

DOI:10.19651/j.cnki.emt.2209953

自适应特征融合的轻量级交通标志检测方法

梁秀满 邵彭娟 刘振东 赵恒斌

(华北理工大学电气工程学院 唐山 063210)

摘要:针对日前交通标志检测方法中网络计算量大、检测效果差的问题,提出一种嵌入坐标注意力机制的轻量级交通标志检测方法。首先在 MobileNetv2 的残差块中嵌入坐标注意力机制 CA(channel attention)模块以保留通道注意力中的坐标信息;其次利用改进的 MobileNetv2 对 YOLOv4 主干网络做轻量化处理,并且在 PANet 中采用深度可分离卷积块降低计算量;然后使用 ASFF 自适应特征融合改进 PANet 结构来均衡不同特征层的不一致性,最后在特征融合模块加入注意力以增加目标信息的权重;并由 K-means++ 算法产生新的先验框聚类中心。实验表明,权重文件由 136 M 降至 54.5 M 削减了 60%,网络体积削减了 80%,精度达到 96.84%,与 YOLOv4 网络相比仅损失了 0.46% 的精度。

关键词:轻量级网络;注意力机制;聚类算法;自适应特征融合

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6

A lightweight traffic sign detection method based on adaptive feature fusion

Liang Xiuman Shao Pengjuan Liu Zhendong Zhao Hengbin

(School of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China)

Abstract: Aiming at the problems of large amount of network computation and poor detection effect in the current traffic sign detection method, a lightweight traffic sign detection method with embedded coordinate attention mechanism is proposed. First, the coordinate attention mechanism CA module is embedded in the residual block of MobileNetv2 to retain the coordinate information in the channel attention; Secondly, the improved MobileNetv2 is used to lighten the YOLOv4 backbone network, and the depthwise separable convolution block is used in PANet to reduce the amount of computation; Then, ASFF adaptive feature fusion is used to improve the PANet structure to balance the inconsistency of different feature layers. Finally, attention is added to the feature fusion module to increase the weight of the target information; and the K-means++ algorithm generates new a priori box cluster centers. Experiments show that the weight file is reduced by 60% from 136 M to 54.5 M, the network volume is reduced by 80%, and the accuracy reaches 96.84%, lose only 0.46% accuracy compared to YOLOv4 network.

Keywords: lightweight network; attention mechanism; clustering algorithm; adaptive feature fusion

0 引 言

随着交通标志检测网络的不断加深,内存占用逐渐增大,不易应用到智能交通系统和自动驾驶等小型平台。轻量级目标检测结构例如 SqueezeNet^[1]、Mobilenet^[2]系列、以及 Efficient^[3]系列,它们的体积较小,可应用到移动平台,为交通标志检测提供了一定的网络架构参考。交通安全干系重大,在实现实时检测的同时,对检测精度有着更高的标准。交通标志检测框架大致分为两种,最早是 2016 年使用滑框检测的 Fast R-CNN^[4]、Faster R-CNN 等二阶段深度学习框架,使用 RPN 生成建议框并利用卷积进行分类

和检测,精度较高但实时性不能满足交通标志检测的要求;

一阶段的 YOLO^[5]系列、SSD^[6]系列等网络将检测和分类合为一体,真正实现了实时检测,成为了近年来目标检测研究的热点。

张毅等^[7]提出的基于改进 FasterR-CNN 模型的交通标志检测方法,使用 ResNeXt 替换原有主干网络 VGG,并使用多维特征融合来提取目标信息,提升了精度但不能满足对实时性的要求。随着一阶段网络框架的出现,大批量基于 YOLO 系列框架的改进方法应用于交通标志检测。Zhang 等^[8]提出了 YOLOv2 交通标志检测方法,实现了端到端的实时检测,但精度有待提升。马露茜等^[9]提出一种

