DOI: 10.19650/j.cnki.cjsi.J2513983

面向复杂工业场景的人体姿态估计性能增强方法*

李帆雅[1],张泽辉[1],陈博洋[2],徐晓滨[1],管 聪[3]
(1.杭州电子科技大学自动化学院 杭州 310018; 2.宁夏石化银骏安全技术咨询有限公司 银川 750000;
3.武汉理工大学船海与能源动力工程学院 武汉 430070)

摘 要:人体姿态估计是工业制造 5.0 的重要支撑技术之一,已经在行为识别、人机交互、数字孪生等多种场景展开了应用。然而,在复杂工业场景下,告示牌、管线、立柱等物品极易对作业人员形成局部或全局遮挡,导致人体姿态估计模型在关键点定位时出现偏差,造成模型对姿态估计的准确率降低。针对该问题,提出了一种面向复杂工业场景的人体姿态估计性能增强方法,该方法首先基于量子化自编码器对人体关键点进行结构化建模,将关节点特征映射到量子化隐空间,以提升人体部分遮挡时姿态估计的准确率。然后,针对工人遮挡数据集构建困难的问题,创新地提出了一种面向人体姿态遮挡的动态数据增强训练方法,在模型训练过程中,通过评估人体姿态估计模型在数据集上对各关键点的估计结果,使用工业场景真实遮挡物动态生成符合工业场景特征的工人遮挡图片用于下一次模型训练,进一步提升模型在人体姿态估计任务中的鲁棒性。实验结果证明,所提的方法在自建数据集上相比先进方法 PCT 平均准确率 AP 和平均召回率 AR 分别提升了 3.8% 和 2.7%,其能够有效地应对复杂工业场景中的作业人员人体遮挡问题。

关键词:人体姿态估计;工业遮挡;数据增强;计算机视觉
中图分类号:TP391 TH86 **文献标识码**:A **国家标准学科分类代码**:510.4050

Human pose estimation performance enhancement method for complex industrial scenes

Li Fanya[1], Zhang Zehui[1], Chen Boyang[2], Xu Xiaobin[1], Guan Cong[3]
(*1. School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China; 2. Ningxia Petrochemical Yinjun Safety Technology Consulting Co., Ltd., Yinchuan 750000, China; 3. School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan 430070, China*)

Abstract: Human pose estimation is one of the important supporting technologies for Industrial Manufacturing 5.0, which has already been applied in various scenarios such as action recognition, human-computer interaction, and digital twin. However, in complex industrial scenes, objects such as notice boards, pipes, and columns can easily cause local or global occlusions for workers, leading to errors in joint points localization by human pose estimation models and a decrease in the performance of the human pose estimation model. To address this problem, this article proposes a human pose estimation performance enhancement method for complex industrial scenes, which firstly structurally models the key points of the human body based on VQ-VAE model, mapping joint features to a quantized latent space to improve the accuracy of human pose estimation when occlusion occurred. Then, to address the problem of insufficient worker occlusion dataset, a dynamic data augmentation and training method is innovatively proposed. In the process of model training, industrial scene-specific worker occlusion images are generated dynamically using real industrial scene occlusion objects by evaluating the human pose estimation results of the model for the next model training, further enhancing the model's robustness in human pose estimation tasks. The experimental results show that the method proposed in this article achieves an average precision (AP) improvement of 3.8% and an average recall (AR) improvement of 2.7% over the advanced method PCT on the self-constructed dataset

收稿日期:2025-04-28 Received Date: 2025-04-28
*基金项目:浙江省自然科学基金(LTGG24F030004)、国家水运安全工程技术研究中心开放基金(A202403)、国家重点研发计划(2022YFE0210700)、国家自然科学基金(52401376)、浙江省科协青年人才托举培养(ZJSKXQT2015049)项目资助

and is able to effectively cope with the human occlusion problem in complex industrial scenes.

Keywords: human pose estimation; industrial occlusion; data augmentation; computer vision

0 引 言

工业 5.0 正在指引着智能制造系统向“人本中心”范式的演进。区别于工业 4.0 的技术驱动模式,工业 5.0 以“人”为核心,优先关注个人的自我实现和个性化需求[1]。通过整合“人-社会-自然-技术”的互动关系,采用价值驱动的方法来促进技术创新,使技术主动适应、观察并学习人类行为,旨在服务于人类需求和社会福祉,从人的需求出发推动技术革新[2]。这要求生产系统能够动态感知作业人员的生理状态与行为意图,从而推动人机协作从物理层面向认知层面延伸。在此背景下,人体姿态估计(human pose estimation, HPE)技术作为实现“人-机-环境”深度交互的基础技术,在工业行为识别、数字孪生建模、人机交互优化等场景具有广泛应用空间[3]。

人体姿态估计通过准确定位人体头部、躯干与四肢等关键部位,分析目标对象的动作语义。Liu 等[4]提出一种基于最小关节自由度模型的改进骨架特征,用于快速准确地对三维人体活动进行识别,以提升工业应用中人体姿态估计的算法效率和实时性。Liu 等[5]提出轻量级骨骼生成网络,通过分析已知的骨骼序列来预测未知的骨骼数据,以实现对即将发生的人类活动的早期预警。人体姿态估计技术在数字孪生建模领域发挥着重要作用。鲍劲松等[6]提出了面向人-机-环境共融的数字孪生协同技术,在人机协同建模层面通过 Openpose 算法提取每帧图像中的人体骨架信息,并采用长短期记忆网络识别动作序列特征,以准确识别人类的操作行为以及动作意图。此外,在人机交互优化领域,Yang 等[7]提出自适应树分解图卷积网络(adaptive tree-decomposition graph convolutional network, ATD-GCN),通过引入新的骨架图结构和特征设计方法,提高模型对复杂活动的识别能力,为智能制造业中的人机协作提供技术支撑。禹鑫燚等[8]融合多相机下关键点检测信息,对人体运动学姿态进行优化估计,实现动作跟随、物品传递、主动避障等人机交互任务。

然而,在真实场景应用时,遮挡会导致手腕、肘等人体关节的错误识别[9]。特别在工业场景中,由于环境的动态性和复杂性,遮挡问题更加突出,工人易受到告示牌、管线、立柱等物品的遮挡,导致人体姿态估计模型对被遮挡部位的识别严重失效。这种遮挡引发的姿态估计精度显著下降的问题,直接制约了人机交互的流畅性、行为识别的准确性及数字孪生模型的保真度。并且,遮挡会造成关键点真实标注的缺失,构建针对遮挡场景的数据集面临显著挑战。因此,如何增强人体姿态估计性能,以有效克服复杂遮挡干扰,确保其在真实工业场景下获得稳定、可靠的人体运动信息亟待研究。

