

基于多特征融合注意力机制的交通标志检测^①

李国强^② 付 乐 查琳琳 王天雷

(燕山大学电气工程学院 秦皇岛 071000)

摘 要 针对复杂环境下交通标志目标检测尺寸较小、特征不明显等问题,在 YOLO v4 算法的基础上,提出了基于多特征融合注意力机制的方法,形成 reSE-BYOLO v4 算法。reSE-BYOLO v4 算法首先利用高分辨率下的特征信息改进检测层部分,提高小目标的检测能力;在特征融合部分利用同一尺度特征信息构建横向连接,使模型在不增加成本情况下融合更多的特征;利用 reSE 模块对有效特征进行增强,该模块对 SE 模块进行优化,利用残差学习的思想对通道维度重新标定。在 TT100K 数据集上的测试结果表明, reSE-BYOLO v4 算法相对 YOLO v4 算法平均精度 (mAP) 提高了 6.57%, 平均精确度提高了 4.78%, 平均召回率提高了 5.26%, reSE-BYOLO v4 算法对交通标志识别能力更强。

关键词 交通标志检测; YOLO v4; 注意力机制; 残差学习; 特征融合

0 引 言

交通标志识别在智能交通领域占据着重要的地位,它能够为车辆提供准确的交通标志信息,在无人驾驶领域拥有广阔的前景,通过实时监测前方交通标志,使无人驾驶技术在遵守交通规则的前提下安全行驶^[1]

对于交通标志识别的研究,早期利用传统图像识别^[2]的方法,采用滑动窗口对图像遍历得到候选区域,然后用基于手工设计的特征提取器提取候选区域的特征,之后利用支持向量机(support vector machine, SVM)对得到的特征进行分类,最后通过目标类别进行边框回归。由于存在光照、遮挡、变形等问题,传统图像方法在实际的交通标志检测任务中很难取得很好的效果。近年来卷积神经网络^[3](convolutional neural network, CNN)被越来越多地应用在目标检测和分类领域,利用 CNN 替代手工特征,提取图像中高层特征的优势,使深度学习算法在目标检测领域取得了很大的成果。这些检测算法可以分

为 2 大类:两阶段方法(R-CNN^[4]、Fast R-CNN^[5]和 Faster R-CNN^[6]等)和单阶段方法(YOLO 系列^[7-10]、SSD(single shot multibox detector)^[11])。其中以 Faster R-CNN 为代表的两阶段方法使用区域候选网络(region proposal network, RPN)通过共享卷积特征在特征层上生成建议框,再利用建议框区域的卷积特征进行分类和目标框的定位,而 YOLO 系列和 SSD 为代表的单阶段检测方法将物体的分类任务和定位任务统一按照回归逻辑在输出层完成预测,并不需要提取候选区,从而大幅提高了运算速度。目前有很多学者将目标检测算法应用于交通标志检测中,文献[12]在 YOLO v3 基础上提出了融合注意力机制与上下文信息的交通标志检测方法。文献[13]提出了一种融合空间金字塔池化(spatial pyramid pooling, SPP)和改进特征金字塔网络(feature pyramid network, FPN)的 YOLOv3 交通标志检测方法。文献[14]以 CSPDarkent53 为基础设计带有反卷积级联结构的特征提取网络。文献[15]提出一种利用扩张卷积在 CNN 中实现多尺度滑动窗的方法并在 GTSDb 数据集^[16]上取得优异的检测结果。

^① 基于神经网络案例匹配的电站锅炉燃烧系统参数整定与优化(F2020203003)资助项目。

^② 男,1984 年生,博士生,副教授;研究方向:神经网络;联系人,E-mail: zhihuiyuan@163.com。
(收稿日期:2021-11-22)

对于实际场景应用下的交通标志检测来说,实时性是算法的一个关键指标,所以提高单阶段检测方法的精度对交通标志检测更有意义。本文以单阶段目标检测算法 YOLO v4 为基础,针对存在小目标漏检的问题,提出一种 reSE-BYOLO v4 算法。该算法对 YOLO v4 检测层进行改进,在减少网络模型参数的同时又增强了对尺寸偏小交通标志的检测能力;在多特征融合中引入横向连接,使模型融合了更多的特征;对通道注意力(SE)模块^[17]进行了优化并与模型相融合。最后本文通过大量的实验证实了该算法的可行性。

1 相关工作

1.1 YOLO v4 网络结构

YOLO v4 算法由进行特征提取的主干网络(Backbone)、进行特征融合的颈部(Neck)和用于分类及回归的检测头(Head) 3 部分组成。与 YOLO v3 算法相比,在 YOLO v3 算法的“Darknet53 + FPN + YOLO-Head”结构上,借鉴了近年来深度学习其他优秀的算法模型和一些训练技巧。YOLO v4 主干网络在 Darknet 基础上,与 CSPNet^[18] 算法相融合,形成 CSPDarknet;颈部由 YOLO v3 采用的特征金字塔网络^[19] (FPN)改成了加入了空间金字塔池化层^[20] (SPP)的路径聚合网络^[21] (PANet),可以将主干网络输出的深层特征向浅层传递,弥补了 FPN 网络特征融合中浅层特征丢失的缺陷;检测头继续延用了 YOLO v3 中的 YOLO-Head。表 1 为 YOLO v3 和 YOLO v4 网络对比图。

