

一种融合注意力机制与上下文信息的交通标志检测方法

王 林, 张 文 卓

(西安理工大学 自动化与信息工程学院, 西安 710048)

摘要: 针对当前交通标志检测中存在小目标检测精度低、检测实时性不高以及目标漏检等问题, 在 YOLOv3 基础上提出了一种融合了注意力机制与上下文信息的交通标志检测方法; 首先通过改进通道注意力机制的压缩方式, 对特征图通道重新进行标定; 然后引入空间金字塔池化模块 SPP; 最后增加特征映射并拼接到原特征融合网络中的小目标部分, 充分利用上下文信息增强对小目标的检测; 实验结果表 TT100K (Tsinghua-Tencent 100K) 交通标志数据集上, 与 YOLOv3 网络相比, 在每秒传输帧数 (FPS, frame per second) 变化不大的情况下, 平均精度均值和小目标的精度均值分别提升 3.03% 和 4.59%; 实验结果证明了改进网络在小目标检测和整体检测中的有效性。

关键词: 小目标检测; YOLOv3; 注意力机制; SPP; 上下文信息

A Traffic Sign Detection Based on Attentional Mechanism and Contextual Information

WANG Lin, ZHANG Wenzhuo

(School of Automation and Information Engineering, Xian University of Technology, Xi'an 710048, China)

Abstract: Aiming at the problems of low accuracy for small targets, low real-time detection and missed target detection in current traffic sign detection, a traffic sign detection method that combines attention mechanism and context information is proposed on the basis of YOLOv3. In this method, firstly, the channel of feature map is recalibrated by improving the compression method of channel attention mechanism; then the spatial pyramid pooling module SPP is introduced; finally, the feature mapping is added and spliced into the small target part of the original feature fusion network, making full use of contextual information to improve the detection of small targets. Compared with the original YOLOv3 network, the experimental results on the TT100K (Tsinghua-Tencent 100K) traffic sign dataset show that the mean average precision and the mean average precision on small target increases respectively by 3.03% and 4.59% with little change in frames per second (FPS). The experimental results demonstrate the effectiveness of the improved network in small target detection and overall detection.

Keywords: small target detection; YOLOv3; Attentional mechanism; SPP; contexts

0 引言

目标检测作为计算机视觉领域的重要课题之一而备受关注。某种意义上来说, 它的发展史也是计算机视觉发展史的缩影。交通标志检测技术是自动驾驶领域的关键技术之一, 被广泛地应用于汽车辅助驾驶系统以及数据地图当中。因此, 交通标志检测具有极高的研究意义和应用价值。

传统的交通标志检测方法以基于颜色^[1]或者基于形状^[2]为代表, 但当面临光照变化、形变以及遮挡等问题时, 检测效果欠佳。卷积神经网络 (CNN, convolutional neural network) 的出现使得目标检测得以迅速发展, 文献 [3] 提出了 R-CNN 模型, 为交通标志检测开辟了新的思路, 随后在此基础上又提出了 Fast R-CNN^[4] 和 Faster R-CNN^[5],

这些方法被称为两阶段目标检测。虽然两阶段目标检测方法检测精度比较高但是实时性却比较差, 针对这一问题文献 [6] 提出 YOLOv1 模型, 这是一个端到端的模型, 因此被称为一阶段目标检测方法, 同时被称为一阶段目标检测的还有 YOLOv2^[7]、YOLOv3^[8]、SSD^[9]、DSSD^[10] 以及 FCOS^[11] 等相关方法。近年来文献 [12] 提出一种基于 CNN 的模型, 该模型采用多尺度滑动窗口策略的同时利用扩张卷积代替原卷积, 取得了不错的检测效果。文献 [13] 按照并行检测原理, 结合目标传感器与神经型卷积分类器, 提出了一种基于卷积神经网络的行人目标检测系统。Zhu^[14] 等人提出一种端到端基于 CNN 的网络, 可同时进行目标的检测和分类任务。

收稿日期: 2021-08-24; 修回日期: 2021-09-24。

基金项目: 陕西省科技计划重点项目 (2017ZDCXL-GY-05-03)。

作者简介: 王 林 (1963-), 男, 江苏东台人, 博士, 教授, 主要从事深度学习、计算机视觉方向的研究。

通讯作者: 张文卓 (1991-), 男, 江苏徐州人, 硕士研究生, 主要从事深度学习、计算机视觉方向的研究。

引用格式: 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.

