

文章编号: 2095-2163(2023)08-0169-06

中图分类号: TP391.41

文献标志码: A

# 改进的 YOLOv5 算法在安全帽检测的应用

王子元, 王国中, 顾嘉城

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 佩戴安全帽可以有效保障建筑施工人员的人身安全。针对目前安全帽检测算法检测效率低,漏检率高的问题,本文提出了一种基于改进 YOLOv5 网络模型的安全帽检测算法。首先,在 YOLOv5 的骨干网络中插入混合注意力模块,降低特征提取网络的数据维度,使网络关注于图片中安全帽特定区域,提高网络的安全帽检测性能;然后,使用双向特征金字塔结构代替原网络中的特征金字塔结构,融合不同层级的特征并保留特征图中浅层信息,提高模型的计算效率;最后,使用  $EIoU$  作为网络的损失函数,提高改进模型的识别准确率。实验结果表明,改进模型在扩充的安全帽数据集上检测精度达到 83.1%,每秒检测速度为 48.1,相比于原始算法模型精度提高了 4.5%,检测速度提高了 1.5 帧。

**关键词:** 安全帽检测; YOLOv5; 混合注意力机制; 双向特征金字塔

## Application of improved YOLOv5 algorithm in hard hat detection

WANG Ziyuan, WANG Guozhong, GU Jiacheng

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** Wearing a safety helmet can effectively ensure the personal safety of construction workers. Aiming at the problems of low detection efficiency and high missed detection rate of the current helmet detection algorithm, this paper proposes a helmet detection algorithm based on the improved YOLOv5 network model. First, a hybrid attention module is inserted into the backbone network of YOLOv5 to reduce the data dimension of the feature extraction network, so that the network can focus on the specific area of the helmet in the image, and improve the performance of the network's helmet detection; then, the bidirectional feature pyramid structure is used to replace the original network. The feature pyramid structure fuses the features of different levels and retains the shallow information in the feature map to improve the computational efficiency of the model; finally,  $EIoU$  is used as the loss function of the network to improve the recognition accuracy of the improved model. The experimental results show that the detection accuracy of the improved model on the expanded helmet dataset reaches 83.1% and the detection speed per second is 48.1. Compared with the original algorithm, the model accuracy is increased by 4.5%, and the detection speed is increased by 1.5 frames.

**[Key words]** helmet detection; YOLOv5; CBAM; BiFPN

## 0 引言

安全帽对于建筑施工人员的人身安全保护至关重要,对于施工场所的意外坠物,安全帽的外壳可以有效减缓冲击,缓解高空坠物对工人头部的伤害,保护施工人员的安全。在以往的施工场所中,通常采用人工监督的方式来保证施工人员佩戴安全帽,但由于施工范围大,施工人员多,部分人员安全意识薄弱,经常出现未按规定佩戴安全帽作业的施工人员。考虑到人工监督方法耗费大量人力物力,并且效率

低下,因此本文利用施工场所的监控视频画面设计了一种自动检测施工人员是否佩戴安全帽的方法。

传统的安全帽检测方法主要是利用人工构造的特征来完成检测任务。如 HOG 特征与 SVM<sup>[1]</sup>相结合的方法, Harr 特征与 Adaboost<sup>[2]</sup>相结合的方法等。这些传统机器学习的安全帽检测算法需要人为设计检测特征和训练分类,对于复杂的施工环境来说主观性较强,鲁棒性差且不易泛化。

基于深度学习的安全帽检测方法<sup>[3]</sup>,不需要人工设置特征,实现方法较为简洁。以 R-CNN<sup>[4]</sup>为代

基金项目: 国家重点研发计划 (2019YFB1802702)。

作者简介: 王子元(1997-),男,硕士研究生,主要研究方向:深度学习、图像处理; 王国中(1962-),男,教授,博士生导师,主要研究方向:视频编解码、图像处理、机器学习; 顾嘉城(1994-),男,硕士研究生,主要研究方向:图像处理、模式识别。

通讯作者: 王国中 Email: wanggz@sues.edu.cn

收稿日期: 2022-08-26

表的双阶段检测算法<sup>[5]</sup>虽然检测精度较高,但检测速度较慢。以 SSD<sup>[6]</sup>和 YOLO<sup>[7]</sup>为代表的单阶段算法在检测速度方面具有优势,徐凯等学者通过改进 YOLOv3 的损失函数来降低检测模型的误差。Zhou 等学者<sup>[8]</sup>通过替换 YOLOv4 的骨干网络使模型轻量化,提高了检测效率。