表 1 YOLOv3 和 YOLOv4 网络对比图

网络	主干网络	颈部	检测头
YOLO v3	Darknet53	FPN	YOLO-Head
YOLO v4	CSPDarknet53	SPP + PAN	YOLO-Head

1.2 双向特征金字塔

双向特征金字塔(BiFPN)^[22]由谷歌大脑团队提出,在 FPN 中加入了跳跃连接,并将 FPN 中自上而下特征融合替换为自上而下和自下到上的特征融

合,使网络不增加额外参数的同时融合更多相同尺度的特征,使不同尺度特征更充分融合。其结构如图 1 所示,BiFPN 有效地融合了不同尺度的特征并增加了同一尺度特征的信息融合。

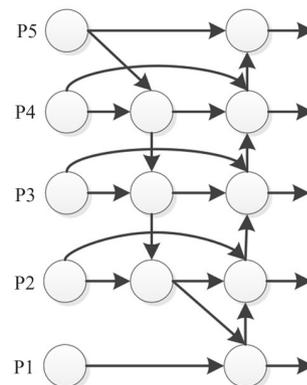


图 1 BiFPN 结构图

1.3 通道注意力

通道注意力机制(squeeze and excitation net, SENet)^[23]其结构如图 2 所示,SENet 分为 3 个步骤。首先进行压缩操作,采用全局平均池化对输入的 $C \times H \times W$ 的特征图进行压缩,使其变为 C 个 1×1 的实数列,使每张特征图具有全局感受野;其次进行激励操作,将实数列送入到由 2 个全连接构成的瓶颈结构,第 1 个全连接层对特征图 C 个通道进行降维,得到 $1 \times 1 \times (C/r)$ 的向量(其中 r 为缩放参数),

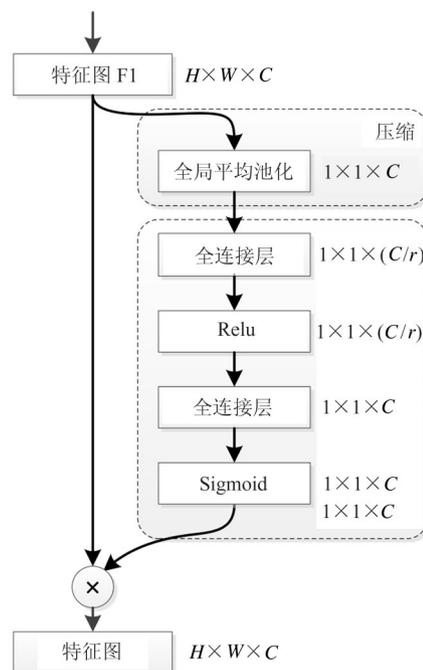


图 2 通道注意力 SE 模块结构

经过 Relu 函数激活后,通过第 2 个全连接层升维,得到 $1 \times 1 \times C$ 的向量,再利用 Sigmoid 激活函数获得对应归一化权重;最后通道权重与输入特征按通道相乘,完成通道注意力的标定。

2 模型分析与改进

2.1 检测层改进

对于 608×608 的输入图像,YOLOv4 的 3 个输出层的网格尺寸为 76×76 、 38×38 和 19×19 ,分别负责预测大、中、小物体,而在自然场景中远距离的标志牌一般尺度较小,为了让模型对小目标检测有更好的表现,在浅层再增加一层输出特征层 152×152 。因为浅层特征图尺度更大,包含更多的像素细节,另一方面考虑到 19×19 的输出层主要负责检测大物体,而交通标志一般尺寸较小,故 19×19 层对小目标的检测意义不大。图 3(a)、(b)、(c)、(d) 分别是 152×152 、 76×76 、 38×38 、 19×19 部分特征图,在尺寸为 19×19 的特征图中可以看出小目标特征丢失严重,无法看出交通标志的轮廓,而尺寸为 152×152 的特征图保留了更多小目标的特征信息。因此本文去掉 19×19 的检测层,添加 152×152 的检测层,改进后的网络依旧是 3 个检测头,主要针对小目标交通标志的检测,分别用 38×38 、 76×76 和 152×152 的检测层来进行预测。

