

DOI:10.19651/j.cnki.emt.2209425

基于改进YOLOX的安全帽佩戴实时检测*

丁田 陈向阳 周强 肖浩樑

(1. 武汉工程大学计算机科学与工程学院 武汉 430205; 2. 武汉工程大学智能机器人湖北省重点实验室 武汉 430205)

摘要: 在建筑行业中,因未佩戴安全帽而导致的安全事故占比较大。针对安全帽检测中存在的干扰性强、小目标准确率低等问题,提出了一种基于YOLOX的改进算法。首先,在加强特征提取网络中加入ECA-Net注意力机制,进行跨通道交互,根据生成的对应通道权重值,抑制干扰信息,加强模型对目标特征的关注度,再将重校准后的特征图进行更深度地特征融合,提高目标特征的表达能。其次,使用CIoU来计算损失,将两框中心点距离和长宽比考虑进惩罚项,不断调整更新损失函数,加快模型收敛速度。最后,构建了一个真实施工场景下的小目标安全帽数据集。实验结果表明,改进后的算法mAP达91.7%,比原YOLOX算高出1.2%,对已佩戴安全帽的工人检测平均精度达93.9%,对未佩戴安全帽的检测平均精度达89.5%,检测速度达到71.9帧/s,保证安全帽佩戴情况实时检测的同时有较高准确率。

关键词: YOLOX;注意力机制;安全帽检测;深度学习

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Real-time detection of helmet wearing based on improved YOLOX

Ding Tian Chen Xiangyang Zhou Qiang Xiao Haoliang

(1. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China;

2. Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China)

Abstract: In the construction industry, safety accidents caused by not wearing helmets account for a relatively large proportion. Aiming at the problems of strong interference and low accuracy of small targets in helmet detection, an improved algorithm based on YOLOX is proposed. Firstly, an ECA-Net attention mechanism is added to the enhanced feature extraction network to carry out cross-channel interaction, suppress the interference information according to the corresponding channel weight value generated, strengthen the model's attention to the target feature, and then fuse the recalibrated feature map more deeply to improve the expression ability of the target feature. Secondly, the CIoU is used to calculate the loss, the distance between the two boxes of center points and the aspect ratio are considered into the penalty term, and the loss function is constantly adjusted and updated to accelerate the model convergence speed. Finally, a small target helmet dataset in a real construction scenario is constructed. Experimental results show that the improved algorithm mAP reaches 91.7%, which is 1.2% higher than the original YOLOX calculation, the average accuracy of the detection of workers who have worn helmets reaches 93.9%, the average accuracy of detection of those who have not worn helmets reaches 89.5%, and the detection speed reaches 71.9 frames/s, which ensures that the real-time detection of helmet wearing has a high accuracy rate.

Keywords: YOLOX;attention mechanism;safety helmet detection;deep learning

0 引言

在施工作业过程中,不时有重物砸落或撞击施工人员的意外情况,施工人员的头部是防护的重点。在建筑行业中佩戴安全帽是一项基本的规章制度,但由于部分施工人

员安全意识薄弱、监管力度不够,传统的人工监督方法不能满足施工安全管理的要求^[1]。因此,有必要采用自动监测的方法^[2]对施工现场佩戴安全帽的情况进行标注和预警。