收稿日期:2022-05-13

复杂环境下交通标志检测方法,在原有 YOLOv3 基础上增加了一个 104×104 的特征检测层,获得更高的精度和速度。张达为等^[10]提出了使用轻量级网络改进 YOLOv3,并引入不同特征层进行融合,精度与以上相比确实提升了,但网络参数量仍需要进一步降低。

陈泉等^[11]提出的基于无预训练卷积神经网络的红外车辆目标检测方法,结合了 YOLOv3 和注意力机制,有助于摆脱对预训练权重的依赖,但网络应用到移动设备仍然具有难度。刘丹^[12]等使用改进的 FCOS^[13]网络对行人进行检测,添加注意力机制增强网络特征提取能力,并利用 GIOU 作为回归损失函数,结果表明注意力在一定程度上可以提升网络检测效果。

针对上述问题,本文提出一种自适应特征融合轻量级交通标志检测方法,基于 YOLOv4^[14]框架使用改进的 MobileNetv2 作为主干特征提取网络,将坐标注意力机制 CA^[15](coordinate attention)嵌入残差块,从而在通道注意力中保留了位置信息。然后使用 ASFF^[16]改进

特征融合结构,更好融合不同特征层的目标信息;在较少计算量的情况下提取到更多的特征信息,保证了精度但没有增加计算量;并在特征融合中加入多个注意力模块 CBAM^[17]形成最终的特征融合模块,实验表明检测效果有一定提升。

1 基于自适应特征融合的交通标志检测方法

1.1 YOLOv4 网络

近几年 YOLO 系列是一阶段中表现较好的网络框架,本文使用的是 2020 年发表的 YOLOv4,它是在 YOLOv3 的基础上改进的,加了许多有效的小技巧,更适合于单 GPU 上训练;并且在主干网络采用跨链接 CSPNet 来进行特征融合保留了更广泛的目标特征;PANet 还增加了一条自上而下的路径使获得特征的梯度更加丰富,扩展了检测空间让特征信息更加完整;SPP 扩大感受野充分捕捉需要的目标特征,最后沿用 YOLOv3 的检测头进行分类和检测,具体网络结构如图 1 所示。

