

改进YOLO11的工人安全帽与反光衣检测算法

谢敏怡¹, 蒋作², 赵周洲³, 万学俊³, 潘文林²

(1. 云南民族大学电气信息工程学院; 2. 云南民族大学数学与计算机科学学院, 云南昆明 650500;
3. 玉溪市公安局科技信息化支队, 云南玉溪 653100)

摘要: 针对工人防护装备检测种类单一且检测精度低的问题, 提出一种改进YOLO11的安全帽与反光衣防护装备检测算法。首先, 在骨干网络中使用ASDCN模块提取特征, 通过并行的非对称卷积和无信息损失的下采样, 以提高特征提取能力; 其次, 结合多样化分支块与分块注意力的DBPA-C3k2, 以提升网络特征提取和融合能力, 从而改善对遮挡小目标的检测; 再次, 新增P2层检测头以有效加强对小目标防护装备的检测能力; 最后, 在训练时采用FI-WIoUv2作为边界框回归损失函数, 以优化参数并提高回归精度。实验结果表明, 改进后的YOLO11算法较原算法在精度mAP上有显著提升, 其值达到了90.6%, 并且相较于其他主流算法其mAP更高, 能够更加精确地检测遮挡小目标, 并提高检测精度。

关键词: YOLO11; 安全帽; 反光衣; ASDCN; DBPA-C3k2; P2检测头; FI-WIoUv2

DOI: 10.11907/rjtk.251139

扫描二维码阅读全文:

中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2026)006-0210-11



An Improved YOLOv11 Algorithm for Safety Helmet and Reflective Vest Detection in Workers

XIE Minyi¹, JIANG Zuo², ZHAO Zhouzhou³, WAN Xuejun³, PAN Wenlin²

(1. School of Electrical & Information Engineering, Yunnan Minzu University; 2. School of Mathematics & Computer Science, Yunnan Minzu University, Kunming 650500, China; 3. Science & Technology Informatization Squad, Yuxi Public Security Bureau, Yuxi 653100, China)

Abstract: To address the issues of limited detection categories and low accuracy in workers' protective equipment detection, an improved YOLO11 algorithm for detecting safety helmets and reflective vests is proposed. Firstly, an ASDCN module is employed in the backbone network to enhance feature extraction capability through parallel asymmetric convolutions and information-preserving downsampling. Secondly, the DBPA-C3k2 module, which combines diverse branch blocks and patch-aware attention, is introduced to improve feature extraction and fusion capabilities, particularly enhancing the detection of occluded small targets. Furthermore, a P2 layer detection head is added to strengthen the detection performance for small protective equipment. Finally, during training, the FI-WIoUv2 bounding box regression loss function is employed to optimize parameters and improve regression accuracy. Experimental results demonstrate that the improved YOLO11 algorithm achieves a significant increase in mAP, reaching 90.6%. Compared with other mainstream algorithms, it attains the highest mAP and exhibits superior performance in detecting occluded small targets while maintaining detection accuracy.

Key Words: YOLO11; helmet; reflective vest; ASDCN; DBPA-C3k2; P2 detection head; FI-WIoUv2

0 引言

建筑工人在工地作业时因违章操作易发生安全事故,

其中尤以未穿戴防护装备如反光衣和安全帽的情况较为常见。安全帽和反光衣在工人从事生产活动时可起到预防事故发生的关键作用, 佩戴安全帽能有效保护工人头部免受伤害, 穿着反光衣可使人更醒目, 从而防止工地交通

收稿日期: 2025-04-09

作者简介: 谢敏怡(2000-), 男, 云南民族大学电气信息工程学院硕士研究生, 研究方向为目标检测; 蒋作(1979-), 男, 博士, 云南民族大学数学与计算机科学学院副教授、硕士生导师, 研究方向为深度学习、软件工程; 赵周洲(1987-), 男, 玉溪市公安局科技信息化支队警务技术三级主管, 研究方向为视频监控建设、视频结构化和人像识别算法; 万学俊(1984-), 男, 玉溪市公安局科技信息化支队民警, 研究方向为图像超分辨率重建、大数据和人工智能; 潘文林(1972-), 男, 博士, 云南民族大学数学与计算机科学学院教授、硕士生导师, 研究方向为智能计算、软件工程。本文通讯作者: 蒋作。

事故发生,从各方面保障工人安全。由近几年全国建筑施工安全生产事故情况分析可知,由于高处坠落和物体打击导致事故发生分别占比为52.41%和14.96%,而佩戴安全帽可在坠落或受到打击时有效保护头部^[1];车辆伤害造成事故发生占比为2.11%,但工地车辆基本都是大型渣土车和挖掘机,作业时噪音较大,一旦司机粗心大意撞到工人往往会造成严重伤害甚至死亡,而穿着反光衣可提醒司机谨慎驾驶。因此,在此类工地作业时穿戴反光衣和安全帽尤为必要。当前安全管理主要依靠人工现场巡查和观看视频监控,存在费时费力、人工成本高的问题,且人长时间巡查后注意力下降容易造成错检,埋下安全隐患。

随着深度学习技术的迅猛发展和硬件设备GPU的不断迭代升级,基于深度学习的目标检测已广泛应用于各领域^[2]。因此,将其引入智慧工地,可实现防护装备的实时精确检测,减少工地安全事故发生并降低人工成本。目前,常用的目标检测算法大致可分为两类:一类是经典的双阶段检测算法R-CNN系列^[3],由于其先对物体区域进行定位,再作分类,故为双阶段且检测时间较长;另一类是目前流行的单阶段检测算法,代表性算法有SSD算法^[6]、RetinaNet算法^[7]和YOLO系列算法^[8],它们直接获取待检测目标的锚框和类别信息,因而速度较快且精度较高。

国内外学者对安全帽和反光衣检测进行了研究,并取得系列成果。Han等^[11]在YOLOv8中引入SEResNeXt检测头以增强特征表示,并在C2f中加入CA注意力以提高模型捕获关键信息的能力,提高对密集小目标安全帽的检测精度。张学立等^[12]对YOLOv5进行轻量化改进以用于安全帽和反光衣检测,将C3模块和部分卷积替换为Ghost模块,在保障准确率的前提下降低了模型复杂度。Du等^[13]提出BLP-YOLOv10模型并将其用于地下矿山的安全帽检测,通过结合稀疏注意力机制、调整骨干通道参数和集成低频增强滤波器以优化安全帽特征提取。谢国波等^[14]对YOLOX进行改进以检测复杂作业场景下的反光衣和安全帽,在特征融合层嵌入带权重的注意力模块,通过权重系数关注特征图的空间维度,增强其表达能力,并引入自适应特征融合模块以促进多尺度特征融合。Zhang等^[15]基于YOLOv9目标检测算法,并结合StrongSORT跟踪算法和YoloPose人体姿态估计模型,提出了一种安全帽佩戴检测方法。许迪等^[16]提出YOLOv8-DSI以解决在安全帽和反光衣检测中遇到复杂天气时检测精度较低的问题,设计了基于并行空洞卷积和残差思想的DR-SPPF模块,以扩大感受野并提升复杂天气下的检测精度。

以上检测算法在工人防护装备检测上都取得了一定成果,但在遮挡小目标安全帽和反光衣检测上仍有不少提升空间。一些算法仅对单一的安全帽进行检测,不能同时检测安全帽与反光衣,忽略了反光衣对保障工人安全的重要性,因而应用范围较窄^[11,13,15]。在安全帽与反光衣检测任务中存在较多遮挡目标和小目标,这是由于工地人员密

集和有较多建筑材料造成遮挡,并且工人流动性较强,距离监控较远时则会形成小目标进而给检测带来困难。现有YOLO11算法的检测精确率和召回率都普遍较低,这是因为现有算法对遮挡小目标的识别效果不理想,故有必要对其进行改进以提升此类目标的检测效果。

