



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 改进 YOLOv5 的高精度跌倒检测算法
作者: 朱胜豪, 钱承山, 阚希
网络首发日期: 2023-09-21
引用格式: 朱胜豪, 钱承山, 阚希. 改进 YOLOv5 的高精度跌倒检测算法[J/OL]. 计算机工程与应用. <https://link.cnki.net/urlid/11.2127.tp.20230920.1124.036>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

改进 YOLOv5 的高精度跌倒检测算法

朱胜豪¹, 钱承山^{1,2}, 阾希²

1.南京信息工程大学 自动化学院, 南京 211800

2.无锡学院 物联网工程学院, 江苏 无锡 214105

摘要：针对原始 YOLOv5 在人体跌倒检测任务中无法有效应对复杂细节捕捉、变形处理、不同尺度目标适应和遮挡检测的困境，提出了一种基于 C2D 改进 YOLOv5 模型的新型高精度跌倒检测算法 C2D-YOLO。首先，提出了一种名为 C2D 的新型特征提取模块，通过融合可变形卷积、标准卷积和通道空间混合注意机制，将其添加到主干网络中，旨在增强特征表征能力、更好地捕捉复杂细节和处理变形。其次，在颈部网络中，采用了 Swin Transformer Block 替代 C3 模块的瓶颈层，旨在最大限度地保留特征信息，以提升对不同尺度目标的检测精度并改善遮挡情况下的性能。最后，在借鉴 YOLOX 解耦结构的基础上对 YOLOv5 的 Head 模块进行改进，旨在优化分类和回归性能。实验结果表明，相比现有的 YOLOv5s，该方法的 mAP0.5 和 mAP0.5:0.95 分别提高了 3.2% 和 6.5%，明显提升了检测精度，减少了误检率。

关键词：YOLOv5; 跌倒检测; C2D; Swin Transformer Block; 解耦结构

文献标志码：A doi: 10.3778/j.issn.1002-8331.2307-0190

High-precision fall detection algorithm with improved YOLOv5

ZHU Shenghao¹, QIAN Chengshan^{1,2}, KAN Xi²

1. School of Automation, Nanjing University of Information Science & Technology, Nanjing 211800, China

2. School of The Internet of Things Engineering, Wuxi University, Wuxi, Jiangsu 214105, China

Abstract: In order to counter the limitations of the original YOLOv5 human fall detection task, a highly accurate fall detection algorithm, called C2D-YOLO, is proposed in this paper. The original task struggled to effectively handle complex detail capture, deformation handling, target adaptation to different scales, and occlusion detection. To overcome these challenges, several improvements are made to the YOLOv5 model. First, a new feature extraction module called C2D is introduced, which improves feature characterisation, captures complex details, and handles deformations by combining deformable convolution, standard convolution, and channel-space hybrid attention mechanisms. Secondly, in the neck network, Swin Transformer Block replace the bottleneck layer of the C3 module to retain more feature information, thereby improving target detection accuracy at different scales and enhancing performance under occlusion. Finally, the head module of YOLOv5 is enhanced based on the decoupled structure of YOLOX borrowed from YOLOv5 to optimise classification and regression performance. Experimental results show that this method achieves a 3.2% improvement in mAP0.5 and a 6.5% improvement in mAP0.5:0.95 compared to existing YOLOv5s. These improvements significantly increase detection accuracy and reduce false alarm rates.

Key words: YOLOv5; fall detection; C2D ;Swin Transformer Block; decoupled structure

基金项目：国家自然科学基金青年基金项目(42105143); 江苏省高等学校基础学科(自然科学)研究面上项目(580221016); 江苏省研究生实践创新计划(SJCX23_0397)。

作者简介：朱胜豪(1996-),男,硕士研究生,研究方向为深度学习、目标检测、智能终端与物联网应用;钱承山(1971-),男,博士,教授,研究方向为智能终端与物联网应用、计算机视觉;阾希(1987-),通讯作者,男,博士,讲师,CCF 会员,研究方向为模式识别与智能系统、深度学习、计算机视觉,E-mail:kanxi@cwu.edu.cn。

随着经济社会的发展,我国正面临日益突出的人口老龄化问题。据预测,到2025年,60岁以上人口将超过3亿,占总人口的20.7%^[1]。独居老年人口不断增加,其日常安全问题引发了子女和社会的广泛关注。国内研究指出,跌倒已成为意外和非故意死亡的次要原因,尤其是65岁以上人群,跌倒导致的伤亡比例最高^[2]。医学研究显示,及时提供有效的跌倒后治疗能降低死亡风险,提高老年人生存率^[3]。因此,利用先进科学技术开发高效实用的跌倒检测系统为老年人提供安全保障,具有重要的社会 and 实际意义^{[4][5]}。

老年人跌倒检测领域的重要意义在于老年人的跌倒事件往往会导致严重的身体伤害和日常生活能力下降。在跌倒检测方法方面,主要分为三类:基于可穿戴传感器设备的方法、基于环境传感器的方法,以及基于计算机视觉的方法^[6]。基于可穿戴传感器设备的跌倒检测方法采用加速度计、陀螺仪、压力传感器、磁针等多种传感器,测量人体各种运动属性,从而检测出异常情况^[7]。尽管这种方法具有安装简便、便携性强的优势,但是需要老年人长时间佩戴,可能对其日常生活产生一定影响^[8]。基于环境传感器的跌倒检测方法利用压力传感器、雷达、红外线等非视频传感器,监测老年人活动环境中是否发生跌倒事件^[9]。通过分析传感器收集到的数据,可以判断老年人是否经历了跌倒。这种方法的优势在于无需老年人佩戴任何设备,不会对其日常行为造成干扰。然而,该方法容易受到环境因素影响,导致检测精度较低^[10]。基于计算机视觉的跌倒检测系统通常依赖摄像机或其他成像设备,捕捉老年人所在环境的图像或视频。这些系统应用多种算法和技术,对视觉数据进行分析,以便实现跌倒事件的检测。相较于基于可穿戴传感器或环境传感器的方法,基于计算机视觉的跌倒检测方法利用深度学习等技术能够提供更为准确和稳健的解决方案^[11]。

