

DOI:10.19651/j.cnki.emt.2313924

基于YOLOv5s和改进质心跟踪的人员跌倒检测*

王新 杨秀梅

(河南理工大学物理与电子信息学院 焦作 454000)

摘要:针对目标检测器检测跌倒时过于依赖卷积网络分类效果、无法利用运动信息的问题,本文设计了一种基于YOLOv5s和改进质心跟踪的跌倒检测模型。为解决耗费资源问题,用MobileNetV3网络和Slim Neck模块对YOLOv5s进行轻量化,同时将MobileNetV3网络中的SE模块替换为更高效的ECA模块,降低网络复杂度的同时保持较高的精度。引入哈希感知算法改进质心跟踪,增加目标关联的依据,提高跌倒检测的准确性。实验结果显示改进YOLOv5s模型大小下降52.2%,计算量下降51.8%,精度高达90.3%。改进质心跟踪的跌倒检测模型准确率提高了4.3%。结果表明了本文提出模型的有效性和优越性。

关键词:跌倒检测;YOLOv5s;质心跟踪;Slim Neck;哈希感知算法

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.60

Personnel fall detection based on YOLOv5s and improved centroid tracking

Wang Xin Yang Xiumei

(School of Physics & Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Aiming at the problem that the object detector relies too much on the classification effect of convolutional network and cannot use motion information when detecting falls, this paper designs a fall detection model based on YOLOv5s and improved centroid tracking. To solve the problem of resource consumption, the MobileNetV3 network and Slim Neck module are used to lightweight YOLOv5s, and the SE module in the MobileNetV3 network is replaced with the more efficient ECA module, which reduces the network complexity while maintaining high accuracy. Hash sensing algorithm is introduced to improve centroid tracking, increase the basis of target association, and improve the accuracy of fall detection. The experimental results show that the size of the improved YOLOv5s model is reduced by 52.2%, the computational capacity is reduced by 51.8%, and the accuracy is as high as 90.3%. The accuracy of fall detection model with improved centroid tracking was increased by 4.3%. The results show the effectiveness and superiority of the proposed model.

Keywords: fall detection; YOLOv5s; centroid tracking; Slim Neck; perceptual Hash algorithm

0 引言

目前大部分国家都面临着人口老龄化挑战。到2050年,老年人的数量预计达到15亿,占全球人口的16%^[1]。跌倒严重威胁高龄人群的生命和健康。当老人跌倒后,发现并提供医疗帮助稍有延迟可能会有致命危险。如何在不影响老人正常行为活动的前提下,采用科学的方法检测人体的跌倒并及时报警救助是当前亟待解决的问题。

随着传感器、网络通信、监控摄像以及人工智能等多领域技术的快速发展,学术界针对人体摔倒检测问题提出了多种解决方法^[2]。主要有基于环境传感器检测法、基于穿

戴传感器检测法和基于机器视觉检测法。其中基于计算机视觉的跌倒检测法由于其低成本、易维护、用户体验好等优点,日渐成为摔倒检测研究的主流方法。在传统的基于计算机视觉的跌倒检测方法中常用手动的方式提取人体轮廓几何特征或者人体加速度运动特征进行跌倒识别^[3],但当区分的类别数目增多,分类和识别问题会变得复杂麻烦甚至难以实现。随着深度学习和神经网络在计算机视觉和行人姿势识别中的发展和设备能力的改善,基于卷积神经网络及基于循环神经网络的人体跌倒检测算法应运而生。

目前已经有许多学者用深度学习的方法对跌倒检测进行了研究,根据深度学习算法的优点将多种算法相结合逐

收稿日期:2023-06-19

* 基金项目:国家重点研发计划(2016YFC0600906)、国家自然科学基金(61403129)项目资助

渐成为主流趋势^[4]。武历展等^[5]用YOLOv5s对人体的跌倒状态进行检测识别,同时对YOLOv5s模型进行优化,提升检测精确度。王鑫等^[6]用MobileNetV3网络得到图像的特征图,根据人体关键点生成3个特征向量,设定跌倒判别特征和阈值来检测跌倒。由于跌倒具有时序性,这些检测方式只能根据静态帧判断目标状态,对于一些与跌倒相似的日常动作,例如躺下、趴下、蹲下等容易造成误判。为了利用时间维度上的运动信息,Adrian等^[7]叠加光流图像作为网络的输入,使用VGG16来分类检测。光流反映了运动物体在图像中的瞬时速度。为了同时利用空间及时间特征,双流网络应运而生。袁智等^[8]通过对视频帧加框标记、计算光流图作为双流网络的输入,使用3D卷积神经网络(convolutional neural networks, CNN)与VGG16网络模型构建双流网络检测跌倒。但是提取光流图像受背景、光照和角度影响比较严重,特征提取难度大,数据处理复杂。金彦亮等^[9]使用帧间差分法提取相邻两帧图像间的人物轮廓信息,再从连续3帧中提取像素点的变化值作为运动信息,最后送入CNN进行分类检测。该类算法过度依赖卷积网络的分类效果,而对于两个相似的动作输入,分类器将很难做出准确判断。