虽然现有的检测方法在交通标志检测上取得了不错的成绩, 但小目标检测仍然没有达到预期效果。首先, 现有的目标检测网络是以检测大目标为主导, 这是小目标检测效果不佳的客观原因; 其次, 小目标在原图中尺寸较小且数量也比较少, 而现有优秀的检测模型所使用的骨干网络都有下采样层, 通过下采样层后小目标的像素达到了个位数, 导致检测效果不佳。因此本文对检测网络进行改进, 以优化模型在交通标志检测中存在小目标检测精度不高、漏检以及误检等问题。

1 相关理论

YOLO 家族的网络虽然不是精度最好的网络, 但在精确度与速度之间达到了理想的平衡。YOLOv3 借鉴了 YOLOv1 和 YOLOv2, 虽然没有太多的创新点, 但在保持 YOLO 家族速度的优势的同时, 提升了检测的精度, 尤其对于小物体的检测能力, 这对于交通标志检测来说是十分友好的。

1.1 YOLOv3 网络框架

YOLOv3 使用 Darknet-53 作为骨干网络, 相对于 ResNet^[15] 网络而言, 使用卷积层替代池化层进行降采样操作, 这样可以有效地减少浅层特征的丢失。如图 1 所示 Darknet-53 包含 53 个卷积层, DBL (Convolutional) 单元包含了卷积层、批归一化^[16] (BN, batch normalization) 和 Leaky Relu 激活函数, Residual 模块由两个 DBL 单元进行残差操作, 通过引入 BN 层和 Residual 模块, 可以加快网络的收敛, 同时防止随着网络层数的增加而出现的梯度爆炸、消失以及网络退化等问题。

YOLOv3 在网络中借鉴了基于多尺度预测的特征金字塔结构, 用多尺度对不同大小的目标进行检测, 不同的是,

没有采取对位相加的操作而是沿着通道进行拼接操作。首先输入大小为 512×512 的图像, 通过 DarkNet-53 后得到 32 倍下采样的特征图, 经相关卷积操作得到尺寸大小为 16×16 的特征图 Y1; 其次将 32 倍下采样获得的特征图, 经卷积操作后进行两倍上采样, 与从 DarkNet-53 中获取输出尺寸大小相同的特征映射沿通道进行拼接, 经相关卷积操作后得大小为 32×32 的特征图 Y2; 最后同理 Y2, 获得大小为 64×64 的特征图 Y3。YOLOv3 共输出 3 个不同尺寸的特征图 Y1、Y2 以及 Y3, 其中 Y1 进行大目标预测, Y2 进行中目标预测, Y3 进行小目标预测。3 个特征图分别的通道数为所用数据集类别数 (num_classes) 加 5 之后与锚框数量 (num_anchor) 的乘积, 5 为置信度、大小和尺寸 5 个信息。

1.2 边界框

锚框 (Anchor box) 是边界框的先验, 是基于数据集通过聚类预测得到。先验锚框的宽和高为 (P_w, P_h) , 中心点为对应网络的左上角 (c_x, c_y) , 通过 YOLOv3 模型预测得出 (t_x, t_y, t_w, t_h) 和置信度 5 个值, t_x, t_y, t_w, t_h 分别为预测框的中心点坐标和尺寸, 通过以上数据进行微调, 计算公式如下:

$$b_x = \delta(t_x) + c_x \tag{1}$$

$$b_y = \delta(t_y) + c_y \tag{2}$$

$$b_w = p_w e^{t_w} \tag{3}$$

$$b_h = p_h e^{t_h} \tag{4}$$

其中: b_x, b_y, b_w, b_h 为调整后的锚框中心坐标和尺度大小, δ 为 sigmoid 函数, 使用 sigmoid 函数可以有效解决预测框可能会出现在图上任意位置这一现象, 从而限制坐标位置, 使得网络更加容易学习。

