DOI:10.13878/j.cnki.jnuist.20230502002

李孟浩[1]　袁三男[1]

基于改进 YOLOv5s 的交通标识检测算法

摘要

针对交通标识在图像中占比小、检测精度低且周围环境复杂等问题,提出一种基于改进 YOLOv5s 的算法.首先,在主干网络部分添加注意力机制 ECA(Efficient Channel Attention,高效通道注意力),增强网络的特征提取能力,有效解决了周围环境复杂的问题;其次,提出 HASPP(Hybrid Atrous Spatial Pyramid Pooling,混合空洞空间金字塔池化),增强了网络结合上下文的能力;最后,修改网络中的 Neck 结构,使高层特征与底层特征有效融合,同时避免了跨卷积层造成的信息丢失.实验结果表明,改进后的算法在交通标识数据集上取得了 94.4% 的平均检测精度、74.1% 的召回率以及 94.0% 的精确率,较原始算法分别提升了 3.7、2.8、3.4 个百分点.

关键词

交通标识检测;小目标检测;YOLOv5s;注意力机制;特征提取;混合空洞空间金字塔池化

中图分类号 TP391.4

文献标志码 A

收稿日期 2023-05-02

作者简介

李孟浩,男,硕士生,研究方向为目标检测、深度学习.menghaoli0411@163.com

袁三男(通信作者),男,博士,副教授,研究方向为视音频处理、嵌入式系统、深度学习.samuel.yuan@shiep.edu.cn

1 上海电力大学 电子与信息工程学院,上海,201306

0 引言

随着人工智能和物联网的迅猛发展,自动驾驶技术随之产生,而交通标识检测在自动驾驶技术中占据十分重要的地位.交通标识的错误分类可能会导致灾难性的后果,甚至危及生命安全.因此,确保对交通标识进行准确检测显得尤为必要,这有助于大大降低事故的发生率.然而,由于交通标识目标小,容易受道路周围树木、车辆等颜色背景相近的目标干扰,导致交通标识的检测难度大大增加.

交通标识的检测最早基于传统的机器学习进行研究,一般可以分为两类,一种是基于支持向量机直接对图像进行研究,另一种是使用不同的预处理技术提取特征,然后将这些特征作为机器学习算法的输入数据.Timofte 等[1]在使用支持向量机之前利用定向梯度直方图(Histogram of Oriented Gradients,HOG)选择特征来进行检测;Zang 等[2]将局部二值模式(Local Binary Pattern,LBP)特征检测器和自适应增强(Adaptive Boosting,AdaBoost)分类器相结合,先提取感兴趣区域,初步筛选后再进行分类检测.

随着计算机技术的不断发展,基于传统的视觉检测方法被发现有许多局限,不能满足交通标识等小目标任务的检测,自此之后,深度学习方法便被应用于视觉检测.基于深度学习的目标检测主要分为一阶段检测算法和两阶段检测算法.一阶段检测算法主要有:YOLOv3[3]、YOLOv4[4]以及 SSD[5]等,它们是将定位和识别任务同时进行,直接回归目标框的定位和分类的概率,较两阶段算法,速度较快但精度有所降低.两阶段检测算法主要有:R-CNN[6]、Faster R-CNN[7]、SPP-Net[8]等,它们先从图像中提取若干候选框,再逐个对这些候选框进行分类和辨别以及坐标调整,最后得出结果,比较准确但速度较慢.

刘安邦等[9]从一维观测向量中提取时域、频域等多个特征,构建高维特征向量,将检测问题转化为二分类问题,提升了小目标的检测性能;陈浩霖等[10]提出由多个卷积组合而成的主干网络,从多个尺度提取目标特征,并通过多尺度融合来学习目标特征;Zhang 等[11]提出了一种基于概率神经网络(Probabilistic Neural Networks,PNN)和中心投影变换(Center Projection Transform,CPT)技术的交通标志检测方法,突出了交通标志的形状特征,实现了比较高的准确率但不适合实时检测;鲍敬源等[12]通过引入通道变化的方法,在减少模型参数的同时加深了模型

的深度,但对小目标的检测效果仍然较差.

