

基于 YOLOv5 的安全帽检测方法研究

张帅帅

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘要:目的 针对工业场所背景复杂导致安全帽的检测精度低、效果不佳等问题,提出了一种基于 YOLOv5 的智能检测安全帽的方法。方法 首先在原模型 YOLOv5 的骨干网络中增加注意力机制,增强对不同尺寸目标特征的提取,使得网络将注意力聚焦在含有安全帽的区域,增强了网络对安全帽信息的提取,以此有效提取安全帽的特征信息;在预测层使用 EIoU 损失函数,考虑宽和置信度的差异、高和置信度的差异,把纵横比拆开,以此改善样本不平衡问题,提升收敛速度的同时提高了回归精度。结果 根据实验结果,改进的算法平均精度达到了 94.7%。相比于 YOLOv5 算法平均检测精度提高了 2.2%,相比于 YOLOv3 算法平均检测精度提高了 12.6%,可以有效地检测安全帽。结论 在同样的背景环境下,改进的算法可以有效地检测出远距离的小目标,对于复杂背景信息的图片,也可以准确地检测出目标。改进的算法有效地改善了原算法中小目标漏检和误检情况,也提高了检测精度。

关键词:安全帽;注意力机制;深度学习;损失函数

中图分类号:TP391 文献标识码:A doi:10.16055/j.issn.1672-058X.2025.0001.006

Research on Safety Helmet Detection Method Based on YOLOv5

ZHANG Shuaishuai

School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective In response to the issues of low detection accuracy and poor performance of safety helmet detection in complex industrial environments, a method for intelligent detection of safety helmets based on YOLOv5 was proposed.

Methods First, an attention mechanism was incorporated into the backbone network of the original YOLOv5 model, which enhanced the extraction of features for targets of different sizes. This modification directed the network's attention toward regions containing safety helmets, thereby improving the network's ability to capture safety helmet information and effectively extract the corresponding features. In the prediction layer, the EIoU loss function was employed, which considered the differences in width and confidence, as well as height and confidence, while splitting the aspect ratio. This approach addressed the issue of sample imbalance, accelerated convergence, and enhanced regression accuracy. **Results** According to experimental results, the improved algorithm achieved an average precision of 94.7%. The improved algorithm improved the average detection accuracy by 2.2% compared with the YOLOv5 algorithm and 12.6% compared with the YOLOv3 algorithm, effectively detecting safety helmets. **Conclusion** Under the same background environment, the improved algorithm can effectively detect small targets at long distances and accurately detect targets in images with complex background information. The improved algorithm effectively addresses the issues of missed detection and false detection of small targets in the original algorithm and also improves detection accuracy.

Keywords: safety helmet; attention mechanism; deep learning; loss function

收稿日期:2023-06-13 修回日期:2023-07-17 文章编号:1672-058X(2025)01-0042-06

作者简介:张帅帅(1996—),女,河南周口人,硕士研究生,从事计算机视觉研究。

引用格式:张帅帅.基于 YOLOv5 的安全帽检测方法研究[J].重庆工商大学学报(自然科学版),2025,42(1):42-47.

ZHANG Shuaishuai. Research on safety helmet detection method based on YOLOv5[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(1): 42-47.

1 引言

安全帽是个人防护设备,应用于建筑、矿山、运输多个行业。在这些场景作业时,安全帽可以很好地保护头部。为保障工人的安全,工作人员必须佩戴安全帽。但由于人数较多,监管和安全意识缺失等因素,有时工作人员未能及时佩戴安全帽,而产生的安全事故时有发生。为规避这种潜在的危险,研究安全帽佩戴的检测方法是有意义的。