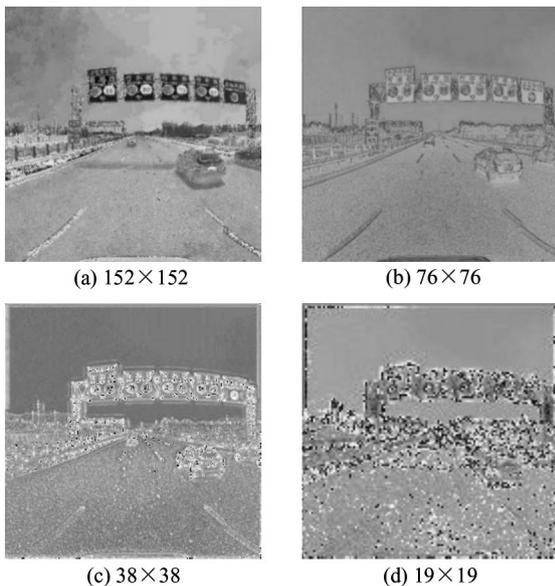


图 3 残差块输出的特征图

2.2 多特征融合

YOLOv4 在多尺度特征融合部分选取了 C_3 、 C_4 、 C_5 3 个不同尺度特征输入到路径聚合网络 (path aggregation network, PAN) 结构中,通过构造由上而下和自底而上的双向通道融合 3 个尺度的特征,最终得到 3 个尺度的输出。针对交通标志尺寸小和尺度变化的问题,原始网络只对 3 个尺度特征进行了提取,对不同大小目标不敏感,不同尺度的特征信息融合较少,且并没有利用信息丰富的特征图 C_2 。本文所提出的多特征融合网络在检测层改进基础上借鉴 BiFPN 网络中的横向连接进行改进。

在图 4 和图 5 中 $P_5 \sim P_2$ 分别是本文中提到的 19×19 、 38×38 、 76×76 、 152×152 的检测层, C_i 则表示前馈网络提取到的多尺度特征。 D_i 和 P_i 分别是由上而下和自底而上的特征融合,虚线是借鉴 BiFPN 横向连接的思想将同一尺度下的特征进一步融合,缓解因网络层级过多造成的特征信息丢失。以 P_4 特征生成为例,原始网络由 P_3 下采样和 D_4 两部分的特征按通道维度拼接得到,生成过程如式 (1) 所示。改进后的网络增加了对 C_4 特征的融合用式 (2) 体现,在不增加成本情况下融合了更多的特征。图中的斜箭头是通过双线性插值来实现 C_3

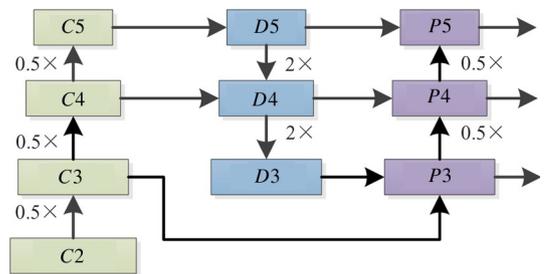


图 4 YOLO v4 的特征融合网络

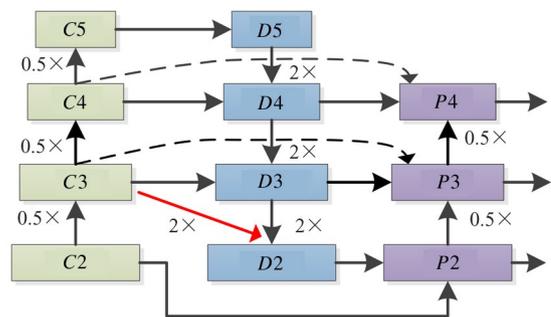


图 5 改进后的特征融合网络

的 2 倍上采样,对高分辨率下 $C3$ 特征充分利用,使网络融入了更多的语义信息。

$$P4 = F([\text{Downsample}(P3); D4]) \quad (1)$$

$$P4 = F([\text{Downsample}(P3); D4; C4]) \quad (2)$$

其中,符号 $[\]$ 表示通道维度的拼接, F 表示用于调整通道的卷积操作,Downsample 是通过卷积操作实现 2 倍下采样。

本文对多尺度特征融合部分进行改进,充分利用了浅层特征图 $C2$ 中的特征信息,提高了对小目标的检测能力,同时借鉴 BiFPN 网络中的横向连接,在同尺度特征层引入跳跃连接融合更多的相同尺度特征,缓解因网络层级过多造成的特征信息丢失,使网络对不同大小的目标更敏感,提升网络的检测能力并降低漏检率。

2.3 通道注意力的构建

由于 SENet 是针对于图像分类而设计的通道注意力,在图像分类中目标物体一般占据图片较大面积,使用全局平均池化可以很好地获得通道特征响应,利用每个通道的平均值来代表该通道的目标情况。然而对于目标检测任务而言,目标物体通常较小,只能占据图片很小的区域,而且图片中包含许多无关物体,这些物体虽然在通道中占据很小的特征响应,但是数量偏多,总的特征响应贡献依旧很大,因此在空间维度上使用全局平均池化并不能很好地反映图片中的前景目标,而每个通道的最大值可能会很好地反映该通道对于前景目标的情况。另一方面,SENet 首先把特征在空间维度上压缩,再经过全连接网络的学习和激活,完成原始特征在通道维度上的重新标定。在激活阶段利用 Sigmoid 函数将输出特征归一化到 0~1 之间,再与主干分支完成点乘,这样会使原始特征图上每一个点的数值变得很小,输出响应变弱,有效特征不明显,从而导致网络可能出现退化现象。