1.3 损失函数

YOLOv3 的损失函数由置信度损失、分类损失和定位

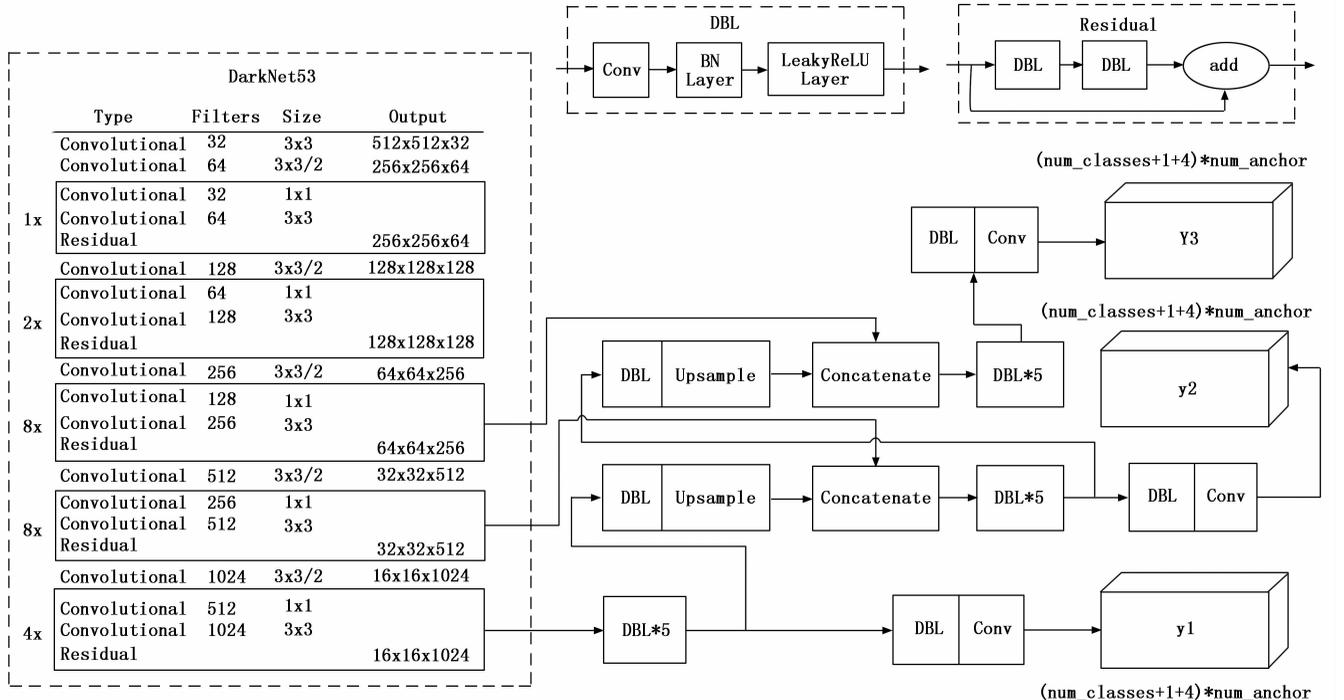


图 1 YOLOv3 网络结构

损失这三部分组成，计算公式如下：

$$L(o, c, O, C, l, g) = \lambda_1 L_{\text{conf}}(o, c) + \lambda_2 L_{\text{cla}}(O, C) + \lambda_3 L_{\text{loc}}(l, g) \quad (5)$$

$$L_{\text{conf}}(o, c) = - \frac{\sum_{i \in N} (o_i * \log_e^{\hat{c}_i} + (1 - o_i) * \log_e^{(1 - \hat{c}_i)})}{N} \quad (6)$$

$$l(O, C) = \sum_{i \in \text{pos}} \sum_{j \in \text{cla}} (O_{ij} * \log_e^{\hat{c}_{ij}} + (1 - O_{ij}) * \log_e^{(1 - \hat{c}_{ij})}) \quad (7)$$

$$L_{\text{cla}}(O, C) = - \frac{l(O, C)}{N_{\text{pos}}} \quad (8)$$

$$L_{\text{loc}}(l, g) = \frac{\sum_{i \in \text{pos}} \sum_{m \in \{x, y, w, h\}} (\hat{l}_i^m - \hat{g}_i^m)^2}{N_{\text{pos}}} \quad (9)$$

其中： $L(o, c, O, C, l, g)$ 、 $L_{\text{conf}}(o, c)$ 、 $L_{\text{cla}}(O, C)$ 、 $L_{\text{loc}}(l, g)$ 分别为总损失、置信度、类别以及定位损失， $l(O, C)$ 为 $L_{\text{cla}}(O, C)$ 的中间变量， $\lambda_1, \lambda_2, \lambda_3$ 为平衡系数； $o_i \in [0, 1]$ 表示预测目标边界框与真实目标边界框的交并比（IOU, intersection over union）， c 为预测值， \hat{c}_i 为 c 经过 sigmoid 后的置信度预测值， N 为正负样本数量的总和； $O_{ij} \in \{0, 1\}$ 表示在第 i 个预测目标的边界框中是否存在第 j 类目标， C_{ij} 为预测值， \hat{C}_{ij} 为 C_{ij} 通过 sigmoid 后的目标概率， N_{pos} 为正样本个数； \hat{l}_i^x, \hat{l}_i^y 为 t_x, t_y 通过 sigmoid 得到的偏移量， \hat{l}_i^w, \hat{l}_i^h 和 t_w, t_h 数值相等， \hat{g}_i^x, \hat{g}_i^y 分别为真值的坐标减去对应的 c_x, c_y 的坐标偏移量， \hat{g}_i^w, \hat{g}_i^h 分别为真值的宽和高除以对应的 p_w, p_h 后取对数的偏移量。