目标检测是视觉领域的一个热点。对于安全帽的检测,是目标检测的一个应用点^[1]。现在目标检测应用的范围广泛,可以用于人脸识别,车辆检测,视频监控等领域。传统的检测是使用人工选择特征,然后使用分类器进行训练。但这一传统的方法设计过程繁琐,泛化能力弱,在工业的运用上存在极大的局限性,近些年,随着卷积神经网络的发展,转而利用深度学习对需要检测的目标上进行特征提取,避免了人工设计特征的繁琐性,检测的精度和速度不断提高。2014年 Girshick 等^[2]设计出了 R-CNN 算法,利用候选框实现物体的检测,代替以往人工设计特征的方法,使得卷积神经网络在物体检测方面的发展更加迅速。基于卷积神经网络的检测算法分为以 R-CNN、Fast R-CNN^[3]、Faster R-CNN^[4] 等代表的两阶段算法和以 SDD^[5]、YOLO^[6-8] 等代表的单阶段算法。徐守坤等^[9]提出了一种基于两阶段网络的 Faster RCNN 的安全帽检测方法,该算法采用多尺度训练,并且增加锚点数量,可以检测不同尺寸目标。但改进的算法计算量大,检测效率低。方明等^[10]基于 YOLOv2 在网络中加入密集块,实现底层语义信息与深层语义信息的融合,提高了对目标检测的灵敏度,采用轻量化的网络,对模型压缩,提高检测的速度,但改进的算法还存在一定的误检率。赵红成等^[11]在 YOLOv3 结构的基础上,添加注意力机制,使用 GIoU 损失函数,通过对实验结果的分析,改进的模型在提高检测的准确度的同时也满足了实时性要求,但改进的算法对于小目标的检测会出现漏检和误检情况。

以上方法虽然提高了模型检测的性能,但运用到复杂背景环境中,还是存在一些困难。复杂的外部环境会给检测模型带来困难。双阶段算法,冗余参数较多,模型本身的计算量较大,导致训练模型的时间长,检测的效率低,并且后期的部署困难。安全帽检测场景复杂、检测目

标较小,漏检和一定程度的误检案例多,因此采用单阶段算法 YOLOv5 作为基础网络模型,在骨干网络中添加 CBAM 注意力模块,增强检测时对目标关键信息的提取,提高网络的泛化能力。采用轻量化的注意力模块,节省计算资源,提高检测速度。然后在预测阶段改变损失函数,使用 EIoU 损失函数,提高收敛速度。实验表明,改进的 YOLOv5 对于安全帽的检测效果得到提升,可以更好地满足安全帽检测的需求。

2 基础网络模型

本文选用 YOLOv5 算法中的 YOLOv5s 模型作为基础模型。YOLOv5s 是单阶段目标检测算法,模型相对较小。其中主要包括四部分输入层、骨干网络层、特征融合层、输出层。

第一部分是 Input 输入层,采用 Mosaic 数据增强的方法,该方法是将 4 张图片随机地拼凑到一起,形成一个大图像里面包含 4 个不同的图像,然后,把包含 4 个不同的图像的大图像作为训练集的图像进行数据增强操作,提高数据的多样性。然后通过自动计算出最合适图像的锚框参数,而不用手动再去设置,训练过程中,自适应地选择样本,降低了样本的影响,提高检测精度。

第二部分是 Backbone 骨干网络层,使用 Focus 和 CSP^[12] 结构。Focus 是网络中第一个卷积层,是对输入层得到的特征图进行下采样,提取重要的信息,实现特征层的下采样和压缩。CSP 结构把输入的特征图一部分经过子网络进行处理,一部分进入下一层的卷积操作,然后,把得到的两部分特征图连接起来,作为下一层的输入。采用多种特征提取模块对安全帽区域进行特征提取,生成目标语义信息,以此加快网络对训练集的训练和推断速度。

第三部分是 Neck 特征融合层,使用 FPN+PAN^[13] 结构,采用自底向上的方法传递位置信息,聚合骨干层中不同检测层的参数,可以减少网络的深度,增加对安全帽信息的融合。

第四部分是 Prediction 输出层,这一层负责检测大尺度目标、中尺度目标和小尺度目标。使用 anchor box 来预测目标边界框的位置和大小,显示预测的类别和类别的概率。YOLOv5s 网络模型大小相对小一点,计算量也少,可用于轻量化研究。网络模型结构如图 1 所示。

