobileNetv3-YOLOv4	416×416	90.42	78	221.9
Finger-YOLOv4	416×416	95.33	78	218.4

2.4 验证模型对上肢区域关键点定位的性能

如表 2 所示,是分别在 COCO-WholeBody 数据集 (正常环境) 和 ULB-P 数据集 (极端环境) 上不同模型对上肢区域 48 个关键点定位性能的对比实验。分别选择 Mediapipe、OpenPose、HMS ML kit^[24]、AlphaPose^[25] 与本文的上肢姿态识别算法进行对比。由于本文只检测人体上肢区域,因此本实验中只考虑上肢区域对应的关键点定位的准确度。OpenPose 姿态识别算法检测速度偏慢,不符合实际需求,Mediapipe 在检测速度上优于所有模型,但是在遮挡、干扰等极端环境下,AP 的平均值只有 49.3%,稳

表 2 上肢关键点检测算法对比实验

Table 2 Comparison experiment of the key point detection algorithm of arm

模型	COCO-WholeBody 数据集 (正常环境)				ULB-P 数据集 (极端环境)			
	AP/%	AP50/%	AP75/%	帧率/FPS	AP/%	AP50/%	AP75/%	帧率/FPS
Mediapipe	58.4	82.3	63.2	52	49.3	79.7	59.9	52
OpenPose	62.3	84.6	68.1	6	61.0	82.4	65.2	6
HMS ML kit	67.0	88.4	73.1	31	63.1	82.7	66.3	31
AlphaPose	73.6	89.3	79.1	20	70.1	84.3	72.6	20
本文方法	75.2	91.2	81.3	33	74.1	90.5	79.4	33

定性差。本文方法对关键点定位的准确度优于上述模型,且检测速度达到 33 FPS,满足实时检测的需求。

2.5 验证模型处理遮挡问题的能力

为了进一步验证本文算法在处理遮挡问题的能力,采用不同方法进行遮挡场景验证,如图 5 所示。常见的遮挡主要分为两种,待检测的目标之间相互遮挡的类内遮挡、待检测目标被干扰物遮挡的类间遮挡。

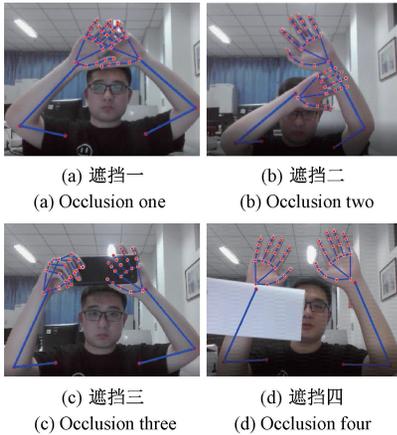


图 5 遮挡实验效果

Fig. 5 Occlusion experiment effect

图 5 中(a)、(b)是类内遮挡实验,分别验证了在手掌对手掌的遮挡、手掌对手臂的遮挡的情况下,算法对上肢关键点定位的准确性。结果显示在类内遮挡的环境下,本文算法可以较好的定位出上肢关键点。

图 5 中(c)、(d)是类间遮挡实验,选择了黑色外物和白色外物分别对手掌和手臂进行遮挡,根据实验结果可以看出在类间遮挡的环境下,算法依然可以较好的定位出上肢关键点。

综上所述,本文提出的上肢姿态识别算法能够有效解决遮挡问题,增强特殊环境下关键点定位准确性。

2.6 在灵巧手机械臂上的人机协作性能测试

人机协作实验如图 6 所示,使用深度相机采集操作人员上肢动作,通过上肢姿态识别算法获取当前帧的人体上肢的 3 维坐标,通过姿态推理获得关节的角度信息,与灵巧手机械臂驱动关节进行映射,使灵巧手机械臂模仿人上肢的动作,进而验证面向人机协作系统的姿态识别算法的有效性。

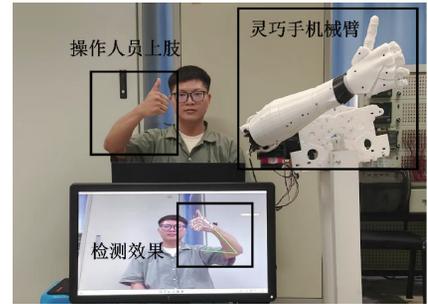


图 6 人机协作实验效果

Fig. 6 Experimental effect of man-machine cooperation

如表 3 所示,是选取 10 个不同受测人员使用不同的姿态识别算法对灵巧手机械臂人机协作的测试实验。分别选择背景较为简单且没有其他人员干扰的纯净场景和有人员流动的复杂场景进行测试。测试动作为每人随机做出 20 个不同的动作。协作完成度是指灵巧手机械臂做的动作和人做的动作相同的次数与整体实验次数的比值;平均关节夹角相似度是指灵巧手机械臂各关节角度与人的上肢各关节角度的平均相似度,协作响应时间是指算法向灵巧手机械臂发送指令间隔的时间。使用 OpenPose 捕捉姿态信息,协作响应时间偏长,影响人机协作的流畅性,不能满足实际的需求,且有人员流动时直接误检多人,无法判断真正的操作人员是谁,会直接造成检测任务的失败。

