

DOI:10.19344/j.cnki.issn1671-5276.2026.03.023

基于 YOLOv7-HHF 的人手脸检测算法研究

郑乐镔, 蒋林轩

(广东工业大学 计算机学院电子信息系, 广东 广州 510006)

摘 要:提出一种改进 YOLOv7-tiny 的人手脸检测改进算法 YOLOv7-HHF, 此算法在人手脸检测数据集上的 mAP 值提高了 2.8%, 实现了性能更好的人手脸检测模型。现有的公开数据集主要聚焦于单独的人体、人脸或手部检测, 只能分步骤独立预测, 造成联合精确度波动大, 推理效率低下, 耗时较长。人手脸目标检测任务中存在“大-小”多尺度目标检测问题, 主流检测算法在面对这种多尺度复杂目标检测任务时往往会失效。为解决以上问题, 构建了一个人体、手部、脸部统一的数据集, 以保留三者之间的相对关系。基于 YOLOv7-tiny 网络, 增加小目标检测层, 提升小目标(手部和人脸)的检测准确性。通过融合 CA 注意力和递归门控卷积, 增强模型对人-手-脸特征的类间联系的挖掘并改善运动情形中目标被部分遮挡的检测效果。改用 Focal-EIoU Loss 损失, 以缓解“大-小”多尺度目标检测带来的优化震荡问题。

关键词:人手脸目标检测; YOLOv7-tiny 网络; 小目标检测; CA 注意力; 递归门控卷积

中图分类号:TP75 **文献标志码:**A **文章编号:**1671-5276(2026)03-0117-07

Research on Human-Hand-Face Detection Algorithm Based on YOLOv7-HHF

ZHENG Lebin, JIANG Linxuan

(Department of Electronic Information, School of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: This paper proposes an improved hand and face detection algorithm based on YOLOv7-tiny, named YOLOv7-HHF. The algorithm achieves a notable 2.8% improvement in mean Average Precision (mAP) for the human-hand-face detection dataset, realizing a better-performing model for hand and face detection. Existing public datasets primarily focus on detecting human bodies, faces, or hands separately, which necessitates independent step-by-step predictions, resulting in high fluctuation in joint accuracy, low inference efficiency, and long processing time. Moreover, the hand and face detection task involves the "large-small" multi-scale object detection problem, and mainstream detection algorithms often fail when facing such complex multi-scale object detection tasks. To overcome these challenges, this study introduces several key innovations. A unified dataset is constructed, encompassing human bodies, hands, and faces, to preserve the relative relationships among them. The YOLOv7-tiny architecture is enhanced by incorporating a small object detection layer tailored for detecting smaller targets, thereby improving the accuracy of detecting hands and faces. Based on the YOLOv7-tiny network, a small object detection layer is added to improve the detection accuracy of small objects (hands and faces). The integration of CA attention and recursive gated convolution augments the model's capability to capture inter-class relationships among human, hand, and face features, leading to improved detection performance, especially in scenarios where targets are partially occluded. The Focal-EIoU Loss is adopted to mitigate the optimization oscillations caused by "large-small" multi-scale object detection.

Keywords: human-hand-face object detection; YOLOv7-tiny network; small object detection; CA attention; recursive gated convolution

0 引言

随着人工智能和计算机视觉技术的快速发展, 人手脸目标检测任务在智能监控、人机交互、机器人技术等多个领域展现出广阔的应用前景。人手脸目标检测任务旨在利用算法和模型从输入的图像或视频序列中自动识别和定位人体、手部

和脸部的位置与范围, 具有高度的精确性和鲁棒性要求。其应用广泛且深远, 如智能监控中的异常行为检测、人机交互中的自然界面设计以及机器人技术中的情感识别与表达等。

人手脸检测虽不能直接获得区别每个人的标识, 但其作为上游任务为姿态估计、手势识别、表情识别、身份认证等下游任务提供基础的人体、人

基金项目:国家自然科学基金项目(62272108)

