

基于注意力机制和特征融合的交通标志检测

韩 珊¹ 朱立忠¹

HAN Shan ZHU Lizhong

摘 要

针对复杂背景信息下不易察觉或相对尺寸较小的交通标志目标,提出一种改进 Faster R-CNN 的新检测算法。首先,用 ResNet50 深度残差网络替代 VGG16 作为基本网络模型,同时在基础网络中嵌入混合注意力机制,使网络专注强化关键位置的信息,抑制无效背景信息对网络的干预,从而加强网络对重要特征的提取;其次,提出改进特征金字塔网络结构进行多尺度融合上下文信息的方法,利用特征层之间的关联性减少低像素小目标的信息损失。改进后的算法在 CCTSDB 数据集上的实验取得了 95.6% 的平均检测精度,较原 Faster R-CNN 网络的检测精度提高了 5.1%,鲁棒性增强。

关键词

交通标志检测; Faster R-CNN; 残差网络; 注意力机制; 特征金字塔

doi: 10.3969/j.issn.1672-9528.2024.09.001

0 引言

智能驾驶系统的迅速发展得益于人工智能在汽车领域内的广泛应用和技术的不断革新。智能驾驶技术基本形成了以“感知-决策-执行”为核心的系统框架^[1]。智能驾驶系统是一个集成了多种先进技术的复杂系统,运用识别技术感知周围环境,主要依靠摄像头、激光雷达和毫米波雷达等传感器实现。通过机器学习算法不断优化驾驶决策和控制系统,使车辆能够做出合理的决策,比如选择最佳的行驶路径或应对突发的交通状况。除此之外,系统还依赖全球定位系统实现精准导航和实时定位,以及车辆智能网联、车辆间信息交互等技术。其中,检测与识别技术是智能驾驶系统的关键,智能汽车依靠车载搭配的各类传感器感知外界事物。限速禁停等多种交通标志、交通信号灯以及行人和非机动车辆的识别检测都属于感知范畴,这就意味着在整体架构中检测算法需要有高精度和较强的泛化能力。

交通标志检测算法作为可自主监督学习的深度学习算法,针对不易察觉的道路标识或绝对尺度较小的交通标志目标,能够辅助驾驶员及时应对复杂环境的变化,最大程度规避可能出现的道路安全问题,避免发生严重的交通事故。

1 研究现状

传统方法受外界限制很多,机器学习法在人工提取特征的过程中也会有局限性。深度学习法因其强大的特征学习能

力,使卷积神经网络模型(convolutional neural network, CNN)^[2]成为深度学习在交通标志目标检测领域里主流模型之一。Shao 等人^[3]提出了一种基于简化 Gabor 小波(SGW)和最大稳定极值区域(MSERs)的第一步区域求解算法,并将其用于改进 Faster R-CNN,针对小目标问题提出一种结合 VGG16 第三层、第四层和第五层特征来丰富小目标特征的方法,并提高了 Faster R-CNN 对目标特征的分类能力;李禹纬等人^[4]提出一种改进 YOLOv7 的轻量化交通标志检测算法,在骨干网络引入大核卷积并提出集中综合深度可分离卷积模块,在检测颈部融合坐标注意力、随机池化等方法,既能构建通道注意力,又能捕捉准确位置,还能提升网络的泛化能力;赵子婧等人^[5]在 Libra R-CNN 网络中引入 GA-RPN 来生成锚框,基于这一改进在训练过程中能产生更为精确且多样化的样本,从而大大改善了交通标志检测对于目标难以定位的问题;Dewi 等人^[6]使用生成对抗网络(GAN)来生成更逼真和多样化的训练图片,在实验中将真实图像与 LSGAN 生成的合成图像混合后,识别性能得到提升,YoloV3 准确率为 84.9%,YoloV4 准确率为 89.33%;Liu 等人^[7]提出一种基于 RetinaNet-NeXt 的交通标志检测方法,采用新型骨干网络 ResNeXt 提高了 RetinaNet 的检测精度和效果,并且采用迁移学习和群体规范化来加速网络训练,同时对数据进行了清理和增强降噪。