2 改进的交通标志检测网络

YOLOv3 是目前应用最广泛的目标检测算法之一，具有较好的检测速度和精度，但直接应用到交通标志检测任务当中还存在一些不足。

2.1 改进的通道注意力机制 ECA-A 模块

SE-Net (Squeeze-and-Excitation Networks) 通道注意力机制，因其复杂度低、新增参数较少以及计算量小，被广泛应用在各个模型当中^[17]。核心思想是增强有用信息的通道，抑制信息较少的通道。SE 模块通过两个全连接层降维操作来减少模型复杂度。但是，降维会对通道的预测产生副作用，得到的通道间的相关性也是低效且不必要的。同时，大量使用 SE 模块也会很大程度上影响网络的实时性。

2020 年，(ECA-Net, efficient channel attention networks) 提出一种无降维的局部跨通道策略，通过考虑每个信道及其 k 个相邻信道捕获局部跨信道交互，在保证实时性的同时也保证了准确性^[18]。但对于大目标物体来说，取特征通道的全局平均值能够很好的反应特征通道的响应情况。不过对于交通标志检测来说，交通标志一般只占图片很小的区域，在剩下的区域当中会存在许多无关的背景信息，这些信息虽然单个响应不大，但是总的信息响应还是不能忽略。因此，对通道进行全局平均，并不能充分代表通道的信息响应，也应当考虑通道的极值响应。本文针对上述情况对 ECA-net 进行的改进如图 2 所示。

其中： k 是通过自适应函数来确定的。如公式 (10) 所示，其中 C 为通道数， b, γ 为调节参数， $|f(x)|_{\text{odd}}$ 为获取值

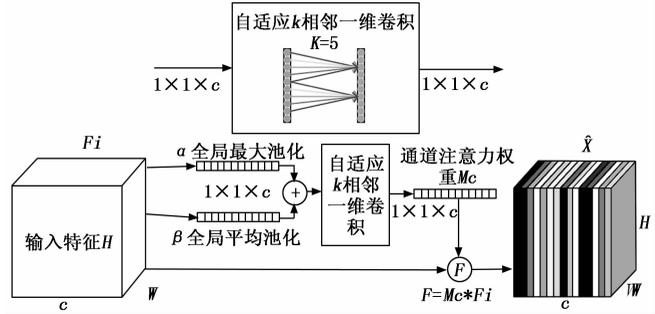


图 2 ECA-A 模块结构

的最近奇数。经过相关卷积操作后得到的 F_i 作为待标定的特征，首先将大小为 $H \times W \times C$ 的输入特征 F_i 在通道上分别进行全局最大池化 (GMaxPool, global max pooling) 和全局平均池化 (GAvgPool, global average pooling)，将两个结果按照一定比例 α 和 β 沿通道对应相加，得到 $1 \times 1 \times C$ 压缩特征，其中 α 和 β 为调节参数且和为 1。然后将压缩特征通过大小为 k 的 1 维卷积，每个信道及其 $k - 1$ 个邻居共同获取跨信道交互信息，即共有 k 个信道参与预测，核大小 k 代表局部跨信道交互的覆盖范围，最后使用 sigmoid 将得到的结果转换成 $0 \sim 1$ 之间的数值，作为原通道上的系数，与 F_i 按通道对应相乘，完成重新标定，得到和 F_i 相同尺寸与维度的 \hat{X} 。

$$k = \left\lfloor \left| \frac{\log_e^C}{\gamma} + \frac{b}{\gamma} \right|_{\text{odd}} \right\rfloor \quad (10)$$

自适应 k 相邻一维卷积的使用即避免降维带来的低效性，也有效地捕捉跨信道交互信息。与 SE 网络相比参数也大大减少，保证了网络的实时性。将 ECA-A 模块加入到原 DarkNet-53 中的残差单元结构当中，更改后的残差模块如图 3 所示。