目前,国内外学者对安全帽等小目标检测做了一系列

收稿日期:2022-03-28

* 基金项目:武汉工程大学教育创新基金(CX2021273)项目资助

的研究。Shrestha 等^[3]使用类似 Haar 的特征检测人脸,再利用边缘检测算法提取安全帽轮廓特征,但是当工人面部朝下时,没有正面图像,该方法就不能检测到人脸和安全帽,而且边缘检测算法在安全帽与背景对比度不高时,提取效果不佳。Wu 等^[4]提出混合描述子提取安全帽特征的方法,用分层支持向量机进行分类,避免了前者出现的问题,但没有考虑到小目标安全帽在网络中丢失细节信息的情况。孙国栋等^[5]通过在双阶段目标检测算法 Faster R-CNN 中添加自注意力层以提取目标的多尺度全局信息,并通过改进锚框来加强训练小目标,强化了网络对小尺度目标的表达能力,但是检测速度偏低。孙备等^[6]针对水上小目标对 SSD 算法进行改进,对浅层网络进行多尺度卷积融合,并添加空间注意力机制,保护了弱纹理目标的特征,与原算法相比有较大提升,但网络变得复杂,冗余参数增多。陈国平等^[7]使用轻量化模型 YOLOv3-Tiny,模型参数量较少,采用增加网络层数、添加注意力机制到特征金字塔中的方法来减少噪音等冗余信息对结果的影响。许凯等^[8]在 YOLOv3 中增加特征图,使用 K-means 聚类算法并加入 Focal Loss 平衡正负样本来改善检测效果,但是检测速度有所下降,不能较好满足实时检测的要求。解尧婷等^[9]在 YOLOv4 的基础上,减少特征层的使用并重新聚类,提高

了对小目标的检测速度。

现有的主流目标检测算法大都是基于锚框的,对小目标的检测有一定困难,且会带来较多计算成本,检测速度较慢。当目标背景复杂、像素较少时,有漏检、错检的情况。因此,采用不使用先验框的 YOLOX^[10]模型进行安全帽佩戴检测,并在网络中添加注意力机制,引导模型更加关注目标特征,对损失函数进行优化,提高模型的定位能力。实验结果表明:改进后的算法对安全帽的检测有较高的检测速度,在检测小目标安全帽的准确率上有较好表现,可以达到实时检测的效果。

1 YOLOX 算法

在 YOLO 系列^[11-15]的基础上,Megvii 公司于 2021 年 8 月提出 YOLOX 网络模型,YOLOX 中最小结构的 YOLOX-Nano 参数仅 0.91 M,约为 YOLOv5n 的一半,推理速度最快可达 9.8 ms。图像输入后,首先通过 CSPDarknet 主干特征提取网络,通过卷积操作进行特征提取,再经过加强特征提取网络,将 3 个有效特征层进行融合,最后通过分类器与回归器对特征点进行判断。模型由主干网络(Backbone),颈部(Neck)和头部(Decoupled Head)组成,如图 1 所示。

