

单伏顺. 基于改进 YOLOv7 的安全头盔检测算法[J]. 智能计算机与应用, 2024, 14(10): 227-230. DOI: 10.20169/j.issn.2095-2163.241034

# 基于改进 YOLOv7 的安全头盔检测算法

单伏顺

(天津职业技术师范大学 电子工程学院, 天津 300222)

**摘要:** 针对原 YOLOv7 模型在工地作业场所检测安全头盔是否佩戴容易出现漏检误检的问题, 提出了一种改进的 YOLOv7 安全头盔目标检测算法。在原模型的主干中加入 GAM 注意力模块, 提高目标有效特征的利用率, 减少无用响应特征, 达到增强网络提取目标有效特征的能力。实验数据表明, 使用了改进后的模型在自制数据集中, 检测安全头盔的  $mAP$  提升了 0.9%, 召回率提升了 1.3%。该研究对于安全头盔检测的应用提供了一定的参考价值。

**关键词:** 安全头盔; 目标检测; YOLOv7 模型; GAM 注意力

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2024)10-0227-05

## Safety helmet detection algorithm based on improved YOLOv7

SHAN Fushun

(School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

**Abstract:** Aiming at the problem that the original YOLOv7 model is prone to missed detection and misdetection when detecting whether the safety helmet is worn at the construction site, an improved YOLOv7 safety helmet target detection algorithm is proposed. The GAM attention module is added to the backbone of the original model to improve the utilization rate of effective features of the target, reduce the useless response features, and enhance the ability of the network to extract the effective features of the target. Experimental results demonstrate that using the improved model in the homemade dataset, the  $mAP$  of the safety helmet is increased by 0.9%, and the recall rate is increased by 1.3%. This study provides a certain reference value for the application of safety helmet detection.

**Key words:** safety helmet; object detection; YOLOv7 model; GAM attention

## 0 引言

在对于头部具有一定潜在危险的工地场所, 如采矿、钻井森林采伐、供电线路、高层建筑等施工区域, 安全头盔作为个人防护装备的重要组成部分, 能够有效保护工作者头部免受意外伤害。但工人却有可能因为一时疏忽或在佩戴安全帽时感觉不舒适而出现未能佩戴安全帽或佩戴不当的问题, 这种情况极大增加了头部受伤的风险。为了解决这个问题, 本次研究则将着重开展对安全帽佩戴检测技术的探索与开发。这种技术可以利用机器学习、深度学习或计算机视觉等先进科技手段, 实时监测和识别工人或参与者是否正确佩戴安全头盔, 及时发现佩戴不当的情况并采取相应的措施, 以此减少意外伤害事件的发生。本次研究可以为建筑工地、煤矿隧道、

机械工业和冲击强度高的室内常温作业场所<sup>[1]</sup>以及其他工作领域的安全管理提供有效的辅助手段。

随着计算机视觉和深度学习技术的迅速发展, 一系列优秀的目标检测算法相继问世, 如 SSD<sup>[2]</sup> 系列 (R-SSD<sup>[3]</sup>、DSSD<sup>[4]</sup>、FSSD<sup>[5]</sup> 等)、RCNN<sup>[6]</sup> 系列 (Fast-RCNN<sup>[7]</sup>、Faster-RCNN<sup>[8]</sup>、Mask-RCNN<sup>[9]</sup> 等) 以及 YOLO 系列。YOLO 系列算法属于 One-Stage 算法, 是直接基于 anchor 进行特征提取并对预测目标进行分类与定位。这种端到端的技术极大提高了算法的检测速度。近几年来, 很多研究人员都在将机器学习和计算机视觉技术应用在安全帽佩戴的问题上, 都取得了非常不错的效果。杨莉琼等学者<sup>[10]</sup>将机器学习算法应用于安全帽佩戴检测。李明山等学者<sup>[11]</sup>改进 SSD300 与 SSD512 算法。王玲敏等学者<sup>[12]</sup>在引入 CA 坐标注意力机制的同时, 还使用

BiFPN 网络结构替换原特征融合模块,使得 YOLOv5 中平均精度达到了 95.9%。但考虑到作业工地场景复杂、人群密集、各种作业材料多样以及安全头盔种类颜色繁多,极易与头戴普通帽子发生混淆,且因为该种检测关乎人员生命安全,检测容错率越低、就越好。

