

融合 ECA 机制的轻量化 YOLOv4 检测模型

刘雅楠, 李维乾

(西安工程大学 计算机科学学院, 陕西 西安 710048)

摘要:近年来,卷积神经网络已在人脸识别、无人驾驶等领域取得重大突破。随着智能移动设备的普及,高精度的大型网络往往伴随着参数量多、计算量大等问题,无法部署在这些资源有限的移动设备平台上。GhostNet 通过简单的线性操作生成更多特征映射,可大幅减少计算成本。为此,提出了一种改进轻量化 YOLOv4 的 GhostNet-YOLOv4 网络模型,该模型将 YOLOv4 的主干网络替换为 GhostNet 残差结构,借助即插即用的 Ghost 模块升级卷积神经网络,并使用 Mosaic 数据增强技术对数据集进行预处理,融合 ECA 机制,加入 Focal Loss 焦点损失函数,在保证一定精度的前提下大幅减少了模型的参数量和计算量。相对于改进前的 GhostNet-YOLOv4 模型,改进后的 GhostNet-YOLOv4 在 PASCAL VOC 2007 数据集上的 mAP (mean Average Precision) 提高 1.65 个百分点,达到 85.09%,且模型参数量只有 11.429 M,相对于原 YOLOv4 模型减少了约 80%,具有更好的综合性能。

关键词:目标检测; YOLOv4; GhostNet; 轻量化神经网络; 注意力机制

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2023)07-0146-08

doi:10.3969/j.issn.1673-629X.2023.07.022

Lightweight YOLOv4 Detection Model Incorporating ECA Mechanism

LIU Ya-nan, LI Wei-qian

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: In recent years, convolutional neural networks have made breakthroughs in applications such as face recognition and autonomous driving. With the popularity of intelligent mobile devices, high-precision large-scale networks are often accompanied by problems such as a large number of parameters and a large amount of computing, and cannot be deployed on these resource-limited mobile device platforms. GhostNet generates more feature maps with simple linear operations, drastically reducing the computational cost. Therefore, we propose an improved lightweight YOLOv4 network model. The backbone network of YOLOv4 is replaced with the GhostNet residual structure, which can upgrade the convolutional neural networks, and the Mosaic data enhancement is used to pre-process the dataset. ECA mechanism is integrated and Focal Loss function is added. On the premise of ensuring certain accuracy, the number of parameters and calculation amount of the model are greatly reduced. Compared with the GhostNet-YOLOv4 before the improvement, the mAP (mean Average Precision) of the improved GhostNet-YOLOv4 on the PASCAL VOC 2007 dataset is increased by 1.65% to 85.09%, and the parameters of the model are only 11.429 M, which reduces by about 80% compared with the original YOLOv4 network, which indicates that the improved model has better overall performance.

Key words: object detection; YOLOv4; GhostNet; lightweight neural network; attention mechanism

0 引言

目标检测作为计算机视觉领域的重要分支,已广泛应用于无人驾驶、行人检测、人脸识别等领域。基于深度学习的目标检测算法通过训练神经网络模型,由浅及深自动提取目标特征,具有更强的鲁棒性,以快速、高效、准确的特点逐渐超越传统算法成为主流。当前目标检测主要分为一阶段和两阶段算法。其中,以

YOLO^[1-3]系列算法为代表的一阶段目标检测直接生成候选框,通过对候选框进行回归生成预测框;以 RCNN^[4-6]系列算法为代表的两阶段目标检测先生成包含感兴趣目标的候选区域,然后对候选区域的候选框进行回归生成预测框。一阶段算法相对于二阶段算法准确率低,但提高了检测速度。

如今,处理计算机视觉任务的总趋势是制作更深、

收稿日期:2022-09-20

修回日期:2023-01-30

基金项目:教育部重点实验室开放基金(NS202118901)

作者简介:刘雅楠(2000-),女,硕士研究生,研究方向为深度学习、目标检测;通讯作者:李维乾(1980-),男,副教授,博士,CCF会员(88255M),研究方向为深度学习、智能决策等。

更复杂的网络,以达到更高的精度。而更深的网络模型通常伴随着巨大的参数量,对存储和计算资源有较高要求。因此,设计更轻量、高性能的网络模型是未来的发展趋势。轻量化神经网络架构的设计主要包括三个不同方向:人工设计轻量化网络模型、基于神经网络架构搜索的自动化神经网络设计、神经网络模型的压缩^[7]。近年来,相关学者设计了一些高效的网络模型,例如 Iandola 等^[8]提出的 SqueezeNet 模型,该模型是将 3×3 卷积替换为 1×1 卷积,通过减少通道数来减少计算量和参数量;Zhang Xiangyu 等^[9]提出了 ShuffleNet 模型,采用深度可分离卷积和通道重排操作代替 ResNet block 相应的层,从而减少计算量,提高效率,解决了因多个 group convolution 叠加造成的边界效应问题;Howard 等^[10]提出了 MobileNet 模型,该模型采用大量的深度可分离卷积设计的神经网络,大幅度减少了模型的参数量和计算量。文献[11-12]分别将 MobileViT 和 MobileNetV3 结构作为主干特征提取网络,通过深度可分离卷积大幅减少主干网络参数量,然而上述改进的轻量化模型对于算力受限的设备仍过于庞大,且此类深度可分离卷积网络模型由于 GPU 访存带宽的限制,读写数据过程占用了大量时间。

