

# 基于 CPM-YOLO 的道路头盔检测方法<sup>\*</sup>

强浩南<sup>1</sup> 邹涌波<sup>1</sup> 马立东<sup>1,2</sup> 李博文<sup>1</sup>

(1. 太原科技大学机械工程学院 太原 030024; 2. 重型装备智能化技术与系统山西省重点实验室 太原 030024)

**摘要:** 头盔检测往往面临交通拥挤、行人干扰、目标严重遮挡等复杂的道路场景,这些情况容易导致检测精度低、检测错误和漏检。提出了一种基于 CPM-YOLO 算法的高性能头盔识别模型。首先,提出新的跨尺度特征融合方法 CS-FPN,更好地整合高阶语义和低阶几何特征信息;其次,提出 PCT 的模块,优化模型的特征提取能力;接着,选用最小点距离的边界框回归损失函数进一步提高模型的收敛速度和准确率;之后,删除骨干网络中  $20 \times 20$  的下采样部分和  $20 \times 20$  的检测头,新增  $160 \times 160$  的小目标检测头;最后,通过消融实验证明各个改进模块对提升模型性能的有效性,通过对比实验证明 CPM-YOLO 模型的优越性与泛化性。实验结果表明,与基线模型相比,本研究方法的 mAP@0.5 提高了 5.5%,参数数量和模型尺寸分别减少了 69.9% 和 67.2%。新模型具有显著降低复杂度和提高道路头盔检测能力的特点。

**关键词:** 双轮车头盔; YOLOv11s; CS-FAN; PCT; MPDIoU; 小目标检测头

**中图分类号:** TN911.73; TP391.41    **文献标识码:** A    **国家标准学科分类代码:** 520.6040

## Road helmets detection method based on CPM-YOLO

Qiang Haonan<sup>1</sup> Zou Yongbo<sup>1</sup> Ma Lidong<sup>1,2</sup> Li Bowen<sup>1</sup>

(1. School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China;

2. Shanxi Key Laboratory of Intelligent Technology and Systems for Heavy Equipment, Taiyuan 030024, China)

**Abstract:** Helmet detection often faces challenges in complex road scenarios such as heavy traffic, pedestrian interference, and severe occlusion of targets. These conditions can easily lead to low detection accuracy, false detections, and missed detections. This paper proposes a high-performance helmet recognition model based on the CPM-YOLO algorithm. First, a novel cross-scale feature fusion method, CS-FPN, is proposed to better integrate high-level semantic and low-level geometric feature information. Next, the PCT module is introduced to optimize feature extraction capabilities of the model. Additionally, a bounding box regression loss function based on the minimum point distance is adopted to enhance the model's convergence speed and accuracy. Furthermore, the  $20 \times 20$  downsampling layer and  $20 \times 20$  detection head in the backbone network are removed, and a new  $160 \times 160$  small-object detection head is introduced. Finally, ablation studies validate the effectiveness of each improved module in enhancing the model's performance, and comparative experiments demonstrate the superiority and generalizability of the CPM-YOLO model. Experimental results show that compared to the baseline model, the proposed method achieves improvements of 5.5% in mAP@0.5. Additionally, the number of parameters and model size are reduced by 69.9% and 67.2%, respectively. The new model significantly reduces complexity while enhancing helmet detection capabilities in road environments.

**Keywords:** two-wheeled vehicle helmet; YOLOv11s; CS-FAN; PCT; MPDIoU; small target detection head

## 0 引言

近年来,电动车和摩托车因其体积小巧、速度轻快等

特点,成为广大市民短途出行的优质选择,但由于骑行人员的一些不安全驾驶习惯,交通事故时常发生,所以如何准确、高效地检测骑行人员是否佩戴头盔,对减少事故伤亡率

收稿日期:2024-11-18

\* 基金项目:国家自然基金(U24A20276)、国家自然基金面上项目(52274389)、山西省关键核心技术和共性技术研发攻关专项(2020XXX009)、山西省重点研发计划(2021020101010, 202202150401010)项目资助

和降低事故伤亡风险有着重大的意义。

目前,头盔检测方法大致可分为两种类型:传统的头盔检测方法和基于深度学习的头盔检测方法。

传统的头盔检测方法主要由在图片中提取候选区域、对候选区域进行特征提取和对识别的区域进行分类并对候选框的坐标进行修正 3 个步骤组成。Waranusast 等<sup>[1]</sup>利用 K-最近邻分类器提取运动物体的区域属性特征,然后根据投影轮廓对识别的摩托车上的骑手头部进行计数和分割,实现对头盔的识别。Silva 等<sup>[2]</sup>提出了一种基于局部二值模式、方向梯度直方图和霍夫变换描述符的混合特征提取描述符对摩托车乘骑人员进行特征提取。然而在实际应用中,传统的头盔检测方法在复杂背景下准确率不高,且因为需要手工设计特征和规则,导致特征鲁棒性不足、检测速度慢、效率低,因此不适用于对实时性和准确性要求高的骑行人员安全头盔检测任务。

近年来,基于深度学习的目标检测技术在道路安全检测领域的应用越来越广泛,它可以自动从输入数据中提取特征,使系统具有了更强的适应性和泛化能力。主流的基于深度学习的目标检测算法有两阶段算法,如基于候选区域提议的卷积神经网络(region-based convolutional neural network, R-CNN)<sup>[3]</sup>、Fast R-CNN<sup>[4]</sup>、Faster R-CNN<sup>[5]</sup>—阶段算法,如 SSD(single shot multibox detector)<sup>[6]</sup>、YOLO (you only look once)<sup>[7]</sup> 算法;基于 transformer 架构的算法,如 DETR(detected transformer)<sup>[8]</sup>。

两阶段算法首先生成候选框,然后对提取区域进行分类,以确定识别对象是否包含在边界框内,并对边界框的位置进行微调。它具有较高的检测准确率,但计算成本大,存在推理速度慢和不易部署的问题。

基于 Transformer 架构的算法由一个编码器和一个解码器组成。编码器处理输入图像,而解码器生成边界框和类别预测。它利用自带的注意力机制捕捉图像中不同区域之间的关系,具有较高的检测精度,然而它通常需要大量的数据和计算资源进行训练,并且在小目标和复杂场景下表现不佳。

单阶段检测算法直接在输入图像上预测边界框和类别,不生成区域建议,虽然牺牲了一定的准确率,但具有速度快、实时性好的优势。由于头盔佩戴检测所需的实时性要求较高,所以利用一阶段算法检测骑行人员是否佩戴头盔越来越流行。