基于深度学习的目标检测算法可分为 One-Stage 和 Two-Stage 两类^[12]。Two-Stage 算法如 Faster R-CNN^[13]、R-FCN^[14]、Mask R-CNN^[15]和 Cascade R-CNN^[16]等,这类算法使用独立的区域选取网络生成候选目标位置,再用另一独立目标检测网络进行分类和边界框精细化,得出最终结果。Two-Stage 算法精度高但速度较慢。One-Stage 算法如 YOLO 系列^[17-20]和 SSD^[21]等,将图像划分为网格单元,直接预测每个单元内的目标类别和边界框,经后处理得最终检测结果。One-Stage 算法速度快,具有较好实时性,尽管可能稍有精度牺牲,但 One-Stage 算法的最新发展大幅提高了其准确性,成为许多实际应用的主流选择。因此,

本研究选取 YOLOv5 进行跌倒检测。

(You Only Look Once version 5, YOLOv5) 是 YOLO 系列的最新代表方法,具备快速、高精度和小体积等显著特点。卷积神经网络结构在演进过程中用了先进的卷积类型(如可变形卷积、转置卷积和扩张卷积)来提升鲁棒性^[22],但在跌倒检测场景下,为应对远距离小目标、人体运动异常、人体遮挡严重等情况,需进一步提高目标检测算法的准确性^[23]。为满足这些场景需求,研究人员可进一步优化 YOLOv5 算法。结合先进卷积技术和流程改进,可提升算法在跌倒检测等特定场景下的性能。Chen T 等人^[24]引入非对称卷积块(Asymmetric Convolution Block, ACB^[25])替代传统基本卷积,增强特征提取能力。在跌倒检测任务中,需考虑目标的空间变换和形状信息,而 ACB 卷积模块主要调整不同位置和通道的特征权重,提升网络表征能力。Li S 等人^[26]提出 MTC-YOLOv5,在 YOLOv5 基础上整合 CA 通道注意力机制(Coordinate attention, CA^[27])和 Transformer^[28]模块,提高小物体检测准确性,减少背景无效信息干扰。CA 注意力机制主要关注通道间关系,对空间位置关注有限,限制了特定任务的性能。Li Y 等人^[29]嵌入 SE(Squeeze-and-Excitation, SE^[30])通道注意力机制,提升特定任务检测效果和准确性,捕捉关键信息 and 提升特征表达能力。SE 注意力机制对空间位置信息有限,主要关注通道间关系,对局部细节和位置变化捕捉较弱。陈彦蓉等人^[31]引入可变形卷积(Deformable Convolutional Networks, DCN^[32])扩大特征图感受野,提高特征识别能力。DCN 通过学习偏移量和权重调整卷积核采样位置,实现自适应空间变换和感受野调整,更好捕捉目标形状和局部细节。胡皓等人^[33]将自注意力机制与卷积模块融合,提升自适应特征融合能力。通过融合自注意力机制和卷积模块,全面捕捉目标图像重要特征,提升感知能力和准确性。自注意力机制和卷积模块融合可在不同层级和尺度上起作用,增强多尺度特征和上下文信息获取,提高目标与背景区分,进一步提高准确性和鲁棒性。

上述方法在人体跌倒检测领域具有重要意义,但也存在以下问题:(1)尽管许多改进算法采用了特殊卷积模块以提高检测精度同时兼顾了目标的适应性和遮挡情况,但这些模块难以确保在目标空间变换和形状信息要求较高时发挥作用。(2)一些模型添加注意力机制以提升精度,但多种注意力机制的有机结合和特定场景的网络设计仍有不足。跌倒检测时对复杂细节的捕捉、变形处理、不同尺度目标的适应性、遮挡

检测以及特定场景任务的准确性和鲁棒性都对模型设计提出了更高要求。

为了实现人体跌倒检测任务下的准确性和鲁棒性要求,提出了一种新的网络:C2D-YOLO。本文的主要贡献如下:

(1) 提出 C2D 模块并将其融入到 YOLOv5 骨干网络中,以增强特征表征能力、更好地捕捉复杂细节和处理目标变形。这一改进旨在提高模型对目标形状和空间变换要求较高的场景的适应性。

(2) 在特征金字塔网络 (Feature Pyramid Networks, FPN^[34]) 的 C3 (CSP Bottleneck with 3 convolutions, C3^[35]) 模块中引入 Swin Transformer Block^[36],以高效提取多尺度特征和上下文信息,从而增强模型的特征融合能力。这一改进旨在改善模型在多尺度目标和遮挡场景下的性能。

(3) 对基于 YOLOv5 模型的检测头进行解耦处理,以降低模型的耦合性、加速模型的收敛速度并提升检测性能。这种解耦结构成功地解决了分类任务和回归任务之间的冲突,使两者能够更好地协同工作。

通过以上改进和整合,C2D-YOLO 在人体跌倒检测任务中能够更好地满足复杂细节的捕捉、目标形状和空间变换的处理、不同尺度目标的适应性以及遮挡情况下的性能要求。

1 基于 C2D 改进的 YOLOv5 跌倒检测算法

提出了一种的高精度人体跌倒检测模型,名为 C2D-YOLO。该模型整合了先进的卷积模块、多种注意力机制融合策略以及高效的解耦头,能够精确地捕捉人体的姿态、动作和细节信息,从而更准确地判断跌倒事件。图 1 展示了 C2D-YOLO 模型的架构,其中红色虚线框表示算法改进的部分。