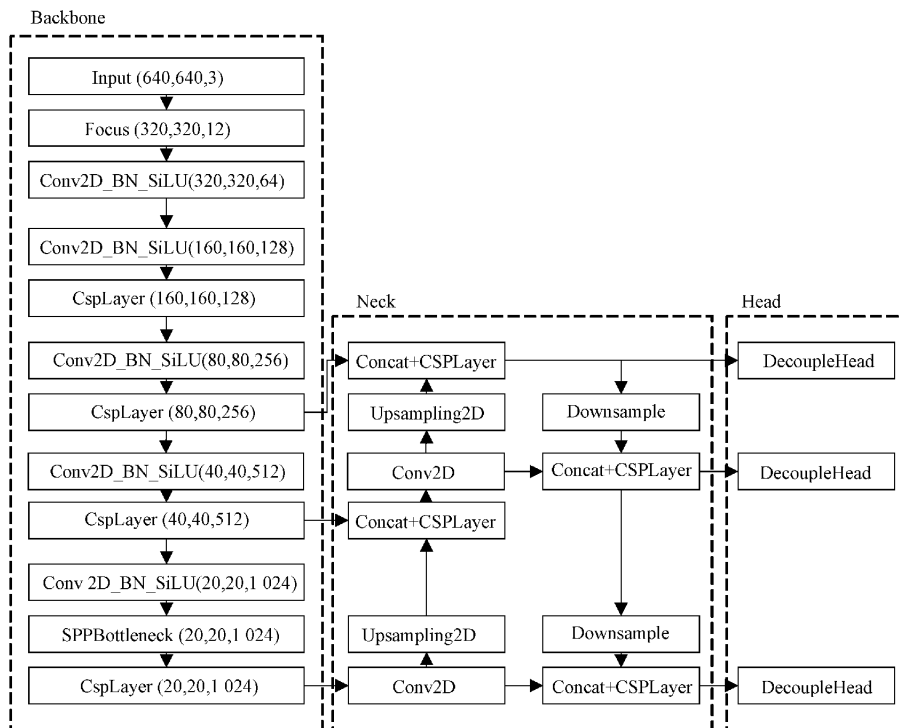


图 1 YOLOX 网络结构

1.1 网络结构

图片首先转换为 RGB 图片,被添加上灰条,即在图像周围进行填充,保证其在调整大小的过程中,不发生失真,再进行归一化处理。数据增强使用了 mosaic 方法。经过

这些处理后,图像进入主干网络。主干网络 SCPDarknet 包含了残差网络 Residual、CSPnet^[16]网络结构、Focus 网络结构和 SPP 结构,并且使用 SiLU^[17]激活函数。残差网络是由 1×1 和 3×3 的卷积构成。CSPnet 结构将残差块的

堆叠分开成两部分处理。Focus 网络结构每隔一个像素拿到一个值,获得四个独立的特征层,将 4 个独立的特征层进行堆叠,输入通道扩充了 4 倍,结构如图 2 所示。

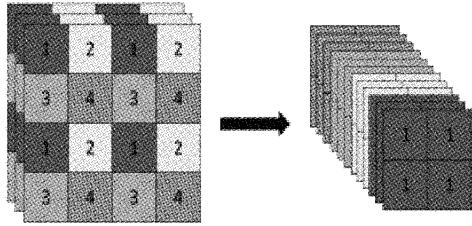


图 2 Focus 结构

SPP 结构通过不同尺寸的池化核对特征图进行池化,再将结果合并,提高网络的感受野。在 YOLOv4 中,SPP 用在 FPN 里,在 YOLOX 中,SPP 模块被用在主干特征提取网络中。SiLU 是基于 Sigmoid 和 ReLU 优化的,具备 Sigmoid 类似的特性:无上界有下界、平滑、非单调。

经过主干网络后,输出 3 个有效特征层。利用输出的特征层构建 FPN 特征金字塔,进行特征融合。

因为分类和定位所侧重的内容不同,YOLOX 在头部将分类和回归分别实现,预测时再整合到一起,如图 3 所示。

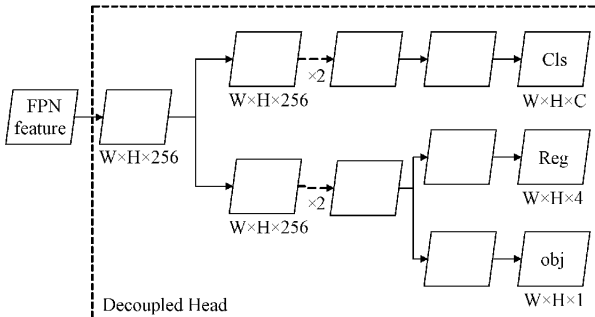


图 3 解耦头结构

1.2 损失函数

损失函数是用来判断神经网络的计算结果与真实值直接的差值,指引下一步的训练走向。YOLOX 在 obj 分支和 cls 分支使用二元交叉熵损失函数(BCELoss),reg 分支使用 IoU Loss。

$$L_{cls} = - \sum_{i=1}^n (t_i \log(p_i) + (1 - t_i) \log(1 - p_i)) \quad (1)$$

$$L_{reg} = -\log(IoU(B_{gt}, B_{pred})) \quad (2)$$

式(1)中 t_i 表示第 i 个样本所属类别, p_i 代表所属类别的概率。

2 改进的 YOLOX 算法

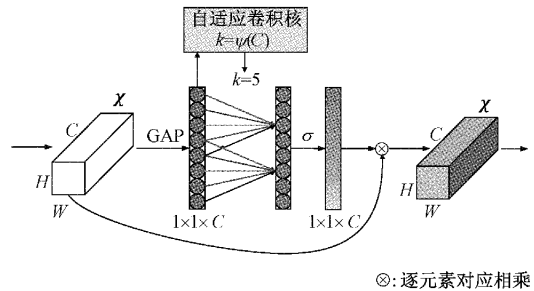
YOLOX 采用无锚框机制,利用 simOTA 方法匹配正样本,检测速度有一定提升。通过主干特征提取网络后,会输出 3 种兼顾大中小尺度物体的特征图,具有一定检测精度。但是对场地复杂、背景多变、像素占比少的安全帽