窦广鉴<sup>[9]</sup>使用轻量级网络 MobileNetV2 替换 SSD 检测模型的 VGG16 主干,将 MobileNetV2 倒置残差中的  $3 \times 3$  卷积替换为多头注意力机制,在满足实时性的同时更好地提取了全局特征信息,减少参数量的同时提高了模型对于安全头盔的检测精度,但模型的泛化能力和在复杂场景中的检测准确率有待进一步提高。

张碧川等<sup>[10]</sup>对 YOLOv8 进行了剪枝轻量化处理,并将

改进的 ConvNeXt 和 CBAM 模块分别嵌入到网络的卷积块中,在增强了网络的表达能力的同时保持了网络的简洁和高效,但在头盔检测精度上依旧存在提升空间。杨璐钱等<sup>[11]</sup>提出了 Yolov8-scG 算法,引入了 SPDCConv 模块和 C2f-CG block 模块,替换 SPPF 模块中的普通卷积为组卷积,提升了模型对道路现场采集的较低分辨率图像的辨识精度,优化了模型对于局部背景的筛选能力,降低了整体模型的计算量,但模型泛化性欠佳,对于道路强光环境等特定场景会出现一定的检测误差。周顺勇等<sup>[12]</sup>提出了 Helmet-YOLO 算法,在渐进式特征金字塔网络结构的基础上融合了 PCA-Head 检测头和 Helmet-IoU 边界框损失函数,提高了模型损失计算的效率和精度,但在远景较小目标的检测方面依然存在错检、漏检问题。袁宇乐等<sup>[13]</sup>通过优化主干网络结构、整合高效的上采样技术和增设专门的 tiny 模块以及改进损失函数强化了 YOLOv8s 的性能,对头盔检测的精度和速度取得了明显提升,但是在极端环境条件下对于小目标的检测效果仍有提升空间。

上述改进算法虽然具有一定的贡献,但检测仍然存在局限性,存在一定的漏检、误检情况。结合实时检测和现实部署的需求,综合考虑模型的检测精度、模型参数量和计算效率,本研究选用 YOLOv11s 模型作为基准模型,在保证检测精度的同时,又兼顾了识别速度,可以满足双轮车头盔实时检测的要求。YOLOv11 是一种基于深度学习和计算机视觉的最先进的实时目标检测模型之一,是 YOLO 系列模型的最新版本,继承了前几代模型的优点,同时引入了新的特性和改进,提高了性能、灵活性和效率。

针对以上问题,本文基于 YOLOv11s 模型进行改进,提出了一种实时高效且能改善上述复杂道路场景下头盔漏检、误检问题的两轮车头盔佩戴检测算法。首先,提出新的跨尺度特征金字塔网络(cross scale feature pyramid networks, CS-FPN),更好地整合高阶语义和低阶几何特征信息;同时,提出一种 CNN 和 Transformer 的融合模块(partially CNN and Transformer block, PCT),保证计算效率的情况下优化特征提取能力;其次,选用最小点距离的边界框回归损失函数(minimum point distance intersection over Union, MPDIoU)<sup>[14]</sup>,进一步提高模型的收敛速度和准确率;最后,删除骨干网络中的下采样部分和  $20 \times 20$  的检测头,在保证小目标检测头不受影响的同时最大限度减少模型的体积。

## 1 YOLOv11 模型

YOLOv11 算法是由 YOLOv8 改进而来,不仅轻量、快速,还可以在多种硬件平台上部署。YOLOv11 网络主要由输入端(Input)、骨干网络(Backbone)、颈部网络(Neck)和检测头网络(Head)4 大部分组成。算法结构如图 1 所示。

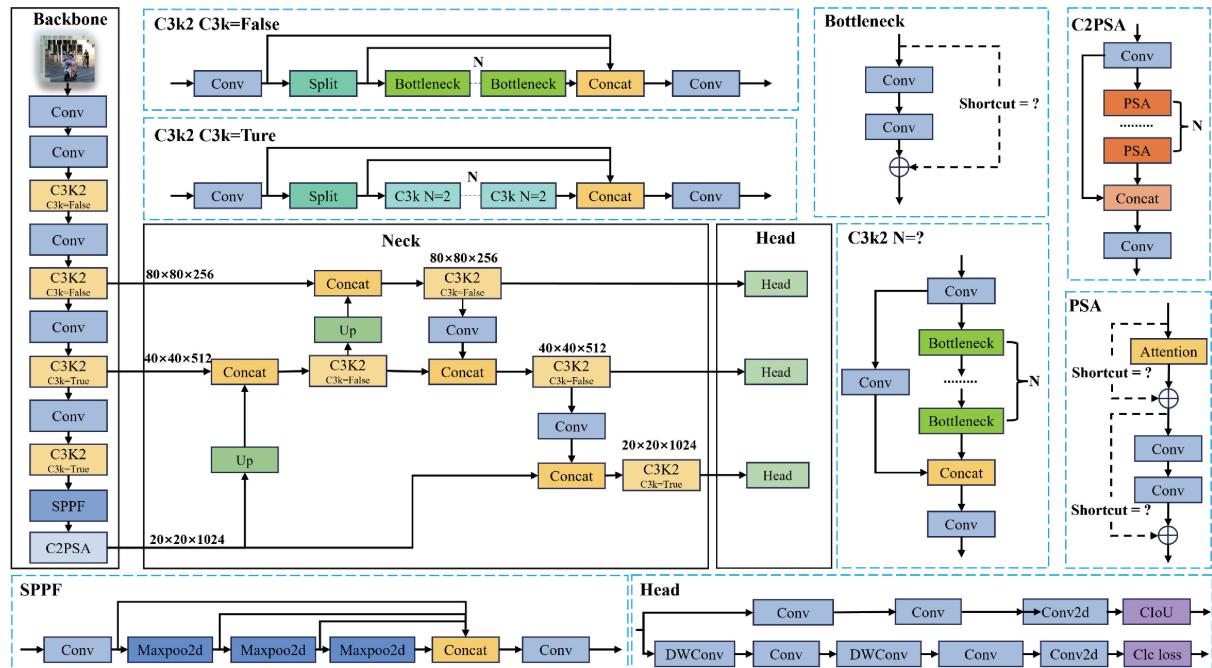


图 1 YOLOv11 网络结构

Fig. 1 Network structure of YOLOv11

Input 层的作用是将图像数据输入到模型中, 输入图像经过预处理会被调整到尺寸大小为  $640 \times 640$ 。

Backbone 层的作用是对输入的图像进行特征提取, 基于 Darknet-53 改进而成, 其中的 3 次卷积模块替换为两次卷积的 CSPLayer<sup>[7]</sup>模块, 相对于 3 次卷积模块添加了更多的跳层连接, 可在保证轻量化的同时获得更加丰富的梯度流信息。