以上网络模型虽然都具有较好的检测精度,但是模型较大,不能满足施工环境下低算力设备的部署需求。因此,本文基于 YOLOv5<sup>[9-10]</sup>网络设计了一种轻量级安全帽检测算法:

(1)提取原网络模型特征图中的通道注意力特征和空间注意力特征,将 CBAM<sup>[11]</sup>模块融合到骨干网络的卷积后,让网络关注于图片中出现安全帽的特定区域,加快网络训练时的收敛速度。

(2)在原模型的颈部网络部分使用双向特征金字塔(BiFPN)<sup>[12]</sup>结构代替原网络中的特征金字塔

结构,融合不同层级的特征并保留特征图中浅层信息,提高模型的计算效率。

(3)使用 GhostConv<sup>[13]</sup>模块替换原始网络中的普通卷积,减少网络的参数量,轻量化模型,并使用  $ElIoU$ <sup>[14]</sup>作为网络的损失函数,提高改进模型的识别准确率。

## 1 YOLOv5 网络结构

YOLO 系列算法是具有代表性的单阶段目标检测算法,与其他两阶段和单阶段算法相比具有检测速度快、检测精度高等优点。YOLOv5 算法将网络模型分为 4 个不同的尺寸,模型由小到大分别是 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x,随着模型的深度和宽度增加,模型体积也依次增大。本文中选择尺寸较小的 YOLOv5s 模型作为基准模型进行实验,YOLOv5s 的网络结构如图 1 所示。