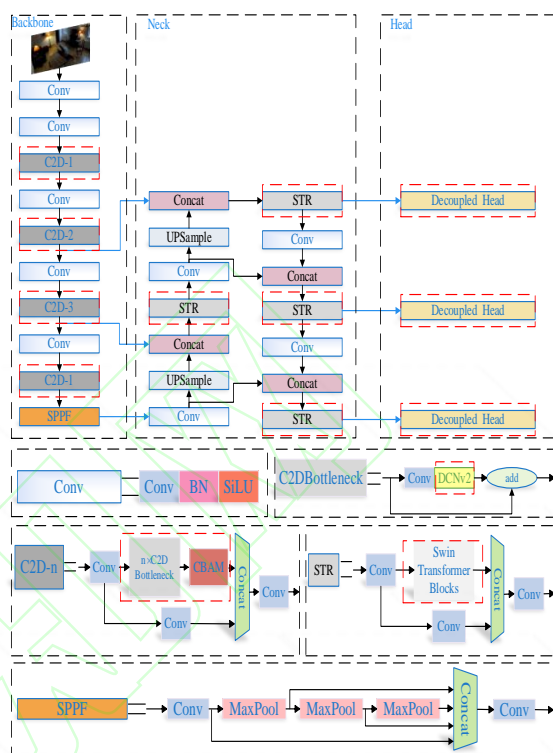


图 1 C2D-YOLO 网络结构图

Fig.1 C2D-YOLO network architecture

C2D-YOLO 模型的改进主要在以下三个方面体现。首先,提出了 C2D 模块,以增强骨干网络的特征提取能力,更好地捕捉复杂细节并处理目标的形变。其次,通过引入 Swin Transformer Block,增强了颈部网络的特征融合能力,从而改善了遮挡检测能力。最后,借鉴 YOLOX 的解耦结构思想,对 YOLOv5 的头部网络进行了改进,进一步优化了分类和回归性能。

1.1 C2D 模块设计

1.1.1 可变形卷积替换标准卷积

跌倒检测任务场景具备独特且复杂的特征,包括广泛的姿态变化、丰富的细节以及目标形变。因此,在特征提取阶段仅使用标准卷积会导致一系列问题。首先,仅采用标准卷积 (Standard Convolution, SC^[37]) 可能未能有效捕捉目标的细粒度细节,尤其是在涉及姿态较大的跌倒动作时,其感受野调整能力有限,导致细节信息的丢失或模糊。其次,标准卷积无法适应目标的形变,无法自适应地调整卷积核的采样位置,导致目标的定位准确性下降,容易产生误检。

为了适应跌倒检测场景的特点并弥补仅使用标准卷积的不足,在特征提取阶段引入了可变形卷积第二版 (Deformable ConvNets version 2, DCNv2^[38])。

DCNv2 通过引入可学习的形变参数,具备更强的感受野调整能力,能够更好地捕捉目标的细粒度细节并适应目标的形变。这种模块的引入填补了标准卷积在跌倒检测中的局限性,提升了特征表示能力和目标定位准确性,从而有效提升跌倒检测系统的性能和鲁棒性。这种优化方法能够更准确地检测和识别跌倒事件,提高整个系统在复杂场景下的可靠性。

可变形卷积 DCN 是一种改进的卷积操作,通过引入偏移量来调整卷积核的形状,以更好地提取输入特征。DCNv2 是对可变形卷积的改进,通过学习偏移和加权,提高了模型从变形物体中提取特征的能力。如图 2 所示,DCNv2 包含两个步骤。首先是偏移量生成:通过卷积操作生成卷积核在输入特征图上沿着 x 和 y 方向的采样点偏移量。其次是采样和卷积:利用输入特征图和计算得到的偏移量进行双线性插值,确定卷积核在输入特征图上的采样点位置。最后,利用这些采样点进行卷积操作。

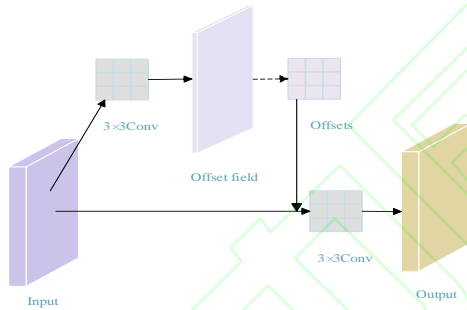


图 2 DCNv2 网络结构图

Fig.2 DCNv2 network architecture

DCNv2 本身对模型的参数数量和计算复杂度没有明显影响,但在实际应用中过度使用调制可变形卷积层会显著增加计算时间。为了在保持模型高效性和有效性之间取得平衡,仅将骨干网络的 C3 瓶颈模块 (Bottleneck) 中的 3×3 标准卷积层替换为 DCNv2 可变形卷积层,就形成了符合要求的瓶颈模块 (C2DBottleneck),如图 3 所示。

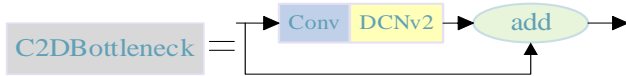


图 3 可变形卷积 DCNv2 替换 3×3 标准卷积 SC

Fig.3 DCNv2 replaces 3×3 SC

1.1.2 CBAM 模块嵌入 C3 网络

在跌倒检测任务中,由于姿态变化、遮挡和背景干扰等因素的存在,目标的特征表示变得更加困难。为了应对这一挑战,CBAM 注意力机制 (Convolutional Block Attention Module, CBAM^[39]) 在图像特征处理中

发挥着重要作用。

CBAM 模块由空间注意力和通道注意力两部分组成,如图 4 所示。空间注意力利用全局平均池化和最大池化捕捉空间特征,通过卷积操作建立空间特征之间的关联性,以增强小目标的空间结构信息。通道注意力则利用多层感知器确定每个通道的注意权重,并将加权的通道信息归一化后融合到原始输入特征图的通道上。这种注意力机制的校准优化了在空间和通道维度上的特征图,使得目标的关键特征更易于被提取。CBAM 模块的引入不仅提升了目标的定位准确性,还使得跌倒检测系统在不同尺度和复杂背景下表现更好。