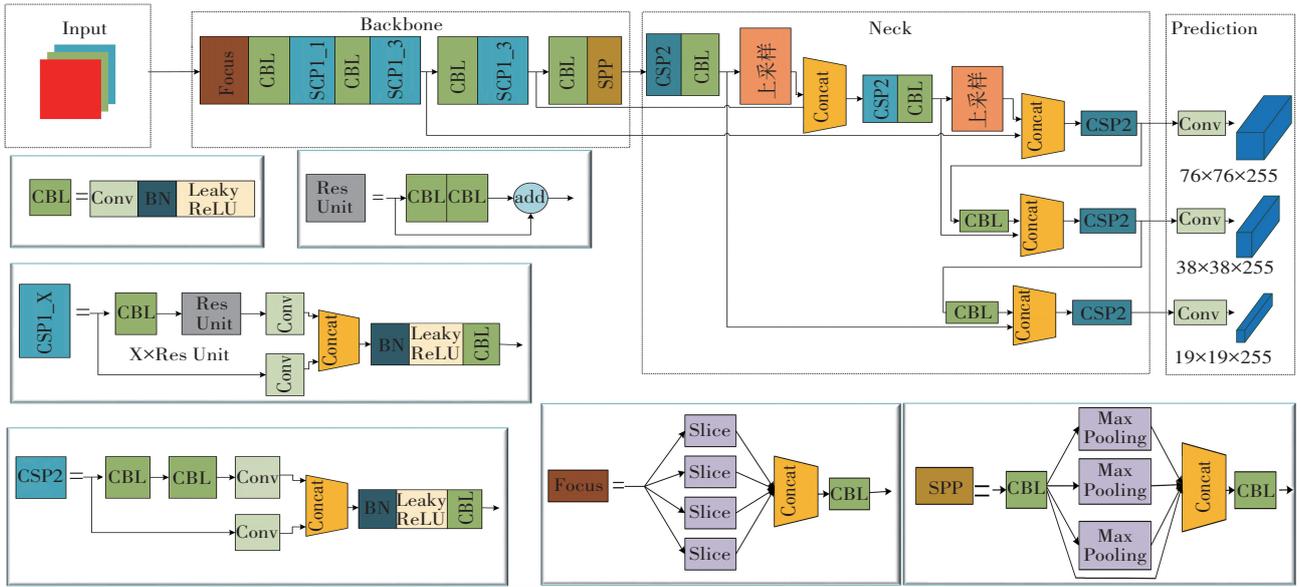


图 1 YOLOv5s 网络结构图

Fig. 1 Structure diagram of YOLOv5s network

3 模型的优化

3.1 添加注意力机制

注意力机制和人类的视觉关注相似,人类眼中通常只会注意到重要的内容,忽略掉一些不相干的信息。注意力机制通常可以对信息部位的特征更加关注,对于杂乱的背景信息可以选择性地抑制,对于检测时特征的提取起到了关键的作用。以此把注意力机制引入到神经网络中,优化网络模型,提高检测的性能。

CBAM^[14] (Convolutional Block Attention Module) 是一个轻量级的网络,操作相对简单,能力高效的模块,网络结构如图 2 所示。CBAM 和别的注意力机制不同,他不仅仅包含通道注意力模块 (Channel Attention Module, CAM),而且还包含了空间注意力模块 (Spatial Attention Module, SAM)。输入的特征传入 CAM 后与最初的输入进行加权,然后经过 SAM 与输入加权后得到最后的结果,可以减少计算量。两个模块网络训练的过程独立,而且利用通道和空间维度,重新加权和压缩安全帽的输入特征,增强对安全帽特征的泛化能力。

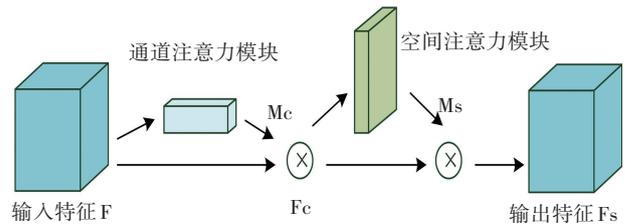


图 2 CBAM 网络结构图

Fig. 2 Structure diagram of CBAM network

通道注意力模块,如图 3 所示,输入特征图,注意力机制关注所需要的特征,将特征图压缩到空间层面上,获得一维矢量,对其进行分析,同时也使用平均池化和最大池化,经过多层感知机 (Multi-Layer Perceptron, MLP)。紧接着对 MLP 输出的特征进行对应元素相加求和操作,最后通过 sigmoid 激活函数,生成通道注意力图。CAM 公式如式(1):

