

郑凯东, 马莉娜. 基于改进 YOLOv5 模型的安全帽佩戴检测方法研究[J]. 智能计算机与应用, 2024, 14(7): 140-144. DOI: 10.20169/j.issn.2095-2163.240721

基于改进 YOLOv5 模型的安全帽佩戴检测方法研究

郑凯东, 马莉娜

(西安石油大学 计算机学院, 西安 710065)

摘要: 建筑施工现场环境复杂, 现有的安全帽检测算法很难在这种复杂环境下保持准确性和实时性。针对此问题, 本文提出了一种基于改进 YOLOv5s 的安全帽佩戴检测模型 (YOLOv5s-REG)。首先, 基于 Res2Net 网络提出了 C3_Res2Block 模块, 替代骨干网络中的 C3 模块, 捕捉更丰富的特征信息; 其次, 在骨干网络中引入 ECA 注意力机制模块, 增强模型对通道特征的关注度; 最后, 引入 GSConv 模块替换颈部网络的 Conv 模块, 减少模型的参数量。实验结果表明, 改进后的模型平均精度均值达到 94.2%, 同时浮点运算次数减少至 14.4 G, 本文模型在满足检测精度需求的同时成功地降低了模型的复杂度, 适用于施工现场算力有限的边缘设备。

关键词: 施工现场; 安全帽; YOLOv5s; Res2Net; ECA; GSConv

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)07-0140-05

Research on safety helmet wearing detection method based on improved YOLOv5 model

ZHENG Kaidong, MA Lina

(School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

Abstract: Due to the complexity of construction sites, existing safety helmet detection algorithms struggle to maintain accuracy and real-time performance in such environments. To address this issue, an improved safety helmet detection model based on YOLOv5s is proposed, named YOLOv5s-REG. Firstly, a C3_Res2Block module based on the Res2Net network is introduced to replace the C3 module in the backbone network, capturing richer feature information. Secondly, an ECA attention mechanism module is incorporated into the backbone network to enhance the model's focus on channel features. Finally, a GSConv module is introduced to replace the Conv module in the neck network, reducing the model's parameter count. Experimental results show that the improved model achieves an average precision mean of 94.2%, while reducing the floating-point operation count to 14.4 G. This indicates that the proposed model successfully reduces model complexity while meeting the requirements of detection accuracy, making it suitable for edge devices with limited computational resources at construction sites.

Key words: Construction site; Safety helmet; YOLOv5s; Res2Net; ECA; GSConv

0 引言

据相关统计, 建筑业施工现场安全事件中, 有 85% 的头部受伤事件是因为未佩戴安全帽的行为引起的^[1]。为了保障施工人员的安全, 要求工人必须佩戴安全帽。目前, 安全帽佩戴的管理都依赖于现场的人工判断, 需要投入大量的人力, 效率低下^[2]。随着高分辨率监控系统的出现, 可对施工现场远程动态监控, 但远程监控仍需要大量人员时时关注视

频画面, 可能因为操作员的视觉疲劳导致目标的漏检和误检。近年来, 计算机视觉领域与图像智能算法发展迅速, 为安全帽佩戴自动检测提供了可能性^[3]。经典的目标检测模型有 Region-CNN (R-CNN)、Fast R-CNN、Faster R-CNN、YOLO (You Only Look Once)、SSD (Single Shot multiBox Detector) 和 CenterNet。

在这一研究领域中, 各种先进的模型和解决方案不断涌现, 旨在提高安全帽检测的准确性和效率。

作者简介: 马莉娜 (1999-), 女, 硕士研究生, 主要研究方向: 计算机视觉, 深度学习。

通讯作者: 郑凯东 (1964-), 男, 硕士, 副教授, 硕士生导师, 主要研究方向: 图形学与虚拟现实, 深度学习与计算机视觉, 程序设计。Email:

