

基于改进 YOLOv5s 的跌倒行为检测

朱正林¹, 钱予阳², 马辰宇², 王悦炜², 史 腾³

(1. 南京工程学院 能源与动力工程学院, 南京 211167; 2. 南京工程学院 电力工程学院, 南京 211167;
3. 南京工程学院 机械工程学院, 南京 211167)

摘要: 为了实现电厂人员跌倒行为的实时检测, 防止跌倒昏迷而无法及时发现并救援的事件发生, 针对跌倒行为检测实时性以及特征提取能力不足的问题, 提出了一种改进 YOLOv5s 的跌倒行为检测算法网络: 在 YOLOv5s 模型中引入 SKAttention 注意力模块, 使得网络可以自动地利用对分类有效的感受野捕捉到的信息, 这种新的深层结构允许 CNN 在卷积核心上执行动态选择机制, 从而自适应地调整其感受野的大小; 同时结合 ASFF 自适应空间融合, 并在其中充分利用不同特征, 又在算法中引入权重参数, 以多层次功能为基础, 实现了水下目标识别精度提升的目标; 加入空间金字塔池化结构 SPPFCSPC, 大幅缩短了推理时间; 实验结果表明, 相比于原始 YOLOv5s, 新网络在 *mAP* 平均精度均值方面提升了 2.1%, 查全率提升了 16%; 改进后的网络在感知细节和空间建模方面更加强大, 能够更准确地捕捉到人员跌倒的异常行为, 检测效果有了显著提升。

关键词: SKAttention 注意力模块; 卷积核; ASFF; 权重参数; 空间金字塔池化

Falling Behavior Detection Based on Improved YOLOv5s

ZHU Zhenglin¹, QIAN Yuyang², MA Chenyu², WANG Yuewei², SHI Teng³

(1. School of Energy and Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China;
2. School of Electric Power Engineering, Nanjing Institute of Technology, Nanjing 211167, China;
3. School of Mechanical Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

Abstract: In order to achieve the real-time detection of fall behavior among power plant personnel, and prevent them from falling and falling into a coma, it cannot be detected and rescued in a timely manner, an improved YOLOv5s falling behavior detection algorithm network is proposed to address the issues of insufficient real-time detection and feature extraction capabilities. The introduction of the SKAttention module in the YOLOv5s model enables the network to automatically utilize the information captured by the effective receptive fields for classification. This new deep structure allows the CNN to perform dynamic selection mechanisms on the convolutional core, thereby adaptively adjusting the size of its receptive field; By combining adaptively spatial feature fusion (ASFF) and fully utilizing different features, and introducing weight parameters into the algorithm, based on multi-level functions, the accuracy of underwater target recognition is achieved; The spatial pyramid pooling structure with context sensitive pyramid convolution (SPPFCSPC) is added to greatly reduce inference time. Experimental results show that compared to the original YOLOv5s, the new network improves the mean average precision (*mAP*) by 2.1%, and the recall rate by 16%. The improved network is more powerful in the perception of details and spatial modeling, and can more accurately capture the abnormal falling behaviors, significantly improving a detection effect.

Keywords: SKAttention module; convolutional Kernel; adaptively spatial feature fusion (ASFF); weight parameter; spatial pyramid pooling

0 引言

基于深度学习实现实时性目标检测已经成为计算机视觉领域的研究热点, 该思想在行人检测^[1]、智能医疗、军事^[2]和行动检测^[3]等领域都受到了重视并不断进行发展与应用。现阶段, 目标识别算法中在研究与运用中比较受关注的有区域折叠神经网络 (R-CNN, region-convolutional neural network), 其可以进一步划分为两个类型: 一类为基于区域推荐的两步检测算法, 该类型中较为经典的包括 Faster

R-CNN^[4]等; 另一类为端到端检测算法, 其中较具有代表性的包括 YOLO (You Only Look Once) 系列^[5]、SSD (Single Shot Multi-Box Detector)^[6]。

能够满足跌倒救援等类似的需求场景, 在公共安全方面具有一定价值。Lu 等人^[7]在研究中进行了跌倒行为的提前检测问题的讨论, 其中主要结合长短期记忆网络, 并深度融合了基于帧序列的三维体积, 但其得到的检测结果有着失真等问题。王兰艳^[8]在其关于跌倒行为检测的相关研究中, 结合了 YOLOv3^[9]和 YOLOv4^[10]这两个算法为基础

收稿日期: 2023-10-18; 修回日期: 2023-10-20。

基金项目: 江苏省产学研合作项目 (BY2019013)。

作者简介: 朱正林 (1971-), 男, 博士, 副教授。

通讯作者: 钱予阳 (2000-), 男, 硕士研究生。

引用格式: 朱正林, 钱予阳, 马辰宇, 等. 基于改进 YOLOv5s 的跌倒行为检测[J]. 计算机测量与控制, 2024, 32(10): 26-31, 38.