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (1)$$

其中, $W_0 \in R^{r \times c}$, $W_1 \in R^{c \times c}$, c 是通道数, r 是缩减率, R 是尺度空间, F_{max}^c 最大池化, F_{avg}^c 是平均池化, σ 是 sigmoid 激活函数。

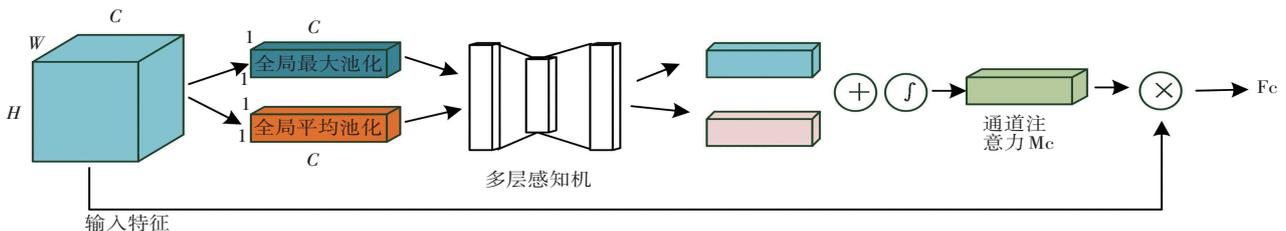


图 3 通道注意力模块图

Fig. 3 Diagram of channel attention module

空间注意力模块,结构图如图 4 所示。SAM 是从通道注意力模块传递的信息进行优化。进行平均和最大池化,平均池化是在通道中提取 $H \times W$ 次中的平均值,最大池化是在通道中提取 $H \times W$ 次中的最大值。然后利用一个卷积层进行连接操作,最后通过激活函数,得到新特征图,最后将这个特征图与该模块的输入特征作乘法运算,得到最后生成的特征图。使用激活函数可以避免输出线性函数,激活函数能给神经元引入非线性的元素,使网络层可以接近于非线性函数,因此网络就可以用到非线性的模型中。SAM 公式如式(2):

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (2)$$

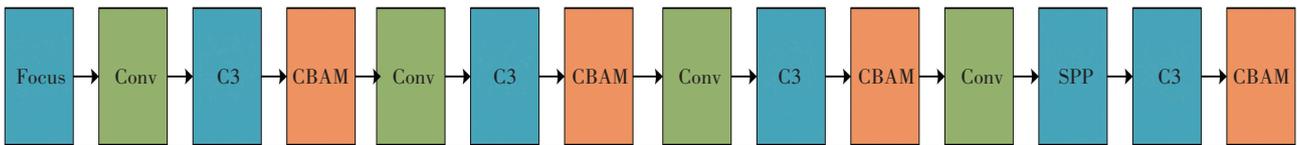


图 4 空间注意力模块图

Fig. 4 Diagram of spatial attention module

骨干层加入 CBAM 坐标注意力机制后的网络结构图,如图 5 所示。注意力模块这一部分利用最大池化和平均池化相加或堆叠。利用通道和空间维度,对输入的信息进行特征提取,增加关键信息安全帽的关注度。

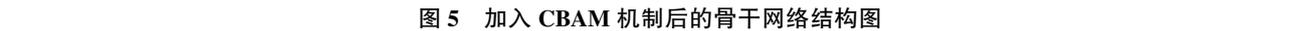


图 5 加入 CBAM 机制后的骨干网络结构图

Fig. 5 Structure diagram of the backbone network after adding CBAM

3.2 优化损失函数

预测层评定模型预测值与真实值之间的差异是采用损失函数来评定的,其收敛速度越快,值越小,说明越好。原模型 Prediction 部分,所使用的损失函数是 GIou^[15],如式(3)所示:

$$L_{GIoU} = 1 - L_{IoU} + \frac{|C| - |A^c \cup B|}{|A^c|} \quad (3)$$

其中, A^c 是预测框和真实框之间的最小外接矩形, $L_{IoU} = \frac{A \cap B}{A \cup B}$,即 IoU 是预测框和真实框交集和并集相比的值。

GIoU 虽然对非重合区域也关注到了,但对于水平框和垂直框,需要更多次地迭代才能收敛,收敛速度慢。为解决这一问题,本文采用 EIou^[16] 作为模型的损失函数,它考虑了锚框的重叠面积、中心点距离以及宽、高长度的真实差异问题。惩罚项的损失函数式为

$$L_{EIou} = L_{IoU} + L_{dis} + L_{asp} = 1 - L_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (4)$$