表 3 姿态识别模型在灵巧手机械臂上人机协作性能测试

Table 3 Human machine cooperation performance test of attitude recognition model on dexterous mobile arm

方法	纯净场景			有人员流动的复杂场景		
	平均夹角相似度/%	完成度/%	协作响应时间/s	平均夹角相似度/%	完成度/%	协作响应时间/s
Mediapipe	46	80	0.025	40	73	0.025
OpenPose	68	95	0.166	-	-	-
HMS ML kit	61	93	0.032	57	85	0.032
AlphaPose	70	97	0.050	-	-	-
本文方法	72	98	0.031	67	93	0.031

Mediapipe 在纯净场景的人机协作的任务中,协作响应时间只需要 0.025 s,但是协作完成度只有 80%,相比

OpenPose 下降了 15%,且平均关节夹角相似度只有 46%。但是该算法能够进行单人检测,可以锁定真正的

操作人员,在有人员流动的场景下也能完成人机协作的任务,实际应用效果优于 OpenPose; HMS ML kit 在纯净场景下协同完成度和关节夹角相似度的指标高于 Mediapipe,但是低于 OpenPose,且在在有人员流动的复杂场景下,协作完成度相比 Mediapipe 提升了 12%; AlphaPose 在纯净场景下协同完成度和关节夹角相似度的指标高于上述 3 种模型,但是协作响应时间是 Mediapipe 的 2 倍,耗时较长,而且在有人员流动的复杂场景下会误检多人,无法判断真正的操作人员是谁,会直接造成检测任务的失败,不满足本文的实际应用需求。而本文提出的上肢姿态识别算法,只检测对应的上肢的姿态信息,无论是协作完成度还是平均关节夹角相似度指标都优于上述 4 种模型,且协作响应时间只有 0.031 s,也能够流畅的完成人机协作任务,特别是在有人员流动的复杂场景下,相比上述模型中效果最好的 HMS ML kit 协同完成度提升了 8%。

3 结 论

本文提出了面向人机协作系统的上肢姿态精准识别算法。针对 YOLOv4 算法在识别上肢的环境下进行改进,提出 Finger-YOLOv4 算法,相比于 YOLOv4 算法在识别上肢的任务中检测速度更快、检测精度更高;采用先框选上肢区域再定位上肢关键点的方法,解决了传统姿态识别需要捕捉全局姿态信息下才能精准检测的问题,可以实现局部姿态的精准检测,大幅度提高了上肢关键点定位的准确率;采用不同特征提取器—图卷积神经网络与卷积神经网络提取操作人员上肢区域 48 个关键位置特征,并注意门通过条件随机场进行推理,解决遮挡导致的类内模糊问题,大幅度提高了上肢关键点定位的准确率;在 Finger-YOLOv4 算法框选上肢区域的基础上引入稀疏性目标提取算法,可以有效排除非操作人员上肢干扰,锁定操作人员上肢区域,增强整体算法的鲁棒性。

参考文献

[1] MILOJEVIC A, LINB S, COJBASIC Z, et al. A novel simple, adaptive, and versatile soft-robotic compliant two-finger gripper with an inherently gentle touch[J]. *Journal of Mechanisms and Robotics*, 2021, 13(1): 1-14.

[2] 刘今越,刘彦开,贾晓辉,等. 基于模型约束的人体姿态视觉识别算法研究[J]. *仪器仪表学报*, 2020, 41(4):208-217.

LIU J Y, LIU Y K, JIA X H, et al. Research on human pose visual recognition algorithm based on model constraints[J] *Chinese Journal of Scientific Instrument*, 2020,41(4):208-217.

[3] 钱银中,沈一帆. 姿态特征与深度特征在图像动作识别中的混合应用[J]. *自动化学报*, 2019, 45(3): 626-636.

QIAN Y ZH, SHEN Y F. Hybrid of pose feature and depth feature for action recognition in static image[J]. *Acta Automatica Sinica*, 2019,45(3):626-636.

[4] YANG Z, YAN S, VAN BEIJNUM B J F, et al. Hand-finger pose estimation using inertial sensors, magnetic sensors and a magnet[J]. *Sensors Journal*, 2021, 21(16): 18115-18122.

[5] KALAITZAKIS M, CAIN B, CARROLL S, et al. Fiducial markers for pose estimation[J]. *Journal of Intelligent & Robotic Systems*, 2021, 101(4): 1-26.

[6] NEEDHAM L, EVANS M, COSKER D P, et al. The accuracy of several pose estimation methods for 3D joint centre localisation[J]. *Scientific Reports*, 2021, 11(1): 1-11.

[7] SHARMA A, SHARMA N, SAXENA Y, et al. Benchmarking deep neural network approaches for Indian sign language recognition[J]. *Neural Computing and Applications*, 2021, 33(12): 6685-6696.

[8] YANG J, XI M, JIANG B, et al. FADN: Fully connected attitude detection network based on industrial video[J]. *IEEE Transactions on Industrial Informatics*, 2020, 17(3): 2011-2020.