的优化想法, 设计了灵敏度更高的算法, 并利用 Plus-Yolov5s 检测网络进行检测, 获得了较为满意的检测精度, 能够达到 95.63%。总体而言, 在人类跌倒行为检测领域, YOLO 系列算法体现了其优越性, 因为其在实时性方面的表现, 与其他算法对比, 更符合需求。

在 YOLOv5 中, 在研究者持续研发下已经出现了多种版本, 以模型深度和宽度作为依据进行划分, 包括 YOLOv5l、YOLOv5m、YOLOv5n、YOLOv5 s (you only look once v5s)、YOLOv5x^[11]等不同版本, 将它们与网络进行对比, 则其中最轻型及运算最快的是 YOLOv5s 模型^[12]。本文考虑课题需求, 为了让检测结果更符合需求, 确定在算法模型中以 YOLOv5s 为核心, 因为其具有较多优势, 其网络结构相对简单, 能够提供更快的速度, 同时模型权重在已有的 YOLOv5 版本中最小^[13]。同时还进行了一定改进, 与初始 YOLOv5s 网络相比, 本文在算法模型设计时加入注意力机制 SKAttention^[14]、自适应空间融合 ASFF^[15] (Adaptively Spatial Feature Fusion)、空间金字塔池结构 sppfpc^[16] (Spatial Pyramid Pooling With Context-Sensitive Pyramid Convolution) 等, 从而支持了对跌倒目标的实时跟踪, 并能够取得更好的检测精度。

1 YOLOv5s-SAS 网络及改进策略

1.1 YOLOv5s 网络结构与改进的 YOLOv5s-SAS 网络结构

1.1.1 YOLOv5s 网络结构

YOLOv5 在 2020 年对外公开, 发行者是美国学者 Alexey 及其带领的团队, 作为著名的实施对象识别网络模型, 已经发行了多个不同版本, 如 YOLOv5s、YOLOv5m、YOLOv5n、YOLOv5l、YOLOv5x 等, 其中相对优势较大的版本为 YOLOv5s, 因为其检测速度最快, 且模型网络较为简单, 同时又能够实现模型权重最小^[17], 从而能够作为研究与开发的基础版本, 为其他网络开发与扩大提供支撑。YOLOv5s 网络可根据功能简单划分为: 输入端、特征提取与融合网络、预测端^[18]。

输入端进行图片输入并通过预处理统一缩放或是填充到网络指定大小, 之后再行归一化操作。同时, 输入端还采用 Mosaic 进行数据增强, 既可以丰富训练数据, 又可以降低模型占用内存, 提升训练速度。在图像输入后, 利用 Focus 层对图像进行切片化处理, 提高检测速度并减少参数量。

Backbone 中的 CBS 模块包括 Conv 卷积层、BN (Batch Normalization) 以及 SiLU (Sigmoid Linear Unit) 激活层组成, BN 层具有防止训练过拟合, 加速收敛的作用; SiLU 激活函数也被称为 Swish 函数, 该函数在输出范围接近零时具有更加平滑的曲线, 在一些应用中具有较好的检测效果。CSP1_X 应用于 Backbone 部分, 在较深层网络中提取烟雾的特征信息, 与其相类似的 CSP2_X 则是将其中的 Res unit 残差模块替换为 2 * X 个 CBS 模块, 应用于更浅层的 Neck 部分; SPPF (Spatial Pyramid Pooling-Fast) 是基

于 SPP 的改进空间金字塔池概念。SPP 结构有效减少了图像区域裁剪和缩放等操作引起的图像失真, 解决了图像相关特征的多个提取问题, 这大大提高了所选框的创建速度并降低了计算成本。SPP 的主要功能是在特征图级别集成局部特征和全局特征, 而应用于 YOLOv5 的 SPPF 结构也是借鉴了该思想, 引入了区域池化的概念, 在保留原有功能的情况下进行改进, 实现了更加轻量级的目标, 提高了模型计算速度。