原始的 YOLOv4 模型以 CSPDarknet53 作为骨干网络,复杂的网络结构和更深的网络层次导致模型占用内存增大,对设备的计算能力也提出了更高的要求。为实现网络轻量化,本文针对目标检测网络计算参数量大、运算复杂度高、小目标检测效果差等问题,设计了一种改进的 GhostNet-YOLOv4 网络模型,将原主干替换为 GhostNet 残差结构,来降低计算量和参数量,提高运行效率。该方法简化了网络结构,有效减少模型计算量,改善模型的检测效果;融合 ECA 注意力机制模块,并引入 Focal Loss,增强模型的特征提取,提高模型的检测精度。

1 相关网络模型

1.1 GhostNet

1.1.1 Ghost 模块

在卷积神经网络中,生成的输入图像的一些特征映射存在许多相似的特征映射对,就像彼此的幻影。一般情况下,这些大量存在的、丰富甚至冗余的特征映射信息,保证了对输入数据的全面理解。因此,作为卷积网络的一个重要组成部分,不能直接将冗余特征去除^[13]。然而,这些冗余也伴随着巨大的参数量和计算量,由于内存和计算资源有限,很难在嵌入式设备上部署。Ghost 模块可以基于一组内在特征映射,应用一系列线性变换生成许多能够充分揭示内在特征背后信息的重影特征映射,以一种高效的、低成本的方式生成

足够的特征映射,从而作为即插即用组件来升级现有的卷积神经网络。

普通卷积模块直接通过卷积得到所有输出信息,如图 1(a)所示;Ghost 模块整体思想分两步:第一,先使用较少的卷积核生成一部分特征映射,如图 1(b)中先通过卷积生成少量特征映射;第二,对生成的固有特征映射应用轻量操作(包括深度卷积、分组卷积或其他方式的卷积)或恒等变换等一系列简单的线性操作,生成更多的特征映射,保留冗余信息;最后,将生成的特征映射和最开始的固有特征进行恒等变换拼接在一起,如图 1(b)中的少量特征映射经过一系列操作得到更多特征映射,二者拼接得到所有输出信息。

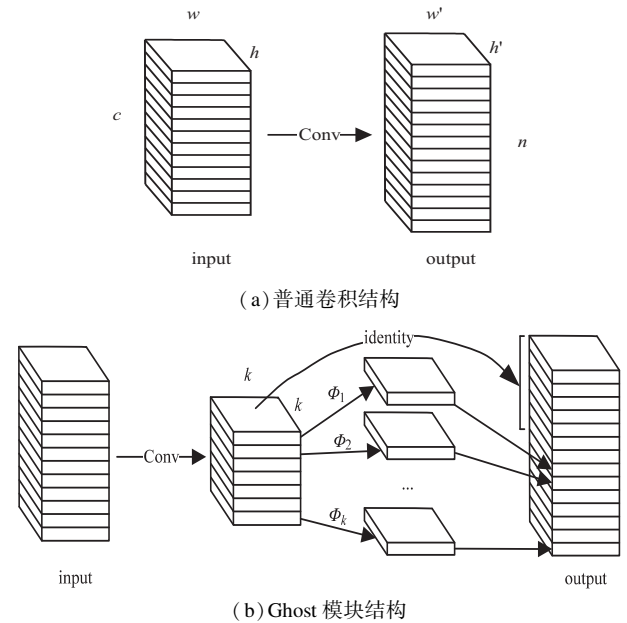


图1 普通卷积结构和 Ghost 模块结构

假定模型输入尺寸为 $h \times w \times c$, 输出为 $h' \times w' \times n$, 卷积核尺寸为 k , Ghost 模块中的轻量操作采用深度卷积,分为 s 个部分。

则普通卷积的计算量为:

$$h' \times w' \times n \times k \times k \times c \quad (1)$$

而采用 Ghost 模块的计算量为:

$$h' \times w' \times \frac{n}{s} \times k \times k \times c + (s-1)h' \times w' \times \frac{n}{s} \times k \times k \times 1 \quad (2)$$

可以得出使用 Ghost 模块后压缩率约为 s :

$$\frac{[h' \times w' \times n \times k \times k \times c]}{[h' \times w' \times \frac{n}{s} \times k \times k \times c + (s-1)h' \times w' \times \frac{n}{s} \times k \times k \times 1]} \approx s \quad (3)$$