针对上述在安全帽和反光衣检测中存在的问题,本文提出了一种改进YOLO11的工人防护装备检测算法。在算法改进和数据集方面作出如下贡献:①提出ASDCN模块,将其替换骨干网络中的Conv模块,在实现下采样的同时捕捉到更复杂的多尺度特征;②提出DBPA-C3k2模块以更新模型中的C3k2模块,加强网络的特征提取和特征融合能力,并在模型入口处引入注意力机制,加强小目标细节特征提取;③添加P2层检测头,有效增强小目标防护装备检测;④提出FI-WIoUv2损失函数以训练模型,提高边界框回归精度;⑤通过自主收集图片和从公开数据集里抽取合适的图片,制作更完善的工人防护装备检测数据集。最后,通过对比实验和消融实验验证了本文算法的优越性,其在检测精度与速度上取得了较好平衡,为工人防护装备检测提供了一种有效的解决方案。

1 YOLO11算法

YOLO11是Ultralytics公司发布的一种目标检测算法,模型结构由骨干网络(Backbone)、颈部网络(Neck)和头部网络(Head)这3部分构成,如图1所示。

YOLO11的主要创新之处在于对骨干网络进行了优化,改进了颈部网络的特征融合机制,并且设计了高效的检测头,实现了精度与速度的双重提升。具体过程如下:首先,使用C3k2替换YOLOv8中的C2f,C3k2是由经典的跨阶段部分瓶颈CSP演变而来,该模块能够进一步优化网络中的信息流,将其应用于特征提取和特征融合网络,可使模型更关注目标的细节特征,在保持性能的同时又减少了计算成本;其次,在SPPF后加入C2PSA以进一步加强骨干网络的特征提取能力,C2PSA是在C2基础上引入多头注意力机制,并运用多层次设计理念,以加强对多层次特征的处理,使模型能够学习更丰富的特征表示;最后,在分类检测头中使用了深度可分离卷积DWConv以降低冗余计算,减少卷积运算的参数量和浮点运算数,以加快检测速度。YOLO11检测头依旧是解耦结构,即2条独立的分支分别用于检测目标的位置和类别。在边界框回归任务中采用的损失函数是CIoU和分布焦点损失DFL,分类时则使用二元交叉熵损失BCE。

2 YOLO11算法改进

由于建筑工地环境复杂,人员离监控的距离有近有远,使得待检目标尺度变化大,且工人作业时易聚集或受

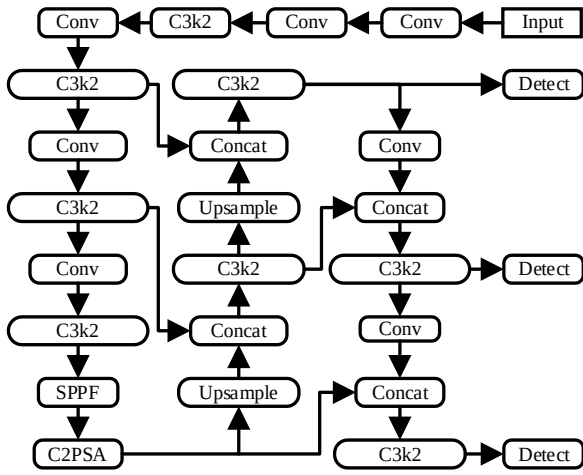


Fig. 1 YOLO11 network structure

图1 YOLO11网络结构

到遮挡,造成特征不明显,使用YOLO11n模型进行防护装备检测时效果不理想,存在一定的漏检和误检问题,故对YOLO11n模型作出改进以提升其检测性能。

2.1 ASDCN

非对称卷积网络(Asymmetric Convolution Network, ACN)^[17]是一种新颖的卷积神经网络,不同于传统的单一方形卷积,它引入两个非对称卷积分支和一个方形卷积分支以增强网络特征提取能力,这种设计能够加强模型对旋转畸变的鲁棒性,提升泛化能力。ACN在训练时采用多分支而推理时将其融合为单分支,因而可以在不增加推理时间的前提下提升防护装备的检测精度。

以算法中使用的3×3卷积核为例说明ACN的构成,首先将单一的方形卷积核拆分为1×3、3×1和3×3的并行结构,每个分支后都加入BN层,然后通过特征图相加操作对特征进行融合。在推理时则先融合各分支的Conv层和BN层,再将各分支卷积核中对应位置的参数相加即可,示意图如图2所示。

具体融合过程推导如下:设 $\mathbf{M} \in \mathbb{R}^{U \times V \times C}$ 是通道数为 C 的输入特征图, $\mathbf{O} \in \mathbb{R}^{R \times T \times D}$ 表示输出特征图,则有 D 个大小为 $H \times W$ 的卷积核,卷积核中的偏置为 $b \in \mathbb{R}^{D \times 1}$,可用 $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$ 表示一个三维卷积核的权重,则对于第 j 个带偏置的卷积核,输出特征图中对应的通道如式(1)所示。

$$\mathbf{O}_{:,j} = \sum_{k=1}^C \mathbf{M}_{:,k} * \mathbf{F}_{:,k}^{(j)} + b^{(j)} \quad (1)$$

其中,*为二维卷积运算。在设计卷积神经网络结构时,常在卷积层后加入BN层用于加速训练过程并减少过拟合的情况,则经过BN层后输出变为如式(2)所示。

$$\mathbf{O}_{:,j} = \left(\sum_{k=1}^C \mathbf{M}_{:,k} * \mathbf{F}_{:,k}^{(j)} + b^{(j)} - \mu_j \right) \frac{\gamma_j}{\sigma_j} + \beta_j \quad (2)$$

其中, μ_j 和 σ_j 分别为此BN层的均值和标准差, γ_j 和 β_j 则为学习到的缩放参数和平移参数,忽略了极小常数 ε 。将式(2)简化后即可得到Conv层与BN层融合后的公式,如式(3)所示。

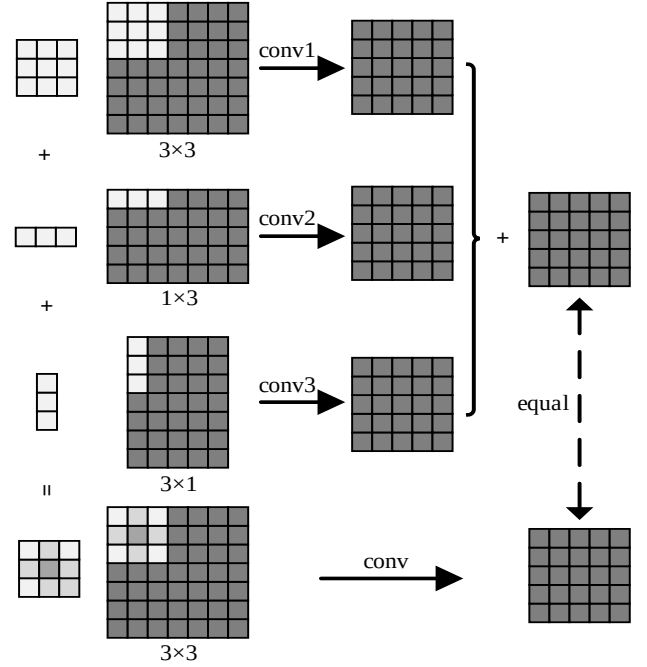


Fig. 2 ACN schematic diagram

图2 ACN示意图

$$\mathbf{O}_{:,j} = \sum_{k=1}^C \mathbf{M}_{:,k} * \mathbf{F}_{:,k}^{(j)} + b_{fused}^{(j)} \quad (3)$$

则融合后卷积层的权重和偏置更新为如式(4)、式(5)所示。

$$\mathbf{F}_{fused:,k}^{(j)} = \frac{\gamma_j}{\sigma_j} \mathbf{F}_{:,k}^{(j)} \quad (4)$$

$$b_{fused}^{(j)} = \frac{\gamma_j (b^{(j)} - \mu_j)}{\sigma_j} + \beta_j \quad (5)$$