第一作者简介:郑乐镔(1999—), 男, 广东汕头人, 硕士, 研究方向为计算机视觉、深度学习、目标检测、模型轻量化研究, 979093003@qq.com。

手、人脸的位置信息,同时与下游任务互相关联和支持,实现更高级别的视觉理解。在人机交互领域,人手脸目标检测技术的应用使得人与机器之间的交互更加自然和智能。人手脸检测与下游任务联合应用,检测人体姿态、面部表情和手部动作,机器可以更全面地理解用户的意图和需求,提供更加个性化和智能的回应。在智能家居和智能车载场景中,用户可以通过简单的肢体动作或面部表情来控制设备或获取信息,享受更加便捷和舒适的使用体验。人手脸目标检测的准确性和鲁棒性对于实现更高级别的视觉理解和应用至关重要。随着技术的不断进步,它在未来将继续发挥重要作用,推动相关领域的持续发展。

目标检测领域中,基于卷积的目标检测算法分为一阶段目标检测算法和二阶段目标检测算法。二阶段目标检测模型以 R-CNN 系列为代表,经历了从 R-CNN^[1] 到 Fast R-CNN^[2] 再到 Faster R-CNN^[3] 的演进。通过引入卷积神经网络和区域候选策略,逐步提高了目标检测的速度和精度。一阶段目标检测算法以 YOLO 系列为代表,适用于实时性要求高的场景。从 YOLO 系列模型从 v1—v6^[4-9] 不断进化,引入如 GoogleNet^[10]、FPN (feature pyramid network)^[11]、SPP (spatial pyramid pooling)^[12]、PANet (path aggregation network)^[13] 和 Darknet 等重要模块结构以及 MixUp、Mosaic 等数据增强方法提升性能,在保持速度优势的同时,逐步提高检测精度和模型鲁棒性。

在人手脸统一检测任务中,现有的数据集主要聚焦于单独的人体、人脸或手部检测。然而,人体、人手和人脸作为一个统一的整体,在结构和功能上存在着密切的关联关系。过往方法在解决这一任务时,往往通过三个独立的目标检测算法分别实现人-手-脸检测,忽视了人体、人手和人脸之间的内在关联性和共享特征,也显著增加了模型的复杂性和计算成本,模型之间存在冗余和重复学习,整体性能不高。

本文采用两种方式分别解决上述两个问题。1)首先构建一个综合数据集,以便构建人体、手部和脸部之间的密切关联。2)基于 YOLOv7-tiny 和上述数据集,提出了人手脸检测的统一框架 YOLOv7-HHF,用于建模人手脸三者之间的内在关联性。实验结果显示该模型能在保持各类别检测性能的同时,显著降低了模型复杂度和计算成本。

1 YOLOv7-tiny 简介

2023 年发布的 YOLOv7 算法^[14], 相较于之前 YOLO 基线检测精度和检测速度都有明显的优势。在预测头 (Head) 网络使用重参数化 REPCONV^[15], 在训练阶段和推理阶段采用不同的结构, 推理阶段将多个计算单元合并减少模型结构的复杂度, 减少了计算量和内存消耗。但其网络结构复杂, 参数量大, 低算力的终端设备上部署仍存在问题。YOLOv7-tiny 同样由骨干网络 (Backbone)、颈部网络 (Neck) 和预测头 (Head) 组成。骨干网络部分, YOLOv7-tiny 采用更简洁的 ELAN 代替 YOLOv7 中的 E-ELAN; 仅保留 MPConv 中池化操作进行下采样, 取代原本池化操作和卷积操作并行的下采样; 颈部网络仍采用 PANet 结构进行特征聚合; 预测头网络, 采用标准卷积结构代替 REPCONV, 调整输出通道数。