针对以上问题,本文基于 YOLOv7 算法的思想提出一种改进的检测算法,于复杂背景中检测安全头盔,提高安全头盔的目标检测精度 (AP), 希望为后续在深度学习领域研究安全头盔目标检测的应用中提供一定的参考价值。

## 1 YOLOv7 检测算法

YOLOv7 是原 YOLOv4 和 YOLOv5 团队推出的新 YOLO 模型结构,当图像帧率在 5~160 之间时,目前绝大部分的图像检测模型的检测性能都不如

YOLOv7,特别是在 30 帧/s 以上的实时检测中, YOLOv7 的准确率处于领先地位。为了更好地适应不同情况下的 GPU 环境,官方发布了 YOLOv7、YOLOv7-tiny 和 YOLOv7-W6 三个模型算法,分别对应普通 GPU、边缘 GPU 和云 GPU。这 3 种基本模型的网络深度和结构模型虽有所不同,但模型中的 *depth\_multiple* 和 *width\_multiple* 全都是 1,基本原理也都相同。因为是在普通 GPU 上使用,所以本次实验使用 YOLOv7 模型。

YOLOv7 与 YOLOv5 在网络结构上非常相像,对各个部分的功能也基本相同,如 backbone 用于提取特征,head 用于预测。但与其不同的是, YOLOv7 将颈部 (neck) 层与 head 层合称为 head 层。因此, YOLOv7 结构模型可以概括为输入 (Input)、骨干网络 (Backbone) 和头部 (Head) 三大组成部分,其结构模型如图 1 所示。

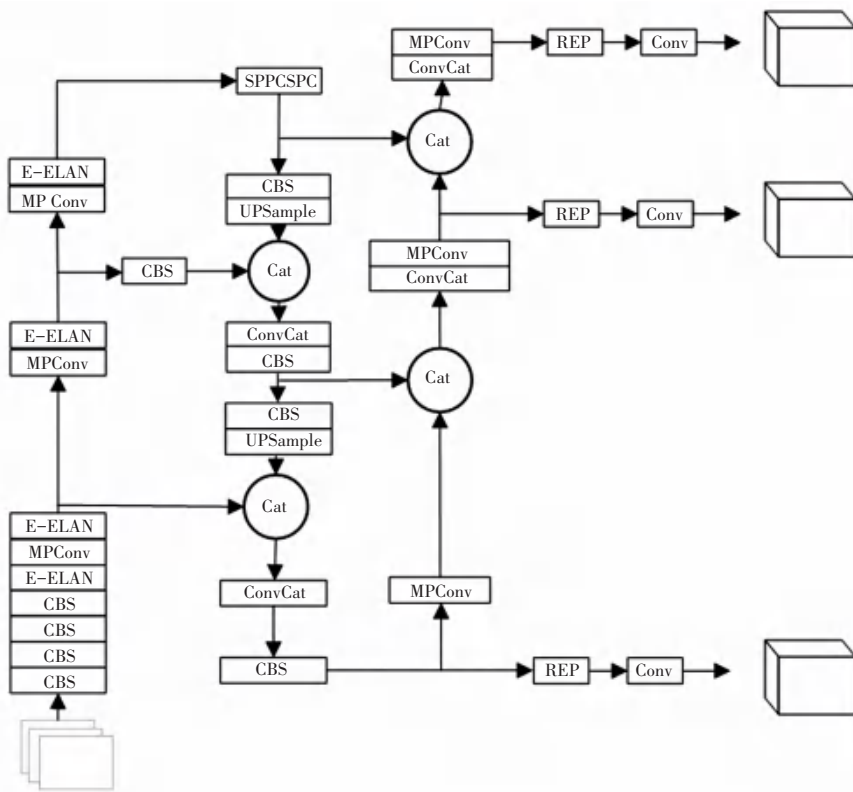


图 1 YOLOv7 模型结构

Fig. 1 YOLOv7 model structure

分析图 1,可以得出其工作流程:首先,图像将被从 Input 端输入进行预处理,将大小调整对齐成  $640 \times 640$  大小的 RGB 图像,随后图像被送入 Backbone 主干中,输入图像会进行特征提取并输出 3 层特征层。在主干部分,获取到的特征层经过 Head 部分中的 Neck 模块进行特征融合处理。最后,再将 Neck 模块融合的特征送入检测头,由此得

