

基于改进YOLOv8的安全帽佩戴检测算法

谢敏怡¹, 蒋作², 万学俊³, 赵周洲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: 针对工地安全帽检测中难以识别被遮挡目标和存在密集小目标的问题, 提出一种基于改进YOLOv8的工地安全帽实时检测算法。首先, 在特征提取网络中使用RVBS模块替换部分Conv模块, 通过多分支结构和无信息损失的下采样操作提升了网络的特征提取能力, 有助于密集小目标的检测; 其次, 采用RepNCSPPELAN4模块代替特征融合网络中所有的C2f模块, 使模型能更有效地融合浅层与深层特征, 从而改善被遮挡目标的检测效果; 再次, 新增小目标检测层以进一步融合多尺度特征, 提升对小目标的检测能力; 最后, 使用WIoU v1损失函数代替原始的CIoU损失函数, 加速模型收敛并提高回归精度。实验结果表明, 改进后的YOLOv8算法在SHWD数据集上的mAP@0.5和mAP@0.5:0.95分别比原算法提高了2.8%和3.2%, 每张图片检测用时为25.7 ms。与其他主流算法相比, 改进算法的mAP@0.5最高。所提算法能有效检测密集遮挡小目标, 具有较高的应用价值。

关键词: YOLOv8; 安全帽; 目标检测; RVBS; 小目标检测层; RepNCSPPELAN4; WIoU v1

DOI: 10.11907/rjtk.241907

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)001-0172-12

Safety Helmet Wearing Detection Algorithm Based on Improved YOLOv8

XIE Minyi¹, JIANG Zuo², WAN Xuejun³, ZHAO Zhouzhou³, 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: A real-time detection algorithm for construction site safety helmets based on improved YOLOv8 is proposed to address the problem of difficulty in identifying occluded targets and the presence of dense small targets in construction site safety helmet detection. Firstly, the RVBS module is used to replace some Conv modules in the feature extraction network, which enhances the network's feature extraction capability through a multi branch structure and information loss free downsampling operation, facilitating the detection of dense small targets; Secondly, the RepNCSPPELAN4 module is used to replace all C2f modules in the feature fusion network, enabling the model to more effectively fuse shallow and deep features, thereby improving the detection of occluded targets; Again, adding a small target detection layer to further integrate multi-scale features and enhance the detection capability for small targets; Finally, the WIoU v1 loss function is used instead of the original CIoU loss function to accelerate model convergence and improve regression accuracy. The experimental results show that the improved YOLOv8 algorithm performs well on the SHWD dataset mAP@0.5 and mAP@0.5:0.95 improved the original algorithm by 2.8% and 3.2% respectively, with a detection time of 25.7 ms per image. Compared with other mainstream algorithms, the improved algorithm mAP@0.5 highest. The proposed algorithm can effectively detect densely occluded small targets and has high application value.

Key Words: YOLOv8; helmet; object detection; RVBS; small object detection layer; RepNCSPPELAN4; WIoU v1

收稿日期: 2024-10-26

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; 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; 赵周洲(1987-), 男, 玉溪市公安局科技信息化支队科技管理科副科长、警务技术三级主管, 研究方向为视频监控系统建设、视频结构化和人像识别算法; 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。

0 引言

安全帽是保障工人头部安全的重要防护装备,能大大减少高空坠物导致的安全隐患。若施工人员不慎从高处坠落,安全帽也可以起到缓冲作用。虽然各个建筑公司近年来都加强了对工程项目的安全管理,但仍有不少事故发生,其中一个主要原因便是工人未按规定佩戴安全帽。当前我国许多施工场所仍采用人工监视的方式来检查工人是否佩戴安全帽,这种方式费时费力,且长时间工作后安全管理员也容易产生误判。因此,采用智能化技术来高效准确地检测工人是否佩戴安全帽十分必要。

安全帽检测算法可大致分为基于传统的检测算法和基于深度学习的检测算法两类。基于传统的检测算法首先通过滑动窗口对图像中的每个像素点进行扫描得到候选区域,其次提取待检测目标的特征,最后采用支持向量机等方法来判定检测目标是否佩戴安全帽。该类方法不仅推理速度慢,而且检测精度较低^[1]。基于深度学习的检测算法的效果明显优于传统检测算法,实时性和准确性均有较大提升,因此被广泛应用于实际检测中。该类算法又可分为两阶段和单阶段检测算法两种。常用的两阶段检测算法为Faster R-CNN(Faster Region-based Convolutional Neural Network)。例如,邓开发等^[2]将深度特征流算法与Faster R-CNN相结合,实现了监控视频中安全帽的检测,但对被遮挡目标的检测效果有待提升。单阶段检测算法包括SSD(Single Shot MultiBox Detector)和YOLO(You Only Look Once)系列等算法。例如,李明山等^[3]在SSD算法中添加分支网络以融合特征,并加入可变参数改变先验框大小,使模型具有较好的实时性和准确性。Li等^[4]提出一种以YOLOv5检测算法为基础,结合多目标跟踪技术的模型,用于监测和追踪实时视频中的工人是否佩戴安全帽,有效减少了被跟踪目标的丢失比例。吕宗喆等^[5]通过优化YOLOv5的损失函数并引入切片辅助推理和微调,提高了模型对小目标的检测精度,能满足安全帽佩戴实时检测的需求。Bao等^[6]将C2f模块和FE模块集成到YOLOv8的网络架构中,形成了一个新的注意力机制模块,该模块可融合来自不同层的特征信息,在减少计算开销的同时也提高了安全帽的识别效率。Lin^[7]在YOLOv8骨干网络中加入坐标注意力机制,以增强模型对安全帽区域的关注;并在颈部网络中采用slim-neck结构融合骨干网络提取的特征,以提高安全帽检测精度。以上检测算法均取得了较好效果,但对密集小目标的检测仍有提升空间。

针对在安全帽检测任务中主流算法难以检测被遮挡目标和密集小目标的问题,本文提出一种基于改进YOLOv8的算法以提升检测精度并满足实时检测的要求。首先,在Backbone网络中使用RVBS(Re-parameterization VGG Block with Space-to-Depth)模块替换Conv模块,以提

升网络的特征提取能力;其次,将Neck网络中的C2f模块替换为RepNCSPeLan4模块,以增强浅层与深层特征的融合;再次,加入小目标检测层,通过融入更丰富的安全帽信息来提高小目标检测效果;最后,在训练时使用WIoU v1替换CIoU以加速模型收敛并提高回归精度^[8-9]。改进后的YOLOv8算法在安全帽佩戴数据集(Safety Helmet Wearing Dataset, SHWD)上的检测性能明显提升,优于对照主流目标检测算法^[10]。

1 YOLOv8算法

YOLOv8是目前流行的一种单阶段目标检测算法,可用于普通目标检测、旋转框定向目标检测和关键点检测等,基于此实现了目标追踪、计数、测距和测速等功能。YOLOv8还提供了5种缩放系数来获得不同大小的模型,用于满足各个场景的需求,本文采用YOLOv8n作为基线模型进行改进。

