徐佩, 陈亚江. 融合 Swin Transformer 的 YOLOv5 口罩检测算法[J]. 智能计算机与应用,2024,14(5):83-92. DOI:10.20169/j.issn.2095-2163.240511

融合 Swin Transformer 的 YOLOv5 口罩检测算法

徐 佩, 陈亚江
(浙江理工大学 理学院, 杭州 310018)

摘 要: 针对口罩佩戴检测算法未平衡模型规模与检测精度的问题,提出了一种口罩佩戴检测改进算法。该算法以 YOLOv5 网络为基础框架:首先,应用轻量级 *Mixup* 数据增强方式和 *Mish* 激活函数以提高模型泛化能力;其次,引入 Swin Transformer 结构和 ECA 注意力机制来增强复杂场景下口罩目标的提取效率;第三,使用 *SIoU* 损失函数以提高检测精度;最后,设计了新的 Neck 网络卷积模块来实现模型轻量化。实验结果表明:相比于原始的 YOLOv5 算法, *mAP* 提升 2.9%,参数量减少 54.2%,模型体积减少 52.1%。该算法很好地平衡了模型规模与检测精度,在口罩检测实际场景中更具优势。

关键词: YOLOv5 算法; 口罩佩戴检测算法; 注意力机制; Swin Transformer; 轻量化
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2024)05-0083-10

Mask detection algorithm based on YOLOv5 integrating Swin Transformer

XU Pei , CHEN Yajiang
(School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Aiming at the problem that mask wearing detection algorithms have not yet balanced the model scale and detection accuracy, an improved mask wearing detection algorithm is proposed. The algorithm is based on the YOLOv5 network. Firstly, lightweight *Mixup* data enhancement and *Mish* activation function are used to improve the model generalization ability; Secondly, Swin Transformer structure and ECA attention mechanism are introduced to enhance the extraction efficiency of mask targets in complex scenes; Thirdly, the *SIoU* loss function are used to improve detection accuracy; Finally, a new Neck convolutional module is designed to realize the lightweight of the model. The experimental results show that compared with the original YOLOv5 algorithm, *mAP* is improved by 2.9%, parameter number is reduced by 54.2%, and model volume is reduced by 52.1%. The proposed algorithm balances model scale and detection accuracy well, which is more advantageous in mask detection scenarios.

Key words: YOLOv5 algorithm; mask wearing detection algorithm; attention mechanism; Swin Transformer; lightweight

0 引 言

诸多传染性强的病毒通过呼吸道飞沫传播,例如新冠病毒、流感病毒等,因此,规范佩戴口罩是降低此类病毒传播风险、阻断疫情扩散蔓延、减少公众交叉感染、保障身体健康最方便、最有效的措施。近年全国人民同心抗疫,自 2023 年 1 月 8 日起国务院应对新型冠状病毒感染疫情联防联控机制综合组发布了《关于对新型冠状病毒感染实施“乙类乙管”的总体方案》,标志着国内疫情防控工作取得了胜利。然而,实施“乙类乙管”的措施后,通过呼吸道飞沫传播的传染病仍时有发生[1],抵抗力较低的人群更需要主动佩戴好口罩,这不仅能减小感染病毒的几率,也可以降低其他呼吸道疾病的发病率[2]。当新一轮病毒来袭,如果能够在养老院、学校、车站等人员聚集的公共场所,对规范佩戴口罩实施高效的自动检测,不仅能提升疫情管控效率,而且还可以节省成本。因此,进一步提高口罩佩戴检测算法的检测效率和速度,对研发和升级自动检测口罩佩戴仪器,以及今后类似传染病的防控具有重要的现实意义[3]。

现有的口罩佩戴检测算法以目标检测算法为依据可分为 2 类。一类是一阶段目标检测算法,一步实现分类和回归任务,生成候选框,并以 SSD(Single shot multibox detector)[4]、RetinaNet[5]、YOLO(You only look once)算法系列为代表。应用较多的典型一阶段检测模型是 YOLO 算法。YOLO 凭借其强大

作者简介: 徐 佩(1999-),女,硕士研究生,主要研究方向:目标检测,机器学习。
通讯作者: 陈亚江(1984-),男,博士,副教授,主要研究方向:计算机视觉,图像处理。Email:yjchen@ zstu.edu.cn

收稿日期: 2023-08-08

的实时检测性能,真正推动目标检测走向大规模落地应用。例如,魏明军等学者[6]改进 YOLOv3 特征金字塔结构,利用跳跃连接和包含通道注意力的位置特征增强模块 LFE,其平均精度均值(mean Average Precision,*mAP*)达到 86.96%。曹域硕等学者[7]基于 YOLOv3 模型,引入注意力机制,同时考虑了口罩佩戴不规范这一类别,*mAP* 达到了 93.33%,但其模型规模较大。谈世磊等学者[8]将图片归一化操作后送入 YOLOv5 标准网络中训练,*mAP* 达到了 92.4%。王艺皓等学者[9]在 YOLOv3 算法中引入改进的空间金字塔池化结构,并采用特征融合策略提升了复杂场景下的口罩检测效果,*mAP* 达到了 90.2%。赵文清等学者[10]在基于 YOLOv5 的遥感目标检测中,引入 Swin Transformer 网络结构和注意力机制,将 *mAP* 提高了 5.3%,达到 88.9%。整体来看,通过改进 YOLO 系统的网络模型,提升了遮挡和小目标检测的效果,*mAP* 较原算法有所提高,但存在模型参数量大,结构冗余的问题。