式(4)中, L_{IoU} 表示重叠损失, L_{dis} 表示中心距离损失, L_{asp} 表示宽、高损失, b, b^{gt} 表示预测边界框与真实边界框的中心点,最小外接框的宽度和高度是 C_w 和 C_h 。EIou loss 中引入 Focal loss 来改善因为回归质量好,导致的样本优化困难这一问题,对于损失相当于一个加权的作用,给回归目标一个好的损失,可以提高回归精度,提高收敛速度。

为了方便看出改进损失函数的优点,选取 GIoU、

DIoU、CIoU 和 EIou 来做对比实验,实验结果如表 1 所示。虽然别的损失函数比原函数也有提升,但提升最大的是 EIou,因此本文选用 EIou 损失函数。

表 1 损失函数对比

Table 1 Comparison of loss functions

损失函数	mAP@ 0.5 (%)
GIoU	89.1
DIoU	91.1
CIoU	92.3
EIoU	92.8

本文将原模型预测层中的 GIou loss 更换为 EIou loss,因为对于边界框回归任务,EIoU 把之前所用的纵横比用计算宽、高的差异值来取代。把算法中所用的 GIou loss 改为 EIou loss,可以加速收敛,解决样本回归中产生的不平衡问题,提高对安全帽检测的准确率。

4 仿真实验与结果分析

4.1 实验环境配置

实验环境配置:CPU Inter Xeon E5-2 620v4@ 2.10 GHz, GPU Tesla T4 (16GB), Cuda 11.0, PyTorch 1.10.0, Python 3.6。

训练过程部分参数设置:图片大小 640×640,动量设置为 0.937,图片批次为 8,权重衰减设置为 0.0005,训练 50 轮。

4.2 数据集

数据集是从公开数据集 SHWD (Safety Helmet

Wearing-Dataset)中筛选出 5 000 张图片进行实验,其中,训练集 4 000 张,测试集 1 000 张。数据集的标注,用的是开源的 labeling 工具进行标注,所储存的格式是 PASCAL VOC 格式,标注后是 xml 文件的格式,所标注的文件名和原图片文件名需要保持一致,以便于后续的实验。

SHWD 数据集中的图片存在不是施工场景中的图片,进行数据的筛选。把不同分辨率的图片格式化为统一大小的图片尺寸,方便实验地读取,也方便后面实验的测试。

使用 labeling 工具标注时,存储图片的文件格式是 xml 格式文件,但实验训练的时候采用的是 txt 文件格式。在进行实验之前把需要进行输入网络的数据集文件格式转换为 txt 格式。

4.3 评价指标

实验中使用精确率(Precision, P)、召回率(Recall, R)和平均精度均值(Mean Average Precision, f_{mAP})作为实验的一个评价指标,具体公式如下:

$$P = \frac{TP}{FP+TP} \quad (5)$$

$$R = \frac{TP}{FP+FN} \quad (6)$$

$$f_{AP} = \int_0^1 P(R) dR \quad (7)$$

$$f_{mAP} = \frac{1}{n} \sum f_{AP} \quad (8)$$

其中, TP 是预测的边界框和基础真实框一样的, FP 是预测的边界框和基础真实框不一样的, FN 是预测时没有检测到的图像中的物体数, n 是所分类的个数, 本文分两个类别。

4.4 结果分析

本文将改进的算法与 Faster RCNN 和 YOLOv3 在数据集上进行对比实验,检测结果如表 2 所示。由表 2 可知,本文算法在 mAP 值上有更高的检测精度。改进的 YOLOv5 算法,相比于 Faster RCNN 算法, $mAP@0.5$ 提高了 10.5%;相比于 YOLOv3 算法, $mAP@0.5$ 提高了 12.6%;相比于原来的 YOLOv5 算法, $mAP@0.5$ 提高了 2.2%,效果显著,优于原有的 YOLOv5 算法。