Neck 部分, YOLOv5 主要用到了 PANet (Path Aggregation Network) + FPN (Feature Pyramid Network) 结构, 通过自下而上的路径改进来提高整个特征级别, 并使用准确的低级别位置信号来提高整个特性级别, 从而缩短低级别和顶级特征之间的信息路径^[19]。然而, 网络层数越深, 图像特征信息越明显的同时, 对小目标的信息就越容易造成忽略。通过使用 FPN 提取网络中的分层特征并进行自上而下的采样, 较低的特征图包含更强的语义信息, 这确保了对不同大小图像的准确预测。在 AN 结构中, 强语义特征与强位置特征的传递分别通过 FPN 层和特征金字塔, 且传递形式具有区别, 强语义特征的传递形式为由上而下, 而强位置特征的传递形式为由下而上。两者共同工作以收集来自不同检测层的不同核心层的参数^[20]。

预测端 (Prediction) 的主要特点为应用了 CIUO 损失函数, 同时在预测框生成和目标框改进时结合了多种不同影响因素, 如位置偏移、尺度偏差等, 从而提高预测回归精度。同时还将算法运行中得到的输出特征图进行卷积, 进而得到对应的预测边界框, 从而能够基于分辨尺度进行预测, 达到预期的检测目标^[21]。

1.1.2 YOLOv5s-SAS 网络结构

基于 YOLOv5 6.0 版本改进之后的 YOLOv5s-SAS 结构如图 1 所示。

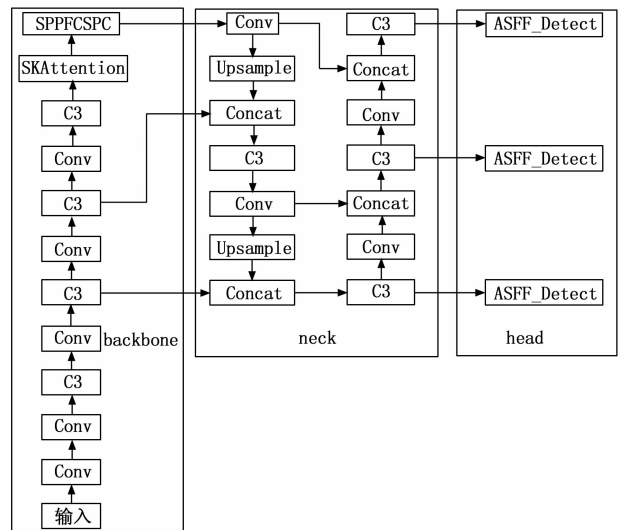


图 1 YOLOv5s-SAS 的网络结构图

首先, 如前所述, 在对 YOLOv5s 模型进行适应性优化过程中, 添加注意力模块 SKAttention, 其作用在于对感受

野捕捉到的信息进行自动追踪和利用,这种结构特性能够为 CNN 卷积核提供优化路径,实现动态选择,丰富了神经元的功能区间,使其能够根据多尺度的输入信息进行动态且即时性的感受野的区间调整;此外,还引入自适应空间融合 ASFF,其支持了权重参数学习,从而能够实现特征融合,并结合层特性,实现算法优化,提高目标识别准确性^[22]。最后,空间金字塔池结构 SPPFCSPC 的加入大大减少了推理时间。

1.2 SKAttention 注意力机制

为了使得网络可以自动地利用对分类有效的感受野捕捉到的信息,提出了一种新的深度结构,这种新的深层结构允许 CNN 在卷积核心上执行动态选择机制,从而自适应地调整其感受野的大小^[23]。称为“选择性核 (Selective Kernel)”,它可以更好地捕捉复杂图像空间的多尺度特征,而不会像一般的 CNN 那样浪费大量的计算资源。SKN 的另一个优点是它可以聚合深度特征,使它更容易理解,同时也允许更好的可解释性。灵感来自这样一个事实,即当我们看到不同大小和距离的物体时,视觉皮层神经元的感受域的大小会随着刺激而调整^[24]。具有不同核大小的几个点以这些点的信息为指导,并与 SoftMax 进行融合。

同时,为了能更好地增加网络的检测性能,现提出一个“选择核”,此 (SK) 卷积可以在原有的基础之上完成自动选择操作,提升了对输入数据提取的能力,使得网络神经元获得在不同的复杂情况之下能够对射频大小自动调整的能力。具体地说,我们通过 3 个运算符——分割、融合和选择来实现 SK 卷积,如图 2 所示。