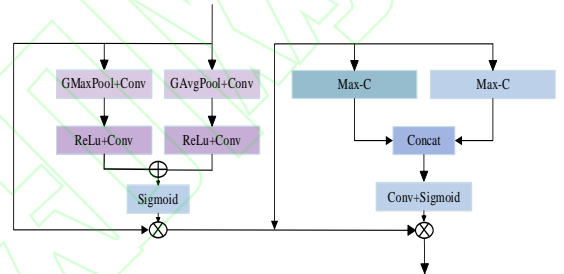


图 4 CBAM 网络结构图

Fig.4 CBAM network architecture

如图 5 所示,为了更好地利用 CBAM 的优势,将 CBAM 嵌入 C3 网络内部。

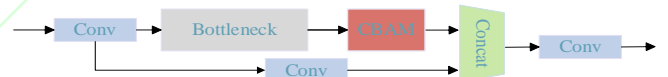


图 5 CBAM 嵌入 C3 内部

Fig.5 CBAM embedded inside C3

1.1.3 融合 DCNv2、SC 和 CBAM

如图 6 所示,提出了 C2D 模块。C2D 模块旨在实现高效的特征提取和目标形变的适应,该模块的设计包括以下关键步骤。首先,C2D 模块采用 1×1 标准卷积和 DCNv2 动态卷积串联的方式,经过批归一化处理 (Batch Normalization, BN^[40]) 和 SiLU 激活函数 (Sigmoid-weighted Linear Unit, SiLU^[41]) 后,再与 1×1 标准卷积进行跳跃连接,形成了新的瓶颈模块 C2DBottleneck。这种设计能够更好地捕捉目标的细粒度细节,并且具备对目标形变的适应能力。其次,在新的瓶颈模块之后,嵌入了 CBAM 注意力机制。该机制旨在更准确地捕捉关键特征,提高目标的定位准确性,并且能够适应不同尺度和复杂背景下的目标检测。最后,把融合了 DCNv2、SC 和 CBAM 的新网络称之为 C2D 模块。

通过将 C2D 模块融合到骨干网络中,可以在较早

的阶段对特征进行增强和选择。这种融合方式使得网络能够更好地捕捉目标的重要特征,从而提高目标检测的性能。

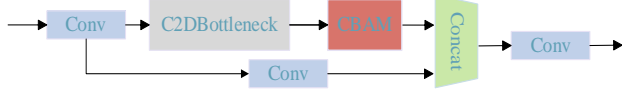


图6 C2D 模块网络结构图

Fig.6 C2D network architecture

1.2 Swin Transformer Block 模块设计

原始的YOLOv5模型在预测小目标时忽略目标与周围上下文特征关系,从而降低不同尺度目标的检测精度。为解决此问题,将原来颈部FPN网络中的C3瓶颈模块替换为Swin Transformer Block,以增强目标检测网络的性能。引入Swin Transformer Block后,模型能够更充分地利用目标与周围环境的关联信息,有效处理小目标和遮挡场景。

Transformer在视觉领域的应用不断发展,Vision Transformer^[42]作为首个引入视觉分类的Transformer模型,在图像检测和分割任务中都具有显著优势。Swin Transformer通过引入局部窗口内的注意力计算并允许跨窗口计算,解决了在Vision Transformer中传统窗口划分策略导致窗口之间缺乏连接性的问题,从而提高了效率并降低了复杂度。图7展示了Swin Transformer的结构,它包括两个基于移动窗口的自注意力机制(Window Multi-head Self-Attention, W-MSA)和(Shifted Window Multi-head Self-Attention, SW-MSA)以及两个多层感知器(Multi-Layer Perceptron, MLP)。为了确保模型的稳定性和准确性,在每个自注意力机制模块和MLP模块之前都添加了层归一化(LayerNorm, LN),并在每个模块之后采用了残差连接。W-MSA采用常规窗口多头自注意力模块,而SW-MSA则采用移位窗口多头自注意力模块。采用移位窗口划分方法,连续的Swin Transformer Blocks计算公式如式(1)至(4)所示。其中, \hat{z}^l 和 z^l 分别是(S)W-MSA和MLP在 l 模块中的特征输出, z^{l-1} 为对应 $l-1$ 层的输出特征。对于公式(1)和(2),输入 z^{l-1} ,将其通过LN层归一化后执行W-MSA操作,再与 z^{l-1} 相加,得到 \hat{z}^l 。进一步地,将 \hat{z}^l 经过LN层归一化后执行MLP操作,再与 \hat{z}^l 相加,得到输出 z^l 。对于公式(3)和(4),输入 z^l ,通过LN层归一化后执行SW-MSA操作,再与 z^l 相加,得到 \hat{z}^{l+1} 。 \hat{z}^{l+1} 经过LN层归一化后执行MLP操作,再与 \hat{z}^{l+1} 相加,最终输出为 z^{l+1} 。

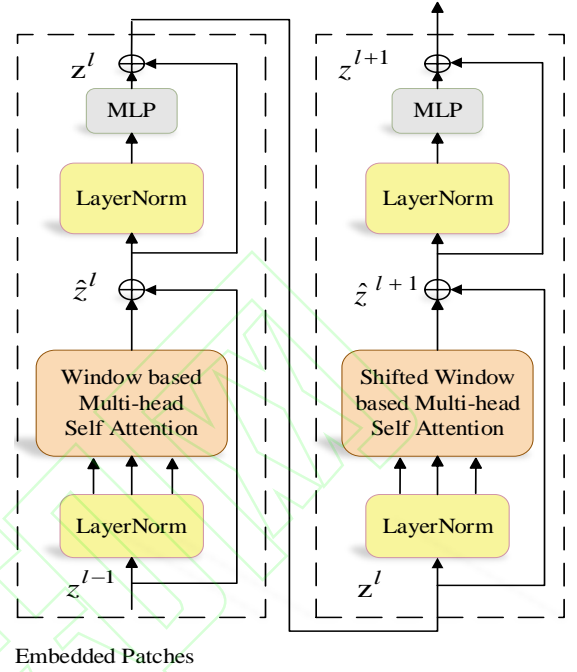


图7 Swin Transformer Block 网络结构图

Fig.7 Swin Transformer Block network architecture

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \quad (1)$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = SW - MSA(LN(z^l)) + z^l \quad (3)$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