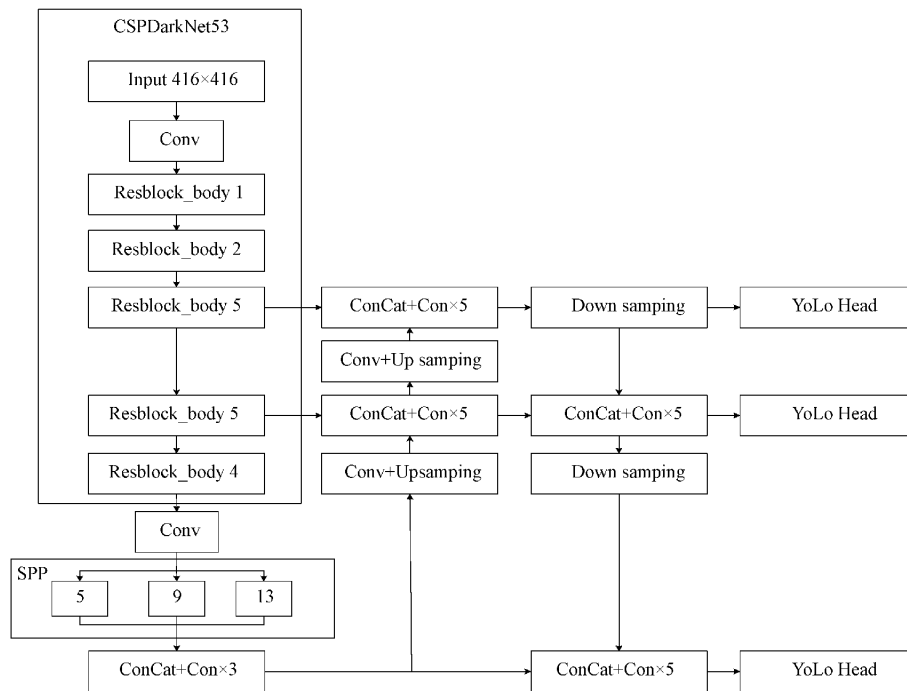


图 1 YOLOv4 网络结构

1.2 轻量化的主干网络

原来 YOLOv4 的网络计算量较大,且网络层数较深,提取到的目标位置信息不够精准。本文中使用轻量级网络 MobileNetv2 代替主干特征提取网络 DarkNet53,整体计算量减少了 80%,更方便应用于移动端和智慧交通系统。2017 年轻量级网络的出现,为深度框架提供了新的网络改进方向。MobileNetv1 中最重要的网络结构是深度可分离卷积:将普通卷积分为两个分卷积:首先经过深度卷积(depthwise convolution),对每个输入通道使用单通道的轻量级滤波器;然后通过点卷积(pointwise convolution)来

结合之前提取到的特征特性,即减少了计算量又保证特征的完整性。

但深度可分离卷积对于输入的通道数不能随意改变,所以大多数特征提取是在低维空间里进行的。而 MobileNetv2 在 v1 的基础上又加入了线性瓶颈的逆残差结构:去除 PW 中的激活函数防止破坏低维空间的线性特征,从而形成 Linear Bottleneck;其先升维再降维的操作来均衡提取高低维度的信息,让提取到的特征更加全面,MobileNetv2 结构如表 1 所示。

本文将与 CSPDarknet53 中相同通道的特征层,融入

表 1 MobileNetv2 网络结构图

输入尺寸	模块	升维倍数	输出通道数	重复次数	卷积步长
416×146	Conv2d	1	16	1	1
208×208	bottleneck	6	24	2	2
52×52	bottleneck	6	32	3	2
26×26	bottleneck	6	64	4	2
26×26	bottleneck	6	96	3	1
13×13	bottleneck	6	160	3	2
13×13	bottleneck	6	320	1	1
1×1×k	Conv2d	—	1 280	1	1

到 YOLOv4 特征融合对应的特征层中,从而实现主干网络的轻量化。

1.3 加入协调注意力 CA 的轻量化网络

MobileNet2 虽实现了参数量的减少,但精度不能满足交通标志检测的要求,所以在其残差块 Inverted Residual Block 中融入坐标注意机制 CA 如图 2 所示,更有效提取感兴趣区域的特征信息。

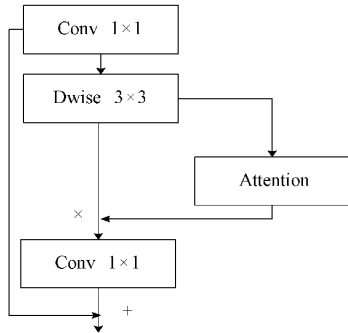


图 2 改进的 block 模块

因为 CA 注意力机制相对于 CBAM、SE 更关注位置信息的特性,更适合应用于特征提取模块。其具体结构如图 3 所示,CA 将通道注意力中嵌入位置信息,保留更多输入特征融合的位置信息,使网络在极少的计算开销情况下提取到更多的特征信息。

如图 3 所示,坐标注意机制主要分为两个部分即注意力的嵌入和注意力的产生。注意力的嵌入:为了避免全局池化的引入带来的位置信息丢失,将通道注意力分为两个平行 1D 特征编码来提升提取坐标信息到特征图的能力。即给定输入 X 从水平和垂直两个方向出发对每个通道编码,因此 C 通道在高度 H 的输出如式(1)所示,同样的宽度为 W 的 C 通道输出如式(2):