若多个大小兼容的卷积核以相同步长对同一输入进行卷积,然后将得到的输出相加,便可直接将多个卷积核对应位置的参数相加融合为一个卷积核,再将其与输入特征图卷积后即可获得相同结果,如式(6)所示。

$$\begin{aligned} & \mathbf{I} * \mathbf{K}^{(1)} + b^{(1)} + \mathbf{I} * \mathbf{K}^{(2)} + b^{(2)} \\ &= \mathbf{I} * (\mathbf{K}^{(1)} \oplus \mathbf{K}^{(2)}) + (b^{(1)} + b^{(2)}) \end{aligned} \quad (6)$$

其中, $\mathbf{I}$ 为输入特征图, $\mathbf{K}$ 和 b 分别为卷积核的权重和偏置, $\oplus$ 表示对应位置参数逐元素相加。ACN中1×3、3×1和3×3的卷积核便属于大小兼容的卷积核,故可按照此公式对其加以融合。

空间到深度层(Space to Depth, SPD)^[18]能够将特征图从空间块重排列至深度维度,即减少特征图的高宽以增加通道数,这样便可像跨步卷积那样实现下采样又能极大地保留通道中丰富的特征信息,适用于含有较多小目标和图像分辨率较低的防护装备检测场景。

设输入SPD层的中间特征图为 $\mathbf{X} \in \mathbb{R}^{S \times S \times C}$,则通过切片操作可将 $\mathbf{X}$ 分割为一系列子特征图 $\mathbf{F}$,如式(7)~式(10)所示。

$$\mathbf{F}_{0,0} = \mathbf{X} [0:S; scale, 0:S; scale] \quad (7)$$

$$\mathbf{F}_{scale-1,0} = \mathbf{X} [scale-1:S; scale, 0:S; scale] \quad (8)$$

$$\mathbf{F}_{0, scale-1} = \mathbf{X}[0:S:scale, scale-1:S:scale] \quad (9)$$

$$\mathbf{F}_{scale-1, scale-1} = \mathbf{X}[scale-1:S:scale, scale-1:S:scale] \quad (10)$$

其中, 比例因子 $scale=2$ 表示对输入特征图进行 2 倍下采样, 故可得到 4 张子特征图 $\mathbf{F} \in \mathbb{R}^{S/2 \times S/2 \times C}$, 将它们按如上排列顺序在通道维度连接起来, 最终 SPD 层的输出特征图为 $\mathbf{X}' \in \mathbb{R}^{S/2 \times S/2 \times 4C}$ 。

综上所述, 将 ACN 与 SPD 相结合, 提出非对称空间到深度卷积网络 (Asymmetric Space to Depth Convolution Network, ASDCN), 其网络结构如图 3 所示。将其用于替换 YOLO11 骨干网络中的后 4 个 Conv 模块, 利用 ACN 的非对称卷积加强水平和垂直方向的特征响应捕获, 并优化 ACN 的网络结构; 在卷积核中引入偏置, 增强模型对数据分布偏移的适应性, 再结合 SPD 层实现无信息损失的下采样操作, 有效地增强了骨干网络的特征提取能力, 并提升了密集小目标防护装备检测精度。

输入特征图先分别经过 1×3 、 3×1 和 3×3 的卷积核进行特征提取, 与 YOLO11 中骨干网络中的 Conv 模块不同, 其步幅全置为 1 且启用了偏置, 输出特征图经 BN 层后再按元素相加进行特征融合, 通过 SiLU 非线性激活函数后执行空间维度上的切片操作, 将特征图一分为四且高宽减半实现下采样的同时又能极大地减少信息损失, 最后在通道维度上拼接即可。此为 ASDCN 训练时的网络结构, 而模型训练好后进行推理时使用重参数化技术, 先将各分支的 Conv 层与 BN 层融合为一个 Conv 层, 再融合各分支新生成的 Conv 层即可在减少参数量与计算量的同时保持精度不变, 同时提高推理速度, 其中 SPD 代表训练时的切片与拼接操作。

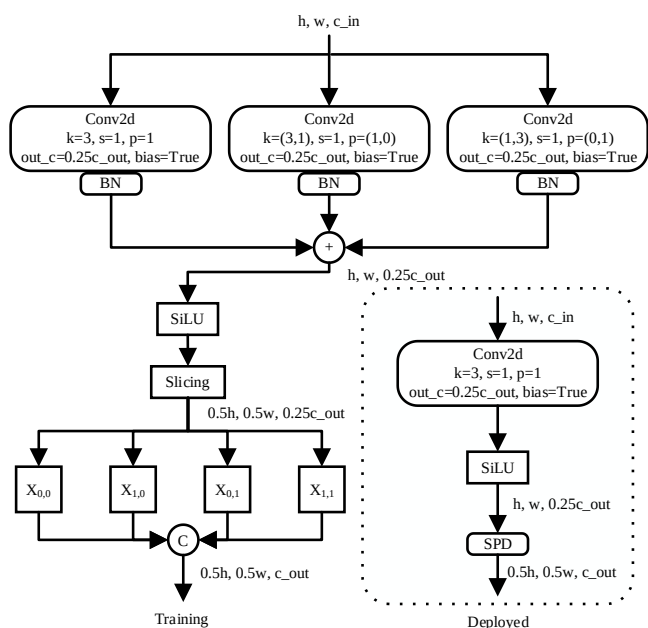


Fig. 3 ASDCN network structure

图3 ASDCN网络结构

2.2 DBPA-C3k2

将 C3k2 用于 YOLO11 的特征提取和特征融合网络中, 其具体结构如图 4 所示, 但其用于检测密集遮挡小目标时面临一定挑战。

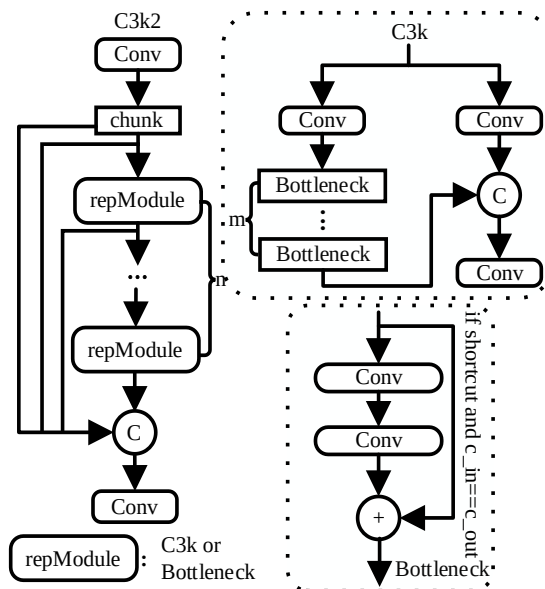


Fig. 4 C3k2 network structure

图4 C3k2网络结构

为提升模型对此类防护装备的检测精度, 提出多样化分支分块注意力 C3k2 (Diverse Branch Patch-aware Attention C3k2, DBPA-C3k2), 其网络结构如图 5 所示。

在 C3k2 中引入多样化分支块 (Diverse Branch Block, DBB)^[19], 通过多分支具有的不同复杂性和尺度, 帮助模型捕获更丰富的特征空间信息。同样使用重参数化技术对各分支进行融合, 使其不影响推理时间。为进一步提升微小目标检测, 在最初的特征提取网络阶段引入并行分块注意力 (Parallelized Patch-aware Attention, PPA)^[20]。PPA 的作用主要体现在如下方面: 首先, 其能够实现多尺度特征融合, PPA 通过并行多分支结构 (局部分支、全局分支、连续卷积分支) 提取不同尺度和层次的特征, 使用较小补丁参数的局部分支可捕获小目标的细节特征, 而较大补丁参数的全局分支可建模目标的上下文信息, 连续卷积分支则可增强局部特征的表达能力, 这种多分支设计使模型能够同时关注目标的局部细节和全局语义, 减少安全防护装备小目标信息丢失; 其次, PPA 通过结合通道注意力和空间注意力实现自适应注意力增强, 通道注意力使用自适应权重筛选重要通道, 抑制无关背景噪声, 而空间注意力则聚焦目标区域的像素位置, 提升小目标的显著性, 两者结合便可动态强化目标相关特征, 弱化复杂背景干扰; 最后, 在 DBPA-C3k2 模块的后端引入 PPA 能够对输出特征进行再校准, 融合多层次信息以优化检测结果的定位精度。