本文选择 Faster R-CNN 网络模型作为主网络结构,在基本架构上改进研究与实验。用 ResNet50^[8]深度残差网络替代 VGG16 作为基本网络模型,并在其 conv2、3 和 4 层中引入混合注意力机制;设计了基于特征金字塔结构的融合相邻层特征的网络结构,实验结果验证了改进后 Faster R-CNN 算法

1. 沈阳理工大学自动化与电气工程学院 辽宁沈阳 110159

[基金项目] 国家重点研发计划(2017YFC0821001-2)

的 mAP 高于原 Faster R-CNN 算法的 mAP。

2 网络改进

2.1 基础网络模型调整

相较于 VGG16 网络, ResNet50 具有更深层的网络结构, 同时它的深度残差结构使其在堆叠层数更多的情况下, 不仅能保证信息的流通, 还不会存在梯度失衡的风险, 模型的性能也会被优化。因此在考虑到交通标志检测任务中, 对模型的精确度和实时性的要求较高, 故选择 ResNet50 作为整个网络的基础网络模型。

2.2 注意力机制模块

车辆在现实场景下接收到的路面信息是杂乱的, 交通标志易受道路两侧树干枝叶、绿化装饰物的遮挡, 而且车辆在收集图像信息时也会受到距离、高度等限制, 这阻碍了中小尺寸目标检测与识别任务的进展。为了显著增强模型的表征能力, 并筛选出不必要的背景信息, 现对 resnet50 的残差结构进行修改, 从而提升模型对中小尺寸目标识别的精准度。如图 1 所示, 即在残差网络的卷积后引入注意力机制模块, CBAM 模块包含通道注意力模块 (channel attention, CA) 和空间注意力模块 (spatial attention, SA)^[9]。引入后在卷积层中实现通道和空间两个维度的联合作用, 同时这两个作用域结合 ResNet50 的残差结构, 不断学习挑选重要位置信息, 并减弱对无效信息的关注, SA 在不改变通道数的情况下, 通过专注有效信息从而使交通标志漏检误检率下降。

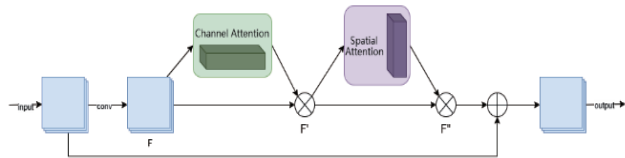


图 1 引入 CBAM 结构图

2.2.1 通道注意力机制

每个特征通道对于特征图上的目标重点信息的关注度都存在差别, 通道注意力模块利用这一特点, 通道间存在互相关联的特性, 从而筛选出贡献率相对较大的重要信息。

如图 2 所示, 首先该模块通过最大池化和平均池化对输入的特征图 F 降维, 并使用多层感知机 (MLP) 来对网络权值进行共享, 再将得到的两个特征相加合并, 激活函数 Sigmoid 后获得通道权重。

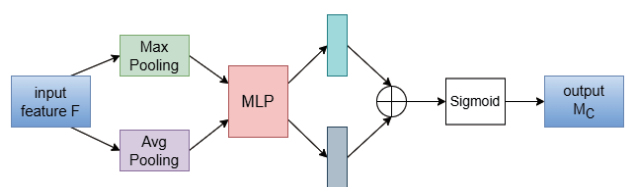


图 2 通道注意力结构图

通道注意力计算公式为:

$$M_c = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (1)$$

式中: σ 表示 Sigmoid 激活函数, W_0 和 W_1 表示 MLP 的权重参数。

2.2.2 空间注意力机制

在空间维度中包含了大量的目标信息和背景信息, 空间注意力模块能在复杂背景信息下筛选出对目标关键信息贡献率更大的空间特征信息, 用于输出空间注意力权重, 如图 3 所示。