YOLOv8主要分为Backbone网络、Neck网络和Head网络3个部分,具体结构如图1所示。其中,Backbone网络用于特征提取,与YOLOv5相比,其将第1个卷积模块中的卷积核由6×6替换成3×3,并将C3模块替换成梯度流更丰富的C2f模块;在Neck网络中移除了1×1卷积的降通道层,同样更换了C3模块;在Head网络中采用解耦头结构,将分类任务与回归任务解耦,同时也将Anchor-Based更换为Anchor-Free,可直接预测目标边界框的关键点。

YOLOv8的损失函数由边界框回归损失和分类损失构成,其中分类损失为二值交叉熵(Binary Cross Entropy, BCE)损失,回归损失又由DFL Loss和CIoU Loss构成^[11]。这3个损失采用一定比例加权后得到最终损失。CIoU的计算公式为:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha\nu \quad (1)$$

$$\alpha = \frac{\nu}{(1 - \text{IoU}) + \nu} \quad (2)$$

$$\nu = \frac{4}{\pi^2} (\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h})^2 \quad (3)$$

式中:IoU为交并比; $\rho^2(b, b^{\text{gt}})$ 为两个边界框中心点之间的欧氏距离; c 为最小外接矩形的对角线长度; ν 为衡量两个边界框宽高比差异的指标; α 为权重系数。

2 改进算法

本文在YOLOv8算法的基础上进行改进,主要包含特征提取模块、特征融合模块和边界框回归损失函数3个方面的改进。

2.1 特征提取模块

2.1.1 RepVGGBlock

RepVGG是一种具有多分支结构的卷积神经网络,由

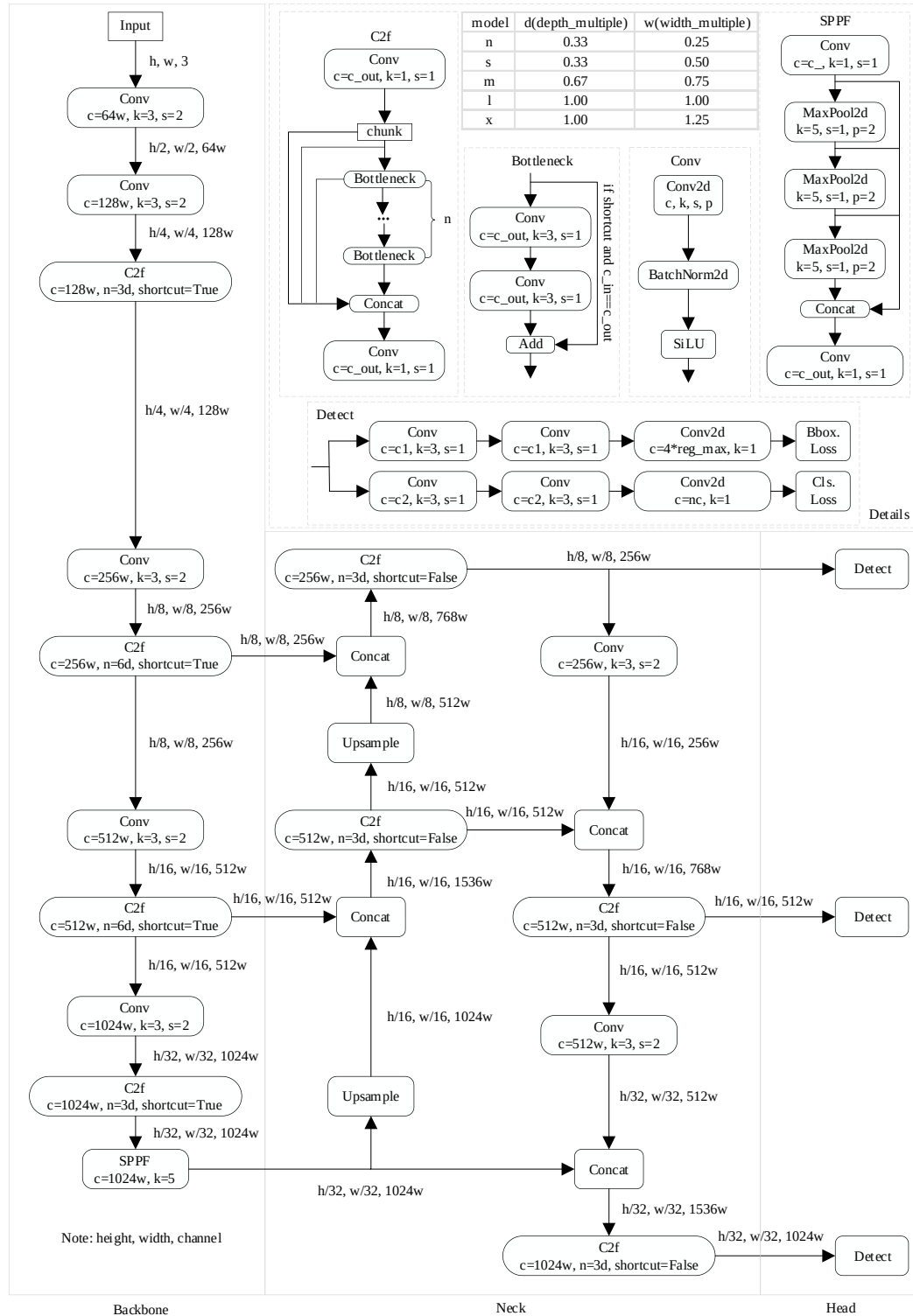


Fig. 1 YOLOv8 network structure

图1 YOLOv8网络结构

3×3的卷积层和激活层堆叠而成,可捕获丰富的特征表示^[12]。其运用结构重参数化(Structural Re-parameterization)技术,即在训练时使用一个类似ResNet-style的多分支模型,当推理时便转化为VGG-style的单路模型。如图2所示,RepVGGBlock由2个卷积模块分支和1个BN层分支并行而成,在训练时并未进行下采样,且BN层分支仅在

RepVGGBlock不改变输入特征图通道数的情况下才会参与训练。之所以在训练时采用多分支结构,是由于并行多个分支通常能够增强模型的特征提取能力,而推理时将多分支转换成单路则会使推理速度更快。

通过结构重参数化将训练完毕的模型转换为推理时的模型,主要分为两步:①将Conv2d与BN融合,由于存在

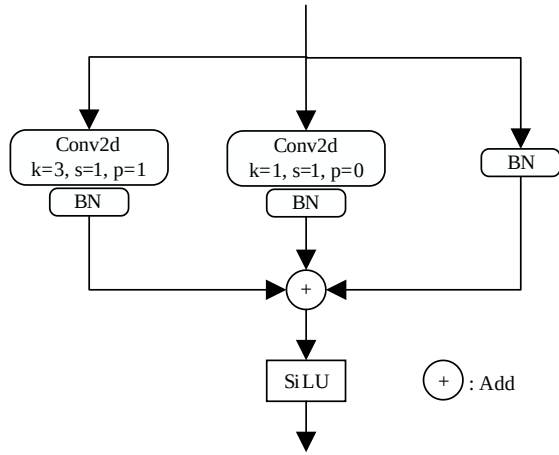


Fig. 2 RepVGGBlock network structure

图2 RepVGGBlock 网络结构

不同大小的卷积核,在融合前要将 1×1 的卷积核转换为 3×3 的卷积核,并在只有 BN 层的分支处加入一个卷积核为 3×3 的 Conv2d;②将每个分支上融合得来的 3×3 卷积层再融合为 1 个卷积层。假设输入特征图的通道数为 2 且 RepVGGBlock 不改变其通道数,则融合过程如图 3 所示。