DBPA-C3k2 中的 PPA 注意力模块被设计为可自由选择是否引入, 以根据实际检测场景和检测运行硬件设备的

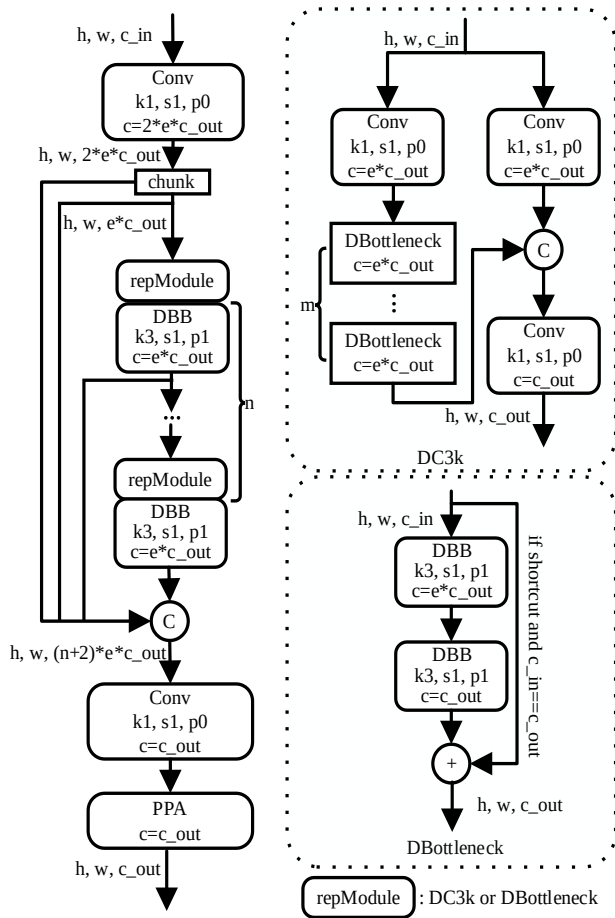


Fig. 5 DBPA-C3k2 network structure

图5 DBPA-C3k2网络结构

计算性能选择合适数量的注意力机制。本文算法仅在改进YOLO11骨干网络的第一个DBPA-C3k2中启用了PPA,如图10所示,以实现精度和速度之间的良好平衡。

相较于C3k2,DBPA-C3k2将Bottleneck中的普通Conv模块用DBB加以替换,并引入GELAN的思想^[8],在每个重复块后添加DBB优化网络结构,使模型在训练时识别到更复杂的特征信息,而推理时则使用重参数化技术将多样化分支融合为一个Conv层。DBB网络结构如图6所示。

其中,第一分支和第四分支为常规的Conv层和BN层;在第二分支的 1×1 卷积层中隐式地加入了恒等映射(Identity Mapping)即残差结构。卷积核由可训练权重和固定的单位矩阵权重相加构成,可使网络在初始阶段保留输入特征不作改变,避免在初始时破坏特征流动,而后续又可训练调整参数,通过直接修改卷积核而非显式跳跃连接,既减少了计算量又便于后续融合。在常规的 3×3 卷积层和平均池化层中,为保持高宽不变,会对输入特征图用零填充一圈,而在DBB中则使用基于BN层参数的填充值以提高网络处理边缘像素的能力,保留边界信息。其值如式(11)所示。

$$p = \beta - \frac{\mu\gamma}{\sqrt{\sigma^2 + \varepsilon}} \quad (11)$$

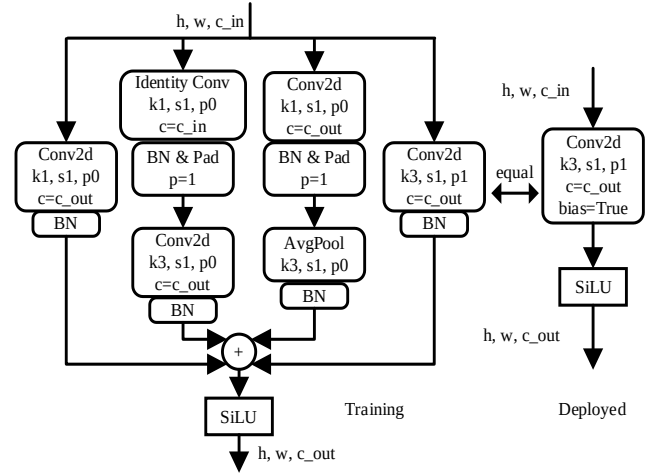


Fig. 6 DBB network structure

图6 DBB网络结构

其中,参数意义同式(2)。

在推理时,DBB会将各分支融合为一个带偏置的Conv层。 1×1 和 3×3 的Conv层与BN层可按式(3)进行融合,而融合后的 1×1 卷积核在权重四周用零填充一圈即可获得等价的 3×3 卷积核。平均池化层等价于一个无偏置而权重值都为 $1/9$ 的 3×3 卷积核,可将池化层转换为卷积层。设第二分支中 1×1 和 3×3 的Conv层与BN层融合后权重分别为 $K_1 \in \mathbb{R}^{c_{in} \times c_{in} \times 1 \times 1}$ 、 $K_2 \in \mathbb{R}^{c_{out} \times c_{in} \times 3 \times 3}$,偏置为 $b_1 \in \mathbb{R}^{c_{in} \times 1}$ 、 $b_2 \in \mathbb{R}^{c_{out} \times 1}$,则对 K_1 进行重排,其维度为 $K_1 \in \mathbb{R}^{c_{in} \times c_{in} \times 1 \times 1}$,再将其与 K_2 进行卷积即可获得 1×1 卷积核与 3×3 卷积核融合后的权重 $K_3 \in \mathbb{R}^{c_{out} \times c_{in} \times 3 \times 3}$,接着将 b_1 进行广播后与 K_2 按元素相乘,最后将其在第1至3维上将所得偏置与 b_2 相加求和即可得到融合后的偏置为 $b_3 \in \mathbb{R}^{c_{out} \times 1}$,第三分支同理。这样便可根据式(6)融合各分支为单一带偏置的卷积层,在保持精度不变的前提下极大地减少了模型推理时的参数量和计算量。

在微小防护装备检测这种需要细节保留的任务中,通过在骨干网络的第1个C3k2中嵌入轻量化的分块注意力PPA,强化对细粒度细节如小目标边缘的捕捉能力,优化网络入口处特征质量,为后续特征提取和融合保留更多关键细节。PPA网络结构如图7所示,其由多分支特征提取和注意力机制两大部分构成。分支结构又可分为序列卷积分支、局部分支和全局分支,序列卷积分支由3个连续

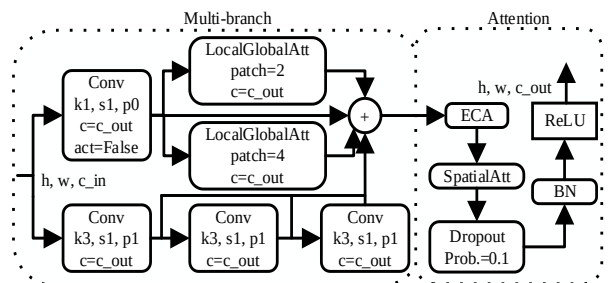


Fig. 7 PPA network structure

图7 PPA网络结构

的 $3 \times 3 \text{Conv}$ 模块构成, 通过堆叠小卷积核扩大感受野, 提升多尺度特征提取能力; 局部分支使用较小的补丁参数, 将特征图分割为非重叠的局部补丁, 加强对细粒度局部特征的提取; 而全局分支则用较大的补丁参数, 以获取更大范围的上下文信息, 减少小目标信息丢失。将各分支逐元素相加得到初步融合特征, 据此生成一维通道注意力图, 以对通道维度进行自适应加权, 强调重要的通道特征, 再生成二维空间注意力图, 在空间维度上突出响应目标区域, 最后通过批归一化和激活函数进一步优化输出最终特征。各注意力结构如图 8 所示。