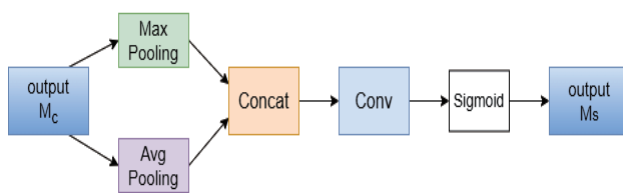


图 3 空间注意力结构图

空间注意力模块把通道注意力模块的输出特征图作为输入, 首先设置最大池化和平均池化分别得到两个二维空间特征图, 并对两个池化特征图进行通道拼接和叠加融合, 再进行 7×7 尺寸卷积降维, 进入函数 Sigmoid 激活后获得 M_s 。空间注意力计算公式为:

$$M_s = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^c; F_{max}^c])) \quad (2)$$

式中: σ 表示 Sigmoid 激活函数, $f^{7 \times 7}$ 表示 7×7 卷积。

2.3 改进 FPN 特征金字塔

在改进之前的 FPN^[10] 结构中, 每一层都自上而下进行一次上采样的操作, 并与卷积层的输出进行相加融合, 如图 4 所示。这种将残差结构与 FPN 结构结合的特征融合方式是很巧妙的, 将分别经过卷积处理和采样处理过的特征在 FPN 层进行相加融合, 这样不仅位于顶端的特征可以获取到更全面的位置信息, 同时也可以输出有较强关联性的信息供预测网络使用。

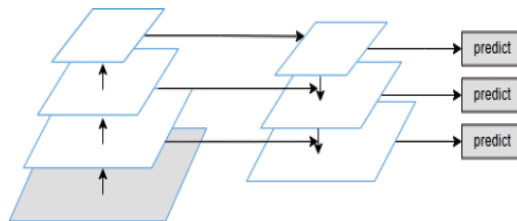


图 4 特征金字塔结构

基于这样的网络结构, 本文在改进的 FPN 中, 对越级特征层的特征进行再次利用, 采用对 M2 特征层再次融合 M3、

M4 特征层的细节信息,对两个特征层之间进行逐级上采样的操作,将邻近的两层向下特征融合,使得能够重复利用到相邻层的语义信息。将 FPN 结构中的 F5' 特征进行四倍上采样的操作,得到与 M2 特征层相同的尺度,并对其进行相加融合;同时将 M4' 采样,同样得到与 M2 特征层相同的尺度,融合三个特征层后,最后经过卷积处理后作为 P2 预测层的输入。改进后的网络结构如图 5 所示。

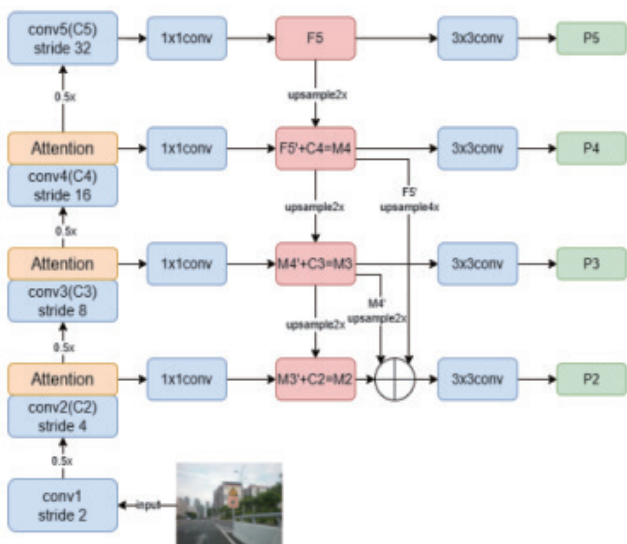


图 5 引入 CBAM 和改进 FPN 网络结构图

3 实验

3.1 数据集

数据集选用由长沙理工大学团队制作的 CCTSDB 交通标志数据集,包含 10 000 多张含有交通标志的图片,其中主要包括我国城市道路上三类常见的交通标志,分别是指示标志(mandatory)、禁令标志(prohibitory)及警告标志(warning)三大类。实验选取数据集内 6000 张图片,并对其中部分图像进行仿射变换、亮度调整等数据增强处理,并按照 5:1 的比例划分训练集和测试集。