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

然后将得到的特征图分别编码为一对方向感知和位

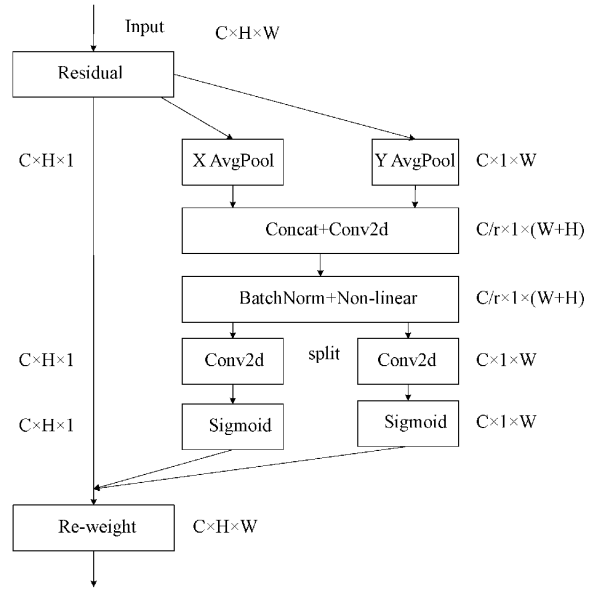


图 3 CA 原理结构

置敏感的注意图层,应用到输入特征层以增强感兴趣对象的表示。

注意力的产生:如式(1)、(2)计算全局特征映射在特征图上的大小来确定精准的位置信息,为了将产生的位置信息传达到网络来精确找到目标信息的区域,需要二次转换。最重要的是获得通道间的依赖关系,先将两个方向结合起来如式(3):

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

然后将空间分为两个独立的张量进行变换,将另外两个 1 * 1 卷积将 f 在水平和垂直方向变换为相同通道的张量输入到 x,即式(4)、(5):

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

最终的坐标注意力可表示如式(6):

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

整个编码过程使得坐标注意力更准确地获取目标对象的位置,让模型更好的检测和分类。产生的计算量极小可忽略不计,可适用于 MobileNet、Efficient 网络的残差块中。

1.4 改进的特征融合模块

原 YOLOv4 中的颈部模块用到的是 PANet、SPP,通过自上而下、自下而上的特征路径,反复融合不同特征层来提取不同特征层的特征信息,SPP 将不同尺寸的特征层进行固定池化,融合多层特征来增加感受野,但 PANet 只是将特征层通过上采样、下采样变换为相同尺度特征层进行融合的,没有充分进行不同特征层之间的融合。将 ASFF 自适应特征融合改进 PANet,通过对不同特征层分配自适应的权重参数来融合多尺度特征图信息,可充分融合特征信息。

如图 4 所示, X1、X2、X3 分别是通过 MobileNet 主干网络得到的通道为 32、96、320 的特征层, 将它们经过特征融合网络之后得到如图 8 所示的 level 1、level 2、level 3 网络层, 以 ASFF-1 为例, 先将 level 2、level 3 通过 1×1 卷积压缩与 level 1 相同的通道数, 然后分别进行 4 倍上采样和 2 倍上采样形成与 level 1 相同维度的特征层, 得到 $resize_level\ 2$ 与 $resize_level\ 3$, 最后利用归一化得到权重参数 $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ 将特征层乘上权重参数如公式 7 所示得到融合后的特征层 y_{ij}^l 。

$$y_{ij}^l = \alpha_{ij}^l \cdot X_{ij}^{l-1} + \beta_{ij}^l \cdot X_{ij}^{2-l} + \gamma_{ij}^l \cdot X_{ij}^{3-l} \quad (7)$$

其中, $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ 三个参数可以自适应低调整而不断更新网络参数, 再通过梯度反向传播更新融合权重, 使得模型的多尺度特征融合更加充分, 更有效地利用特征, 提高模型对于小目标物体的检测能力。

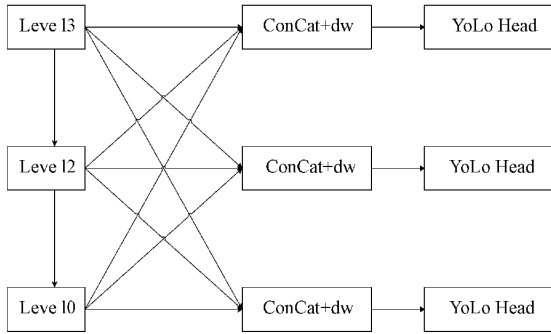


图 4 改进的 PANet 网络结构

1.5 嵌入注意力的特征融合模块

网络越深而导致提取到的底层特征不够全面, 从而检测精度达不到更高的水平。注意力的加入有助于底层特征信息的保留, 其原理是模仿人类有选择的关注目标信息模式, 自适应地去关注一些重要的特征并避开干扰信息, 从而提取到更多的目标特征, 其结构如图 5 所示。