鉴于此,本研究提出一种面向复杂工业场景的人体姿态估计性能增强方法,以提高人体姿态估计模型在遮挡下的鲁棒性。本研究的主要贡献为:

1) 提出了一种面向人体姿态遮挡的动态数据增强方法,重点解决工人在复杂场景下的姿态遮挡数据集构建困难的问题,使用真实工业场景遮挡物构建图像数据增强数据集,并在训练过程中评估姿态估计模型的输出结果,动态调整人体姿态图片关键点遮挡率,提升遮挡状态下的人体姿态估计准确度;

2) 提出了用于人体姿态估计模型性能提升的多轮迭代微调方法,应用迁移学习和早停机制,基于动态数据增强训练集不断迭代训练,在增强人体姿态估计模型在遮挡因素下识别准确率的同时有效防止了模型过拟合。

3) 构建了工业场景下作业人员姿态识别图像数据集,并在自建数据集上对本文所提方法进行实验。实验结果证明,本研究所提出的方法在自建数据集上相比先进方法平均准确率(average precision, AP)和平均召回率(average recall, AR)分别提升了 3.8% 和 2.7%,说明其能够有效地应对复杂工业场景中的人体遮挡问题。

1 相关工作

1.1 图像数据增强

图像数据增强是一种通过传统图像处理和深度学习技术,对原始图像进行变换和扩增,以提升图像特征表达质量和扩大数据集规模的方法。按其实现的原理大体可分为两类:基于基本图像操作的图像数据增强方法和基于深度学习的图像数据增强方法[10]。

基于基本图像操作的图像数据增强方法通过低计算代价的图像操作来生成数据,有基于几何变换(如旋转、平移、缩放等)的数据增强方法和基于像素级图像变换(如对比度增强、亮度增强、颜色增强等)的数据增强方法[11]。然而,这些方法能产生的数据量有限,难以生成具有多样性的数据。随着研究的深入,基于图像擦除和图像混合的确定式图像数据方法被提出。基于图像擦除的方法通过对图像中的部分区域进行像素擦除实现数据增强,用于模拟遮挡、缺失信息等情况。基于图像混合的方法通过将多个图像混合到一张图像中生成新数据,以使模型学习到更多关于不同图像组合的特征[12]。

为进一步增强数据的多样性,基于深度学习的图像数据增强方法被提出,分为自动增强类、特征增强类和深度生成式方法类[13]。其中,自动增强类[14-15]通过强化学习智能搜索最优的数据增强策略和组合方式,从而生成丰富多样的样本。特征增强类[16-17]直接在特征层面进行数据变换,相比传统图像空间增强能有效降低噪声干扰。此外,深度生成式方法类[18-19]借助生成对抗网络等深度生成模型自动化地生成新数据。

总体而言,基于基本图像操作的图像数据增强方法通过进行基本图像操作实现,计算代价小但数据增强的效果受限。而基于深度学习的图像数据增强方法性能较基于基本图像操作的方法有所提升,但所需的计算资源较大。基于此,本研究提出一种面向复杂工业场景人体姿态估计的动态数据增强方法。该方法通过图像融合技术,在关键点层级使用真实工业场景遮挡物进行遮挡,以较少的计算资源完成工人遮挡数据集的构建。同时,基于模型的姿态估计结果动态生成遮挡图片用于下一轮训练,增强模型对真实遮挡情况的特征重建能力,从而提高人体关键点识别的准确性。

1.2　人体姿态估计

人体姿态估计旨在从图像中估计身体关节的位置并预测人体关键点坐标,其能够为计算机视觉的多个下游任务提供支持,比如人体解析和动作识别等[20]。根据对关节点的建模方式,人体姿态估计可分为自顶向下的方法[21-23]和自底向上[24-26]的方法。自顶向下的方法采用分而治之的策略,其核心思想是先检测人体的位置,再对每个检测到的人体分别估计关键点。自底向上的方法与自顶向下的方法相反,先检测图像中所有的关键点,再将关键点聚类为不同个体的骨架。

如何在遮挡情况下准确定位关节点是当前人体姿态估计领域的研究重点,为提升模型在面对遮挡因素时的鲁棒性,研究者提出了多种解决方案。Wang 等[27]结合单目相机和稀疏惯性传感器数据,利用部分特定的跨模态融合机制和多尺度时间模块,解决了人机协作场景中人体姿态估计的遮挡问题。代钦等[28]提出一种基于部位遮挡级别的可形变姿态估计方法,通过定义遮挡级别、建立遮挡级别的部位检测器和部位间形变模型,有效减少了有遮挡情况下的部位误匹配问题。Jiang 等[29]提出一种双通道人体姿态估计网络,通过增强遮挡区域特征和补偿遮挡特征来提高人体姿态估计在遮挡环境下的准确性。然而,现有的人体姿态估计方法对每个关节点的处理通常是独立的,这忽略了关节之间的相互依赖关系。这种处理方式在遇到遮挡时,往往导致不合理的姿态预测结果。

针对复杂工业场景中人体姿态估计模型所面临的遮挡问题挑战,本文使用量子化自编码器结构(vector quantised-variational auto encoder, VQ-VAE)[30]学习姿态特征编码,设计基于量子化自编码器的人体姿态估计模型,将姿态估计任务转化为预测量子化密码本中特征的分类问题,用以学习关节点之间的依赖关系,缓解人体姿态估计模型遇遮挡导致的性能下降问题。

2　人体姿态估计方法

2.1　方法框架

所提出的面向复杂工业场景的人体姿态估计性能增强方法包括:基于量子化自编码器的人体姿态检测方法、动态图像数据增强方法以及对应的多轮迭代训练方法,具体如图 1 所示。

