DOI:10.3969/j.issn.1671-024x.2026.04.010

基于二维和三维人体姿态估计的跌倒检测

李现国[1,2],段显霖[1],杨彦利[1,2],蒋 鹏[3],张 磊[3]
(1.天津工业大学 电子与信息工程学院,天津 300387;2.沧州市天津工业大学研究院,河北 沧州 061000;3.山东新巨龙能源有限责任公司,山东 巨野 274000)

摘 要: 针对现有跌倒检测技术普遍存在的误检漏检问题,提出联合二维和三维人体姿态估计的跌倒检测策略:基于Mamba模型设计Mamba-Attention注意力机制、改进C3_M模块,提出二维人体姿态估计算法MambaPose,提高二维人体关键点检测能力;融合Flash-Attention、DCT变换与低通滤波,提出三维人体姿态估计算法FlashPose,提高二维人体关键点映射到三维空间的准确性和实时性;将MambaPose、FlashPose和K近邻分类器(KNN)结合,构建出一种高准确率实时跌倒检测方法MFKF。公共数据集上的实验结果表明:在COCO数据集上,MambaPose算法与YOLOv7Pose相比,AP值提高了2.5个百分点,显著提升了二维人体关键点检测的准确性,而且小目标检测的置信度更高、处理低分辨率图像时鲁棒性更强;在Human3.6M数据集上,FlashPose算法的MPJPE和FPS值分别达到47.2 mm和314,优于PoseFormerV2,准确性更高、计算速度更快;在UR、Le2i和Multicam这3个跌倒检测数据集上,本文MFKF方法的F_1、灵敏度、特异度、准确率均为最优,且FPS值均高于22.7,实现了跌倒的准确和实时检测。代码及演示视频链接:https://github.com/xgli411/MFKF.

关键词: 跌倒检测;二维人体姿态估计;三维人体姿态估计;K近邻分类器(KNN);Mamba模型
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1671-024X(2026)04-0080-09

Fall detection based on 2D and 3D human pose estimation

Li Xianguo[1,2], Duan Xianlin[1], Yang Yanli[1,2], Jiang Peng[3], Zhang Lei[3]
(1. School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China; 2. Cangzhou Institute of Tiangong University, Cangzhou 061000, Hebei Province, China; 3. Shandong Xinjulong Energy Co., Ltd., Juye 274000, Shandong Province, China)

Abstract: To address the problem of false positives and false negatives in the existing fall detection technologies, a fall detection strategy integrating two-dimensional and three-dimensional human pose estimation is proposed. Based on the Mamba model, the Mamba-Attention mechanism is designed, the C3_M module is enhanced, and the 2D human pose estimation algorithm MambaPose is introduced. By integrating Flash-Attention, DCT transform and low-pass filtering, the 3D human pose estimation algorithm FlashPose is proposed to improve the accuracy and real-time performance of mapping 2D human keypoints to 3D space. Finally, MFKF, a high accuracy real-time fall detection method, is constructed by combining MambaPose, FlashPose, and the KNN classifier. The experimental results on the public datasets show that: On the COCO dataset, the MambaPose algorithm improves the AP value by 2.5 percentage points compared with YOLOv7Pose, which significantly improves the accuracy of 2D human keypoint detection, and has higher confidence in small target detection and greater robustness when processing low-resolution images; on the Human3.6M dataset, the FlashPose algorithm′s MPJPE and FPS values reach 47.2 mm and 314, respectively, which are better than PoseFormerV2, with higher accuracy and faster computation speed; on the three fall detection datasets of UR, Le2i, and Multicam, the proposed MFKF method has the optimal F_1, sensitivity, specificity, and accuracy indexes, and the FPS values are higher than 22.7, which realizes accurate and real-time detection of falls. Code and demonstration videos are available at https://github.com/xgli411/MFKF.

收稿日期: 2024-11-11 **基金项目:** 天津市技术创新引导专项(23YDTPJC00290)
通信作者: 李现国(1981—),男,博士,教授,主要研究方向为智能信息处理、光电检测。E-mail:lixianguo@tiangong.edu.cn

Key words: fall detection; 2D human pose estimation; 3D human pose estimation; KNN classifier; Mamba model

据世界卫生组织报告,人体跌倒已成为意外死亡的第二大原因[1]。每年约有 30%的 65 岁以上老年人经历跌倒,危害巨大[2-3]。相关研究显示,跌倒也是 65 岁以下成年人面临失去独立性的重要风险因素之一[4]。此外,近年来大中城市的人员密集场所跌倒踩踏事件频发,如 2022 年韩国首尔市踩踏事故导致 156 人死亡、152 人受伤。因此,及时准确的跌倒检测,对于保护人的生命安全、提供紧急救助、减少伤害程度至关重要。

现有的跌倒检测技术主要分为 2 类:接触式检测和非接触式检测。接触式检测技术主要依赖于加速度传感器或陀螺仪,通过捕捉人体运动变化识别跌倒行为,但需要使用者持续佩戴,影响体验,且易受佩戴位置影响而产生误报。非接触式检测技术主要采用红外传感器、超声波或毫米波雷达以及可见光摄像头等检测人体活动,进而判断是否发生跌倒。其中,采用红外传感器和雷达的方法只适用于简单场景,在实际应用中鲁棒性和可靠性低。近年来,随着人工智能和深度学习技术的快速发展,基于可见光摄像头的机器视觉跌倒检测方法取得了显著进步,且适用于复杂场景,已成为当前研究热点。

Lu 等[5]结合视频数据,通过三维卷积神经网络(3D convolutional neural network,3D-CNN)检测三维人体关键点,使用长短期记忆网络(long short-term memory,LSTM)判别当前人体状态。Chen 等[6]采用 OpenPose 算法[7]和骨骼状态分析,通过设定阈值实现跌倒检测。Zhang 等[8]利用 BlazePose 提取关键点,结合改进的时空图卷积网络(spatial-temporal graph convolutional networks,ST-GCN)分析人体状态,同时融入视频帧时序信息,取得了良好的效果。Amsaprabhaa 等[9]先以 YOLOv3 定位人体位置,再用 OpenPose 提取关键点,最后通过 ST-GCN 分析人体状态。Beddiar 等[10]使用支持向量机(SVM)分类器和时间卷积网络(TCN)进行跌倒检测。Kan 等[11]和 Marcos 等[12]分别利用改进的 YOLOv5 和 Transformer 网络进行跌倒检测。尽管这些方法取得了一定效果,但仍存在一定的局限性,如跌倒数据集规模有限、低分辨率环境或检测小目标时性能下降等。而且当前常见的基于人体姿态估计的跌倒检测方法主要依赖二维人体姿态估计,通过对二维人体的关键点分类检测跌倒事件[7-8]。然而,这种方法因缺乏深度信息导致跌倒与未跌倒状态的关键点坐标分布混淆,增加分类器误判风险,影响跌倒检测的准确性。这些问题共同制约了基于机器视觉的跌倒检测技术在实际生活场景中的普及应用。