3.2 实验环境及参数设置

硬件环境为 Intel(R) Core(TM) i7-7700HQ CPU、NVIDIA GTX 1080 GPU。软件环境为 Windows10 系统、Python3.7、CUDA11.6 以及 PyTorch 深度学习框架。动量值设为 0.9,初始学习率设为 0.005,学习率衰减值为 0.000 5。

3.3 性能评价指标

实验采用的模型性能评价指标为三类交通标志的平均精度(Average Precision)、精确率(Precision)和召回率(Recall)。同时,当交通标志检测边框与真实框相关性,即交并比 IOU 大于 0.7 时,检测结果记为正样本,小于 0.3 时记为负样本。平均精度 AP、精确率 P 和召回率 R 的表达式为:

$$AP = \int_0^1 P(R) dR \quad (3)$$

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

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

式中: TP 表示交通标志被检测为正样本, FN 表示交通标志被检测为负样本, FP 表示负样本被检测为正样本。

3.4 结果与分析

如表 1 所示,与原 Faster R-CNN 网络相比,采用深度残差网络 ResNet50 以及改进特征金字塔网络的改进算法平均检测精度提高了 3.1%。由此可见,特征融合算法能够将细节信息再次利用,并提高对目标的检测效果。CBAM 机制的引入将更多注意力放在有用信息上,在没有增加过多检测时间的基础上, mAP 值比原网络提高了 5.1%。

表 1 平均检测精度结果

Method	mAP/%	P/%	R/%
Faster R-CNN	90.50	78.43	89.45
Faster R-CNN+G-FPN	93.63	86.31	93.17
Faster R-CNN+CBAM+G-FPN	95.60	89.67	93.88

图 6 为原网络与改进网络对三类交通标志目标的类别 AP 值对比结果。其中,改进后禁止标志取得了 94.41% 的检测精度,较原网络提高了 8.1%。

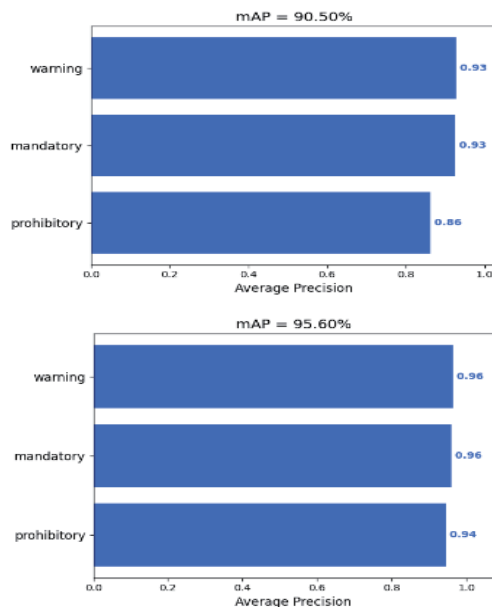


图 6 类别 AP 比较数据

为了验证算法的可靠性,使用相同数据集和环境下,对另两种深度学习模型进行对比实验,对比结果如表 2 所示,本文算法的 mAP 值均高于另两种模型。

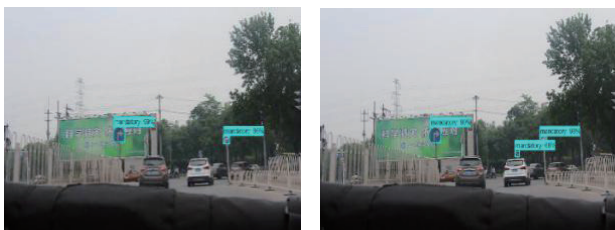
表 2 对比实验结果

Method	mAP/%	Time/s
SSD	86.20	0.028
YOLOv3	90.81	0.025
原网络	90.50	0.037
本文算法	95.60	0.039