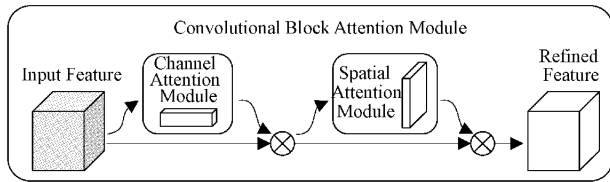


图 5 CA 原理结构

首先将特征变量经过通道和空间两个维度输入, 然后乘上注意力输出到网络, 继而提升获取更多有用信息的能力。CBAM 由两个部分组成即 CAM 和 SAM, 通道注意力 CAM 的结构如图 6 所示, 将输入特征图分别经过平均池化、最大池化两个方向, 之后共同输入到共享的全连接层进行逐元素求和, 然后进行归一化得到每个特征层的权重, 最后乘上输入层输入到 SAM。

而空间注意力 SAM 是对 CAM 的输出层做全局池化和平均池化, 然后进行降维卷积和归一化操作, 最后与输

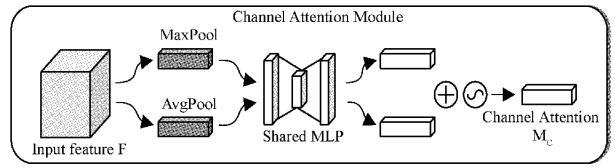


图 6 通道注意力模块

入特征层相乘得到最终的特征, SAM 的结构图如图 7 所示。由于 CBAM 是一个轻量级的通用模块, 所以它可以嵌入任何卷积网络中实现端到端的检测训练。

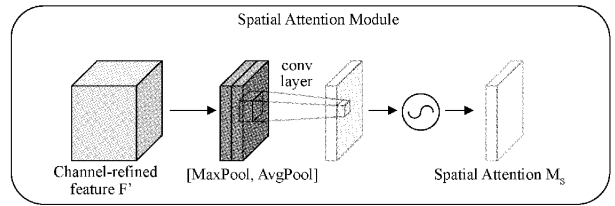


图 7 空间注意力模块

本文中结合 CBAM 与 PANet 形成新的特征融合模块, 具体结构如图 8 所示, 在主干特征输出层 32 倍、16 倍及 8 倍上采样和下采样后加入 CBAM 注意力机制, 来增强通道和空间的特征提取能力, 使目标权重增大, 增强提取目标特征的能力。

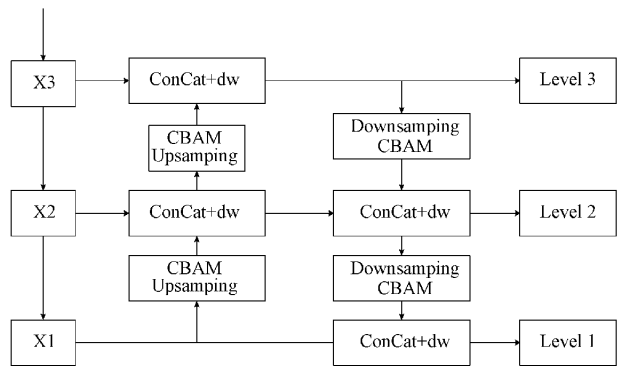


图 8 改进的特征融合网络结构

1.6 K-Means++ 聚类算法优化

最初 YOLOv4 的先验框是根据在 VOC 数据集中采用 K-means 算法得到的, 而当前的数据集中目标物体的长宽和形状与 VOC 有所不同, 所不能使用原来的 anchor box, 该方法将采用 K-means++ 来确定先验框。