针对上述问题,本文提出一种基于二维和三维人体姿态估计的跌倒检测方法。首先提出一种联合二维和三维人体姿态估计的跌倒检测策略,有效避免由于跌倒数据集规模有限、RGB 相机缺乏深度信息导致的模型泛化能力不足和准确性低的问题。然后,基于此策略提出二维人体姿态估计算法 MambaPose 及网络结构,通过融合 Mamba 模型设计 Mamba-Attention 注意力机制,改进 C3_M 模块替代传统 C3 模块,以提高模型检测二维人体关键点的能力;加入空间深度转换卷积(SPDConv),有效缓解下采样过程中信息丢失的问题,提升模型处理低分辨率图像和检测小目标时的性能。在此基础上,提出三维人体姿态估计算法 FlashPose,将二维人体关键点映射至三维空间,解决 RGB 相机拍摄视频或图像缺乏深度信息的问题;融合 Flash-Attention,降低 Transformer 中 self-attention 的计算复杂度,提升模型训练推理效率;采用离散余弦变换(discrete cosine transform,DCT)和低通滤波技术,增强模型抗噪能力。最后,结合 MambaPose、FlashPose 以及 K 近邻分类器(K-nearest neighbor,KNN)(KNNPose),构建出一种高准确率的实时跌倒检测方法,简称为 MFKF。此外,本文还自制了含 5 500 个三维人体关键点的跌倒检测数据集,用于 KNNPose 训练,并在 UR、Le2i 和 Multicam 等 3 个跌倒检测数据集上验证了 MFKF 的检测效果。

1　MFKF 跌倒检测方法

1.1　整体框架

MFKF 跌倒检测方法的整体框架如图 1 所示。

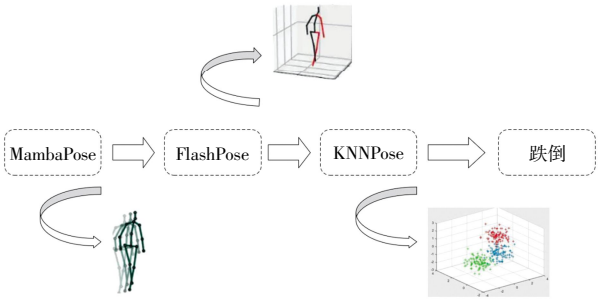


图 1　MFKF 跌倒检测方法的整体框架

Fig.1　Overall framework of the MFKF fall detection method

该方法主要由 3 部分组成:首先,使用 MambaPose 对视频进行二维人体关键点提取;然后,将提取的二维人体关键点通过 FlashPose 实现向三维空间的映

射,生成三维人体关键点;最后,将三维人体关键点送入预先训练好的 KNNPose 模型进行分类,从而高效准确地实现跌倒检测。

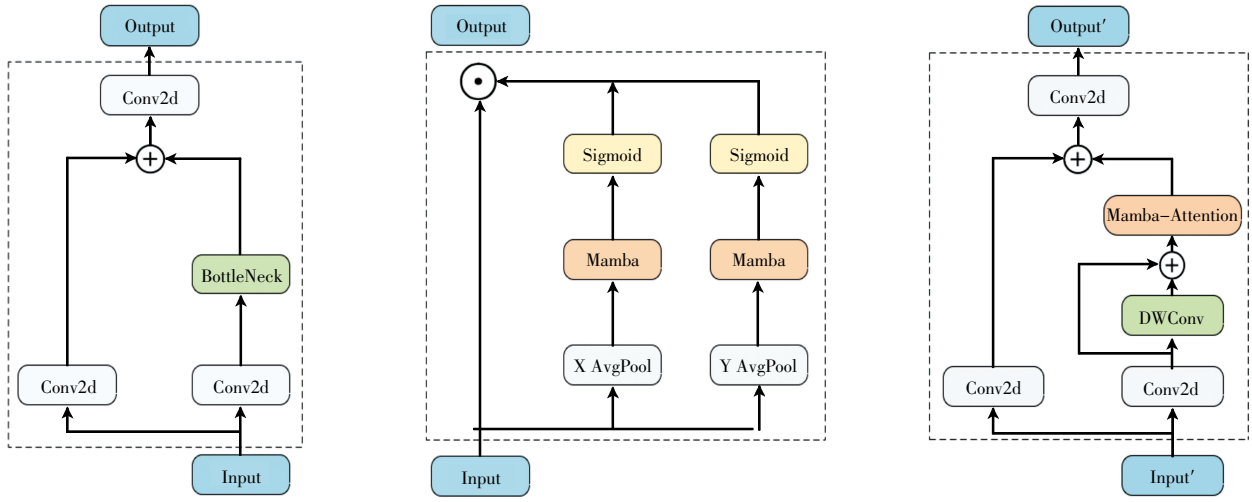
为克服 RGB 相机无法获取深度信息导致误检和漏检的缺陷，本文将二维人体关键点映射至三维空间,并对三维人体关键点进行分类,以提升检测准确率。同时,针对直接使用三维人体姿态估计网络可能导致的训练与推理速度下降问题，对其进行了优化,确保在满足实时性的前提下,准确实现人体关键点二维至三维的映射。此外,选择简单高效的 KNN 作为分类器,进一步简化方法复杂度,保证分类过程中准确性与效率的平衡。

1.2　二维人体姿态估计算法 MambaPose

本文提出一种基于改进 YOLOv7Pose[13]的二维人体姿态估计算法 MambaPose，实现视频中二维人体关键点的准确提取。

1.2.1　改进的 C3_M 模块

在 YOLOv7Pose 架构中,传统 C3 模块在捕捉长距离依赖关系方面存在明显不足,使用全局注意力机制或 Transformer 常因忽视通道维度或 Transformer 固有的计算复杂性，而面临准确度提升有限与计算负担加重的双重挑战。2023 年,Mamba 模型[14]因其跨领域的优越性能而备受关注。该模型采用线性时间序列,结合选择性状态空间，通过智能选择传播或丢弃序列中元素的信息，有效应对长序列处理的计算瓶颈,推理速度较 Transformer 显著提升。本文将 Mamba 模型引入人体姿态估计，设计了 Mamba-Attention 注意力机制及全新的 C3_M 模块,以替代传统的 C3 模块。本文改进的 C3_M 与传统的 C3 模块结构比较如图 2 所示。