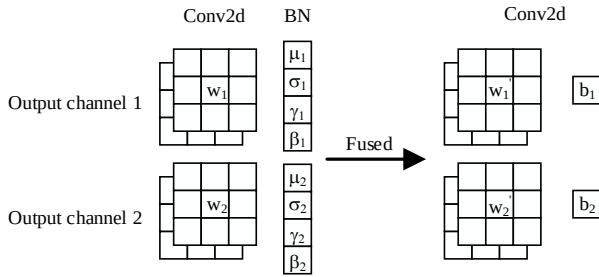


Fig. 3 The fusion process of Conv2d and BN

图3 Conv2d 与 BN 的融合过程

设 $X_{:,i,:}$ 为输入 BN 层的特征图的第 i 个通道,则对应的输出通道 $Y_{:,i,:}$ 为:

$$Y_{:,i,:} = \frac{X_{:,i,:} - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} \gamma_i + \beta_i \quad (4)$$

式中: μ 、 σ^2 、 γ 和 β 分别为 BN 层的均值、方差、权重和偏差; ε 为一个非常小的常量,用于防止分母为零。

卷积操作是一种线性运算,Conv2d 与 BN 融合即等价于 1 个带偏置的 Conv2d。表示为:

$$W'_{i,:} = W_{i,:} \frac{\gamma_i}{\sqrt{\sigma_i^2 + \varepsilon}} \quad (5)$$

$$b_i = \beta_i - \mu_i \frac{\gamma_i}{\sqrt{\sigma_i^2 + \varepsilon}} \quad (6)$$

式中: $W_{i,:}$ 为训练好的输出通道 i 对应的卷积核权重; $W'_{i,:}$ 为融合后的卷积核权重, b_i 为其偏置。

用 0 对 1×1 的卷积核四周进行填充,将其等价转化为 3×3 的卷积核。在仅有 BN 层的分支处加入 1 个卷积核为 3×3 的 Conv2d。对于第 i 个输出通道,只需令对应卷积核中的元素 $w_{i,i,1,1}$ 为 1 而其余为 0,即可实现恒等映射。将 3

个分支上由 3×3 的 Conv2d 与 BN 融合得来的 3×3 卷积层再融合为 1 个卷积层,即将各个分支对应的参数相加即可,包括权重和偏置。公式为:

$$O = (I \otimes W_1 + b_1) + (I \otimes W_2 + b_2) + (I \otimes W_3 + b_3) \quad (7)$$

$$= I \otimes (W_1 + W_2 + W_3) + (b_1 + b_2 + b_3)$$

式中: $\otimes$ 为卷积运算; I 、 O 分别为输入和输出特征图; W 、 b 分别为卷积核的权重与偏置。

2.1.2 SPD

在卷积神经网络中使用步幅大于 1 的卷积运算或池化操作会导致信息丢失和特征提取效率降低,且当输入分辨率较低或目标较小时上述缺陷会被放大。在安全帽检测任务中存在较多小目标,因此本文在特征提取中引入 SPD(Space-to-Depth)代替卷积模块中的下采样操作,用于提高模型处理小目标和低分辨率图像等困难任务时的效率^[13]。SPD 结构见图 4。

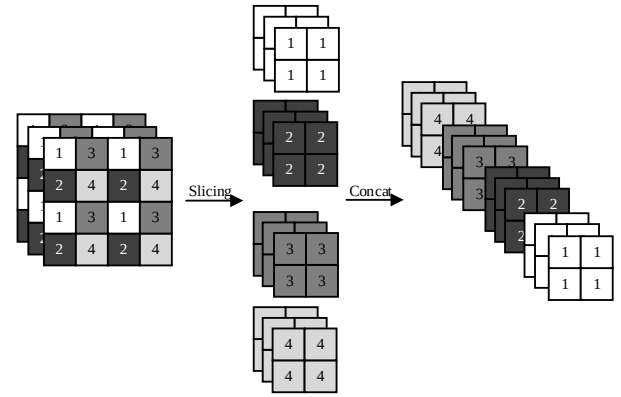


Fig. 4 Structure of SPD

图4 SPD 结构

设输入特征图 X 的形状为 (S, S, C) ,则切片出来的一系列子特征图为:

$$\begin{aligned} X_{0,0} &= X[\dots, 0: S/2, 0: S/2] \\ X_{1,0} &= X[\dots, 1: S/2, 0: S/2] \\ X_{0,1} &= X[\dots, 0: S/2, 1: S/2] \\ X_{1,1} &= X[\dots, 1: S/2, 1: S/2] \end{aligned} \quad (8)$$

式中: 2 表示在切片时每隔一个元素取一个,其形状皆为 $(S/2, S/2, C)$ 。将得到的所有子特征图在通道维上进行拼接即可获得 2 倍下采样后的特征图 $X'(S/2, S/2, 4C)$ 。

2.1.3 RVBS

RepVGGBlock 的多分支结构能增强模型的特征提取能力;SPD 可用于替代卷积运算中的下采样操作,有助于小目标的检测,将这两者结合起来构成 RVBS 模块,用于替换 Backbone 网络中的部分 Conv 模块可提升特征提取网络性能,增强模型对密集小目标的检测能力。RVBS 网络结构如图 5 所示,该网络在 Conv2d 步幅为 1 的情况下仍能对特征图进行 2 倍下采样。