综上所述,虽然现有的目标检测算法取得了比较不错的效果,但交通标识在图像中占比小、检测精度低且周围环境复杂等问题仍没有得到有效的解决.因此,本文提出一种基于改进 YOLOv5s 的算法,主要针对以上存在的问题进行改进.本文的改进思路及创新点如下:

1)在主干部分加入注意力机制 ECA 模块,在少量增加参数的情况下增强主干网络的特征提取能力;

2)提出混合空洞空间金字塔池化(Hybrid Atrous Spatial Pyramid Pooling,HASPP)模块,将原有的快速空间金字塔池化(Spatial Pyramid Pooling-Fast,SPPF)模块替换,防止了因下采样造成的信息丢失,增强网络区域上下文的能力;

3)改进 Neck 结构,采用加权双向特征金字塔(Bidirectional Feature Pyramid Network,BiFPN)与空间深度层(Space-to-Depth,SPD)相结合的方式来融合深层次与浅层次的信息,避免了经过跨卷积层而导致的细粒度信息丢失,有利于小目标的位置信息的学习.

1 YOLOv5s 网络

YOLOv5 是当前广泛应用的一阶段目标检测网络之一.一阶段网络相比于两阶段网络少了区域候选过程,速度更快,且 YOLOv5s 是 YOLOv5 的轻量化版本,更适用于实时交通标识检测.因此本文选用 YOLOv5s 网络为基础进行交通标识检测.

YOLOv5s 的网络结构如图 1 所示.

YOLOv5s 网络结构主要分为 3 个部分:主干网络(Backbone)、颈部(Neck)和检测头(Head).其中主干网络主要负责提取特征,由 CBS、CSP1 和 SPPF 3 部分组成:CBS 是由卷积(Conv)、批量归一化(Batch Normalization,BN)和 SiLU 激活函数构成;CSP1 是一种残差结构[13],可以使计算过程中的参数量变小,速度更快,并且通过残差模块可以控制模型的深度,图中 CSP1_X,CSP2_X 的 X 表示该模块使用的串接次数,即深度;SPPF 的作用是对特征图进行多次池化,对高层特征提取并融合,比 SPP-Net 拥有更快的推理速度.颈部采用 PANet[14]结构,主要作用是进行特征融合,PANet 由 CBS、上采样(Upsample)、CSP2 组成.因为在特征融合时不需要一味地加深网络,所以 CSP2 与 CSP1 相比去掉了残差结构.

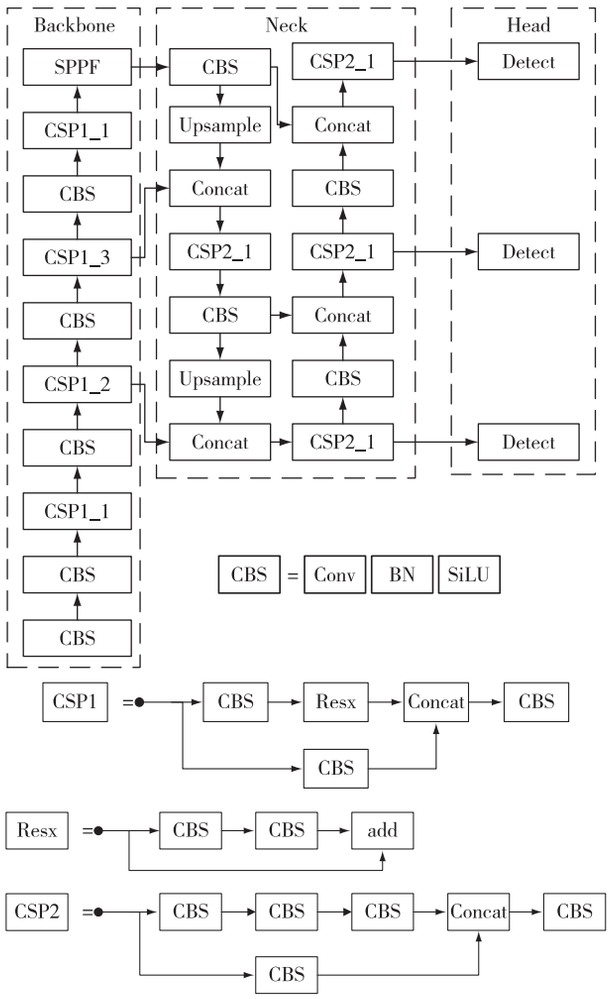

图 1 YOLOv5s 结构

Fig. 1 Structure of YOLOv5s

2 改进的 YOLOv5s

2.1 ECA 注意力机制

针对交通标识检测任务中容易受周围颜色背景相近目标干扰的问题,在网络中加入注意力机制.注意力机制可以定位到感兴趣的信息,通过在主干网络提取的特征层的通道维度上进行权值训练,让网络集中关注输入图片的交通标识部分,从而抑制周围复杂环境中无用的信息,减轻复杂环境对交通标识检测任务的影响,使深度卷积网络在性能上有所提高.网络的主干部分(Backbone)是特征提取的关键部分,因此本文将注意力机制 ECA 加入了网络的主干部分.