Neck 层的作用是对骨干网络提取的特征进行融合和处理, 延续了 FPN+PAN 结构, 实现了多尺度特征融合, 有效地避免了对图像区域进行裁剪、缩放操作而导致的图像失真, 大大提高了产生候选框的速度, 且节省了计算成本。

Head 层的作用是对特征进行预测, 得到目标的位置信息和类别信息。检测头网络采用解耦头 (Decoupled-Head) 结构将分类和检测头分离, 从基于锚框方法 (Anchor-Based) 改为无锚框 (Anchor-Free) 方法, 通过动态分配正负样本<sup>[15]</sup>, 改善了正负样本不均的问题。

YOLOv11 属于单目标检测网络, 根据模型尺寸的不同, 分为 n、s、m、l、x 5 个版本, 这 5 个模型的深度和宽度依次增加, 检测精度逐渐提升, 但训练花费时间也依次提高。本研究结合道路实际情况, 选用 YOLOv11s 模型作为基准模型, 在保证检测精度的同时, 又兼顾了识别速度, 可以满足双轮车头盔实时检测的要求。

## 2 CPM-YOLO 检测算法

本文以 YOLOv11s 算法为基准, 对其骨干网络、颈部

网络、检测头网络和损失函数都进行了针对性改进, 最终得到改进的 CPM-YOLO 模型, 其整体结构如图 2 所示。

### 2.1 CS-FAN

YOLOv11 的基本模型结合了特征金字塔网络 (feature pyramid networks, FPN) 和路径聚合网络 (path aggregation network, PANet)<sup>[16]</sup> 的特点, 使用 FPN+PAN<sup>[17-18]</sup> 的结构对不同尺度特征进行融合。PAN 网络中的输入特征经 FPN 网络处理后会发生原始特征丢失, 导致特征融合效率低, 同时, 多次下采样会使图像中相邻像素的信息逐渐融合, 造成特征图的空间分辨率下降, 导致图像的物体位置信息弱化和边缘信息模糊, 使网络对于细节特征的捕捉能力下降。

为了解决这个问题, 本文提出了新的跨尺度特征融合方法 CS-FPN, 在原先融合策略的基础上采用跨层融合策略, 将 PAN 自底向上路径中的特征图与骨干网络中相对应层次的特征图进行融合, 其结构如图 3 所示。

通过将 FPN 自底向上路径上的特征图与 PAN 自底向上路径上对应层次的特征图进行融合, 可以在特征图中平衡语义与空间细节, 从而在保持高语义表达的同时, 不丢失关键的空间位置信息, 确保模型在多尺度下的特征提取更加完整, 形成更全面的特征表示。

### 2.2 PCT

多项研究表明, 骨干网络在提取图像特征时使用的卷积神经网络 (CNN) 的感受野较小, 导致其只能专注于提取输入图像的局部特征, 但 Transformer 结构却因其强大的全局特征提取能力一直受到广泛关注。本文提出了一种

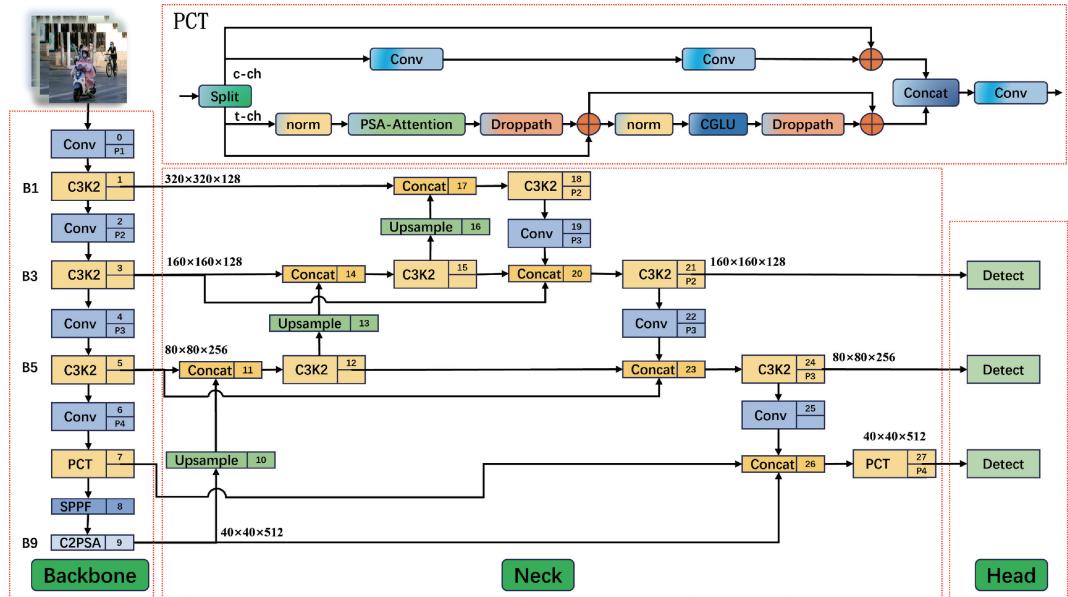


图 2 CPM-YOLO 网络结构

Fig. 2 Network structure of CPM-YOLO

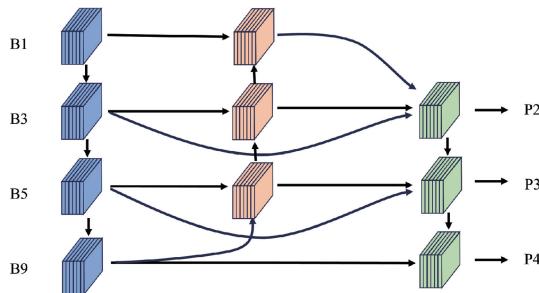


图 3 CS-FAN

Fig. 3 CS-FAN

名为 PCT 的混合结构, 将输入特征图分为两部分, 分别由 CNN 和 Transformer 处理, 旨在结合 CNN 和 Transformer 的优势来增强特征提取能力。其整体结构如图 2 的 PCT 模块所示。

由图 2 可知, Transformer 这一特征提取分支上有位置敏感注意力机制 (positional sensitive attention, PSA) 和卷积门控线性单元 (convolutional gated linear unit, CGLU) 两个关键模块, 其中 PSA 负责提取全局特征, CGLU 用于增强非线性特征表达能力。