2.2 特征融合模块

2.2.1 RepNCSPELAN4

GELAN (Generalized Efficient Layer Aggregation Net-

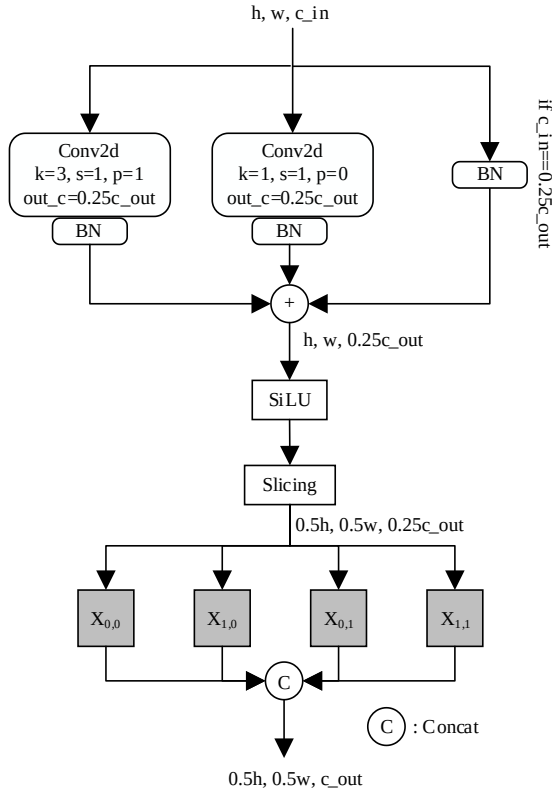


Fig. 5 RVBS network structure

图5 RVBS网络结构

work)是通过结合 CSPNet (Cross Stage Partial Network) 与 ELAN (Efficient Layer Aggregation Network) 这两种具有梯度路径规划的神经网络架构而提出的^[14-16]。ELAN网络中仅使用卷积层堆叠,而 GELAN 则对其进行拓展,继承了 CSPNet 的梯度路径优化思想和 ELAN 的高效拓扑结构,形成了一个允许使用任意类型计算块的通用框架,即 GELAN 可根据不同的应用场景及其需求灵活地选择计算块。基于 GELAN 的框架特性,本文在安全帽检测任务中选择将 RepNCSP 模块和 Conv 模块作为计算块构成 RepNCSP-SPELAN4,其网络结构如图 6 所示。将 Neck 网络中的 C2f 全部替换为 RepNCSP-SPELAN4 可使模型更有效地进行特征融合,改善被遮挡目标的检测情况。

如图 7 所示,RepNCSP 的一路分支由 n 个 RepNBottleneck 堆叠而成的,与 YOLOv8 的 Bottleneck 相比,RepNBottleneck 将第一个 Conv 模块替换为 RepConvN,而 RepConvN 即上文所述的无 BN 层分支的 RepVGGBlock,同样是在训练时使用多分支结构而在推理时将其融合为单分支。

2.2.2 小目标检测层

为进一步提升模型检测小目标的能力,再增加 1 个小目标检测层。在 Neck 网络中进行特征融合会得到 3 种尺寸的特征图用于检测不同大小的目标,当输入图片尺寸为 640×640 时,这 3 种尺寸分别为 80×80 、 40×40 和 20×20 ,其中较大尺寸的特征图含有较多的局部细节特征和目标位置信息,对小目标的检测较为友好;而小尺寸特征图经过

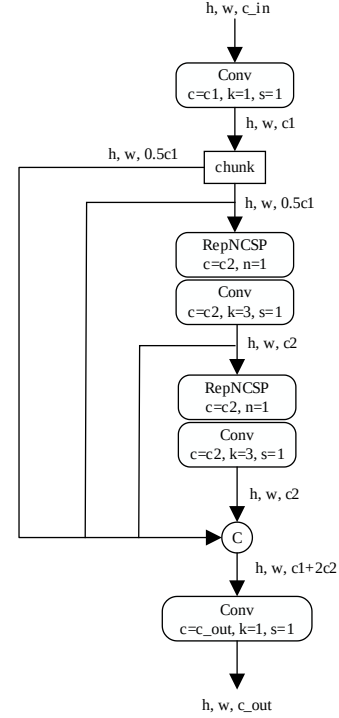


Fig. 6 RepNCSP-SPELAN4 network structure

图6 RepNCSP-SPELAN4网络结构

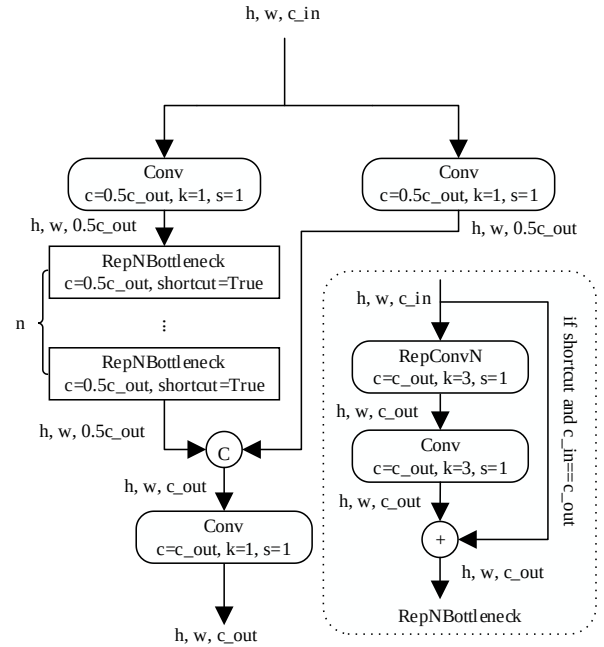


Fig. 7 RepNCSP network structure

图7 RepNCSP网络结构

一系列的特征提取后虽然有丰富的语义信息但局部细节不明显,因而更适用于检测大目标^[17]。

将 Backbone 中的浅层特征图与 Neck 中的深层特征图进行拼接,再经过 RepNCSP-SPELAN4 进一步进行特征融合后输出尺寸为 160×160 的特征图,以增强特征传递与特征融合,使模型更加关注于小目标的检测,有效降低了其误检率和漏检率。

2.3 损失函数

CIoU 虽然考虑了边界框回归的距离、重叠面积、中心点偏移和宽高比, 但仍存在一些缺点, 如未考虑训练数据集中高质量样本与低质量样本之间的边界框回归平衡问题, 且当预测框与真实框的宽高比相同但宽度和高度的值不同时, 将宽高比作为惩罚因子不能体现两个框之间的真实差异。为应对上述问题, 本文使用基于注意力的边界框损失函数 WIoU v1 替换 CIoU。在预测框与真实框能较好地重合时, WIoU v1 会削弱几何度量的惩罚, 从而不过多地干预训练, 使模型具有更好的泛化能力。此外, 其在训练过程中更关注那些能提高模型性能的样本, 同时利用距离度量构建了距离注意力, 得到具有两层注意力机制的损失函数, 帮助模型更好地捕捉预测框与真实框之间的细微差异。IoU 损失函数用公式表示为:

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{h_{gt} w_{gt} + hw - W_i H_i} \quad (9)$$

式中: W_i 和 H_i 分别为相交矩形的宽和高; h_{gt} 和 w_{gt} 分别为目标框的高和宽; h 和 w 分别为锚框的高宽。

IoU 示意图如图 8 所示。图中, (x_{gt}, y_{gt}) 为目标框中心点 A 的坐标, 锚框中心点 B 的坐标为 (x, y) 。首先, 构造距离注意力 R_{WIoU} , 用于衡量两个框之间的相对位置关系, 实现对几何因素的考虑; 其次, 将基本 IoU 与距离注意力权重相乘得到 WIoU v1。表示为:

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (10)$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (11)$$

式中: W_g 和 H_g 分别为最小外接矩形的宽和高; “*” 表示将 W_g 和 H_g 从计算图中脱离, 以避免在计算回归损失时二者的值产生的梯度影响模型收敛, 从而使模型在训练时更专注于与目标框匹配的锚框, 而不受到外部尺寸和位置因素的干扰, 因此无需引入新的度量指标, 如宽高比。