因此,与普通卷积相比,Ghost 模块在不改变输出特征映射大小的情况下,大幅降低了计算成本。

1.1.2 Ghost Bottleneck

Ghost Bottleneck 结构类似于 ResNet 中的基本残

差块,集成了几个卷积层和 shortcut。该部分主要由两个堆叠的 Ghost 模块组成,两个模块之间用 shortcut 连接,包括 stride 为 1 和 stride 为 2 两种情况,如图 2 所示。当 stride 为 1 时,第一个 Ghost 模块充当扩展层,用于增加通道数,第二个 Ghost 模块用于减少通道数,以匹配 shortcut,shortcut 用来连接这两个模块的输入和输出,且第二个模块不使用 ReLU 激活函数;当 stride 为 2 时,shortcut 路径由下采样层和 stride 为 2 的深度卷积实现,用于加速卷积操作。此外,Ghost 模块中的初始卷积均为 1×1 卷积,用于提高效率。

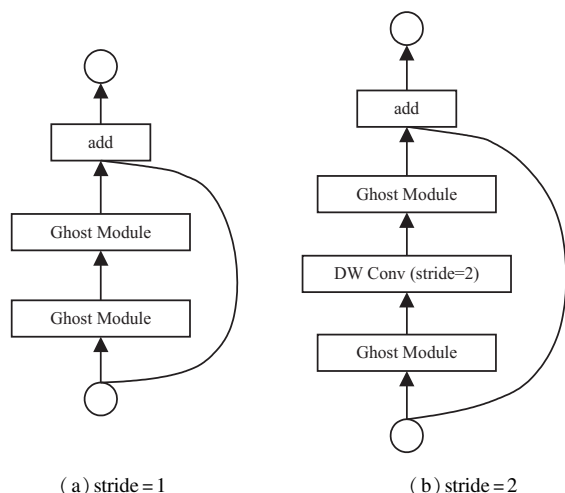


图 2 Ghost Bottleneck 结构

1.1.3 GhostNet

GhostNet 是基于 Ghost Bottleneck 建立的一种高效神经网络结构,第一层是具有 16 个滤波器的标准卷积层,然后通过一系列 Ghost Bottleneck 逐渐压缩特征映射的尺寸,扩大通道数。这些 Ghost Bottleneck 根据输入特征图的大小分为六个不同的阶段,见表 1 中输入为 (208,208,16) (104,104,24) (52,52,40) (26,26,80) (26,26,112) (13,13,160) 时的六个阶段,除了前五个阶段中每个阶段的最后一个瓶颈步长为 2 之外,其余各部分 GhostNet Bottleneck 的步长均为 1。整体结构参数如表 1 所示。

表 1 GhostNet 结构参数

输入	操作	输出	stride
416, 416, 3	Conv2d	208, 208, 16	2
208, 208, 16	Ghost Bottleneck	208, 208, 16	1
208, 208, 16	Ghost Bottleneck	208, 208, 16	2
104, 104, 24	Ghost Bottleneck	104, 104, 24	1
104, 104, 24	Ghost Bottleneck	52, 52, 40	2
52, 52, 40	Ghost Bottleneck	52, 52, 40	1
52, 52, 40	Ghost Bottleneck	26, 26, 80	2
26, 26, 80	Ghost Bottleneck	26, 26, 80	1
26, 26, 80	Ghost Bottleneck	26, 26, 80	1
26, 26, 80	Ghost Bottleneck	26, 26, 80	1

续表 1

输入	操作	输出	stride
26, 26, 80	Ghost Bottleneck	26, 26, 112	1
26, 26, 112	Ghost Bottleneck	26, 26, 112	1
26, 26, 112	Ghost Bottleneck	13, 13, 160	2
13, 13, 160	Ghost Bottleneck	13, 13, 160	1
13, 13, 160	Ghost Bottleneck	13, 13, 160	1
13, 13, 160	Ghost Bottleneck	13, 13, 160	1
13, 13, 160	Ghost Bottleneck	13, 13, 160	1
13, 13, 160	Ghost Bottleneck	13, 13, 160	1

1.2 YOLOv4 模型

当前主流的目标检测网络模型主要由主干网络 Backbone、颈部网络 Neck 和头部网络 Head 构成。其中,主干网络用于特征提取,颈部网络用于加强特征提取,头部网络用于预测。YOLOv4 模型^[14]的主干网络采用 CSPDarknet53, CSPDarknet53 是在 Darknet53 的基础上融入了 CSPNet 结构,由五个残差块构成,分别包含 1、2、8、8、4 个残差单元,用于提取输入数据的特征;加强特征提取网络由 SPPNet 和 PANet 构成,用于将主干网络得到的三个初步特征层进行特征融合,从而获得更有效的三个特征层。当主干网络的最后一个特征层完成三次卷积操作后,SPPNet 对最后一个特征层进行四个不同尺度的最大池化操作来增加感受野,分离出显著的上下文特征。随后将提取得到的特征层与主干网络得到的另两个特征层送入 PANet,执行自下而上和自上而下的双向特征融合,实现特征的反复提取;头部网络则利用从 PANet 获得的多尺度特征进行回归和分类,最终输出的维度包含样本类别值、预测框在 x 轴和 y 轴的偏移量、预测框的高度和宽度及置信度。