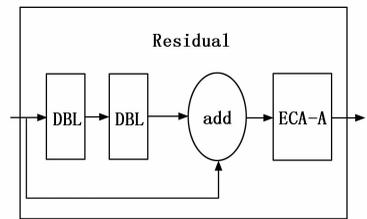


图 3 改后的残差单元结构

2.2 空间金字塔

空间金字塔池化 (SPP, spatial pyramid pooling)^[19] 是何凯明等人提出的一种多尺度特征融合模型。传统的神经网络只能接受以固定的尺寸作为输入，所以经常会对输入图片进行裁剪以及缩放等操作，以此来满足输入的要求，但这些操作会导致信息的丢失以及图像的失真，从而导致检测效果不佳。如图 4 所示，本文引入 SPP 模块以 YOLOv3 框架为基础，在 DarkNet-53 后面的第三至第四个 DBL 单元之间加入 SPP 模块，SPP 模块是由 4 个分支构成，分别由步距为 1 尺寸大小为 $5 \times 5, 9 \times 9, 13 \times 13$ 的最大池化操作和一个跳跃连接，随后将 4 个分支进行拼

接。这样可以获取多尺度局部特征信息, 将其融入到后续的全局特征当中, 从而得到更加丰富的特征表示, 最终提高检测精度。

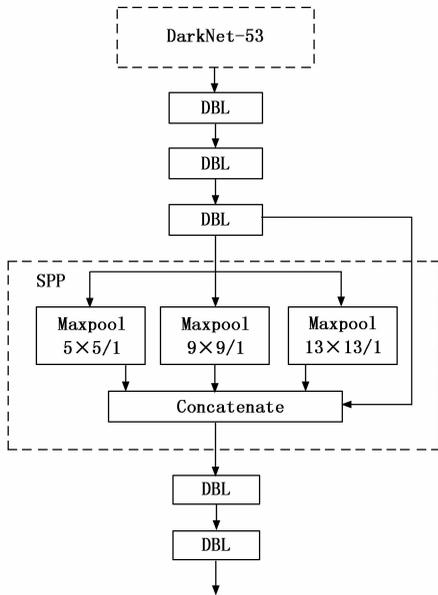


图 4 SPP 模块结构

2.3 增强上下文信息

对于目标检测而言, 待检测目标不可能单独存在, 它一定会和周围的其他目标或多或少存在某种关系, 这就是通常所说的上下文信息。如何挖掘它们之间的关系, 利用这个关系来增强特征表示是上下文信息的核心问题。

对于交通标志检测而言, 大部分的交通标志都悬挂在交通标志杆上, 交通标志杆就是交通标志的上下文信息。因此, 学习交通标志杆与交通标志之间的潜在关系是十分重要的。YOLOv3 原有的 Y3 特征图包含了两个拼接, 但都是深层特征, 缺乏浅层特征表示。为此本文在原来的基础

上, 从 DarkNet-53 中获取输出尺寸大小为 128×128 的特征映射, 经过大小为 3×3 步距为 2 的最大池化操作后得到 64×64 的特征图, 将得到的 3 个尺寸相等的特征图进行拼接, 经相关卷积操作后得到新的 Y-3 特征图。选择最大池化操作是因为交通标志杆和周围的颜色区别很大, 最大池化后能选取到更多信息。整体更改后的网络结构如图 5 所示。相当于拼接处共拼接 3 个特征图, 包含了最深层的特征信息、中层特征信息以及相对浅层的信息, 这样既能获得更精细度的信息, 也能获得更加丰富的上下文信息。

3 实验结果与分析

本文实验所使用的环境配置为 Windows 10、CUDA (Compute Unified Device Architecture) 10.0、CUDNN (CUDA Deep Neural Network library) 7.0。硬件配置为 Intel-CPU-i5-10400F 处理器, 显卡为 Nvidia Ge Force GTX 1070Ti, 调用 GPU 进行加速训练。

3.1 数据集与 Anchor 的重新设定

本文采用 TT100K 数据集^[20], 该数据集是由清华和腾讯联合打造。包含了中国各大城市的实景街拍交通标志图像。其中, 交通标志共有 150 多种, 本文选取实例数大于 50 的 48 个类别作为使用的类别, 经删减和数据增强后, 共得到 15 800 张图片, 训练使用 10 400 张图片, 测试使用 5 400 张图片。

Anchor 值的设定对最终的检测性能至关重要, 针对不同的数据集应当设定不同的 Anchor 值。合适的 Anchor 值能加快网络的收敛速度、降低误差。通过对 TT00K 数据集尺寸进行统计, 发现 55% 左右的目标像素小于 $45 \text{ 像素} \times 45 \text{ 像素}$, 因此存在大量的目标相对尺寸较小的情况。使用 K-means 算法对参数进行重新选择, 更改后的参数对比如表 1 所示。从表中可以看出修改后的 Anchor 大小比原来缩小了很多, 这也符合 TT100K 数据集中的目标相对尺寸比较小这一情况。