通道注意力机制是注意力机制的一种,其中最具代表性的是 SENet[15],主要是通过特征聚集和特征重新校准来提取特征,这种方法虽然获得了较高的精度,但是复杂度高、计算量大,其原因是 SENet

用降维的方法来控制模型复杂度,但是降维会对通道注意预测产生负面影响,在所有通道中获取依赖关系的效率低.而 ECA 可以有效地解决这个问题,ECA 结构如图 2 所示.

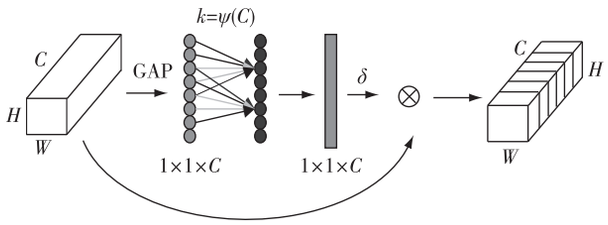

图 2　ECA 结构

Fig. 2　Structure of ECA

首先将输入的特征图进行平均池化得到聚合特征,然后通过卷积核大小为 k 的一维卷积来进行局部跨通道信息融合.其中,卷积核大小 k 代表了局部跨通道信息融合的覆盖率,k 与通道维数 C 成非线性比例,通过非线性映射,高维通道的相互作用距离较长,而低维通道相互作用距离较短,即:

$$C=\varphi(k)=2^{(\lambda k-d)}. \tag{1}$$

给定通道维数 C,卷积核大小 k 便可以自适应确定,即:

$$k=\psi(C)=\left|\frac{\log_2(C)}{\lambda}+\frac{d}{\lambda}\right|_{\text{odd}}, \tag{2}$$

式中:λ,d 是调节参数;$|b|_{\text{odd}}$ 表示 b 的最近的奇数.经过反复实验,ECA 层加在 HASPP 前一层效果最好.

2.2　HASPP 模块

捕获区域上下文信息在目标检测任务中是十分重要的.待检测的目标与其周围环境中的其他目标是同时存在的,它们之间一定会存在着某种联系,学习它们之间的关系有利于对待检目标进行检测.

为了更好地检测小目标物体,本文提出使用 HASPP 模块替换原有的 SPPF 模块,在 YOLOv5s 中,SPPF 起到捕获区域上下文信息的作用,SPPF 由 SPP 改进而来,在获得同样信息的情况下比 SPP 有更快的速度,SPPF 模块的结构如图 3 所示.

输入特征图依次通过 3 个 5×5 的最大池化层,得到了 3 个不同大小的感受野,以此来达到捕获区域上下文信息的目的.最大池化虽然可以扩大感受野,但会降低特征图的分辨率,损失一些有用的信息.本文参考 ASPP[16],提出 HASPP 模块,即改进 SPPF.改进后的 SPPF 包含了 3 个分支,如图 4 所示.

一条支路经过 1×1 的卷积来获得 1×1 的感受野,另一条支路依次通过 3×3 的普通卷积、空洞卷积扩张率(dilation rate,d)为 3 的卷积和空洞卷积扩张率为 5 的卷积,来获得 3 个不同大小的感受野,最后一条支路通过池化和上采样操作来获得全局感受野,并将 5 种不同大小的感受野融合输出.