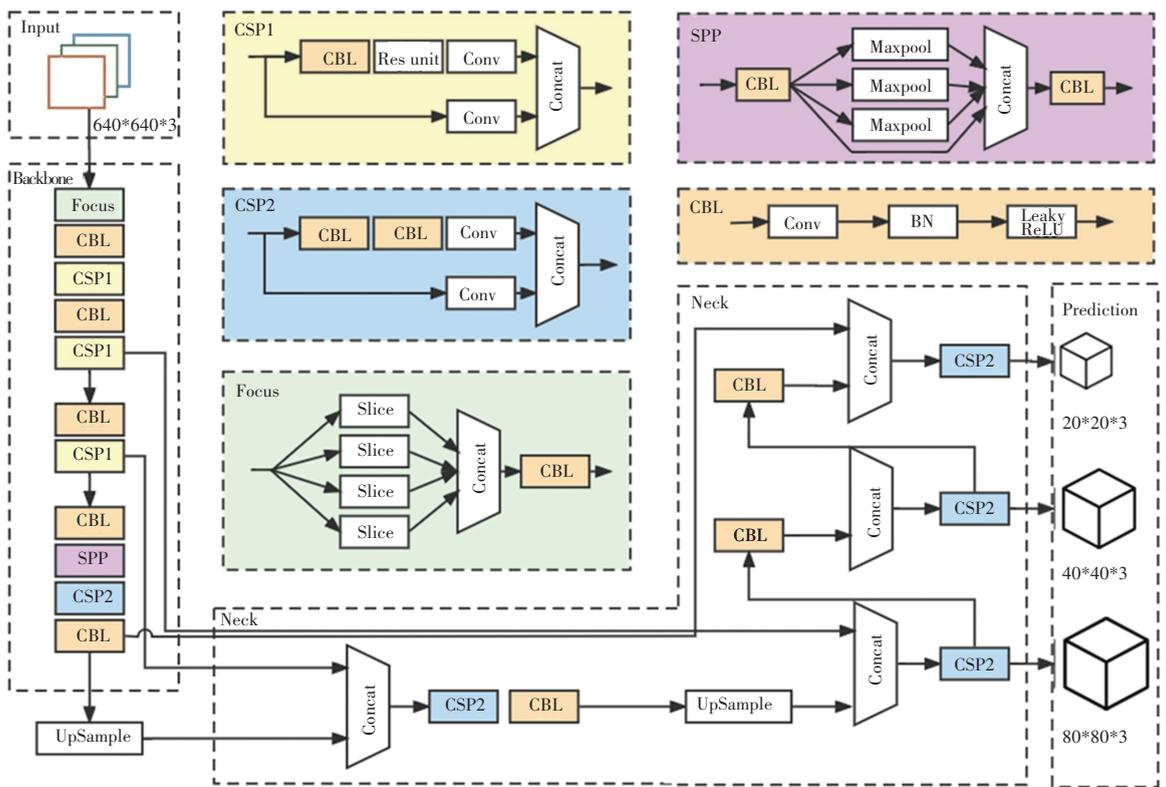


图 1 YOLOv5s 网络结构

Fig. 1 YOLOv5s network structure

从 YOLOv5s 的网络结构图中可以看出, YOLOv5s 网络模型主要由 4 部分构成,分别是图片输入端 (Input)、骨干网络 (Backbone)、颈部模块 (Neck)、预测模块 (Prediction)。在图片输入端使用 Mosaic 数据增强、图片缩放和自适应锚框计算增强输入数据的多样性。骨干网络由 Focus 下采样、CSP

结构和 SPP 池化金字塔结构组成,用于提取图像特征信息。颈部模块使用特征金字塔 (FPN)<sup>[15]</sup>与金字塔注意力网络 (PAN)<sup>[16]</sup>相结合的方式融合特征,传递不同尺度的特征信息。预测模块输出 3 组不同尺寸的预测信息。

## 2 改进 YOLOv5 算法

### 2.1 添加 CBAM 注意力机制

CBAM(Convolutional Block Attention Module)是一种集成了通道注意力和空间注意力的轻量化混合注意力机制。通过学习关注特征图中与安全帽相关的信息,抑制不重要的特征来提升检测效率。CBAM 的结构如图 2 所示。

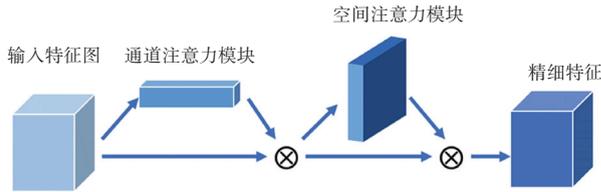


图 2 CBAM 结构图

Fig. 2 CBAM structure diagram

在施工环境中,由于安全帽的背景复杂多样,原始网络对无效信息的剔除能力有限,而背景信息对网络的检测性能影响较大,因此通过在 Focus 下采样后引入 CBAM 注意力机制来确定模型重点关注的的目标特征。

### 2.2 改进特征融合网络

在 YOLOv5 原始网络的特征融合部分,使用自下而上和自上而下的方式融合高层级的特征信息如图 3 所示。而对于安全帽检测网络来说,安全帽在画面中多为小尺寸目标,使用 FPN+PAN 结构会丢失图像中的小尺度特征信息,限制了网络模型的小尺度检测能力。

本文中使用的 BiFPN 替换原始特征融合结构见图 3,通过删去边缘融合节点来加快网络计算速度,并使用跳跃连接传递底层特征信息,通过融合底层特征与高级特征的信息来增强网络对安全帽尺度目标的检测能力。

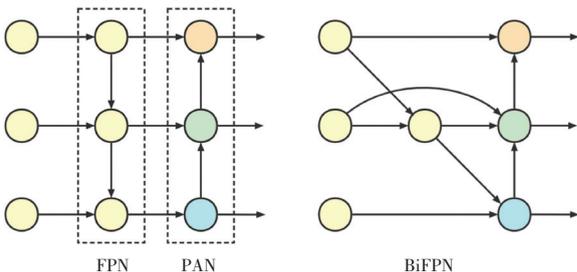


图 3 改进特征融合结构

Fig. 3 Improved feature fusion structure

BiFPN 对不同的输入特征分配不同的权重,使网络通过调整权重来衡量不同特征的重要程度,并采用快速归一化方法,可按式(1)来计算:

$$Out = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot In_i \quad (1)$$

其中,  $Out$  表示输出特征;  $In$  表示输入特征;  $\omega_i$  采用  $ReLU$  激活函数;  $\varepsilon = 0.0001$  为数值稳定小量。

### 2.3 改进卷积计算

GhostModel 是一种新型的卷积计算方法,基于原始特征图使用一系列的线性变换从原始特征图中生成相似特征,使用该方法可以减少网络模型的参数量,搭建轻量化网络模型。GhostModel 的结构如图 4 所示。