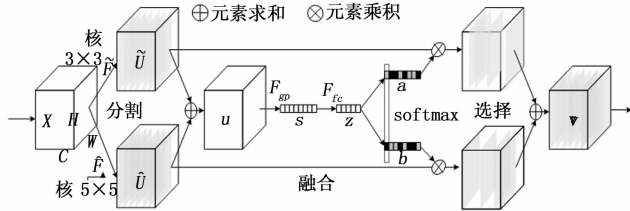


图 2 选择性核卷积

分割:对于任何给定的特征映射 $X \in R^{H \times W \times C}$,默认情况下,首先进行两个转换: $\tilde{F}: X \rightarrow \tilde{U} \in R^{H \times W \times C}$ 和 $\hat{F}: X \rightarrow \hat{U} \in R^{H \times W \times C}$ (其核大小分别为 3 和 5)。二者虽说核大小不同,但是结构都具有批量归一化,并且具有 ReLU 激活函数,获得了该激活函数优异的性能。为了进一步提高效率,将具有 5×5 核的传统卷积替换为具有 3×3 核和膨胀大小为 2 的膨胀卷积。

融合:我们的目标是使神经元能够根据刺激内容自适应地调整 RF 大小^[25]。为了实现这一目标,我们必须在每一点上集成信息,通过元素求和融合来自多个(图 2 中的两个)分支的结果:

$$U = \tilde{U} + \hat{U} \quad (1)$$

对于信息统计方面,我们使用全局平均池化来生成:通过 s 元素中的第 C 个元素,把 U 元素由原来的空间维度

缩小到 $H \times W$,这样更有利于我们在训练过程中获取数据全局信息,提升网络精度:

$$s_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (2)$$

该结构在此通过一个简单的全连接层对相应功能进行实现的,这一操作带来了紧凑的功能 $z \in R^{d \times 1}$,减小了空间维度,使得网络简便,大大增加了网络的学习效率。

$$z = F_{fc}(s) = \delta(B(Ws)) \quad (3)$$

其中: δ 是 ReLU 函数, B 表示批量归一化, $W \in R^{d \times C}$ 。为了研究 d 对模型效率的影响,我们使用折减率 r 来控制其值:

$$d = \max(C/r, L) \quad (4)$$

其中: L 表示 d 的最小值 ($L=32$ 是实验的常规设置)。

选择:跨通道软注意用于在紧凑的特征描述符 z 的指导下自适应地选择信息的不同空间尺度^[26]。具体地,在通道数字上应用 softmax 算子:

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \quad (5)$$

其中: $A, B \in R^{C \times d}$ 和 A, B 分别表示 \tilde{U} 和 \hat{U} 的软注意向量。注意: $A_c \in R^{1 \times d}$ 是 A 的第 c 行, a_c 是 a 的第 c 个元素,同样的 B_c 和 b_c 也是如此。由于 $a_c + b_c = 1$,所以在两个分支的情况下,矩阵 B 是冗余的。最终得到特征图 V :

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c \quad (6)$$

其中: $V = [V_1, V_2, \dots, V_C], V_c \in R^{H \times W}$ 。注意,这里提供了两个分支情况和一个分支情况的公式,可以通过扩展等式 (1) (5) (6) 来容易地推断具有更多分支的情况。

1.3 ASFF 自适应空间融合

在现有的对象识别算法中,最普通、最经典的体系结构为特征金字塔网络 (FPN, feature pyramid network),其特点是分别利用了高级、低级特征的不同特性,前者的语义信息,后者的细化融合。而前文提到的自适应空间特征融合 (ASFF),其实质就是通过一定算法机制综合高级和低级特征特性,从而获得了不同层次的功能。基于上述分析,本文在 YOLOv5s 的适应性改进中融合了 ASFF,利用其特性与优势,提高水下目标识别质量。FPN 详细结构如图 3 所示,ASFF 结构如图 4 所示。