1.3 解耦头设计

目标检测任务中,分类和回归任务存在冲突,需要解耦头来解决这个问题^[43]。尽管YOLOv5在主干和特征金字塔方面有所改进,但其检测头仍然是耦合的,可能对整个网络性能造成影响^[44]。因此,对原始YOLOv5的检测头进行解耦处理。如图8所示该解耦头包含一个 1×1 转换层,用于降低信道维度,并包含两个平行分支,每个分支都有两个 3×3 转换层。首先,将骨干网络输出的人体检测特征图作为解耦头的输入,然后使用 1×1 转换层将通道减少到256。接下来,增加了两个平行分支,一个用于人体检测图的分类任务,另一个用于人体检测图的回归任务。此外,在回归分支上还增加了一个IoU分支。通过这种替换,网络能

够更好地处理分类和回归任务之间的冲突,提高网络性能和目标检测的准确性。

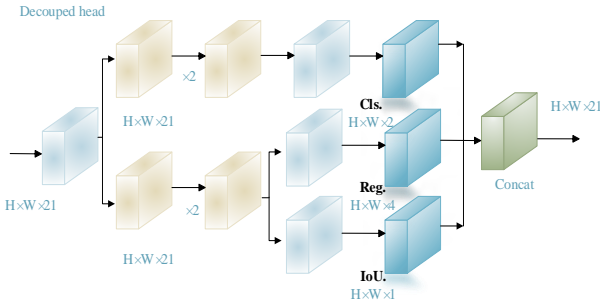


图8 解耦头网络结构图

Fig.8 Decoupled Head network architecture

2 实验与结果分析

2.1 所用数据集

为解决目前跌倒检测领域数据集规模有限、无法涵盖各种情况和变化的问题,综合利用 UR Fall Detection Dataset、Fall detection Dataset、Le2i Fall Detection Dataset 和 AI Studio 平台的 Fall Detection Dataset 等多个数据集。针对性地收集不同光照条件、高速运动场景和不同遮挡场景的数据,并将得到的数据集进行数据清洗和增强。如图9所示,展示了一些典型跌倒场景作为参考,用于评估 C2D-YOLO 在真实情况下的性能。



图9 典型数据集展示

Fig.9 typical dataset display

2.2 实验过程

构建基于 PyTorch 1.13.0 的深度学习框架,对跌倒检测模型进行训练和测试。测试平台的相关配置如下: AMD Ryzen9 5900HX CPU 和 NVIDIA GeForce RTX 3060 GPU,操作系统为 Windows 10,开发环境是 PyCharm 2022.1.2 和 Python 3.8。模型输入图像大小为 640×640 像素。训练参数如下: 批量大小为 32,训练迭代次数为 100,动量为 0.937,初始学习速率为 0.001,衰减系数为 0.9。

原始 YOLOv5s 和 C2D-YOLO 的训练过程使用相

同的数据集和参数,如图10所示,经过10轮迭代,两种模型的损失值下降速度开始趋于平缓,60轮迭代后两种模型损失值误差趋于稳定。实验结果表明,C2D-YOLO 和原模型相比,收敛速度更快,损失值更小,网络收敛能力明显提高。

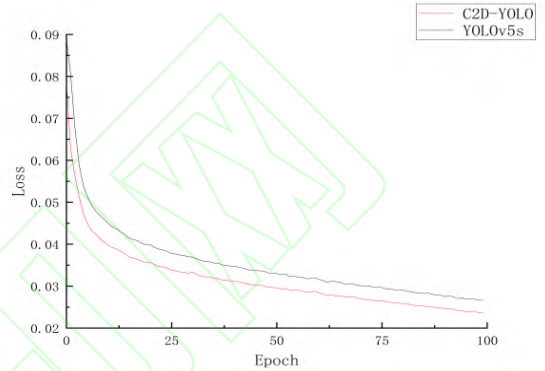


图10 两种模型训练的损失曲线

Fig.10 Loss curves trained by two models

2.3 评价标准

以精确率 (Precision, P) 来衡量所提出的 C2D-YOLO 跌倒检测算法的准确性,召回率 (Recall, R) 测量所有数据集中正样本的检测,平均正确率 (Average precision, AP) 测量检测器在每个类别中的性能,平均精度 (Mean average precision, mAP) 是所有类别 AP 的平均值。P、R、AP 和 mAP 定义为式(5)-(8)。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$AP = \int_0^1 P(R) dR \times 100\% \quad (7)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (8)$$

其中, TP 是正确预测的正样本的数量, TN 是正确预测的负样本的数量, FP 是被划分为正样本的负样本的数量, FN 是被划分为负样本的正样本的数量,并且 k 是类别的数量。

2.4 实验结果和分析

为了验证 C2D-YOLO 模型的性能,设计了两组对比实验。第一组比较原始 YOLOv5s 模型和所有改进后的模型的性能和准确性。第二组比较改进后的模型

与所有 YOLO 系列模型和当前主流物体检测算法的性能和准确性。

2.4.1 消融实验

消融实验的目的是验证各个改进模块的优化效果。在本实验中,对 C2D-YOLO 进行了消融研究。具体而言,改进后的模型代 1 表示在骨干网络中引入了 C2D 模块进行特征提取。模型 2 表示在颈部网络中使用 Swin Transformer Block 替代原来 C3 网络的 Bottleneck 进行特征融合。模型 3 表示对头部网络进行解耦处理。最终的 C2D-YOLO 模型则是在原始 YOLOv5s 网络中综合了上述所有改进方法。实验结果如表 1 所示,每个模块都在不同程度上提高了 C2D-YOLO 的精度。

在使用 C2D 模块改进骨干网络后,精确率、召回率、mAP0.5 和 mAP0.5:0.95 分别提高了 4.3%、2.2%、2.1%和 4.3%。这表明 C2D 模块有效增强了骨干网络的特征提取能力,更为精准地捕捉复杂细节并有效处理目标变形。模型在正样本的识别上展现出更高的准

确性,从而降低了误报率。此外,模型对真实正样本的把握能力也得到了增强,进而提升了检测的综合性性能。这种优化在不同 IoU 阈值下都表现出积极的影响,显著提升了模型在各种场景下的精度。