2 改进的 GhostNet-YOLOv4 网络模型

目前为止,对图像进行实时目标检测面临着挑战和困难,首要就是深度神经网络存在复杂性和存储量高的问题。针对这个问题,该文所做的工作主要分为以下三个方面:

(1)改进前的 GhostNet-YOLOv4 模型以 YOLOv4 模型为基础,将 YOLOv4 原本的主干网络 CSPDarknet53 替换为 GhostNet 中的 Ghost 模块;

(2)在 GhostNet-YOLOv4 基础上进行模型优化,加入 ECA 注意力机制和 Focal Loss,得到改进后的 GhostNet-YOLOv4 模型,在不影响模型运行速度的前提下,提高模型的检测精度,具体网络结构如图 3 所示;

(3)将原 YOLOv4 模型、改进前后的 GhostNet-

YOLOv4 模型分别在 PASCAL VOC 2007 数据集上进行目标检测实验,验证该模型的优越性。

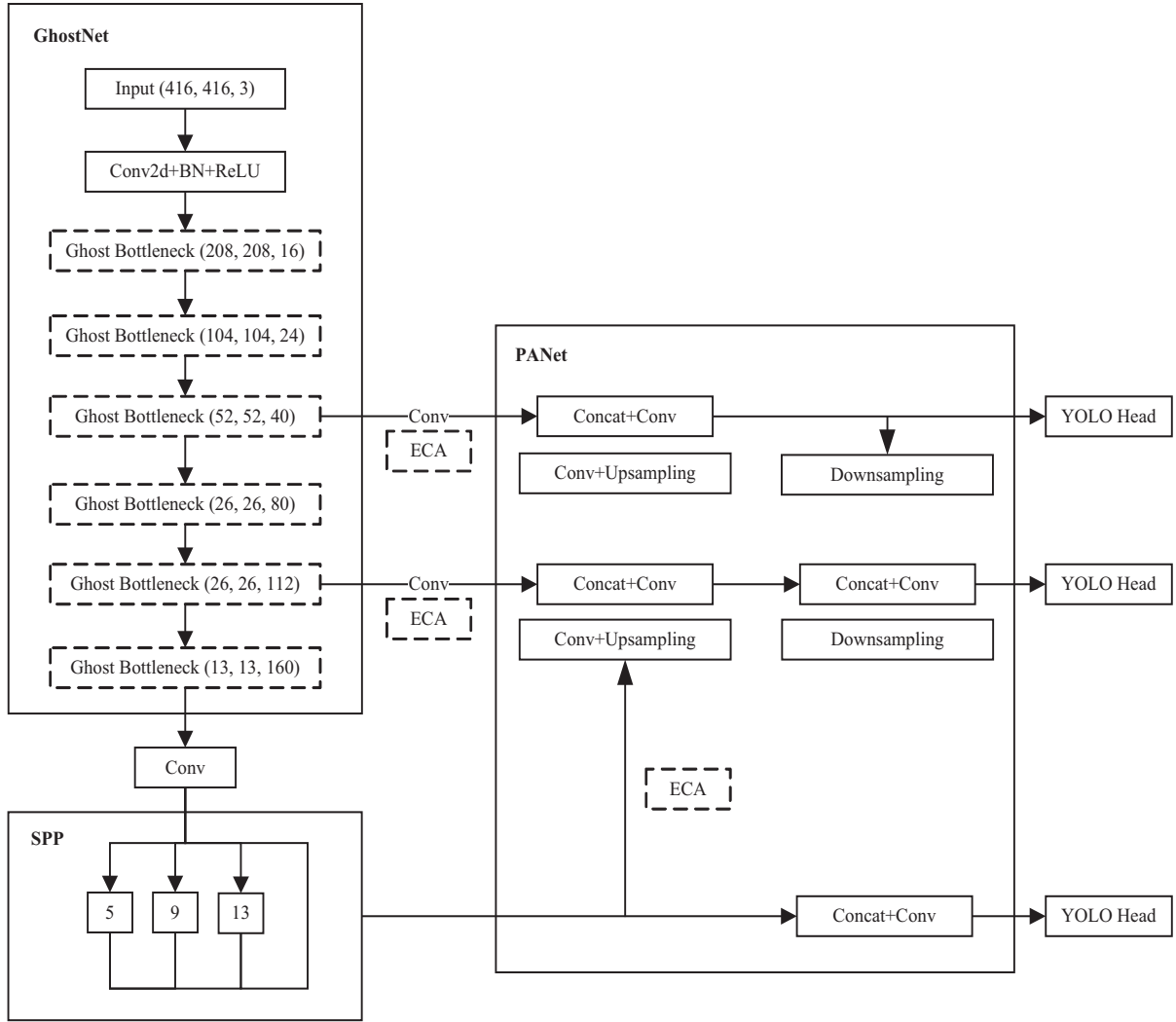


图 3 改进后的 GhostNet-YOLOv4 网络模型

2.1 ECA 注意力机制

近年来,通道注意机制在改善深度卷积神经网络性能方面显示出巨大的潜力。然而,大多数现有方法致力于开发更复杂的注意模块,以获得更好的性能,这不可避免地增加了模型的复杂性。为了克服性能和复杂性权衡的矛盾,ECA 注意力机制只涉及少量的参数,提出了一种无降维的局部交叉信道交互策略^[15],进行适当的跨通道交互,从而在显著降低模型复杂性的同时保持性能。通过实验表明 ECA 注意力机制效率更高,同时性能优于其他模块,在图像分类、目标检测和实例分割方面表现出卓越的能力。

ECA 模块通过全局平均池化将输入特征图的信息压缩为一个具有全局感受野的通道描述符,使用尺寸为 k 的一维卷积,从该通道描述符中获取局部通道之间的相关性,再利用 Sigmoid 激活函数根据通道间的相关性计算出各通道的注意力值,最后作用在输入的特征图上,实现对特征通道的重标定。

其中一维卷积核的尺寸 k 由通道维数的函数来自

适应确定,公式为:

$$k = \varphi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (4)$$

其中, C 表示通道数, odd 表示向上取最近的奇数,该文 γ 和 b 分别取 2 和 1。

2.2 损失函数

2.2.1 置信度损失函数

置信度损失函数如下所示:

$$L_{\text{obj}} = - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} [\hat{C}_j^i \log(C_j^i) + (1 - \hat{C}_j^i) \log(1 - \hat{C}_j^i)] - \lambda_{\text{noobj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} [\hat{C}_j^i \log(C_j^i) + (1 - \hat{C}_j^i) \log(1 - \hat{C}_j^i)] \quad (5)$$