本文为了利用时间维度上的运动信息,运用了质心跟踪来实现跌倒检测模型的设计。质心跟踪基于质心法,由于质心法的计算量小,实时性高,上世纪的美国“小牛”实时导弹就运用了质心法。现在质心跟踪仍应用广泛,例如叶阳等^[10]提出一种基于质心法的车辆轮廓检测与跟踪方法,文中使用了传统的质心跟踪算法,根据区域内所有点的灰度值来求区域内的质心的位置。许多学者对质心跟踪算法进行了改进,例如高璇等^[11]提出加权质心跟踪算法,在实时空中目标跟踪中进行应用,用灰度差异和距离表示像素点对于目标的隶属度,以此为权值进行质心计算。Rahman等^[12]提出一种将目标检测器与质心跟踪结合的跟踪检测违规车辆的算法,将目标检测器检测到的目标框中心点作为质心点,将前后帧的质心点

通过计算欧式距离进行关联,达到跟踪目的。这种低复杂度的质心跟踪算法能更好地适用于经济型嵌入式设备,本文选择此质心跟踪算法来设计跌倒检测模型。

本文主要的工作内容如下:

1) 为了解决过于依赖卷积网络分类效果的问题和利用时间维度上的运动信息,本文以YOLOv5s网络模型为基础,引入质心跟踪模块设计了跌倒检测模型。YOLOv5s对目标进行检测分类,质心跟踪模块关联前后帧目标共同判断,减少模型完全依赖卷积网络的分类效果。

2) 为了解决模型复杂问题,对YOLOv5s进行轻量化,轻量化的同时保持了较高的检测精确度。本文选择用轻量化网络MobileNetV3替换YOLOv5s的特征提取网络,用Slim Neck改进YOLOv5s的特征增强网络。

3) 为了解决质心跟踪出现错误跟踪导致误检问题,本文引入哈希感知算法改进质心跟踪,增加质心移动的欧几里得距离、哈希值相似比例、IOU作为目标关联的依据,对质心跟踪算法进行改进,减少错误跟踪。

1 YOLOv5s 算法及改进

1.1 YOLOv5s 算法

YOLOv5是目前比较优秀的单阶段目标检测网络。由于本文的跌倒检测模型每帧都需要运行一次目标检测器,计算量大的对象检测器在资源受限的设备上帧处理将大大减慢,所以本文选用算法模型的体积最小、单帧处理时间最短的YOLOv5s版本^[13]。网络结构如图1所示。输入的图片首先在Backbone部分进行特征提取,获取了3个有效特征层,输入到Neck部分结合不同尺度的特征信息进行特征融合,使用了Panet的结构,对特征进行上采样实现特征融合后,再进行下采样实现特征融合,获得3个加强过的有效特征层,将这3个特征图看作一个又一个特征点的集合,每一个特征点都有通道数个特征,最后对特征点进行判断,判断特征点是否有物体与其对应。

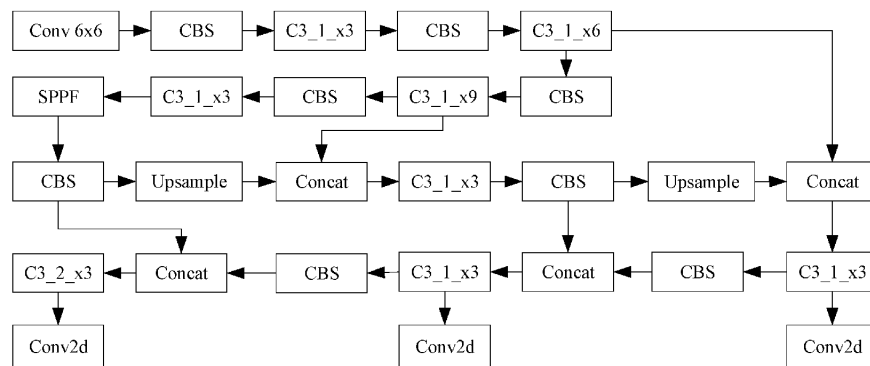


图1 YOLOv5s网络结构

轻量化不可避免的要损失精确度,本文在保持较高精度的情况下对YOLOv5s进行了轻量化。用MobileNetV3对YOLOv5s的Backbone进行改进,与已有

的轻量化网络ShuffleNetV2、Ghostnet、PP-LCNet对YOLOv5s改进的模型对比,精确度更高。本文选择Slim Neck对YOLOv5s的Neck进行改进,Slim Neck中的