检测任务,表现不佳。本文在加强特征提取网络中加入高效通道注意力模块(efficient channel attention for deep convolutional neural networks, ECA-Net)^[18],增强模型对目标信息的把控能力,对损失函数进行优化,整体提升模型检测效果。

2.1 注意力机制的引入

在施工现场检测的画面中,安全帽和人体头部占整个图像的比例较小,在往网络深层进行传递时,可用的特征信息丢失严重,会出现无法定位导致漏检、错检的情况。为了提高对此类小目标的检测精度,需要在基本的特征金字塔结构中加入注意力机制,来保留更多细节信息。

ECA-Net 将通道与注意力权重直接对应,通过分组卷积实现跨通道交互,根据注意力权重大小对特征信息进行增强或抑制,提高模型对有用信息的利用程度。图像通过主干特征提取网络后,会生成 3 种不同尺度的特征图。ECA-Net 首先将特征图进行全局平均池化操作,通过核大小为 k 的卷积进行局部跨通道交互,充分利用通道信息,经过 sigmoid 激活函数获取每个通道的权重,再对原特征图进行重校准,可以削弱或者加强某个特定区域的特征,对小目标检测有一定帮助。如图 4 所示。



⊗: 逐元素对应相乘

图 4 ECA-Net 结构图

在通道神经元中,跨通道交互的带状卷积如下:

$$\begin{bmatrix} w^{1,1} & \dots & w^{1,k} & 0 & 0 & \dots & \dots & 0 \\ 0 & w^{2,2} & \dots & w^{2,k+1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & w^{C,C-k+1} & \dots & w^{C,C} \end{bmatrix}$$

k 为卷积核大小, C 为通道数。卷积核与通道数映射函数为:

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (3)$$

其中, $\gamma = 2, b = 1, k$ 取临近奇数结果。

带状卷积中,每行元素较上一行元素向后平移一位,在进行通道神经元的计算时,既实现了通道间的交互,也减少了参数量。卷积计算过程简化如图 5 所示。

在 3 个不同尺度的特征层上添加注意力机制,使网络更精准地抓取大中小目标信息,在特征金字塔上采样部分也添加注意力机制,修改后网络结构如图 6 所示。

浅层特征图有较多细节信息,感受野小,适合检测小

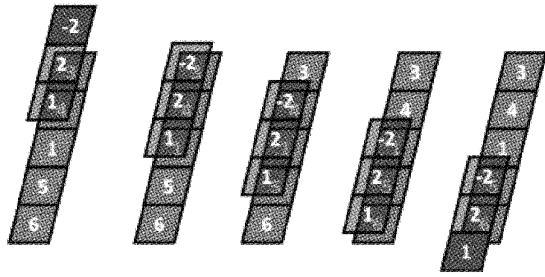


图 5 计算过程简化图

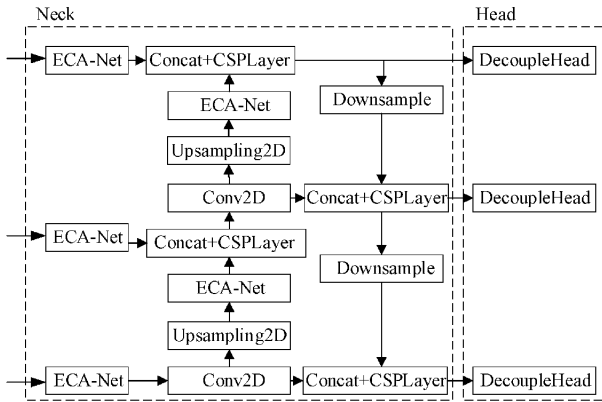


图 6 添加 ECA-Net 网络结构图

目标,但是语义信息不足,易受到干扰;在网络加深过程中,深层特征图感受野大,语义丰富,但是细节信息丢失严重。特征金字塔上采样部分经过注意力模块后,强化有用信息,抑制无关信息;将浅层特征和深层特征进行更深度地融合,增强网络对目标的感知能力。