(a) C3　　(b) Mamba-Attention　　(c) 改进的 C3_M

图 2　改进的 C3_M 与传统的 C3 模块结构比较

Fig.2　Comparison of improved C3_M and traditional C3 module structures

图 2(a)为 YOLOv7Pose 中的 C3 模块,其核心在于堆叠 BottleNeck 模块以增强网络的特征提取能力。然而,受限于传统卷积操作的“窗口平移”特性,该模块主要聚焦于局部特征的提取,难以捕捉全局特征信息。本文设计的 Mamba-Attention 注意力机制能够有效弥补感知全局特征信息能力的不足。如图 2(b)所示，首先将输入特征图在通道维度上进行平均池化,然后经过 Mamba 模型扫描,获取特征的长距离依赖关系,再经过 Sigmoid 函数激活。特征图 X 的计算式为

$$X = \mathrm{Sigmoid}(\mathrm{Mamba}(\mathrm{AvgPool}(\mathrm{Input}X))) \tag{1}$$

特征图 Y 的计算式为:

$$Y = \mathrm{Sigmoid}(\mathrm{Mamba}(\mathrm{AvgPool}(\mathrm{Input}Y))) \tag{2}$$

最后将尺度特征图 X、特征图 Y 与输入特征图执行逐元素相乘,得到输出特征图 Output:

$$\mathrm{Output} = \mathrm{Input}*X*Y \tag{3}$$

由于 Mamba-Attention 缺乏局部特征提取能力,本文融合了深度可分离卷积(DWConv)以强化局部特征信息的提取并降低计算复杂度。同时采用残差结构策略,防止特征信息在传递过程中的丢失,如图 2(c)所示。改进的 C3_M 模块计算过程为:

$$A = \mathrm{Mamba\ Attention}(\mathrm{DWConv}(\mathrm{Conv2d}(\mathrm{Input}')) + \mathrm{Conv2d}(\mathrm{Input}')) \tag{4}$$

$$B = \mathrm{Conv2d}(\mathrm{Input}') \tag{5}$$

$$C = A + B \tag{6}$$

$$\mathrm{Output}' = \mathrm{Conv2d}(C) \tag{7}$$

1.2.2　空间深度转换卷积

图 3 为二维人体姿态估计算法 MambaPose 的网络结构。

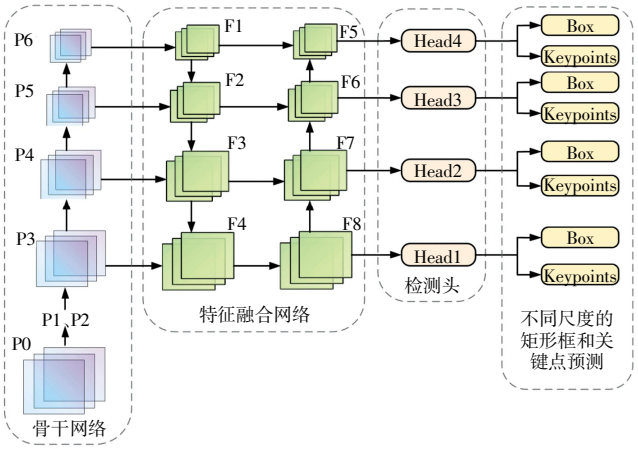


图 3　MambaPose 的网络结构

Fig.3　Network structure of MambaPose

在 YOLOv7Pose 的骨干网络和特征融合网络部分,大量的下采样操作导致细粒度信息的丢失和不充分的特征表示学习,最终导致低分辨率图像和小目标的人体姿态估计准确率较低。为缓解下采样操作导致的信息丢失问题，本文采用空间深度转换卷积(SPDConv)[15]替代传统的下采样卷积。SPDConv 由一个空间到深度(Space-to-Depth)层和一个非步长卷积(Non-Strided Convolution)层组成。

空间到深度(SPD)层是一种转换层,旨在将输入图像的空间维度转换为深度维度,在不损失信息的前提下增加特征图深度。在处理低分辨率图像和小目标时,该层尤为重要,通过转换至深度维度,避免传统步长卷积和池化操作导致的信息丢失。

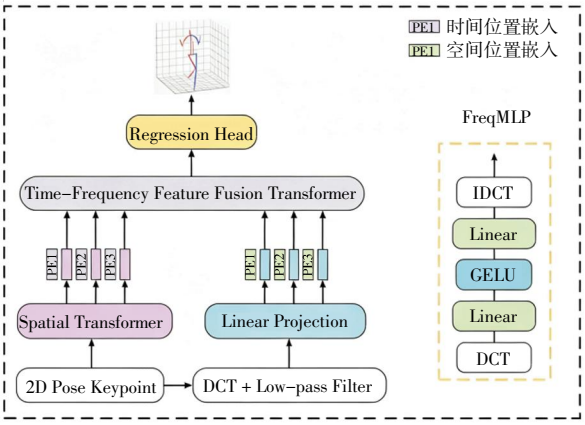
非步长卷积层不采用步长卷积以保留细粒度信息。在 SPD 层后使用非步长卷积层,能够在不缩减特征图尺寸的前提下维护图像的细粒度信息。

在图 3 的 P0→P3、P3→P4、P4→P5、P5→P6、F8→F7、F7→F6、F6→F5 过程中，使用 SPDConv 替换原始下采样卷积,通过这一改进,增强了模型对小目标和低分辨率图像的人体姿态感知能力。

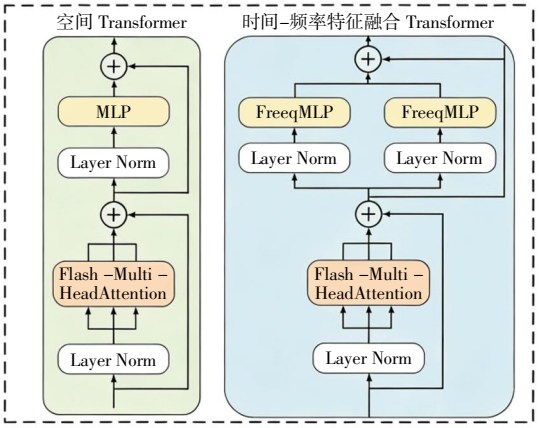
1.3　三维人体姿态估计算法 FlashPose

本文提出了一种三维人体姿态估计算法 FlashPose,旨在丰富人体姿态信息,提升后续跌倒检测任务的准确性与可靠性。图 4 所示为三维人体姿态估计算法 FlashPose 的网络结构。