2 YOLOv7-HHF

为构建一个统一的人手脸检测框架,本文在YOLOv7-tiny 基础上进行了改进,提出了 YOLOv7-HHF 网络,如图 1 所示。

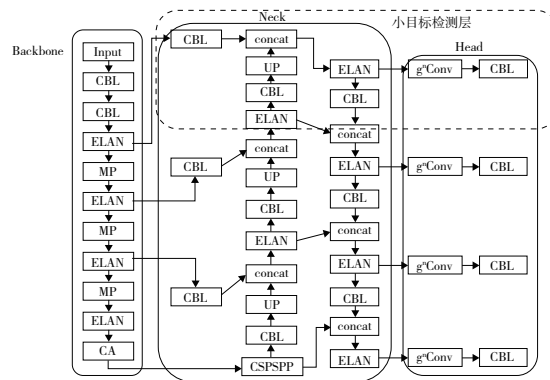


图 1 YOLOv7-HHF 网络结构

YOLOv7-HHF 算法流程如图 2 所示。

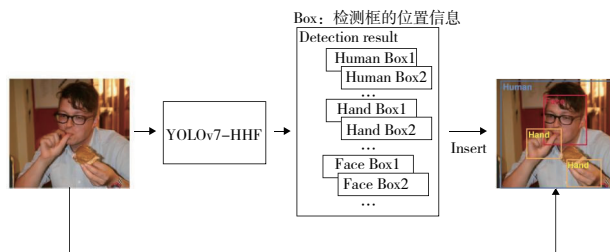


图 2 YOLOv7-HHF 算法流程

针对输入的图像, YOLOv7-HHF 算法能够准

确地进行回归预测,不仅输出检测框的详细类别信息,还精确提供位置信息。这些位置信息涵盖了标注框中心点与图像宽高之间的相对 X 坐标、 Y 坐标,以及检测框与图像宽度和高度的相对尺寸。通过这些回归结果,算法能够实现对目标区域的精确锁定和分类,为图像分析提供了坚实的基础。

2.1 增加小目标检测层

人手脸检测统一检测任务中,人手和人脸这两个类别相对于整个人体尺寸明显较小。人手脸目标检测任务中存在“大-小”多尺度目标检测问题。主流的检测网络常使用下采样来构建“大”尺度目标特征。而“小”尺度目标的信息极易丢失。网络难以充分提取到手和脸这两个小目标类别的特征信息。为解决这一问题,本文在骨干网络的第一个最大池化层之前增加一个输出,用以提取浅层的小目标特征,并在此基础上增加一个专门用于小目标检测的新输出头。小目标检测层如图1虚线所示。

在颈部网络增加上采样层和卷积层,同时用Concat操作融合前一层上采样得到的特征图和第一个最大池化层前输出的浅层特征;在检测头网络处生成一个富含小目标特征信息的小目标检测头,利用更多的小目标特征信息,改善小目标类别的检测效果。

2.2 引入CA模块

正如前文所述,人手、人脸和完整人体之间存在相对的空间联系。然而,传统的卷积方法往往难以准确地建模这三者之间的位置关系。SE等^[16]注意力在提高模型性能方面具有显著效果,但它们通常忽略了位置信息。

为了更好地捕捉人手、人脸和完整人体之间的相对特征,本文引入了CA (coordinate attention)^[17]模块,在YOLOv7-tiny骨干网络末端进行融合,如图1所示。CA模块如图3所示,暂定 C 、 H 、 W 分别为输入特征张量 $\mathbf{X} = [x_1, x_2, x_3, \dots, x_c]$ 的通道数、高度和宽度。为了充分捕获位置信息,首先分别将全局池化操作按水平和垂直两个方向,分解为两个独立方向的池化操作。如式(1)和式(2)所示,特征图按 X 方向和 Y 方向进行池化,分别生成 z^h 和 z^w 两个特征图,其中两个特征图大小分别为 $C \times H \times 1$ 和 $C \times 1 \times W$ 。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq w} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq h} x_c(j, w) \quad (2)$$

将两个方向分别池化得到的特征图在第三个维度进行Concat操作,得到 $C \times 1 \times (W+H)$ 的特征图,同时利用 1×1 卷积对 $C \times 1 \times (W+H)$ 的特征图执行 r 倍下采样,结合归一化操作和非线性激活函数,得到了同时包含水平和垂直两个方向信息的 $C/r \times 1 \times (W+H)$ 的特征向量,记为 $\mathbf{f}$ 。再者,沿着上文执行Concat操作的维度,对特征向量 $\mathbf{f}$ 执行Split操作,并使用两个 1×1 卷积分别对Split操作得到的两个特征向量进行 r 倍上采样,得到大小为 $C \times H \times 1$ 的 $\mathbf{f}^h$ 和 $C \times 1 \times W$ 的 $\mathbf{f}^w$;最后,结合Sigmoid函数分别对 $\mathbf{f}^h$ 和 $\mathbf{f}^w$ 计算,得到注意力向量 $\mathbf{g}^h$ 和 $\mathbf{g}^w$ 并相乘,得到CA注意力机制模块的输出。CA注意力机制提高了模型对空间信息的利用,加强了表征能力。