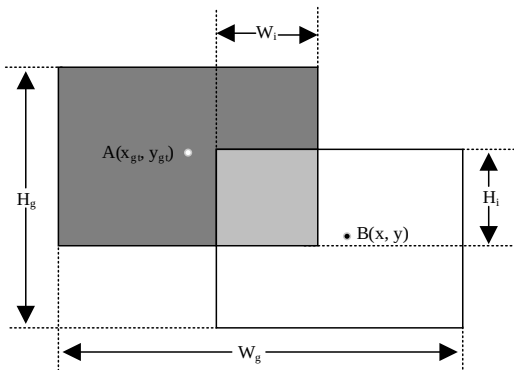


Fig. 8 Illustration of IoU

图 8 IoU 示意图

综上所述, 改进后的 YOLOv8 网络结构如图 9 所示。

3 实验方法与结果分析

3.1 实验环境与参数设置

本文所有实验均在表 1 所示的环境中进行。

实验相关参数设置见表 2。

3.2 数据集

本文使用的数据集为公开的 SHWD, 该数据集包含工地上的各种复杂环境, 待检目标有多种尺寸, 亦有许多被遮挡和密集小目标满足本文对实际场景的需求。SHWD 中共有 7 581 张图片, 包含 2 个类别, 一种是正样本, 即佩戴安全帽的人头 (helmet); 另一种是负样本, 即未佩戴安全帽的人头 (person)。将该数据集按照 8:1:1 的比例随机划分为训练集 (6 064 张)、验证集 (759 张) 和测试集 (758 张)。

3.3 评价指标

采用精确率 (Precision, P)、召回率 (Recall, R)、平均精度均值 (mean Average Precision, mAP)、单张图片检测用时 (Times)、参数量 (Params) 和浮点运算量 (FLOPs) 来衡量模型性能。二分类混淆矩阵见表 3。

P 为在所有被预测为正例的样本中真正例所占的比例, 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的比例, 其计算公式分别为式 (12) 和式 (13)。mAP 为常用于评估目标检测模型性能的指标, 其综合评价了各个类别的 P 和 R 。首先计算每个类别 P-R 曲线下的面积求得平均精度 (Average Precision, AP), 然后对各个 AP 取平均值即可得到 mAP, 其值越高表明检测性能越好, 计算公式见式 (14)。mAP@0.5 指在 IoU 阈值为 0.5 的情况下计算 AP, 而 mAP@0.5:0.95 是指在计算 mAP 时使用从 0.5~0.95 (步长为 0.05) 的多个 IoU 阈值, 然后再对所有 IoU 阈值下的 mAP 取平均。

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$mAP = \frac{1}{N} \sum_{n=0}^{N-1} \int P_n(R) dR \quad (14)$$

式中: TP 为真正例, FP 为假正例, FN 为假负例, n 为类别数, 对 P_n 积分即为某类的 AP。

Times 用于衡量模型的检测速度, 即推理时将批量大小设为 1 时检测 1 张图片平均所需要的时间; Params 通过参数量来衡量模型大小; FLOPs 则用于评价模型的复杂度。

3.4 与主流算法的比较

为验证本文算法的有效性, 将其与性能优越的主流目标检测算法进行比较, 包括经典的 Faster R-CNN 和 SSD, 基于 Transformer 的模型 RT-DETR^[18], 被广泛应用的 YOLOv5、YOLOv6^[19] 和 YOLOv7^[20], 以及 YOLO 系列的最新算法 YOLOv9^[16] 与 YOLOv10^[21]。所有实验均使用同一训练

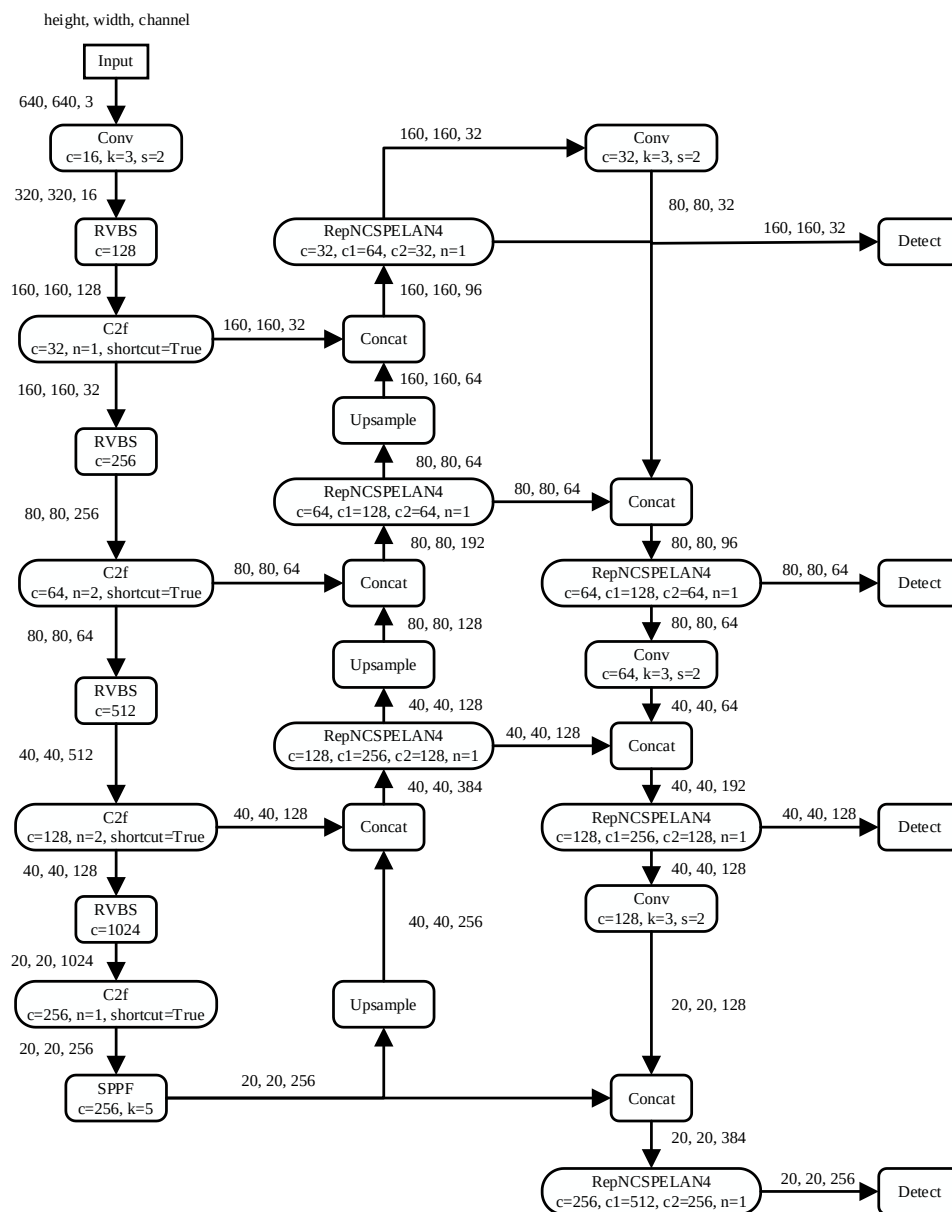


Fig. 9 Improved YOLOv8 network structure

图9 改进后的YOLOv8网络结构

Table 1 Experimental environment

表1 实验环境

项目	设置
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 parameter settings

表2 实验参数设置

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	0.005

Table 3 Binary classification confusion matrix

表3 二分类混淆矩阵

真实值	预测值	
	正例	负例
正例	TP	FN
负例	FP	TN

集和验证集,各算法评价指标比较结果见表4。

可以看出,本文算法的 $mAP@0.5$ 为 94.8%, R 为 90.3%,均优于其他对照算法; P 为 91.6%,仅次于 YOLOv5 和 YOLOv9,说明改进算法能有效降低原算法的漏检率和