为了降低实际应用的部署成本,众多学者对 YOLO 系列模型的轻量化改进做了研究。例如,张烈平等学者[11]将预训练的 MobileNetv2 特征提取网络与 YOLOv2 网络相结合,构成了简化的口罩佩戴检测网络模型,将检测速度提升了 2.5 倍。薄景文等学者[12]使用 ShuffleNetv2 替换 YOLOv3 的主干特征提取网络,模型体积压缩了 63.1%。王艺霏等学者[13]在 YOLOv4 算法基础上,研究引入 Ghost module 模块搭建特征提取网络,模型参数减少了 82.05%,检测速度达到了38.23 fps。彭成等学者[14]基于 YOLOv5 模型将参数量更小的 Ghost Bottleneck CSP 和 Shuffle Conv 模块代替原网络的 C3 和 Conv 模块,参数量减少为原来的34.24%,CPU 平台的运行速度提升了 28.25%。王圣雄等学者[15]将卷积注意力机制、Ghost 卷积技术与 YOLOv5 模型结合,以增强特征提取能力,获得了89.1%的检测精度,模型大小减少了 19.63%。总体而言,基于 YOLO 系列的轻量化改进模型在网络参数量和检测速度上均有所提升,却同时伴随平均检测精度的下降,甚至在实际场景下仍会漏检小目标人脸[12]。

另一类口罩佩戴检测算法则基于两阶段的目标检测算法,该类算法根据生成的候选区域,通过判断前景与背景,使用边界框回归分类校准检测结果,以 R-CNN(Region-Based Convolution Neural Network)[16]、Faster R-CNN[17]、Mask R-CNN[18]等为代表。例如,李泽琛等学者[19]在 Faster R-CNN 框架中,研究引入残差结构和注意力机制,在 FMDD(Face Mask Detection Dataset)数据集上的平均精度均值为88.69%,但并未将口罩佩戴不规范数据单独分组。任钰等学者[20]基于 Faster R-CNN 采用 Res Net-101 特征提取网络,将平均精度均值提升到89.41%。刘玉国等学者[21]为获得更多语义信息,使用 ResNet 残差网络替代 Faster R-CNN 算法的 VGG16 网络,平均精度均值提升至 93.06%。该类算法具有精度高的特点,但李泽琛等学者[19]针对自然场景的小目标检测同样存在漏检、误检情况。任钰等学者[20]设计的 FMD-RCNN 模型,其检测速度为 310 ms,*FPS* 为3.23,网络结构的复杂性导致检测速度不够理想。以上基于两阶段检测算法的改进模型,虽然提升了检测精度,但是速度有待提高。对于口罩检测场景,*FPS* 至少应达到 30,才能较好满足实时检测的要求。

由于口罩检测算法对实时性要求较高,检测精度和模型大小同时兼顾,才能更好地应用于实际场景。本文选择速度较快的 YOLOv5 一阶段目标检测算法,融入 Swin Transformer 结构。本文的训练实验显示,改进后的算法将口罩目标检测 *mAP* 提升 2.9%,参数量减少 54.2%,模型体积减少 52.1%,同时达到了 53.3 fps 的检测速度,较好地平衡了模型规模与检测精度。

1 YOLOv5 算法原理

YOLO 是一种通用的目标检测模型,YOLOv5 为其第 5 个版本,与以往几个版本相比,检测速度更快、精度也更高。YOLOv5 根据模型从大到小可以分为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 等 4 个版本。为满足精度和实时检测要求,本文选择 YOLOv5s 中精度较高的 6.0 版本。

YOLOv5s 的总体框架如图 1 所示,由 Input 输入端、Backbone 特征提取网络、Neck 特征融合网络和 Head 输出端四个部分组成,对各部分功能、这里可做阐释分述如下。

YOLOv5 的输入端负责对数据集图片进行预处理操作,将输入图片缩放至网络指定的输入尺寸,然后进行自适应锚框计算,比较预测锚框和初始锚框的差距,找到最佳锚框值,送至检测网络。

Backbone 网络可以分为 Conv 卷积结构、跨阶段局部结构(Cross Stage Partial,CSP)和空间金字塔池化(SPPF)结构等 3 个部分。SPPF 模块采用多个小尺寸池化核级联代替空间金字塔池化(Spatial Pyramid Pooling,SPP)[22]模块中单个大尺寸池化核,进一步提高运行速度。

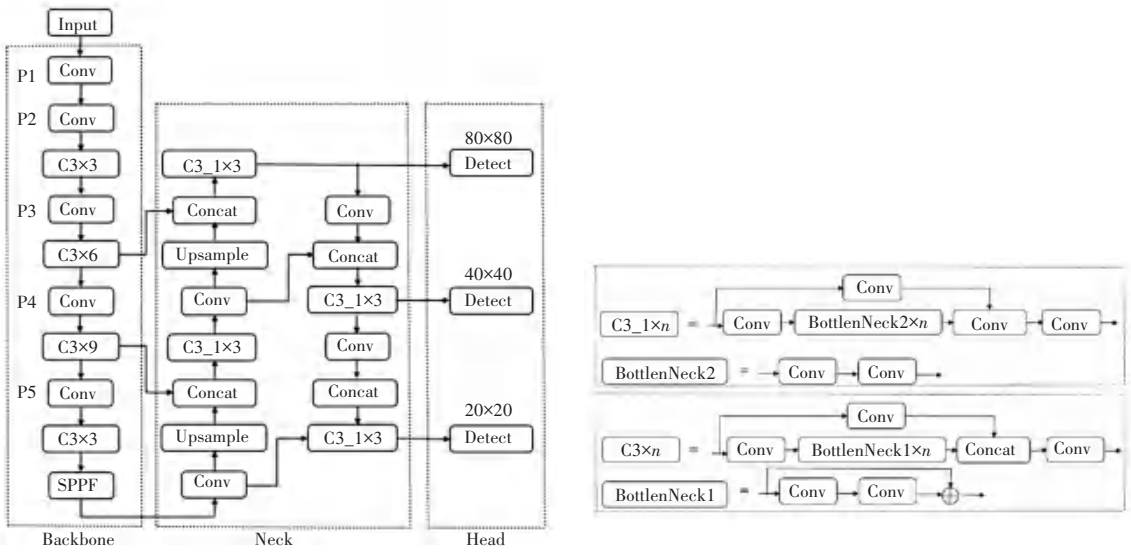

图 1　YOLOv5s 网络结构图

Fig. 1　YOLOv5s network structure diagram

Neck 网络使用特征金字塔(Feature Pyramid Networks, FPN)[23] 与路径聚合网络(Path Aggregation Net-works, PAN)[24]结合的方式进行特征融合,FPN 将语义信息通过自顶向下的结构向浅层传递,PAN 是在 FPN 的基础上又引入一条自底向上的路径,使位置信息能传到深层。采用两者结合的方式,可增强不同网络层的语义特征和位置特征的表达。