在颈部网络使用 Swin Transformer Block 后,精确率、召回率、mAP0.5 和 mAP0.5:0.95 分别提高了 3.2%、1.5%、1.4%和 2.7%。这说明在 FPN 网络的 C3 模块中引入 Swin Transformer Block 替换原有的 Bottleneck 增强了颈部网络的自适应融合,有效提取多尺度特征和上下文信息,从而增强模型的感知能力和准确性。

在对头部网络进行解耦处理后,精确率、召回率、mAP0.5 和 mAP0.5:0.95 分别提高了 1.6%、2.1%、0.7%和 1.4%。这表明解耦头的引入成功地解决了分类任务和回归任务之间的冲突,使两者能够更好地协同工作。

从消融实验结果来看,改进的三个模块都提高了整个模型的精度,说明这些改进方法具有良好的耦合性。因此,C2D-YOLO 相对于原始的 YOLOv5s 在性能上有了显著的提升。

表 1 不同设计策略的消融实验

Table 1 Ablation experiments with different design strategies

模型	C2D	STR	Decoupled Head	精确率/%	召回率/%	mAP_0.5/%	mAP_0.5:0.95/%
YOLOv5s	×	×	×	82.5	83.4	88.5	58.8
YOLOv5s_1	√	×	×	86.8	85.6	90.6	63.1
YOLOv5s_2	×	√	×	85.7	84.9	89.9	61.5
YOLOv5s_3	×	×	√	84.1	85.5	89.2	60.2
C2D-YOLO	√	√	√	89.5	87.6	91.7	65.3

2.4.2 对比实验

为了比较不同高性能模型在跌倒检测中的准确性,选择 12 个具有代表性的高精度网络模型 YOLOv 3、YOLOv 4、YOLOv 5s、YOLO5s6、YOLOv5s-ACB(文献 17 所提模型)、YOLOv5s-FPD^[45]、YOLOX-s、YOLOv 7、YOLOv 7-Tiny、YOLOv8s、SSD 和 Cascade R-CNN 与 C2D-YOLO 进行比较和测试,所有模型使用相同的数据集进行训练和测试。选择精准率、召回率、mAP 作为模型精度评价指标。

表 2 显示了不同模型的跌倒检测结果。结果表明,与上述 12 个模型相比,C2D-YOLO 模型的精准率、召回率和 mAP 值最高,分别为 89.5%、87.6%和 91.7%。与上述 12 个网络模型相比,C2D-YOLO 模型的 mAP

分别提高了 7.4%、4.6%、3.2%、1.6%、1.3%、0.8%、6.8%、2.8%、3.6%、0.5%、13.1%和 8%,精准率分别提高了 8.6%、5.6%、7%、4.3%、3.4%、2.4%、9.4%、6.7%、8.3%、1.7%、14.5%和 9.5%,召回率分别提高了 8.2%、5.5%、4.2%、4%、2.3%、1%、8.7%、4.9%、7.7%、0.3%、14.7%和 9.5%。与未经专门针对跌倒检测场景优化的高精度算法(包括最新的 YOLOv8 算法)相比,C2D-YOLO 算法表现出明显的性能优势。那些已经经过跌倒检测场景优化的 YOLO 系列算法,虽然在增强网络的特征表征能力方面做出了改进,但却未充分考虑目标的空间变换和形状信息,这导致它们在应对当前主流基于 YOLOv5 的跌倒检测任务时存在一定的局限性。综上所述,C2D-YOLO 在跌倒检测场景下,相较于诸多主流算法表现出了更优良的精度。

表2 高精度模型跌倒检测结果对比实验

Table 2 Comparative experiments on fall detection results of high-precision models

模型	精确率/%	召回率/%	mAP_0.5/%
YOLOv3	80.9	79.4	84.3
YOLOv4	83.9	82.1	86.9
YOLOv5s	82.5	83.4	88.5
YOLOv5s6	85.2	83.6	90.1
YOLOv5s-ACB	86.1	85.3	90.4
YOLOv5s-FPD	87.9	86.6	90.9
YOLOX-s	80.1	78.9	84.9
YOLOv7	82.8	82.7	88.9
YOLOv7-Tiny	81.2	79.9	88.1
YOLOv8s	87.8	87.3	91.2
SSD	75.0	72.9	78.6
Cascade R-CNN	80.0	78.1	83.7
C2D-YOLO	89.5	87.6	91.7

2.5 场景测试

在多场景下,我们对原始 YOLOv5s 和 C2D-YOLO 算法进行了可视化对比,针对不同的跌倒检测场景,包括光线干扰、高速运动形变、目标遮挡以及复杂中小目标场景,进行了详细的对比展示,如图 11 所示。每组测试图包含三组图,左图展示了原始图片,中间两图展示了原始 YOLOv5s 算法的检测结果以及相应的 GradCam 可视化检测结果,右边两图展示了 C2D-YOLO 算法的检测结果和相应的 GradCam 可视化检测结果。

在图 11a 中,光线干扰导致原始 YOLOv5s 网络未能精确定位人体,出现了漏检现象。然而,引入 C2D 模块后,增强了网络感受野和特征提取能力,使得 C2D-YOLO 能够集中关注人体目标,从而显著提升了预测框的定位准确性,避免了漏检,进一步在跌倒检测任务中取得了更高的预测精度。

在图 11b 中,高速运动形变情况下,原始 YOLOv5s 发生误检,将其他物体误识别为人体。通过引入 DCNv2 模块, C2D-YOLO 能够有效处理高速运动形变,准确识别高速运动中的人体与物体,从而避免了误检情况的发生。

在图 11c 中,原始 YOLOv5s 在目标遮挡时出现了漏检和误检现象,无法正确识别被遮挡的人体。引

入 CBAM 机制嵌入 C3 网络后, C2D-YOLO 增强了网络的自适应融合能力,成功识别被遮挡的目标,避免了漏检和误检,显著提升了跌倒状态的预测准确度。

在图 11d 中,原始 YOLOv5s 在复杂中小目标场景中漏检严重且预测精确度低。Swin Transformer Block 的引入增强了模型提取多尺度特征和上下文信息能力,使得 C2D-YOLO 在复杂场景下能够更准确地识别目标。