到最终输出结果。YOLOv7 的 Backbone 部分概括为 4 个卷积模块,分别是: CBS 卷积、E-ELAN (Extended-ELAN) 卷积、MPConv 卷积和 SPPCSPC 卷积。CBS 是由二维卷积 (Conv2d)、批归一化处理层 (BN 层) 和 SiLU 激活函数三部分组成。E-ELAN 模块是在原 ELAN 基础之上改进的,也称作扩展的 ELAN。E-ELAN 是在不破坏原有梯度路径的基础

之上,用扩大通道数(*expand*)、深浅层特征间的通道进行 *shuffle* (*shuffle*)、深浅层特征做进一步融合 (*merge cardinality*) 方法来实现增强模型学习能力的目的。对于 MPCConv 模块,有 2 个分支,其作用就是进行下采样。在第一条分支中,先进行最大池化 (*MaxPool*) 下采样来扩张当前特征层的感受野,再经过一个  $1 \times 1$  的卷积来改变通道数,以此提高网络泛化性。第二条分支先改变通道数,再经过卷积核来达到下采样的效果。最后把第一条分支和第二条分支的结果加在一起,得到了超过一般下采样的效果。SPPCSPC 模块在第一条分支通过四层的最大池化来获得不同感受野,加强算法对多分辨率图像的适应能力。同时结合第二条分支,不仅使其计算量减少一半,而且精度也得到了提升。在最后的 head 部分的 Neck 模块中,YOLOv7 也采用了与 YOLOv4 和 YOLOv5 一样的 FPN+PAN 网络并多尺度融合上层特征。在检测头部分,YOLOv7 选择使用 IDetect 检测头,RepConv 模块在 RepVGG 构造的基础上引入了结构重参数化的思想,这样一来网络的复杂度就下降了。

## 2 YOLOv7 检测算法的改进

注意力机制本质上就是将人的视觉感知应用到机器模型上,使模型能够提高目标有效特征的利用率,并减少无用响应特征,达到增强网络提取目标有效特征的能力,以使其能够聚焦于局部感兴趣信息。注意力机制的提出最初是在视觉图像领域,后来

Google DeepMind 团队<sup>[13]</sup> 和 Bahdanau 等学者<sup>[14]</sup> 分别将注意力机制使用在 NLP 领域及 RNN 模型上进行图像分类。自此,注意力机制在 CV 相关问题的网络设计上获得了广泛应用。

GAM 注意力采用 CBAM 的顺序通道-空间注意机制并重新设计了子模块。通道注意子模块使用三维排列以此在 3 个维度上保留信息。接着,由于多层感知器 (MLP) 能够建模通道之间的非线性关系,因此使用一个两层的 MLP 放大跨维通道 - 空间依赖性。在空间注意力子模块中,使用 2 个  $7 \times 7$  的卷积来更多地注意到空间信息。2 个  $7 \times 7$  的卷积能够建模  $7 \times 7$  方块中像素点的非线性关系并且在该模块中使用缩减比  $r$  (其值和 BAM 的一样),因此参数可以更多地捕获像素空间之间的关系。考虑到最大池化操作使检测的图片变得模糊,所以删除了最大池化操作来最大程度地保留其特性,所以空间注意力模块的参数数量将有明显增加。为了减少参数数量,此模块将采用 Group 卷积。通过改进,该注意力不仅能在减少模型信息弥散情况出现的同时还能进一步增强全局跨维度的交互作用。GAM 注意力流程如图 2 所示。给定输入特征图  $F_1 \in RC \times H \times W$ ,定义中间状态  $F_2$  和输出  $F_3$ , 其计算公式分别写为:

$$F_2 = M_c(F_1) \otimes F_1 \tag{1}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{2}$$