最后,Head 输出端包含 3 个检测层,输出维度为 20×20×75、40×40×75 和 80×80×75,分别检测小、中、大目标,利用非极大值抑制算法(Non Maximum Suppression, NMS)算法去除冗余的预测框,最后输出预测的类别与位置信息。

2　改进网络模型

针对 YOLOv5 网络对密集场景下口罩小目标检测存在误检、漏检等情况,而且参数量大,难以部署的问题,做出了改进。使用更轻量化的数据增强方式 *Mixup*、将 *SiLU* 替换为 *Mish* 激活函数,在降低计算开销的同时提高模型的泛化能力;然后使用 *SIoU* 边框损失函数代替原始的 *CIoU*,提高模型训练的收敛速度与精度。

将 Swin Transformer 结构引入 Backbone 层,使用 C3STR 替换后 2 层的 C3 模块,更好地识别底层信息,在第一个 C3STR 模块前以及 SPPF 特征融合前添加 ECA 模块,增强网络对有用信息提取,减少复杂背景下的误检和漏检等情况,该模块改进前后的对比结构如图 2 所示。

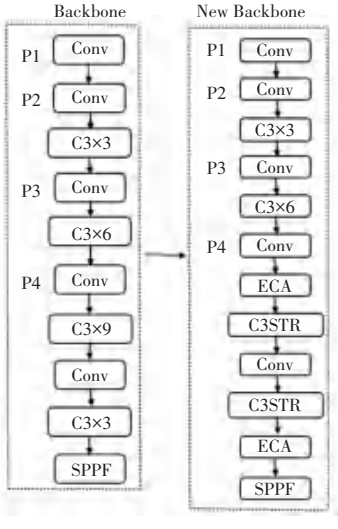

图 2　Backbone 结构改进图

Fig. 2　Backbone structure improvement diagram

将 Neck 层普通卷积替换为 DWConv,且将 C3 替换为轻量级的 C3Ghost 模块,同时在 C3 模块之后添加 ECA 模块,降低网络的参数量,实现模型轻量化,Neck 层改进前后对比如图 3 所示。改进后的 YOLOv5s 网络总体结构如图 4 所示。

2.1　Mixup 样本数据增强

Mixup 是一种轻量化的数据增强方式,将原始数据中的随机 2 个数据进行正负样本的融合生成新的一组样本,由此使得样本量翻倍,提高了小样本模型的泛化能力[25]。大多数数据增强方法没有多个图像之间的融合,而且计算成本较大。本文采用的 Mixup 数据增强方法,通过线性插值的方法实现新样本和标签的构建,降低了计算开销,具有更好的实

用性。Mixup 的计算公式分别如下：

$$\lambda = Beta(\alpha, \beta) \quad (1)$$

$$mixed\_batch_x = \lambda \cdot batch_{x1} + (1-\lambda) \cdot batch_{x2} \quad (2)$$

$$mixed\_batch_y = \lambda \cdot batch_{y1} + (1-\lambda) \cdot batch_{y2} \quad (3)$$