2.2 损失函数的改进

对于安全帽中尺度小、像素少的部分目标,边框回归精度较低、模型收敛速度较慢。边框回归损失函数直接影响目标的定位效果。

YOLOX 在位置回归中使用 IoU 作为判断真实框与预测框重合程度的指标,来反馈预测框的检测效果,公式如下:

$$IoU = \frac{A \cap B}{A \cup B} \quad (4)$$

当预测框与真实框不相交的时候,IoU 结果为 0,损失值也为 0,不能进行下一步的训练,网络权值无法更新,模型不能收敛到全局最优。且存在不同重叠方式下,IoU 值相同的情况,图 7 为其中的一种情况,虚线框为真实框,实线框为预测框。相交形式不具唯一性,无法准确说明两框覆盖情况。

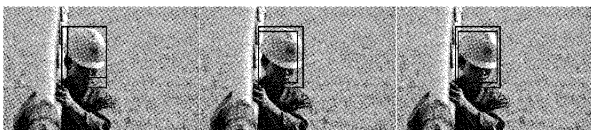


图 7 IoU 相同,重叠方式不同

考虑两框中心点距离这一因素的 DIoU^[19],损失函数为:

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \quad (5)$$

b, b^{gt} 代表预测框和真实框的中心点, $\rho(\cdot)$ 计算欧式距离, c 为包含预测框和真实框最小框的对角线距离如图 8 所示。

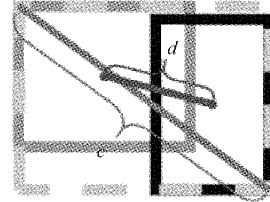


图 8 距离示意图

如果两框中心点重合,而长宽比不同时,如图 9 所示,也无法继续优化。

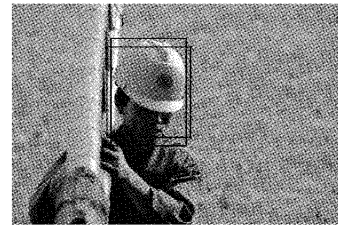


图 9 两框中心点重合

合适的损失函数可以提高模型的回归性能,让预测值逐渐逼近真实值。本文使用 CIoU 来计算损失函数,其中加入了惩罚项 R_{CIoU} ,并且关注预测框与真实框的长宽比,使边框快速稳定回归。CIoU 计算公式如下:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \quad (6)$$

其中惩罚项:

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \quad (7)$$

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \quad (8)$$

$$\nu = \frac{4}{\pi^2} \left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w}{h}\right) \right)^2 \quad (9)$$

α 为权重系数, ν 测量长宽比的一致性, w^{gt}, h^{gt}, w, h 为两框的长和宽。

CIoU 在两框不相交时,梯度不为 0,可以提供移动方向继续训练;当两框中心点重合,两框长宽不同时,可以继续调整预测框。加入惩罚项具有更快收敛速度,关注框的长宽比能更准确反映两框位置关系,使模型的预测效果更好。

3 实验及结果

3.1 实验平台及数据

本文实验对硬件设备有一定的要求,需要使用 GPU

进行加速运算。在服务器上搭建实验所需环境,CPU 为 Intel(R) Xeon(R) Gold 5218 CPU,GPU 为显存 16 G 的 Tesla T4,选用 pytorch 框架,实验所需具体环境如表 1 所示。

表 1 实验环境配置

配置	参数
CPU	Intel(R) Xeon(R) Gold 5218 CPU
显卡	Tesla T4
系统	Ubuntu18.04
Python 版本	3.7
深度学习框架	Pytorch
CUDA	10.0.130