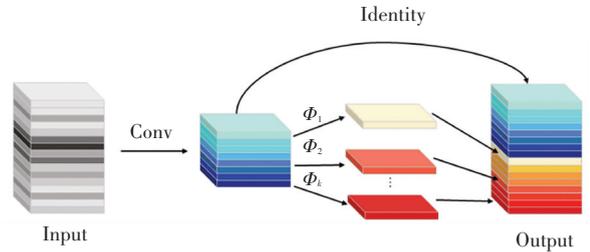


图 4 Ghost 模块

Fig. 4 Ghost model

对于给定输入数据,  $X \in \mathbb{R}^{c \times h \times w}$ ,  $c$  表示输入通道数,  $w$  和  $h$  分别表示输入数据的宽度和高度。GhostNet 的输出特征图可以表示为:

$$Y' = X * f' \quad (2)$$

其中,  $Y' \in \mathbb{R}^{w' \times h' \times m}$  使用一般卷积  $f' \in \mathbb{R}^{c \times k \times k \times m}$  生成  $m$  个特征图,然后对  $m$  个特征图进行线性运算可以生成  $s \times m$  个特征图,用到的数学公式为:

$$y_{ij} = \Phi_{i,j}(y'_i), \forall i = 1, \dots, m, j = 1, \dots, s \quad (3)$$

其中,  $y'_i$  表示  $Y'$  中第  $i$  个原始特征图;  $\Phi_{i,j}$  表示第  $j$  个线性运算;  $y_{ij}$  表示生成的特征图。最终得到  $n = m \cdot s$  个特征图  $Y = [y_{11}, y_{12}, \dots, y_{ms}]$ , 计算可得 GhostModel 的计算量和参数量约为一般卷积的  $\frac{1}{s}$ 。因此,使用 Ghost 模块可以在保证输出特征图

一致的情况下,降低模型的计算量和参数量。

Ghost bottleneck 可以根据步长分为 2 种结构。当步长为 1 时,在输入端的首个 GhostModule 增加通道数,起到扩展作用。在经过靠近输出端的 GhostModule 后,数据通道数减少与输入相匹配并与输入相加。对于步长为 2 的情况,添加了步长为 2 的深度卷积来实现。

在本文中使用的步长为 1 的 Ghost bottleneck 代替原 YOLOv5 模型中的 CSP2 模块,使用步长为 2 的 Ghost bottleneck 代替原模型中的 CBL 卷积模块。

## 2.4 改进损失函数

YOLOv5 中使用  $GIoU$  (Generalized IoU loss) 作为损失函数。当预测框与标签框未完全重叠的情况下  $GIoU$  的方法可以衡量 2 个标签框的重叠程度,但是当预测框在标签框内部的时候,  $GIoU$  退化成了  $IoU$ , 无法衡量预测框与标签框之间的距离。

本文中使用了  $EIOU$  (Efficient IOU) 作为边界框损失函数,  $EIOU$  损失函数可以分为 3 个部分, 分别是: 预测框与真实框之间的中心距离损失、重叠损失和宽高损失, 数学定义公式具体如下:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{C_\omega^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (4)$$

其中,  $b^{gt}$  与  $b$  分别表示真实边界框与预测边界框的中心点;  $\rho$  表示中心点的欧氏距离;  $c$  表示包围真实框与预测框的最小外接矩形的对角线距离;  $C_\omega$  表示该最小外接矩形的宽度;  $C_h$  表示该最小外接矩形的高度。

$EIOU$  对于目标框和预测框的长和宽分别进行计算, 综合评估预测框与目标框的中心点距离、重叠面积, 使收敛速度更快, 回归精度更高。

## 3 实验结果与分析

### 3.1 实验环境与参数

网络训练在 Ubuntu18.04 系统上进行, 系统的运行内存为 16 GB, 显卡为英伟达 RTX3060。处理器为 AMD Ryzen 7 5800H 3.20 GHz, 使用深度学习框架 PyTorch 编写改进的 YOLOv5 模型进行训练。编程环境为 Python 3.7。GPU 的加速库为 CUDA11.3 和 CUDNN8.2.1。

在网络的训练过程中, 使用自适应锚框、图片尺寸缩放和 Mosaic 数据增强技术对输入图片进行数据增强, 网络模型的图片输入 (input size) 大小设置为  $640 \times 640$ , 动量参数 (momentum) 设置为 0.9, 迭代次数 (epoch) 设置为 100, 批量大小 (batch size) 设置为 8。

### 3.2 数据集

本文中所使用的安全帽数据集基于网络图片和施工现场监控视频截取, 共收集到 6 545 张图片。使用 Labelimg 软件将收集到的图片进行标注, 对于佩戴安全帽的工人标注为“hat”, 未佩戴安全帽的工人标注为“person”。最后将标注信息保存为 VOC