kdzheng@xsyu.edu.cn

收稿日期: 2023-12-24

哈尔滨工业大学主办 ◆ 专题设计与应用

Huang 等^[4]提出了一种基于改进YOLOv3的模型,对预测的锚框进行像素特征统计,并结合权重系数,输出每个预测的锚框区域中佩戴安全帽的置信水平,虽然该模型在训练速度和特征增强方面取得了巨大的进步,但在一些复杂的施工现场,模型的鲁棒性还需要提升;Deng 等^[5]提出了一种改进的基于YOLOv4的模型,采用了多尺度的训练策略,旨在增强模型在不同检测尺度下的性能,但该模型在检测特征不明显的目标时仍需要进一步优化;Xu 等^[6]提出了一种引入挤压激励模块的YOLOv5模型,该模型不仅能够获取图像通道的权重,还能精确分离图像的前景和背景,该模型虽然提升了检测性能但增加了模型的参数量,使得该模型难以部署在边缘低算力终端上;邓开发等^[7]结合Faster R-CNN模型和深度特征流算法,实现对施工现场监控视频中安全帽的检测;刘瑞昊等^[8]提出了一种基于YOLOv3的改进多尺度特征融合模型,提高了工业现场小目标检测的准确率;吕宗喆等^[9]提出了一种面向安全帽的YOLOv5小目标检测模型,通过优化损失函数和引入切片辅助推理技术,在密集场景下显著提高了检测精度和置信度;Fang 等^[10]在YOLOv2模型的基础上,将密集网络与特征提取网络相融合,同时采用轻量级的MobileNet模型,以提升检测精度和速度;梅国新等^[11]提出了一种基于EfficientDet的安全帽检测模型,该模型具备小参数量和快速推理等优势;金源等^[12]基于YOLOv5s模型,通过结合DenseBlock和Res2Net网络的核心设计思想,实现了多级分类目标检测;Chen 等^[13]将K-means++算法与Faster R-CNN结合,提高了小尺寸头盔检测准确率,但检测速度较慢,无法满足实时需求;Rao 等^[14]通过在YOLOv3-Tiny中集成改进的SPP模块以实现实时安全帽佩戴检测,但其精度不足以满足铁路安全标准。

1 YOLOv5 模型

YOLOv5是由Ultralytics提出的实时目标检测模型。为了适应各种算力资源,官方给出了几种不同参数量的模型。与YOLO历史版本相比,YOLOv5在网络架构、数据增强和模型训练等方面进行了改进和优化,从而在多个目标检测竞赛中成绩表现卓越。YOLOv5被广泛应用到物体检测、行人检测、车辆检测等任务上。

YOLOv5的网络结构包含4个部分,即输入层(Input)、骨干网络(Backbone)、颈部网络(Neck)和

输出端(Head)^[15]。YOLOv5结构如图1所示。YOLOv5使用Leaky ReLU作为激活函数,通过引入小的负斜率解决了ReLU在负数部分输出为零的问题;使用综合损失函数,组合目标位置的均方误差、目标存在的二值交叉熵以及类别概率的交叉熵。

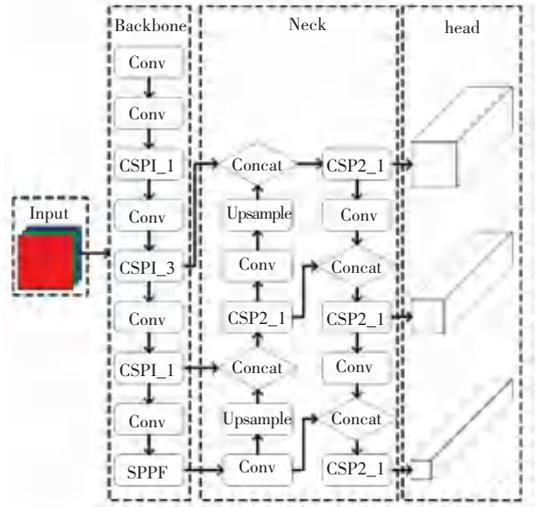


图1 YOLOv5结构

Fig. 1 YOLOv5 architecture

2 模型改进与优化

2.1 多尺度残差网络

残差网络(Residual Network, ResNet)是He等^[16]于2015年提出的一种深度残差网络。其设计目的是解决深度神经网络训练过程中网络退化和梯度消失/爆炸的问题。传统观点认为,在深度神经网络中,网络层数越深,其性能和表达能力就越好。但在实际训练中,随着网络层数的增加,准确率会饱和并逐渐下降,导致退化问题,这是由于深层网络难以优化梯度消失或爆炸问题。残差连接的核心思想是将输入特征与输出特征直接相加,即通过跳跃连接将输入直接传递给后续层,使得梯度更容易传播,有效解决梯度消失或爆炸问题,ResNet网络的Bottleneck模块的结构如图2所示。然而,ResNet在网络结构中并未明确考虑如何融合不同尺度的特征信息。每个卷积层只能捕捉一定范围内的特征,导致在处理多尺度目标时存在一定困难,影响网络的表达能力和泛化能力。

Res2Net(Residual-Residual Network)是一种基于ResNet改进的网络,通过将卷积分解成多个子模块并连接起来,扩展感受野的范围,从而捕捉到更丰富的多尺度特征信息^[17]。在一些低算力设备上进行推理任务时,Res2Net相较于其他更深的模型具