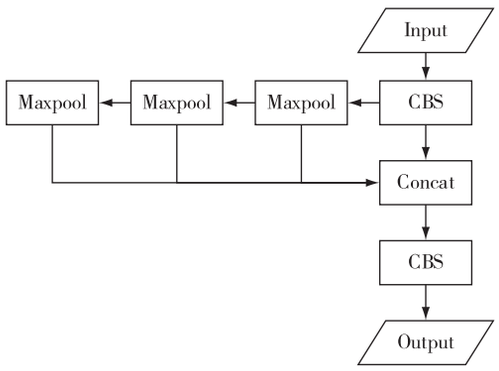

图 3　SPPF 结构

Fig. 3　Structure of SPPF

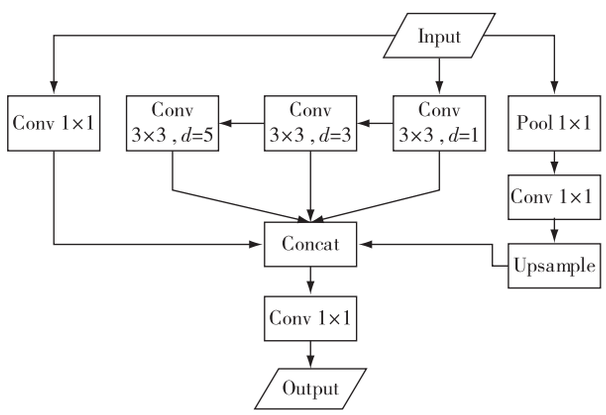

图 4　HASPP 结构

Fig. 4　Structure of HASPP

而经过多次空洞卷积后会造成感受野不连续的问题,可能会丢失大量的信息,因此借鉴混合扩张卷积[17](Hybrid Dilated Conv,HDC)的方法来选择空洞卷积扩张率,获得连续的感受野.将 2 个像素点非零值之间的最大距离 M_i 定义为

$$M_i=\max[M_{i+1}-2d_i,M_{i+1}-2(M_{i+1}-d_i),d_i], \tag{3}$$

并且

$$M_n=d_n. \tag{4}$$

这样设计的目的是为使

$$M_2\leqslant k, \tag{5}$$

式中,d_i 是第 i 个空洞卷积扩张率,k 为卷积核大小.经过多次实验,空洞卷积扩张率采用 1、3 和 5 的效

果最佳.

特征图输入到此模块后,通过控制不同的空洞卷积扩张率来获得不同尺度的特征信息,同时保证不丢失分辨率,并通过 HDC 算法来获得连续的感受野,防止因连续空洞卷积造成的信息丢失,因此可以学习到交通标识周围其他目标的相关信息,对小尺度交通标识的特征信息起到补充作用.

2.3 改进的 Neck 结构

融合不同尺度的特征是提高目标检测性能的一个关键手段.分辨率更高的底层特征包含更多的细节信息,但是其包含语义信息较少、噪声较多;分辨率低的高层特征则包含更多的语义信息,但是所包含的细节信息较少.因此,如何高效地融合不同尺度的特征对于交通标识的检测是十分重要的.

YOLOv5s 中的 Neck 采用的是 PANet 的结构(图 5)来进行特征融合的.先经过上采样传递高层的语义特征,再经过下采样来补充底层的信息,弥补了定位信息.

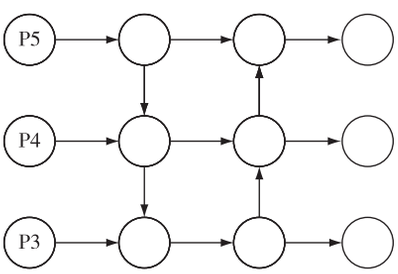

图 5 PANet 结构

Fig. 5 Structure of PANet

经过多层网络后,底层的信息比较模糊,并且下采样过程中跨步长卷积造成了细粒度信息的丢失,特征表示比较低效.因此,本文在 P4、P5 节点和输出节点之间增加了 2 条新的路径,用原始特征去补充经过多次网络后比较模糊的信息,并且用 SPD 模块替换掉跨步长卷积,结构如图 6 所示.

SPD 模块是由一个从空间到深度的层和一个非跨步长卷积组成的,其结构如图 7 所示.对特征图利用切片的方法进行下采样,保留了通道维度上的所有信息,因此不存在信息的丢失,然后用一个非跨步长的卷积来减少切片操作增加的通道数量.

切片操作如下:

把任意大小为 $S \times S \times C_1$ 的中间特征映射为 X,将子特征映射序列切片为

$$f_{s-1,0} = X[s-1:S:s,0:S:s],$$

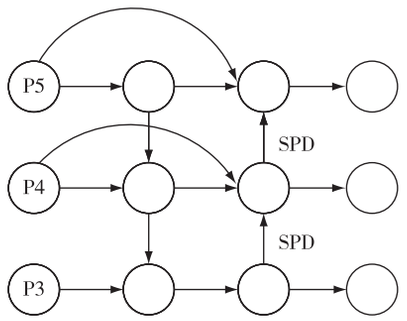

图 6 改进后的 Neck 结构

Fig. 6 Improved Neck structure

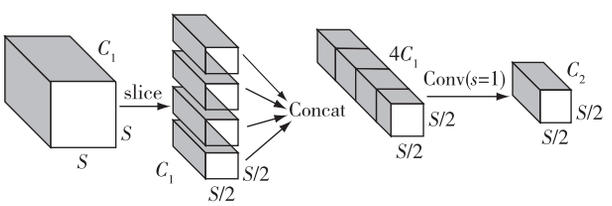

图 7 SPD 结构

Fig. 7 Structure of SPD

$$f_{s-1,1} = X[s-1:S:s,1:S:s],$$
$$\vdots$$
$$f_{s-1,s-1} = X[s-1:S:s,s-1:S:s], \quad (6)$$