GSCConv 模块比标准卷积(the standard convolution, SC)模块计算量小,比深度可分离卷积(depthwise separable convolution, DSC)模块精确度高,在复杂度和精确度之间达到了很好的平衡,满足本文需求。

1.2 基于 MobileNetV3 的特征提取网络改进

针对 YOLOv5s 特征提取网络 CSPDarknet53 存在体积大、计算量大等问题,本文选用轻量级 MobileNetV3^[14] 替换 YOLOv5s 的特征提取网络。舍弃 MobileNetV3 中的最后 3 个网络层:平均层、池化层和 softmax 层,将其替换到 YOLOv5s 的 Backbone 部分进行特征提取。选取 MobileNetV3 中的逐点卷积层以及 MobileNetV3 的最终输出作为 3 个尺度的特征图,进行多尺度预测。在保证模型精确度的基础上,减小模型体积和参数量。

MobileNetV3 特有的 bneck 结构^[15] 如图 2 所示,首先输入通过标准 1×1 的卷积进行升维,然后进行 3×3 深度可分离卷积,提取更多的信息,深度可分离卷积参数量下降比例如式(1)所示。再通过压缩与激励网络(squeeze-and-excitation networks, SE-Net),在 SE-Net 中,将全局平均池化后的特征依次输入激活函数为 ReLU 和激活函数为 h-swish 的两个全连接层得到特征向量。h-swish 函数如式(2)所示。将得到的通道权重进行权重赋值得到输出特征。最后使用 1×1 卷积降维。

$$\frac{D_{in} \times k \times k + D_{in} \times D_{out}}{D_{in} \times k \times k \times D_{out}} = \frac{1}{D_{out}} + \frac{1}{k^2} \quad (1)$$

$$h-swish[x] = x \frac{ReLU6(x+3)}{6} \quad (2)$$

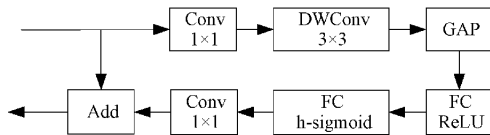


图 2 bneck 结构

针对替换 Backbone 后精确度降低的问题,本文将 Mobilenetv3 中的 SE-Net^[16] 替换为高效的通道注意力(efficient channel attention networks, ECA-Net)^[17],降低精确度损失。SE-Net 中的降维操作在降低模型的复杂度的同时破坏了通道及其权重之间的直接对应关系。ECA-Net 用一维卷积实现了局部跨通道交互,提取通道间的依赖关系,避免降维,降低参数量。首先将输入特征图进行全局平均池化,其次进行卷积核大小为 K 的一维卷积操作,其中 K 通过式(3)来自适应变化。然后经过 Sigmoid 激活函数得到各个通道的权重,如式(4)所示。最后将得到的归一化权重进行权重赋值得到输出特征图。

$$K = \varphi(C) = \left\lfloor \frac{lbC}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (3)$$

$$\omega = \sigma(C1D_k(y)) \quad (4)$$

式中: $C = 2^{\lfloor \gamma \times k - b \rfloor}$, 映射参数 $\gamma = 2$ 和 $b = 1$ 。C1D 表示一维卷积, $y \in \mathbb{R}^c$, k 代表 C1D 的卷积核大小。

1.3 基于 Slim Neck 的特征增强网络的改进

针对卷积神经网络对图像进行特征融合时,每一次空间压缩和通道扩展都会导致语义信息的丢失问题,使用 GSCConv 替代 SC。SC 最大限度地保留了每个通道之间的隐藏连接,DSC 操作则完全切断这些隐藏连接。SC、DSC 和 GSCConv^[18] 的时间复杂度为: $Times_{SC} \sim O(W \cdot H \cdot K_1 \cdot K_2 \cdot C_1 \cdot C_2)$, $Times_{DSC} \sim O(W \cdot H \cdot K_1 \cdot K_2 \cdot 1 \cdot C_2)$, $Times_{GSCConv} \sim O[W \cdot H \cdot K_1 \cdot K_2 \cdot \frac{C_2}{2}(C_1 + 1)]$ 。