借鉴 PoseFormerV2[16]的方法,引入离散余弦变换(DCT)将输入的二维骨架序列转换至频域,并仅采用低频系数表征序列。此方法能有效捕捉序列的整体趋势,同时滤除高频噪声,如关键点检测中的抖动与异常值。此外，还设计了一个时间-频率特征融合模块



(a) 算法整体结构



(b) 空间 Transformer 及时间-频率特征融合 Transformer 结构

图 4　三维人体姿态估计算法 FlashPose 的网络结构

Fig.4　Network architecture of FlashPose, a 3D human pose estimation algorithm

(time-frequency feature fusion Transformer)，该模块融合了时域特征(源自少数中心帧的空间 Transformer 编码器)与频域特征(源自完整序列的低频 DCT 系数),使模型能同时利用局部精细动作信息与全局序列信息。

为优化 Transformer 的复杂结构，本文借鉴 PoseFormerV2 的处理方式,将多层感知器(MLP)改进为快速多层感知器(FreqMLP),以更好地处理时域和频域特征。FreqMLP 在频域特征的前馈网络中运用 DCT 变换与逆 DCT 变换,调整各频率成分的权重。考虑到跌倒检测方法的实时性需求,引入 Flash-Attention[17]改进 Multi-Head Attention，提出全新的 Flash-Multi-Head Attention,在提升模型训练和推理速度的同时,也提高了模型的检测性能。Flash-Attention 通过优化注意力计算,降低内存占用,解决了 Transformer 模型中 self-attention 层处理长序列时速度慢、内存消耗高的问题;通过重新排序注意力计算、运用平铺与重计算等技术,显著提升计算速度,并将内存使用从平方关系降

低到线性关系。

2　实验验证

本文实验硬件环境为 NVIDIA GTX 3090 GPU（两块），软件环境为 Ubuntu16.04、PyTorch = 2.1.1、CUDA = 11.9。

2.1　实验数据

本文提出的二维人体姿态估计算法 MambaPose 的训练和测试分别在 COCO 数据集的 train 集和 val 集上进行。COCO 数据集[18]是目前公开数据集中人体实例数量最多的人体姿态估计数据集，涵盖 200 000 张图片，涉及 250 000 个人体实例，每个人体标注 17 个关键点。该数据集分为 train 集（57 000 张图片）、val 集（5 000 张图片）和 test-dev 集（20 000 张图片）。本文与 MambaPose 对比的先进算法均在 COCO 数据集的 train 集上进行训练。训练策略与 YOLOPose[19]一致，即 300 轮训练，基础学习率设为 $1/e^{-2}$，并应用随机尺度（[0.5，1.5]）、随机平移（±10）、概率为 0.5 的随机翻转等数据增强技术。

本文提出的三维人体姿态估计算法 FlashPose 的训练和测试均在 Human3.6M 数据集上进行。Human3.6M 数据集[20]是三维人体姿态估计领域的重要公开数据集。该数据集包含超过 360 万张图像，对应丰富的三维人体姿态样本，由 11 名专业演员（6 男 5 女）参与录制，涵盖讨论、吸烟、拍照、打电话等多种日常活动。通过不同角度的摄像机拍摄，确保了数据的多样性和完整性。FlashPose 的训练及测试策略与 PoseFormerV2 保持一致。

本文制作了一个包含 3 000 个跌倒场景和 2 500 个非跌倒场景的三维人体关键点数据集（共 5 500 个样本）。该数据集通过收集公共跌倒检测数据集及其他现实场景中跌倒与未跌倒姿态的图像，并使用本文提出的二维和三维人体姿态估计算法，提取三维人体关键点。然后，对三维人体关键点进行优化和后处理，包括使用平滑算法减少抖动和噪声，使用几何约束保持人体关节的合理性。基于此数据集，训练了 KNN 分类器，构建了 KNNPose 模型。由于该数据集包含不同环境及不同姿态的三维人体关键点，且广泛覆盖蹲、坐、弯腰等易混淆姿态，能够有效提高 KNNPose 模型对不同环境中不同人体姿态的感知能力，提升跌倒检测的准确性。选择 KNN 作为二分类器，旨在优化 MFKF 跌倒检测方法的执行速度，相较于其他复杂分类器，KNN 在效率上具有明显优势。鉴于跌倒与非跌倒状态下的三维人体关键点坐标分布差异较大，KNN 分类器足以高效准确地完成分类任务。

UR Fall Detection Dataset[21]是一个公开的跌倒检测数据集，包含 30 个跌倒视频和 40 个正常生活视频，总计 70 个视频。Le2i Fall Detection Dataset [22]共包含 191 个视频，其中包含咖啡厅、居家、讲堂与办公室不同场景下的数据。视频中存在假摔部分，且视频中存在没有人的帧数。Multicam Fall Detection Dataset 共有 24 个场景，每个场景用 8 个摄像机记录，前 22 个场景包含跌倒和正常时间，后 2 个场景仅包含其他混杂事件，共 192 个视频。

2.2　评估指标

采用灵敏度（Sensitivity）、特异度（Specificity）、准确率（Accuracy）、F_1 分数、精确度（Precision）、召回率（Recall）等评价指标进行跌倒检测评估，计算公式如下：

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

$$F_1 = 2*\frac{\text{Precision*Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

式中：TP（真阳性）、FP（假阳性）、TN（真阴性）及 FN（假阴性）分别表示跌倒类别检测为跌倒、非跌倒类别检测为跌倒、非跌倒类别检测为非跌倒、跌倒类别检测为非跌倒的样本数。AP（average precision）为平均精度，是人体姿态估计的核心评价指标，综合衡量关键点检测的整体准确率；AP_{50}、AP_{75} 分别为关键点相似度阈值取 0.5、0.75 时的平均精度；AP_S、AP_M、AP_L 分别为小尺寸、中等尺寸、大尺寸人体目标的平均精度；AR（average recall）为平均召回率，反映模型对人体关键点的召回能力。

2.3　实验结果分析与对比

表 1 为二维人体姿态估计算法 MambaPose 在 COCO val 数据集上的实验结果。MambaPose 相较于原始 YOLOv7Pose 的 AP 值提升了 2.5 个百分点，尽管参数量增加了 6.9×10^6，但单张图片推理时间仅延长 5.4 ms。与采用 Swin-L 作为骨干网络的 PETR 和 ED-Pose 算法相比，MambaPose 在准确率上更胜一筹，且在推理速度和参数量上均表现更优，验证了本文设计的 C3_M 模块及 Mamba-Attention 注意力机制相较于 Transformer 更为高效。