PSA 是一种改进的注意力机制, 通过在计算注意力权重时考虑每个查询的位置信息, 使得模型能更好地捕捉到序列中元素之间的相对位置关系, 增强了模型对位置信息的敏感性。其整体结构如图 4 所示。

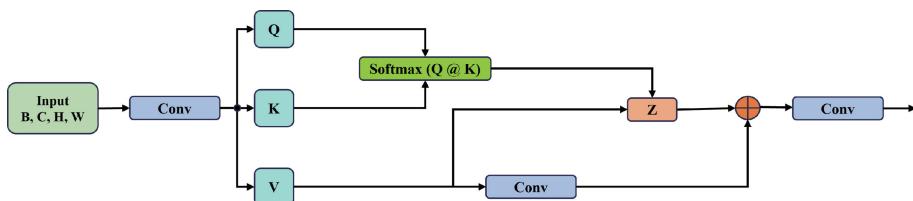


图 4 PSA-Attention 模块

Fig. 4 PSA-Attention module

PSA 不仅考虑了输入元素之间的相关性, 还显式考虑了它们的相对位置信息, 这种机制可以通过相对位置编码和相对位置嵌入实现。在注意力机制中, 输入通常是  $X = [x_1, x_2, \dots, x_n]$  的一个序列, 对每个输入序列的位置  $i$  处的元素  $x_i$  进行线性变换, 生成查询 (Query)、键 (Key) 和值 (Value) 矩阵。

$$Q_i = x_i W_Q, K_i = x_i W_K, V_i = x_i W_V \quad (1)$$

式中:  $W_Q, W_K, W_V$  分别是可学习的查询、键和值的权重矩阵,  $x_i$  是输入的特征。

PSA 通常使用缩放点积进行常规的注意力权重计算, 然后归一化到  $[0, 1]$  范围内的概率分布:

$$A_{ij} = \text{softmax} \left( \frac{Q_i K_j^T}{\sqrt{d_k}} \right) \quad (2)$$

式中:  $A_{ij}$  表示第  $i$  个位置与第  $j$  个位置之间的注意力权

重,  $d_k$  是 Key 的维度, 用于缩放稳定训练。

使用注意力权重矩阵  $\mathbf{A}$  对 Value 特征进行加权求和, 得到增强后的特征表示:

$$Z_i = \sum_{j=1}^{N=n} \mathbf{A}_{ij} \mathbf{V}_j \quad (3)$$

式中:  $Z_i$  是经过增强后的特征。

将增强后的特征  $Z$  与原始输入特征  $X$  进行融合, 得到最终的输出特征  $Y$ :

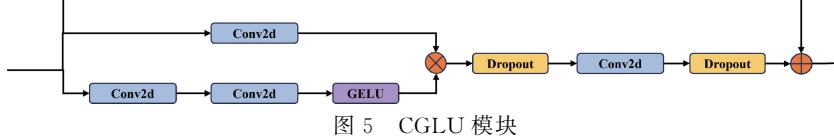


图 5 CGLU 模块

Fig. 5 CGLU module

假设输入为  $X$ , 其形状为  $(B, C, H, W)$ , 其中  $B$  是批大小,  $C$  是通道数,  $H$  是高度,  $W$  是宽度。输入  $X$  通过两个卷积层, 分别生成两个特征图  $A$  和  $B$ 。

$$A = \text{Conv1}(X), B = \text{Conv2}(X) \quad (5)$$

式中:  $\text{Conv1}$  和  $\text{Conv2}$  是具有相同卷积核大小的卷积操作。

对第 2 个特征图  $B$  应用  $\text{GELU}$  激活函数生成门控值  $G$ ,  $G$  的输出在  $0 \sim 1$  之间。

$$G = \sigma(B) \quad (6)$$

式中:  $\sigma$  是  $\text{GELU}$  函数。

将第一个特征图  $A$  和门控值  $G$  进行逐元素相乘, 抑制不重要的特征, 提高重要特征的突出性, 生成输出:

$$Y = A \odot G \quad (7)$$

式中:  $\odot$  表示逐元素乘法。

CGLU 中的卷积操作能够有效地捕捉输入中的局部特征, 门控机制可以通过控制信息流动引入非线性特性, 动态地选择需要通过的特征, 从而增强模型的表达能力。通过组合卷积和门控机制, CGLU 在需要捕捉复杂特征模式的任务中表现尤为出色。

### 2.3 MPDIoU 损失函数

基准模型依赖基于完整交并比的边界框回归损失函数(complete IoU, CIoU)<sup>[19]</sup>确定目标的位置, 但当预测框和目标框的长宽比相同时, CIoU 会失去其效用。为了提升模型精度, 本文引入了 MPDIoU 优化基准模型的损失函数, 计算示意图如图 6 所示。

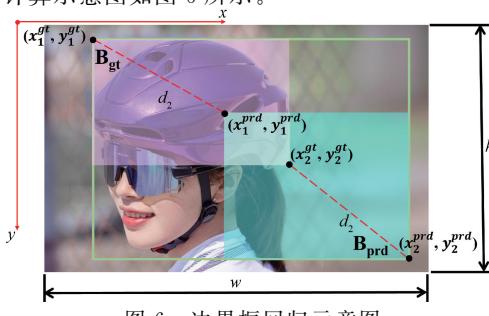


图 6 边界框回归示意图

Fig. 6 Schematic diagram of bounding box regression

$$Y = X + Z \quad (4)$$

这种融合操作有助于保留输入的原始信息, 同时增强不同空间位置之间的信息交互。

CGLU 是一种结合了卷积操作和门控机制的神经网络模块, 通过捕捉局部特征并动态控制信息流动, 增强了特征提取的有效性和灵活性, 加强了模型的表达能力和稳定性。其整体结构如图 5 所示。