有一定的优势,可以在不增加网络深度的情况下,提升网络的表达能力和性能,在保持计算资源消耗较低的同时提供更好的效果。本文在骨干网络的 C3 模块中引入 Res2Net 网络的 Bottleneck 模块。Res2Net 网络的 Bottleneck 模块的结构图如图 3 所示。

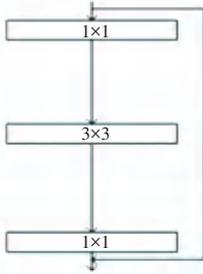


图 2 ResNet 网络的 Bottleneck 模块
Fig. 2 Bottleneck module of ResNet

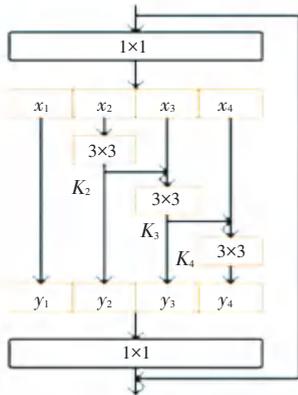


图 3 Res2Net 网络的 Bottleneck 模块
Fig. 3 Bottleneck module of Res2Net

假设输入特征图表示为 $X \in R^{H \times W \times C}$, 其中 H 是高度, W 是宽度, C 是通道数。首先, 将输入特征图通过 1×1 卷积层处理, 将其按通道数均匀分成 s 个子集, 用 x_i 表示, 其中 $i \in \{1, 2, \dots, s\}$ 。在这里, 除了第一个子集 x_1 之外, 其余的子集都需要经过一个 3×3 卷积核, 记为 $K_i(\cdot)$, 用 y_i 表示 $K_i(\cdot)$ 的输出。具体而言, 对于第 i 个子集, 当 i 为 $2 \sim s$ 之间时, 将 x_i 与前一个子集的输出 y_{i-1} 相加, 然后将结果送入卷积核 $K_i(\cdot)$ 进行处理。这一过程可以用公式(1)表示:

$$y_i = \begin{cases} x_i, & i = 1 \\ K_i(x_i), & i = 2 \\ K_i(x_i + y_{i-1}), & 2 < i < s \end{cases} \quad (1)$$

2.2 高效的通道注意力机制

ECA (Efficient Channel Attention) 注意力机制是一种在 SE (Squeeze-and-Excitation) 注意力机制基

础上改进的通道注意力机制^[18]。SE 注意力机制是先对输入特征图进行通道压缩, 以降低维度。通道压缩通过全局信息的整合来提取特征的重要性, 但在这个过程中, 有可能会丢失一些细粒度的通道间关联信息。基于这一理念, ECA 通过避免降维的策略, 移除了原有 SE 注意力机制中的全连接层, 并利用了 1×1 卷积核, 实现了对局部跨通道之间的高效交互, 更有效地提取通道之间的依赖关系。ECA 注意力机制在不增加过多计算资源的情况下, 增强了通道间的交互和特征表示能力。相较于传统的注意力机制, 其具有更高的计算效率和更少的参数量。因此, 本文在骨干网络的 SPPF (SPP-Fast) 模块前引入 ECA 注意力机制, ECA 注意力机制的结构如图 4 所示。

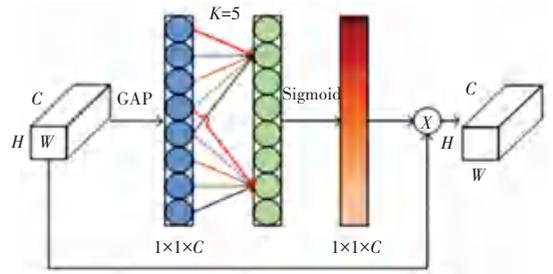


图 4 ECA 注意力机制的结构
Fig. 4 ECA attention mechanism structure

实现原理如下:

假设输入特征图表示为 $X \in R^{H \times W \times C}$, 其中 H 是高度, W 是宽度, C 是通道数。

全局平均池化 (Global Average Pooling) 操作通过对每个通道进行全局平均池化, 计算每个通道维度的平均值, 表示为 $Avg(X) \in R^C$;

采用 1×1 的卷积核进行变换, 表示为 $W_1 \in R^{C \times r}$, 其中 r 是缩放因子 (通常小于 1 的数值), 得到 $Conv_{1 \times 1}(X) \in R^{H \times W \times r}$; 对卷积输出进行 Sigmoid 操作, 得到通道维度的权重, 表示为 $Sigmoid(Conv_{1 \times 1}(X)) \in R^{H \times W \times r}$; 将输入特征图 X 与得到的权重相乘, 表示为 $ECA(X) = X \odot Sigmoid(Conv_{1 \times 1}(X))$, 其中 \odot 表示逐元素相乘。