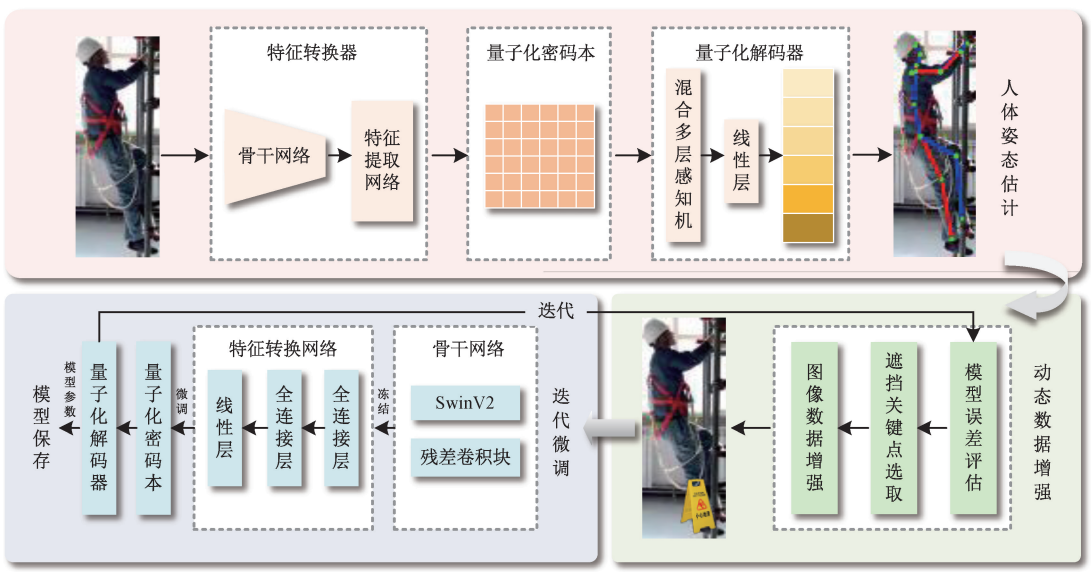


图 1　面向复杂工业场景的人体姿态估计性能增强方法框架

Fig. 1　The framework for human pose estimation performance enhancement method for complex industrial scenes

首先,本研究使用基于量子化自编码器的人体姿态估计模型检测作业人员的人体关键点坐标。然后,依据人体姿态估计模型得出的检测结果,利用真实工业场景中的遮挡物,采用图像融合方法进行动态图像数据增强。特别地,本研究提出一种适用于模型训练的多轮迭代性能增强方法,进一步提升模型在人体姿态估计任务中的准确性和可靠性。

2.2 基于量子化自编码器的人体姿态检测

针对工人作业过程中易受到机械设备、管线、告示牌等物品的遮挡问题,根据文献[31]提出的方法,构建基于量子化自编码器的人体姿态估计模型,将人体姿态表示为 M 个离散的 tokens,而每个 token 表征由几个相互关联的关节组成的一个子结构,以此降低重建误差,解决身体在复杂的工业环境中被遮挡的问题。

第 1 阶段,学习一个组合编码器 $F_e(G)$ 将原始人体姿态 G 转换成 M 个 tokens 特征,每个 tokens 特征 t_i 对应于姿态的一个子结构:

$$F_e(G)=(t_1,t_2,\cdots,t_M) \tag{1}$$

构建共享的密码本 $C=(c_1,\cdots,c_V)^{\mathrm{T}}$ 对所提取的 token 特征量化并记录,其中 V 是所述密码本记录的离散化特征条目数。

第 2 阶段,使用 SwinV2(swin Transformer V2)和 2 个残差卷积块构建一个转换器模块,从输入图像数据中提取人员作业姿态的 K 个 tokens 特征,并转换至密码本 C 所在的特征空间。进一步,利用构建的密码本 C 使用最近邻搜索量化每个 token 特征 t_i,将所述量化后的 token 特征送入解码器恢复原始姿态:

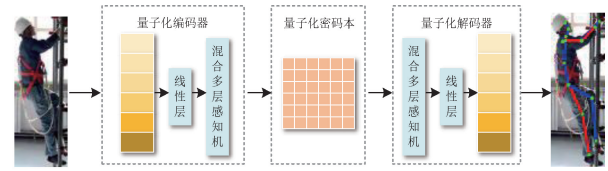
$$\hat{G}=F_d(c_{q(t_1)},c_{q(t_2)},L,c_{q(t_K)}) \tag{2}$$

式中:$q(t_i)$ 是第 i 个 token 在所述密码本中匹配到的最近特征索引;$F_d(\cdot)$ 是量子化解码器;利用量化后的 K 个 token 重构所述人体关键点坐标 $\hat{G}$。

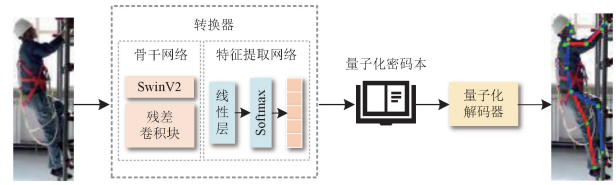
最终,输入待识别图像,将图像通过转换器生成 token 特征,将每个特征与密码本中的特征库做最邻近匹配,并把匹配结果输入到解码器中重构关键点坐标。基于量子化自编码器的人体姿态估计模型构建步骤如图 2 所示,这种结构化的表示方式允许模型在面对复杂场景时,能基于可见部分和视觉特征预测完整的人体姿态,将人体姿态约束在一个低维、紧凑和本质的表达空间,以提升作业人员人体关键点预测精度。

2.3 动态数据增强方法

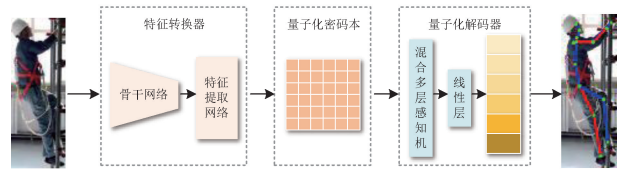
常用的轻量级数据增强方法,如:随机遮挡增强、单一色块增强等,与工业场景真实遮挡模式存在显著分布差异。为解决该问题,本研究提出一种基于真实工业场景的动态数据增强策略,通过图像融合技术在关键点层级使用真实工业场景遮挡物进行遮挡。



(a) 人体姿态估计模型训练第1阶段
(a) First stage of training for the human pose estimation model



(b) 人体姿态估计模型训练第2阶段
(b) Secondstage of training for the human pose estimation model



(c) 人体姿态估计模型训练第3阶段
(c) Third stage of training for the human pose estimation model

图 2 基于量子化自编码器的人体姿态估计模型

Fig. 2 The human pose estimation model based on VQ-VAE

1) 使用均方误差(mean squared error, MSE)对人体姿态估计模型对各个关节点的识别准确率进行评估,即:

$$Err_{i,j}=\frac{1}{2}\left[(x_{i,j}^{(1)}-x_{i,j}^{(2)})^2+(y_{i,j}^{(1)}-y_{i,j}^{(2)})^2\right] \tag{3}$$