针对以上 2 个问题,通道注意力的改进结构如图 6 所示。以输入特征图 F_1 为例,先对 F_1 特征在通道方向上分别进行全局平均池化和全局最大池化,并将池化结果按通道进行拼接,此时压缩后的特征向量为 $1 \times 1 \times 2C$ 。利用 1×1 的卷积核调整通道为 $1 \times 1 \times C$,经过第 1 个全连接进行特征融合和降

维,并使用 ReLU 函数进行激活,此时将 C 个通道压缩成 C/r 个通道,通过全局特征编码,达到降低计算量的目的;第 2 个全连接层将特征通道数恢复成 C 个通道并用 Sigmoid 函数激活,进而得到通道权重 M_c 。其中 F 为残差学习,将通道权重 M_c 与特征图 F_1 先进行残差连接再按通道相乘,即 $F = (1 + M_c) F_1$,得到最终的通道注意力特征,以确保网络不会出现退化问题。

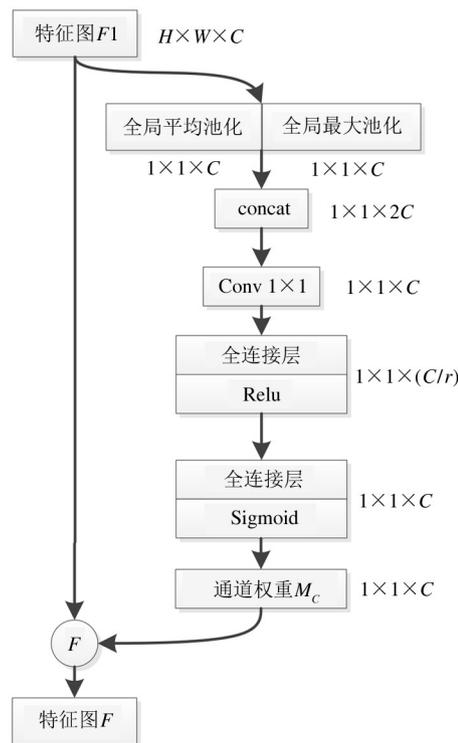


图 6 改进后的通道注意力模块结构

本文利用全局平均池化和全局最大池化的并行组合结构替代原始的全局平均池化,使网络在关注全局特征的同时又兼顾了特征图中响应最大的局部特征;利用 1×1 的卷积核用较少的参数量完成特征的压缩;同时将残差学习的思想引入通道注意力中,以保证引入的注意力机制不会导致网络退化,避免了 SENet 中输出权重归一化与主干分支进行点乘会导致降低特征值的问题。将 M_c 作为特征图 F_1 的选择器,使特征图中的有效特征增强而噪声被抑制,将基于残差学习的通道注意力模块记为 reSE;利用 reSE 模块对多尺度特征进行通道间的重新标定,以达到增强有效通道特征、抑制冗余通道特征的目的。考虑到通道拼接方式的特征融合会给模型带来丰富

的特征,此时利用 reSE 模块可以对特征更好地进行选择,对特征图中的有效特征进行增强。本文选择在模型的颈部上采样和下采样的特征融合区域之后分别嵌入该模块并进行对比实验。

3 实验与结果分析

实验平台为操作系统 Windows 10, 学习框架 Pytorch 1.4, GPU 为 Nvidia TITAN Xp, 显存为 12 GB。

3.1 实验数据集

本文实验数据集采用 Tsinghua-Tencent 100K Tutorial (TT100K) 交通标志数据集^[24], 剔除未标注和标注错误的图片, 训练集和验证集共包含 9170 张图片, 实例数超过 100 的共有 43 类交通标志, 分别为 pn、pne、i5、p11、pl40、pl50、pl80、io、pl60、p26、i4、pl100、pl30、il60、pl50、i2、w57、p5、p10、ip、pl120、il80、p23、pr40、ph4. 5、w59、p12、w55、p3、pm20、pl20、pg、pl70、pm55、il100、p27、w13、p19、ph4、ph5、p6、pm30 和 w32。随机抽取 1500 张图片作为测试集, 剩余 7670 张图片按照 8:2 进行训练集和验证集的划分。针对 TT100K 数据集中存在不同类别样本数目不平衡问题产生的长尾效应, 本文对数目实例数少于 200 的类别利用 Mixup 数据增强^[25]进行数据扩充。Mixup 数据增强将 2 张不同的图像混合生成新图像, 新样本的标签由原标签混合而来, 实现方式如式 (3) 所示。最终得到的训练集数据共有 14 460 张图片。

$$(x_n, y_n) = \lambda(x_i, y_i) + (1 - \lambda)(x_j, y_j) \quad (3)$$

式中, $\lambda \in [0, 1]$, 其取值符合 Beta 分布 $\beta(\alpha, \alpha)$, 参数 α 满足 $\alpha \in (0, \infty)$; (x_i, y_i) 和 (x_j, y_j) 是训练集中随机抽取的 2 个样本。图 7 为不同混合系数 λ 的 Mixup 数据增强示意图。

3.2 锚框参数的设置

YOLO v4 模型预设的 9 个目标框 (12, 16; 19, 36; 40, 28; 36, 75; 76, 55; 72, 146; 142, 110; 192, 243; 459, 401) 涵盖了大到汽车、火车, 小到鸟类、盆栽等多尺度的目标尺寸, 而对于目标大小较为极端的数据集, 需要重新对边界框聚类分析。