2.3 广义稀疏卷积

GSCConv (Generalized-Sparse Convolution), 由 Conv 模块、DWConv 模块、Concat 模块和通道混洗模块组成^[19-20]。GSCConv 卷积层采用了深度可分离卷积结合标准卷积的方式, 旨在减轻计算负担的同时, 提升模型对复杂特征的抽取和融合能力, 更好地平

衡模型的性能和效率。本文将 GSConv 模块嵌入到颈部网络, 替换原有的标准卷积 Conv, 从而进一步降低模型的参数量。GSConv 模块的结构如图 5 所示。

假设从上一层所获取到的特征图通道数为 C_1 , 期望处理后特征图通道数为 C_2 。使用 Conv 模块进

行标准卷积处理后获得一个通道数为 $C_2/2$ 的特征图; 使用 DWConv 模块对其进行深度可分离卷积操作, 得到一个新的通道数为 $C_2/2$ 的特征图; 将得到的特征图进行拼接并使用通道混洗算子对拼接特征图进行处理, 输出具有目标通道数 C_2 的最终特征图。

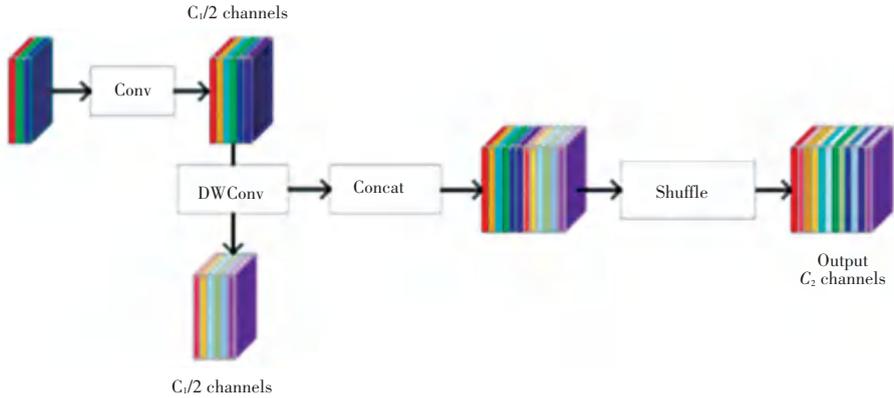


图 5 GSConv 模块结构

Fig. 5 GSConv module structure

3 实验及结果分析

3.1 实验环境与参数

本文实验环境: 操作系统为 Linux, CPU 为 Intel (R) Xeon (R) Silver 4214R CPU@ 2.40 GHz, GPU 为 NVIDIA GeForce RTX 3080 Ti, Pytorch 版本为 2.0.0, cudnn 为 11.8, Python 为 3.8。在模型训练时, 初始学习率为 0.01, 采用线性调整策略, 最终学习率为 0.000 1, 采用了 epoch 为 3 的预热学习率的策略, 动量参数为 0.937, 权重衰减为 0.000 5, batch size 设置为 16, 训练轮数为 150 epochs, 使用 SGD 优化器对网络进行优化。

3.2 实验数据集

本文所采用的数据集在开源数据集 (Safety Helmet Wearing-Dataset, SHWD) 的基础上进行筛选和补充而得。由于 SHWD 数据集中存在大量非施工场景下的数据, 因此额外收集了一部分施工现场的图片, 使用图像标注工具对这些图像进行标注, 标注标签信息存储为 txt 数据格式。整合后的数据集共有 6 000 张图片, 图像里标注的目标分别是: person (行人); hat (佩戴安全帽的头部); head (未佩戴安全帽的头部)。其中 person 目标有 19 321 个标签, hat 目标有 5 602 个标签, head 目标有 14 990 个标签。数据集按 8 : 2 的比例划分为训练集和验证集, 部分数据集图片如图 6 所示。



图 6 部分数据集图片

Fig. 6 Partial dataset showcase

3.3 评价指标

本文采用精确率、召回率、平均精度均值、浮点运算量以及模型大小作为评价指标。

3.4 消融实验

为了验证本文对 YOLOv5s 模型改进的有效性, 在确保各实验参数相同的前提下进行了消融实验。在这一系列实验中, 针对不同的改进方法进行比较, 旨在深入探讨这些改进对模型性能的影响, 消融实验的结果见表 1, 表中使用“√”表示实验中使用了该方法来改进模型, 而“×”表示没有使用该方法。由表 1 可知, 本文算法在 mAP 值上有更高的检测精度, 同时降低计算量, 优于原有的 YOLOv5s 模型, 更好地适应施工现场环境。