一般来说, 一个矩形可以由左上角点和右下角点的坐标唯一定义, MPDIoU 受边界框几何特性的启发, 直接最小化预测边界框与真实边界框之间的左上和右下点距离, MPDIoU 的计算为:

$$\text{MPDIoU} = \frac{B_{gt} \cap B_{prd}}{B_{gt} \cup B_{prd}} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (8)$$

$$B_{gt} = (x_2^{gt} - x_1^{gt}) \times (y_2^{gt} - y_1^{gt}) \quad (9)$$

$$B_{prd} = (x_2^{prd} - x_1^{prd}) \times (y_2^{prd} - y_1^{prd}) \quad (10)$$

$$x_1^l = \max(x_1^{pre}, x_1^{gt}), \quad x_2^l = \min(x_2^{pre}, x_2^{gt}) \quad (11)$$

$$y_1^l = \max(y_1^{pre}, y_1^{gt}), \quad y_2^l = \min(y_2^{pre}, y_2^{gt}) \quad (12)$$

$$B_{gt} \cap B_{prd} = \begin{cases} (x_2^l - x_1^l) \times (y_2^l - y_1^l), & x_2^l > x_1^l, y_2^l > y_1^l \\ 0, & \text{其他} \end{cases} \quad (13)$$

$$B_{gt} \cup B_{prd} = B_{gt} + B_{prd} - B_{gt} \cap B_{prd} \quad (14)$$

$$d_1^2 = (x_1^{prd} - x_1^{gt})^2 + (y_1^{prd} - y_1^{gt})^2 \quad (15)$$

$$d_2^2 = (x_2^{prd} - x_2^{gt})^2 + (y_2^{prd} - y_2^{gt})^2 \quad (16)$$

式中:  $B_{gt}$  表示真实框面积,  $B_{prd}$  表示预测框面积,  $(x_1^{gt}, y_1^{gt})$  和  $(x_1^{prd}, y_1^{prd})$  分别表示真实框和预测框的左上角点坐标,  $(x_2^{gt}, y_2^{gt})$  和  $(x_2^{prd}, y_2^{prd})$  分别表示真实框和预测框的右下角点坐标,  $d_1$  和  $d_2$  分别为对应顶点的距离,  $w$  和  $h$  表示输入图像的宽和高。

模型训练时, 通过最小化损失函数  $\mathcal{L}$ , 迫使各个预测边界框  $\mathfrak{B}_{prd} = [x_{prd}, y_{prd}, w_{prd}, h_{prd}]^T$  趋近于真实边界框  $\mathfrak{B}_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]^T$ , 设  $n$  一范数  $\mathcal{L}$  表达式如式(17)所示。

$$\mathcal{L} = \min_{\Theta} \sum_{\mathfrak{B}_{gt} \in \mathfrak{B}_{gt}} \mathcal{L}(\mathfrak{B}_{gt}, \mathfrak{B}_{prd} | \Theta) \quad (17)$$

式中:  $\mathfrak{B}_{gt}$  表示真实边界框的集合,  $\Theta$  为深度回归模型的参数。

基于 MPDIoU 的损失函数定义为:

$$\mathcal{L}_{MPDIoU} = 1 - \text{MPDIoU} \quad (18)$$

从式(18)中可以看出, 所有影响损失函数的因素都可以通过左上角点和右下角点的坐标来确定,  $\mathcal{L}_{MPDIoU}$  能在预

测框和真实框长宽比相同且真实长度不同的情况下有效反映损失数值的变化。

#### 2.4 小目标检测

考虑到在实际道路场景中存在摄像头分辨率较低、相机拍摄角度造成的遮挡以及远景模糊等因素, YOLOv11 原始模型输出的  $80 \times 80$  特征图中可能出现检测不到的目标, 导致小目标检测的准确度不够, 因此, 本文将特征图尺寸为  $20 \times 20$  的检测头替换为  $160 \times 160$  的小目标检测头来适应较小的目标尺寸, 提高网络对小目标的关注度, 减少漏检情况的出现。

改进后增加的大尺度特征图能够捕捉到图像中更大范围内的上下文信息, 图中的每个元素对应原始图像中的较大区域, 可以将目标划分更加细致, 对目标细节关注度更高, 利用浅层位置信息和深层语义信息相结合的方法, 更加准确地定位和识别小目标。

### 3 实验结果与分析

#### 3.1 训练环境

本文建立了一个实验平台来训练模型并进行测试, 实验平台、硬件和软件配置如表 1 所示。在实验中, 所有模型的训练、测试和验证过程使用的超参数保持一致。在模型训练中, 使用 SGD 优化器训练, 动量 (momentum) 为 0.937, 权重衰减率 (weight decay) 为 0.0005, 初始学习率 (Lr0) 设置为 0.01。所有训练都是用 150 个 epoch 进行的, 批次大小设置为 32。

表 1 实验平台及软硬件配置

Table 1 Experimental platform and software and hardware configuration

配置环境	版本型号
操作系统	Ubuntu 20.04
框架	PyTorch 1.13.1
计算架构	Cuda 11.6
语言	Python 3.9
CPU	AMD EPYC 7J13 64-Core Processor
GPU	NVIDIA GeForce RTX 4090(24 GB)

#### 3.2 数据集

本文选取了双轮车(摩托车和电动车)佩戴头盔检测数据集 (two wheeler helmet dataset, TWHD) 进行实验, 按照双轮车与驾乘人员整体 (two\_wheeler)、未戴头盔的人头 (without\_helmet)、戴头盔的人头 (helmet) 3 个类别进行定位与分类标注。将总计 5 448 张图像样本与标注 txt 文件按照 8:1:1 划分, 选取其中 4 358 张作为训练集, 544 张作为验证集, 546 张作为测试集进行实验, 图 7 展示的是 TWHD 数据集中的样本图像。

#### 3.3 评价标准

本实验选用查准率 (precision, P)、查全率 (recall, R)、



图 7 双轮车佩戴头盔检测数据集样本

Fig. 7 The sample of TWHD dataset

平均精度均值 (mean average precision, mAP)、参数量 (parameters) 和模型尺寸 (size) 对模型性能进行综合评价, 其中 mAP@0.5 表示 IoU 阈值为 0.5 时各类别的平均精度如图 8 所示, Precision 指全部检测出为正样本中检测正确的概率, Recall 指全部正样本中被正确预测出来的概率, 计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (19)$$

$$R = \frac{TP}{TP + FN} \quad (20)$$

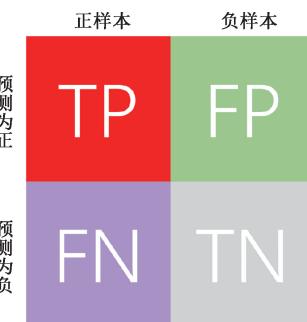


图 8 指标示意图

Fig. 8 Indicator diagram

其中, 正样本指包含目标物体的图像区域, 负样本指不包含目标物体的图像区域, 预测为正指模型识别分类出物体, 预测为负指模型未识别分类出物体, TP 指被预测出物体的正样本数量, FP 指被预测出物体的负样本数量, FN 指未被预测出物体的正样本数量。