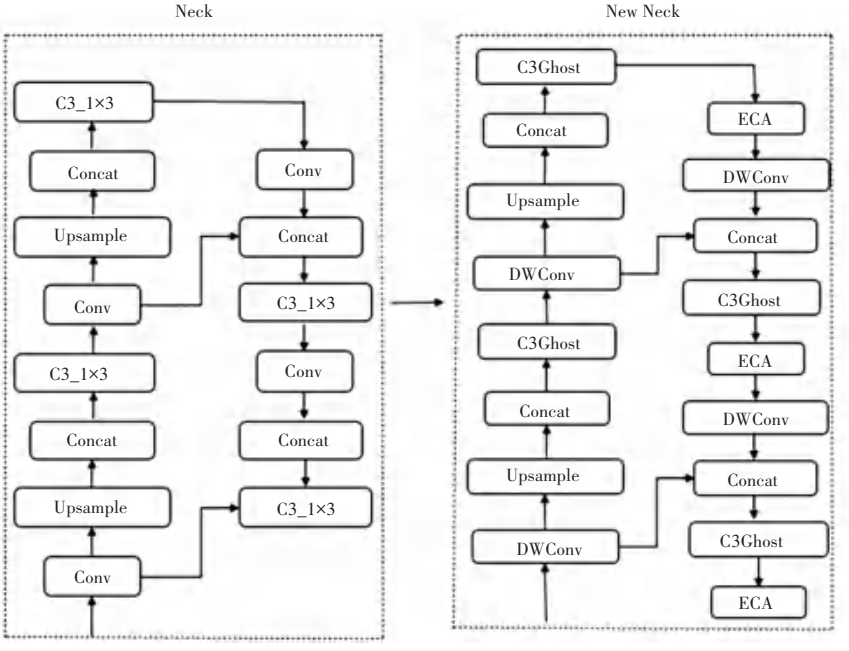

图 3 Neck 结构改进图

Fig. 3 Improvement of Neck structure

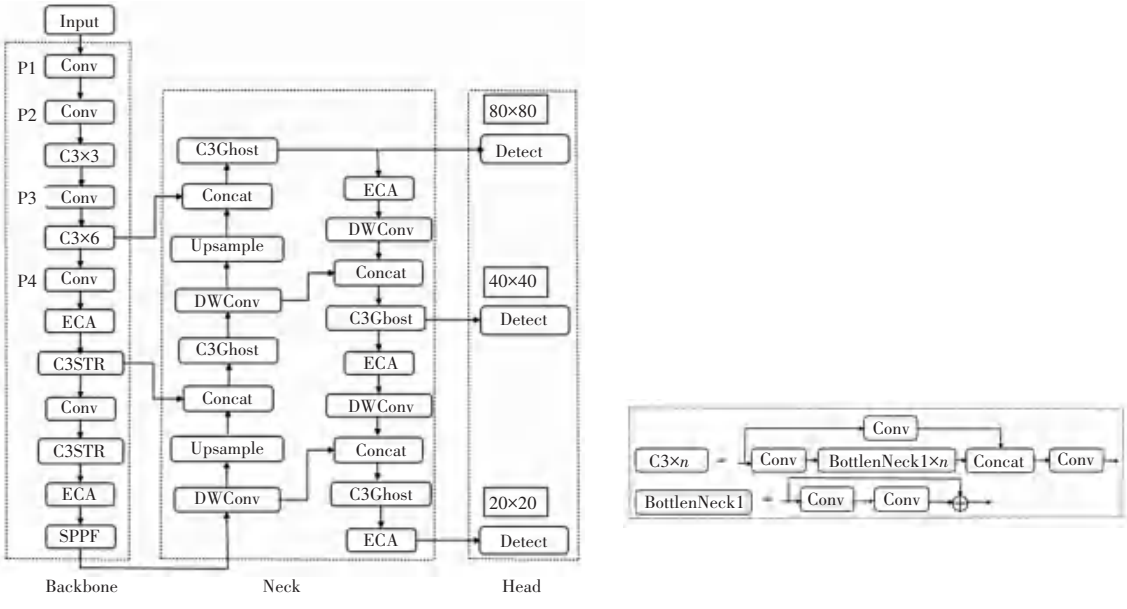

图 4 YOLOv5s 改进网络结构图

Fig. 4 Improved network structure of YOLOv5s

其中，*Beta* 表示贝塔分布；λ 为参数 α、β 由贝塔分布得出的混合系数；$batch_{x1}$ 和 $batch_{x2}$ 表示 2 个 *batch* 样本；$batch_{y1}$ 和 $batch_{y2}$ 表示样本对应的标签；混合后的 *batch* 样本和对应的标签分别是 $mixed\_batch_x$ 和 $mixed\_batch_y$。当 α、β 取值不同时，其 *Beta* 分布概率密度曲线如图 5 所示。

2.2 *Mish* 激活函数

激活函数将神经网络的输入映射到输出，为网络提供非线性，从而提高了神经网络对于模型相关的表达能力。目前应用较广的激活函数包括

Sigmoid、*Softplus*、Tanh、*ReLU* 和 *Mish* 等，曲线如图 6 所示。

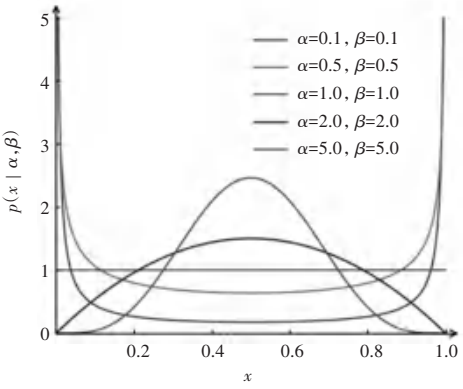

图 5　Beta 分布概率密度曲线

Fig. 5　Beta distribution probability density curve

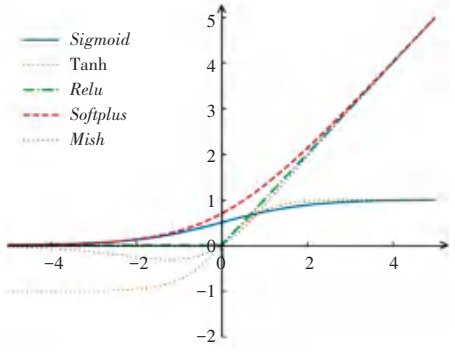

图 6　激活函数曲线

Fig. 6　Activation function curve

然而采用 *Sigmoid* 和 *Softplus* 激活函数，训练时易发生梯度消失现象。Tanh 激活函数的输出极值趋近 −1 和 1 时，模型会产生梯度饱和。*ReLU* 激活函数对大梯度的反向传播会产生大量无效神经元。相比之下，*Mish* 函数都具有非单调、平滑、有下界、无上界等良好特性，可提高网络的可解释性和梯度流。*Mish* 函数的相关公式为：

$$f(x) = x \cdot \tanh[Softplus(x)] \tag{4}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{5}$$

$$Softplus(x) = \ln(1 + e^{x}) \tag{6}$$

2.3　Backbone 网络改进

2.3.1　基于 Swin Transformer 编码的 C3STR 模块

本文将 Swin Transformer 结构[26]嵌入到 C3 卷积块中，构成新模块 C3STR，从而引入 Transformer 的离散参数，借助 Swin Transformer 结构中的窗口自注意力模块，增强小目标的语义信息和特征表示。

Swin Transformer 模块包括成对的窗口多头注意力层、滑动窗口多头注意力层、MLP 层和归一化层。输入嵌入层包括 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 三个矩阵信息[27]，可由式(7)进行描述：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = SoftMax(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d} + \boldsymbol{B})\boldsymbol{V} \tag{7}$$

其中，$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示查询、键以及值矩阵；d 表示通道数；$\boldsymbol{B}$ 表示相对位置偏差。

通过归一化层处理后送入窗口多头注意力层，经融合以及归一化层处理后，传至多层感知机 MLP，将堆叠融合后的输出作为另一个 Swin Transformer 模块的输入，通过归一化层处理后送入滑动窗口多头注意力层，经 LN 层处理后，传至多层感知机 MLP，作为网络输出。C3STR 与 Swin Transformer 结构如图 7 所示。