CA模块提高了模型对空间信息的利用,加强模型的表征能力,其灵活性和有效性使其成为移动网络中值得探索的注意力机制。

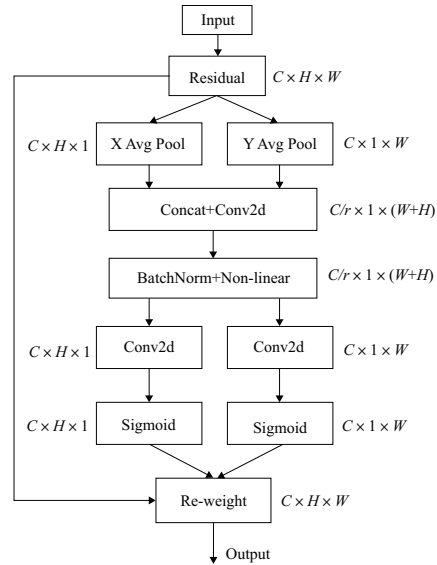


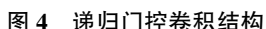
图3 CA模块结构

2.3 融合递归门控卷积 $g^n\text{Conv}$

在人手脸检测过程中,人体运动状态的多样性和复杂性导致各种遮挡情况的存在,造成目标特征的混淆和丢失。如图1所示,本文在网络预测头处引入递归门控卷积,实现了输入自适应、远程和高阶的空间交互。

递归门控卷积($g^n\text{Conv}$)^[18]由标准卷积、深度可分离卷积(DWConv)、线性投影(Proj)和元素乘法(Mul)组成。在人手脸检测任务中,挥手遮挡

为了解决以上问题,本文使用 Focal EIoU Loss^[19]作为损失函数。不同于 CIoU Loss 针对宽和高相对比例进行惩罚,Focal EIoU Loss 将方位损失分解为直接对预测框的宽和高回归结果进行惩罚,以更好地反映了预测框和真实框的相似度。它使本文方法在反向传播过程中更加关注与目标框重叠度高、高质量的锚框,减少了低质量锚框所产生梯度的干扰。Focal EIoU Loss 如式(5)所示


$$p_{k+1}=f_k(q_k)\odot g_k(p_k)/\alpha,(k=0,1,\cdots,n-1) \quad (3)$$
$$g_k = \begin{cases} \text{Identity}, & k=0 \\ \text{Linear}(C_{k-1}, C_k), & 1 \leq k \leq n-1 \end{cases} \quad (4)$$

2.4 优化过程中的冲突

人手脸检测数据集的标注信息涵盖了检测框的类别和精确位置信息,旨在为训练目标检测模型提供丰富的数据支撑,使其具备对未知图片中

的人体、手部及脸部进行精准检测的能力。鉴于运动状态的多样性和复杂性,数据集中不乏众多小目标物体以及遮挡情况的存在,这就要求我们在目标检测算法层面上进行针对性的优化,以提升模型在各类实际场景中的检测性能和稳定性。本实验将整个数据集按照4:1的比例随机划分为训练集和测试集。具体来说,训练集包含了10 000张图像,用于模型的训练和优化;而测试集包含了2 500张图像,用于评估模型的检测性能。

3.2 实验环境与配置

本文实验使用Ubuntu 18.04 LTS操作系统,搭载24G显存的NVIDIA GeForce RTX3090显卡,深度学习框架为Pytorch v1.10.0。实验过程,训练批次大小(bs)设置为64,训练轮数(epoch)为300,优化器使用SGD,初始学习率为0.001。实验环境、实验参数配置分别如表1—表2所示。

表1 实验环境

实验环境	版本
操作系统	Ubuntu 18.04 LTS
CPU	intel(R) Core(TM) i9 - 10900K
GPU	NVIDIA GeForce RTX3090
内存	24 G
深度学习框架	Pytorch v1.10.0
IDE	Pycharm v2020.1

表2 实验参数配置

bs	epoch	lr	Mo-mentum	Weight_decay	Input_size
64	300	0.001	0.937	0.000 5	640

3.3 评价指标

评价网络性能的指标主要有召回率(recall, R)、平均精度(average precesion, AP)和平均精度均值(mean average precesion, mAP),各指标的定义如式(6)—式(8)所示

$$R = \frac{T_p}{T_p + F_N} \quad (6)$$

$$A_p = \int_0^1 P(R) dR \quad (7)$$

$$m_{AP} = \frac{1}{N} \sum A_p \quad (8)$$

式中: T_p 为预测正确的正样本数量; F_N 为预测错误的负样本数; F_p 为预测错误的正样本数; N 为待检测目标的类别数。

3.4 实验结果与分析

1) 对比试验

为了进一步验证本文提出方法的优势,在本文构造的人手脸数据集上与当前主流的目标检测算法进行对比,包括二阶段目标检测算法Faster RCNN以及单阶段目标检测算法SSD^[20]、YOLOv4-tiny、YOLOv5s和YOLOv7-tiny,实验结果如表3所示。