表1 消融实验结果对比
Table 1 Comparison of ablation experiment results

模型	Res2Net	ECA	GSCConv	精确率	召回率/%	平均精度 均值(50)	平均精度 均值(50-95)	浮点 运算量/G	模型大小/ MB
YOLOv5s	×	×	×	93.437	89.224	93.755	63.410	15.800	14.400
YOLOv5s-R	√	×	×	93.496	89.447	94.047	64.115	14.900	13.900
YOLOv5s-E	×	√	×	94.149	88.817	94.077	63.689	15.800	14.400
YOLOv5s-G	×	×	√	93.304	89.511	93.827	63.807	15.200	13.500
YOLOv5s-REG	√	√	√	94.453	89.244	94.227	64.466	14.400	13.000

4 结束语

在真实的施工场景中进行安全帽佩戴检测面临两个主要挑战:由于拍摄视角问题,监控到的目标通常较小,导致目标特征与背景混合,难以准确区分;不同的施工场景和天气条件会降低目标的可见性,可能导致目标被遮挡。本文在 YOLOv5s 模型的基础上进行了改进,引入了 Res2Net 和 ECA 模块来优化检测网络,以提高模型的检测性能;同时,采用 GSCConv 卷积层来进一步减少模型的参数量,以提升模型的检测效率和速度。

参考文献

[1] USUKHBAYAR R, CHOI J. Critical safety factors influencing on the safety performance of construction projects in Mongolia[J]. Journal of Asian Architecture and Building Engineering, 2020, 19(6):600-612.

[2] 赵羽. 基于轻量级神经网络的智慧工地视频监控系统研究[D]. 南京:南京邮电大学,2021.

[3] CHEN S, TANG W, JI T, et al. Detection of safety helmet wearing based on improved Faster R-CNN[C]// Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, UK:IEEE, 2020: 1-7.

[4] HUANG L, FU Q, HE M, et al. Detection algorithm of safety helmet wearing based on deep learning[J]. Concurrency and Computation: Practice and Experience, 2021, 33(13): e6234.

[5] DENG Benyang, LEI Xiaochun, YE Miao. Safety helmet detection method based on YOLOv4[C]// Proceedings of 2020 16th International Conference on Computational Intelligence and Security (CIS). IEEE, 2020: 155-158.

[6] XU Z P, ZHANG Y, CHENG J, et al. Safety helmet wearing detection based on YOLOv5 of attention mechanism[J]. Journal of Physics: Conference Series. IOP Publishing, 2022, 2213(1): 012038.

[7] 邓开发,邹振宇.基于深度学习的安全帽佩戴检测实现与分析

[J]. 计算机时代,2020(7):12-15,21.

[8] 刘瑞昊,于振中,孙强.改进多尺度特征融合的工业现场目标检测算法[J].机械与电子,2022,40(11):40-45.

[9] 吕宗喆,徐慧,杨骁,等.面向小目标的 YOLOv5 安全帽检测算法[J].计算机应用,2023,43(6):1943-1949.

[10] FANG Qi, LI Heng, LUO Xiaochun, et al. Detecting nonhardhat-use by a deep learning method from far-field surveillance videos[J]. Automation in Construction,2018,85(1):1-9.

[11] 梅国新,姚庆华,陈瑶,等.一种边缘环境下基于 EfficientDet 的施工人员安全帽检测方法[J].数字通信世界,2020(9):77-78.

[12] 金源,张长鲁.扩展 YOLOv5 安全帽多级目标分类检测[J].计算机系统应用,2023,32(2):139-149.

[13] CHEN S, TANG W, JI T, et al. Detection of safety helmet wearing based on improved faster R-CNN[C]// Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-7.

[14] RAO C, HE X N, ZHENG Z L, et al. Multi-scale safety helmet detection based on SAS-YOLOv3-tiny[J]. Applied Sciences, 2021,11(8):3652.

[15] 谢轩,康丽.融合 GhostNet 和 Yolov5 的遥感图像目标检测[J].东莞理工学院学报,2023,30(3):70-77.

[16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[17] GAO S H, CHENG M M, ZHAO K, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.

[18] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 11534-11542.

[19] LI H, LI J, WEI H, et al. Slim-neck by GSCConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022.

[20] 张政超.改进 YOLOv5 的轻量级带钢表面缺陷检测[J].计算机系统应用,2023,32(6):278-285.