其中, λ_{noobj} 表示参数因子,实验算法中取 1; $K \times K$ 代表输入图像的网格数,每个网格有三个 anchor,即 13×13 、 26×26 、 52×52 ; M 代表当前第几个 anchor,因此取值范围为 0、1、2; I_{ij}^{obj} 和 I_{ij}^{noobj} 代表当前 anchor 中是否检测到目标,对于 I_{ij}^{obj} ,若 anchor 中检测到目标则置 1,

反之则为 0; 对于 I_{ij}^{noobj} , 若 anchor 中检测到目标则置 0, 反之则为 1; C_j^i 表示第 i 个网格中第 j 个预测框置信度, \hat{C}_j^i 表示第 i 个网格中第 j 个 anchor 的置信度。

2.2.2 分类损失函数

分类损失函数如下所示:

$$L_{\text{cls}} = - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \alpha (1 - \hat{y})^\gamma [- \hat{p}_j^i(c) \log(p_j^i(c)) - (1 - \hat{p}_j^i(c)) \log(1 - p_j^i(c))] \quad (6)$$

其中, c 表示目标对应的类别, classes 表示所有类别; $p_j^i(c)$ 表示当前待检测目标中第 i 个网格中第 j 个预测框为 c 类别的概率, $\hat{p}_j^i(c)$ 表示当前待检测目标中第 i 个网格中第 j 个 anchor 为 c 类别的概率。

2.2.3 回归损失函数

回归损失函数如下所示:

$$L_{\text{loc}} = \lambda_{\text{iou}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (1 - \text{IoU} + \frac{d^2}{c^2} + \alpha v) \quad (7)$$

其中, $\alpha = \frac{v}{1 - \text{IoU} + v}$, $v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2$, λ_{iou} 表示参数因子, 文中算法中取 1, IoU 为交并比, w 、 h 表示检测框的宽、高, w^{gt} 、 h^{gt} 表示真实框的宽、高; c 表示被检测目标的真实框(图 4 中实线框)与预测框(图 4 中点线框)的最小外包矩形(图 4 中虚线框)的对角线长, d 表示真实框与预测框的中心点距离。

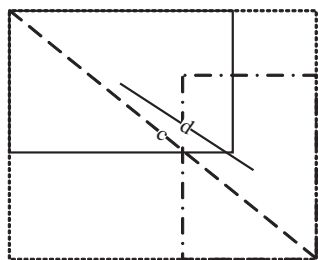


图 4 c 、 d 示意图

2.2.4 Focal Loss

一阶段算法的检测精度比不上两阶段, 主要原因是训练过程样本不平衡^[16]。样本不平衡包括正负样本不平衡、难易样本不平衡等。目前解决这类问题主要有两种方案, 一是 hard sampling 方法, 即从所有样本中选择一定量的正样本和负样本, 只计算被选择的样本的 loss, 一般会倾向选择一些困难负例样本, 比如 OHEM^[17]; 另外一类方法是 soft sampling 方法, 计算所有样本的 loss, 但是针对不同样本赋予不同权值, 比如 Focal Loss。