实验在线上收集了现有的一些数据集(SHWD)和网络图像,对图像做部分筛选,线下拍摄少量施工图像,使数据集更具专一性,建立了有 8 500 张图片的安全帽数据集。安全帽数据集分为两类,佩戴安全帽的人(hat)和未佩戴安全帽的人(person)。数据集采用 PASCAL VOC 格式。设置训练集、测试集与验证集比例为 8 : 1 : 1。

3.2 网络训练

在训练过程中,使用迁移学习的思想,先冻结模型的主干,以加快训练速度,迭代次数为 50,迭代批量为 8,学习率初始值设为 0.001,衰减系数 0.000 5,学习动量 0.92;解冻阶段迭代次数为 100,网络结构发生改变,迭代批量为 4,学习率初始值 0.000 1。在 0 到 50 轮时,损失值下降并逐渐缓慢;50 轮后,损失值迅速下降,在 90 轮之后损失值趋于稳定,如图 10 所示。图中 train loss 与 val loss 下降趋势较为同步,train loss 平稳地位于 val loss 上方,是较为理想的状态,网络得到良好的训练。

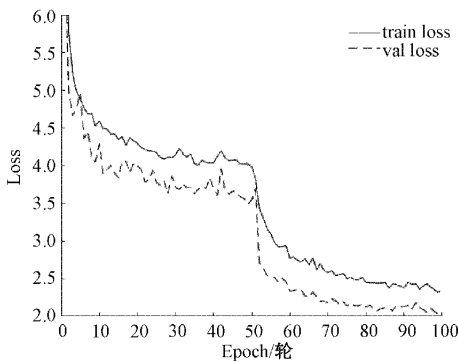


图 10 损失图

3.3 实验结果与分析

模型训练结束后,用精确率(precision)、召回率(recall)和均值平均精度(mAP)等评价指标来评估模型性能。精确度、召回率、平均精度值和均值平均精度计算过程如式(10)~(13)所示。

$$precision = \frac{TP}{TP + FP} \tag{10}$$

$$recall = \frac{TP}{TP + FN} \tag{11}$$

$$AP = \frac{TP + TN}{TP + TN + FP} \tag{12}$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP(i) \tag{13}$$

在本实验中,TP 为真正例,即佩戴了安全帽的人被预测为佩戴了安全帽,FP 为假正例,即未佩戴安全帽被预测为佩戴了安全帽,TN 为真反例,即未佩戴安全帽被检测为未佩戴安全帽,FN 为假反例,即佩戴了安全帽被检测为未佩戴安全帽。C 为类别数,AP(i)为第 i 个类别的精确率。蓝色框为未佩戴安全帽的人,标记为 person 并给出置信度,红色框为佩戴安全帽的人,标记为 hat。

对本文算法的性能进行验证,结果如表 2 所示,佩戴安全帽与未佩戴安全帽检测的准确率均高于其他算法,mAP 达到了 91.7%,比原 YOLOX 算法高出 1.2%,检测速度达到 71.9 帧/s,能够达到实时检测的效果。

表 2 不同检测算法性能对比

算法	AP/%		mAP/	FPS
	hat	person	%	
Faster RCNN	86.8	43.4	65.1	8.5
YOLOV3	90.2	86.9	88.6	40.8
YOLOV4	89.8	88.4	89.1	34.0
SE-Net-YOLOV5S	89.0	86.8	87.9	20.5
YOLOX	93.0	88.1	90.5	72.2
本文算法	93.9	89.5	91.7	71.9

表 3 为消融实验结果,来定性定量观察各改进模块的效果,加入 ECA-Net 注意力机制后,网络特征提取能力增强,网络能够更好的捕获目标信息,一定程度上改善小目标漏检情况。对损失函数进行优化后,解决两框不相交时无法回归的问题,比原来算法有提高,验证方案有可行性。