Table 4 Comparison results of evaluation indicators for various algorithms

表 4 各算法评价指标比较结果							
算法	P	R	mAP@0.5	mAP@0.5: 0.95	Times	Params	FLOPs
	/%	/%	/%	/%	/ms	/M	/G
Faster R-CNN	55.8	76.1	70.8	36.6	89.8	28.3	474.1
SSD	90.5	63.5	84.1	44.9	37.9	23.7	136.8
RT-DETR	87.2	82.9	88.9	56.8	44.3	15.5	37.4
YOLOv5	92.9	85.8	91.5	57.8	20.6	1.8	4.1
YOLOv6	89.6	87.8	91.4	58.7	29.0	4.6	11.3
YOLOv7	90.8	87.4	92.4	59.2	24.7	2.3	6.6
YOLOv8	90.7	86.3	92.0	60.5	19.0	3.0	8.1
YOLOv9	92.5	89.3	94.8	64.0	42.0	13.9	52.3
YOLOv10	90.5	86.4	91.9	59.8	19.9	2.7	8.2
本文算法	91.6	90.3	94.8	63.7	25.7	6.5	25.5

误检率。虽然 YOLOv9 算法的 mAP@0.5 与本文算法一样,均达到最大值 94.8%,但其参数量和浮点运算量较大,导致检测时需要更多算力,单张图片检测用时高达 42.0 ms,而本文算法仅为 25.7 ms。总体来看,本文算法更适用于对精度和实时性均要求较高的安全帽检测场景。其他对照算法,如 Faster R-CNN、SSD 和其他 YOLO 算法的检测速度略快于本文算法,但检测精度较低,不适用于安全帽检测。

部分算法的 P-R 曲线如图 10 所示,通过该曲线可以直观了解各算法在不同置信度阈值下的精确率和召回率表现。P-R 曲线越接近右上角,其与坐标轴围成的面积越大,即精准率和召回率越高,表明算法性能越好。由图 10 可知,本文算法和 YOLOv9 算法的 P-R 曲线最接近左上角,二者的检测性能优于其他算法。

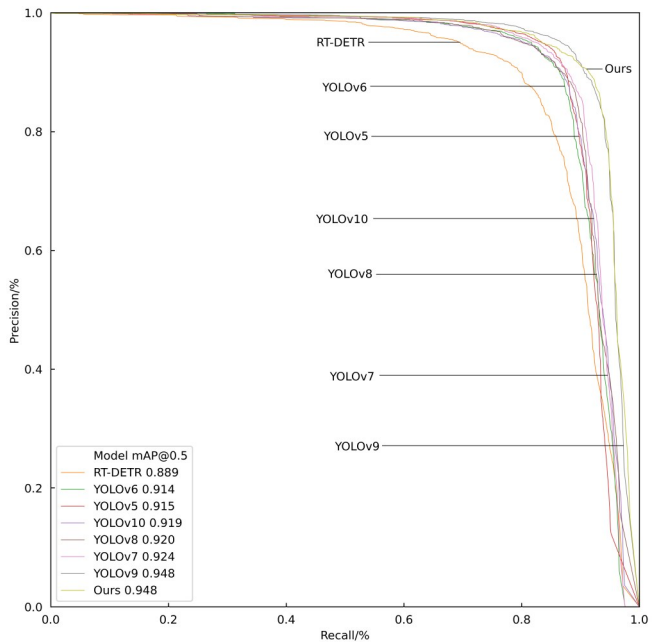


Fig. 10 P-R curve of some algorithms
图 10 部分算法的 P-R 曲线

3.5 消融实验

3.5.1 客观数据分析

为验证本文算法各个改进点的有效性,选用基线模型 YOLOv8n 在测试集上进行 10 组消融实验,结果见表 5。可以看出,原始算法的 mAP@0.5 和 mAP@0.5: 0.95 均为最低,分别为 92.0% 和 60.5%,但其检测速度最快,单张图片检测用时仅为 19.0 ms,参数量 Params 和浮点运算数 FLOPs 分别为 3 M 和 8.1 G。仅加入 RVBS 模块时,模型的 mAP@0.5 和 mAP@0.5: 0.95 分别提高了 1.1% 和 1.2%,在改进后的 YOLOv8 算法中去掉该模块后 mAP@0.5 和 mAP@0.5: 0.95 分别减少了 0.6% 和 0.8%,可见加入多分支训练和采用 SPD 中的切片取代大步幅卷积能提高模型的特征提取能力。仅采用 RVBS 模块替换 Conv 模块后,模型参数量上升了 0.3 M,这是由于 RVBS 模块输出特征图的通道数比 Conv 模块更多,导致下一模块 C2f 的输入特征图通道数同样加深,从而增加了卷积层数量,使参数量略微上升。RVBS 模块中虽然采用了多分支结构,但模型训练完成后会进行结构重参数化来合并各分支的参数,因此仅增加少量参数便可获取较大的特征提取能力提升。引入 RVBS 模块后,模型的浮点运算数增加了 3.4 G,这是由于在卷积运算后对特征图新增了切片和拼接操作,后续 C2f 模块中增加的卷积层也会导致浮点运算数上升。然而,引入 RVBS 模块替换 Conv 模块后,单张图片检测用时仅增加了 1.7 ms。

Table 5 Ablation experiment results
表 5 消融实验结果

RVBS	RepNCS-PELAN4	小目标检测层	WIoU v1	mAP@0.5 /%	mAP@0.5: 0.95 /%	Times /ms	Params /M	FLOPs /G
×	×	×	×	92.0	60.5	19.0	3.0	8.1
√	×	×	×	93.1	61.7	20.7	3.3	11.5
×	√	×	×	92.5	61.4	24.7	6.2	14.6
×	×	√	×	94.0	62.1	22.7	2.9	12.2
×	×	×	√	92.4	60.4	19.5	3.0	8.1
×	√	√	√	94.2	62.9	28.9	6.2	22.0
√	×	√	√	94.6	62.7	24.1	3.2	15.6
√	√	×	√	93.4	62.2	23.4	6.4	18.1
√	√	√	×	94.6	63.3	28.3	6.5	25.5
√	√	√	√	94.8	63.7	25.7	6.5	25.5

仅加入 RepNCSPELAN4 模块时,模型的 mAP@0.5 和 mAP@0.5: 0.95 分别提高了 0.5% 和 0.9%,而在改进后的 YOLOv8 中去掉该模块后, mAP@0.5 和 mAP@0.5: 0.95 分别下降了 0.2% 和 1.0%,可见该模块比 C2f 模块能更有效地传播和聚合来自不同层级的特征信息。加入该模块后,模型的参数量和浮点运算数均有所增加。这是由于 1 个 C2f 模块中仅有 1 个计算块,且为单分支结构;而 RepNCSPELAN4 中有 2 个计算块,每个计算块又引入了多分支结构,这些操作在获得丰富梯度流信息的同时也增加了卷积层个数,因而导致参数量和浮点运算数增加。