Focal Loss 由 Ross Girshick、He Kaiming 等于 2017 年提出。Focal loss 主要对交叉熵损失函数进行了改进, 解决负样本不平衡问题^[16]。

交叉熵如公式(8)所示。

$$\text{CE} = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1 - p), & \text{if } y = 0 \end{cases} \quad (8)$$

为解决难负样本不平衡问题, Focal Loss 中引入表达式(9), 使难样本的 loss 进一步增加, 负样本的 loss 进一步降低。

$$\text{FL} = \begin{cases} -(1 - p)^\gamma \log(p), & \text{if } y = 1 \\ -p^\gamma \log(1 - p), & \text{if } y = 0 \end{cases} \quad (9)$$

除此之外, Focal Loss 中还通过 α 来解决正负样本不平衡问题, 如公式(10)所示。

$$\text{FL} = \begin{cases} -\alpha \log(p), & \text{if } y = 1 \\ -(1 - \alpha) \log(1 - p), & \text{if } y = 0 \end{cases} \quad (10)$$

综上, Focal Loss 的综合表达式为:

$$\text{FL} = \begin{cases} -\alpha (1 - p)^\gamma \log(p), & \text{if } y = 1 \\ -(1 - \alpha) p^\gamma \log(1 - p), & \text{if } y = 0 \end{cases} \quad (11)$$

公式(9)~(11)中, γ 和 α 为超参数, 文中分别取 2 和 0.25。

3 实验及分析

3.1 数据集及参数设置

实验采用 PASCAL VOC 2007 数据集, 包括 aeroplane、bicycle、bird、boat、bottle、bus、car、cat、chair、cow、diningtable、dog、horse、motorbike、person、pottedplant、sheep、sofa、train、tvmonitor 共 20 个类别, 9 963 张图像, 其中训练集 5 011 张, 测试集 4 952 张, 累计标注目标 24 640 个。实验在 Windows 10 操作系统下进行, 使用 Python 3.7 进行编译, 开发平台为 PyCharm, 深度学习框架为 PyTorch, GPU 为 NVIDIA GeForce RTX 3090。

实验使用 SGD 优化器, 动量设置为 0.937, 初始学习率设置为 0.01, 学习率衰减方式采用余弦退火算法; 迭代次数为 300, batch size 为 128, 采用冻结训练, 50 个 epoch 后 batch size 为 64; 数据增强方式采用 Mosaic 增强, 在训练小模型时适当弱化^[18], 将 Mosaic 系数设置为 0.7。

3.2 性能评价指标

该文使用 AP 和 mAP 作为衡量检测精度的指标。在多类别物体检测中, 每一个类别都可以根据准确率 P (Precision) 和召回率 R (Recall) 绘制一条 PR 曲线。准确率和召回率的公式如下:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

Precision 和 Recall 是一对矛盾的度量, 一般来说, Precision 高时, Recall 值往往偏低; 而 Precision 值低

时,Recall 值往往偏高。当分类置信度高时,Precision 偏高;分类置信度低时,Recall 偏高。为了综合考虑这两个指标,使用 F1-score 对整体效果进行评估,公式如下:

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (14)$$

AP 表示 PR 曲线下的面积,mAP 表示多个类别的 AP 的均值,公式如下:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP \quad (15)$$

其中, n 表示类别总数,TP 表示被预测为正样本的正样本数量,FP 表示被预测为正样本的负样本数量,FN

表示被预测为负样本的正样本数量。

此外,为评测模型的轻量化效果,也将参数量和计算量纳入评价指标。

在计算 mAP 时,为更好地测量模型对目标区域的定位精度,判断检测目标区域的预测框是否精确,使用 IoU (Intersection over Union) 作为衡量预测框与真实框的误差标准。实验在计算 mAP 时设置 IoU 为 0.5。图 5、图 6 表示改进后的 GhostNet-YOLOv4 模型训练产生的 P-R 曲线、F1 曲线(此处选取 aeroplane、bicycle、bird、boat 四类为例进行展示)。