式中,s 是卷积步长,本文要替换的跨卷积步长为 2,因此 $s=2$. 综上所述,改进后的 Neck 结构会使输出的特征图包含更多的信息,有利于对交通标识中小目标的检测.改进后的 YOLOv5s 结构如图 8 所示.

改进后的 Neck 结构用 SPD 模块替换掉原网络中的跨步长卷积的下采样模块,避免了跨步长卷积造成的信息丢失,同时通过两条新的路径,将经过 SPD 模块处理后的特征图与经过骨干网络残差模块处理后的特征图进行特征融合,从而弥补了信息丢失的问题,因此提高了对小目标的检测精度.

3 实验及评价指标

实验环境为 Windows10 操作系统,计算机内存为 16 GB,处理器为 E5-2678 V3,显卡为 RTX2080Ti,显存为 11 GB,深度学习框架为 Pytorch1.2,编程语言为 Python3.9.

3.1 数据集

数据集采用交通标识数据集 TT100K[18],它是腾讯和清华大学合作制作的交通标识公共数据集(https://cg.cs.tsinghua.edu.cn/traffic-sign/).TT100K 提供了 10 000 多张包含交通标识的图片,每个交通标识

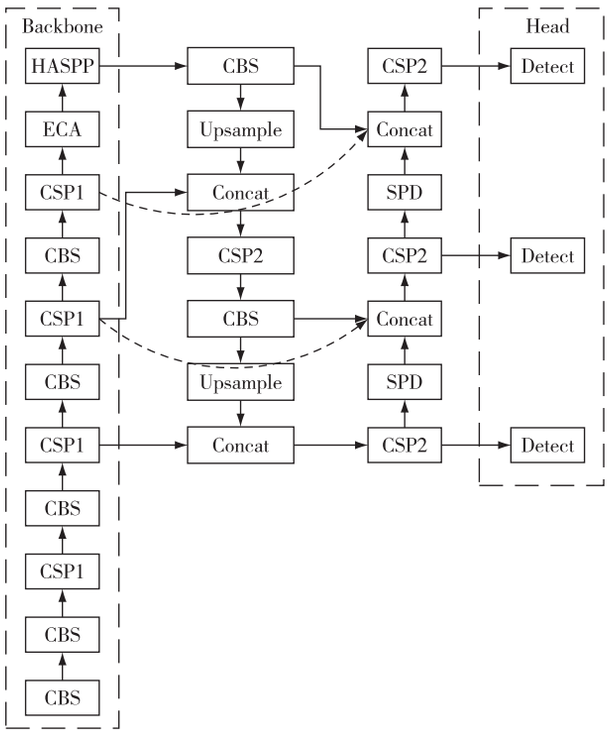

图 8　改进后的 YOLOv5s 结构

Fig. 8　Improved YOLOv5s structure

都使用一个类标签.本文从中选取数据量较多的 45 个类别,共 6 105 张图片,按照 8∶2的比例划分训练集和测试集,其中训练集有 4 814 张,测试集 1 291 张.

3.2　参数设置与评价指标

训练使用 Mosaic 数据增强方法,使用 YOLOv5s 的预训练权重进行初始化,优化函数采用随机梯度下降(Stochastic Gradient Descent,SGD),初始学习率大小为 0.01,输入尺寸大小为 640×640,输入批次大小为 16,动量为 0.934,一共训练 200 轮.

实验的结果采用平均精度均值(mean Average Precision,mAP)、召回率(Recall,R)和精确率(Precision,P)等作为评价指标,具体的计算公式如下:

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{7}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{8}$$

$$\mathrm{AP}=\int_0^1 P(R)\,\mathrm{d}R, \tag{9}$$

$$\mathrm{mAP}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i, \tag{10}$$

式中,AP 是 P-R(Precision-Recall)曲线的面积,TP 为正确预测的正类样本数量,FP 为错误预测的正类样本数量,FN 为错误预测的负类样本数量,n 为类别数目.本文设置评价指标为 mAP@0.5,即设置 IOU 检测阈值为 0.5.