表3 不同网络对比实验

Model	mAP/%	R/%	P_{arms}/M
Faster RCNN	89.4	83.9	185.6
SSD	77.9	74.1	23.5
YOLOv4-tiny	86.8	83.6	22.8
YOLOv5s	90.0	84.2	14.0
YOLOv7-tiny	90.5	84.4	6.0
YOLOv7-HHF	93.3	87.1	6.8

由实验结果可知,在人手脸任务中,本文提出的改进算法的性能相比于主流模型都有一定优势。从mAP指标上看,YOLOv7-HHF相对于Faster RCNN提高了3.9个百分点,且相对于同样为YOLO系列目标检测算法的YOLOv4-tiny和YOLOv5s分别提高了6.5个百分点和3.3个百分点,相较于YOLOv7-tiny则提高了2.8个百分点;从参数量看,得益于统一的人手脸检测框架,本文提出的YOLOv7-HHF算法也有明显优势,仍然具有轻量的优势。

本文基于YOLOv7-tiny网络做出改进,由于增加小目标检测层引入了一定数量的卷积模块和上采样层,且改进融合的递归门控卷积结构存在一定计算量,故改进后的YOLOv7-HHF算法在参数量上相较于原本的YOLOv7-tiny网络提升了0.8 M;但从mAP指标上看,YOLOv7-HHF相对于YOLOv7-tiny在人手脸检测数据集上提高了2.8个百分点。总体上看,改进后的检测网络仍适合部署在边缘网络,应用在对实时性有要求的人机交互场景,在参数量少量提升的前提下,进一步提升了模型的检测准确度。

2) 小目标检测层改进对比实验

为了验证改进小目标检测层的效果,本文在人手脸检测数据集上,对改进前后的结果进行对比,实验结果如表4所示。

表4 小目标检测层改进前后结果对比实验

是否改进 小目标 检测层	$mAP_{(all\ classes)}/\%$	$mAP_{(hand)}/\%$	$mAP_{(person)}/\%$	$mAP_{(face)}/\%$	P_{arms}/M
否	90.5	86.8	89.0	95.6	6.0
是	91.2	89.1	88.7	96.7	6.1

由实验结果可见,改进小目标检测层之后人手、人脸这两个尺寸较小的类别 mAP 值相较于改进前有了较为明显的提高,分别提高了 2.3 个百分点和 1.1 个百分点,由于改进后 YOLOv7-tiny 网络的颈部网络增加了卷积层和上采样层,以及检测头网络增加了一个小目标检测分支,故参数量相较于改进前有少量增加,从 6.0 M 增加至 6.1 M;另外,人体(Person)类因其目标尺寸较大,mAP 值相较于改进前降低了 0.3 个百分点。总体来说,通过增加对 YOLOv7-tiny 网络浅层中小目标信息的利用,实现了对小目标更友好的检测模型,在实际的人机交互场景中对于加强手势、表情等信息的利用也有重要的意义。

3) 消融实验

为了进一步验证本文提出改进点的有效性和合理性,本文在人手脸检测数据集上设计消融实验,并保持不同组别之间参数设置一致。实验结果如表 5 所示。

表 5 消融实验

实验	小目标检测层	CA 注意力	g^n Conv	Focal EIoU	mAP/%	P_{params}/M
1	—	—	—	—	90.5	6.0
2	✓	—	—	—	91.2	6.1
3	✓	✓	—	—	91.7	6.1
4	✓	✓	✓	—	92.9	6.8
5	✓	✓	✓	✓	93.3	6.8