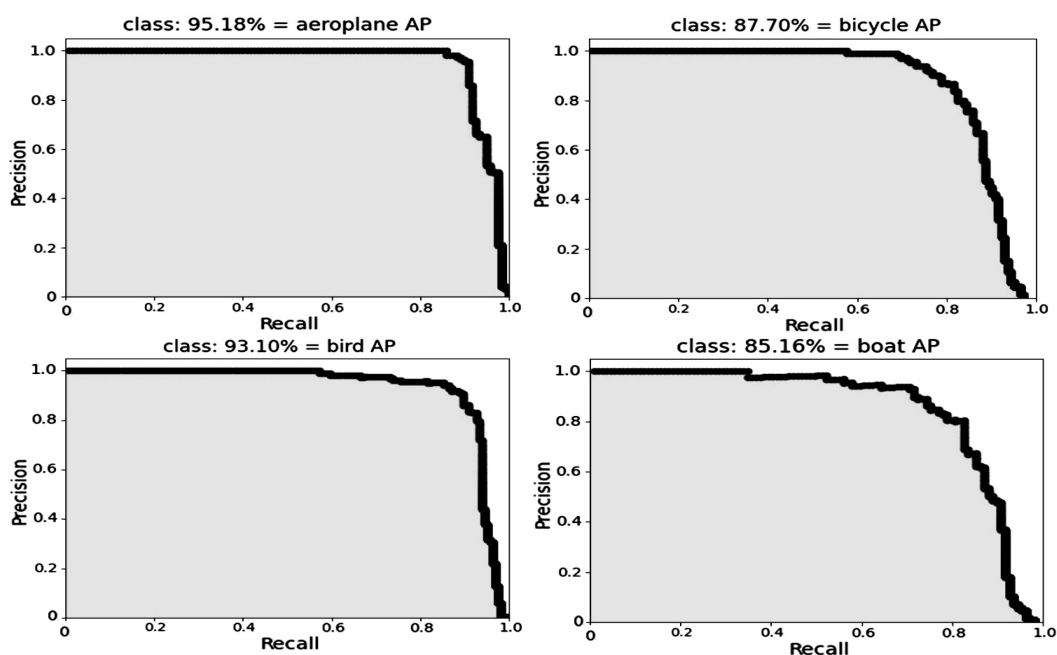


图5 模型训练产生的 P-R 曲线

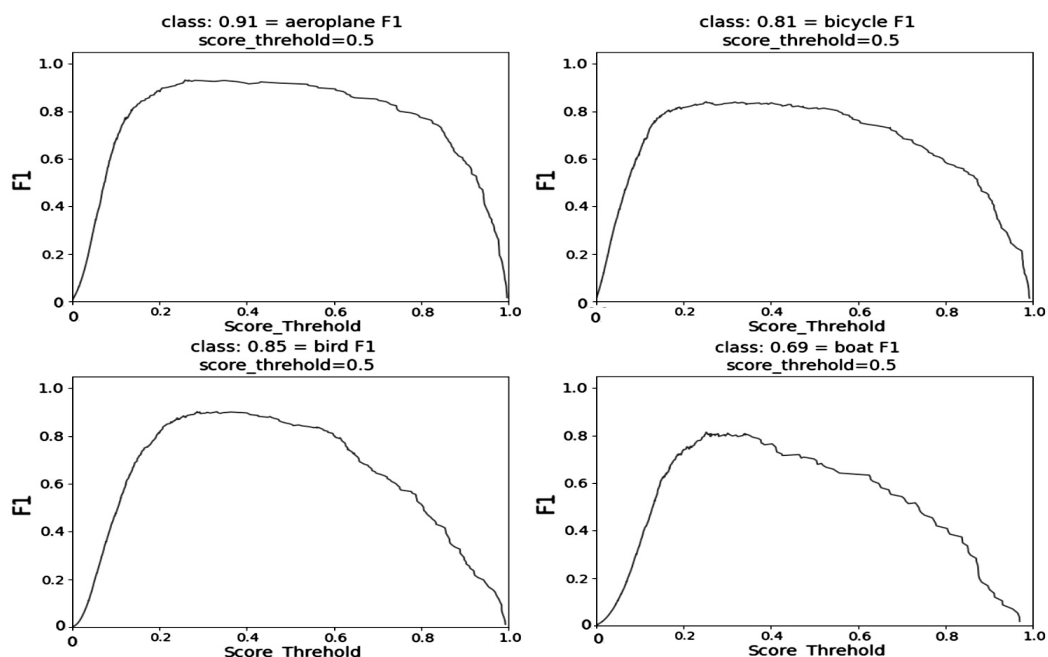


图6 模型训练产生的 F1 曲线

3.3 实验结果分析

分别将 YOLOv4 原网络模型、改进前的 GhostNet-YOLOv4 模型和改进后的 GhostNet-YOLOv4 模型在

PASCAL VOC 2007 数据集上运行,3 种模型在 20 种目标类别上的 AP 如图 7 所示。

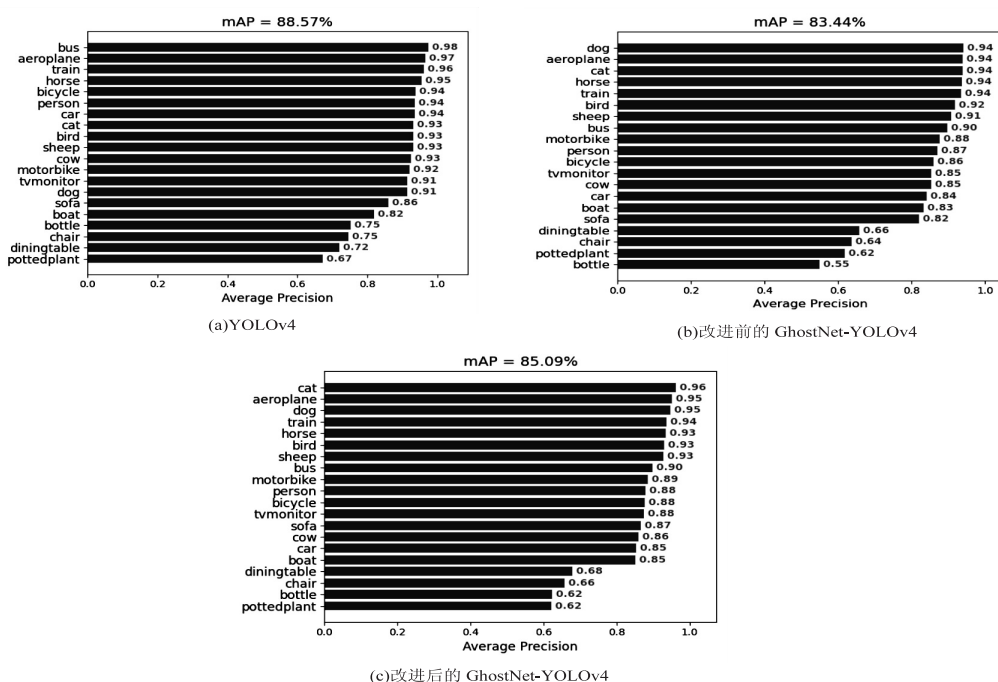


图 7 YOLOv4、改进前的 GhostNet-YOLOv4、改进后的 GhostNet-YOLOv4 在 20 种目标类别上的 AP 对比

表 2 YOLOv4、改进前的 GhostNet-YOLOv4、改进后的 GhostNet-YOLOv4 的参数量、计算量、mAP 对比

网络模型	参数量/M	计算量/GFLOPs	mAP/%
YOLOv4	64.363	60.334	88.57
改进前的 GhostNet-YOLOv4	11.429	7.066	83.44
改进后的 GhostNet-YOLOv4	11.429	7.068	85.09

为直观展示文中算法的有效性,算法随机抽取了三张图片,将改进前和改进后的 GhostNet-YOLOv4 网络模型的检测效果进行对比,从左至右依次为原始图像、改进前的 GhostNet-YOLOv4 检测效果、改进后的

GhostNet-YOLOv4 检测效果,如图 8 所示。可以看出,改进后的 GhostNet-YOLOv4 检测结果良好,能够完成目标距离较远、重叠、被遮挡、较小等情况下的检测任务,且相比改进前的模型有更高的置信度。



图 8 改进前后的 GhostNet-YOLOv4 模型检测效果对比

4 结束语

该文设计了一种基于 YOLOv4 的轻量化网络模型,该模型主干网络采用 GhostNet 残差结构,并融合 ECA 注意力机制模块,在保证一定的准确率的同时,极大地压缩了模型占用内存,解决了因大量卷积操作造成计算成本高、推理速度慢的问题。此外,还在模型中添加了焦点损失 Focal Loss,帮助网络提高检测精度。结果表明,改进后的 GhostNet-YOLOv4 在 PASCAL VOC 2007 数据集上的 mAP 达到 85.09%,相对于改进前的 GhostNet-YOLOv4 提高了 1.65 百分点;相对于 YOLOv4,参数量、计算量减少为原来的六分之一。

参考文献:

- [1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas; IEEE, 2016: 779-788.
- [2] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu; IEEE, 2017: 7263-7271.