在实验结果中采用参数量(Parameters,N_p)作为模型复杂度的评价指标,计算公式如下:

$$N_p=C_1\times(k_1\times k_2\times C_2+1), \tag{11}$$

其中,C_1 和 C_2 分别代表输出和输入的通道数,k_1 和 k_2 分别代表卷积核的高和宽.

用 FPS 作为检测速度的指标,FPS 代表每秒处理的帧数,单位为帧/s.

3.3　实验结果分析

本文共设置了 4 组实验,分别为改进部分的消融实验、注意力机制的对比实验、空洞卷积扩张率 d 的对比实验以及改进算法和主流算法的对比实验.每次实验均进行 10 次,取平均结果作为实验的最终结果.

3.3.1　改进部分消融实验

为了验证本文所提出的改进算法对交通标识检测的有效性,设置了一组消融实验,包含 6 个部分:1)原 YOLOv5s 模型;2)YOLOv5s 模型+ECA 模块;3)YOLOv5s 模型+HASPP 模块;4)YOLOv5s 模型+改进颈部模块;5)YOLOv5s 模型+ECA 模块+HASPP 模块;6)YOLOv5s 模型+ECA 模块+HASPP 模块+改进颈部模块.具体对比分析结果如表 1 所示.

表 1　消融实验

Table 1　Ablation experiment

序号	算法	N_p/10^6	mAP/%	R/%	P/%	FPS/(帧/s)
1	YOLOv5s	7.1	90.7	71.3	90.6	43
2	YOLOv5s+ECA	7.1	91.8	73.7	91.4	50
3	YOLOv5s+HASPP	15.4	91.7	73.4	91.7	43
4	YOLOv5s+改进 Neck	7.2	92.1	74.6	92.3	47
5	YOLOv5s+ECA+HASPP	15.4	93.2	73.7	93.1	45
6	YOLOv5s+ECA+HASPP+改进 Neck	**15.5**	**94.4**	**74.1**	**94.0**	**43**

由表 1 可以看出:仅添加 ECA 模块,mAP 值提升 1.1 个百分点,R 提升 2.4 个百分点,P 提升 0.8 个百分点;仅添加 HASPP 模块,mAP 值提升 1 个百分点,R 提升 2.1 个百分点,P 提升 1.1 个百分点;仅添加改进 Neck 模块,mAP 值提升 1.4 个百分点,R 提升 3.3 个百分点,P 提升 1.7 个百分点;3 个模块同时添加,mAP 值提升 3.7 个百分点,同时 FPS 与原算法模型相同,均为每秒 43 帧,可以同时满足车辆标识检测的精度和实时性.

3.3.2 注意力机制的对比实验

为了验证注意力机制 ECA 的有效性,本文将其分别与主流注意力机制 SE(Squeeze-and-Excitation)、CBAM[19](Convolutional Block Attention Module)以及 CA[20](Coordinate Attention)进行了对比实验,对比结果如表 2 所示.

表 2 注意力机制对比实验

Table 2 Comparative experiment on attention mechanism

序号	算法	$N_p/10^6$	mAP/%	R/%	P/%
1	YOLOv5s+SE	7.1	91.1	73.3	91.0
2	YOLOv5s+CBAM	7.1	91.4	72.8	91.2
3	YOLOv5s+CA	7.1	91.2	71.2	91.0
4	YOLOv5s+ECA	7.1	91.6	73.1	91.2
5	在 HASPP 前加入 ECA 模块	15.4	91.8	73.7	91.4
6	在 HASPP 后加入 ECA 模块	15.4	91.2	73.3	91.1

从表 2 中可以看出,添加 ECA 模块 mAP 值为 91.6%,R 为 73.1%,P 为 91.2%,与其他注意力机制模型相比,在模型复杂度相同的情况下,mAP 值和 P 均达到最优.为了验证 ECA 模块在不同位置的效果,本文进行了对比实验.然而,在特征提取网络较浅的位置加入注意力机制并不会带来良好的效果,甚至会导致性能下降.因此,本文将注意力机制放置在提取网络的最后两层,即 HASPP 模块前后分别加入 ECA 模块.实验结果如表 2 所示,在 HASPP 模块之前加入 ECA 模块的效果优于在 HASPP 模块之后加入 ECA 模块,最终本文选择采用 ECA 加入到 HASPP 模块之前.