一般情况下, P 与 R 呈负相关, P 与 R 不能单独评估模型的整体性能, 但基于 P 和 R 的曲线与坐标轴围成的面积可以表示平均精度 (average precision, AP), 计算公式为:

$$AP = \int_0^1 P(R) dR \quad (21)$$

AP 是对所有分类目标的 AP 取均值, 计算公式为:

$$mAP = \frac{\sum_i^N AP_i}{N} \quad (22)$$

式中:  $N$  代表检测种类的数量。

### 3.4 消融实验

本文将 PCT 模块加入到 YOLOv11s 模型的网络中来确定 PCT 模块最合理的部署位置,实验结果如表 2 所示。表中“BackboneX”和“NeckX”分别表示骨干网络和颈部网络中 PX 层的 C3K2 模块被替换为 PCT 模块,本文设计了不同的变换组合。

由表 2 可以看出,当 PCT 模块应用于 P4 层时,模型的检测精度来到最高,mAP@0.5 提升了 1.1%,并且参数量和模型体积也有所下降。Transformer 机制在中等分辨率的  $40 \times 40$  特征图上可以充分捕捉长距离依赖关系,增强目标与背景之间的区分能力,改善目标检测的精度。 $20 \times 20$  特征图分辨率过低,Transformer 机制会引入不必要的计算复杂度,反而对性能造成负面影响,所以 PCT 模块应用于 P5 层时,检测精度反而会下降。为保证最高的检测精度,选择将 PCT 模块应用于整个网络的 P4 层。

为了验证小目标检测头的有效性,本文将 YOLOv11s 模型  $20 \times 20$  的检测头替换为  $160 \times 160$  的小目标检测头,并尝试删除 P5 下采样层,实验结果如表 3 所示,表中“Small head”和“Delete”分别表示上述两种操作。由表 3 可以看出,在更换小目标检测头后,mAP@0.5 提升了 2.8%,并且参数量和体积有一定减少。删除 P5 下采样层在几乎不影响检测精度的情况下使模型的参数量和体积减小了 70.2% 和 68.2%,证明了这一操作的可取。

为了验证提出的改进策略在算法中的有效性,本文以

YOLOv11s 为基准模型,将所有改进加入到网络结构中进行了一系列消融实验,结果如表 4 所示。表中“CS-FAN”、“PCT”、“MPDIOU”和“Delete + Small head”分别代表本研究提出的 4 个改进点,符号“√”表示采用该改进点。

如表 4 所示,第 1 组实验直接使用 YOLOv11s 模型,其 mAP@0.5、参数量和体积分别为 0.851、9.4 M 和 19.2 MB。第 2 组采用 CS-FAN 结构,更好地整合了高阶语义和低阶几何特征信息,mAP@0.5 增加了 2%;第 3 组用 PCT 模块代替了 P4 层的 C3K2 模块,优化了特征提取能力,mAP@0.5 增加了 1.1%;第 4 组更换模型的边界框回归损失函数为  $mAP$ ,提高了模型收敛速度和准确率,mAP@0.5 增加了 0.9%。

第 5 组和第 6 组分别在第 2 组的基础上融合了第 3 组和第 4 组的优势,检测精度都进一步提高,mAP@0.5 分别在第 2 组基础上上升了 1% 和 0.7%;第 7 组将 3 个改进点最终结合,与基准模型相比,mAP@0.5 提高了 3.7%,有效证明了各改进方法的有效性以及它们在算法中的相容性。

第 8 组实验在第 7 组实验的基础上进一步调整了网络结构,删除 P5 下采样层并更换小目标检测头,模型的 mAP@0.5 最终为 0.906,相较基准模型增加了 5.5%,参数量和体积仅为原来模型的 30.1% 和 32.8%。

上述消融实验验证了本文提出的 4 个改进点的有效性,并展示了每个改进点对道路头盔检测和模型轻量化的贡献。

表 2 PCT 模块的消融结果

Table 2 Ablation results of the PCT module

组别	模型	AP@0.5		Without-helmet	mAP@0.5	Parameters/M	Size/MB
		Two-wheeler	helmet				
1	YOLOv11s	0.969	0.851	0.732	0.851	9.4	19.2
2	Backbone4	0.965	0.858	0.747	0.857	9.3	19.0
3	Backbone5	0.966	0.842	0.725	0.844	9.0	18.3
4	Backbone45	0.967	0.847	0.735	0.850	8.9	18.2
5	Neck4	0.965	0.858	0.744	0.856	9.2	18.8
6	Backbone4+Neck4	0.966	0.863	0.756	0.862	9.1	18.6

表 3 删除下采样层和更换检测头的消融结果

Table 3 Ablation results of removing the subsampling layer and replacing the detection head

组别	模型	AP@0.5		Without-helmet	mAP@0.5	Parameters/M	Size/MB
		Two-wheeler	helmet				
1	YOLOv11s	0.969	0.851	0.732	0.851	9.4	19.2
2	Small head	0.967	0.891	0.779	0.879	7.1	14.7
3	Delete+Small head	0.966	0.890	0.774	0.877	2.8	6.1

表 4 消融实验结果  
Table 4 Results of ablation experiment

组别	模型			AP@0.5		mAP@0.5	Parameters/M	Size/MB
	CS-FAN	PCT	MPDIoU	two_wheeler	helmet			
1				0.969	0.851	0.732	0.851	9.4
2	✓			0.965	0.865	0.782	0.871	9.8
3		✓		0.966	0.863	0.756	0.862	9.2
4			✓	0.965	0.862	0.752	0.860	9.4
5	✓	✓		0.967	0.878	0.799	0.881	9.7
6	✓		✓	0.965	0.865	0.782	0.878	9.8
7	✓	✓	✓	0.966	0.892	0.805	0.888	9.7
8	Delete+Small head			0.968	0.91	0.841	0.906	2.9
								6.3

### 3.5 可视化结果

为了更直观地比较本文算法与 YOLOv11s 算法的检测效果, 使用相同测试集对两种算法进行测试, 并选取了其中具有代表性的图片进行可视化分析, 其对比可视化结果如图 9 所示。其中, 图 9(a)~(c) 为原始图像, 图 9(d)~(f) 为 YOLOv11s 检测图像, 图 9(g)~(i) 为 CPM-YOLO 检测图像, 圆点线框表示 two\_wheeler, 方点线框表示 without\_helmet, 曲形线框表示 helmet, 实线框表示 CPM-YOLO 优于 YOLOv11s 的检测部分。