式中:Err 表示 MSE 误差;i, j 表示第 i 张数据集图片的第 j 个人体关键点;$(x^{(1)},y^{(1)})$ 表示所述关键点在图片中的真实坐标;$(x^{(2)},y^{(2)})$ 为所述姿态估计模型预测出的关键点坐标。

2) 对数据集中所有关键点误差由小到大排序,设 $\alpha\%$ 是所提出的动态遮挡率,$\beta\%$ 是为防止模型误识别设定的固定遮挡率,每次数据集中遮挡关键点的选取方法为:获取前 $\alpha\%$ 最小误差关键点所在的集合 S_1,和后 $\beta\%$ 最大误差关键点所在的集合 S_2。为了增强人体姿态估计模型的防遮挡能力,避免网络朝着无遮挡方向收敛,本策略对集合 S_1 中的关键点进行遮挡;为了提高模型对人体的识别准确度,避免错误的对象识别,本策略同时对集合 S_2 中的所有关键点进行遮挡。遮挡关键点选取方案如图 3 所示 。

3) 为获取真实工业场景遮挡物,本研究通过视频监控系统采集工业场景图像,从采集的工业场景图像中获取具有代表性的遮挡物,并利用图像融合策略将所述代表性遮挡物融合至原始数据图像待被遮挡的关键点上:设 R 代表原始图像,O 代表遮挡物体图像,M 代表 O 的二进制掩膜(0 或 255)。

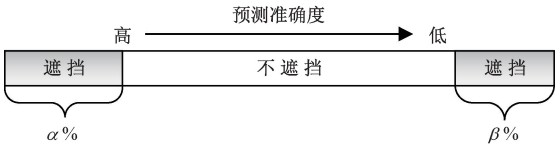


图 3　遮挡关键点选取方案

Fig. 3　Occlusion key point selection scheme

为将 O 融到 R 中,首先对 M 进行形态侵蚀得到 M_0,然后利用 M_0 得到掩膜 M 的物体边缘 $E=M-M_0$,接着将 E 所含的像素点设为 191,得到灰度掩膜 M_3(非对象:0,边:191,对象:255),最后用 M_3 除以 255 得到一个 0、0.75、1 的非整数掩膜 M_4,并利用 M_4 将 O 融入 R 中,R_o 表示生成的数据增强图像:

$$R_o=(1-M_4)\times R+M_4\times O \tag{4}$$

式中:R_o 表示生成的数据增强图像。使用该方法生成的图像数据增强示例如图 4 所示。

图 4　图像数据增强策略示例

Fig. 4　Example of image data augmentation strategy

假设对单个关键点进行遮挡的算法复杂度为 O_f,模型训练过程中数据集规模大小确定,可以设为 S。根据所提出的动态数据增强算法,被遮挡的关键点数量为 $k\cdot S$,其中 k 为确定整个数据集上被遮挡人体关键点数量的比例系数,与动态遮挡率 $\alpha\%$ 和固定遮挡率 $\beta\%$ 相关。所以,所提算法的时间复杂度为 $O(O_f\cdot k\cdot S)$。据分析,算法复杂度和数据集规模呈线性相关,能在较少计算资源下完成工人遮挡数据集的构建,增强模型在特定工业场景下的泛化能力,具体步骤见算法 1。

算法 1:动态数据增强方法

输入: 人体姿态估计模型 P, 数据集 D, 遮挡图集 O, 动态遮挡率 $\alpha\%$, 固定遮挡率 $\beta\%$

输出: 数据增强数据集 D_{aug}

1　　**1) 人体姿态估计模型误差评估:**

2　　$error_{dataset}\leftarrow$ CalMSE(P, D); //使用 MSE 误差对模型在数据集上对各关键点识别准确率进行评估

3　　**2) 遮挡关键点选取:**

4　　S_1, $S_2\leftarrow$ Sort($error_{dataset}$, $\alpha\%$, $\beta\%$); //对误差进行排序,获取前 $\alpha\%$ 误差最小关键点集合 S_1 和后 $\beta\%$ 误差最大关键点集合 S_2

5　　**3) 图像数据增强:**

6　　**for** *keypoint* **in** S_1, S_2 **do**

7　　R $\leftarrow$ Augment(*keypoint*, O); //使用遮挡图集中的遮挡物对关键点进行遮挡

8　　$D_{aug}\leftarrow$ Append(D_{aug}, R); //将生成的数据增强图像添加进数据增强数据集中

9　　**end for**

10　　**return** D_{aug};

2.4　模型训练方法

在人体关键点识别模型训练第 2 阶段,本研究迁移在 COCO2017(common objects in context 2017)数据集上预训练的 SwinV2 骨干网络参数至当前模型。同时,为增强模型在特定工业场景中的防遮挡能力,本研究通过人体姿态估计模型的识别结果生成增强数据集进行训练,实现遮挡区域潜在表示重建。

每一迭代轮次人体姿态估计模型的训练集为基于上一轮识别结果生成的增强数据集,模型的初始阶段基于第 $N-1$ 轮训练完成后的骨干网络,迁移源任务(第 $N-1$ 轮)的权重和偏置到本阶段(第 N 轮)人体姿态识别任务中。迁移学习中,只对特征转换网络部分参数进行更新,而模型的其他网络被冻结,即这些层的权重在训练过程中不被更新,迭代训练流程如图 5 所示。

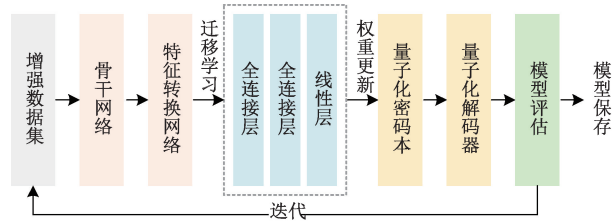


图 5　迭代训练方法流程

Fig. 5　Flowchart of the iterative training method

为了防止模型过拟合,在本训练方法中引入早停机制。设置标准平均准确率 AP 作为评估模型性能的指标,设置早停容忍度 p,若模型在连续 p 个 epoch 上 AP 值均没有提升,则提前终止本轮训练。此外,若该迭代轮次结束后,在验证集上人体姿态估计模型的性能不再提升或提升效果小于阈值 λ,则终止训练,得到最终的模型参数。进一步,考虑到随着训练轮次的推进,人体姿态估计模型识别和适应遮挡情况的能力会逐步增强的特点,本研究使用的动态遮挡率选取方法如图 6 所示。