为了在精度和速度之间实现了极好的权衡,本文选用 GSCConv 模块对网络进行改进,以较低的时间复杂度尽可能多地保留这些连接。

为了克服 DWConv 计算过程中信道信息分离的缺点, GSCConv 采用通道混洗来融合从 DWConv 和 SC 获得的特征图,使 SC 的输出完全融合到深度卷积中,同时也降低了计算成本。使用一次性聚合方法设计了跨层部分网络模块 VoV-GSCSP,降低计算量和网络结构的复杂性的同时保持了足够的特征增强能力。

在 YOLOv5s 的 Neck 起始端应用 GSCConv 和 VoV-GSCSP,对主干网络提取到的特征层进行增强,使每一个特征层都能同时兼顾深层特征的语义信息和浅层特征的细节信息,提高对目标的特征响应。GSCConv 与 VOV-GSCSP 结构如图 3 所示。

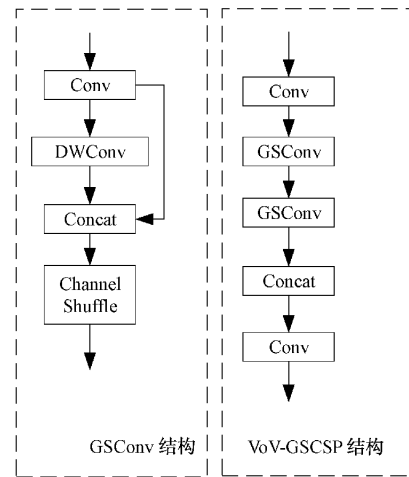


图 3 GSCConv 与 VOV-GSCSP 结构

2 质心跟踪算法及改进

2.1 质心跟踪算法

基于质心的跟踪算法依赖于现有对象质心和视频中后续帧之间的新对象质心之间的欧几里得距离,假设两个相邻帧中欧几里得距离最近的两个质心表示同一个对象,本文中质心为人物边界框的中心点。欧几里得距离^[19] 衡量的是多维空间中两个点之间的绝对距离。欧几里得的二维空间公式为如式(5)所示。

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (5)$$

式中: d 表示两个中心点 (x_i, x_j) 和 (y_i, y_j) 之间的欧几里得距离。

首先 YOLOv5s 将每一帧中检测到的所有人体的坐标信息传递到质心跟踪模块中,根据坐标信息计算边界框的中心 (x, y) 坐标,并给它们分配唯一的 ID,如图 4(a)演示。其次计算每对原始质心(黑色)和新质心(白色)之间

的欧几里得距离,如图 4(b)演示。然后选择具有最小欧几里得距离的新质心与旧质心关联起来,并更新现有对象的 (x, y) 坐标,如图 4(c)演示。当检测到的对象多于正在跟踪的现有对象,需要注册新对象。图 4(d)演示了关联现有对象 ID 并注册新对象的过程。当旧对象无法与任何现有对象匹配时,将注销旧对象。