3.3.3 卷积扩张率 d 的对比实验

为了验证 HASPP 模块的有效性,本文设置了不同空洞卷积扩张率 d 的对比实验,结果如表 3 所示.其中,卷积扩张率 2,2,2;3,3,3;4,4,4 并不满足 $M_2 \leqslant k$ 的条件,为普通的空洞卷积;而卷积扩张率 1,2,3;1,2,5;1,3,5 满足 $M_2 \leqslant k$ 条件,为混合空洞卷积.

表 3 HASPP 的部分对比实验

Table 3 Partial comparative experiments of HASPP

序号	d	P/%	R/%	mAP/%
1	2,2,2	91.0	74.1	91.4
2	3,3,3	91.2	73.8	91.2
3	4,4,4	90.4	73.7	90.1
4	1,2,3	92.8	73.6	93.6
5	1,2,5	93.7	73.7	94.1
6	1,3,5	94.0	74.1	94.4

从表 3 中可以看出:混合空洞卷积的 mAP 值和 P 均大于普通空洞卷积,而混合空洞卷积中 $d=1,3,5$ 的 mAP 值、R 和 P 分别达到 94.0%、74.1% 和 94.4%,明显最优.因此,本文选取 $d=1,3,5$ 作为 HASPP 模块的卷积扩张率.

3.3.4 与主流算法的对比实验

为了验证本文改进后的算法,本文分别与 YOLOv3、YOLOv4、YOLOv5、SSD、RetinaNet、Faster R-CNN 等主流算法在 TT100K 数据集上进行对比实验,结果如表 4 所示.

表 4 与主流算法对比实验

Table 4 Comparison with mainstream algorithms

序号	算法	$N_p/10^6$	mAP/%	FPS/(帧/s)
1	YOLOv3	61.8	83.4	29.6
2	YOLOv4	96.9	87.8	41.7
3	YOLOv5	7.1	90.7	43.0
4	SSD	25.1	70.4	25.0
5	RetinaNet	66.7	49.5	28.6
6	Faster R-CNN	167.7	92.0	18.0
7	YOLOX-l[21]		94.6	37.3
8	AutoAssign[22]		91.2	43.04
9	本文	15.1	94.4	43.0

从表 4 中可以看出:改进后的算法 mAP 值达到 94.4%,模型复杂度为 15.1×10^6,FPS 为每秒 43 帧,与之相比其他主流算法中最好的 YOLOX-l[21] 的 mAP 值达到 94.6%,仅比本算法高 0.2 个百分点,但其推理速度比本文算法慢大约每秒 6 帧;与比较前沿的 AutoAssign[22] 相比,在推理速度上基本相同,但 mAP 值高出大约 3 个百分点;与 Faster R-CNN 相比,mAP 值高出 2.4 个百分点,而模型仅为 Faster R-CNN 的 1/10,同时推理速度也更快.基于以上结果,可以看出改进后的算法具有很大的先进性,可以满足交通标识检测的准确度和实时性要求.

3.3.5 与原始算法对比实验

算法改进前后在交通标识检测任务上的测试结果如图 9 所示.对比图 9a 和 9b 可以发现,在类别为 pn 的待检目标上平均精度分别为 0.87、0.93,p27 的待检目标上平均精度分别为 0.89、0.90,表明改进算法的平均精度高于原始算法;对比图 9c 和 9d,原 YOLOv5s 在复杂环境下检测效果差,而改进后的 YOLOv5s 在复杂环境下依然有较高的平均精度,表明改进后的 YOLOv5s 在加入注意力机制后可以有效抑制无用的信息;对比图 9e 和 9f 可以发现,原

a. 整体效果(原YOLOv5s)

b. 整体效果(改进后的YOLOv5s)

c. 复杂环境下(原YOLOv5s)

d. 复杂环境下(改进后的YOLOv5s)

e. 小目标(原YOLOv5s)

f. 小目标(改进后的YOLOv5s)

图 9　改进前后算法的对比效果

Fig. 9　Performance comparison of algorithms before and after improvement

YOLOv5s 对远处的小目标检测的平均精度只有 0.69,而改进后 YOLOv5s 对小目标检测的平均精度为 0.81,表明改进后的 YOLOv5s 在 HASPP 模块和改进 Neck 的作用下可以进一步加强网络对小目标特征信息的提取能力.因此,改进后的算法可以有效提升交通标识检测精度.