表 1　MambaPose 在 COCO val 数据集上的实验结果

Tab.1　Experimental results of MambaPose on the COCO val dataset

算法	骨干网络	参数量/10^6	时间/ms	AP/%	AP_{50}/%	AP_{75}/%	AP_M/%	AP_L/%	AR/%
DirectPose[23]	ResNet-101	—	—	63.3	86.7	69.4	57.8	71.2	—
FCPose[24]	ResNet-101	60.5	93	65.6	87.9	72.6	62.1	72.3	—
InsPose[25]	ResNet-101	—	100	66.3	89.2	73.0	61.2	73.9	—
PETR[26]	ResNet-50	43.7	89	67.6	89.8	75.3	61.6	76.0	—
PETR[26]	Swin-L	213.8	133	70.5	91.5	78.7	65.2	78.0	—
CID[27]	HRNet-w32	29.4	84	68.9	89.9	76.9	63.2	77.7	74.6
CID[27]	HRNet-W48	65.4	94.8	70.7	90.4	77.9	66.3	77.8	76.4
ED-Pose[28]	ResNet-50	50.6	135.2	69.8	90.2	77.2	64.3	77.4	—
ED-Pose[28]	Swin-L	218.0	265.6	72.7	92.3	80.9	67.6	80.0	—
KAPAO-s[29]	CSPNet	12.6	26.9	63.8	88.4	70.4	58.6	71.7	71.2
KAPAO-m[29]	CSPNet	35.8	37.0	68.8	90.5	76.5	64.3	76.0	76.3
KAPAO-l[29]	CSPNet	77.0	50.2	70.3	91.2	77.8	66.3	76.8	77.7
YOLO-Pose-s[19]	CSPDarknet	10.8	7.9	63.2	87.8	69.5	57.6	72.6	67.6
YOLO-Pose-m[19]	CSPDarknet	29.3	12.5	68.6	90.7	75.8	63.4	77.1	72.8
YOLO-Pose-l[19]	CSPDarknet	61.3	20.5	70.2	91.1	77.8	65.3	78.2	74.3
YOLOv7Pose[13]	CSPDarknet	74.2	21.8	71.4	90.5	78.7	66.6	79.2	77.6
本文算法(MambaPose)	CSPDarknet	81.1	27.2	73.9	93.2	81.7	69.0	80.5	78.9

表 2 为 MambaPose 算法各模块的消融实验结果，表中数据表明单独加入空间深度转换卷积(SPDConv)后，AP_S 指标提升 2 个百分点，凸显了该算法对小目标人物检测的有效性。

表 2　MambaPose 各模块消融实验结果

Tab.2　Results of MambaPose ablation experiments by module

SPDConv	C3_M	AP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	参数量/10^6
		71.4	90.5	78.7	52.1	66.6	79.2	74.2
√		72.2	91.8	80.6	54.1	67.5	79.4	75.8
	√	73.4	92.8	81.3	52.6	68.4	80.4	79.4
√	√	73.9	93.2	81.7	54.5	69.0	80.5	81.1

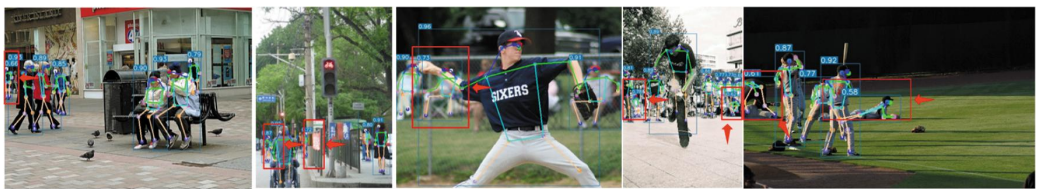
表 3 为图像尺寸对模型性能的影响对比。实验数据显示，随着图像输入分辨率的降低，YOLOv7Pose 的 AP 值下降 2.8 个百分点，而 MambaPose 的 AP 值仅下降 1.8 个百分点。这一对比凸显了 MambaPose 在处理低分辨率图像时具有更强鲁棒性。

表 3　图像尺寸对模型性能的影响

Tab.3　Influence of image size on model performance

算法	输入大小/像素	AP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	AR/%
YOLOv7Pose[13]	960×960	71.4	90.5	78.7	52.1	66.6	79.2	77.6
本文算法(MambaPose)	960×960	73.9	93.2	81.7	54.5	69.0	80.5	78.9
YOLOv7Pose[13]	640×640	68.6	87.2	76.0	45.4	65.1	78.1	75.4
本文算法(MambaPose)	640×640	72.1	91.7	80.3	52.4	67.4	79.2	78.0

图 5 为 MambaPose(第 1 行)与 YOLOv7Pose(第 2 行)的检测效果对比。由图 5 可以看出，MambaPose 置信度更高，整体表现更佳，特别是在小目标人物检测上，效果明显优于 YOLOv7Pose，进一步证明了 MambaPose 的优越性和实用性。



(a) MambaPose(第 1 行)



(b) YOLOv7Pose(第 2 行)

图 5　MambaPose(第 1 行)与 YOLOv7Pose(第 2 行)的效果对比

Fig.5　Comparison of the effect of MambaPose (first row) and YOLOv7Pose (second row)

表 4 为本文提出的三维人体姿态估计算法 FlashPose 在 Human3.6M 数据集上的实验结果对比。由表 4 可以看出，在采用相同二维人体姿态估计算法的前提下，相较于 PoseFormerV2，本文提出的 FlashPose 在 MPJPE 指标上（即预测关节点与真实关节点坐标间的平均欧几里得距离）上降低了 0.4 mm，帧率（FPS）则提升了 42。这一结果验证了 FlashPose 的预测准确性和计算速度均得到了较大提升。

表 4　FlashPose 在 Human3.6M 数据集上的实验结果

Tab.4　Experimental results of FlashPose on the Human3.6M dataset

算法	参数量/10^6	MFLOPs	MPJPE/mm	FPS
MHFormer[30]	—	342.9	47.8	—
PoseFormer[31]	9.58	150.0	49.9	320
PoseFormerV2[16]	10.86	351.7	47.6	272
本文算法（FlashPose）	11.12	351.7	47.2	314