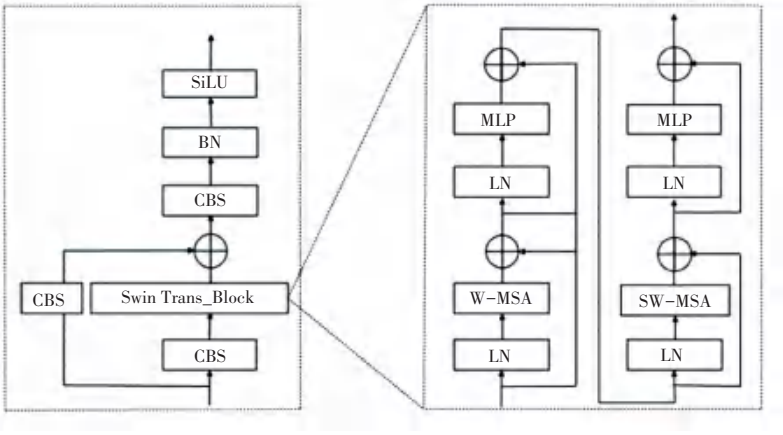

图 7　C3STR 与 Swin Transformer 结构图

Fig. 7　Structure diagram of C3STR and Swin Transformer

与传统的 Transformer 结构相比，C3STR 模块划分多个局部窗口控制计算区域，而且引入了滑动窗口多头注意力层，实现信息在相邻窗口的传递，降低网络计算量的同时也不会隔绝不同窗口的信息交流。

2.3.2 ECA 注意力机制

在 YOLOv5 的 C3 模块后面加入 ECA 注意力模块,能对输入特征图进行通道特征加强,避免计算时丢失部分有用特征,而且在不增加大量参数的情况下,提高了检测网络的性能。

ECA 原理如图 8 所示。图 8 中,首先将输入特征图进行 GAP 全局平均池化,再通过尺寸为 K 的 1×1卷积,使相邻层通道进行信息交互,共享权重[28]。最后使用 *Sigmoid* 函数获得每个通道的权重比例,将输入特征图与处理好的特征图权重相乘,获得特征信息。

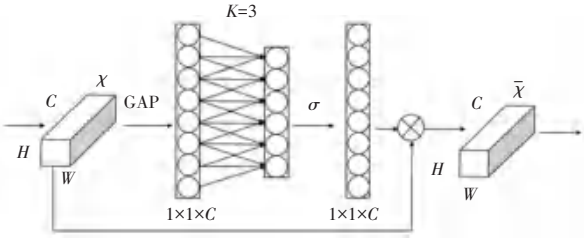

图 8 ECA 注意力模块结构图

Fig. 8 ECA attention module structure

2.4 Neck 轻量化改进

2.4.1 深度可分离卷积模块

本文将 Neck 网络的普通卷积替换成深度可分离卷积(Dwconv),该卷积分深度卷积与逐点卷积两步实现,如图 9 所示。深度卷积主要进行滤波,图 9 中深度卷积层为 3 个单通道、3×3 大小的卷积,每个深度卷积作用于特征图的单通道,然后融合其输出特征图。逐点卷积进行通道转换,图 9 中为 4 个 3 通道、1×1 大小来实现模型轻量化。

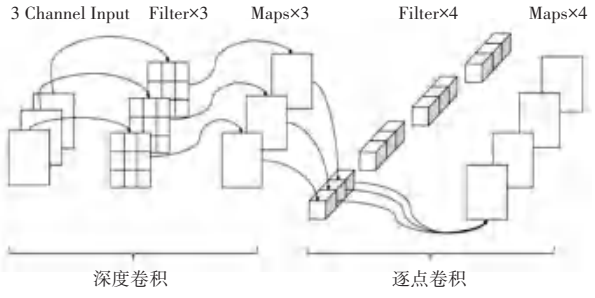

图 9 深度可分离卷积

Fig. 9 Deep separable convolution

2.4.2 Ghost-Net

由于嵌入式设备计算资源的规模,往往需要降低所部署神经网络的大小和计算资源的占用。Ghost-Net[29]是由华为公司诺亚实验室提出的轻量级卷积网络,可以在最大限度减少网络计算损耗的同时提高网络的检测准确率。

Ghost 模块将原有的卷积操作分成 2 个阶段。其中,第一阶段通过卷积计算生成部分特征层,第二阶段则是将第一阶段的输出进行分块单独线性卷积,生成特征层,最后将其组合得到大量的特征图[30],如图 10 所示。

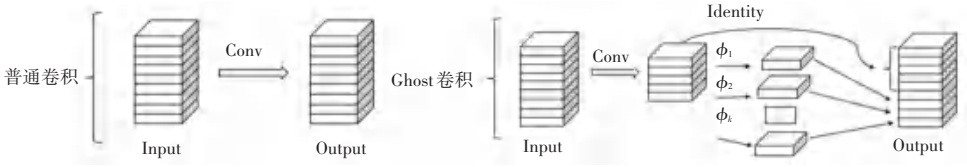

图 10 普通卷积与 Ghost 卷积结构图

Fig. 10 Structure diagram of ordinary convolution and Ghost convolution

Ghost-Net 由多个 Ghost-Bottleneck 组成,其中 Ghost-Bottleneck 由 2 个 Ghost 模块构成,其结构图与 YOLOv5 网络中的 C3 模型结构相似,本文将 neck 网络中的 C3 模型中的普通卷积替换成 Ghost-Net,其结构如图 11 所示,改进后的结构为 C3Ghost。

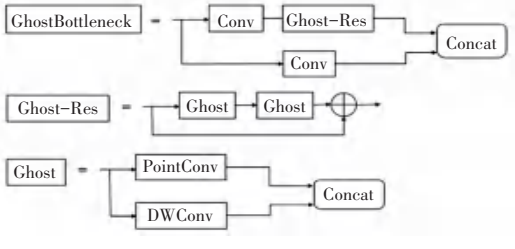

图 11 Ghost-Bottleneck 结构图

Fig. 11 Structure of Ghost-Bottleneck

2.5 *SIoU* 边框回归损失函数

损失函数是一种衡量模型预测结果准确度的方法。YOLOv5 使用 *CIoU Loss* 作为回归损失函数,该方法忽略了目标框与预测框的向量角度关系,这会导致收敛速度较慢。针对以上不足,本文摒弃 *CIoU* 损失函数,而是采用 *SIoU Loss*[31] 损失函数,如式(8)、式(9) 所示:

$$SIoU = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{8}$$

$$IoU = \frac{| B \cap B^{GT} |}{| B \cup B^{GT} |} \tag{9}$$

其中,B^{GT}、B 分别表示真实框和预测框;Ω 表示形状样本;Δ 为重新定义的距离样本。Ω、Δ 的定义公式具体如下:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^{\theta} \tag{10}$$