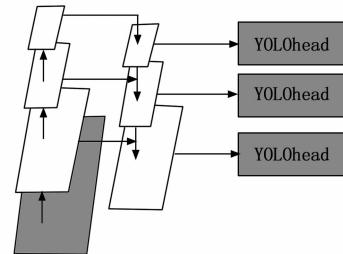


图 3 FPN 结构

ASFF 的核心特性就是需要基于学习参数进行特征过滤,因此在应用该结构时,如果将信息对象的范围限制在当前级别的注意力信息,则基于该结构的算法逻辑,信息

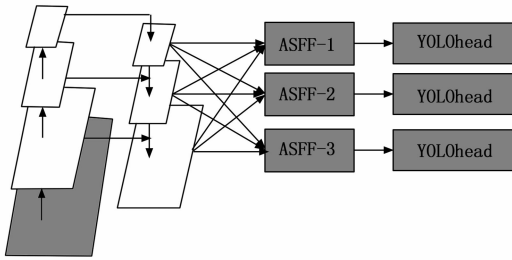


图 4 ASFF 结构

将分层, 从而有助于优化模型学习效率。如 ASFF-1 其中就包含了不同尺度的特征层, 并与各自对应的权重参数 α, β, χ , 相乘, 之后再得将所得信息相加, 可获得如下公式:

$$y_{ij}^1 = \alpha_{ij}^1 \times x_{ij}^{1-1} + \beta_{ij}^1 \times x_{ij}^{2-1} + \chi_{ij}^1 \times x_{ij}^{3-1} \quad (7)$$

式中, ASFF-1 获得的新特征图 y_{ij}^1 是不同层的权重 $\alpha_{ij}^1, \beta_{ij}^1, \chi_{ij}^1$ 和不同特征层 $x_{ij}^{1-1}, x_{ij}^{2-1}, x_{ij}^{3-1}$ 的输出。由于使用了加法, 需要确保 ASFF 层从不同级别获得相同的输出特征维度和通道数。卷积核的大小采用 1×1 相同通道数的卷积层。权重参数 α, β, χ 之和为 1, $[0, 1]$ 的值由归一化函数锁定。

1.4 空间金字塔池化结构 SPPFCSPC

该结构是基于 SPPF 对 SPP 结构的改进, 所以这里先对 SPP 模块结构进行说明。YOLO 算法中的 SPP 结构如图 5 所示。该模块将输入的原特征图按不同分块进行操作, 然后将 3 个子图中每一个框的最大值取出, 完成了最大池化 (Maxpool), 最后再将得出的不同特征图 Concat 叠加在一起。空间金字塔结构 SPP 的最早提出者为何凯, 最初设计该结构的目的是为了解决输入图像的尺寸问题, 在应用中发现其具有 3 个优点: 首先, 不需要提供输入变量, 直接可以根据内部逻辑生成固定长度的输出; 然后, 使用了 Maxpool 的多尺度空间容器; 最后, 可以从不同尺度的特征图中提取特征信息, 有效地提高了检测精度。

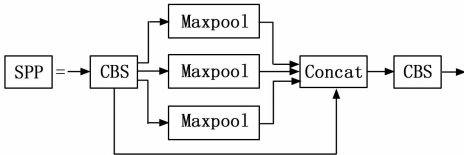


图 5 SPP 结构图

以往图片在输入进网络时需要保持一个固定的大小, 若图片不能满足规定大小尺寸, 网络将无法继续进行学习计算。此时就需要对图片进行裁剪、拉伸等操作, 使图片大小符合规定输入尺寸, 但是这样常常会导致输入图片失真, 改变了输入图片原有的特征, 从而影响网络学习训练的精度。

YOLOv5 6.0 版本中特别加入了 SPPF 结构, 其与 SPP 结构对比, 特殊之处在于应用了若干个较小的池核, 并将它们用级联的方式组合起来, 用以替代较大的池核。从结果来看, SPPF 结构的采用, 实现了多尺度特征图的信息提取, 从而在检测精度上表现更佳。能够在融合功能维持不

变的情况下, 减少 51.6% 的推测时间^[27]。

YOLOv7 版本中的改进之处是引入基于 SPPCPC 的池化结构, 其特殊之处在于, 与原始 SPP 结构对比, 特征图 f_1, f_2, f_3, f_4 的获取过程, 分别需要通过一个核心为 1×1 的卷积通道、两个核心为 3×3 和 1×1 卷积层、SPP 结构、两个核心为 3×3 和 1×1 的卷积通道, 然后在输入端得到的 1×1 卷积后的特征图 f_5 进行融合处理, 得到的结果再进行 1×1 卷积, 输出的结果就是该结构运营的最终值。SPPCSPC 模块分为两个部分, 小残差结构正常通过 SPP 模块, 另一模块经过常规操作, 最终二者通过 concat 叠加。这样的结构使得网络在具有 SPP 模块优势的基础上, 对图片原有的计算量进行调节, 提取出更多有用的特征信息, 检测精度上升。