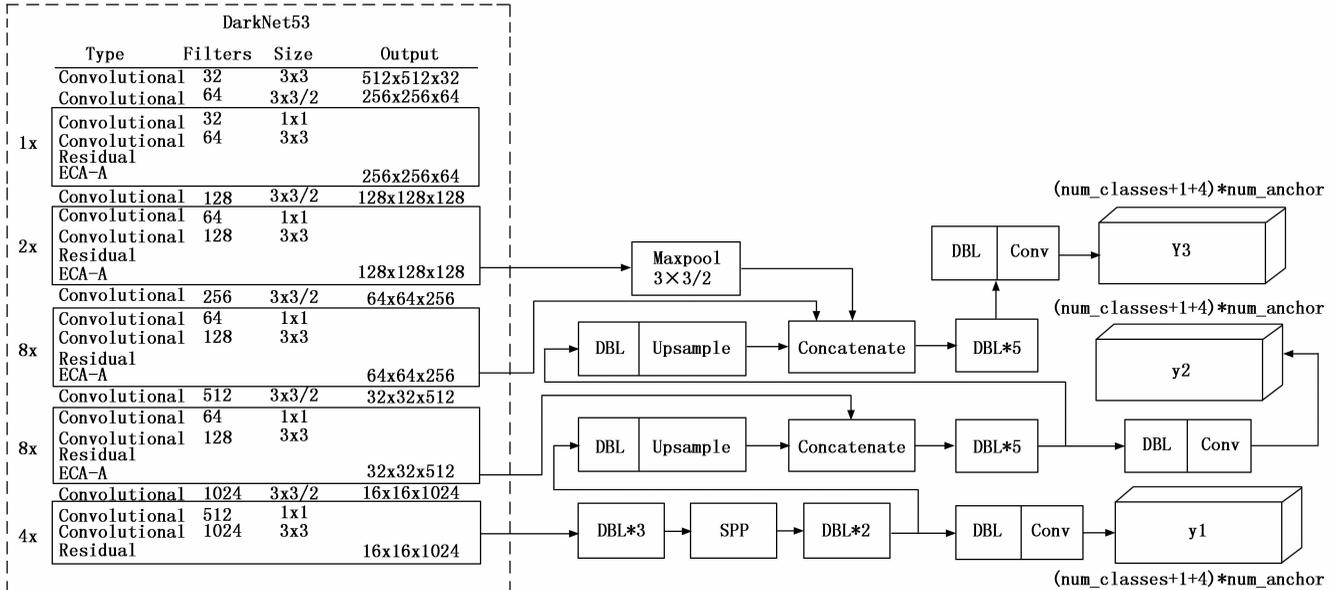


图 5 更改后的网络结构

表 1 Anchor 值重定

特征层	原数据集 Anchor 值	修正后的 Anchor 值
16×16	(116,90) (156,198) (373,326)	(24,28) (30,35) (45,76)
32×32	(30,61) (62,45) (59,119)	(13,13) (16,18) (25,31)
64×64	(10,13) (16,30) (33,23)	(7,9) (9,14) (13,20)

3.2 评价指标与参数设置

实验指标采用平均精度均值 (MAP, mean average precision)、小目标精度均值 (AP-s, small object of average precision)、中型目标精度均值 (Medium Object of Average Precision, AP-m)、大目标精度均值 (AP-l, large object of average precision) 以及画面每秒传输帧数 (FPS, frame per second)。其中, MAP 的数值越高表明模型的性能越好, FPS 的数值越大表明模型的实时性就越好。在训练阶段, 本文以 YOLOv3 作为基础框架, 每种实验都进行 100 个 epochs, 初始学习率为 0.001, 动量和衰减分别设置为 0.9 和 0.000 5, 批处理数 BatchSize 设置为 16。

3.3 结果与分析

本文首先基于 YOLOv3 对所提出的改进通道注意力机制压缩方式进行试验, 找出最佳的 α 和 β 参数 ($\alpha + \beta = 1$), 实验结果如图 6 所示。结果表明在 TT100K 数据集中取 $\alpha = 0.6$ 和 $\beta = 0.4$ 时模型的 MAP 达到最佳。后续实验的 ECA-A 模块都取 $\alpha = 0.6$ 和 $\beta = 0.4$ 。