图 9 展示了基准模型 YOLOv11s 和 CPM-YOLO 模型检测的可视化结果, 通过图 9(d) 和 (g) 可知, 针对因为对焦问题导致的边缘成像模糊, CPM-YOLO 在模糊图像的检测上要优于 YOLOv11s; 通过图 9(e) 和 (h) 可知, CPM-YOLO 在未出现人脸的头盔背后检测上效果远远优于 YOLOv11s; 通过图 9(f) 和 (i) 可知, YOLOv11s 出现了错检, 将黄色头盔错误检测成了未佩戴头盔, 而 CPM-YOLO 没有出现误检问题, 证明在存在遮挡的情况下, CPM-YOLO 的检测性能优于 YOLOv11s。由此可见, CPM-YOLO 较好地解决了对骑行人员头盔检测的漏检、误检等问题, 在检测小目标、远距离和被遮挡物体等方面均具有良好的效果。

### 3.6 对比实验

通过对比实验验证算法的性能优势, 选取已公开代码的检测算法 (Faster-RCNN<sup>[5]</sup>, SSD<sup>[6]</sup>, RT-DETR<sup>[8]</sup>, RTDETR-RESNET101<sup>[20]</sup>, YOLOv3<sup>[21]</sup>, YOLOv5s<sup>[22]</sup>, YOLOv6s<sup>[23]</sup>, YOLOv8s<sup>[24]</sup>, YOLOv8s-WORLDv2<sup>[25]</sup>, YOLOv9s<sup>[26]</sup>, YOLOv10s<sup>[27]</sup>, YOLOv11 1, YOLOv11x) 和个别自己复现的最新改进 YOLO 模型 (MobileNetV2-MHSA-SSD<sup>[9]</sup>, Yolov8n-sCG<sup>[11]</sup>, Helmet-YOLO<sup>[12]</sup>, 改进 YOLOv8s<sup>[13]</sup>) 作对比, 实验结果如表 5 所示。

通过对比表 5 中数据可得, CPM-YOLO 比起其他 18



图 9 CPM-YOLO 和 YOLOv11s 检测效果对比

Fig. 9 Comparison of detection effects between CPM-YOLO and YOLOv11s

个模型, Parameters 最少, Size 最小, 各个类别的 mAP 和总 mAP 均为最高, 在各个评价指标上都具有明显优势。与其他典型算法相比, 本文提出的 CPM-YOLO 模型不仅具备轻量化模型的参数量, 还具有较高的 MPDIoU, 在平衡检测精度和复杂度方面取得了很好的效果。

### 3.7 泛化性对比实验

为验证算法泛化能力, 将基准模型 YOLOv11s 与改进后的 CPM-YOLO 模型分别用于另外 3 个头盔检测数据集 (Hard\_hat\_workers、Helmet 和 Helmet-detection) 的训练。CPM-YOLO 模型的改进效果主要基于 TWHD 数据集, 并在该数据集上表现出明显的性能提升。

由表 6 可以看出, CPM-YOLO 模型在不同数据集上均展示出良好的泛化能力, 同时显著减少了模型参数量和存储大小, 在精度和效率上相较于 YOLOv11s 具有明显的优势, 适合在资源受限的环境中应用。

表 5 不同模型在 TWHD 数据集上的对比实验结果

Table 5 Comparative experimental results of different models on the TWHD dataset

组别	模型	AP@0.5			mAP@0.5	Parameters/M	Size/MB
		two_wheeler	helmet	without_helmet			
1	Faster-RCNN	0.826	0.166	0.044	0.345	28.7	108
2	SSD	0.685	0.430	0.218	0.444	26.3	91.6
3	RT-DETR	0.923	0.788	0.615	0.775	28.4	59.0
4	RTDETR-RESNET101	0.924	0.809	0.619	0.784	41.9	86.0
5	YOLOv3	0.954	0.865	0.767	0.862	103.7	207.8
6	YOLOv5s	0.954	0.842	0.722	0.839	9.1	18.5
7	YOLOv6s	0.95	0.818	0.694	0.821	16.3	32.8
8	YOLOv8s	0.965	0.844	0.718	0.842	9.8	19.9
9	YOLOv8s-WORLDv2	0.967	0.847	0.72	0.845	12.7	25.8
10	YOLOv9s	0.959	0.84	0.735	0.845	7.2	15.2
11	YOLOv10s	0.965	0.847	0.733	0.848	7.2	16.5
12	YOLOv11s	0.969	0.851	0.732	0.851	9.4	19.2
13	YOLOv11l	0.963	0.861	0.768	0.864	25.3	51.2
14	YOLOv11x	0.966	0.87	0.771	0.869	56.8	114.4
15	MobileNetV2-MHSA-SSD	0.963	0.891	0.787	0.88	6.1	21.2
16	Yolov8n-sCG	0.968	0.888	0.788	0.881	3.2	6.6
17	Helmet-YOLO	0.966	0.892	0.796	0.885	12.4	25.3
18	改进 YOLOv8s	0.968	0.902	0.828	0.899	7.4	15.1
19	CPM-YOLO	<b>0.968</b>	<b>0.910</b>	<b>0.841</b>	<b>0.906</b>	<b>2.9</b>	<b>6.3</b>

注: 对比模型的最好效果已经加粗表示。

表 6 YOLOv11s 和 CPM-YOLO 在不同数据集上的对比实验结果

Table 6 Comparative experimental results of YOLOv11s and CPM-YOLO on different datasets

数据集	模型	mAP@0.5	Parameters/M	Size/MB
TWHD	YOLOv11s	0.851	9.4	19.2
	CPM-YOLO	0.906	2.8	6.2
Hard_hat_workers	YOLOv11s	0.623	9.4	19.2
	CPM-YOLO	0.635	2.9	6.3
Helmet	YOLOv11s	0.807	9.4	19.2
	CPM-YOLO	0.831	2.9	6.3
Helmet_detection	YOLOv11s	0.936	9.4	19.2
	CPM-YOLO	0.961	2.9	6.3

## 4 结 论

针对双轮车驾乘者的头盔佩戴在目标遮挡严重、车辆密集和存在行人干扰的复杂道路场景下,出现的误检、漏检以及检测精度低的问题,本文基于 YOLOv11s 模型进行了一系列改进,提出一种基于 CPM-YOLO 算法的头盔智能识别模型。

首先,改进了原有的 FPN-PAN 结构,提出了新的跨尺度特征融合方法 CS-FPN,更好地整合了高阶语义和低阶几何特征信息,提升了头盔检测的精度。同时,通过结合