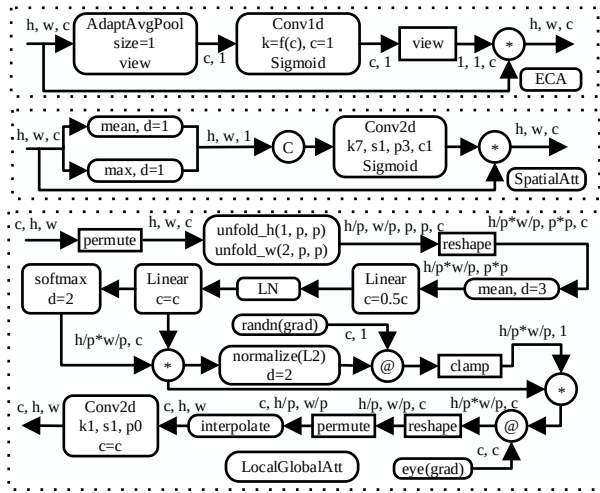


Fig. 8 Attention structure

图 8 注意力结构

通过在骨干网络前端引入 PPA, 以较低计算代价加强对低级特征的判别力, 使用并行多分支结构捕获不同尺度的上下文信息, 并结合通道-空间注意力机制增强小目标防护装备的显著性。相比其他经典注意力机制, 其有效地提高了检测精度。

2.3 P2 检测层

Chen 等^[21]对小目标作如下定义: 目标框的宽高与图像的宽高比值小于某一定值, 通常取 0.1。如图 11 所示, 防护装备检测任务中有较多小目标, 引入 P2 检测层以精准捕获小目标, 提高对小目标的检测能力。

输入模型的图片分辨率是 640×640 , P3 检测层特征图的最大尺寸仅为 80×80 。对于有较多小目标的防护装备检测而言, 这种尺寸的特征图其语义特征依旧不明显, 经多次卷积后对小目标的特征信息保留也较少, 从而给小目标检测带来较大困难。故引入 P2 检测层, 即在原模型上新增一个尺寸为 160×160 的检测特征图, 以充分利用浅层特征图易捕捉小目标的细节特征和其高分辨特性, 有效提高小目标检测效果。

2.4 FI-WIoUv2 边界框损失函数

在训练数据中包含低质量的示例不可避免, 如标注质量较差或因遮挡等难以被模型有效学习。YOLO11 模型采用的 CIoU 损失函数包含距离和纵横比这样的几何度量,

会加剧模型对低质量示例的惩罚, 导致泛化能力下降。因此, 引入 WIoUv2 损失函数, 使模型在目标框与锚框有较好重合的情况下减弱几何度量的影响, 不过多地干预训练以提高模型泛化能力^[22]。同样, 样本不均衡问题也广泛存在于目标检测任务中, 在防护装备检测中便有简单样本即一般的检测目标, 还有另一种困难样本, 如难以定位的小目标。因此, 引入 Focaler-IoU 通过聚焦困难样本的边框回归以提升小目标检测效果^[23]。由于防护装备检测中有较多遮挡小目标, 故模型检测遮挡目标时可能仅检测到可见部分, 而目标框却包含完整目标, 且小目标尺寸较小, 容易造成锚框稍有偏差便导致 IoU 大幅下降。这些因素都导致检测遮挡小目标时含有较多低 IoU 样本, 故引入 Inner-IoU, 通过大尺度的辅助框进一步优化模型对此类防护装备的检测^[24]。鉴于此, 提出聚焦内嵌式 WIoUv2 (Focaler Inner WIoUv2, FI-WIoUv2) 边界框损失函数以提高模型检测精度, 而不增加参数量和计算量, 也不影响推理速度, 如式 (12) 所示。

$$L_{FI-IoUv2} = \left(\frac{L_{FI-IoU}}{L_{FI-IoU}} \right)^{\gamma} R_{WIoU} L_{FI-IoU} \quad (12)$$

其中, L_{FI-IoU} 为聚焦困难样本和使用较大辅助边框计算而来的交并比损失值, γ 为 0.5, * 表示将此张量与计算图分离, 防止产生阻碍收敛的梯度。引入 L_{FI-IoU} 的均值 $\overline{L_{FI-IoU}}$ 作为归一化因子以加快收敛速度, 其初始值为 1, 按式 (13) 进行更新, 表示动量为 $m = 1 - 0.5^{(1/7000)}$ 的滑动平均值。

$$\overline{L_{FI-IoU}}_{t+1} = (1 - m) \overline{L_{FI-IoU}}_t + m L_{FI-IoU}^{batch_mean} \quad (13)$$

式 (12) 中的 R_{WIoU} 计算公式如式 (14) 所示。

$$R_{WIoU} = \exp \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2} \right) \quad (14)$$

其中, (x, y) 和 (x_{gt}, y_{gt}) 分别表示锚框和目标框的中心点, W_g 和 H_g 为最小外接矩形的宽和高, 如图 9 所示。

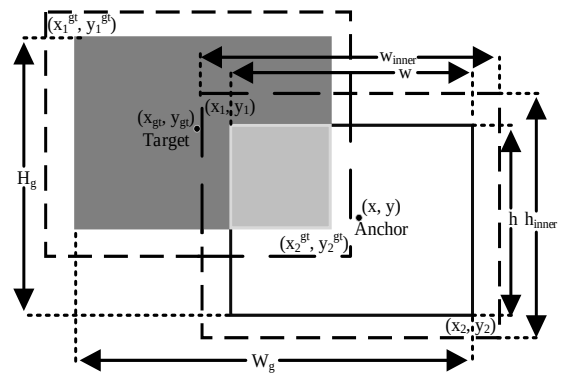


Fig. 9 IoU schematic diagram

图 9 IoU 示意图

由式 (14) 可知, $R_{WIoU} \in [1, e]$, 故将放大普通质量锚框的损失值 L_{FI-IoU} , 而 $L_{FI-IoU} \in [0, 1]$, 则将显著降低高质量锚框的 R_{WIoU} , 即降低对中心点距离的关注, 减弱几何度量

的影响,使模型在降低高质量锚框竞争力的同时,更有效地减少了低质量示例产生的有害梯度对模型训练产生的破坏,更关注于普通质量的锚框,从整体上提高检测性能。式(12)中的 L_{FI-IoU} 计算公式如式(15)所示。

$$L_{FI-IoU} = 1 - IoU^{FI} \quad (15)$$

IoU^{FI} 计算公式如式(16)所示。

$$IoU^{FI} = \frac{IoU^I - d}{u - d} \quad (16)$$

其中, $d=0$ 、 $u=0.95$,且 IoU^{FI} 的值被限制为0~1,通过线性区间映射调整交并比,聚焦困难样本的边框回归, IoU^I 表示使用辅助框计算而来的交并比。如图9所示, w^{gt} 和 h^{gt} 分别为目标框的宽和高, w 和 h 表示锚框的宽和高,则目标框以中心点不变用尺度因子 $r \in [0.5, 1.5]$ 进行缩放后所得框的左上角坐标为 (x_1^{gt}, y_1^{gt}) 即 $(x^{gt} - 0.5rw^{gt}, y^{gt} - 0.5rh^{gt})$,右下角坐标为 (x_2^{gt}, y_2^{gt}) 即 $(x^{gt} + 0.5rw^{gt}, y^{gt} + 0.5rh^{gt})$ 。同理,锚框缩放后左上角坐标为 (x_1, y_1) 即 $(x - 0.5rw, y - 0.5rh)$,右下角坐标为 (x_2, y_2) ,即 $(x + 0.5rw, y + 0.5rh)$,则这两个框的交集和并集分别如式(17)、式(18)所示。

$$Inter^I = [\min(x_2, x_2^{gt}) - \max(x_1, x_1^{gt})] * [\min(y_2, y_2^{gt}) - \max(y_1, y_1^{gt})] \quad (17)$$

$$Union^I = w^{gt}h^{gt}r^2 + whr^2 - Inter^I \quad (18)$$

故得辅助边框的交并比如式(19)所示。

$$IoU^I = \frac{Inter^I}{Union^I} \quad (19)$$

当 $r = 1$ 时,辅助边框的交并比等同于常规的交并比; $r < 1$ 时,将生成小于实际框的辅助框,对于高IoU样本可加速边框回归过程; $r > 1$ 时,生成的较大辅助边框则适用于低IoU样本,因为其扩大了回归的有效范围,对于低IoU的回归有增益作用。本文取 $r=1.2$,使模型对遮挡小目标的回归有增益作用。本文取 $r=1.2$,使模型对遮挡小目标的定位更加精准。