本实验使用 K-means ++ 聚类算法^[26] 对 TT100K

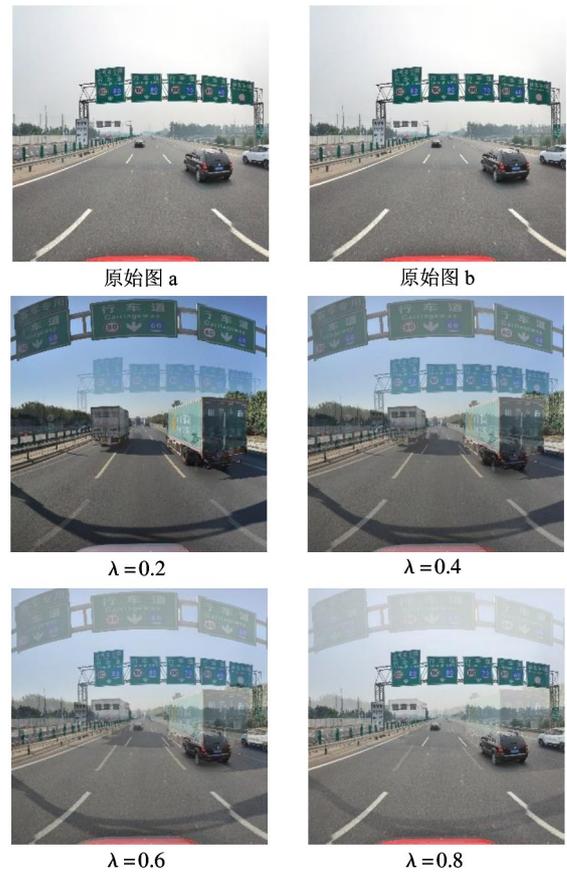


图 7 不同混合系数 λ 的 Mixup 数据增强图

数据集进行聚类分析。与传统的 K-means 算法^[27] 相比, K-means ++ 优化了初始点的选择, 能显著改善分类结果的误差, 以获得更适合的 Anchor box, 提高交通标志检测的精度。K-means ++ 算法首先随机选取某一个目标框作为初始聚类中心 C_1 ; 然后计算每个样本与已有聚类中心点的距离 $D(x)$, 并计算每个样本被选为下一个聚类中心的概率 $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$, 以此类推直至选出 k 个聚类中心; 后不断重复计算每个样本到 k 个聚类中心的距离, 并将它划分到最近的聚类中心再重新计算聚类中心, 直至得到的 Anchor 大小不再发生改变。在输入图片尺寸为 608×608 时, 聚类结果如图 8 所示, 分别用不同的颜色标注出聚类簇, 其中 9 个星星代表每个聚类簇的聚类中心, 其坐标即为初始候选框的宽高 $[5 \times 6, 7 \times 8, 10 \times 10, 8 \times 15, 13 \times 14, 17 \times 19, 24 \times 27, 35 \times 37, 56 \times 56]$ 。

3.3 评价指标及模型训练

以精确度 (precision, P), 召回率 (recall, R) 和平

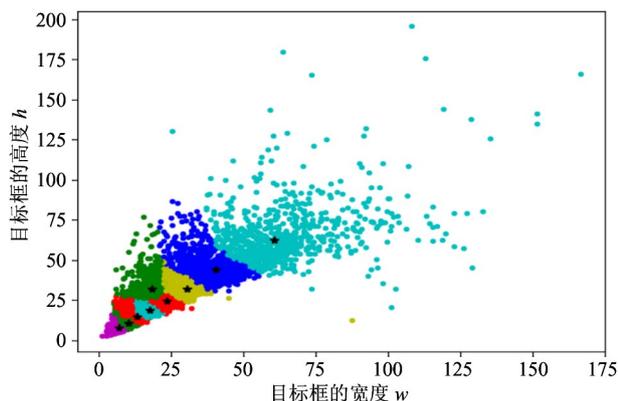


图8 k-MEANS ++ 聚类可视化结果

均检测精确度 (mean average precision, mAP) 作为评价指标。精确度 (P) 和召回率 (R) 的公式分别为式(4)和(5)所示。

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

其中, TP 为实际为正样本被系统判定为正样本, FP 为实际是负样本但是被系统误判为正样本, FN 为实际为正样本但被系统误判为负样本。本文设置 0.5 的置信度阈值, 对比模型的精确度和召回率。

设置初始学习率为 0.001, 每 25 次迭代后将学习率衰减为原来的 0.1 倍, 动量为 0.9, 衰减系数为 0.0005, 使用 SGDM 梯度优化算法, 批处理 (batch-size) 前 50 次迭代设为 4, 后 50 迭代设为 2, 设置迭代次数 (epochs) 为 100 次。

3.4 实验结果对比及分析

本文在 YOLO v4 的基础上提出了 reSE-BYOLO v4, 并在 TT100K 交通标志数据集上进行了大量实验及消融实验。其中包括 2 部分: 第 1 部分对 reSE-BYOLO v4 的各个改进点进行对比实验; 第 2 部分对模型整体的改进策略进行效能评估及分析。

