

复杂场景下的改进 YOLOv8n 安全帽佩戴检测算法

雷源毅,朱文球,廖欢



(湖南工业大学计算机学院,湖南 株洲 412007)

✉ yuanyi_lei@foxmail.com; wenqiu_zhu@126.com; huanliao@foxmail.com

摘要:为解决工地场景下安全帽检测因背景复杂、目标密集等导致 YOLOv8n 算法识别精度低、易出现漏检及误检等问题,提出一种改进 YOLOv8n 的 RDCA-YOLO 安全帽检测算法。通过改进 Backbone 结构提升模型特征提取能力;嵌入 CBAM 注意力机制,增强对小目标信息提取的能力;设计一种 Coord-BiFPN 结构,增强网络特征融合能力;提出一种 OD-C2f 结构,实现提取不同形状和大小安全帽的关键特征;设计一种 FR-DyHead 检测头替换原始 Detect 结构,提升了检测精度并降低了延时。使用 GDUT-HWD 数据集进行训练测试,结果表明所提算法的 *mAP* 达到 85.8%,相比 YOLOv8n 提升了 2.6%,能有效提高复杂场景下的安全帽佩戴检测精度。

关键词:YOLOv8;CBAM 注意力机制;Coord-BiFPN;FR-DyHead 检测头

中图分类号:TP391.4 **文献标志码:**A

Helmet Wearing Detection Algorithm in Complex Scenes Based on Improved YOLOv8n

L EI Yuanyi, ZHU Wenqiu, LIAO Huan

(School of Computer, Hunan University of Technology, Zhuzhou 412007, China)

✉ yuanyi_lei@foxmail.com; wenqiu_zhu@126.com; huanliao@foxmail.com

Abstract: Complex background and dense targets of helmet detection in construction site scenes lead to low recognition accuracy, easily missed detection and false detection of YOLOv8n algorithm. To solve these problems, this paper proposes a RDCA-YOLO safety helmet detection algorithm based on improved YOLOv8n. Firstly, model feature extraction ability is enhanced by improving Backbone structure and CBAM (Convolutional Block Attention Module) attention mechanism is embedded to improve small target information extraction. Secondly, a Coord-BiFPN structure is designed to enhance network feature fusion capability, and an OD-C2f structure is proposed to extract key features of safety helmets with different shapes and sizes. Finally, a FR-DyHead detection head is designed to replace the original Detect structure, so as to improve detection accuracy and reduce latency. Training and testing are conducted on GDUT-HWD helmet dataset. Test results show that the *mAP* of the proposed algorithm achieves 85.8% accuracy, which is 2.6% higher than YOLOv8n, which verifies that the proposed algorithm can effectively improve the accuracy of helmet wearing detection in complex scenes.

Key words: YOLOv8; CBAM attention mechanism; Coord-BiFPN; FR-DyHead detection head

0 引言(Introduction)

由于建筑行业施工环境复杂,高空物体坠落的伤害风险高,因此要求工人在施工现场佩戴安全帽,最大限度地降低工

人被坠落物体伤害的风险,保护工人的生命安全。目前,用于工地经典的目标检测算法分为两种,即一阶段目标检测算法和两阶段目标检测算法。一阶段目标检测算法包括 YOLO 系列

算法^[1]、SSD(单激发多框检测)算法^[2]等。两阶段目标检测算法包括 R-CNN^[3](基于区域的卷积神经网络)、Faster R-CNN^[4]、Mask R-CNN 等。YOLO 系列算法是目前为止发展最快、最好的算法,尤其是 2023 年发布的最新的 YOLOv8 算法,达到了目前为止最高的检测准确率。然而,YOLO 虽然检测全尺寸的目标效果较好,但是当目标为具有不同尺寸的特殊场景时,其性能不如当前的一些小尺寸目标检测算法。工地安全帽目标的尺寸较小,使得检测器很难准确、全面地提取其特征,并且当前各种算法对小物体的检测准确率普遍较低。

1 相关研究(Related research)