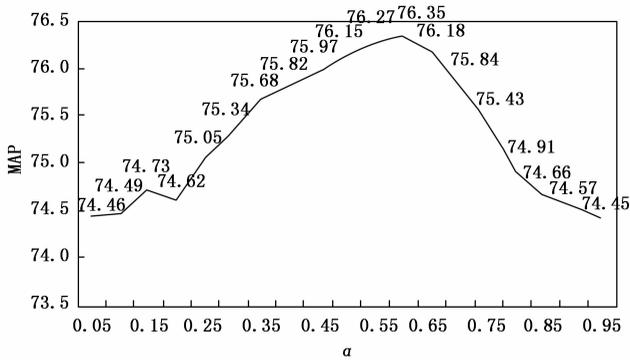


图 6 参数选择

随后对于特征压缩方法的选择, 在相同的实验环境下, 本文所采用的方法与 GMaxPool、GAvgPool 以及 GMaxPool 和 GAvgPool 沿通道维度拼接这 3 个压缩方法进行对比, 得到的 MAP 结果对比结果如表 2 所示。表明取 $\alpha = 0.6$ 和 $\beta = 0.4$ 时的 GMaxPool 与 GAvgPool 沿通道相加的压缩方法性能最佳。

表 2 各压缩方法对比结果

特征层	MAP%
GAvgPool	74.42
GMaxPool	74.45
GAvgPool 与 GMaxPool 拼接	75.68
GAvgPool 与 GMaxPool 相加	76.35

标尺寸更小且图像分辨率更高, 以 YOLOv3 为基础架构使用消融实验来验证改进与添加模块的有效性。评价指标分别采用 IOU 取值为 0.5 时的 AP-s、AP-m、AP-l 以及 MAP, 实验后的结果如表 3 所示。实验结果表明, 单个模块中 ECA-A 对整体检测准确度的影响最大 MAP 提升了 2.2%; SPP 对整体都有提升; 新的 Y3 特征图对小目标提升最大, 提升了 2.61%。YOLOv3 在将 3 个改动全部添加后达到了最佳检测性能, MAP 提升 3.03%、AP-s 提升 4.59%、AP-m 提升 2.66% 以及 AP-l 提升 2.26%, 可以看出整个改进后小目标检测的精度要优于单个改进后的精度。而后续的实验则使用改进后的网络。

表 3 消融实验数据

实验	ECA-A	SPP	Y-3	AP-s%	AP-m%	AP-l%	MAP%
A	—	—	—	35.76	75.29	84.32	74.15
B	√	—	—	37.21	77.54	85.14	76.35
C	—	√	—	36.31	76.82	86.69	76.21
D	—	—	√	38.37	75.25	84.26	74.84
E	√	√	√	40.35	77.95	86.58	77.18

将改进后的网络与其他主流方法进行实验比对, 包括 Faster R-CNN、SSD 以及 YOLOv3。取 IOU 值为 0.5 和 0.75 时的 MAP 和 FPS 作为评价指标。其中 IOU 为 0.5 时和 IOU 为 0.75 时的 MAP 分别为对检测能力和定位精度的有效评估指标。

从表 4 可以看出, 当 IOU 取值为 0.5 时, 改进后的 YOLOv3 的 MAP 值比第二的 Faster R-CNN 高出 0.64%; 当 IOU 取值为 0.75 时, 改进后的 YOLOv3 的 MAP 值达到最佳, 比第二的 Faster R-CNN 高出 1.45%; 在 FPS 指标上, 改进后的 YOLOv3 只比最佳的 YOLOv3 每秒差 5 张图片, 但高出 YOLOv3-SE 每秒 11 张图片。可以看出改进后的网络在模型性能、定位精度以及实时性方面都表现很好。

表 4 4 种检测网络对比

算法	MAP-IOU=0.5%	MAP-IOU=0.75%	FPS
Faster R-CNN	76.54	66.37	11
SSD	63.67	52.09	20
YOLOv3	74.15	63.59	38
YOLOv3-SE	76.23	64.12	22
改进后的网络	77.18	67.82	33

最后, 为了进一步验证改进后的网络的模型性能, 本文从数据集中选出三张图片进行测试, 结果如图 7 所示, 其中 (a1)、(a2)、(a3) 为 YOLOv3 网络的检测结果, (b1)、(b2)、(b3) 为改进后网络的检测结果。

首先对比 (a1) 和 (b1) 的结果可以看出, 两个都将目标完全地检测出来, (b1) 在形变的交通标志上的预测框覆盖区域要优于 (a1) 的预测框; 其次在 (a2) 与 (b2) 的结果上来看, (b2) 检测出来而 (a2) 出现了漏检, 这是由于这张图像上的目标与周边的环境相似, 而改进后的网络丰富了上下文信息利用了交通标志杆这个信息, 从而避免