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \tag{11}$$

其中，$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$；$w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$；$\theta$ 表示对 Ω 的关注程度；$\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w}\right)^2$；$\rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h}\right)^2$，进一步可以得到：

$$\gamma = 1 + 2\sin^2\left(\arcsin\frac{|b^{gt} - b_{cy}|}{\sqrt{(b_{cx}^{gt} - b_{cx})^2 + (b_{cy}^{gt} - b_{cy})^2}} - \frac{\pi}{4}\right) \tag{12}$$

其中，b_{cx}^{gt}、b_{cy}^{gt} 和 b_c、b_{cy} 分别表示真实框与预测框的中心坐标；w^{gt}、h^{gt} 和 w、h 分别表示真实框与预测框的宽度和高度。*SIoU Loss* 考虑了目标框与预测框的向量角度关系，改写了损失函数，加快了网络的收敛速度，从而提高了算法精度。

3　实验及结果分析

3.1　数据集处理

由于目前含不正确佩戴口罩样本的公开口罩数据集较少，本文采用自制的数据集，包括由网络搜集、摄像头拍摄等方式收集的 8 500 张人脸图片，并分为正确佩戴、没有佩戴和不正确佩戴口罩等 3 类。通过 Labelimg 软件对数据集进行标注后，转换成 YOLO 格式的 txt 文件后，将数据集按照 9∶1 分为训练集和测试集进行训练。部分数据集图片和标注图片如图 12 所示。

(a) 数据集图片

(b) 标注图片

图 12　自制数据集图片

Fig. 12　Image of self-made dataset

3.2　实验环境

本实验平台的操作环境为 Linux 64 位系统，CPU 为 Intel Core i7-10700CPU @ 2.90 GHz，内存 32 GB，显卡 NVIDIA A100-PCIE-40 GB，搭配 CUDA11.6 和 cudnn8.4.1，选用 Pytorch1.12.0 深度学习框架，以及 Python3.7.5。

3.3　评价指标

为全面定量分析算法的检测性能，本文采用 3 类模型评价指标。第一类衡量模型的检测精度，包括精确度（*Precision*，*P*）、召回率（*Recall*，*R*）、平均精度（Average Precision，*AP*）、平均精度均值（mean Average Precision，*mAP*）、*P* - *R* 曲线（Precision - Recall Curve）；第二类衡量模型速度、即每秒检测帧数（Frames Per Second，*FPS*）[32]；第三类是衡量模型大小，包括权重文件大小、参数量大小（parameters）。平均精度 *AP* 由 *P* - *R* 曲线的面积确定，而精确度 *P*、召回率 *R*、平均精度均值 *mAP* 计算公式见如下：

$$P = TP/(TP + FP) \tag{13}$$

$$P = TP/(TP + FN) \tag{14}$$

$$AP = \int_0^1 P(R)\,\mathrm{d}r \tag{15}$$

$$mAP = (AP_{\text{mask}} + AP_{\text{nomask}} + AR_{\text{wrongmask}})/3 \tag{16}$$

其中，*TP* 表示被正确预测的样本数；*FP* 表示被错误预测的样本数；*FN* 表示未被检测出的样本数。

3.4　实验结果与分析

3.4.1　ECA 模块添加数量与位置分析

确定在网络中添加高效的 ECA 注意力模块后，需确定 ECA 模块添加的位置及数量[33]，ECA 模块添加的位置及数量会影响口罩佩戴特征提取效果，因此，本文对其做如下实验分析，以达到更好的检测效果。实验 1～3 在 Backbone 网络 C3 模块后分别添加 2、3、4 个 ECA 结构；实验 4 在 Neck 网络中增加 4 个 ECA 结构；实验 5 在 Backbone、Neck 网络分别添加 2、3 共计 5 个 ECA 结构。实验结果见表 1。

表 1　ECA 数量与位置实验对比结果

Table 1　Comparison results of ECA quantity and location experiment

实验编号	ECA 数量	mAP_{50}/%	AP_{mask}/%	AP_{nomask}/%	$AP_{\text{wrongmask}}$/%
1	2	94.9	**94.8**	93.9	96.0
2	3	94.9	94.0	94.0	96.6
3	4	94.6	93.5	94.1	96.2
4	4	93.9	92.6	93.3	95.7
5	5	**95.2**	94.0	**94.5**	**97.2**

从实验 1～3 可看出，ECA 仅添加在 Backbone 网络中时，口罩佩戴的平均检测精度随着融合 ECA 的数量增加而有所降低。从实验 3、4 的结果可知，口罩佩戴的平均检测精度与 ECA 添加的位置也相

关,在 Neck 网络中添加 4 个 ECA 模块比在 Backbone 网络中添加的 *mAP* 要低 0.7%。而实验 5 在 Backbone 网络中加入 2 个 ECA 模块的同时,在 Neck 网络中加入 3 个 ECA 模块,实现了在不增加较大参数量的前提下,获得较高的精度。

3.4.2 损失函数对比实验

为检测使用 *SIoU* 损失函数是否有更好的效果,将原始 *CIoU* 损失函数与本文所提出的 *SIoU* 损失函数进行对比实验,实验结果见表 2。

表 2 不同损失函数对比

Table 2 Comparison of different loss functions

损失函数	mAP_{50}/%	*FPS*
CIoU loss	94.00	51.28
SIoU loss	**94.60**	51.28

在表 2 的实验中,仅替换了算法的损失函数。从结果可看出在相同的检测速度情况下,模型 *mAP* 提升了 0.60%,即 *SIoU* 损失函数更具优势,更有利于提升模型性能。