安全帽图像具有目标分布密集、目标尺度变化大及背景复杂等特点,给安全帽佩戴检测带来了很大的难度。孙国栋等^[5]通过在 Faster R-CNN 中添加自注意力机制获取多尺度的全局信息,强化网络对小目标的表达信息,但该方法的检测速度较低。许凯等^[6]使用参数量较少的轻量化模型 YOLOv3-tiny,采用增加网络层数、添加注意力机制到特征金字塔中的方法,减少噪音等冗余信息对安全帽检测结果的影响,但在密集情况下,出现较多漏检情况。杜晓刚等^[7]在 YOLOv5 的基础上选择性能更好的 Swin Transformer 作为主干网络,提取更深层次的语义信息,捕获更多的安全帽细节特征,但在颜色相同情况下存在误检问题。赵敏等^[8]通过在 YOLOv7-tiny 的基础上设计参数量更少的 tiny-BiFPN 结构,作为原模型特征融合模块中的特征金字塔结构,增强模型多尺度特征融合,但在反光情况下漏检较多。上述改进算法的检测精度仍不够准确,背景特征提取也不够丰富,未能很好地解决建筑工地中小目标密集、安全帽和背部同色等场景下的漏检、误检问题。针对这些问题,提出一种改进的 RDCA-YOLO 方法,该方法能够提高建筑工地复杂场景中小目标的检测精度,对正常尺度目标的检测精度也有较大的提高,实现了工地工人安全帽佩戴的自动化检测。

2 YOLOv8 算法(YOLOv8 algorithm)

Ultralytics 于 2023 年发布了 YOLOv8 算法,与之前的网络相比,YOLOv8 在减少网络参数量的同时,提高了检测精度和实时性。针对建筑工地实际场景,选择一个轻量版本的 YOLOv8n 作为基础网络进行改进。YOLOv8 的特点在于融合了许多实时目标检测中优异的技术,仍然采用了 YOLOv5 中的 CSP(跨阶段局部网络)思想、特征融合方法(PAN-FPN)和 SPPF 模块;其主要改进如下:(1)提供了全新的基准模型,包括分辨率为 640×640 的 P5 和分辨率为 1280×1280 的 P6 目标检测网络。为了满足不同项目的需求,还基于与 YOLOv5 相似比例系数设计了不同比例的模型。(2)在保留 YOLOv5 原有思想的前提下,参考 YOLOv7 中的 ELAN(高效层聚合网络)结构设计了 C2f(跨阶段局部网络卷积块)模块。(3)检测头部分也采用了目前流行的方法,将分类头和检测头分开,其他大部分仍然是基于 YOLOv5 的思想。(4)分类损失使用 BCE 损失(多标签分类任务的损失函数),回归损失的形式为 CIoU 损失函数。与以往的 YOLO 算法相比,YOLOv8 具有很强的扩展性。

在 YOLOv8n 算法的基础上,建立了一个 RDCA-YOLO

复杂建筑工地情况下的安全帽检测模型。通过改进 Backbone 主干结构,在网络中嵌入 CBAM 注意力机制,提出一种 Coord-BiFPN 结构,增强网络的特征融合能力,设计了一种 OD-C2f 结构代替 head 中的 C2f 组件,提出一种新的 FR-DyHead 检测头替换原始的 Detect 结构,其网络结构如图 1 所示,该方法可以适应建筑工地复杂场景,并获得了更好的安全帽检测精度。

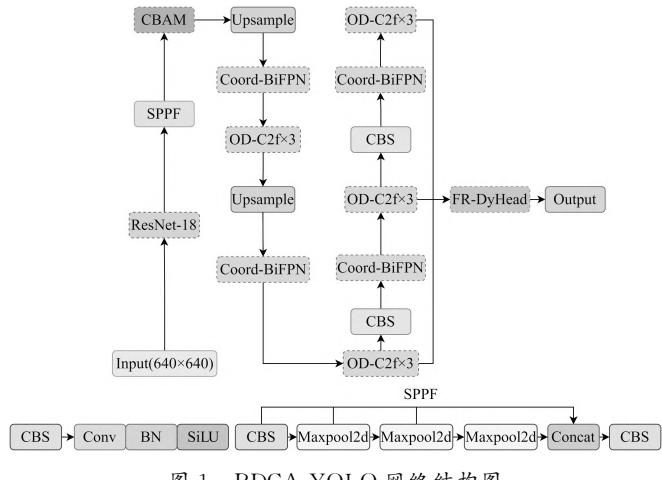


Fig.1 Diagram of RDCA-YOLO network structure

3 RDCA-YOLO 安全帽检测模型(RDCA-YOLO helmet detection model)

3.1 改进 Backbone 结构

原始网络中的主干网络较为复杂,提取信息的能力较低,因此模型通过 ResNet-18^[9] 网络改进原有的 Backbone 结构,该网络由一个卷积层和八个残差块组成,能够大幅度提高模型的特征提取能力,残差块是组成网络的基本结构,输出如公式(1)所示:

$$y = F(x) + x \quad (1)$$

其中: $F(\cdot)$ 表示残差函数, x 和 y 分别表示输入和输出。网络由 17 个卷积层、一个滤波器大小为 3×3 的最大池化层(max pooling)和一个全连接层组成。传统的 ResNet-18 模型涉及 3 316 万个参数,其中 ReLU 激活函数和批量归一化(Batch Normalization)应用于整个卷积层的后面,ResNet-18 结构如表 1 所示。