综上所述,改进后的YOLO11网络结构如图10所示,并在训练时采用FI-WIoUv2作为边界框回归损失函数以优化参数。

3 实验与结果分析

3.1 实验平台与参数

本文所有消融实验和对比实验环境如表1所示。不同的参数设置会对实验结果产生一定影响,实验相关参数设置如表2所示。

3.2 数据集

由于在百度飞桨平台公开的安全帽反光衣数据集仅有1 083张原始图片,故在此基础上从公开的SHWD安全帽数据集中抽取合适的图像,再从网络上收集相应图片扩充数据集,最终获得包含4 000张图片的安全帽与反光衣数据集。使用Labelimg进行标注,共包含佩戴安全帽、未佩戴安全帽、穿着反光衣和未穿反光衣这4个类别。按8:

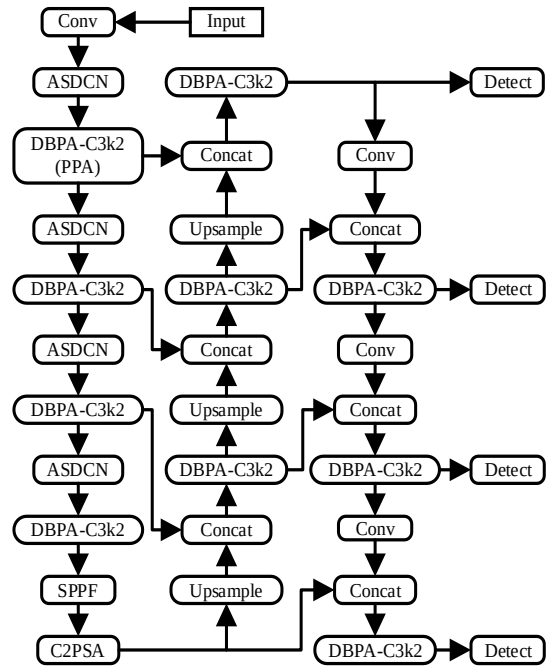


Fig. 10 Improved YOLO11 network structure

图10 改进后的YOLO11网络结构

Table 1 Experimental environment

表1 实验环境

项目	内容
操作系统	Windows 11
CPU	Intel(R) Core(TM) i7-13700H 2.40GHz
GPU	NVIDIA GeForce RTX 4060 Laptop
显存	8 GB
内存	16 GB
CUDA	12.2
开发环境	Python 3.8
深度学习框架	Pytorch 2.0.0+cu118

Table 2 Experimental hyperparameters

表2 实验超参数

参数名	参数值
迭代次数	200
批量大小	8
图像大小	640×640
优化器	SGD
初始学习率	0.01
动量	0.937
权重衰减	0.000 5

2的比例将数据集随机划分为训练集与验证集,如图11所示,框与图片的宽高比值在0.1处的颜色最深,可见在数据集中含有较多小目标。本文算法和数据集均已开源。

3.3 评价指标

使用以下6个指标对算法进行评价:①精确率(Precision, P)为在所有被预测为正例的样本中真正例所占比例;②召回率(Recall, R)为在所有真实为正例的样本中被预测为正例的样本所占比例;③平均精度均值(mean Average Precision, mAP)衡量算法在所有类别上的检测性能,对

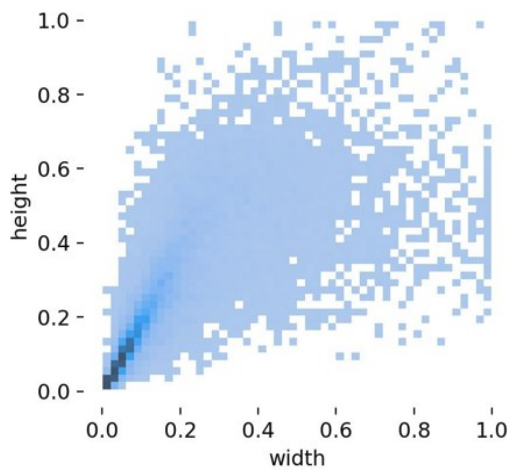


Fig. 11 Scale value of marker box
图 11 标记框比例值

P 和 R 进行了综合评价;④检测延迟 Latency 表示检测一张图片需要耗费的时间;⑤浮点运算量 (Floating-Point Operations Per Second, FLOPs) 体现计算复杂度的浮点运算次数;⑥参数量 (Parameters, Param.) 为可训练参数的个数。

3.4 对比实验

为验证本文改进算法的有效性, 将其与如下算法进行对比: 经典的 Faster R-CNN 和 SSD、基于 Transformer 的模型 RT-DETR^[25]、YOLO 系列 YOLOv5~YOLOv10 和 YOLO11s, 以及 YOLO 最新的算法 YOLOv12。所有实验均使用同一数据集, 各指标如表 3 所示, 迭代次数都为 200 轮, 均未加载预训练权重进行训练。其中, 部分模型的 P-R 曲线如图 12 所示, P-R 曲线越接近右上角则与坐标轴围成的面积越大, 即精准率和召回率都更高, 表明算法性能越好。

Table 3 The results of comparative experiments
表 3 对比实验结果

Model	P/%	R/%	mAP@0.5/%	Latency/ms	Param./M	FLOPs/G
Faster RCNN-resnet50	55.3	70.6	69.3	91.5	28.3	474.1
SSD-vgg	86.6	64.7	75.7	38.4	24	137.2
RT-DETR-l	90.6	83.9	87.8	35.0	15.5	37.5
YOLOv5-n	89.5	80.2	86.5	21.1	1.8	4.1
YOLOv6-n	92.2	76.9	85.6	29.7	4.6	11.3
YOLOv7-tiny	91.6	80.0	87.4	22.1	6.0	13.0
YOLOv8-n	89.3	80.7	87.1	18.3	3.0	8.1
YOLOv9-t-gelan	90.5	80.5	88.9	44.8	1.9	7.1
YOLOv10-n	90.0	77.4	86.3	18.8	2.7	8.2
YOLO11-n	88.3	80.0	86.9	18.7	2.6	6.3
YOLO11-s	90.6	82.5	89.1	19.2	9.4	21.3
YOLOv12-n	90.9	77.5	86.6	29.2	2.6	6.3
Ours	91.5	83.3	90.6	23.9	3.0	20.7

通过对比可知, 本文算法的检测精度 mAP@0.5 最高达到了 90.6%, 相比 YOLO11n 提高了 3.7%, 精确率和召回率也分别提高了 3.2% 和 3.3%。由此可见, 改进后的算法能够有效改善原算法的漏检和误检情况, 虽然浮点计算量有