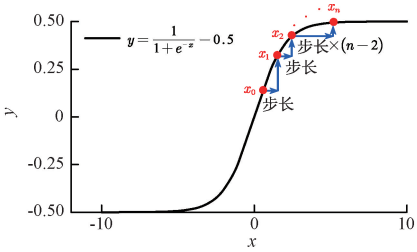


图 6 遮挡率选取方案

Fig. 6 Diagram of the occlusion rate selection scheme

具体来说,在模型的多轮迭代训练过程中,所提出的动态遮挡率 $\alpha\%$ 的选取方法为:在经过平移处理后零中心化的 Logistic 函数上按照固定的步长获取每一轮迭代训练集里的人体关键点遮挡概率。该方法能够平滑地模拟人体关键点的遮挡概率随训练迭代轮次的增加而逐渐增大的趋势,以逐步地扩大数据集上的人体关键点遮挡区域,循序渐进地增强模型在遮挡因素下的预测准确度。综合而言,所提出的模型多轮迭代训练算法的具体流程见算法 2。

算法 2:模型多轮动态迭代训练方法

输入: 训练集 D_{train}, 验证集 D_{val}, 大迭代轮次 N, 训练轮数 epoch, 预训练人体姿态估计模型 P_{coco}, 早停容忍度 p, 性能阈值 λ, 遮挡图集 O, 动态遮挡率选取起点自变量 x_0, 动态遮挡率选取步长 $step$, 固定遮挡率 $\beta\%$

输出: 最终模型 P_{final}

初始化:
$count \leftarrow 0$; //初始化早停计数器
$AP_{\text{best}} \leftarrow 0$; //初始化最优平均准确率
$P_0 \leftarrow P_{\text{coco}}$; //初始化人体姿态估计模型
$AugSet_{\text{train},0} \leftarrow D_{\text{train}}$; //初始化图像增强数据集
for $i = 1$ **to** N **do**
  while $count < p$ **do**
  for $j = 1$ **to** $epoch$ **do**
    $P_i \leftarrow \text{TransferLearning}(AugSet_{\text{train},i-1}, P_{i-1})$; //迁移参数并在数据增强训练集上训练
    $AP \leftarrow \text{Validate}(D_{\text{val}})$; //在验证集上验证模型
    if $AP < AP_{\text{best}}$ **then**
      $count \leftarrow count + 1$; //早停计数器增 1
    else
      $AP_{\text{best}} \leftarrow AP$; //更新最优平均准确率
      $count \leftarrow 0$; //归零早停计数器重新开始计数
    end if
  end for
end while
$promotion \leftarrow \text{Assess}(P_i, P_{i-1})$ //评估模型性能提升率
if $promotion > \lambda$ **then**
  $\alpha\% \leftarrow \text{Logistic}(x_0, step, i)$; //获取当轮动态遮挡率
  $AugSet_{\text{train},i} \leftarrow \text{DynamicAug}(P_{i-1}, D_{\text{train}}, O, \alpha\%, \beta\%)$; //使用算法 1 生成数据增强训练集
else
  $P_{\text{final}} \leftarrow P_i$;
  return P_{final};
 end if
end for
$P_{\text{final}} \leftarrow P_N$;
return P_{final};

3 数据集与实验评估

3.1 实验数据集

工业生产环境变化多样,作业人员的姿态识别相较日常姿态识别具更大的难度。目前,由于公共数据集与实际工业场景在环境特征、作业姿态等方面存在显著差异,基于公共数据集训练的姿态识别算法在工业场景中表现不佳,难以满足实际生产中的识别需求。

鉴于此,本研究针对复杂工业场景构建作业人员姿态识别图像数据集,在照明条件良好的情况下拍摄人员作业视频,并将视频抽取成图片,示例如图 7 所示。在完成数据采集后,参考 COCO 数据集格式,对原始图片中的人体关键点进行标注,包括鼻子、左眼、右眼、左耳、右耳、左肩、右肩、左肘、右肘、左手、右手、左髋、右髋、左膝、右膝、左脚和右脚,每个人体均标注 17 个关键点,每个关键点都对应一组精确的坐标信息。

图 7 采集的工业场景数据集

Fig. 7 The collected industrial scene dataset

最终,所构成的复杂工业场景作业人员姿态识别图像数据集共包含训练集图像 814 张、测试集图像 232 张。特别说明,采集该数据集所使用的拍摄设备为佳能 EOS 77D 和索尼 HDR-PJ50E,详细的设备参数如表 1 所列。

表 1　数据集拍摄设备参数表

Table 1　Parameter table of data set collecting equipment

参数类别	佳能 EOS 77D	索尼 HDR-PJ50E
传感器	2 420 万像素 APS-C CMOS	229 万像素 1/5.8 英寸 Exmor R CMOS
影像处理器	DIGIC 7	BIONZ
镜头	EF-S 18~55 mm f/4~5.6 IS STM	具体参数
视频录制	全高清 1 080 p@60 fps	全高清 1 080 p@60 fps

3.2　评估标准与训练测试

标准的评估指标由目标关键点相似度(object keypoint similarity, OKS)来定义,即:

$$OKS_p = \frac{\sum_i \exp(-d_i^2/2s^2k_i^2) \times \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \tag{5}$$

式中:p 为第 p 个人;i 为人体第 i 个关键点;d_i 为预测关键点和实际关键点的欧式距离;v_i 为标签是否可见的标志(v_i = 0 表示人体关键点存在遮挡且未标注;v_i = 1 表示人体关键点不存在遮挡且标注);s 代表第 p 个目标的边界框面积;k_i 为关键点的控制衰减常数;δ 为可见性。

本实验利用标准平均准确率 AP、OKS 阈值为 0.5 的标准平均准确率 $AP^{0.5}$ 和 OKS 阈值为 0.75 标准平均准确率 $AP^{0.75}$,以及标准平均召回率 AR、OKS 阈值为 0.5 的标准平均召回率 $AR^{0.5}$ 和 OKS 阈值为 0.75 标准平均召回率 $AR^{0.75}$ 作为姿态估计模型性能的评估指标。

所提出的动态数据增强方法使用 Python3.8 版本和 OpenCV-Python 4.7.0 实现,具体实验中该方法动态遮挡率起点自变量、迭代步长和固定遮挡率分别设为 0.1、0.1 和 0.005。模型多轮迭代微调训练的实验配置为:NVIDIA GeForce RTX 4080 SUPER,采用 Python3.8 版本和 PyTorch1.8.0, MMCV1.7.0, MMPose0.29.0 框架训练和测试模型。在训练阶段采用与 PCT[31](human pose as compositional tokens)方法相同的设置,在测试时采用自顶向下的方法,先用人体检测器检测人体对象,然后在人体检测器框出的范围内预测人体关键点。