表 1 ResNet-18 结构表

Tab.1 ResNet-18 structure table

层名	输出大小	18 层
Conv1	112×112	$7 \times 7, 64, \text{stride } 2$ $3 \times 3, \text{max pool, stride } 2$
Conv2_x	56×56	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$
Conv3_x	28×28	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 2$
Conv4_x	14×14	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 2$
Conv5_x	7×7	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$

3.2 注意力机制

通过在网络中 Head 前嵌入 CBAM 注意力机制^[10],使网络更加关注感兴趣的目标,CBAM 原理图如图 2 所示,其中 x 表示中间变量。

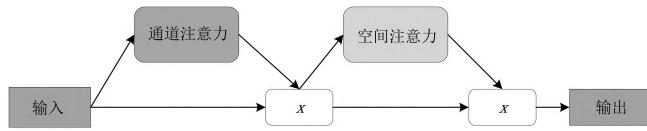


图 2 CBAM 原理图

Fig.2 Diagram of CBAM principle

通道注意力实现原理:输入 $H \times W \times C$ 的特征图, H 和 W 表示输入图像的高和宽, C 表示通道数, 使用全局平均池化和最大平均池化收集空间信息, 然后用两个 1×1 的卷积(MLP, 多层感知机)和 Sigmoid 激活函数获得每个通道的权重, 通道注意力模型结构图如图 3 所示。

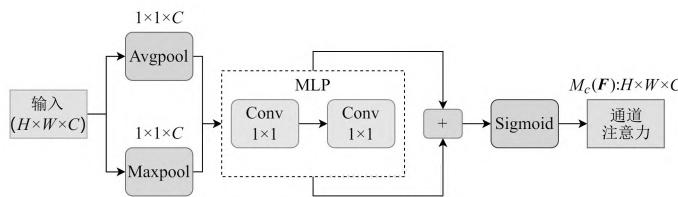


图 3 通道注意力模型结构图

Fig.3 Diagram of channel attention model structure

空间注意力实现原理:首先对通道信息进行最大池化和平均池化操作, 并将结果连接起来, 其次由一个 7×7 的卷积和 Sigmoid 激活函数得到每个点的坐标权重, 设置填充 Padding 为 3, 空间注意力模型结构图如图 4 所示。

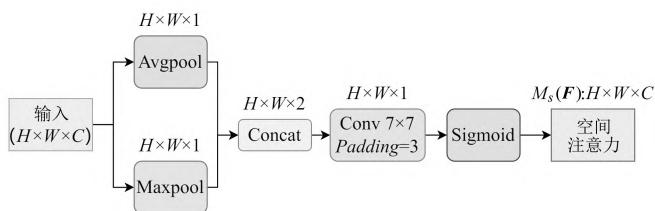


图 4 空间注意力模型结构图

Fig.4 Diagram of spatial attention model structure

注意力模块的计算方法如公式(2)和公式(3)所示, \mathbf{F} 为输入特征矩阵, $M_c(\mathbf{F})$ 和 $M_s(\mathbf{F}')$ 分别表示通道注意力和空间注意力所获得的权重, \otimes 表示逐元素相乘, \mathbf{F}' 表示通道加权结果, \mathbf{F}'' 表示 CBAM 的最终结果。

$$\mathbf{F}' = \mathbf{F} \otimes M_c(\mathbf{F}) \quad (2)$$

$$\mathbf{F}'' = \mathbf{F}' \otimes M_s(\mathbf{F}') \quad (3)$$

空间注意力模块和通道注意力模块的具体计算方法如公式(4)和公式(5)所示, $f^{n \times n}$ 表示一个 $n \times n$ 的卷积。“ \oplus ”表示拼接操作, “ $+$ ”表示逐元素求和。

$$M_s(\mathbf{F}') = \delta(f_1^{7 \times 7}(AvgPool(\mathbf{F}') \oplus MaxPool(\mathbf{F}'))) \quad (4)$$

$$M_c(\mathbf{F}) = \delta(f_0^{1 \times 1}(f_0^{1 \times 1}(AvgPool(\mathbf{F}) + f_0^{1 \times 1}(MaxPool(\mathbf{F})))))) \quad (5)$$