3.4.1 对比实验及分析

在输入图片分辨率为 608×608 时, 本节对改进后的 YOLO v4 模型进行实验分析。其中主要包括检测层改进、多特征融合以及 reSE 模块的消融实验 2 部分, 其实验结果如表 2 和 3 所示。

表 2 为本文 3.1 和 3.2 节提出的模型优化实验数据, 其中 YOLO v4-A 是在 YOLO v4 基础上舍弃

19×19 的检测层, 增加了 152×152 的检测层。YOLO v4-B 是在 YOLO v4-A 的基础上增加了同一尺度下特征的横向连接和对高分辨率下特征信息的融合。从实验可以看出, YOLO v4-A 相比于原始 YOLO v4 其 mAP 提升了 2.3%, 表明将原始模型大目标检测层去掉, 并不影响检测交通标志, 同时增加高分辨率的检测层能够利用图像中小目标的信息, 更好地检测交通标志; YOLO v4-B 相比于原始 YOLO v4 其 mAP 提升了 4.53%, 检测精度优于 YOLO v4-A, 这表明多特征融合能够在不增加模型复杂度的情况下使模型融入更多的特征, 从而进一步提高模型的检测精度。

表2 YOLO v4 模型优化对比实验

模型	检测层改进	多特征融合	mAP/%
YOLO v4	--	--	73.91
YOLO v4-A	✓	--	76.21
YOLO v4-B	✓	✓	78.44

表 3 主要研究 SE 和 reSE 2 种模块嵌入在 YOLO v4 模型的不同特征融合位置所带来的性能变化, 其中“Upsampling_concat”表示在每一个颈部上采样特征融合区域之后添加该模块, 而“Downsampling_concat”则是每次完成下采样特征融合之后添加。

表3 通道注意力消融实验

实验	注意力模块	Upsampling_concat	Downsampling_concat	mAP/%
1	SE	✓	--	74.12
2		--	✓	75.69
3	reSE	✓	--	75.84
4		--	✓	77.02

从表 3 可以看出, 注意力模块嵌入在模型的不同特征融合位置带来的性能有所不同, 从实验 1、2 和实验 3、4 的对比中, 可以发现在“Downsampling_concat”处添加注意力模块要比“Upsampling_concat”处效果提升明显, 本文认为网络更深层的特征图表现出更丰富的语义特征, 特征信息高度融合, 每

个通道上的特征分布更加明显。从实验 1、3 和实验 2、4 的对比中,可以发现 reSE 模块所带来的性能提升要优于 SE 模块,进一步说明了将残差学习的思想引入注意力模块中能够帮助模型对有效特征的选择。实验 4 相比于原始 YOLOv4 提升了其 mAP 3.1%,表明 reSE 模块的嵌入有效提升模型对交通标志的检测性能。

3.4.2 reSE-YOLOv4 模型性能测试以对比实验

结合 3.4.1 的实验结果,本文发现 YOLOv4-B 和实验 4(即表 4 中的“reSE-YOLOv4”)都可以有效

表 4 reSE-BYOLOv4 和相关实验结果对比

模型	P/%	R/%	mAP/%
YOLO v4	80.48	66.82	73.91
YOLO v4-B	84.91	68.61	78.44
reSE-YOLO v4	83.59	68.10	77.02
reSE-BYOLO v4	85.26	72.08	80.48

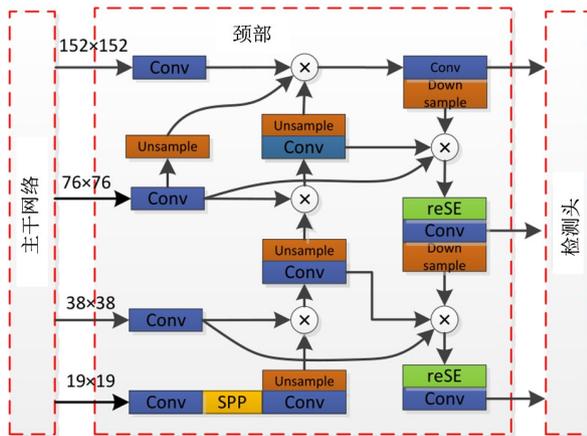


图 9 reSE-BYOLO v4 模型结构图

提升交通标志的检测性能,因此将两者相结合,提出一种基于残差注意力的多特征融合的交通标志算法——reSE-BYOLOv4 模型,该模型具体结构如图 9 所示(图中“⊗”代表特征融合的 concat 操作)。

从表 4 中可以看出,本文提出的改进方法检测效果均比 YOLO v4 模型好,其中 reSE-BYOLO v4 模型在交通标志检测中较 YOLO v4 相比其 mAP 提升了 6.57%,并对比于其他改进模型, reSE-BYOLO v4 达到了 80.48% 的最高 mAP 值,并且其准确度和召回率均有不同程度的提升。实验证实了本文所提出的改进方法能够有效提升对交通标志的检测。