本文中引入的 SPPFCSPC 结构, 其最早提出者为迪菲赫尔曼, 其具体框架如图 6 所示, 从实质上解析该机构, 其是基于 SPPF 对 SPP 的优化版本, 同时还将 YOLOv7 中的该结构的最大池化部分进行了一定调整, 其中采用了多个小卷积核的级联处理方式。实验证实, 如果参数集保持不变, 推理时间将减少到 SPPCPC 的 72.7%。

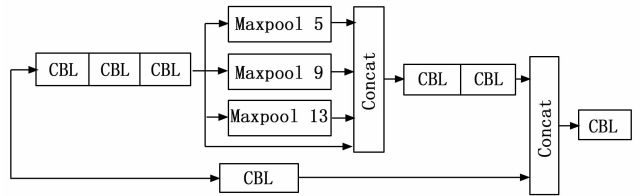


图 6 SPPFCSPC 结构示意图

2 平台搭建与参数处理

2.1 数据集的处理

本实验使用的自建数据集主要为在不同场景下人员摔倒的图片, 共包含 1 443 张不同场景与不同尺度的图像。对于准备好的数据集, 利用了 LabelImg 软件的功能对其按照标签为 fall 进行手工标注生成相对应的 xml 文件, 并将数据集以 9 : 1 的比例进行训练集和验证集划分。

2.2 实验平台搭建

本文的实验训练平台配置如表 1 所示。

表 1 平台参数配置

平台	型号参数
操作系统	Ubuntu16.04
CPU	AMD Ryzen ThreadRipper 3970X
GPU	NVIDIA GeForce RTX 3090 显卡 * 3
显存	24 G
编程环境	Python
GPU 加速库	CUDA CUDNN

2.3 参数设置及评价指标

模型训练时输的图片大小 (Img-Size) 设定为 640×640 , 批次数量 (Batch-Size) 设定为 16, 同时调用 3 台

GPU 进行训练，轮次 (Epochs) 设定为 300 轮，初始学习率设定为 0.01。当每轮训练结束后，利用验证集对模型的训练情况进行检测精度测试，依据评价指标来确定最佳的训练模型^[28]。

本文的模型精度使用查准率 (Precision)，查全率 (Recall) 及平均精度均值 (Mean Average Precision) 来进行评估。其中评价指标的计算需要用到以下公式：

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_i + 1) \quad (10)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (11)$$

其中： n 表示样本类别数目；对 $p(r)$ 函数积分表示 P-R 曲线所围成的面积，亦即 AP (Average Precision)， mAP 的计算则是在此基础上对所有类别的 AP 进行平均；TP (True Positives) 表示被正确识别为正样本的目标数，FP (False Positives) 表示被错误识别为正样本的目标数，FN (False Negatives) 表示正样本被识别为负样本的目标数。由于 Precision 和 Recall 两者之间处于此消彼长的状态，其中一方的增高就意味着另一方的降低，因此一般都会选取平均精度均值 (mAP) 来进行模型评估。

3 实验结果评估分析

3.1 目标检测算法对比

为进一步证明本文算法的有效性，本文同时将其与经典网络模型如 SSD、YOLOv3、YOLOv4、Mask RCNN 与 EfficientDet-D2 网络，分别进行横向比较，实验结果如表 2 所示。

表 2 主流目标检测算法对比

算法	P/%	R/%	mAP/%
SSD ^[29]	93.1	90.8	92.2
YOLOv3 ^[29]	94.0	90.4	90.5
YOLOv4 ^[29]	96.2	92.8	93.3
YOLOv5s ^[29]	97.0	92.2	95.4
Mask RCNN ^[30]	74.2	70.9	73.3
EfficientDet-D2 ^[8]	90.2	74.2	88.4

由表 2 可以看出，相比于经典的 SSD、YOLOv3、YOLOv4Mask RCNN 与 EfficientDet-D2 网络，YOLOv5s 不论是在模型大小程度，还是在查准率和查全率方面都具有相当大的检测优势，因而选取 YOLOv5s 进行改进具有一定的实际价值。

3.2 注意力机制实验

为了证明更换 SKAttention 注意力机制的有效性，在 YOLOv5s 中除了使用 SKAttention 注意力机制，还分别使用 SE、CBAM、SimAM、S2-MLPv2、NAMAttention 种