在通道注意力中, 两个 1×1 的卷积在经过池化操作后表示为 $f_0^{1 \times 1}$ 和 $f_1^{1 \times 1}$, 然后经过 Sigmoid 激活函数获得了取值范围为 $0 \sim 1$ 的权重 $M_c(\mathbf{F})$ 。在空间注意力中, 采用一个 7×7 的卷积 $f_1^{7 \times 7}$ 获取池化和拼接操作后空间维度的权重信息 \mathbf{F}' , 如公式(4)所示, 用相同方式获取每一个点的权重 $M_s(\mathbf{F}')$ 。从上述运算可以发现 CBAM 参数主要来自于卷积运算, 卷积的参数与卷积核的大小和输入、输出通道数有关, 为了尽可能地减少 CBAM 的计算开销, 降低 CBAM 外部参数, 可以将通道数 $f_0^{1 \times 1}$ 设置为输入通道的 $1/r$, 再用 $f_1^{1 \times 1}$ 恢复通道数(r 为压缩比)。针对不同的情况, 设置也不同。

3.3 Coord-BiFPN 结构

模型设计了一个 Coord-BiFPN 的颈部网络, 融合添加了 BiFPN^[11] 和 CoordConv^[12], 增强了网络的特征融合能力和特征映射的鲁棒性。Coord-BiFPN 的特点在于高效的双向跨尺度连接和加权特征融合(图 5)。首先, 从主干网络输入 P_1 、 P_2 、 P_3 三个不同大小的特征图, P_1 和 P_3 的特征直接融合到输出, 这两个阶段中只有一个输出边缘, 并且对合并不同的特征信息影响很小。其次, 在 P_2 层的输出节点上添加一条额外的边, 融合更多特征的同时, 也不会增加计算成本。最后, 增加 P_3 和 P_2 的输入层与 P_1 的输出层之间的连接, 以获得更高级别的特征融合。

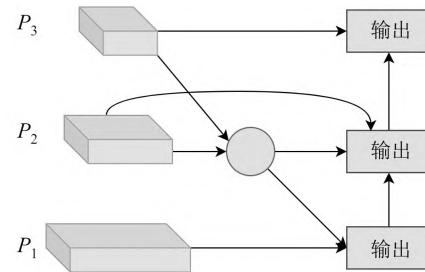


图 5 Coord-BiFPN 结构图

Fig.5 Diagram of Coord-BiFPN structure

当融合不同分辨率的特征时, 通常的做法是将它们调整到相同的分辨率, 然后将它们相加。然而, 当输入特征出现变化时, 它们的输出特征的权重通常是不同的。向 Coord-BiFPN 的每个输入添加额外的权重, 让网络学习每个输入特征的重要性, 利用快速归一化融合调整 Coord-BiFPN 中的权重, 如公式(6)所示:

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i \quad (6)$$

其中: w_i 为可学习的权重, 并且通过在每个 w_i 后应用 SiLU 激活函数保证 $w_i \geq 0$, 并且设置 $\epsilon = 0.0001$ 是为了防止值不稳定, 归一化权重的每个值为 $0 \sim 1$ 。Coord-BiFPN 的每个模块的输出特征图可以表示如下:

$$P_1^{\text{out}} = C \left(\frac{w'_1 \cdot P_1^{\text{in}} + w'_2 \cdot R(P_2^{\text{id}})}{w'_1 + w'_2 + \epsilon} \right) \quad (7)$$

$$P_2^{\text{out}} = C \left(\frac{w'_1 \cdot P_1^{\text{in}} + w'_2 \cdot P_2^{\text{id}} + w'_3 \cdot R(P_1^{\text{out}})}{w'_1 + w'_2 + w'_3 + \epsilon} \right) \quad (8)$$

$$P_3^{\text{out}} = C \left(\frac{w'_1 \cdot P_3^{\text{in}} + w'_2 \cdot R(P_2^{\text{out}})}{w'_1 + w'_2 + \epsilon} \right) \quad (9)$$