经过综合前后对比实验结果的观察,可以明显地看出 C2D 模块在提升模型性能方面取得了显著的改善效果。DCNv2 的引入增强了网络的感受野调整能力,使其更好地捕捉目标的细节和形变情况。特别在高速运动形变的场景下, DCNv2 能够准确识别目标,从而有效避免了误检问题的出现。与此同时, CBAM 嵌入到 C3 网络中,进一步强化了模型对关键特征的注意力,由此提升了预测框的准确定位能力。除此之外,引入 Swin Transformer Block 进一步加强了模型对多尺度特征和上下文信息的提取能力,使得模型在复杂多变的中小尺寸目标场景中表现出色。

综合以上分析,我们明确验证了 C2D-YOLO 算法在各个场景下的显著优势。该算法通过改善预测框定位精度、增强注意力以及强化特征提取、形变处理和自适应融合能力,有效地提升了跌倒检测任务的性能。



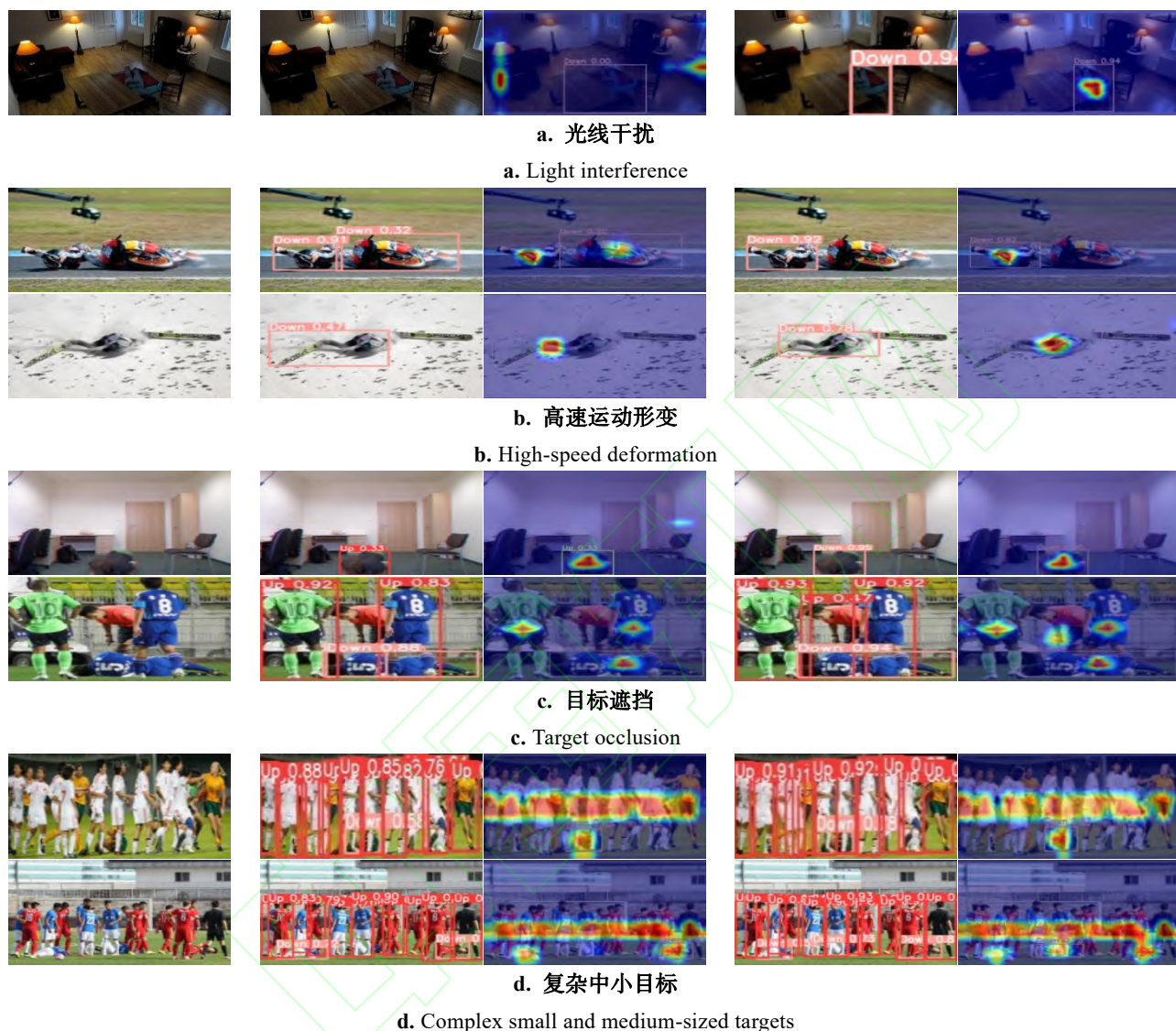


图 11 场景实测

Fig.11 Scene measurement

3 结束语

介绍了一种名为 C2D-YOLO 的新型高精度人体跌倒检测方法,它解决了将传统的 YOLOv5 应用于跌倒检测时所面临的复杂细节捕捉、变形处理、不同尺度目标适应和遮挡检测的困境。本文主要贡献可以概括如下:

(1) 使用 C2D 模块重新设计了 YOLOv5 骨干网络,增强了模型的特征提取能力、使其能够更好地捕捉复杂细节和处理目标变形。

(2) 引入 Swin Transformer Block 优化了颈部网络,增强了模型提取多尺度特征和上下文信息能力,从而提升模型的感知能力和准确性。

(3) 对基于 YOLOv5 模型的检测头进行了解耦

处理,降低了模型的耦合性、加速模型的收敛速度并提升检测性能。

(4) 相比于现有的高精度算法(例如, YOLOv3、YOLOv4、YOLOv5s、YOLOv5s6、YOLOv5s-ACB、YOLOv5s-FPD、YOLOX-s、YOLOv7、YOLOv7-Tiny 和 YOLOv8s 等),本文算法表现更好。