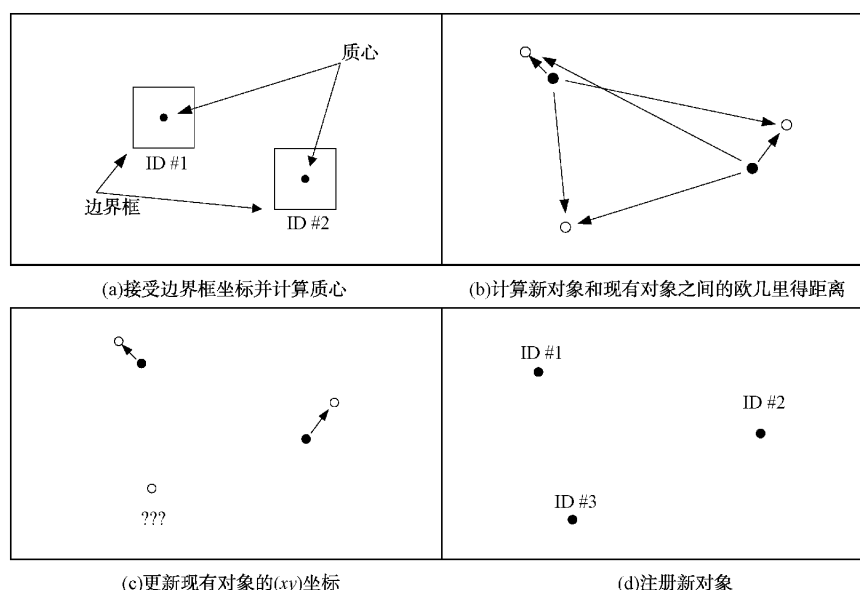


图4 质心跟踪过程

由于质心跟踪严格依赖质心之间的欧几里得距离,没有额外的指标、启发式或学习模式,极易产生错误跟踪。YOLOv5s 结合质心跟踪算法与已有的 YOLOv5s 结合 DeepSort 的跟踪算法对比,本文算法计算量小实时性高,但是精确度不足,所以本文提出一种基于哈希感知的质心跟踪方法,哈希感知算法简单高效,可以快速对前后帧关联目标进行相似度判断,减少错误跟踪。同时增加质心移动的欧几里得距离、IOU 作为目标关联的依据,提高模型的鲁棒性和准确性。

2.2 基于哈希感知的质心跟踪方法

本文跌倒检测系统的设计严格依赖每一帧间的质心移动距离,所以每一帧的跟踪失误都会影响对跌倒的判断。针对此问题,用哈希感知算法^[20]对质心跟踪算法进行改进,提高质心跟踪的准确性。该算法不考虑目标的颜色和形状,使计算简单快速,同时保持一定的准确率,可以有效快速地对两幅图像的相似度进行判断。

感知哈希算法涉及的主要 3 个步骤是图像预处理、感知特征提取和量化或压缩,以生成它们的“指纹”字符串哈希值。图像预处理包括缩小尺寸和简化色彩。将图片缩小到 32×32 的尺寸,快速的去除高频和细节,只保留结构明暗,摒弃不同尺寸、比例带来的图片差异。将缩小后的图像灰度化,进一步简化计算量,灰度转化公式如式(6)所示。在预处理之后,计算 32×32 图像的 DCT,产生 32×32

系数矩阵,其中每个系数由 $C_{ij} (i = 0 \dots 31, j = 0 \dots 31)$ 表示。DCT 变换过程如式(7)所示。然后,选择左上 8×8 低频系数 $C_{ij} (i = 0 \dots 7, j = 0 \dots 7)$ 用于最终哈希值计算。通过与所有 $8 \times 8 = 64$ 个系数的中值 (C_m) 进行比较来量化每个系数,如式(8)所示。最后得到的哈希值是一个 64 位的字符串。

$$L = R \times 0.299 + G \times 0.587 + B \times 0.114 \quad (6)$$

$$F(u, v) = c(u)c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) \times \cos \left[\frac{(j+0.5)\pi v}{N} \right] \quad (7)$$

式中:

$$c(u) = \begin{cases} \sqrt{\frac{1}{N}}u = 0 \\ \sqrt{\frac{2}{N}}u \neq 0 \end{cases}, c(v) = \begin{cases} \sqrt{\frac{1}{N}}v = 0 \\ \sqrt{\frac{2}{N}}v \neq 0 \end{cases}$$

$u = 1, 2, \dots, N-1; v = 1, 2, \dots, N-1; F(u, v)$ 是变化系数矩阵; u, v 为频域采样值; i, j 为空间域采样值。

$$C_{ij} = \begin{cases} 0, & C_{ij} \leq C_m \\ 1, & C_{ij} > C_m \end{cases} \quad (8)$$

给定两个图像 I 和 I' , 它们对应哈希值分别为 $h_I = H(I)$ 和 $h_{I'} = H(I')$, $D(h_I, h_{I'})$ 是相似性度量, τ 是经验确定的阈值, $D(h_I, h_{I'}) < \tau$ 表示 I 和 I' 相似。本文选择汉明距离作为相似性度量。汉明距离表示两个相同长度

字符串对应位置的不同字符的数量。汉明距离计算公式如式(9)。相似比例为 $(HD(H_1, H_2))/64$ 。

$$HD(H_1, H_2) = \sum_{i=0}^{n-1} h_1(i) XOR h_2(i) \quad (9)$$

式中: H_1 和 H_2 分别表示两个哈希值, $H_1 = \{h_1(0), h_1(1), \dots, h_1(n-1)\}$, $H_2 = \{h_2(0), h_2(1), \dots, h_2(n-1)\}$ 。