其中, P_1 、 P_2 和 P_3 处的输出特征图分别由 P_1^{out} 、 P_2^{out} 和 P_3^{out} 表示。 P_1^{in} 、 P_2^{in} 和 P_3^{in} 分别表示 P_1 、 P_2 和 P_3 处的输入特征图, C 表示卷积运算, w'_1 、 w'_2 和 w'_3 表示每一层的权重, R 表示调整大小操作, 中间特征层 P_2^{id} 可以表示如下:

$$P_2^{\text{id}} = C \left(\frac{w_1 \cdot P_2^{\text{in}} + w_2 \cdot R(P_3^{\text{in}})}{w_1 + w_2 + \epsilon} \right) \quad (10)$$

其中, w_1 和 w_2 为对应的权重。Coord-BiFPN 在同一层的输入和输出特征之间增加了跳跃连接, 由于尺度相同, 因此增加跳跃连接可以更好地提取和传递特征信息。此外, 使用加权特征融合可以融合不同分辨率的输入层, 不同分辨率的输入层对应的权重也不同。将改进后的双向特征金字塔网络 Coord-BiFPN 模块引入特征融合部分, 在不增加过多计算量的情况下, 能够更好地实现多尺度特征融合。各输入层的权值参数通过网络自动学习, 能更好地表示整体特征信息。

3.4 OD-C2f 结构

动态卷积^[13] 基于卷积层的输入确定每个卷积核的权重, 并且使用注意力机制对这些卷积核进行加权和求和, 以获得适合该输入的动态卷积核, 动态卷积输出可以表示如下:

$$\text{Output}(y) = x \times (\alpha_1 W_1 + \alpha_2 W_2 + \dots + \alpha_n W_n) \quad (11)$$

其中: x 和 y 分别表示输入特征和输出特征变量, α_i ($i = 1, 2, \dots, n$) 为注意力标量, n 为卷积核数量, 并且每个卷积核 W_i ($i = 1, 2, \dots, n$) 具有与标准卷积核相同的大小。

多维动态卷积 ODConv 利用多维注意力机制和并行策略, 学习沿着内核空间的所有四个维度的任何卷积层处卷积内核的互补注意力。四个注意力为卷积核的输入通道数量、卷积核的感受野、卷积核输出通道数以及卷积核数量。这四个注意力相互补充, 并且按照位置、通道、滤波器和内核的顺序与卷积内核相乘, 显著增强了对上下文信息的捕获, ODConv 的输出可以用公式(12)表示:

$$\text{Output}(y) = x \times (\alpha_{wi} \odot \alpha_{fi} \odot \alpha_{ci} \odot \alpha_{si} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) \quad (12)$$

公式(12)中 α_{wi} 、 α_{fi} 、 α_{ci} 、 α_{si} ($i = 1, 2, \dots, n$) 表示四种注意力, 卷积核 W_i ($i = 1, 2, \dots, n$) 为变量, 表示卷积核沿着空间中不同维度进行乘法运算。

RDCA-YOLO 安全帽检测模型在 ODConv 和 C2f 的基础上提出了一种新的 OD-C2f 结构, 将 C2f 的 Bottleneck 的普通卷积优化为 ODConv 动态卷积, 根据不同的输入图像数据自适应调整卷积核, 并针对性地提取安全帽的特征, 显著增强了模型的特征提取能力, 加强了网络的学习能力, 提高了模型的识

别精度, OD-C2f 结构如图 6 所示。

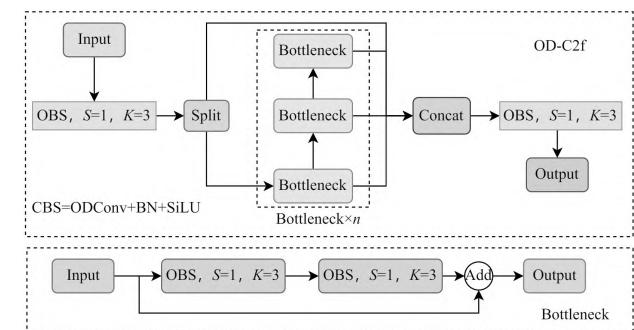


图 6 OD-C2f 结构图

Fig.6 Diagram of OD-C2f structure

3.5 FR-DyHead 检测头

在复杂的工地场景下, 工人安全帽佩戴检测面临位置变化、角度切换、尺度变化的挑战。YOLOv8n 的检测头不能很好地收集并融合多尺度信息、空间信息和任务信息。研究人员尝试改进已有目标检测器的检测头, 以增强模型性能, 导致不同的检测头对应不同对象检测器。然而, 还没有一种方法可以标准化检测头的结构。