其中,  $M_c$  和  $M_s$  分别表示通道注意力和空间注意力,“ $\otimes$ ”表示按元素进行乘法操作。

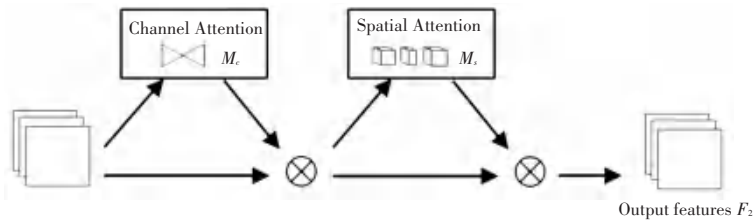


图 2 GAM 注意力流程图

Fig. 2 GAM attention flowchart

本文在主干网的最后一个 E-ELAN 模块中添加双层 GAM 注意力,修改后的 Backbone 网络结构如图 3 所示。通过 GAM 模块后得到一个可以全局

跨维度的交互作用特征图,提升了各个维度之间的联系,增强网络提取有效特征的能力。对有效的通道特征上赋予更高的权重,更多地关注重要信息。

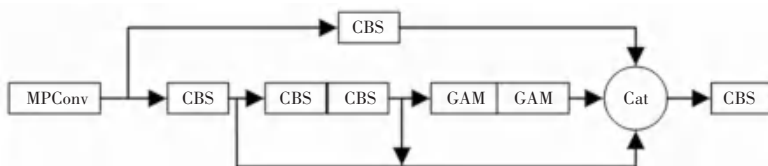


图 3 改进的 Backbone 网络结构

Fig. 3 Improved Backbone network structure

### 3 实验结果与分析

#### 3.1 数据集制作

本次实验使用的数据集采用了 Github 上开源的 SHWD 数据集图片,从中筛选出总计 1 000 张佩戴安全头盔和未佩戴安全头盔两种类别标签的图片。由于原 SHWD 数据集图片标签标注存在漏标



图4 数据集标注

Fig. 4 Dataset labels

#### 3.2 实验环境及评价指标

本次实验采用 Pytorch 学习框架搭建,操作平台是 Windows11,使用双核显卡 Nvidia GeForce RTX 3080 10 GB, CUDA11.1 加速库及 Pytorch1.9.1、Anaconda\_python3.9。

为了客观评价实验改进的性能,本文使用准确率(Precision)、召回率(Recall)和平均精度均值(mAP)三个指标来评估模型效果。对此,研究给出阐释分述如下。

(1) 召回率。表示正确预测为正的占全部实际为正的比例,也称为查全率,推得的数学公式具体如下:

$$R = \frac{TP}{TP + FN} \quad (3)$$

(2) 准确率。表示正确预测为正的占全部预测为正的的比例,也称为查准率,推得的数学公式具体如下:

$$P = \frac{TP}{TP + FP} \quad (4)$$

其中,  $TP$  表示在所有正例中被正确划分的数量,即真正例(True Positives);  $FP$  表示在所有正例中被错误划分的数量,即假正例(False Positives);  $FN$  表示在所有负例中被错误划分的数量,即假反例(False negatives)。

(3)  $mAP$ 。是各类的平均准确率( $AP$ )的均值,推得的数学公式具体如下:

标错等情况,使用 Labeling 图片标注软件自行标注与修改完善,将佩戴安全头盔的位置标注为 helmet,将未佩戴安全头盔的位置标注为 no-helmet,数据集标注案列如图 4 所示。2 类目标组成: 2 838 个 helmet 的标签和 4 460 个 no-helmet 的标签。将 810 张图划为训练集,90 张图划为验证集,100 张图划为测试集。

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (5)$$

其中,  $N$  表示检测物体类别的总数,  $AP_i$  表示第  $i$  个类别的物体的实际准确率。

#### 3.3 实验结果分析

本次实验对目标检测模型共训练了 150 次,原 YOLOv7 模型与添加注意力后的改进模型结果见表 1,各阶段改进的检测精度对比如图 5 所示。图 5 中,横坐标表示训练次数(epoch),纵坐标表示  $IoU$  阈值为 0.5 时的平均精度( $mAP@0.5$ )。

表 1 仿真对比结果

算法	$mAP@0.5$	$mAP@0.5;0.95$	Recall
YOLOv7	95.5	63.0	91.6
YOLOv7+GAM	96.4	63.7	92.9

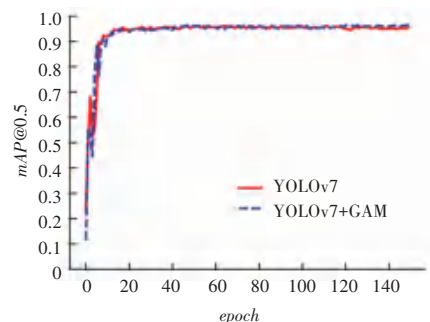


图5  $mAP$  曲线

Fig. 5  $mAP$  curve