3.4.3 消融实验

为检测 Mixup 数据增强方式、*Mish* 激活函数以及 C3STR 模块、DW 卷积模块、C3Ghost 模块的优劣性,本文设置了消融实验,见表 3。表 3 中,"×"表示未使用该方法,"√"表示使用了该方法。表 3 中,"①"表示 YOLOv5s 标准网络;"②"表示用 *SIoU* 替换原始的边框损失函数;"③"表示采 Mixup 数据增强方式;"④"表示将原始激活函数替换为 *Mish* 函数;"⑤"表示融入 Swin Transformer 结构,使用 C3STR 替换部分 C3 模块;"⑥"表示加入 ECA 注意力机制;"⑦"将 Neck 层卷积替换为 DWcon 模块;"⑧"表示将 Neck 层 C3 替换为轻量级的 C3Ghost 模块。

表 3 消融实验设计方案

Table 3 Design scheme of ablation experiment

编号	YOLOv5s	*SIoU*	Mixup	Mish	C3STR	ECA	DWcon	C3Ghost
①	√	×	×	×	×	×	×	×
②	√	√	×	×	×	×	×	×
③	√	√	√	×	×	×	×	×
④	√	√	√	√	×	×	×	×
⑤	√	√	√	√	√	×	×	×
⑥	√	√	√	√	√	√	×	×
⑦	√	√	√	√	√	√	√	×
⑧	√	√	√	√	√	√	√	√

接下来,研究得到的平均精度对比结果见表 4。模型规模与检测速度对比结果见表 5。由表 4、表 5 可知,在算法①基础上,将 YOLOv5s 的 *CIoU Loss* 换成 *SIoU Loss* 后,即成为算法②,*mAP* 由 94.0%提升到 94.6%,增加了 0.6%,算法③在算法②基础上,加入了 Mixup 数据增强方式,在不增加参数量和模型体积的情况下,*mAP* 继续增加了 0.7%;算法④使用 *Mish* 替换原始的激活函数,在保持参数量不变的情况下,与算法 ① 相比,*mAP* 提高了 2.3%,达到了 96.3%;算法⑤将主干网络后两层的 C3 替换为 C3STR 结构,使得参数量由 7 018 216 降为 4 826 994,模型体积减少了 4.3 M,检测速度达到了 53.7 fps;算法⑥加入了 5 个 ECA 高效注意力模块,*mAP* 由 94.0%提升到 97.2%,增加了3.2%;算法⑦继续将 Neck 网络的普通卷积替换为 DWconv;算法⑧将 C3 替换为 C3Ghost,最终 *mAP* 达到了 96.9%,较 YOLOv5s 网络提高了 2.9%,参数量减少为原来的 45.8%,模型体积减少为原来的47.9%,检测速度达到了 53.3 fps,达到了实时检测的要求。

表 4 平均精度对比

Table 4 Comparison of average accuracy

模型编号	mAP_{50}/%	AP_{mask}/%	AP_{nomask}/%	$AP_{wrongmask}$/%
①	94.0	92.8	93.3	96.0
②	94.6	93.5	94.8	95.6
③	95.9	95.5	95.6	96.5
④	96.3	95.7	96.1	97.2
⑤	96.4	96.2	96.1	97.0
⑥	**97.2**	**96.9**	**97.3**	97.5
⑦	96.9	96.5	96.6	**97.7**
⑧	96.9	96.4	96.9	97.3

表 5 模型规模与检测速度对比结果

Table 5 Comparison results of model scale and detection speed

模型编号	Parameters	Weight/MB	*FPS*
①	7 018 216	14.4	51.2
②	7 018 216	14.4	49.0
③	7 018 216	14.4	48.8
④	7 018 216	14.4	48.8
⑤	4 826 994	10.1	53.7
⑥	4 827 009	10.1	50.5
⑦	4 178 945	8.8	51.0
⑧	**3 216 657**	**6.9**	**53.3**

3.4.4 主流算法结果分析

为展示本文算法在口罩检测场景的有效性,将本文的改进算法模型与单阶段的 YOLOv5s、SSD、YOLOv3[34] 等主流算法以及两阶段的主流算法 Faster RCNN 进行对比,在本文自制口罩佩戴数据集下,设置 *epoch* = 100,*batchsize* = 8 进行模型训练。

使用口罩佩戴检测的各个类别 *AP*、平均精度的均值 *mAP* 以及检测速度 *FPS* 衡量实验效果。

对比实验的结果见表 6。本文提出的改进算法比单阶段主流算法 YOLOv5s、SSD、YOLOv3 的口罩检测 *mAP* 分别提高了 2.9%、9.5%、4.2%，*FPS* 达到了 53.3，满足实时口罩佩戴检测。与两阶段的主流检测算法 Faster RCNN 相比，口罩检测的 *mAP* 提高了 19.6%。由此可见，改进后的算法给出了更高的检测精度与速度，更适合公共场合中的实时口罩佩戴检测。

表 6　主流目标检测算法检测性能对比

Table 6　Comparison of detection performance of mainstream target detection algorithms

模型	mAP_{50}/%	AP_{mask}/%	AP_{nomask}/%	$AP_{wrongmask}$/%	*FPS*
Faster RCNN	77.3	81.0	69.0	82.0	33.9
SSD	87.4	91.7	75.2	95.4	53.0
YOLOv3	92.7	93.8	88.8	95.6	48.6
YOLOv5s	94.0	92.8	93.3	96.0	51.2
Ours	**96.9**	**96.4**	**96.9**	**97.3**	**53.3**

3.4.5　测试图片效果对比

图 13 是 YOLOv3、YOLOv5s 与本文改进算法在遮挡严重的密集人脸口罩佩戴场景实际场景下的口罩检测效果对比图。可以看出与 YOLOv3、YOLOv5s 相比，本文算法引入高效注意力 ECA 模块后，图 13 中的左边两位未佩戴口罩男士的口罩检测置信度得分高于其他模型。因此，本文提出的口罩检测方法适合于公共场合的密集场景口罩佩戴检测任务。