良好的目标检测头应该具有三种能力。一是尺度感知能力。图像中往往存在多个不同尺度的物体, 一些物体检测器使用特征金字塔网络增强模型的尺度感知, 但并不能提高检测头本身的尺度感知。二是空间感知能力。在不同的视点下, 物体会发生变形和旋转, 其轮廓和位置也会发生变化。增强空间感知能力可以使检测器具有更强的泛化能力。三是任务感知能力。由于待检测对象可以用不同的形式(锚框、中心点、角点)表示, 这使得相应的检测头具有不同的结构和损失函数, 增强模型的任务感知能力可以使模型检测不同形式的对象。FR-DyHead 通过在每个维度使用三种不同的注意力机制进行统一, 分别为特征图、空间和通道三种感知, 从而使模型能够以更深的尺度关注信息。动态检测头 DyHead^[14] 是一个新颖的动态检测头结构, 它使用了多个注意力。本研究在动态头 DyHead 的基础上设计了 FR-DyHead 检测头, 运用 FReLU^[15] 激活函数替换原有的 ReLU 激活函数, 增强了模型对目标检测任务的适应性, 提高了检测的精确度。改进的 FR-DyHead 结构如图 7 所示。

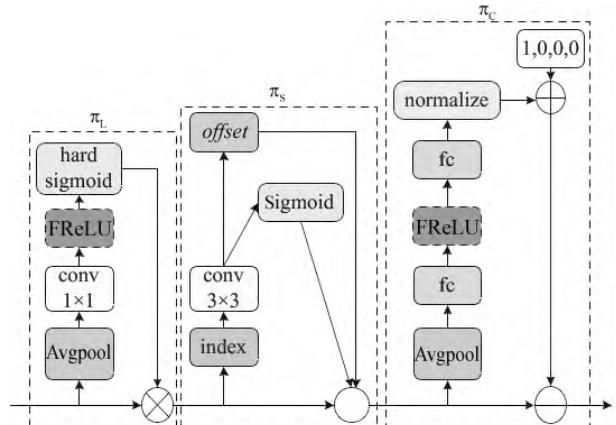


图 7 FR-DyHead 结构图

Fig.7 Diagram of FR-DyHead structure

图 7 中, π_l 是尺度感知注意力; π_s 是空间感知注意力; π_c 是任务感知注意力; $offset$ 表示偏移量; (1, 0, 0, 0) 为超参数, 用于控制激活阈值。

4 实验环境配置及数据准备 (Experimental environment setting and data preparation)

4.1 实验环境部署

模型基于 PyTorch 框架, 使用 GPU 进行训练, 实验所使用的 GPU 为 NVIDIA GTX 2070, CUDA 版本 11.1, PyTorch 版本 1.7.1, Python 版本 3.7, CPU 为 i5-9400F, 操作系统为 Windows 10 家庭中文版。

4.2 数据准备

实验所用安全帽数据集为 GDUT-HWD, 该数据集共有 3 174 张图像, 实验分别以 7 : 2 : 1 的比例将图片划分为训练集、验证集和测试集, 共包含 18 893 个对象, 包括建筑工地常见的大部分场景, 代表性和广泛性较强。所有对象实例根据安全帽的颜色分为蓝色、白色、黄色、红色和无类别, 不同颜色代表不同人员的工作属性, 无类别表示个人未戴安全帽, 共 5 个类别。设置实验迭代次数为 200 次, 训练批次为 8, 学习率为 0.01, 优化器为 SGD, $decay$ (学习率衰减因子) 为 0.000 5。

4.3 评价指标

评估网络性能的指标主要有准确率(Precision, P), 召回率(Recall, R), 平均准确率(Average Precision, AP), 平均准确率均值(mean Average Precision, mAP), 计算方法如公式(13)和公式(14)所示:

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

其中: TP 为预测正确的正样本数量, FP 为预测错误的正样本数, FN 为预测错误的负样本数。模型使用 $mAP@0.5$ 作为评估模型准确性的指标, 表示将 IoU(Intersection over Union) 阈值取 50% 时的 mAP 值, 通过平均每个类别的 AP 计算得到, 计算方法如公式(15)和公式(16)所示。

$$AP = \int_0^1 P(r) dr \quad (15)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (16)$$

5 实验结果与分析 (Experimental results and analysis)