表 5 为本文提出的 MFKF 跌倒检测方法在 UR 数据集上的实验结果对比。

表 5　MFKF 在 UR 数据集上的实验结果

Tab.5　Experimental results of MFKF on the UR dataset

%

方法	F_1	灵敏度	特异度	准确率
Beddiar et al.[10]	93.4		—	95.0
Egawa et al.[32]	98.5	97.3	—	99.7
Amsaprabhaa et al.[9]	96.2	—	—	95.8
Zheng et al.[8]	—	97.4	97.3	97.2
Zhang et al.[8]	99.1	—	99.1	98.6
Fei et al.[33]	—	96.6	—	98.5
Marcos et al.[12]	94.8	—	—	95.5
Li et al.[34]	94.7	93.9	98.6	97.1
本文方法（MFKF）	99.7	100	99.5	99.9

由表 5 可以看出，本文 MFKF 方法在关键性能指标 F_1 分数、灵敏度、特异度及准确率上均超越表 5 中其他方法。特别地，MFKF 在该数据集上的帧率（FPS）达到了 23.0，满足实时检测需求。与 Beddiar 等[10]、Zhang 等[8]仅依赖二维人体关键点分类的方法相比，MFKF 通过二维至三维的关键点映射后分类，性能显著提升。

表 6 和表 7 分别为 MFKF 在 Le2i 与 Multicam 跌倒检测数据集上的实验结果对比。结果显示，MFKF 在 Le2i 和 Multicam 数据集上均表现出较大优势，证明了 MFKF 方法具备良好的泛化能力。此外，MFKF 在 Le2i 和 Multicam 跌倒检测数据集上的帧率（FPS）分别达到 24.6 和 22.7，满足跌倒检测的实时性要求。

图 6 为在 Le2i 跌倒检测数据集上采用三维人体姿态估计算法的 MFKF（MambaPose + FlashPose + KNNPose）与未使用三维人体姿态估计算法的 MambaPose+KNNPose 的对比实验，该实验结果进一步验证了将二维人体关键点升至三维空间的有效性。

表 6　MFKF 在 Le2i 数据集上的实验结果

Tab.6　Experimental results of MFKF on the Le2i dataset

%

方法	F_1	灵敏度	特异度	准确率
Beddiar et al.[10]	87.8	86.8	88.4	85.8
Kan et al.[11]	—	89.1	—	90.4
Zheng et al.[8]	—	96.7	96.8	96.8
Zhang et al.[8]	99.0	—	99.0	99.3
Fei et al.[33]	—	98.7	—	98.9
Marcos et al.[12]	94.8	—	—	95.5
本文方法（MFKF）	99.4	100	99.5	99.7

表 7　MFKF 在 Multicam 数据集上的实验结果

Tab.7　Experimental results of MFKF on the Multicam dataset

%

方法	F_1	灵敏度	特异度	准确率
Zhang et al.[8]	98.1	—	98.1	99.2
本文方法（MFKF）	99.1	99.9	99.2	99.4

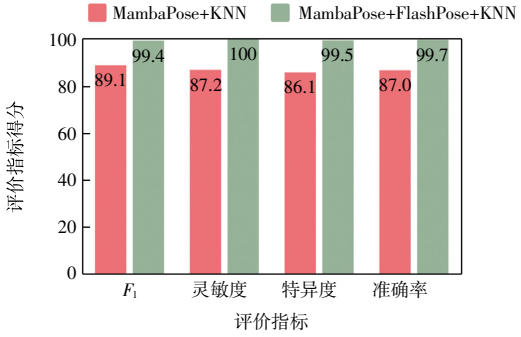


图 6　在 Le2i 数据集上是否使用 FlashPose 的实验结果对比

Fig.6　Comparison of experimental results with and without FlashPose on Le2i dataset

图 7 为三维人体姿态估计算法 FlashPose 实现的二维人体关键点到三维空间的映射效果。由图 7 可以看出，图中转换后的三维人体关键点增加一维深度信息后更好地反映了原始人体姿态，丰富了人体姿态变化信息。

上述实验及结果均是针对室内环境，由于目前缺少公开的户外场景跌倒检测数据集，本文方法 MFKF 未能针对户外环境进行实验验证。然而，MambaPose 和 FlashPose 通过融合 Mamba-Attention 注意力机制、DCT 变换与低通滤波技术，能够有效提升模型在复杂环境下的二维人体关键点检测能力、降低二维人体姿态估计带来的噪声影响，从而提高三维人体关键点检测的鲁棒性；KNNPose 依据人体 17 个关键点组成的人体姿态进行分类，尽管复杂环境二维或三维人体姿

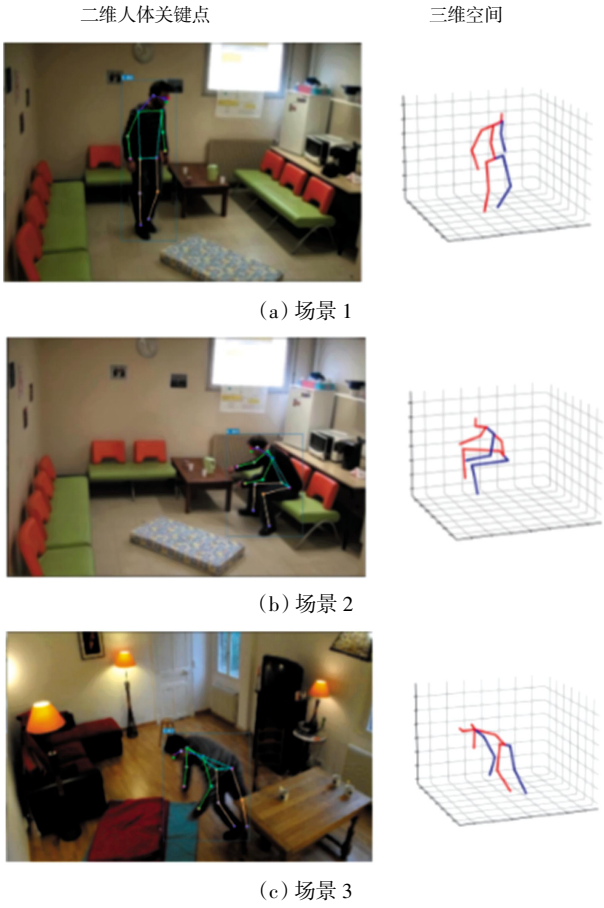


图 7　FlashPose 实现的二维人体关键点到三维空间的映射效果

Fig.7　Mapping effect of FlashPose from 2D human body keypoints to 3D space

态估计可能出现关键点预测偏差，但并不影响 KNNPose 对整体姿态的分类。此外,通过涵盖复杂户外场景的数据及图像,扩充自制的三维人体关键点数据集,能够进一步提高 KNNPose 模型对复杂姿态的感知能力。这些特点说明 MFKF 具有在复杂户外环境中应用的可能。同时,在复杂户外环境中,遮挡是影响检测准确率的关键因素。为应对这一挑战,可采用多摄像头从不同角度采集视频数据,综合分析各角度视频中的人体姿态信息,弥补单一视角的不足,从而提升 MFKF 方法的泛化能力和鲁棒性。