引入交并比(intersection over union, IOU)、哈希值相似比例与质心移动的欧几里得距离共同作为跟踪的判定依据,对质心跟踪算法进行改进,达到正确跟踪的目的。IOU 计算的是“预测的边框”和“真实的边框”的交叠率,最理想情况是完全重叠。计算如式(10)所示。

$$IOU = \frac{area(B_p \cap B_g)}{area(B_p \cup B_g)} \quad (10)$$

式中: B_p 和 B_g 表示两个目标检测框。

当目标与前帧中欧几里得距离最近的目标锁定后,将两者的欧几里得距离 D_{min} 与欧几里得距离阈值 D 相比较,当 $D_{min} < D$ 时,计算两个目标框内图像的哈希值,并计算出两者的相似比例 M_{min} 与哈希相似比例阈值 M 相比较,当 $M_{min} > M$ 时,计算两个目标框的 IOU 值 N_{min} 并与 IOU 阈值 N 比较,当 $N_{min} > N$ 时,将两个目标相关联,实现跟踪。改进后的整个跟踪算法流程如图 5 所示。

3 实验分析

3.1 数据集制作

本文对 UR Fall Detection Dataset、Le2i Fall detection Dataset 和 Multiple cameras fall dataset 这 3 个数据集的视频数据集进行视频帧截取,人工选取具有代表性的帧共 4 693 张作为本文模型训练的数据集,为了增加数据的可靠性,在网络选取 884 张多人场景下的图像。将数据集按照训练集与测试集 7:3 的比例进行划分。

3.2 实验环境

本文以 YOLOv5s 模型作为实验对象,使用搭载 6 块 NVIDIA GeForce RTX 3090 显卡的 Linux 服务器,实验环境的 Python 版本为 3.7.15,torch 版本为 1.7.1,CUDA 版本为 11.0。

3.3 结果分析

在自建数据集上对比改进 YOLOv5s 与其他单阶段目标检测模型的检测效果。统一设定 epochs 为 300, batch-size 为 8, imgsz 为 640。通过比较改进前后算法的 mAP、FLOPs、模型大小来评价检测模型的整体性能。mAP 表示所有类别平均检测精确度的均值。FLOPs 表示模型计算量,用来衡量模型的复杂。

实验结果如表 1 所示。改进的 YOLOv5s 与原 YOLOv5s 相比, mAP 仅下降 0.7% 且高达 90.3% 的情况下,模型体积下降了 52.2%, 计算量下降 51.8%, 大大降低了模型的复杂度,同时保持了足够的 mAP。改进的 YOLOv5s 与轻量化模型 YOLOv3-tiny 和 YOLOv4-tiny

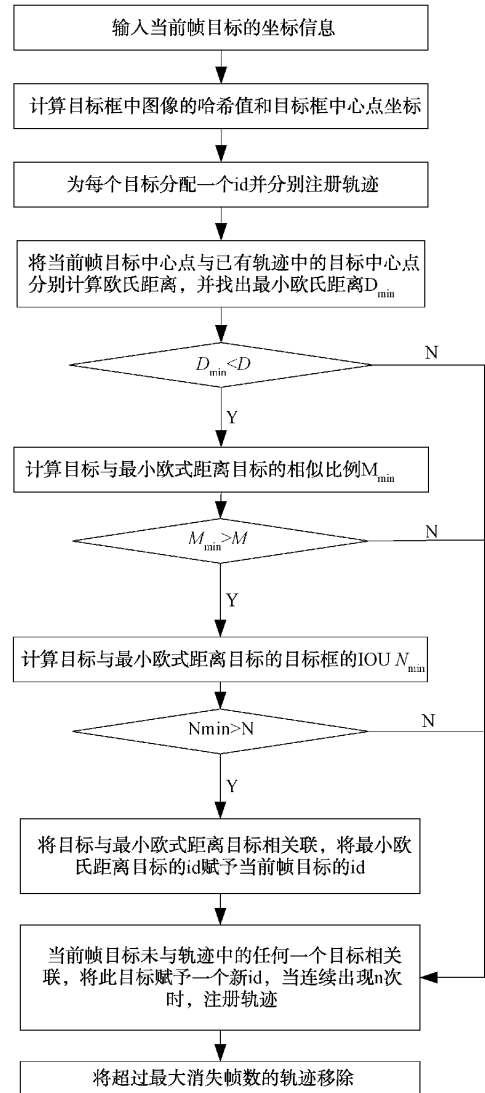


图 5 改进的质心跟踪流程

模型体积小,参数量低,改进的 YOLOv5s 的模型体积和参数量比轻量化模型更低, mAP 反而分别高了 1.2% 和 0.7%。从中可以看出本文改进的 YOLOv5s 模型复杂度低,准确率高,相比较于其它模型具有优越性。