在交通标志检测任务中，不仅要克服小尺寸目标的检测问题，还要面对不同光线和天气变化等复杂情况，因此在测试时重点关注了这几类目标的检测效果，现举例展示改进前后算法的检测效果。

（1）小尺寸目标检测效果

由图 7 所示，图片中包含三个指示标志牌，原算法检测出了左右两个标志牌，并未检测出中间较小尺寸的标志牌。本文算法则顺利检测出了三个标志牌，可见本文算法对远距离较小标志的检测效果增强。



(a) 原 Faster R-CNN 算法 (b) 本文算法

图 7 改进前后算法对多目标的检测

（2）图片较暗及有遮挡目标检测效果

由图 8 所示，原算法并未检测出中间存在遮挡的指示标志牌，存在漏检现象，本文改进算法不仅在弱光的场景下顺利检出标志牌，同时被路边物体遮挡的标志也被检测出来，因此可见本文算法在面对较为复杂的路面信息时也具备良好的性能。



(a) 原 Faster R-CNN 算法 (b) 本文算法

图 8 改进前后算法对遮挡目标的检测

4 结论

本文针对处于交通标志图像较远位置且目标较小的交通标志，提出了一种改进的网络结构的 Faster R-CNN 算法，嵌入混合注意力机制的 ResNet50 深度残差网络，能够专注地强化重要位置的特征，减少背景信息的干预。通过改进特征金字塔网络结构进行多层次联合预测前的输出信息的方法，使

得提取层的深、浅层语义信息能够被更全面地利用，从而提高算法对交通标志的检测效果。在 CCTSDB 数据集上的实验结果表明，相较于原 Faster R-CNN 算法，本文方法 mAP 值达到了 95.6%，提升了 5.1%。

参考文献：

- [1] 王晓,张翔宇,周锐,等.基于平行测试的认知自动驾驶智能架构研究[J].自动化学报,2024,50(2):356-371.
- [2]KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image-Net classification with deep convolutional neural networks[J]. Communication of the ACM,2017,60(6):84-90.
- [3]SHAO F, WANG X, MENG F, et al.Improved faster R-CNN traffic sign detection based on a second region of interest and highly possible regions proposal network[J].Sensors, 2019, 19(10):2288.
- [4]李禹纬,付锐,刘帆.改进 YOLOv7 的交通标志检测算法[J].太原理工大学学报,2023,32:157-165.
- [5]赵子婧,刘宏哲,曹东璞.基于 Libra R-CNN 改进的交通标志检测算法[J].机械工程学报,2021,57(22):255-265.
- [6]DEWI C, CHEN R C, LIU Y T, et al. Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN[J]. IEEE Access, 2021, 9: 97228-97242.
- [7]LIU S, CAI T, TANG X, et al.Visual recognition of traffic signs in natural scenes based on improved RetinaNet[J].Entropy, 2022, 24(1):112.
- [8]HE K M, ZHANG X Y, REN S Q, et al.Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision Pattern Recognition.Piscataway: IEEE,2016:770-778.
- [9]WOO S, PARK J, LEE J Y, et al.CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision.Berlin:Springer,2018:3-19.
- [10]LIN T Y, DOLLÁR P, GIRSHICK R, et al.Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision Pattern Recognition.Piscataway: IEEE,2017:2117-2125.

【作者简介】

韩姗(2000—),女,吉林长春人,硕士研究生,研究方向:目标检测。

朱立忠(1967—),男,辽宁沈阳人,硕士,教授,研究方向:图像处理、智能控制。

(收稿日期:2024-06-02)