3　结　论

本文提出了一种基于二维和三维人体姿态估计的高准确率实时跌倒检测方法 MFKF。该方法主要由二维人体姿态估计算法 MambaPose、三维人体姿态估计算法 FlashPose 和 KNNPose 分类器组成。实验验证结果如下:

(1) 在 COCO 数据集上，相较于 YOLOv7Pose，MambaPose 算法将 AP 和 AP_S 值分别提升了 2.5 和 2.4 个百分点，有效提高了二维人体姿态估计的准确率，尤其对小目标感知能力更强、对低分辨率图像鲁棒性更好。

(2) 对于 Human3.6M 数据集,FlashPose 算法的 MPJPE 和 FPS 指标分别为 47.2 mm 和 314，准确性和速度均优于 PoseFormer V2。

(3) 在 UR、Le2i 及 Multicam 等 3 个跌倒检测数据集上,本文提出的 MFKF 方法表现最优,且 FPS 值均不低于 22.7，验证了本文所提出方法的正确性、有效性和鲁棒性。

实际应用中,低分辨率和遮挡问题是影响跌倒检测准确率的重要因素。下一步将基于现有算法,针对特征提取过程中信息丢失问题进一步优化,结合传统或深度学习的图像超分或增强方法缓解图像低分辨率影响,通过在实际应用中部署多摄像头降低遮挡的影响。此外,本文自制数据集规模有限,未能涵盖所有情况下的人体姿态,而且,本文跌倒检测实验主要基于室内公共数据集，尚未涵盖复杂的户外环境数据,可能导致模型在复杂环境中泛化能力受限。因此,未来将拓展数据集范围,覆盖更多场景,尤其是复杂的户外场景,完善跌倒检测方法,提高检测的准确性和实时性,为公共安全及人体姿态监测提供技术支持。

参考文献:

[1] World Health Organization (WHO). Falls[EB/OL]. (2021-04-26)[2024-11-06]. https://www.who.int/news-room/fact-sheets/detail/falls.

[2] Burns E, Kakara R. Deaths from falls among persons aged ≥ 65 years: United States, 2007-2016[J]. MMWR Morbidity and Mortality Weekly Report, 2018, 67(18): 509-514.

[3] Kelsey J L, Procter-Gray E, Hannan M T, et al. Heterogeneity of falls among older adults: Implications for public health prevention[J]. American Journal of Public Health, 2012, 102(11): 2149-2156.

[4] Yu M, Rhuma A, Naqvi S M, et al. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment[J]. IEEE Transactions on Information Technology in Biomedicine, 2012, 16(6): 1274-1286.

[5] 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.

[6] Chen W M, Jiang Z J, Guo H L, et al. Fall detection based on key points of human-skeleton using OpenPose[J]. Symmetry,

2020, 12(5): 744.

[7] Cao Z, Simon T, Wei S H, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 1302-1310.

[8] Zhang Y, Gan J S, Zhao Z W, et al. A real-time fall detection model based on BlazePose and improved ST-GCN[J]. Journal of Real-Time Image Processing, 2023, 20(6): 121.

[9] Amsaprabhaa M, Nancy Jane Y, Khanna Nehemiah H. Multimodal spatiotemporal skeletal kinematic gait feature fusion for vision-based fall detection[J]. Expert Systems with Applications, 2023, 212: 118681.

[10] Beddiar D R, Oussalah M, Nini B. Fall detection using body geometry and human pose estimation in video sequences[J]. Journal of Visual Communication and Image Representation, 2022, 82: 103407.

[11] Kan X, Zhu S H, Zhang Y H, et al. A lightweight human fall detection network[J]. Sensors, 2023, 23(22): 9069.

[12] Núñez-Marcos A, Arganda-Carreras I. Transformer-based fall detection in videos[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107937.

[13] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.

[14] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[PP/OL]. V2. arXiv (2024-05-31)[2024-11-06]. https://doi.org/10.48550/arXiv.2312.00752.

[15] Sunkara R, Luo T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects[C]//Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2023: 443-459.

[16] Zhao Q T, Zheng C, Liu M Y, et al. PoseFormerV2: Exploring frequency domain for efficient and robust 3D human pose estimation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 8877-8886.

[17] Dao T, Fu D, Ermon S, et al. FlashAttention: Fast and memory-efficient exact attention with IO-awareness[C]//Advances in Neural Information Processing Systems 35. New Orleans, Louisiana, USA: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022: 16344-16359.

[18] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]//Computer Vision-ECCV 2014. Cham: Springer, 2014: 740-755.

[19] Maji D, Nagori S, Mathew M, et al. YOLO-pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, LA, USA: IEEE, 2022: 2636-2645.

[20] Ionescu C, Papava D, Olaru V, et al. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1325-1339.

[21] Kwolek B, Kepski M. Human fall detection on embedded platform using depth maps and wireless accelerometer[J]. Computer Methods and Programs in Biomedicine, 2014, 117(3): 489-501.

[22] Charfi I, Miteran J, Dubois J, et al. Optimized spatio-temporal descriptors for real-time fall detection: Comparison of support vector machine and AdaBoost-based classification[J]. Journal of Electronic Imaging, 2013, 22(4): 041106.

[23] Tian Z, Chen H, Shen C H. DirectPose: Direct end-to-end multi-person pose estimation[PP/OL]. V2. arXiv (2019-11-24)[2024-11-06]. https://doi.org/10.48550/arXiv.1911.07451.

[24] Mao W A, Tian Z, Wang X L, et al. FCPose: Fully convolutional multi-person pose estimation with dynamic instance-aware convolutions[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 9030-9039.

[25] Shi D H, Wei X, Yu X D, et al. InsPose: Instance-aware networks for single-stage multi-person pose estimation[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3079-3087.

[26] Shi D H, Wei X, Li L Q, et al. End-to-end multi-person pose estimation with transformers[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 11059-11068.

[27] Wang D K, Zhang S L. Contextual instance decoupling for robust multi-person pose estimation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 11050-11058.