在自制测试集上的实验结果表明, C2D-YOLO 的 mAP 为 91.7%, 精确率为 89.5%, 召回率为 87.6%, mAP_{0.5:0.95} 为 65.3%。与原始的 YOLOv5s 网络相比, C2D-YOLO 将 mAP、精确率、召回率和 mAP_{0.5:0.95} 分别提高了 7%、4.2%、3.2%和 6.5%。综合实验对比表明, C2D-YOLO 具有高效的结构和更高的检测精度,可以满足实时检测需求。与现有模型相比, C2D-YOLO 提供了更高的检测精度以应对各种复杂的跌倒场景。

总之, C2D-YOLO 提高了人体跌倒检测的准确性, 并具有实时跌倒检测和预警的实用价值。未来的研究可以集中于在资源受限的嵌入式跌倒设备上部署 C2D-YOLO, 并进一步完善所提出的算法。

参考文献:

- [1]Chen Y, Liu Z, Huang Y. The aging trend of Chinese population and the prediction of aging population in 2015-2050[J]. Chinese Journal of Social Medicine, 2018, 35(5): 480-483.
- [2]Chen Y, Du R, Luo K, et al. Fall detection system based on real-time pose estimation and SVM[C]//2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE). IEEE, 2021: 990-993.
- [3]Wang X, Ellul J, Azzopardi G. Elderly fall detection systems: A literature survey[J]. Frontiers in Robotics and AI, 2020, 7: 71.
- [4]Ren L, Peng Y. Research of fall detection and fall prevention technologies: A systematic review[J]. IEEE Access, 2019, 7: 77702-77722.
- [5]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.
ZHAO Z Z, DONG Y R, CAO H, et al.Research Status of Elderly Fall Detection Algorithms[J].Computer Engineering and Applications, 2022, 58(5): 50-65.
- [6]Wang X, Ellul J, Azzopardi G. Elderly fall detection systems: A literature survey[J]. Frontiers in Robotics and AI, 2020, 7: 71.
- [7]HU L, WANG S, CHEN Y, et al. Fall detection algorithms based on wearable device: a review[J]. Journal of Zhejiang University (Engineering Science), 2018, 52(9): 1717-1728.
- [8]Er P V, Tan K K. Wearable solution for robust fall detection[M]//Assistive Technology for the Elderly. Academic Press, 2020: 81-105.
- [9]Bhattacharya A, Vaughan R. Deep learning radar design for breathing and fall detection[J]. IEEE sensors journal, 2020, 20(9): 5072-5085.
- [10]Ma L, Liu M, Wang N, et al. Room-level fall detection based on ultra-wideband (UWB) monostatic radar and convolutional long short-term memory (LSTM)[J]. Sensors, 2020, 20(4): 1105.
- [11]Inturi A R, Manikandan V M, Garrapally V. A novel vision-based fall detection scheme using keypoints of human skeleton with long short-term memory network[J]. Arabian Journal for Science and Engineering, 2023, 48(2): 1143- 1155.
- [12]Delgado-Escano R, Castro F M, Cozar J R, et al. A cross-dataset deep learning-based classifier for people fall detection and identification[J]. Computer methods and programs in biomedicine, 2020, 184: 105265.
- [13]Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440- 1448.
- [14]Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in neural information processing systems, 2016, 29.
- [15]He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]// Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [16]Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.
- [17]Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [18]Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [19]Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [20]Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [21]Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multi-box detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [22]Diwan T, Anirudh G, Tembhurne J V. Object detection using YOLO: Challenges, architectural successors, datasets and applications[J]. Multimedia Tools and Applications, 2023, 82(6): 9243-9275.
- [23]Mei X, Zhou X, Xu F, et al. Human Intrusion Detection in Static Hazardous Areas at Construction Sites: Deep Learning-Based Method[J]. Journal of Construction Engineering and Management, 2023, 149(1): 04022142.
- [24]Chen T, Ding Z, Li B. Elderly Fall Detection Based on Improved YOLOv5s Network[J]. IEEE Access, 2022, 10: 91273-91282.

- [25] Ding X, Guo Y, Ding G, et al. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1911-1920.
- [26] Li S, Li K, Qiao Y, et al. A multi-scale cucumber disease detection method in natural scenes based on YOLOv5[J]. Computers and Electronics in Agriculture, 2022, 202: 107363.
- [27] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- [28] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [29] Li Y, Ma R, Zhang R, et al. A tea buds counting method based on YOLOV5 and Kalman filter tracking algorithm[J]. Plant Phenomics, 2023, 5: 0030.
- [30] 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.
- [31] 陈彦蓉,高刃,吴文欢等.[J/OL].电子测量与仪器学报: 1-11[2023-06-01].<http://kns.cnki.net/kcms/detail/11.2488.TN.20230307.1403.010.html>.
CHEN Y R, GAO R, WU W H, et al. Defect detection method for new energy battery collector disc based on improved YOLOv5 network[J/OL]. Journal of Electronic Measurement and Instrumentation: 1-11[2023-06-01]. <http://kns.cnki.net/kcms/detail/11.2488.TN.20230307.1403.010.html>.
- [32] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.
- [33] 胡皓,郭放,刘钊.改进 YOLOX-S 模型的施工现场目标检测[J].计算机科学与探索, 2023, 17(5): 1089-1101.
HU H, GUO F, LIU Z. Object Detection Based on Improved YOLOX-S Model in Construction Sites[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(5): 1089-1101.
- [34] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [35] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.
- [36] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [37] Hu J, Liu B, Peng S. Forecasting salinity time series using RF and ELM approaches coupled with decomposition techniques[J]. Stochastic Environmental Research and Risk Assessment, 2019, 33: 1117-1135.
- [38] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: More deformable, better results[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9308-9316.
- [39] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [40] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448-456.
- [41] Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks, 2018, 107: 3-11.
- [42] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [43] Song G, Liu Y, Wang X. Revisiting the sibling head in object detector[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11563-11572.
- [44] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.
- [45] 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.
WU L Z, WANG X L, ZHANG Q, et al. An object detection method of falling person based on optimized YOLOv5s[J]. Journal of Graphics, 2022, 43(5): 791-802.