表 3 消融实验结果

分组	ECA-Net	优化损失函数	mAP/%
1			90.5
2	✓		91.2
3		✓	91.7

为了更直观了解不同算法的检测效果,将同一张图像分别在 YOLOX 网络和改进后的网络中进行检测,检测图分别如图 11~13 所示,左边为 YOLOX 检测效果图,右边为改进后检测效果图,可以看出改进后的 YOLOX 对小目标的检测效果更好,减少了小目标漏检情况,精度有一定提升。

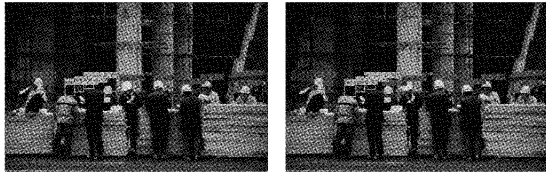


图 11 远距离目标检测



图 12 密集目标检测



图 13 小目标检测

4 结 论

为弥补人工监督施工现场安全防护的不足,提出一种基于 YOLOX 的改进算法,特征金字塔部分加入 ECA-Net 模块,进行跨通道交互,充分利用特征信息,优化损失函数,加快模型收敛速度,提升模型对小目标安全帽检测准确率。经过实验测试,有较高的准确率和实时性,与实验中其他算法相比具有一定优势,是弥补传统人工检测不足的一种可行方法。模型变得复杂后,参数增多,后续会对模型进行剪枝和量化,降低计算和存储成本。目前,实验在云服务器上通过测试,如何将其部署到多个真实场景下,进行多角度的检测,也是后期继续努力的方向。

参考文献

[1] 李坤伦. 改进的小目标检测方法在安全帽佩戴检测中的应用研究[D]. 成都:四川师范大学,2021.

- [2] 张锦,屈佩琪,孙程,等. 基于改进 YOLOv5 的安全帽佩戴检测方法[J]. 计算机应用,2021,12(6):1-11.
- [3] SHRESTHA K, SHRESTHA P P, BAJRACHARYA D, et al. Hard-hat detection for construction safety visualization [J]. Journal of Construction Engineering, 2015,2015(1):1-8.
- [4] WU H, ZHAO J. An intelligent vision-based approach for helmet identification for work safety[J]. Computers in Industry,2018, 100(7):267-277.
- [5] 孙国栋,李超,张航,等. 融合自注意力机制的安全帽佩戴检测方法[J]. 计算机工程与应用,2021,28(7):1-7.
- [6] 孙备,左震,吴鹏,等. 面向无人艇环境感知的改进型 SSD 目标检测方法[J]. 仪器仪表学报,2021,42(9):52-61.
- [7] 陈国平,彭之玲,黄超意,等. 基于改进 YOLOv3-Tiny 的毫米波图像目标检测[J]. 电子测量技术,2021,44(21):163-167.
- [8] 许凯,邓超. 基于改进 YOLOv3 的安全帽佩戴识别算法[J]. 激光与光电子学进展,2021,58(6):300-307.
- [9] 解尧婷,张丕状. 基于改进的 YOLOv4 输电线路小目标检测[J]. 国外电子测量技术,2021,40(2):47-51.
- [10] GE Z, LIU S T, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [11] REDMON J, DIVVALS S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2016:779-788.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2017:6517-6525.
- [13] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOV4: optimal speed and accuracy of object detection [J]. Computer Vision and Pattern Recognition,2020,17(9):198-215.
- [15] 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. salt lake City: IEEE Computer Society and CVF, 2018: 8759-8768.
- [16] WANG C Y, 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, New York: IEEE, 2020:390-391.
- [17] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximate-on in reinforcement learning[J]. Neural Networks, 2018,107:3-11.
- [18] WANG Q, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]. Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2020, Piscataway: IEEE, 2020: 11531-11539.
- [19] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

作者简介

丁田, 硕士研究生, 主要研究方向为图像处理。

E-mail: 1589206033@qq.com

陈向阳(通讯作者), 副教授, 主要研究方向为大数据处理、图像处理等。

E-mail: 04005074@wit.edu.cn