3.3　实验验证

本研究的实验验证共分为 2 个部分,实验验证 1 中验证了在动态遮挡率(dynamic ratio)设置下,本研究所提方法较当前人体姿态估计主流模型的优越性;实验验证 2 中对比了固定遮挡率(fixed ratio)和动态遮挡率下的人体姿态估计结果,验证了动态遮挡率选取策略的有效性。

1) 实验验证 1

在本实验中,为验证所提出的动态数据增强方法对提升模型防遮挡能力的有效性,特别设计一种基于迭代过程中逐步增加关键点遮挡概率的动态遮挡率选取策略。为了验证所提出方法的优越性,选取当前人体姿态估计主流模型与本文提出的方法进行对比,所提出的面向复杂工业场景的人体姿态估计性能增强方法在自建工业场景数据集上和不同文献的方法在准确度的对比如表 2 所示,对比结果如图 8 所示。

表 2　不同方法在验证集上的实验结果

Table 2　The experimental results of different methods on the validation set　(%)

方法	骨干网络	*AP*	$AP^{0.5}$	$AP^{0.75}$	*AR*	$AR^{0.5}$	$AR^{0.75}$
IPR[32]	ResNet50	57.8	96.7	61.3	66.7	97.8	77.6
SimCC[33]	MobileNetV2	73.1	99.0	84.4	76.3	99.1	87.9
ViTPose[34]	ViT-Base	82.6	98.0	94.6	84.6	98.7	95.3
PCT[31]	Swin-Base	87.1	100.0	97.6	90.7	100.0	98.7
本文(dynamic ratio)	Swin-Base	90.9	100.0	100.0	93.4	100.0	100.0

本研究提出的方法平均准确率 AP 为 90.9%,优于其他人体姿态估计方法。具体地,和 IPR(integral human pose regression), SimCC(simple coordinate classification perspective for human pose estimation)和 ViTPose(simple vision transformer baselines for human pose estimation)方法相比,本研究提出的方法准确率分别提高了 33.1%, 17.8%和 4.7%。与同样使用 Swin-Base 为骨干网络的 PCT 方法相比,本研究提出的方法准确率提高了 3.8%。由此可见,所提出的面向复杂工业场景的人体姿态估计性能增强方法在应对工业遮挡因素上的优势。通过 $AP^{0.5}$ 和 $AP^{0.75}$ 的对比反映出本方法对于小尺度目标的准确率相比于其他网络也有明显提升,通过 AR, $AR^{0.5}$, $AR^{0.75}$ 的对比说明算法在对关键点定位的准确性上也具有显著提升。该模型的计算复杂度(giga floating-point operations per second, GFLOPs)和每秒可处理图像帧(frames per second, FPS)分别为 15.2 GFLOPs 和 115.1 fps,在图像处理与姿态估计上有较高实时性,对实际工业场景应用具较高价值。

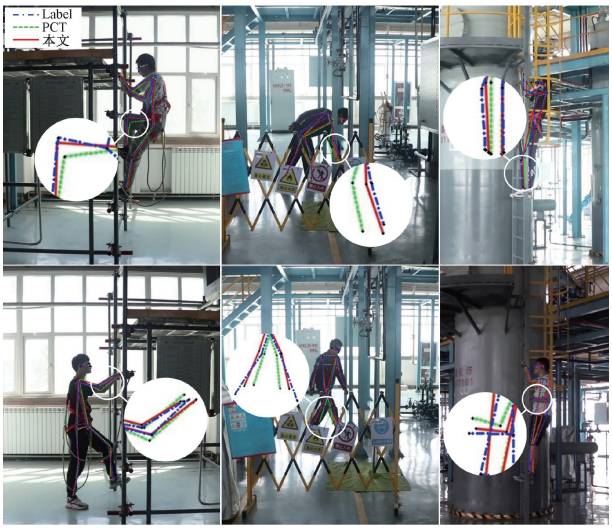


图 8 所提方法和 PCT 模型识别结果对比

Fig. 8 Comparison of the proposed method and the PCT model

2）实验验证 2

为验证本研究提出的动态遮挡率选取策略在提升姿态估计性能方面的有效性，该实验聚焦于探究固定遮挡率和动态遮挡率的使用对姿态估计模型准确度的影响，旨在明确动态遮挡率策略的独特优势。具体而言，设置 4 组固定遮挡率（0.05，0.10，0.15，0.20）作为对照组，与采用动态遮挡率的实验组进行对比，详细结果如表 3 所示，对比结果如图 9 所示。

表 3 本文所提方法不同遮挡率设置的实验结果

Table 3 The experimental results of different occlusion rate settings of the proposed method （%）

遮挡类型	遮挡率	AP	$AP^{0.5}$	$AP^{0.75}$	AR	$AR^{0.5}$	$AR^{0.75}$
固定遮挡	0.05	89.3	100.0	99.0	92.4	100.0	99.6
	0.10	90.6	100.0	100.0	93.1	100.0	100.0
	0.15	89.4	100.0	99.0	92.3	100.0	99.6
	0.20	87.1	100.0	97.6	90.8	100.0	98.7
动态遮挡		90.9	100.0	100.0	93.4	100.0	100.0

实验结果说明，与各种固定遮挡率设置相比，本研究提出的动态遮挡率选取策略对提升模型的预测准确度具有显著且积极的作用。此创新策略根据训练进程的迭代阶段，动态地、逐步地增加关键点被遮挡的概率，这种渐进式的难度递增机制，有效模拟了现实场景中可能遇到的不同程度遮挡，迫使模型在训练过程中持续学习如何应对越来越困难的遮挡情况。因此，模型在面对遮挡时的适应性和鲁棒性得到增强，从而在最终的预测任务中取得了更优的精度表现。

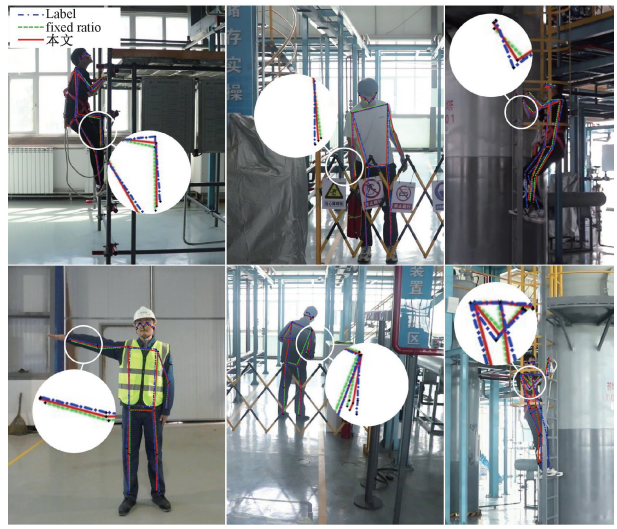


图 9 固定遮挡率（ratio=0.10）和动态遮挡率识别结果对比

Fig. 9 Comparison of fixed occlusion rate （ratio=0.10） and dynamic occlusion rate