本文修改之后的 anchor box 为 (6, 17)、(8, 22)、(8, 15)、(11, 18)、(11, 30)、(15, 24)、(20, 32)、(31, 45)、(59, 83), 使用更贴合本数据集的先验框可降低检测的难度, 加快收敛速度。

2 实验与分析

2.1 实验配置

自适应特征融合的轻量化交通标志检测方法基于

python 语言,使用 Pytorch 深度学习框架进行测试和实验,租用云服务器 GeForce RTX A4000 GPU,python3.8 版本,Cuda11.0,Pytorch1.7.0,显存为 11 G,训练 200 批次,初始学习率设为 0.001,batch_size 设为 16,实验在 IOU=0.5 的条件下进行的,使用到了余弦退火学习率,最后利用迁移学习的思想,借鉴大数据集 COCO 训练出来的权重文件,来加强检测交通标志目标的能力,减低网络收敛的时间。

选取中国交通数据集 CCTSDB 的部分 6 000 张图片,其中有图像模糊、曲折变形、以及来自不同天气的交通标志图片,主要分为警告标志(warning)、禁止标识(prohibitory)、指示标志(mandatory)三大类,按照 8:2 的比例划分测试集和训练集,共选取 1 200 张图片作为最后的测试集。部分数据集如图 9 所示。



图 9 部分交通标志图片

2.2 实验结果与分析

检测速度 FPS 是每秒检测的图片帧数一个代表速度的指标,越大说明速度越快。平均精度 mAP 是一个代表模型准确率指标,越大说明模型表现越好。实验以表明,优化后的网络在增加的计算量较少的情况下,精度提升了 1.4%,权重文件由原来 YOLOv4 训练出来的 136 M 降至 54.5 M,最终的精度可达到 96.84%,此网络因更小的体积且更高的精度可应用到智能交通平台或者驾驶辅助系统。对比不同网络模型如表 2 所示,文献[8]中的模型精度有

所提升,但是网络没有做轻量化处理;文献[9]是针对特殊天气环境下的交通标志检测,但模型精度需要继续提升;文献[10]的检测效果较好,但是应用到实际是网络参数量仍需要优化。

表 2 不同方法模型实验对比表

模型	警告标志	禁止标识	指示标志	mAP/%	FPS
YOLOv4	99%	96%	96%	97.30%	50.2
YOLOv2 ^[8]	96%	94%	96%	95.65%	46.0
YOLOv3 ^[9]	85%	80%	80%	82.73%	34.2
YOLOv3 ^[10]	97%	95%	96%	96.30%	35.2
本文	98%	96%	95%	96.84%	40.2

2.3 实验模型对比

将 YOLOv4、以 MobileNetv2 为主干网络的 YOLOv4,以及融入 CA(CoordAttention)、ASFF、CBAM 等网络模块对比,来研究不同模型对于交通标志检测性能的影响。本实验进行了 5 组实验对比如表 3 所示,分别为 YOLOv4、模型 A、模型 B、模型 C、模型 D。在训练条件相同的条件下,首先模型 A 是轻量化的 YOLOv4 网络,并在颈部使用深度卷积网络代替普通卷积,虽然实现了削减了网络体积的 85%,但精度方面降低了 5.14%,警告、禁止、指示标识分别下降了约 3%、3%、7%;之后模型 B 在原有网络基础上在轻量化网络中融入了注意力机制 CA,保留更多位置信息来提升 1.78%检测精度,增加的计算量极小;其次模型 C 在之前网络对基础上改进特征融合模块,使用自适应特征融合(ASFF)来加强不同特征层之间的融合,并利用自适应检测目标权重来提升检测目标的精确度,增加的计算量忽略不计,但精度提升了 0.98%;最后模型 D 在特征融合模块加入注意力机制 CBAM 来加强通道和空间的目标权重,进一步提升网络精度,实验表明损失了一定的速度,但精度提升到了 96.84%与 YOLOV4 仅仅相差 0.46%,类别标识精度与 YOLOv4 对比相差极小,计算量减低了约 80%满足了应用到移动系统的要求。