5.1 消融实验和对比实验

针对 GDUT-HWD 数据集设计消融实验, 均采用“4.2”小节中的实验参数, 本文方法在 GDUT-HWD 数据集的消融实验结果比较如表 2 所示。其中: 模型 2 表示改进 Backbone 结构, 模型 3 表示嵌入 CBAM 注意力机制, 模型 4 表示设计一种

Coord-BiFPN 结构, 模型 5 表示在动态卷积的基础上设计了 OD-C2f 结构, 模型 6 表示提出了一种新的 FR-DyHead 检测头替换原始的 Detect 结构。从表 2 中的数据可以看出, 与原 YOLOv8n 算法相比, 本文模型的 mAP 提高了 2.6%。

表 2 本文方法在 GDUT-HWD 数据集的消融实验结果比较

Tab.2 Comparison of ablation experimental results of the proposed method on GDUT-HWD dataset

模型	AP/%				mAP@0.5/%	
	蓝色	白色	黄色	红色	无类别	0.5/%
YOLOv8n	81.2	83.7	87.5	82.0	81.7	83.2
YOLOv8n+ResNet-18	83.6	86.6	88.2	83.0	82.7	84.8
YOLOv8n+CBAM	82.1	84.6	87.3	82.5	81.5	83.6
YOLOv8n+Coord-BiFPN	83.1	84.5	88.0	82.5	81.0	83.8
YOLOv8n+OD-C2f	82.3	84.2	88.7	81.1	81.3	83.5
YOLOv8n+FR-DyHead	83.8	84.4	88.1	81.2	82.0	83.9
本文模型	85.2	87.0	89.2	84.2	83.5	85.8

为验证改进模型的有效性, 采用相同的验证集做了相关对比实验, 将其与目前主流的目标检测算法进行对比, 不同算法在 GDUT-HWD 数据集的检测结果比较如表 3 所示, 表明改进后的方法获得了最优的 mAP 值。

表 3 不同算法在 GDUT-HWD 数据集的检测结果比较

Tab.3 Comparison of detecting results of different algorithms on GDUT-HWD dataset

模型	P/%	R%	mAP@0.5/%
YOLOv3-tiny	88.3	68.8	75.5
YOLOv5s	88.8	78.1	83.6
YOLOv7	88.4	76.3	84.2
YOLOv7-tiny	91.5	75.4	83.9
YOLOv8n	88.2	76.6	83.2
YOLOv8s	89.7	77.9	85.2
本文模型	88.3	79.1	85.8

5.2 改进前后检测效果对比

为更加直观地评价改进前后的效果, 使用 GDUT-HWD 数据集中部分图片进行测试。对比图 8(a)、图 8(b)可以看出, YOLOv8n 在密集目标场景下存在 1 处漏检, 没有将图中远处的红色安全帽检测到。对比图 8(c)、图 8(d)可以看出, YOLOv8n 在小目标场景下存在误检, 将图中栏杆误检为红色安全帽。对比图 8(e)、图 8(f)可以看出, YOLOv8n 在反光场景下存在漏检, 没有将图中的白色安全帽检测到。对比图 8(g)、图 8(h)可以看出, YOLOv8n 在黑暗场景下存在漏检, 没有将图中没戴安全帽的人员情况检测到。可以看出, 改进后的模型检测准确率更高, 检测性能明显提升。