仅加入小目标检测层时,模型的 mAP@0.5 和 mAP@0.5: 0.95 分别提高了 2.0% 和 1.6%,而在改进后的 YOLOv8

中去除该模块后 $mAP@0.5$ 和 $mAP@0.5: 0.95$ 分别降低了 1.4% 和 1.5%; 在训练时使用 WIoU v1 替换 CIoU 后模型的 $mAP@0.5$ 提高了 0.4%, 在改进后的 YOLOv8 算法中去除该损失函数后模型的 $mAP@0.5$ 和 $mAP@0.5: 0.95$ 分别降低了 0.2% 和 0.4%, 证实了二者对于提升目标检测精度的有效性。

通过比较单独引入各模块时对模型 mAP 的提升程度可知, 小目标检测层的贡献最为突出, RVBS 模块和 RepNCSPELAN4 次之, 而 WIoU v1 损失函数虽然进一步优化了回归精度, 但与前 3 个模块相比贡献略逊。加入全部改进模块的本文算法的 $mAP@0.5$ 和 $mAP@0.5: 0.95$ 最高, 相较于基线模型分别提高了 2.8% 和 3.2%, 且单张检测用时仅为 25.7 ms, 能较好地平衡检测精度与速度, 满足实时检测需求。

各消融实验对应的 P-R 曲线如图 11 所示, 其中“+”表示在原始 YOLOv8 算法的基础上仅加入某一个改进点, “-”表示在本文算法的基础上仅去掉某一个改进点。可以看出, 加入所有改进点后, 模型的 P-R 曲线下面积最大, 而去掉任何一个改进点面积均会变小, 表明各个改进点对于提高检测性能均起到了一定作用。

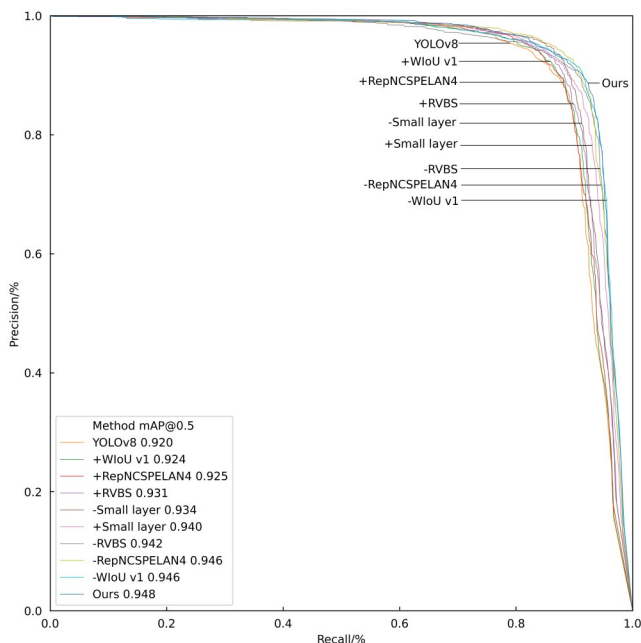


Fig. 11 P-R curves of ablation experiments

图 11 消融实验的 P-R 曲线

图 12 为本文算法和原始算法在验证集上的 box_loss 曲线比较。可以看出, 与原始算法相比, 本文算法的边界框回归损失值更低, 收敛速度更快, 曲线也更平稳。

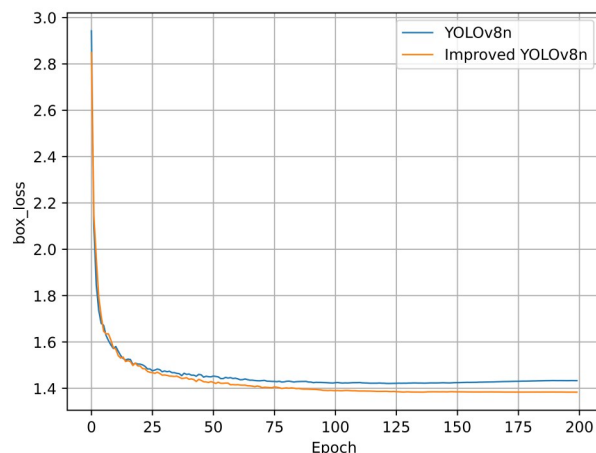


Fig. 12 box_loss curves

图 12 box_loss 曲线

图 13 为改进前后算法的 $mAP@0.5$ 曲线比较。可以看出, 随着迭代次数的增加, 二者的曲线均逐渐收敛, 趋于平稳, 但本文算法的 $mAP@0.5$ 值更高。

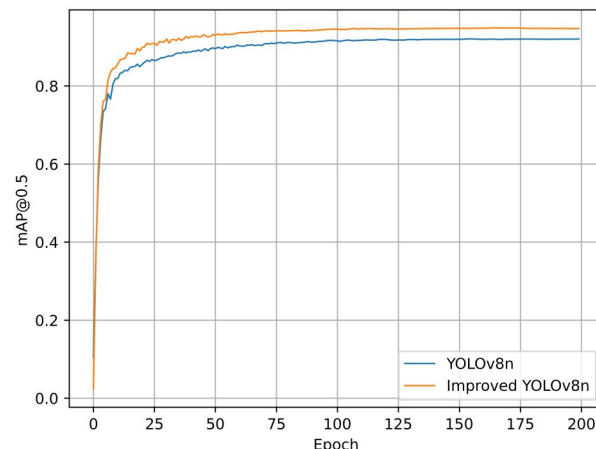


Fig. 13 mAP@0.5 curves

图 13 mAP@0.5 曲线

3.5.2 检测结果可视化分析

为更直观地展现改进前后 YOLOv8 算法的检测效果, 采用实际工地图片进行验证实验。检测结果如图 14 所示, 其中蓝圈表示误检, 黄圈表示漏检。



(a) Dense scene (before improvement)

(a) 密集场景 (改进前)



(b) Dense scene (after improvement)

(b) 密集场景 (改进后)



Fig. 14 Comparison of detection results of YOLOv8 algorithm before and after improvement