表 5 展示了 reSE-BYOLO v4 模型与其他模型的性能效率对比。模型的运行环境均为 Nvidia TITAN Xp GPU,由于训练数据量较大耗时普遍较长,在相同输入尺寸 608 × 608 时,检测效果最佳的 reSE-BYOLO v4 模型训练时间相比于原 YOLO v4 额外增加约 1 min,以牺牲很小的训练时间来换取更好的检测效果,检测速度相比于原 YOLO v4 慢了 0.02 s,但仍满足实时检测要求。

表 5 模型性能效率对比

尺寸	模型	训练时间/epoch	时间/s
608 × 608	YOLO v4	6'55"	0.05
	YOLO v4-B	7'28"	0.06
	reSE-YOLO v4	7'42"	0.05
	reSE-BYOLO v4	8'07"	0.07

图 10 展示了 4 种模型在 TT100K 测试集上对 43 类交通标志识别的 AP 值,因为训练数据分布存

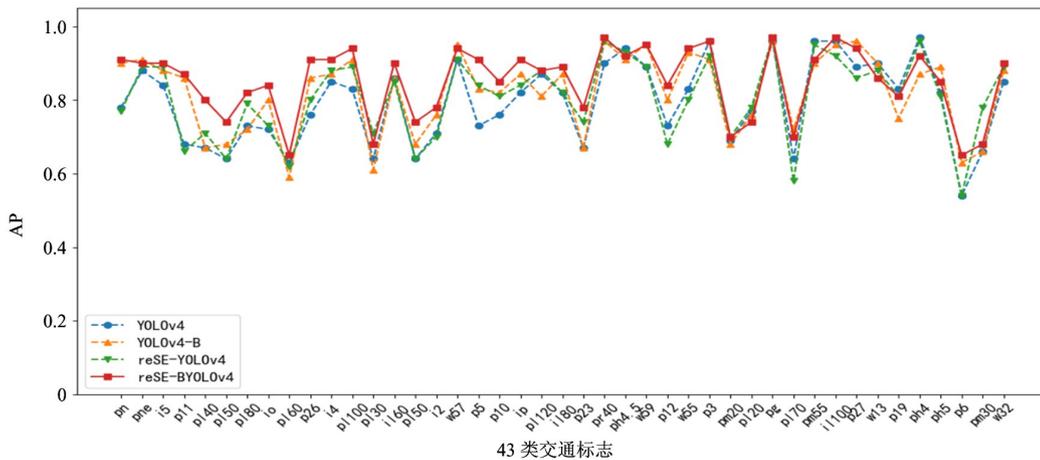


图 10 不同模型的 AP 对比

在差异,即使同一个模型在相同的训练策略下训练2次也会出现同一类别 AP 值的小波动。从图中可以看出, YOLO v4-B 和 reSE-YOLO v4 模型相比于 YOLO v4 模型整体 AP 略有提升,极少类别 AP 出现下滑,属于数据的正常波动。本文提出的 reSE-BY-OLO v4 模型相比于其他模型在测试各个交通标志 AP 值上表现更优,证明了该模型的在交通标志检测上的有效性。

3.5 图片检测效果

为了证明 reSE-BY-OLO v4 的检测效果,从数据集集中随机抽取了3张图片进行比较并将图片较小部分进行局部放大。

从图 11 可以看出,子图(a)只检测出了尺寸较大的6个交通标志,忽略了较小的1个交通标志;子图(b)检测出了较小的交通标志;子图(c)中出现了漏检情况,漏检了2个交通标志;子图(d)中可以检

测出图中的所有的交通标志;子图(e)中只检测出了2个较大的交通标志;子图(f)则将远处较小的2个交通标志也检测了出来。因此 reSE-BY-OLO v4 模型大幅提高了小目标交通标志的成功率。

3.6 reSE-BY-OLO v4 模型与其他模型对比

为进一步验证该模型的有效性,将 reSE-BY-OLO v4 模型与其他先进的模型进行对比实验,实验结果如表 6 所示。

从表 6 可以看出, reSE-BY-OLO v4 相比两阶段检测的 Faster-RCNN,其 mAP 提升了 10.95%,精准度和召回率均有大幅提升;相比于基于 Anchor-free 系列的 CenterNet、FCOS 其 mAP 分别提升了 8.03%、4.61%,精准度和召回率也均有不同程度的提升。文献[13]以 YOLO v3 为基础网络通过融合 SPP 结构和改进 FPN,使 mAP 达到 75.20%。文献[14]利用带有反卷积级联结构的特征提取网络,提高了小目标的检测能力,使 mAP 高达 77.45%,仅次于 reSE-YOLO v4 模型。实验证明,本文所述的 reSE-BY-OLO v4 模型对小目标交通标志有更好的检测效果。

表 6 不同模型实验对比

尺寸	模型	P/%	R/%	mAP/%
608 × 608	YOLOv4	80.48	66.82	73.91
	CenterNet-Resnet50	78.54	69.31	72.45
	FCOS-Resnet101	81.34	70.72	75.87
	Faster-RCNN	79.65	65.41	69.53
	文献[13]方法	81.92	69.07	75.20
	文献[14]方法	83.04	72.10	77.45
	reSE-BY-OLOv4	85.26	72.08	80.48