3.4 讨 论

人体姿态估计的准确度受多重因素干扰，包括光照条件、相机参数畸变、背景复杂度等。而由于遮挡直接破坏了人体结构的视觉连续性，使模型失去推断姿态所依赖的物理先验知识，在该领域遮挡问题被认为是造成人体姿态估计准确度最大误差的原因之一。遮挡场景下，误差主要从 3 个方面被显著放大：被遮挡部位的视觉信息归零，模型依赖的局部纹理与轮廓特征彻底消失；生物力学连接关系被切断，导致单个关键点误差通过骨骼链传递，引发整体姿态的漂移；同一遮挡模式对应多种真实姿态，迫使模型从确定性推断退化为概率性猜测。

因此，为有效解决人体姿态估计中的遮挡问题，本文提出一种面向复杂工业场景的人体姿态估计性能增强方法，具体通过动态数据增强方法和模型多轮动态迭代方法实现。实验结果证明，本文所提方法在自建数据集上相比先进方法 PCT 平均准确率 AP 和平均召回率 AR 分别提升了 3.8% 和 2.7%。针对工业场景中的遮挡挑战，所提方法有效提升了模型的适应性和预测可靠性。注意到相机成像从 3D 到 2D 映射时，当相机远离主体会存在肢体长度和关节角度因透视投影发生畸变的问题。由于本文方法的设计无空间限定性，其可兼容三维扩展。

本研究的方法在实际部署中，利用全局快门工业相机采集数据输入人体关键点检测模型，将训练模型部署于 NVIDIA Jetson 平台进行边缘计算，通过工业级显示器实时可视化预测结果。同时，以关键点数据为基础，衔接行为识别模块并联动报警装置，可实现工人不安全行为监控与工业安全事故预防。

4 结　论

本研究提出了一种面向复杂工业场景的人体姿态估计性能增强方法,用以解决工业场景下模型面对遮挡因素性能下降的问题。通过量子化编码器对姿态进行结构性建模,能突破现有方法对显式视觉信息的依赖。同时,所提出的基于真实工业场景遮挡物的图像数据增强方法,一方面有效克服了遮挡数据集构建困难的问题,另一方面显著提升了人体姿态估计模型在特定工业场景下的泛化能力。特别地,提出一种模型多轮迭代性能增强方法,力求实现更加精准的姿态估计。实验结果表明,对比先进方法 PCT,所提出的方法在自建的工业场景数据集上,其平均准确率 AP 提升 3.8%,平均召回率 AR 提升 2.7%,说明所提方法具有优越性,能够有效增强模型在工业环境中的抗遮挡性能,提升姿态估计的稳定性。

未来工作将从以下方向进一步深化人体姿态估计技术:通过多目相机融合以校正投影几何误差,同时为分析模型在不同场景与不同类别姿态下的性能表现差异,未来将拍摄标注更多不同工业场景的数据集,并对人体姿态进行分类处理,分别针对不同场景、不同类别姿态进行专项实验,以充分挖掘模型价值。

参考文献

[1] 郑继贵. 精密测量, 智能制造的基石[J]. 仪器仪表学报, 2017, 38(8): 1821.
ZHU J G. Precision measurement, the cornerstone of intelligent manufacturing[J]. Chinese Journal of Scientific Instrument, 2017, 38(8): 1821.

[2] 马南峰, 姚锡凡, 陈飞翔, 等. 面向工业 5.0 的人本智造[J]. 机械工程学报, 2022, 58(18): 88-102.
MA N F, YAO X F, CHEN F X, et al. Human-centric smart manufacturing for industry 5.0[J]. Journal of Mechanical Engineering, 2022, 58(18): 88-102.

[3] 李佳宁, 王东凯, 张史梁. 基于深度学习的二维人体姿态估计:现状及展望[J]. 计算机学报, 2024, 47(1): 231-250.
LI J N, WANG D K, ZHANG SH L. Deep-learning-based 2D human pose estimation: Present and future[J]. Chinese Journal of Computers, 2024, 47(1): 231-250.

[4] LIU T Y, WENG C Y, JIAO L, et al. Toward fast 3D human activity recognition: A refined feature based on minimum joint freedom model (Mint)[J]. Journal of Manufacturing Systems, 2023, 66: 127-141.

[5] LIU T Y, WENG C Y, HUANG J, et al. A lightweight future skeleton generation network (FSGN) based on spatio-temporal encoding and decoding[J]. Knowledge-Based Systems, 2024, 306: 112717.

[6] 鲍劲松, 张荣, 李婕, 等. 面向人-机-环境共融的数字孪生协同技术[J]. 机械工程学报, 2022, 58(18): 103-115.
BAO J S, ZHANG R, LI J, et al. Digital-twin collaborative technology for human-robot-environment integration[J]. Journal of Mechanical Engineering, 2022, 58(18): 103-115.

[7] YANG X J, WENG C Y, JIAO L, et al. ATD-GCN: A human activity recognition approach for human-robot collaboration based on adaptive skeleton tree-decomposition[J]. Robotics and Computer-Integrated Manufacturing, 2025, 95: 103019.

[8] 禹鑫燚, 王正安, 吴加鑫, 等. 满足不同交互任务的人机共融系统设计[J]. 自动化学报, 2022, 48(9): 2265-2276.
YU X Y, WANG ZH AN, WU J X, et al. System design for human-robot coexisting environment satisfying multiple interaction tasks[J]. Acta Automatica Sinica, 2022, 48(9): 2265-2276.

[9] 杨旭升, 吴江宇, 胡佛, 等. 基于渐进高斯滤波融合的多视角人体姿态估计[J]. 自动化学报, 2024, 50(3): 607-616.
YANG X SH, WU J Y, HU F, et al. Multi-view human pose estimation based on progressive gaussian filtering fusion[J]. Acta Automatica Sinica, 2024, 50(3): 607-616.

[10] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.

[11] MAHARANA K, MONDAL S, NEMADE B. A review: Data pre-processing and data augmentation techniques[J]. Global Transitions Proceedings, 2022, 3(1): 91-99.