图 14 改进前后 YOLOv8 算法检测结果比较

图14(a)和图14(b)为密集场景,可以看出,原始算法在检测密集人群时易发生误检,即将一个安全帽误认为多个安全帽,或将背景误检为安全帽,误检数达到3个,同时有3个安全帽被漏检。而改进算法未发生误检现象,仅有1个戴安全帽的人头因遮挡过多而未被检出。图14(c)和图14(d)为遮挡场景,戴安全帽的人头被前面的人脸遮挡一半后,原始算法便将两者重叠后的部分误检为安全帽,而改进算法能准确标注安全帽和人头,将两者区分开来。图14(e)和图14(f)为小目标场景,含有较多难以检测的小目标,也易将人的上半部分误认为安全帽。原始算法对安全帽的漏检数高达16个,改进算法将漏检数降低至8个,同样也减少了误检数。图14(g)和图14(h)为多尺寸场景,在同一张图中含有大、中和小3种不同尺寸的安全帽。原始算法漏检了部分中等尺寸和小尺寸的安全帽,漏检数达5个,而改进算法仅漏检了某个过于微小的目标。图14(i)和图14(j)为复杂场景,可以看出地上摆放了多种建筑材料,并有大量垃圾,检测环境极其复杂。原始算法易将未戴安全帽的人头或背景误检为安全帽,误检数为7个,漏检情况更严重,高达11个。改进算法将误检数降至4个,漏检情况也得到大幅改善,仅有4个目标未被检测出来。图14(k)和图14(l)为暗光场景,光照不足会增加检测难度。原始算法漏检现象严重,而改进算法在黑暗环境下仍能准确识别。图14(m)和图14(n)为俯视场景。原始算法将人的背部误检为安全帽,而真正的安全帽却未被检测到,改进算法则能准确无误地检测出安全帽。

3.6 不同损失函数的比较

为验证训练时使用的WIoU v1损失函数的有效性,将其与主流损失函数进行比较。实验均在同一硬件环境和验证集下进行,结果如图15所示。与原始模型的CIoU损失函数相比,使用DIoU、EIoU、GIoU、和MPDIoU训练后的模型的mAP@0.5分别下降了0.318%、0.168%、0.258%和0.177%,说明这些主流损失函数并不适用于本文场景。使用SIoU训练后,模型的mAP@0.5仅增加了0.117%;使用WIoU v1训练后,模型的mAP@0.5可增加0.422%。可以看出,WIoU v1可较大程度上提升检测精度。此外,使用

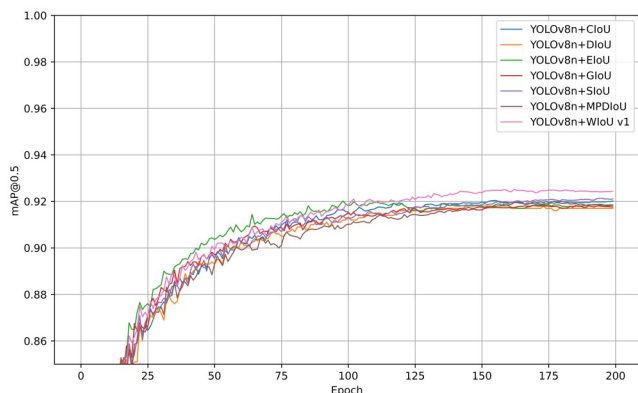


Fig. 15 Comparison of different loss function curves

图15 不同损失函数曲线比较

WIoU v1的模型训练至Epoch为100~125之间时,其mAP@0.5便开始领先于其他损失函数,然后逐渐趋于平稳,可见WIoU v1可使用更少的训练时间获得同样的检测精度,能加速模型收敛。

4 结语

针对工地安全帽检测中对密集遮挡小目标容易误检和漏检的问题,本文提出一种基于改进YOLOv8的安全帽实时检测算法,其通过以下改进措施提高了检测性能,包括:①将RVBS模块引入特征提取网络中,通过多分支结构训练模型,然后重参数化合并各分支以降低参数量,并使用切片与拼接替换大步幅下采样卷积来减少特征信息丢失,最终减少了对密集小目标的误检和漏检现象;②使用聚合特征信息更强的RepNCSPeLan4模块,使模型捕捉更多全局特征,缓解目标被遮挡时的误检和漏检问题,并新增小目标检测层,提高对小目标的识别能力;③使用WIoU v1损失函数更新模型权重,加快网络收敛并提升回归精度。实验结果表明,改进算法比原始算法和其他主流检测算法更适用于对精度和实时性要求较高的安全帽检测。未来工作要在保证高检测精度的前提下进一步降低模型参数量和计算量,从而缩短检测时间,实现更即时的检测;同时扩充数据集,引入不同天气、光照条件下的检测场景,进一步增强模型的泛化能力。

参考文献:

- [1] ZHANG L Y, WU W H, NIU H M, et al. Summary of application research on helmet detection algorithm based on deep learning[J]. Computer Engineering and Applications, 2022, 58(16): 1-17.
张立艺, 武文红, 牛恒茂, 等. 深度学习中的安全帽检测算法应用研究综述[J]. 计算机工程与应用, 2022, 58(16): 1-17.
- [2] DENG K F, ZOU Z Y. Realization and analysis of helmet wearing detection based on deep learning[J]. Computer Era, 2020(7): 12-15, 21.
邓开发, 邹振宇. 基于深度学习的安全帽佩戴检测实现与分析[J]. 计算机时代, 2020(7): 12-15, 21.
- [3] LI M S, HAN Q P, ZHANG T Y, et al. Safety helmet detection method of improved SSD[J]. Computer Engineering and Applications, 2021, 57(8): 192-197.
李明山, 韩清鹏, 张天宇, 等. 改进SSD的安全帽检测方法[J]. 计算机工程与应用, 2021, 57(8): 192-197.
- [4] LI F F, CHEN Y, HU M, et al. Helmet-wearing tracking detection based on strongSORT[J]. Sensors, 2023, 23(3): 1682.
- [5] LYU Z Z, XU H, YANG X, et al. Small object detection algorithm of YOLOv5 for safety helmet[J]. Journal of Computer Applications, 2023, 43(6): 1943-1949.
吕宗喆, 徐慧, 杨晓, 等. 面向小目标的YOLOv5安全帽检测算法[J]. 计算机应用, 2023, 43(6): 1943-1949.
- [6] BAO J J, LI S H, WANG G L, et al. Improved YOLOv8 network and application in safety helmet detection[J]. Journal of Physics: Conference Series

- ries, 2023, 2632: 012012.
- [7] LIN B Y. Safety helmet detection based on improved YOLOv8[J]. IEEE Access, 2024, 12(1): 28260–28272.
- [8] 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>.
- [9] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993–13000.
- [10] NJVISIONPOWER. Safety-helmet-wearing-dataset [EB/OL]. <https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset>.
- [11] LI X, WANG W H, WU L J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020, 33(1): 21002–21012.
- [12] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style convnets great again [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733–13742.
- [13] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022: 443–459.
- [14] 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.
- [15] WANG C Y, LIAO H Y M, YEH I H. Designing network design strategies through gradient path analysis [J]. Journal of Information Science and Engineering, 2023, 39(4): 975–995.
- [16] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information [DB/OL]. <https://arxiv.org/abs/2402.13616>.
- [17] WEI C H, YANG R, LIU Z B, et al. YOLOv8 with bi-level routing attention for road scene object detection [J]. Journal of Graphics, 2023, 44(6): 1104–1111.
- 魏陈浩, 杨睿, 刘振丙, 等. 具有双层路由注意力的 YOLOv8 道路场景目标检测方法[J]. 图学学报, 2023, 44(6): 1104–1111.
- [18] ZHAO Y A, LV 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.
- [19] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.
- [20] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464–7475.
- [21] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection [DB/OL]. <https://arxiv.org/abs/2405.14458>.

(责任编辑:尹晨茹)