由实验结果可知,实验 2 在实验 1 基础上增加了小目标检测层,提高了对小尺寸目标的检测效果,mAP 提升了 0.7 个百分点,但同时参数量增加了 0.1 M;实验 3,通过在骨干网络末端融合 CA 模块,在复杂的背景中,提升模型对重要信息的关注,提高对空间信息的利用,加强模型学习能力,其 mAP 值相对于实验 2 提升了 0.7 个百分点,且由于 CA 注意力机制简洁轻量,因此模型改进之后参数量并未显著增加;实验 4 在实验 3 基础上,于 YOLOv7-tiny 检测头融合递归门控卷积 g^n Conv,通过捕捉图像中的高阶空间互动,实现长距离建模,mAP 值相对于实验 3 也提升了 1.2 个百分点,参数量增加了 0.7 M。最后,实验 5 改进了 Focal EIoU 损失函数,mAP 指标提升了 0.4 个百分点。由消融实验可知,在原 YOLOv7-tiny 模型上改进的不同模块都有体现了相应的优势。

4 检测效果分析

为了直观地展示本文改进算法 YOLOv7-

HHF 相较于原始 YOLOv7-tiny 算法在人手脸检测任务上的性能提升,我们选取了人手脸检测数据集测试集中部分包含小目标和遮挡情况的困难样本,并分别使用两种算法进行检测。图 6 展示了这两种算法在相同图像上的检测效果对比。



图 6 YOLOv7-HHF 和 YOLOv7-tiny 算法检测效果可视化对比

在图 6(a)中,我们可以看到原始 YOLOv7-tiny 算法的检测效果。虽然该算法能够检测到大部分的人手脸目标,但在某些情况下仍然存在一些问题。例如,对于远处尺寸较小的人脸,原始算法往往会出现漏检的情况。此外,在目标人物重叠时,原始算法的检测效果也不尽如人意,往往无法准确检测出所有的目标。相比之下,图 6(b)展示了本文改进后的 YOLOv7-HHF 算法的检测效果。通过对比可以看出,改进后的算法在多个方面都表现出了明显的优势。首先,对于远处尺寸较小的人脸,改进算法能够成功检测出更多的目标,这有效缓解了因目标尺寸小而导致的漏检问题。其次,在目标人物重叠的情况下,改进算法也能够更好地分离并检测出每一个目标,从而实现了更好的检测效果。此外,改进算法还减少了将远处物体误识别为人的情况,降低了小目标的错检率。从图 6 中可以看出,改进算法在检测到的目标上普遍获得了更高的置信度得分,这进一步

证明了其检测效果的提升。

5 结语

本文建立了一个专门的人手脸检测数据集以建模人手脸之间存在的相对空间关系。基于YOLOv7-tiny构建一个多任务级联式人手脸统一检测框架YOLOv7-HHF。实验结果表明:得益于建模人手脸之间的空间关联性,该框架不仅有助于减少多模型检测带来的部署困难,还将精确度稳定提升了2.7个百分点。

然而,实际的人手脸检测在室外场景中面临更多挑战。如逆光、阴影、光照不均等问题,这些因素都可能对模型的性能产生负面影响。针对逆光问题,一种可能的解决方案是在模型训练过程中加入更多的逆光场景数据,以便让模型学会在这种条件下进行准确的检测。此外,还可以考虑采用更先进的图像预处理技术,如自适应直方图均衡化(adaptive histogram equalization)或Gamma校正,来改善输入图像的质量,减少光照不均对模型的影响。

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014:580-587.
- [2] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015:1440-1448.
- [3] REN S Q, HE K M, 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.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016:779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017:6517-6525.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. 2018; arXiv: 1804.02767. <https://arxiv.org/abs/1804.02767>
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020; arXiv: 2004.10934. <https://arxiv.org/abs/2004.10934>
- [8] Glenn J. YOLOv5: state of the art object detection at 140 FPS [J]. arXiv preprint arXiv:2010.03167, 2020.
- [9] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [EB/OL]. 2022; arXiv: 2209.02976. <https://arxiv.org/abs/2209.02976>
- [10] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015:1-9.
- [11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017:936-944.
- [12] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [13] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.
- [14] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [15] DOBARIYA N J. Machine Learning for Autonomous Fault Detection in Wind Turbine Blades [D]. Kingsville: Texas A&M University-Kingsville, 2022.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:7132-7141.
- [17] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021:13708-13717.
- [18] RAO Y, ZHAO W, TANG Y, et al. Hornet: efficient high-order spatial interactions with recursive gated convolutions [J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.
- [19] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [M]//Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.