TT100K 数据集相比于其他交通标志数据集来说, 目

了这次漏检; 最后对比了 (a3) 和 (b3), (a3) 出现了两个漏检 (b3) 出现了一个漏检, 第三张图片的 4 个标志全都是像素小于 10×10 的, 图中右边的两个目标边上的背景非常复杂, 这也是 YOLOv3 都漏检的原因, 而改进后的网络检测没有检测出“禁止鸣笛”的交通标志, 原因可能是目标太小以及和周边背景颜色相近。综合来看, 改进后的网络在小目标检测以及漏检方面要优于原 YOLOv3。

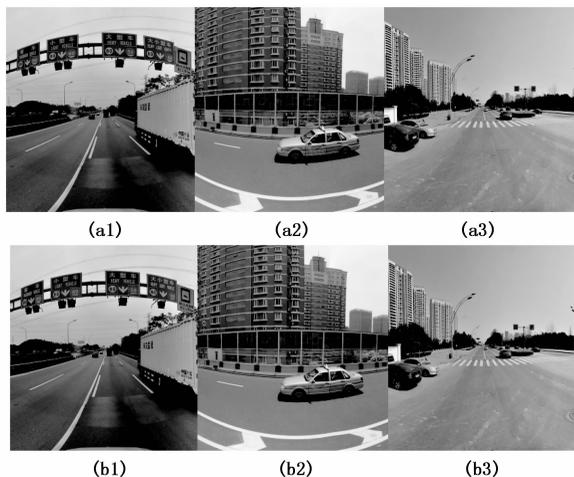


图7 实验测试结果

4 结束语

本文主要介绍了改进后的交通标志网络, 针对在目前检测算法中存在小目标检测精度不高、漏检及误检等问题, 改进了 ECA 的压缩方式, 将改进后的 ECA-A 加入到骨干网络 DarkNet-53 当中的残差模块, 对通道进行重新标定, 增强重要的通道抑制信息量较少的通道, 显著地提高算法的检测性能; 引入金字塔池化模块, 以增强最终检测精度; 重新拼接后的特征图包含了最深层的特征信息、中层特征信息以及相对浅层的信息, 这样既能获得更细粒度的信息, 也能获得更加丰富的上下文信息。实验表明改进后的网络在小目标检测、整体性能、定位精度以及实时性上的有效性。

参考文献:

- [1] DE IA ESCALERA A, MORENO L E, SALICHS, M A, et al. Road traffic sign detection and classification [J]. IEEE Transactions on Industrial Electronics, 1997, 44 (6): 848 - 859.
- [2] GONZALEZ A, GARRIDO M A, LLORCA D F, et al. Automatic Traffic Signs and Panels Inspection System Using Computer Vision [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12 (2): 485 - 499.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580 - 587.
- [4] GIRSHICK R. Fast R-CNN [C] //Proceedings of 2015 IEEE International Conference on Computer Vision and Pattern Recognition. IEEE Press, 2015: 1440 - 1448.
- [5] 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.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, 2016: 779 - 788.
- [7] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C] //Proceedings of 2017 Conference on Computer Vision and Pattern Recognition. Nevada: IEEE, 2017: 6517 - 6525.
- [8] REDMON J, FARHADI A. YOLOV3: an incremental improvement [C] //Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2018: 1 - 6.
- [9] LIU W, ANGELOV D, ERHAN D, et al. SSD: single shot multi-box detector [C] //Proceedings of 2016 European Conference on Computer Vision, Berlin, Germany: Springer, 2016: 21 - 37.
- [10] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional Single Shot Detector [Z]. 2017.
- [11] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection [C] //Proceedings of the Conference on Computer Vision, 2019: 9627 - 9636.
- [12] AGHDAM H H, HERAVI E J, PUIG D. A practical approach for detection and classification of traffic signs using Convolutional Neural Networks [J]. Robotics and Autonomous Systems, 2016, 84: 97 - 112.
- [13] 王 林, 刘 盼. 基于卷积神经网络的行人目标检测系统设计 [J]. 计算机测量与控制, 2020, 28 (7): 64 - 68.
- [14] LIU Z, SHEN C, FAN X, et al. Scale-aware limited deformable convolutional neural networks for traffic sign detection and classification [J]. IET Intelligent Transport Systems, 2020, 14 (12): 1712 - 1722.
- [15] HE K, ZHANG X, REN S, et al. Residual Learning for Image Recognition [C] // IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [16] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C] //International Conference on Machine Learning, 2015.
- [17] JIE H, LI S, GANG S, et al. Squeeze-and-Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42 (8): 2011 - 2023.
- [18] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [19] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (9): 1904 - 1916.
- [20] ZHU Z, LIANG D, ZHANG S, et al. Traffic-Sign Detection and Classification in the Wild [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.