图 11 效果对比图

4 结论

本文基于 YOLO v4 提出一种 reSE-BY-OLO v4 模型。针对多数交通标志尺寸偏小的问题,优化检测层并在特征融合区域引入同一尺度特征的横向连接,使模型学习更多的特征,有效提高了对小目标的检测能力;同时将 reSE 模块和 YOLO v4 模型进行结合,使模型从诸多特征中选择有效特征来进行学习。但 reSE-BY-OLO v4 会增加模型的参数量,下一步研究中将会考虑对模型进行压缩和剪枝,在保证

算法精度的基础上提高模型的实时性。

参考文献

- [1] MOGELMOSE A, TRIVEDI M M, MOESLLND T B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012,13(4):1484-1497
- [2] REINA A V, SASTRE R J L, ARROYO S L, et al. Adaptive raffic road sign panels text extraction[C]//Proceedings of the 5th WSEAS International Conference on Signal Processing, Robotics and Automation, Stevens Point, USA, 2006: 295-300
- [3] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998,86(11):2278-2324
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014:580-587
- [5] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE International Conference on Computer Vision, Santiago, Chile, 2015:1440-1448
- [6] REN S Q, HEK 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*, 2015,39(6):1137-1149
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016:779-788
- [8] REDMON J, FARHADI A. YOLO 9000: better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2017:7263-7271
- [9] REDMON T, FARHADI A. Yolov3: an incremental improvement [EB/OL]. <https://arxiv.org/pdf/1804.02767.pdf>; arXiv, (2018-04-08), [2021-01-28]
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. <https://arxiv.org/abs/2004.10934>; arXiv, (2018-04-08), [2020-01-28]
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // European Conference on Computer Vision, Amsterdam, Netherlands, 2016:21-37
- [12] 王林, 张文卓. 一种融合注意力机制与上下文信息的交通标志检测方法[J]. *计算机测量与控制*, 2022,30(3):54-59
- [13] 刘紫燕, 袁磊, 朱明成, 等. 融合 SPP 和改进 FPN 的 YOLOv3 交通标志检测 [J]. *计算机工程与应用*, 2021, 57(7):164-170
- [14] 李焯, 顾晨峰. 基于注意力机制的多尺度小目标交通标志检测[J]. *小型微型计算机系统*, 2022,43(2):381-386
- [15] AGHDAM H H, HERAVI E J, PUIG D. A practical approach for detection and classification of traffic signs using convolutional neural network [J]. *Robotics and Autonomous Systems*, 2016, 84:97-112
- [16] HOUBEN S, STALLKAMP J, SALMEN J, et al. Detection of traffic signs in real-world images; the German traffic sign detection benchmark[C]//The 2013 International Joint Conference on Neural Networks, Piscataway, USA, 2013: 1-8
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018,42(8):2011-2023
- [18] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSP-Net: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, USA, 2020: 390-391
- [19] 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, Piscataway, USA, 2017: 2117-2125
- [20] HE K, ZHANG X, REN S, 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
- [21] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // Proceedings of The IEEE Conference on Computer Vision And Pattern Recognition, Salt Lake City, USA, 2018: 8759-8768
- [22] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//IEEE Conference on Com-

- puter Vision and Pattern Recognition (CVPR), Seattle, USA, 2020: 10778-10787
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132-7141
- [24] ZHE Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2110-2118
- [25] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. <https://arxiv.org/abs/1710.09412>; arXiv, (2017-10-25), [2021-01-28]
- [26] ARTHUR D, VASSILVITSKII S. K-Means + + : the advantages of careful seeding[C]//Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, USA, 2007: 1027-1035
- [27] HARTIGAN J A, WONG M A. A K-means clustering algorithm: algorithm as 136[J]. *Applied Statistics*, 2013, 28(1): 100-108

Traffic sign detection based on multi-feature fusion attention mechanism

LI Guoqiang, FU Le, ZHA Linlin, WANG Tianlei

(College of Electrical Engineering, Yanshan University, Qinhuangdao 071000)

Abstract

Aiming at the problems of small size and inconspicuous feature of traffic sign target detection samples in a complex environment, based on the YOLO v4 algorithm, a method based on multi-feature fusion attention mechanism is proposed to form reSE-BYOLO v4 algorithm. The reSE-BYOLO v4 algorithm firstly improves the detection layer by using the feature information in high resolution to improve the detection ability of small targets. In the feature fusion part, the same scale feature information is used to construct the horizontal connection, so that the model can fuse more features without increasing the cost. The effective features are enhanced by reSE module, which optimizes SE module and recalibrates channel dimension by residual learning. The test results on TT100K dataset show that compared with YOLO v4 algorithm, reSE-BYOLO v4 algorithm has improved mean average precision (mAP) by 6.57%, average accuracy by 4.78%, and average recall by 5.26%. It can be seen that reSE-BYOLO v4 has a stronger ability to recognize traffic signs.

Key words: traffic sign detection, YOLO v4, attention mechanism, residual learning, feature fusion