表 3 不同模型实验对比表

模型	CA	ASFF	CBAM	警告标志/%	禁止标识/%	指示标志/%	计算量/M	mAP/%	FPS
YOLOv4	—	—	—	99	96	96	64.01	97.30	50.20
A	—	—	—	95	92	88	10.38	91.89	62.06
B	✓	—	—	95	94	92	11.21	93.67	50.88
C	✓	✓	—	97	94	93	11.83	94.65	46.34
D	✓	✓	✓	98	96	95	12.97	96.84	40.21

3 结论

针对目前交通标志检测的网络结构复杂、精度低问

题,首先使用轻量化网络 MobileNetv2 作为主干特征提取网络,为将来方便嵌入一些移动设备减少网络占用内存。通过嵌入坐标注意力机制(CA)来对目标信息进行精确定

位,保留更多通道注意力中的位置信息用于更有效的特征融合和检测;使用自适应特征层来改进颈部网络,更好的融合了不同层之间的特征;然后在特征融合位置加入多个注意力模块(CBAM)来强调目标信息而减少其他物体信息的干扰;最后将 neck 的普通卷积全部替换为 DW 卷积降低参数量,最终形成自适应特征融合的轻量级交通标志检测方法。实验表明,该网络在计算量较少的情况下,提升了精度,检测效果整体性能表现良好。

参考文献

- [1] 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 Preprint, 2016, ArXiv:1602.07360.
- [2] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [3] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, PMLR, 2019: 6105-6114.
- [4] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015, 1440-1448.
- [5] 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, 2016, Las Vegas, 779-788.
- [6] LIU W, ANGUOLOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [7] 张毅, 龚致远, 韦文闻, 等. 基于改进 Faster R-CNN 模型的交通标志检测[J]. 激光与光电子学进展, 2020, (18): 173-181.
- [8] ZHANG J, HUANG M, JIN X, et al. A real-time chinese traffic sign detection algorithm based on modified YOLOv2[J]. Algorithms, 2017, (4): 127.
- [9] 马露茜, 吴钦木. 基于改进 YOLOv3 的复杂环境下交通标志检测[J]. 微处理机, 2022, (1): 39-42.
- [10] 张达为, 刘绪崇, 周维, 等. 基于改进 YOLOv3 的实时交通标志检测方法[J]. 计算机应用, 2022, 42(7): 2219-2226.
- [11] 陈皋, 王卫华, 林丹丹, 等. 基于无预训练卷积神经网络的红外车辆目标检测[J]. 红外技术, 2021, (4): 342-348.
- [12] 刘丹, 汪慧兰, 曾浩文, 等. VoVNet-FCOS 道路行人目标检测算法研究[J]. 国外电子测量技术, 2021, 40(11): 64-71, DOI:10.19652/j.cnki.femt.2103000.
- [13] TIAN Z, SHEN C, CHEN H, et al. fully convolutional one-stage object detection. In 2019 IEEE[C]. Conference: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv:2004.10934.
- [15] GU R, WANG G, SONG T, et al. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 40(2): 699-711.
- [16] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [J]. ArXiv Preprint, 2019, ArXiv:1911.09516.
- [17] 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.

作者简介

梁秀满, 硕士, 副教授, 主要研究方向为检测技术与智能装置、机器视觉等。

E-mail: 2171301551@qq.com

邵彭娟, 硕士研究生, 主要研究方向为机器视觉、目标检测等。

E-mail: 2235946780@qq.com

刘振东, 硕士, 主要研究方向为检测技术与智能装置、机器视觉等。

E-mail: 3464649039@qq.com

赵恒斌, 硕士研究生, 主要研究方向为机器视觉、目标检测等。

E-mail: 3464649039@qq.com