- [3] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767, 2018.
- [4] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Venice; IEEE, 2017: 2961-2969.
- [5] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago; IEEE, 2015: 1440-1448.
- [6] REN S, HE K, 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.
- [7] 葛道辉, 李洪升, 张亮, 等. 轻量级神经网络架构综述[J]. 软件学报, 2019, 31(9): 2627-2653.
- [8] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv:1602.07360, 2016.
- [9] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City; IEEE, 2018: 6848-6856.
- [10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861, 2017.
- [11] 熊李艳, 涂所成, 黄晓辉, 等. 基于 MobileVIT 轻量化网络的车辆检测方法[J/OL]. 计算机应用研究: 1-6 [2022-07-21]. <https://doi.org/10.19734/j.issn.1001-3695.2021.12.0684>.
- [12] 孔维刚, 李文婧, 王秋艳, 等. 基于改进 YOLOv4 算法的轻量化网络设计与实现[J]. 计算机工程, 2022, 48(3): 181-188.
- [13] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle; IEEE, 2020: 1580-1589.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [15] 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. Seattle; IEEE, 2020: 11534-11542.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. Venice; IEEE, 2017: 2980-2988.
- [17] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas; IEEE, 2016: 761-769.
- [18] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[J]. arXiv:2107.08430, 2021.