注意力机制进行对比，根据最终得到的实验数据进行评价，改进注意力实验结果如表 3 所示。由表 3 看出，SKAttention 注意力机制比原网络 mAP 提升了 1.8%，提升效果明显。CBAM、SimAM 注意力在各方面性能都不如 SKAttention，SKAttention 在 mAP 方面分别高于 SE、CBAM、SimAM、S2-MLPv2、NAMAttention 注意力 1.3、3.5、1.5、2.9、4.2%，具有明显优越性。同时，SKAttention 注意力机制的检测速度的优势十分明显。

表 3 注意力机制对比实验结果

算法	P/%	R/%	mAP/%	检测速度/FPS
YOLOv5s	92.6	80.0	89.3	227.272
YOLOv5s+SE	87.9	84.8	89.8	204.082
YOLOv5s+CBAM	86.5	80.3	87.6	370.37
YOLOv5s+SimAM	88.6	82.4	89.6	370.37
YOLOv5s+S2-MLPv2	90.0	78.3	88.2	204.082
YOLOv5s+NAMAttention	90.4	81.0	86.9	153.846
YOLOv5s+SKAttention	87.3	92.0	91.1	131.578

3.3 消融实验

由表 3 可以看出，YOLOv5s 算法作为对比基准，虽然在查准率方面表现较好，但在查全率和 mAP 方面相对较低。添加了 SKAttention 注意力机制之后，在查全率上有了显著提升 (92%)，但查准率有下降 (87.3%)，整体 mAP 为 91.1%；将 ASFF 自适应空间融合应用于 YOLOv5s，相较于基准算法，查准率有轻微提升 (93.5%)，查全率有较大提升 (93%)， mAP 值为 91.6%；在空间金字塔池化结构中用 SPPFCSPC 代替 SPPF 后，在查全率上有了显著提升 (96%)，但查准率有下降 (83.5%)，整体 mAP 为 91.9%；同时结合 SKAttention 注意力机制和空间金字塔池化 SPPFCSPC 后，查准率、查全率与 mAP 值都有一定提升；在前述改进基础上，再加入 ASFF 自适应空间融合，查全率和 mAP 值均有显著提升，同时模型的大小也基本不变，因而本文提出的改进算法在保持一定轻量化的同时能够有效地提高精度。在检测速度方面，与作为基准的 YOLOv5s 算法相比，YOLOv5s-SAS 算法的检测速度也有所提升。

表 4 消融实验结果

算法	查准率 /%	查全率 /%	mAP@0.5/%	检测速度 /FPS
YOLOv5s	92.6	80.0	89.3	227.272
YOLOv5s+SKAttention	87.3	92.0	91.1	131.578
YOLOv5s+ASFF	93.5	93.0	91.6	204.082
YOLOv5s+SPPFCSPC	83.5	96.0	91.9	344.828
YOLOv5s+SKAttention+ASFF	81.5	92	91.6	181.818
YOLOv5s+SKAttention+SPPFCSPC	95.4	92.0	90.5	175.439
YOLOv5s+ASFF+SPPFCSPC	90.4	94.0	92.6	181.818
YOLOv5s-SAS	88.6	96.0	91.4	201.141

为了更加直观地感受到实际模型在检测过程中的现场效果, 选取了部分验证图片进行模型改进前后对照。对比如图 7 所示。



(a) 原 YOLOv5s 检测效果 (b) YOLOv5s-SAS 检测效果

图 7 YOLOv5s 网络与 YOLOv5s-SAS 网络效果对比图

可以看出在原始网络上, 对于人员跌倒的识别置信度较低, 对于多人同时跌倒目标容易造成漏检, 同时对于跌倒人员身体重叠也会产生漏检行为。在改进之后网络在出现以上问题的图像识别能力均有明显的提高。因而该改进网络在感知细节和空间建模方面更加强大, 能够更准确地捕捉到人员跌倒的异常行为, 从而有效地解决了跌倒行为检测实时性以及特征提取能力不足的问题, 防止跌倒昏迷而无法及时发现并救援的事件发生, 保障了电厂的人员安全。这进一步证明了改进网络在人员跌倒检测任务中的优越性。