表 2 实验结果对比

Table 2 Comparison of experimental results

算法	$P/\%$	$R/\%$	$mAP@0.5/\%$
Faster RCNN	84.7	77.1	84.2
YOLOv3	85.4	74.0	82.1
YOLOv5	91.3	86.9	92.5
改进算法	93.1	89.3	94.7

实验的最后对训练好的模型进行测试,实验结果如图 6 所示。在第 1 行图片中,原算法中存在把目标 person 误检成 hat 的情况,改进的算法可以正确检测;第 2 行图片中,改进的算法可以检测出漏检的 hat 目标。第 3 行图片中,原算法对于远距离的物体不仅出现了误检还出现了漏检的情况,把远处的路灯误检成 hat,对于近处的戴面罩的工作人员漏检,改进的算法可以正确检测。因此改进的算法对远距离的目标可以检测到,对小目标的漏检和误检都有一定程度的改善,置信度也相对较高,相比于原始 YOLOv5 算法在安全帽检测上具有良好的效果。



图 6 实验结果

Fig. 6 Experimental results

5 结论与展望

(1) 结论。本文在网络中增加了对安全帽注意的同时优化了损失函数,提升了远距离小目标的泛化能力。具体是在原模型的骨干网络 C3 后面添加注意力模块,使得网络将注意力聚焦在含有安全帽的区域;同时,改变损失函数,使用 EIou 损失函数替换原模型的 GIoU 损失函数,优化预测过程中样本不平衡问题,可以使模型更加专注好地锚框,加速了收敛速度。实验结果证明,改进的算法平均精度达到了 94.7%。相较于原模型 YOLOv5 平均检测精度提高了 2.2%;相比于 YOLO v3 模型检测精度提高了 12.6%,减少了复杂场

景中的漏检和误检情况,对于多尺寸目标和远距离的小目标有较好的检测效果。

(2) 展望。改进的算法有效地改善了小尺寸安全帽的漏检和误检。但网络的计算量还是很大的,后续将继续优化本文模型,换成轻量化的网络,保证检测准确率的同时,提升检测速度,更高效地检测安全帽。可以尝试在输出层添加一个更小目标输出层,检测更小的目标。

参考文献(References):

- [1] YAN G, SUN Q, HUANG J, et al. Helmet detection based on deep learning and random forest on UAV for power construction safety[J]. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2021, 25(1): 40-49.
- [2] 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*. Piscataway: IEEE Press, 2014: 580-587.
- [3] GIRSHICK R. Fast R-CNN [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2015: 1440-1448.
- [4] 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(6): 1137-1149.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//*Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2016: 21-37.
- [6] REDMON 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*. Piscataway: IEEE Press, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2017: 6517-6525.
- [8] REDMON J, FARHADI A. Yolov3: an incremental improvement [C]//*Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 89-95.
- [9] 徐守坤, 王雅如, 顾玉宛, 等. 基于改进 Faster RCNN 的安全帽佩戴检测研究[J]. *计算机应用研究*, 2020, 37(3): 901-905.
XU Shou-kun, WANG Ya-ru, GU Yu-wan, et al. Safety helmet wearing detection study based on improved Faster RCNN [J]. *Application Research of Computers*, 2020, 37(3): 901-905.
- [10] 方明, 孙腾腾, 邵桢. 基于改进 YOLOv2 的快速安全帽佩戴情况检测 [J]. *光学 精密工程*, 2019, 27(5): 1196-1205.
FANG Ming, SUN Teng-teng, SHAO Zhen. Fast helmet-wearing-condition detection based on improved YOLOv2 [J]. *Optics and Precision Engineering*, 2019, 27(5): 1196-1205.
- [11] 赵红成, 田秀霞, 杨泽森, 等. 改进 YOLOv3 的复杂施工环境下安全帽佩戴检测算法 [J]. *中国安全科学学报*, 2022, 32(5): 194-200.
ZHAO Hong-cheng, TIAN Xiu-xia, YANG Ze-sen, et al. Safety helmet wearing detection algorithm in complex construction environment based on improved YOLOv3 [J]. *China Safety Science Journal*, 2022, 32(5): 194-200.
- [12] WANG C Y, MARK LIAO H Y, 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*. Piscataway: IEEE Press, 2020: 1571-1580.
- [13] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2018: 8759-8768.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//*Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2018: 3-19.
- [15] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2019: 658-666.
- [16] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.

责任编辑:陈 芳