[12] NAVEED H, ANWAR S, HAYAT M, et al. Survey:

Image mixing and deleting for data augmentation[J]. Engineering Applications of Artificial Intelligence, 2024, 131: 107791.

[13] 杨锁荣, 杨洪朝, 申富饶, 等. 面向深度学习的图像数据增强综述[J]. 软件学报, 2025, 36(3): 1390-1412.
YANG S R, YANG H CH, SHEN F R, et al. Image data augmentation for deep learning: A Survey[J]. Journal of Software, 2025, 36(3): 1390-1412.

[14] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: Learning augmentation strategies from data[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 113-123.

[15] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 3008-3017.

[16] 程思雨, 陈莹. 基于 ViT 的细粒度特征增强无监督行人重识别方法[J]. 电子测量与仪器学报, 2024, 38(9): 24-35.
CHENG S Y, CHEN Y. Fine-grained feature enhancement unsupervisedperson re-identification method based on ViT[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 24-35.

[17] LI B Y, WU F, LIM S, et al. On feature normalization and data augmentation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12378-12387.

[18] 王云艳, 周志刚, 罗帅. 基于数据增强的太阳能电池片缺陷检测[J]. 电子测量与仪器学报, 2021, 35(1): 26-32.
WANG Y Y, ZHOU ZH G, LUO SH. Defect detection of solar cell based on data augmentation[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1): 26-32.

[19] CHINBAT V, BAE S. GA3N: Generative adversarial AutoAugment network[J]. Pattern Recognition, 2022, 127: 108637.

[20] 杨傲雷, 周应宏, 杨帮华, 等. 基于 Transformer 的三维人体姿态估计及其动作达成度评估[J]. 仪器仪表学报, 2024, 45(4): 136-144.
YANG AO L, ZHOU Y H, YANG B H, et al. Transformer-based 3D human pose estimation and action achievement evaluation[J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 136-144.

[21] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-3364.

[22] 张小娜, 吴庆涛. 基于深度学习的自顶向下人体姿态估计算法[J]. 电子测量技术, 2021, 44(9): 105-109.
ZHANG X N, WU Q T. Top-down human pose estimation algorithm based on deep learning[J]. Electronic Measurement Technology, 2021, 44(9): 105-109.

[23] 刘宏哲, 陶相如, 徐成, 等. 一种优化多尺度特征融合的人体姿态估计方法[J]. 机械工程学报, 2024, 60(16): 306-313.
LIU H ZH, TAO X R, XU CH, et al. Human pose estimation method based on optimized multi-scale feature fusion[J]. Journal of Mechanical Engineering, 2024, 60(16): 306-313.

[24] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186.

[25] LI M Y, ZHAO J. CE-HigherHRNet: Enhancing channel information for small persons bottom-up human pose estimation[J]. IAENG International Journal of Computer Science, 2022, 49(1): 260-269.

[26] NEWELL A, HUANG ZH AO, DENG J. Associative embedding: End-to-end learning for joint detection and grouping[J]. Advances in Neural Information Processing Systems, 2017, 30: 2278-2288.

[27] WANG B C, SONG C, LI X Y, et al. A deep learning-enabled visual-inertial fusion method for human pose estimation in occluded human-robot collaborative assembly scenarios[J]. Robotics and Computer-Integrated Manufacturing, 2025, 93: 102906.

[28] 代钦, 石祥滨, 乔建忠, 等. 结合遮挡级别的人体姿态估计方法[J]. 计算机辅助设计与图形学学报, 2017, 29(2): 279-289.

DAI Q, SHI X B, QIAO J ZH, et al. Articulated human pose estimation with occlusion level [J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(2): 279-289.

[29] JIANG J H, XIA N. A dual-channel network based on occlusion feature compensation for human pose estimation[J]. Image and Vision Computing, 2024, 151: 105290.

[30] VAN DEN OORD A, VINYALS O, KAVUKCUOGLU K. Neural discrete representation learning[J]. Advances in Neural Information Processing Systems, 2017, 30: 6309-6318.

[31] GENG Z G, WANG C Y, WEI Y X, et al. Human pose as compositional tokens[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 660-671.

[32] SUN X, XIAO B, WEI F Y, et al. Integral human pose regression[C]. ECCV-2018, 2018: 536-553.

[33] LI Y J, YANG S, LIU P D, et al. Simcc: A simple coordinate classification perspective for human pose estimation[C]. ECCV-2022, 2022: 89-106.

[34] XU Y F, ZHANG J, ZHANG Q M, et al. Vitpose: Simple vision transformer baselines for human pose estimation[J]. Advances in Neural Information Processing Systems, 2022, 35: 38571-38584.

作者简介

李帆雅,现为杭州电子科技大学本科生,主要研究方向为人体姿态估计。
E-mail:22061703@ hdu. edu. cn

Li Fanya is currently a B. Sc. candidate at Hangzhou Dianzi University. Her main research interest is human pose estimation.

张泽辉(通信作者),2012 年于上海海事大学获得学士学位,2019 年于武汉理工大学获得硕士学位,2022 年于南开大学获得博士学位,现为杭州电子科技大学副研究员,主要研究方向为人体姿态估计和工业安全。
E-mail:zhangzehui@ hdu. edu. cn

Zhang Zehui (Corresponding author) received his B. Sc. degree from Shanghai Maritime University in 2012, his M. Sc. degree from Wuhan University of Technology in 2019, and his Ph. D. degree from Nankai University in 2022. He is currently an associate research fellow at Hangzhou Dianzi University. His main research interests include human pose estimation and industrial safety.

陈博洋,2013 年于宁夏大学获得学士学位,现为宁夏石化银骏安全技术咨询有限公司法人、总经理,主要研究方向为安全生产信息化系统研发和安全评价。
E-mail:64868053@ qq. com

Chen Boyang received his B. Sc. degree from Ningxia University in 2013. He is currently legal person and the general manager of Ningxia Petrochemical Yinjun Safety Technology Consulting Co., Ltd. His main research interests include the development of information systems for safety production and safety evaluation.

管聪,2010 年于武汉理工大学获得学士学位,2012 年于武汉理工大学获得硕士学位,2015 年于武汉理工大学获得博士学位,现为武汉理工大学副教授,主要研究方向为智能控制和行为识别。
E-mail:guancong2008@ 126. com

Guan Cong received his B. Sc., M. Sc., and Ph. D. degrees all from Wuhan University of Technology in 2010, 2012, and 2015, respectively. He is currently an associate professor at Wuhan University of Technology. His main research interests include intelligent control and behavior recognition.