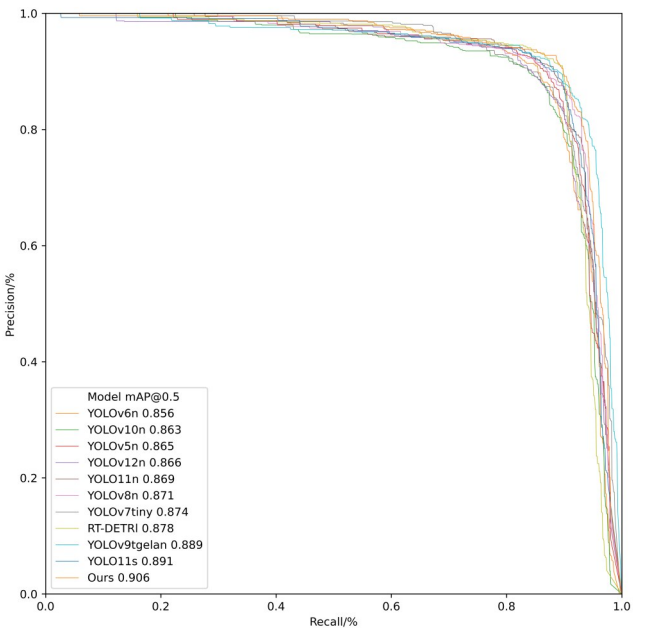


Fig. 12 P-R curve of comparative experiments
图 12 对比实验的 P-R 曲线

所增加, 但参数量仅增加了 0.4 M, 单张图片检测耗时为 23.9 ms 相较于 YOLO11n 也仅增加了 5.2 ms, 而其他算法检测精度 mAP@0.5 与所提算法都有一定差距, 如通过简单地增加卷积层通道数和 C3k2 中重复块个数的 YOLO11s 在增加 6.8 M 参数量的代价下依旧未能获得高于所提算法的检测精度, 最新的 YOLOv12n 在防护装备检测上效果也不理想。可以证明, 本文算法是一种高精度且具有较好实时性的检测算法, 更适用于工人防护装备检测任务。

3.5 消融实验

为验证本文各改进点的有效性, 开展消融实验以充分评测改进点 1~4 对算法检测性能的影响, 如表 4 所示。其中, 1 代表使用 ASDCN 替换 Conv, 2 代表使用 DBPA-C3k2, 3 表示添加 P2 检测层, 4 为使用 FI-WIoUv2 作为边界框回归损失函数。

Table 4 The results of ablation experiments
表 4 消融实验结果

1	2	3	4	mAP@0.5/%	Latency/ms	Param./M	FLOPs/G
×	×	×	×	86.9	18.7	2.6	6.3
√	×	×	×	88.2	17.4	2.3	9.3
×	√	×	×	88.5	24.2	3.1	13.7
×	×	√	×	87.9	26.7	2.7	10.2
×	×	×	√	87.5	19.1	2.6	6.3
√	√	×	×	89.2	23.7	2.9	16.6
√	√	√	×	90.2	25.0	3.0	20.7
√	√	√	√	90.6	23.9	3.0	20.7

使用 ASDCN 模块替换 Conv 模块后, mAP 提高了 1.3%, 参数量减少了 0.3 M, 而计算量增加了 3.0 G, 但检测耗时减少了 1.3 ms。由此可见, ASDCN 通过 3 路异构卷积并行分支结构和无信息损失的空间特征重组策略, 增强了

骨干网络的特征提取能力。在此基础上引入 DBPA-C3k2 模块以提取和融合特征后, mAP 提高了 2.3%, 相较于原算法单独使用此模块时参数量增加了 0.5 M, 但计算量增加较多, 达到了 7.4 G, 故检测延迟也增加了 5.5 ms, 可见多样化分支块和并行分块注意力可进一步提升网络性能。再新增一个 P2 层检测头, mAP 继续提高了 3.3%, 相较于原算法仅引入新的检测头时参数量略微增加了 0.1 M, 计算量增加了 3.9 G, 检测延迟则多增加了 8 ms, 表明 P2 层检测头有效地捕获了微小目标。最后使用 FI-WIoUv2 损失函数以优化参数, 提高了边界框回归精度。由于损失函数仅参与训练而不影响推理, 故参数量和计算量都不会改变, 检测延迟也与原算法相差无几。经过各模块组合, 使检测精度 mAP 最终增加了 3.7%, 相比于原算法在增加 0.4 M 参数的代价下检测精度得到大幅提升。而单张图片检测耗时也仅增加了 5.2 ms, 依然满足实时性要求。由此可见, 本文算法在精度与速度上实现了较好平衡, 是一种高效的防护装备检测算法。算法改进前后 mAP 随迭代次数变化如图 13 所示, 可知改进后算法检测效果得以改善。

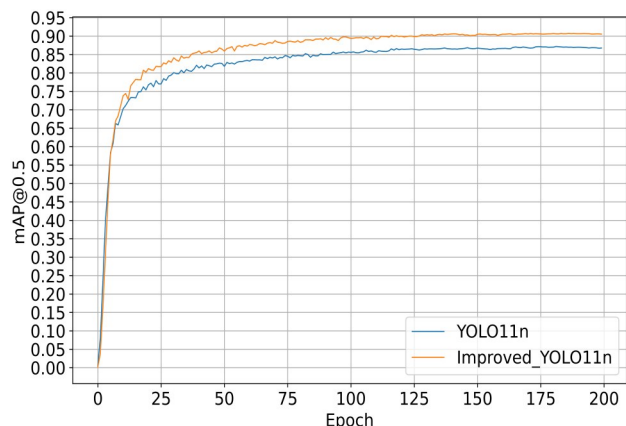


Fig. 13 mAP@0.5 curve

图 13 mAP@0.5 曲线

在第 1~12 轮内改进前后的 YOLO11 算法其 mAP 相差无几, 也都在快速提升, 但随后改进后的算法在精度 mAP 上则与原算法拉开了较大差距, 并在训练到第 100 轮左右时 mAP 便已达到 90% 左右趋于收敛, 而原算法依旧未收敛还在轻微上涨, 且在最后的第 200 轮时已收敛达到稳定但其 mAP 却低了 3.7%, 表明改进后的算法不仅收敛速度更快一些而且检测精度得以大幅提升。

3.5.1 重参数化前后参数对比

由上述分析可知, ASDCN 模块和 DBPA-C3k2 模块中的 DBB 模块在推理时都会将多个并行分支结构使用重参数化技术融合为单一的卷积层, 具体变化如表 5 所示。在 ASDCN、DBPA-C3k2 和改进后的算法中使用重参数化后, 参数分别减少了约 12 万个、115 万个和 131 万个, 可见重参数化能够有效降低参数量和计算量, 从而降低检测耗时且不影响精度。

Table 5 Reparameterization experiment

表 5 重参数化实验

Method	mAP@0.5/%	Latency/ms	Param.	FLOPs/G
ASDCN(Not)	88.2	20.3	2 447 252	12.1
ASDCN(Fused)	88.2	17.4	2 319 508	9.3
DBB(Not)	88.5	66.9	4 302 626	16.1
DBB(Fused)	88.5	24.2	3 147 082	13.7
Ours(Not)	90.6	84.3	4 287 206	26.7
Ours(Fused)	90.6	23.9	2 971 558	20.7

3.5.2 DBPA-C3k2 中使用不同注意力机制比较

如图 10 所示, 在骨干网络的第一个 DBPA-C3k2 模块中引入了 PPA 注意力以额外增强对微小目标的特征提取。与引入其他注意力的精度对比如表 6 所示, 在 C3k2 中仅使用 DBB 模块以加强特征提取和融合时, mAP 仅提升 0.7%, 而在骨干网络的入口处引入 PPA 注意力机制后精度提升了 1.6%, 可见 PPA 能够帮助模型有效地检测小目标防护装备。

Table 6 Comparison of attention mechanisms

表 6 注意力机制比较

Method	mAP@0.5/%	Method	mAP@0.5/%
DBPA-C3k2(No Att)	87.6(+0.7)	+CA	87.8(+0.9)
+ECA	87.0(+0.1)	+LSK	87.7(+0.8)
+CBAM	87.0(+0.1)	+PPA	88.5(+1.6)

3.5.3 不同损失函数比较

图 14 为改进前后算法在验证集上的 box_loss 曲线对比情况。可以看出, 使用 FI-WIoUv2 损失函数训练改进后的 YOLO11n 模型其损失值更低, 收敛也更快。