[9] SUH Y, WANG J, TANG S, et al. Part-aligned bilinear representations for person re-identification[C]. *ECCV*, 2018: 402-419.

[10] SUBRAMANIAN B, OLIMOV B, NAIK S M, et al. An integrated mediapipe-optimized GRU model for indian sign language recognition[J]. *Scientific Reports*, 2022, 12(1): 1-16.

[11] HEATH C D C, VENKATESWARA H, PANCHANATHAN S. Are you paying attention? Classifying attention in pivotal response treatment videos[C]. *CVPR*, 2019: 1-9.

[12] 蔡轶珩,王雪艳,马杰,等. 基于随机森林误分类处理的 3D 人体姿态估计[J]. *自动化学报*, 2020, 46(7): 1457-1466.

CAI Y H, WANG X Y, MA J, et al. 3D human pose estimation based on random forest misclassification processing mechanism [J]. *Acta Automatica Sinica*, 2020,46(7):1457-1466.

[13] 冯文宇,朱洪堃,殷佳炜,等. 无人 CT 智能姿态识别算法研究[J]. *仪器仪表学报*, 2020, 41(8):188-195.

FENG W Y, ZHU H K, YIN J W, et al. Research on intelligent posture recognition algorithm of unmanned CT[J]. *Chinese Journal of Scientific Instrument*, 2020,

- 41(8):188-195.
- [14] 朱洪堃,殷佳炜,冯文字,等.一种轻量化实时人体姿势检测模型研究与应用[J].系统仿真学报,2020,32(11):2155-2165.
ZHU H K, YIN J W, FENG W Y, et al. Research and application of a lightweight real-time human posture detection model[J]. Journal of System Simulation, 2020, 32(11):2155-2165.
- [15] NIU Z, ZHONG G, YU H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
- [16] PAOLETTI M E, HAUT J M, PEREIRA N S, et al. Ghostnet for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10378-10393.
- [17] KULKARNI U, MEENA S M, GURLAHOSUR S V, et al. Quantization friendly mobilenet (qf-mobilenet) architecture for vision based applications on embedded platforms[J]. Neural Networks, 2021, 136: 28-39.
- [18] XU Y, DU B, ZHANG L, et al. A low-rank and sparse matrix decomposition-based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(7): 1248-1252.
- [19] CHEN T, ZHANG X, YOU M, et al. A GNN-based supervised learning framework for resource allocation in wireless IoT networks [J]. IEEE Internet of Things Journal, 2021, 9(3): 1712-1724.
- [20] LOISEAU N, KECHICHIAN R, VALETTE S, et al. Learning 3D medical image keypoint descriptors with the triplet loss [J]. International Journal of Computer Assisted Radiology and Surgery, 2022, 17(1): 141-146.
- [21] LIU S, HE T, DAI J. A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese[J]. Mobile Networks and Applications, 2021, 26(5): 1891-1903.
- [22] SHENG J, LUMIN X, JIN X, et al. Whole-body human pose estimation in the wild[C]. ECCV, 2020: 196-214.
- [23] LV F, LI Y, LU F. Attention guided low-light image enhancement with a large scale low-light simulation dataset[J]. International Journal of Computer Vision, 2021, 129(7): 2175-2193.
- [24] 杨涛,杨博雄,尹萍,等.基于高性能嵌入式 AI 计算平台的人机交互手势控制识别研究[J].信息记录材料,2019,20(11):175-177.
YANG T, YANG B X, YIN P, et al. Research on human-computer interaction gesture control recognition based on high-performance embedded AI computing platform [J] Information Recording Materials, 2019, 20(11):175-177.
- [25] FANG H S, XIE S, TAI Y W, et al. Rmpe: Regional multi-person pose estimation [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334-2343.

作者简介



张堃,2005 年于江苏大学获得学士学位,2010 年于浙江大学获得硕士学位,2016 年于上海大学获工学博士学位,2017~2018 年在英国贝尔法斯特女王大学从事博士后研究。现为南通大学教授。主要研究方向为智能控制、人工智能、机器视觉。

E-mail: zhangkun_nt@163.com

Zhang Kun received his B. Sc. degree from Jiangsu University in 2005, received his M. Sc. degree from Zhejiang University in 2010, and received his Ph. D. degree from Shanghai University in 2016. He was a post-doctoral at Queen's University Belfast from 2017 to 2018. He is currently a professor at Nantong University. His main research interests include intelligent control, artificial intelligent, and machine vision.



华亮(通信作者),2001 年于南通工学院获得学士学位,2007 年和 2014 年于浙江工业大学分别获得硕士学位和博士学位,现为南通大学教务处处长、教授,主要研究方向为机器人及控制、图像处理与模式识别。

E-mail: hualiang@ntu.edu.cn

Hua Liang (Corresponding author) received his B. Sc. degree from Nantong Instituted Technology in 2001, received his M. Sc. degree and Ph. D. degree both from Zhejiang University of Technology in 2007 and 2014, respectively. He is currently the director of the academic affairs office and a professor at Nantong University. His main research interest include robot control, machine vision, and pattern recognition.