(a) 原图　(b) YOLOv3

(c) YOLOv5　(d) 本文 YOLOv5 改进算法

图 13　不同方法检测效果对比

Fig. 13　Comparison of detection effects of different methods

4　结束语

科学正确佩戴口罩能极为有效地阻断呼吸道传染病的传播，在重点场所实施口罩佩戴实时监测至关重要。为提高实时检测口罩佩戴的效率，降低部署成本，本文提出了一种基于 YOLOv5s 的口罩佩戴检测改进算法。首先针对主流口罩检测算法对复杂场景中口罩检测存在误检、漏检的问题，将 Swin Transformer 结构和 ECA 注意力模块融入 YOLO 主干网络，以增强算法的目标提取能力；其次，将 *CIoU* 替换为 *SIoU* 边框损失函数，提升检测精度；使用轻量级 Mixup 数据增强方式和 *Mish* 激活函数，提高模型的泛化能力；最后，考虑到部署时模型的规模问题，分别用 Dwconv 卷积和轻量的 C3Ghost 模块代替原框架下 Neck 网络的卷积和 C3 模块，大幅降低网络的参数量，保证 *mAP* 的同时，实现口罩检测模型的轻量化。针对实际中存在不规范佩戴口罩的场景，增加了此类别数据集，对所设计的算法与

YOLOv5s、SSD、YOLOv3、Faster RCNN 等主流算法进行了比较。结果表明,改进后模型的平均精度均值提高了 2.9%,达到了 96.9%,检测速度达到了 53.3 fps,参数量减少为原来的 45.8%,模型体积减少为原来的 47.9%,很好地平衡了检测速度与精度。同时,相比于原始 YOLOv5s 算法,改善了错检和漏检的情况;相比于其他主流单阶段和两阶段算法,新算法在检测速度和精度上都具有更大优势,为后续应用于嵌入式设备中提供一种有效的技术手段。

参考文献

[1] 龚震宇,龚训良. 世界卫生组织关于 2023—2024 年流感季节北半球使用流感疫苗病毒株组成成分推荐[J]. 预防医学, 2023, 35(5): 460.

[2] 中华预防医学会新型冠状病毒肺炎防控专家组. 新型冠状病毒肺炎流行病学特征的最新认识[J]. 中国病毒病杂志, 2020, 10(2): 86-92.

[3] 王欣然,田启川,张东. 人脸口罩佩戴检测研究综述[J]. 计算机工程与应用, 2022, 58(10): 13-26.

[4] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//14th European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.

[5] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020,42(2):318-327.

[6] 魏明军,周太宇,纪占林,等. 基于 YOLOv3 的公共场所口罩佩戴检测方法[J]. 广西师范大学学报(自然科学版), 2023, 41(1): 76-86.

[7] 曹城硕,袁杰. 基于 YOLO-Mask 算法的口罩佩戴检测方法[J]. 激光与光电子学进展, 2021, 58(8): 211-218.

[8] 谈世磊,别雄波,卢功林,等. 基于 YOLOv5 网络模型的人员口罩佩戴实时检测[J]. 激光杂志, 2021, 42(2): 147-150.

[9] 王艺皓,丁洪伟,李波,等. 复杂场景下基于改进 YOLOv3 的口罩佩戴检测算法[J]. 计算机工程, 2020, 46(11): 12-22.

[10] 赵文清,康怿瑾,赵振兵,等. 改进 YOLOv5s 的遥感图像目标检测[J]. 智能系统学报, 2023, 18(1): 86-95.

[11] 张烈平,李智浩,唐玉良. 基于迁移学习的轻量化 YOLOv2 口罩佩戴检测方法[J]. 电子测量技术, 2022, 45(10): 112-117.

[12] 薄景文,张春堂. 基于 YOLOv3 的轻量化口罩佩戴检测算法[J]. 电子测量技术, 2021, 44(23): 105-110.

[13] 王艺霏,贺利乐,何林. 基于 YOLOv4 的轻量化口罩佩戴检测模型设计[J]. 西北大学学报(自然科学版), 2023, 53(2): 265-273.

[14] 彭成,张乔虹,唐朝晖,等. 基于 YOLOv5 增强模型的口罩佩戴检测方法研究[J]. 计算机工程, 2022, 48(4): 39-49.

[15] 王圣雄,刘瑞安,燕达. 基于改进 YOLOv5 的轻量型口罩佩戴检测算法[J]. 计算机时代, 2023(5): 109-112.

[16] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 580-587.

[17] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[18] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy:IEEE , 2017: 2961-2969.

[19] 李泽琛,李恒超,胡文帅,等. 多尺度注意力学习的 Faster R-CNN 口罩人脸检测模型[J]. 西南交通大学学报, 2021, 56(5): 1002-1010.

[20] 任钰,刘全金,黄忠,等. 基于 Faster R-CNN 与迁移学习的口罩佩戴检测算法[J]. 安庆师范大学学报(自然科学版), 2021, 27(4): 25-30.

[21] 刘玉国,张晶. 基于改进的 Faster R-CNN 的行人口罩检测[J]. 现代计算机, 2021, 27(26): 73-76.

[22] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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.

[23] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA :IEEE, 2017: 2117-2125.

[24] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 8759-8768.

[25] 李昂,孙士杰,张朝阳,等. 改进 YOLOv5s 的轨道障碍物检测模型轻量化研究[J]. 计算机工程与应用, 2023, 59(4): 197-207.

[26] LIU Ze, LIN Yuyong, CAO Yue, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2021: 10012-10022.

[27] 谢静,杜耀文,刘志坚,等. 基于轻量化改进型 YOLOv5s 的可见光绝缘子缺陷检测算法[J]. 电网技术, 2023, 47(12): 5273-5283.

[28] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2020: 11531-11539.

[29] HAN Kai, WANG Yunhe, TIAN Qi, et al. Ghostnet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2020: 1580-1589.

[30] HAN Kai, WANG Yunhe, TIAN Qi, et al. Ghostnet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2020: 1577-1586.

[31] GEVORGYAN Z. SIoU Loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv: 2205.12740, 2022.

[32] 张明路,郭策,吕晓玲,等. 改进的轻量化 YOLOv4 用于电子元器件检测[J]. 电子测量与仪器学报, 2021, 35(10): 17-23.

[33] 周旗开,张伟,李东锦,等. 基于改进 YOLOv5s 的光学遥感图像舰船分类检测方法[J]. 激光与光电子学进展, 2022, 59(16): 476-483.

[34] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767,2018.