[28] Yang J, Zeng A, Liu S, et al. Explicit box detection unifies end-to-end multi-person pose estimation[C]//2023 International Conference on Learning Representations (ICLR). Kigali, Rwanda: IEEE, 2023: 598-607.

[29] McNally W, Vats K, Wong A, et al. Rethinking keypoint representations: Modeling keypoints and poses as objects for multi-person human pose estimation[C]//Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 37-54.

[30] Li W H, Liu H, Tang H, et al. MHFormer: Multi-hypothesis transformer for 3D human pose estimation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 13137-13146.

(下转第 95 页)

Zhang Zhen, Xiao Peng. Application of industrial CT in aviation[J]. Measurement & Control Technology, 2023, 42(2): 1–6(in Chinese).

[6] 汤戈, 赵欣雨, 王宇翔, 等. 工业 CT 技术在地球科学中的应用[J]. CT 理论与应用研究(中英文), 2024, 33(1): 119–134.
Tang Ge, Zhao Xinyu, Wang Yuxiang, et al. Applications of industrial computed tomography technology in the geosciences[J]. CT Theory and Applications, 2024, 33(1): 119–134(in Chinese).

[7] Withers P J, Bouman C, Carmignato S, et al. X–ray computed tomography[J]. Nature Reviews Methods Primers, 2021, 1: 18.

[8] 邹晶, 蒋一鸣, 胡晓东, 等. 光学耦合 CT 系统几何位姿参数标定方法[J]. 光子学报, 2022(4): 051.
Zou Jing, Jiang Yiming, Hu Xiaodong, et al. Calibration method for geometric pose parameters of optically coupled CT system[J]. Acta Photonica Sinica, 2022(4): 051(in Chinese).

[9] Francken N, Sanctorum J, Huyge B, et al. Correcting for focal spot drift in edge illumination X–ray phase contrast imaging[J]. e–Journal of Nondestructive Testing, 2025, 30(2): 1–7.

[10] Messner I M, Keuschnigg P, Stöllinger B, et al. Investigating focal spot position drift in a mobile imaging system equipped with a monobloc–based X–ray generator[J]. Medical Physics, 2024, 51(5): 3578–3589.

[11] Bircher B A, Meli F, Küng A, et al. X–ray source tracking to compensate focal spot drifts for dimensional CT measurements[J]. e–Journal of Nondestructive Testing, 2020, 25(2): e02043 .

[12] Shimakawa Y, Nishiki M, Yanagita S, et al. Evaluation of X–ray effective focal spot size dependency on X–ray exposure settings using edge response analysis[J]. Radiological Physics and Technology, 2023, 16(1): 39–47.

[13] Lillaney P, Shin M, Hinshaw W, et al. Electrostatic focal spot correction for X–ray tubes operating in strong magnetic fields[J]. Medical Physics, 2014, 41(11): 112302.

[14] Obodovskiy A V, Bessonov V B, Larionov I A. Shift focal spot X–ray tube to the imposition anode under long exposure[J]. Journal of Physics: Conference Series, 2018, 967: 012010.

[15] Tilley II S, Zbijewski W, Stayman J W. High–fidelity modeling of shift–variant focal–spot blur for high–resolution CT[C]// The 14th International Meeting on Fully Three–Dimensional Image Reconstruction in Radiology and Nuclear Medicine. Xi′an, China: Fully3D 2017 Proceedings, 2017: 752–759.

[16] Vavřík D, Jandejsek I, Pichotka M. Correction of the X–ray tube spot movement as a tool for improvement of the micro–tomography quality[J]. Journal of Instrumentation, 2016, 11(1): C01029.

[17] Gwerder D, GmbH M, Willner M, et al. Automated drift compensation procedure for high–resolution XCT imaging based on simultaneous measurement of marker sphere[J]. e–Journal of Nondestructive Testing, 2024, 29(3): 29255.

[18] Liu M N, Han Y, Xi X Q, et al. Horizontal drift correction by trajectory of sinogram centroid fitting for laboratory X–ray nanotomography[C]//Proceedings of SPIE Conference on International Conference on Optics and Image Processing (ICOIP 2021). Guilin, China: SPIE, 2021: 127.

[19] Liu M N, Yu H, Xi X Q, et al. Drift correction in laboratory nanocomputed tomography using joint feature correlation [J]. Applied Optics, 2023, 62(11): 2784–2791.

[20] Tilley I S, Zbijewski W, Siewerdsen J H, et al. Modeling shift–variant X–ray focal spot blur for high–resolution flat–panel cone–beam CT[C]//International Conference on Image Formation in X–ray Computed Tomography, 2016: 463–466.

[21] Sun Y, Hou Y, Zhao F Y, et al. A calibration method for misaligned scanner geometry in cone–beam computed tomography[J]. NDT & E International, 2006, 39(6): 499–513.

本文引文格式:

赵金涛, 吕寒玉, 杨晓霞, 等. 光学耦合 CT 系统的焦斑漂移实时监测与补偿方法[J]. 天津工业大学学报, 2026, 45(4): 89–95.
Zhao Jintao, Lü Hanyu, Yang Xiaoxia, et al. Real–time monitoring and compensation method for focal spot drift in optically coupled CT systems[J]. Journal of Tiangong University, 2026, 45(4): 89–95(in Chinese).

(责任编辑:孙广宇)

(上接第 89 页)

[31] Zheng C, Zhu S J, Mendieta M, et al. 3D human pose estimation with spatial and temporal transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 11636–11645.

[32] Egawa R, Miah A S M, Hirooka K, et al. Dynamic fall detection using graph–based spatial temporal convolution and attention network[J]. Electronics, 2023, 12(15): 3234.

[33] Fei K X, Wang C, Zhang J X, et al. Flow–pose Net: An effective two–stream network for fall detection[J]. The Visual Computer, 2023, 39(6): 2305–2320.

[34] Li S Y, Song X, Xu S Y, et al. Dilated spatial–temporal convolutional auto–encoders for human fall detection in surveillance videos[J]. ICT Express, 2023, 9(4): 734–740.

本文引文格式:

李现国, 段显霖, 杨彦利, 等. 基于二维和三维人体姿态估计的跌倒检测[J]. 天津工业大学学报, 2026, 45(4): 80–88, 95.
Li Xianguo, Duan Xianlin, Yang Yanli, et al. Fall detection based on 2D and 3D human pose estimation[J]. Journal of Tiangong University, 2026, 45(4): 80–88, 95(in Chinese).

(责任编辑:孙广宇)