数据集格式, 并按照 9 : 1 的比例划分为训练集和测试集, 训练集图片共有 5 890 张, 测试集共有 655 张。

### 3.3 评价指标

在计算机视觉领域, 对于目标检测的性能通常使用召回率 (Recall)、精确率 (Precision) 和准确率 (Accuracy) 等指标来进行评价。

精确率表示在预测到的正样本中实际为正样本的概率, 召回率表示在实际的正样本中预测为正样本的概率, 准确率表示预测正确的概率, 研究推得的公式见如下:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \end{aligned} \quad (5)$$

以召回率为 X 轴坐标, 以准确率为 Y 轴坐标, 设置相应的阈值后画出的曲线就是  $P-R$  曲线。 $P-R$  曲线与坐标轴围成图形的面积被称为平均精确度 (AP)。

AP 就是平均精确度, 对  $PR$  曲线上的 Precision 值求均值。对于  $PR$  曲线来说, 使用积分来进行计算, 即:

$$AP = \int_0^1 P(r) dr \quad (6)$$

### 3.4 实验结果

为了验证本文改进方法对 YOLOv5s 基准检测模型的性能提升, 设计如下消融实验, 所有实验均在英伟达 RTX3060 GPU 和相同数据集下进行。 $mAP(0.5)$  表示在检测框与真实框交并比大于等于 0.5 的情况下, 2 种类别检测结果的平均精度。消融实验见表 1。

表 1 消融实验

Tab. 1 Ablation experiment

实验	CBAM	BiFPN	GhostNet	$EIoU$	$mAP(0.5)\%$
1					78.6
2	✓				81.2
3	✓	✓			82.5
4	✓	✓	✓		82.8
5	✓	✓	✓	✓	83.1

从表 1 中可以看出, 在加入 CBAM 注意力机制后, 网络模型的平均精度提升了 2.6%, 证实了注意力机制是可以关注到图片中有效特征的, 而在使用

BiFPN 特征融合机制后平均精度增长了 1.3%, 这是由于安全帽在图片中的尺寸较小, 使用 BiFPN 特征融合机制可以有效地提取特征图像中的小尺度信息, 更有利于安全帽的检测。在使用 Ghost 卷积计算后, 网络精度仅提升了 0.3%, 因为 Ghost 模块的加入为轻量化模型做出较多贡献, 对精度的提升有限。在实验 5 中使用所有改进策略后平均精度较原始网络提升了 4.5%, 表明组合改进策略可以有效提升安全帽检测模型的性能。

为了在不同检测算法中横向对比改进网络的安全帽检测性能, 本文选取了传统检测算法和基于深度学习的检测算法与改进模型进行对比, 其中添加所有改进的模型记为 I-YOLOv5。

对比实验结果见表 2。从表 2 中可以看出, 单阶段算法在检测速度方面具有明显优势, 其中 YOLOv5 算法速度最为突出, 改进的算法模型 I-YOLOv5 与原始算法相比平均精度提高了 4.5 个百分点, FPS 提升了 1.5。传统机器学习算法 HOG+SVM 检测效率低下, 对于复杂背景下的安全帽目标检测精度较低。由于安全帽目标多为小尺寸目标, 两阶段安全帽检测算法 Faster RCNN 表现不佳。SDD 算法虽然也达到较高的检测速度, 但是与 YOLOv5 算法相比仍存在一定的差距。本文中 I-YOLOv5 算法在检测速度与检测精度方面表现良好。

表 2 对比实验结果

Tab. 2 Comparative experimental results

实验	骨干网络	mAP (0.5) %	FPS
HOG+SVM	-	42.8	0.9
Faster RCNN	VGG-16	53.6	12.5
SSD	VGG-16	57.1	23.2
YOLOv5s	CSPNet	78.6	46.6
I-YOLOv5	GhostNet	83.1	48.1

为了更直观地对比检测性能, 实验选取了不同施工环境下的安全帽图片, 分别使用 YOLOv5 算法检测, 并与本文 I-YOLOv5 算法检测结果做对比, 检测结果如图 5、图 6 所示。