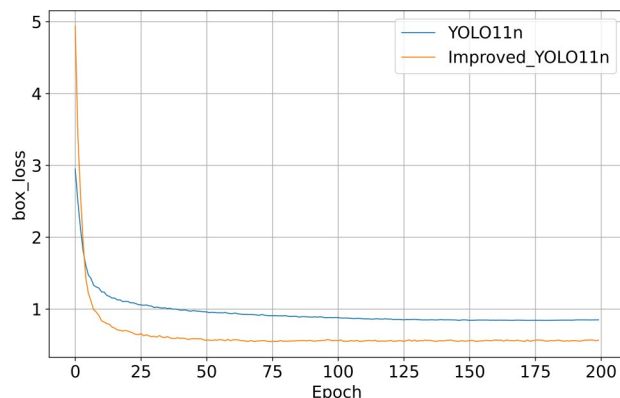


Fig. 14 box_loss descent curve

图 14 box_loss 下降曲线

为进一步验证所提损失函数在训练防护装备检测算法上的有效性, 将其与主流的损失函数进行对比, 如表 7 所示。结果表明, FI-WIoUv2 损失函数相比其他损失函数能够有效提高边界框回归精度, 并且通过实验发现, 缩放尺度因子 $r = 1.2$ 时其效果更好。

3.6 检测结果可视化分析

为了更直观地看到改进前后 YOLO11n 算法的检测效果, 对工地上的图片用改进前后算法进行检测, 其结果如

Table 7 Comparison of loss functions
表 7 损失函数比较

Method	mAP@0.5/%	Method	mAP@0.5/%
YOLO11n(CIoU)	86.9	DIoU	86.8(−0.1)
EIoU	86.7(−0.2)	GIoU	86.4(−0.5)
SIoU	87.0(+0.1)	MPDIoU	86.8(−0.1)
WIoUv1	87.1(+0.2)	WIoUv2	87.2(+0.3)
WIoUv3	87.1(+0.2)	FIWIoUv2(r=0.8)	86.9
FIWIoUv2(r=0.9)	87.0(+0.1)	FIWIoUv2(r=1.1)	87.1(+0.2)
FIWIoUv2(r=1.2)	87.5(+0.6)	FIWIoUv2(r=1.3)	86.8(−0.1)

图 15 所示。

其中,“h”代表佩戴安全帽,“nh”即未佩戴,“r”代表着反光衣,“nr”即未穿。图 15(a)左图是原算法检测结果,在白框位置发生了错检或漏检,如将佩戴安全帽识别成未

戴,未识别出被遮挡的反光衣,改进后的算法则减少了这些误识别情况,但依旧未识别出一些被遮挡的安全帽。如图 15(b)左图所示,原算法未检测到远处两个人,而改进后的算法则精准识别到了远处小目标的防护装备穿戴情况。图 15(c)中有多个尺寸大小的目标时,原算法能够检测出大尺寸的目标但未能发现小尺寸目标,而改进后的算法则改善了小尺寸反光衣的检测情况,但安全帽因过小依旧未检出。在图 15(d)的黑暗场景中,原算法未能检测出反光衣和安全帽,而改进后的算法检测出了一部分。由上述检测结果对比可知,本文算法对防护装备检测效果更佳。

4 结语

针对工地防护装备检测中大多数仅检测单一的安全帽和对遮挡小目标检测效果不理想的问题,本文提出了改进 YOLO11 的安全帽和反光衣防护装备检测算法。采用 ASDCN 模块替换传统的卷积模块,以加强网络特征提取能力;提出结合多分支块和注意力机制的 DBPA-C3k2,用于特征提取与融合;引入专门用于加强小目标检测的 P2 层检测头;使用 FI-WIoUv2 损失函数优化训练参数以提高边界框回归精度。实验结果表明,改进后的算法在检测精度上明显优于原算法,检测速度也依然满足实时性需求,对比其他检测算法具有显著优势。

下一步工作重点是在保持高精度的同时对模型实行轻量化,如采用神经网络剪枝、量化、知识蒸馏和轻量级架构设计,如深度可分离卷积等手段实现轻量化,以降低计算复杂度,实现更高更快的检测效果。同时,进一步扩充数据集,引入恶劣天气、黑夜等场景下的图片,增强模型对各种场景的适应性。

参考文献:

[1] MA Z. Statistical analysis of safety production accidents in the field of construction[J]. Construction and Architecture, 2022(3): 52-55.
马哲. 建筑施工领域安全生产事故情况统计分析[J]. 建筑, 2022(3): 52-55.

[2] JIAO L C, ZHANG F, LIU F, et al. A survey of deep learning-based object detection[J]. IEEE Access, 2019(7): 128837-128868.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[4] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017(39): 1137-1149.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the European Conference on Computer Vision,



Fig. 15 Comparison of detection results of YOLO11 algorithm before and after improvement

图 15 改进前后 YOLO11 算法检测结果对比

- 2016:21-37.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017:2980-2988.
- [8] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information [C]//Proceedings of the European Conference on Computer Vision, 2024:1-21.
- [9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection [J]. Advances in Neural Information Processing Systems, 2024, 37:107984-108011.
- [10] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors [DB/OL]. <https://arxiv.org/abs/2502.12524>.
- [11] HAN D G, YING C L, TIAN Z H, et al. YOLOv8s-SNC: an improved safety-helmet-wearing detection algorithm based on YOLOv8 [J]. Buildings, 2024, 14(12):3883-3883.
- [12] ZHANG X L, JIA X C, WANG M G, et al. Lightweight detection of helmets and reflective clothings: improved YOLOv5s algorithm [J]. Computer Engineering and Applications, 2024, 60(1):104-109.
张学立, 贾新春, 王美刚, 等. 安全帽与反光衣的轻量化检测:改进 YOLOv5s 的算法 [J]. 计算机工程与应用, 2024, 60(1):104-109.
- [13] DU Q, ZHANG S H, YANG S J. BLP-YOLOv10: efficient safety helmet detection for low-light mining [J]. Journal of Real-Time Image Processing, 2024, 22(1):10.
- [14] XIE G B, XIAO F, LIN Z Y, et al. Method for detecting reflective vests and safety helmets in complex operational environments [J]. Journal of Safety and Environment, 2024, 24(9):3513-3521.
谢国波, 肖峰, 林志毅, 等. 复杂作业场景下的反光衣和安全帽检测方法 [J]. 安全与环境学报, 2024, 24(9):3513-3521.
- [15] ZHANG Y, HUANG S Z, QIN J Z, et al. Detection of helmet use among construction workers via helmet-head region matching and state tracking [J]. Automation in Construction, 2025(171):105987.
- [16] XU D, ZHANG S Q, GE C. YOLOv8 security equipment inspection for complex environments [J]. Electronic Measurement Technology, 2024, 47(7):121-129.
许迪, 张淑卿, 葛超. 面向复杂环境的 YOLOv8 安全装备检测 [J]. 电子测量技术, 2024, 47(7):121-129.
- [17] DING X H, GUO Y C, DING G 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.
- [18] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022:443-459.
- [19] DING X H, ZHANG X Y, HAN J G, et al. Diverse branch block: building a convolution as an inception-like unit [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:10886-10895.
- [20] XU S B, ZHENG S C, XU W H, et al. HCF-Net: hierarchical context fusion network for infrared small object detection [DB/OL]. <https://arxiv.org/abs/2403.10778>.
- [21] CHEN C, LIU M Y, TUZEL O, et al. R-CNN for small object detection [C]//Asian Conference on Computer Vision, 2017:214-230.
- [22] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [DB/OL]. <https://arxiv.org/abs/2301.10051>.
- [23] ZHANG H, ZHANG S J. Focaler-IoU: more focused intersection over union loss [DB/OL]. <https://arxiv.org/abs/2401.10525>.
- [24] ZHANG H, XU C, ZHANG S J. Inner-IoU: more effective intersection over union loss with auxiliary bounding box [DB/OL]. <https://arxiv.org/abs/2311.02877>.
- [25] ZHAO Y A, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:16965-16974.

(责任编辑:孙娟)