将本文的跌倒检测模型与其他文献的模型在公开数据集 UR Fall Detection Dataset 上进行对比测试。选择该数据集的 cam0 角度的视频图像,包含 70 个视频数据,分别为 30 个跌倒和 40 个日常生活活动视频。对改进质心跟踪的跌倒检测系统阈值进行设置:由于质心移动速度用前后帧距离来表示,所以将最大错过帧数设置为 0。由于目标离摄像头的距离、目标人物高度不同,每一帧质心移动距离 D 与目标的检测框的长 H 和宽 W 相关,根据多次实验结果表示,取 $D = \sqrt{H^2 + W^2}$ 来代表移动速度 V ,可以取得较好效果。注册轨迹连续出现帧数设置为 1, IOU 阈值设置为 0.78, 哈希相似比例阈值设置为 0.68, V 设置为 0.04。

表1 不同目标检测算法性能对比

算法模型	AP%		mAP%	FLOPs/G	模型大小/MB
	fall	person			
YOLOv3-tiny	93.9	84.3	89.1	12.9	16.6
YOLOv4-tiny	94.1	85.2	89.6	16.18	22.4
YOLOv5s	95.4	86.5	91	16	13.6
改进的YOLOv5s	94.9	85.7	90.3	7.7	6.5

跌倒检测中通常使用的敏感度(Sensitivity)、特异度(Specificity),准确率(Accuracy)作为评价标准,具体计算方法如式(11)~(13)所示。

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (13)$$

其中,TP表示跌倒的样本被检测为跌倒的次数。FP表示未跌倒的样本被检测为跌倒的次数。TN表示未跌倒的样本没有被检测为跌倒的次数。FN表示跌倒的样本没有被检测为跌倒的次数。

表2为各个模型在公开数据集上面的跌倒检测效果对比。Harrou等^[21]通过寻找视频兴趣点并计算其光流特征,再设置阈值进行跌倒判断。Abobakr等^[22]使用Kinect采集的深度图像进行特征处理,然后送入Random Forest分类器进行分类。Adrian等^[23]将光流图像作为输入,送入CNN进行跌倒检测。陈京荣^[24]通过姿态估计方法获取人体关键点,然后从中提取用于跌倒检测的新特征,使用Bi-LSTM分类网络进行跌倒行为判断。

表2 不同跌倒检测模型性能对比 %

跌倒检测方法	敏感度	特异度	准确率
光流图像特征+阈值判断 ^[21]	96.7	95	95.7
深度图像特征+随机森林 ^[22]	100	93	95.7
光流图像特征+CNN ^[23]	100	94.8	95
姿态估计+Bi-LSTM ^[24]	100	95	97.1
本文未改进质心跟踪	100	90	94.2
本文改进质心跟踪	100	97.5	98.5

表2从敏感度、特异度和总准确率这3个评价标准将本文改进后的跌倒检测方法与其他跌倒检测方法进行对比。本文改进的模型敏感度已经达到了最高的100%,特异度和准确率都高于其他几种方法。与本文未改进质心跟踪的跌倒模型进行对比,改进后的模型特异度提高了7.5%,准确率提高了4.3%。从表中可以看出,本文改进的跌倒检测模型已经达到比较优秀的水平。

4 结 论

本文提出一种基于YOLOv5s和改进质心跟踪的人员跌倒检测模型,考虑到跌倒是一个具有时序性的动作,运用质心跟踪将前后帧目标关联起来,根据YOLOv5s检测到的人体目标检测框的中心点移动速度以及YOLOv5s的检测结果共同判断跌倒。由于每一帧运行一次YOLOv5s耗费资源,用MobileNetV3替换YOLOv5s的Backbone,并将MobileNetV3中的SE模块替换为ECA模块,用Slim Neck替换YOLOv5s的Neck,降低YOLOv5s计算量和网络复杂性的同时保持足够的精度。由于质心跟踪算法仅依赖欧几里得距离进行跟踪,当出现YOLOv5s错检漏检等情况时容易造成错误跟踪、前后帧间目标质心移动距离计算错误,从而导致跌倒的误判,用哈希感知算法对质心跟踪算法进行改进,用感知哈希算法计算目标间的相似比例,添加相似比例阈值、IOU阈值、欧几里得距离阈值多个判断条件进行目标关联,减少错误跟踪。改进后的跌倒检测系统模型体积减小,参数量下降,误判减少,具有很好的推广性。