CNN 和 Transformer,提出了一种名为 PCT 的模块,通过对输入通道进行部分分配,在保证计算效率的同时,优化了特征提取能力。其次,更换模型的损失函数为  $40 \times 40$ ,进一步提高了模型收敛速度和准确率。最后,将下采样的最终尺寸定为  $40 \times 40$ ,删除  $20 \times 20$  的下采样层和  $20 \times 20$  的检测头,在保证小目标检测头不受影响的同时最大限度减少了模型的体积。

实验结果表明,相较于基准模型 YOLOv11s,CPM-YOLO 的 mAP@0.5 提升了 5.5%,参数数量和模型尺寸减少了 69.9% 和 67.2%,在大幅度降低模型复杂度的同时有

效提升了对骑行人员是否佩戴头盔的检测能力,在完全满足检测要求的情况下,便于在资源受限的设备上轻松部署。通过与多种主流算法对比,CPM-YOLO在各个指标上都达到了最优解,能够为实际场景检测双轮车骑行人员头盔佩戴检测任务提供良好的解决方案。

## 参考文献

- [1] WARANUSAST R, BUNDON N, TIMTONG V, et al. Machine vision techniques for motorcycle safety helmet detection [C]. 2013 28th International Conference on Image and Vision Computing New Zealand(IVCNZ 2013), 2013: 35-40.
- [2] SILVA R, AIRES K, SANTOS T, et al. Automatic detection of motorcyclists without helmet[C]. 2013 XXXIX Latin American Computing Conference (CLEI), 2013: 1-7.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN[J]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN SH Q, HE K M, 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 (6): 1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [9] 窦广鉴. 基于MobileNetV2-SSD的安全头盔检测研究[D]. 石家庄: 石家庄铁道大学, 2024.
- DOU G J. Research on safety helmet detection based on MobileNetV2-SSD[D]. Shijiazhuang: Shijiazhuang Railway University, 2024.
- [10] 张碧川, 刘卫东, 米浩, 等. 基于轻量化YOLOv8的安全帽检测[J]. 电脑与电信, 2024(Z1): 35-39.
- ZHANG B CH, LIU W D, MI H, et al. Safety helmet detection based on lightweight YOLOv8[J]. Computer and Telecommunications, 2024 (Z1): 35-39.
- [11] 杨琚钱, 胡平, 戴家树. 基于YOLOv8-scG神经网络的电动车头盔佩戴检测算法[J/OL]. 重庆工商大学学报(自然科学版), 1-10 [2025-01-24]. <http://kns.cnki.net/kcms/detail/50.1155.N.20240710.2030.002.html>.
- YANG J Q, HU P, DAI J SH. Electric vehicle helmet wearing detection algorithm based on YOLOv8 scG neural network [J/OL]. Journal of Chongqing Technology and Business University (Natural Science Edition), 1-10 [2025-01-24]. <http://kns.cnki.net/kcms/detail/50.1155.N.20240710.2030.002.html>.
- [12] 周顺勇, 彭梓洋, 张航领, 等. Helmet-YOLO:一种更高精度的道路安全头盔检测算法[J]. 计算机工程与应用, 2025, 61(2): 135-144.
- ZHOU SH Y, PENG Z Y, ZHANG H L, et al. Helmet-YOLO: A higher precision algorithm for detecting road safety helmets [J]. Computer Engineering and Applications, 2025, 61(2): 135-144.
- [13] 袁宇乐, 汤文兵. 基于改进YOLOv8s模型的电动车骑乘人员头盔佩戴检测[J]. 湖北民族大学学报(自然科学版), 2024, 42(3): 355-360, 367.
- YUAN Y L, TANG W B. Helmet wearing detection of electric vehicle riders based on improved YOLOv8s model [J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2024, 42(3): 355-360, 367.
- [14] MA S, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv: 2307.07662, 2023.
- [15] FENG CH J, ZHONG Y J, GAO Y, et al. Tood: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 3490-3499.
- [16] WANG K X, LIEW J H, ZOU Y T, et al. PANet: Few-shot image semantic segmentation with prototype alignment [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 9196-9205.
- [17] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [18] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

- [19] DU SH J, ZHANG B F, ZHANG P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection[C]. 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning, 2021:92-98.
- [20] PANIGRAHI U, SAHOO P K, PANDA M K, et al. A ResNet-101 deep learning framework induced transfer learning strategy for moving object detection[J]. Image and Vision Computing, 2024, 146: 105021.
- [21] 邓杰, 万旺根. 基于改进 YOLOv3 的密集行人检测[J]. 电子测量技术, 2021, 44(11): 90-95.  
DENG J, WAN W G. Dense pedestrian detection based on improved YOLOv3 [J]. Electronic Measurement Technology, 2021, 44(11): 90-95.
- [22] 张宏飞, 冯永利, 黄金凤. 改进 YOLOv5 的输送带缺陷检测[J]. 电子测量技术, 2024, 47(22): 161-168.  
ZHANG H F, FENG Y L, HUANG J F. Improving YOLOv5s conveyor belt defect detection [J]. Electronic Measurement Technology, 2024, 47(22): 161-168.
- [23] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.
- [24] 曲晨阳, 吕进, 卫策. 基于改进 YOLOv8s 的无人机目标检测算法[J]. 国外电子测量技术, 2024, 43(7): 14-23.  
QU CH Y, LYU J, WEI C. UAV target detection algorithm based on improved YOLOv8s[J]. Foreign Electronic Measurement Technology, 2024, 43 (7): 14-23.
- [25] 师皓, 王澍, 韩健鸿, 等. 基于视觉-文本损失的开放词汇检测大模型对抗样本生成方法[J]. 图学学报, 2024, 45(6): 1222-1230.  
SHI H, WANG SH, HAN J H, et al. Large model adversarial sample generation method for open vocabulary detection based on visual-text loss [J]. Journal of Graphic Science, 2024, 45(6): 1222-1230.
- [26] WANG CH Y, YEH I H, LIAO M H Y, et al. YOLOv9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv:2402.13616, 2024.
- [27] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2025, 37: 107984-108011.

### 作者简介

强浩南, 硕士研究生, 主要研究方向为视觉引导的移动机械臂作业、机器视觉、深度学习。

E-mail: q13111084859@163.com

邹涌波, 硕士研究生, 主要研究方向为非平坦地形中移动机器人的自主导航。

E-mail: zybung@163.com

马立东(通信作者), 博士, 教授, 主要研究方向为智能机器人系统、视觉检测技术、先进检测技术、智能化弯曲与矫直工艺及装备。

E-mail: mald@tyust.edu.cn

李博文, 硕士研究生, 主要研究方向为移动机器人的路径规划。

E-mail: 1203937723@qq.com