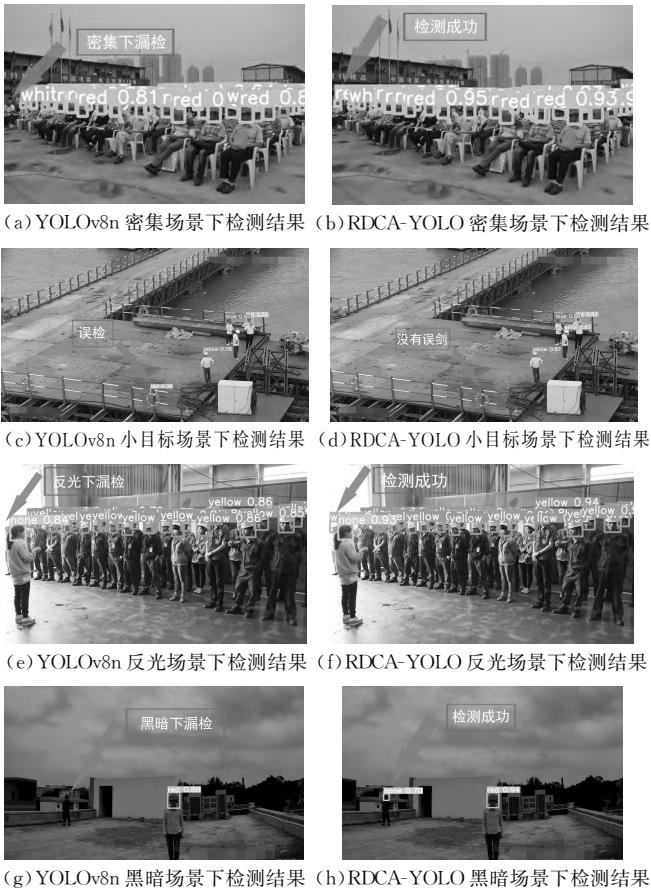


图 8 改进前后检测效果对比图

Fig.8 Comparison of effect before and after improvement

6 结论(Conclusion)

针对复杂工地场景下安全帽检测存在的模型识别精度低易出现漏检以及误检等问题。相比于原始的 YOLOv8n 模型, RDCA-YOLO 模型在 GDUT-HWD 数据集上的 mAP 提升了 2.6%、 R 提高了 2.5%, 证明本文所提方法在工人安全帽佩戴检测中有效,与其他目标检测模型相比具有显著优势,实现了工地工人安全帽佩戴的自动化检测。

参考文献(References)

- [1] TAGARELLI A, ZUMPANO E, ANASTASIU D C, et al. “Managing, mining and learning in the legal data domain”[J]. Information Systems, 2022, 106:101981.
- [2] 张云川, 姜麟, 林莉. 基于单次双向特征金字塔网络的目标检测模型[J]. 激光与光电子学进展, 2023, 60(2):273-283.
- [3] BHARATI P, PRAMANIK A. Deep learning techniques—R-CNN to mask R-CNN: a survey[C]//Das A, Nayak J, Naik B, et al. Computational Intelligence in Pattern Recognition. Singapore: Springer, 2020:657-668.
- [4] 王凤随, 王启胜, 陈金刚, 等. 基于注意力机制和 Soft-NMS 的改进 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2021, 58(24):405-416.

[5] 孙国栋, 李超, 张航. 融合自注意力机制的安全帽佩戴检测方法[J]. 计算机工程与应用, 2022, 58(20):300-304.

[6] 许凯, 邓超. 基于改进 YOLOv3 的安全帽佩戴识别算法[J]. 激光与光电子学进展, 2021, 58(6):300-307.

[7] 杜晓刚, 王玉琪, 娄润冰, 等. 基于 YOLO-ST 的安全帽佩戴精确检测算法[J]. 陕西科技大学学报, 2022, 40(6):177-183, 191.

[8] 赵敏, 杨国亮, 王吉祥, 等. 改进 YOLOv7-tiny 的安全帽实时检测算法[J]. 无线电工程, 2023, 53(8):1741-1749.

[9] ZHOU Q, ZHU W J, LI F C, et al. Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning[J]. Current Pharmaceutical Design, 2022, 28(4):287-295.

[10] WANG W, TAN X N, ZHANG P, et al. A CBAM based multiscale transformer fusion approach for remote sensing image change detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15:6817-6825.

[11] LI T G, ZHANG Y Z, LI Q Q, et al. AB-DLM: an improved deep learning model based on attention mechanism and BiFPN for driver distraction behavior detection [J]. IEEE Access, 2022, 10:83138-83151.

[12] ZHANG H T, TIAN M, SHAO G P, et al. Target detection of forward-looking sonar image based on improved YOLOv5 [J]. IEEE Access, 2022, 10:18023-18034.

[13] LIU C Y, HUANG D, WANG T, et al. Road traffic small target detection algorithm based on lightweight YOLO V5[J]. Frontiers in Computing and Intelligent Systems, 2023, 3(1):134-137.

[14] ZHANG H T, ZHANG P C, HU X W, et al. GLIPv2: unifying localization and vision-language understanding[DB/OL]. (2022-10-11)[2023-02-14]. <https://arxiv.org/abs/2206.05836>.

[15] ASANO H, HIRAKAWA E, HAYASHI H, et al. A method for improving semantic segmentation using thermographic images in infants[J]. BMC Medical Imaging, 2022, 22(1):1.

作者简介:

雷源毅(1998-),男,硕士生。研究领域:深度学习,目标检测。
朱文球(1969-),男,硕士,教授。研究领域:人工智能,计算机视觉,大数据技术。本文通信作者。

廖欢(1997-),男,硕士生。研究领域:图像处理,目标检测。