从图 5 中可以看出, 对于图片中的大部分安全帽 YOLOv5 算法都能够成功检出, 但是将施工人员佩戴的头套错误识别为安全帽, 而且在人员遮挡的图像部分, 遮挡部分施工人员未佩戴安全帽没有被模型检出。从图 6 的 I-YOLOv5 算法的检测结果中可以看出, 对于遮挡的施工人员也能够成功检出, 而且佩戴头套的施工人员也被正确检出。在 I-YOLOv5 模型中, 由于添加了 CBAM 注意力机制, 能够使模型关注于施工人员头部以及安全帽特征, 抑制背景或遮挡对模型检测产生的影响。同时, 使用了 *EIoU* 损失函数, 可以使安全帽检测框的尺寸更加精准, 准确定位到施工人员的头部。



图 5 YOLOv5s 算法检测效果

Fig. 5 YOLOv5s algorithm detection effect



图 6 I-YOLOv5 算法检测效果

Fig. 6 I-YOLOv5 algorithm detection effect

### 4 结束语

为了保障施工人员的安全, 本文对目前安全帽

检测算法进行分析, 并针对复杂环境下的安全帽检测算法做出改进。在 YOLOv5 骨干网络部分插入 CBAM 注意力机制, 让网络关注于图片中的小尺寸特征, 提升网络模型对安全帽区域的敏感程度。通

过使用 BiFPN 替换原始特征融合结构,删去边缘融合节点来加快网络计算速度。使用 GhostNet 卷积替换原始网络中的一般卷积,使模型轻量化,提高检测速度。使用 *EIoU* 损失函数,改善了定位精度低和训练过程中检测框回归速度慢的问题。相比于 YOLOv5s 算法模型,改进的算法模型的精度提高了 4.5%。每秒检测速度达到 48 帧。未来将在保持安全帽检测性能的情况下对模型做进一步优化,在保证准确率的前提下,降低模型的复杂度。

## 参考文献

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 05). Washington, DC, USA: IEEE, 2005, 1: 886-893.
- [2] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001). Kauai, HI, USA: IEEE, 2001, 1: I-I.
- [3] 陈科圻,朱志亮,邓小明,等.多尺度目标检测的深度学习研究综述[J].软件学报,2021,32(04):1201-1227.
- [4] 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. Columbus, OH, USA: IEEE, 2014: 580-587.
- [5] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 39(6): 1137-1149.
- [6] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]//European Conference on Computer Vision. Cham; Springer, 2016: 21-37.
- [7] RDEMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [8] ZHOU M, FANG Z, ZHAO B, et al. Safety Helmet Wearing Detection and Recognition Based on YOLOv4 [C]//2021 3<sup>rd</sup> International Academic Exchange Conference on Science and Technology Innovation (IAECST). IEEE, 2021: 798-802.
- [9] DENG Benyang, LEI Xiaochun, MIAO Ye. Safety helmet detection method based on YOLO v4 [C]//2020 16<sup>th</sup> International Conference on Computational Intelligence and Security (CIS). Guangxi, China: IEEE, 2020: 155-158.
- [10] WANG Junfan, CHEN Yi, GAO Mingyu, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection [J]. arXiv preprint arXiv:2112.08782, 2021.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: dblp, 2018: 3-19.
- [12] TAN Mingxing, PANG Ruoming, LE Q V. Efficientdet: Scalable and efficient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 10781-10790.
- [13] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 1580-1589.
- [14] ZHANG Yifan, REN Weiqiang, ZHANG Zhang, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2117-2125.
- [16] WANG Wenhai, XIE Enze, SONG Xiaoge, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 8440-8449.
- [7] SODHRO A H, OBAIDAT M S, ABBASI QH, et al. Quality of service optimization in an IoT-driven intelligent transportation system [J]. IEEE Wireless Communications, 2019, 26(6): 10-17.
- [8] 李俊伯,秦品乐,曾建潮,等.基于超分辨率网络的CT三维重建算法[J].计算机应用,2022,42(02):584-591.
- [9] 余玛俐,张海.基于低秩矩阵填充与全变分约束的HDR成像[J].计算机工程,2019,45(04):262-266,274.
- [10] 陈伟.运动捕获技术在3D计算机动画制作中的应用[J].自动化与仪器仪表,2017(12):143-144.
- [11] 徐艾明,黄宇星,沈秋.基于三维Saab变换的高光谱图像压缩方法[J].北京航空航天大学学报,2022,48(08):1505-1514.
- [12] 张迅,李建胜,王安成,等.无人平台视觉导航算法验证仿真系统的设计与实现[J].测绘科学技术学报,2021,38(01):9-14,20.

(上接第168页)