参考文献

- [1] UNITED NATIONS. World population ageing 2019[M]. New York: Department of Economic and Social Affairs, 2019.
- [2] 曹梓耕,周辰雨,杨珊,等. 居家老年人跌倒检测技术的应用进展[J]. 护理管理杂志, 2022, 22(9): 644-647, 654.
- [3] BASAVARAJ G M, KUSAGUR A. Vision based surveillance system for detection of human fall[C]. Proceedings of the IEEE International Conference on Recent Trends in Electronics Information & Communication Technology, 2017: 1516-1520.
- [4] 赵珍珍,董彦如,曹慧,等. 老年人跌倒检测算法的研究现状[J]. 计算机工程与应用, 2022, 58(5): 50-65.
- [5] 武历展,王夏黎,张倩,等. 基于优化YOLOv5s的跌倒人物目标检测方法[J]. 图学学报, 2022, 43(5): 791-802.
- [6] 王鑫,郑晓岩,高焕兵,等. 基于卷积神经网络和多判别特征的跌倒检测算法[J]. 计算机辅助设计与图形学学报, 2023, 35(3): 452-462.

- [7] NÚÑEZ-MARCOS A, AZKUNE G, ARGANDA-CARRERAS I. Vision-based fall detection with convolutional neural networks [J]. *Wireless Communications and Mobile Computing*, 2017;1-16.
- [8] 袁智,胡辉.一种基于双流卷积神经网络跌倒识别方法[J].*河南师范大学学报(自然科学版)*,2017,45(3):96-101.
- [9] 金彦亮,陈刚.基于双流卷积的实时跌倒行为检测[J].*计算机工程与设计*,2021,42(9):2621-2626.
- [10] 叶阳,卢奇,程时伟.基于质心法的车联网目标跟踪方法与应用[J].*计算机科学*,2021,48(S2):340-344,359.
- [11] 高璇,邵瑞.改进质心跟踪算法在实时空中目标跟踪中的应用研究[J].*信息化研究*,2014,40(3):36-39.
- [12] RAHMAN Z, AMI A M, ULLAH M A. A real-time wrong-way vehicle detection based on YOLO and centroid tracking [C]. 2020 IEEE Region 10 Symposium(TENSYMP), 2020;916-920.
- [13] 张银胜,杨宇龙,吉茹,等.改进YOLOv5s的风力涡轮机表面缺陷检测[J].*电子测量与仪器学报*,2023,37(1):40-49.
- [14] HOWARD A, PANG R, ADAM H, et al. Searching for mobilenetv3 [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1314-1324.
- [15] 张奇,陈梦蝶,赵杰.融合特征金字塔和通道注意力的轻量车辆检测算法[J].*国外电子测量技术*,2023,42(1):41-48.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.
- [17] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,2020: 11531-11539.
- [18] 孙福临,李振轩,梁允泉,等.基于改进YOLOv5算法和边缘设备的电动车违规载人检测[J].*现代计算机*,2023,29(8):1-11.
- [19] 刘明洋,万九卿.基于深度测量的行人体态特征提取与再识别方法[J].*仪器仪表学报*,2023,44(1):201-211.
- [20] 李文举,王子杰,崔柳.基于SIFT和感知哈希改进的CamShift跟踪算法[J].*电子测量技术*,2023,46(4):184-192.
- [21] HARROU F, ZERROUKI N, SUN Y, et al. Vision-based fall detection system for improving safety of elderly people [J]. *IEEE Instrumentation & Measurement Magazine*,2017,20(6):49-55.
- [22] ABOBAKR A, HOSSNY M, NAHAVANDI S. A skeleton-free fall detection system from depth images using random decision forest [J]. *IEEE Systems Journal*,2017,12(3):2994-3005.
- [23] NÚÑEZ-MARCOS A, AZKUNE G, ARGANDA-CARRERAS I. Vision-based fall detection with convolutional neural networks [J]. *Wireless Communications and Mobile Computing*,2017.
- [24] 陈京荣.基于姿态估计与Bi-LSTM网络的跌倒行为识别[J].*现代计算机*,2021(20):80-85.

作者简介

王新(通信作者),教授,博士生导师,主要研究方向为信号处理,故障诊断。

E-mail:wangxin@hpu.edu.cn

杨秀梅,硕士研究生,主要研究方向为信号处理,目标检测。

E-mail:330242747@qq.com