4　结束语

本文介绍了改进的 YOLOv5s 交通标识检测模型.针对交通标识在图像中占比小、检测精度低且周围环境复杂等问题,在原始算法的主干部分加入了注意力机制 ECA 模块用于增强网络的特征提取能力,解决了交通标识周围环境复杂的问题;提出 HASPP 模块,提高了网络区域上下文的能力,解决了小目标特征提取难的问题;改进了原算法的 Neck 结构,避免了细粒度信息的丢失,并结合深层和浅层信息,增强了网络学习小目标位置信息的能力.通过实验分析,改进后的算法在平均精度均值上有所提升,速度也达到了实时检测的标准.未来的研究方向是使用剪枝、知识蒸馏的手段压缩模型,以达到更加轻量化的效果.

参考文献
References

[1] Timofte R, Zimmermann K, Gool L V. Multi-view traffic sign detection, recognition, and 3D localisation [J]. Machine Vision and Applications, 2014, 25(3): 633-647
[2] Zang D, Zhang J Q, Zhang D D, et al. Traffic sign detection based on cascaded convolutional neural networks [C]// 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). IEEE, 2016: 201-206
[3] Redmon J, Farhadi A. YOLOv3: an incremental improvement[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95
[4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection [C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020. DOI: 10. 48550/arXiv. 2004. 10934
[5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37
[6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587
[7] 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
[8] 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
[9] 刘安邦,施赛楠,杨静,等.基于虚警可控梯度提升树的海面小目标检测[J].南京信息工程大学学报(自然科学版),2022,14(3):341-347
LIU Anbang, SHI Sainan, YANG Jing, et al. Sea-surface small target detection based on false-alarm-controllable gradient boosting decision tree [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2022, 14(3): 341-347
[10] 陈浩霖,高尚兵,相林,等.FIRE-DET:一种高效的火焰检测模型[J].南京信息工程大学学报(自然科学版),2023,15(1):76-84
CHEN Haolin, GAO Shangbing, XIANG Lin, et al. FIRE-DET: an efficient flame detection model [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2023, 15(1): 76-84
[11] Zhang K, Sheng Y, Li J. Automatic detection of road traffic signs from natural scene images based on pixel vector and central projected shape feature[J]. IET Intelligent Transport Systems, 2012, 6(3): 282-291
[12] 鲍敬源,薛榕刚.基于 YOLOv3 模型压缩的交通标志实时检测算法[J].计算机工程与应用,2020,56(23):202-210
BAO Jingyuan, XUE Ronggang. Compression algorithm of traffic sign real-time detection based on YOLOv3 model [J]. Computer Engineering and Applications, 2020, 56(23): 202-210
[13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778
[14] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768
[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141
[16] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40(4): 834-848
[17] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 1451-1460
[18] Zhu Z, Liang D, Zhang S H, et al. Traffic-sign detection and classification in the wild [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2110-2118
[19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19
[20] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722
[21] Ge Z, Liu S T, Wang F, et al. Yolox: exceeding Yolo series in 2021[J]. arXiv e-print, 2021, arXiv: 2107. 08430
[22] Zhu B J, Wang J F, Jiang Z K, et al. Autoassign: differentiable label assignment for dense object detection [J]. arXiv e-print, 2020, arXiv: 2007. 03496

南京信息工程大学学报,2024,16(1):11-19
Journal of Nanjing University of Information Science & Technology,2024,16(1):11-19

Traffic sign detection based on improved YOLOv5s

LI Menghao[1] YUAN Sannan[1]
1 College of Electronics and Information Engineering,Shanghai University of Electric Power,Shanghai 201306,China

Abstract An algorithm based on improved YOLOv5s is proposed to address the problems of small percentage of traffic signs in the image,low detection accuracy and complex surrounding environment.First,the attention mechanism of ECA (Efficient Channel Attention) is added to the backbone network part to enhance the feature extraction ability of the network and effectively solve the problem of complex surrounding environment.Second,the HASPP (Hybrid Atrous Spatial Pyramid Pooling) is proposed,which enhances the network's ability to combine context.Finally,the neck structure in the network is modified to allow efficient fusion of high level features with underlying features while avoiding information loss across convolutional layers.Experimental results show that the improved algorithm achieves an average detection accuracy of 94.4%,a recall rate of 74.1% and an accuracy rate of 94.0% on the traffic signage dataset,which were 3.7,2.8,and 3.4 percentage points higher than the original algorithm,respectively.

Key words traffic sign detection;small target detection;YOLOv5s;attention mechanism;feature extraction;hybrid atrous spatial pyramid pooling (HASPP)