4 结束语

本文分析主要是立足实时监测个人跌倒行为的相关研究, 对人群中跌倒行为检测中出现的一些问题进行了阐述。基于 YOLOv5s 网络, 参考已有成果, 对其进行了优化和改进, 在其中加入注意力模块, 从而提高了感知分类信息的利用质量和效率, 这种结构特性能够为 CNN 卷积核提供优化路径, 实现动态选择, 丰富了神经元的功能区间, 使其能够根据多尺度的输入信息进行动态且即时性的感受野的区间调整; 此外, 也加入了自适应空间融合 ASFF, 其支持了权重参数学习, 因此能够实现多特征融合, 并结合层次函数, 实现算法优化, 提高目标识别准确性; 从本文实验结果来看, 利用改进的 YOLOv5s, 进行设计和开发的跌倒检测算法在实验中得到了正向反馈, 在查准率、查全率和平均精度均值等维度上都有较好表现。本文之后将完善数据集, 并不断加入新检测场景, 以继续对该算法进行进一步完善。

参考文献:

[1] 刘彦博, 贾瑞生, 徐志峰. 基于尺度空间金字塔网络的人群计

数算法 [J]. 中国科技论文, 2021, 16 (3): 276-280.

- [2] 栗俊杰, 毛鹏军, 方 骞, 等. 基于 K210 的火点检测与喷头定向控制 [J]. 科学技术与工程, 2021, 21 (35): 15136-15143.
- [3] 刘 悦, 张 雷, 辛 山, 等. 融入时空注意力机制的深度学习网络视频动作分类 [J]. 中国科技论文, 2022, 17 (3): 281-287.
- [4] 李 波, 杨 慧, 石 实, 等. 基于改进 Faster R-CNN 的吸烟检测算法研究 [J]. 现代信息科技, 2023, 7 (15): 69-72.
- [5] 邓亚平, 李迎江. YOLO 算法及其在自动驾驶场景中目标检测研究综述 [J/OL]. 计算机应用: 1-12 [2023-09-05].
- [6] 葛海波, 李 强, 周 婷, 等. 改进 SSD 特征融合的目标检测算法研究 [J]. 计算机工程与应用, 2023, 59 (22): 193-201.
- [7] 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.
- [8] 王兰燕. 面向办案区倒地行为检测识别的深度学习算法研究 [D]. 西安: 西安电子科技大学, 2021.
- [9] 卢志博, 徐澄宇, 杨 罡, 等. 基于改进 YOLOv3 的输电线路部件实时检测 [J]. 电测与仪表, 2023, 60 (7): 138-144.
- [10] 郝新泽, 施一萍, 邓 源, 等. 基于轻量化 YOLOv4 的火灾检测识别算法 [J]. 传感器与微系统, 2023, 42 (8): 143-147.
- [11] 王浩臣, 辛月兰, 盛 月, 等. 基于改进 YOLOv5x 的遥感图像目标检测算法 [J]. 激光杂志, 2024, 45 (2): 95-100.
- [12] 王悦炜, 焦良葆, 高 阳. 基于改进 YOLOv5s 的遥感目标检测 [J]. 计算机测量与控制, 2023, 31 (8): 70-76.
- [13] 陈 扬, 吕艳辉. 基于改进 YOLOv5s 的头盔佩戴检测算法 [J]. 沈阳理工大学学报, 2023, 42 (5): 11-17.
- [14] 鲁 鑫. 基于改进的 YOLOX 的中药材检测算法研究 [D]. 荆州: 长江大学, 2023.
- [15] 张陆峰, 吕 青, 张秋菊. 基于改进 YOLOX 的 2D 激光雷达行人检测技术研究 [J]. 应用激光, 2023, 43 (8): 112-121.
- [16] 李晓松. 基于改进 YOLOv5s 的健身动作检测与应用研究 [D]. 广州: 广州大学, 2023.
- [17] 农 健, 陈伟业. 基于视频对象和深度学习的异常行为分析系统设计和实现 [J]. 长江信息通信, 2022, 35 (3): 127-129.
- [18] 王 彝, 江 灏, 陈 静. 改进的前端轻量级网络工业手套缺陷检测研究 [J]. 福建电脑, 2023, 39 (5): 16-20.
- [19] 周 涛, 刘 超. 基于 DPPC-YOLO 的交通标志检测算法研究 [J]. 软件导刊, 2022, 21 (12): 26-32.
- [20] 谷玉海, 曹梦婷, 修嘉芸, 等. 基于 YOLOv4 网络的违章行为检测算法 [J]. 重庆理工大学学报 (自然科学), 2021, 35 (8): 114-121.

(下转第 38 页)