

DOI:10.19651/j.cnki.emt.2210870

# 基于改进 SSD 的交通标志检测算法

赵友章 吕进

(长安大学信息工程学院 西安 710000)

**摘要:** 为了解决真实交通场景下交通标志因目标较小而导致检测精度低的问题,提出了一种改进 SSD 的交通标志检测算法。首先使用更深层次的 ResNest 网络替换原始 SSD 算法的主干网络 VGG16 来增强弱目标特征的强表征能力,然后在 SSD 的额外添加层使用 RFB 模块来增加小目标的感受野。其次使用 Bi-FPN 加权双向特征金字塔网络有效结合深层与浅层的特征信息,改善小目标的检测性能。最后使用 K-means++ 聚类算法调整默认窗口的大小,有效避免因原始默认窗口太大但交通标志较小而无法匹配的问题,以改善检测效率。实验结果表明,本文提出的模型在中国交通标志数据集(CCTSDb)上获得了 95.33% 的 mAP,与原始 SSD 模型相比,本文所构建的模型能更好的适应自然背景下的交通标志检测。

**关键词:** 交通标志检测;SSD;ResNest;K-means++;RFB 模块;加权特征融合

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.6040

## Traffic sign detection algorithm based on improved SSD

Zhao Youzhang Lyu Jin

(School of Information Engineering, Chang'an University, Xi'an 710000, China)

**Abstract:** In order to solve the problem of low detection accuracy of traffic signs due to small targets in real traffic scenes, a traffic sign detection algorithm based on improved SSD was proposed. First, a deeper ResNest network was used to replace VGG16, the backbone network of the original SSD algorithm, to enhance the strong characterization of weak target features. Then, the RFB module was used in the additional layer of SSD to increase the receptive field of small targets. Secondly, bi-FPN weighted bidirectional feature pyramid network is used to effectively combine deep and shallow feature information to improve the detection performance of small targets. Finally, K-means++ clustering algorithm was used to adjust the size of the default window, which effectively avoided the problem that the original default window was too large but the traffic signs were small and could not be matched, so as to improve the detection efficiency. Experimental results show that the proposed model achieves 95.33% mAP on the China Traffic Signs Data Set (CCTSDb). Compared with the original SSD model, the proposed model can better adapt to traffic signs detection under natural background.

**Keywords:** traffic sign detection;SSD;ResNest;K-means++;RFB module;weighted feature fusion

## 0 引言

随着人工智能的飞速发展与不断完善,越来越多的人工智能技术被应用于智能交通,作为智能交通领域主要组成部分的交通标志检测在面对复杂交通情况时能做出准确且及时的反应,为人们的安全出行提供了重要保障。但目前在国内真实复杂的交通环境下,交通标志往往都是以小目标的形式存在,同时会受到天气、光线变化、物体遮挡等特殊因素的影响,导致驾驶人员与智能辅助驾驶系统对前方路况判断不清晰,从而引发交通事故。

在真实交通场景下的交通标志牌都具有明显的颜色和形状特征,故传统的目标检测方法通常利用相应的图像处理技术来提取颜色或形状特征,并对其进行分类检测。Gudigar 等<sup>[1]</sup>通过计算被检测图片全局亮度的均值,比较不同光照条件的阈值与均值来判断当前的光照环境,然后对不同的光照环境分别进行红色角度空间转换,最后根据一系列不同的阈值对图像进行分割,从而确定不同种类标志在图像中的区域。但当交通标志牌褪色或者处于曝光状态下时,基于颜色特征的交通标志检测效率会受到严重影响。Boumediene 等<sup>[2]</sup>使用灰度图像检测三角交通标志,通过交

通标志的编码梯度得到相应的角点,其次在得到的角点上检测出对称线,该方法很好的检测出了三角交通标志。虽然基于形状特征的交通标志检测有效避免了光照与褪色的影响,但当交通标志牌受损、被遮挡或者检测视角发生变化时,检测效果还是会受到影响,无法适应自然场景下的交通标志检测。近年来,随着深度学习进入大众的视野,越来越多的研究者将深度学习应用于目标检测。相较于传统算法,深度学习算法能够有效避免在图像特征提取时出现的语义信息不充分的问题,从而改进对目标的检测效果。基于深度学习的目标检测算法有两类,其中一类是以 R-CNN<sup>[3]</sup>,Fast R-CNN<sup>[4]</sup>,Faster R-CNN<sup>[5]</sup>为代表的 two-stage 算法,首先通过算法过滤掉大量无用的背景信息,生成一系列交通标志的候选区域,再通过卷积神经网络对其进行回归和分类。因为其在分类时只需要训练生成的候选区域,不受背景环境的影响,所以其检测精度较高,但因为其前期通过大量的计算去过滤背景信息,所以其检测速度相对较慢,不能很好的满足实时检测的需求。另一类是以单激发多框探测器 (single shot multibox detector, SSD)<sup>[6-8]</sup>,YOLO(you only look once, YOLO)<sup>[9-10]</sup>为代表的 one-stage 算法,one-stage 省略了生成候选区域的步骤,直接通过卷积神经网络计算交通标志目标的类别概率以及在图像中的位置坐标。YOLO 算法因为没有候选区域的限制,所以在检测速度方面有不错的提升,但对于物体长宽比的泛化能力较弱,当一类物体出现新的长宽比时,检测的准确率降低。SSD 算法结合了 YOLO 与 Faster R-CNN 的优点,并且加入了多尺度特征的思想,在检测精度和实时性方面都有不错的提升。但是在真实复杂的交通背景下,交通标志多以小目标的形式出现,而 SSD 算法在小目标的检测效果上还存在一些不足,故本文在 SSD 算法的基础上做出如下改进:

1)采用更深层次的注意力拆分网络(ResNest)<sup>[11]</sup>来代替原始的 VGG(visual geometry group, VGG)<sup>[12]</sup>网络,ResNest 网络使用分组卷积以及通道注意力机制,将通道进行分组,对每组运用注意力机制,同时保留了残差结构,增强了弱目标特征的强表征能力。然后将原始 SSD 模型的额外添加层的前两层更换为通过增加小目标感受野来增强特征提取能力的模块(receptive field block network,

RFB)<sup>[13]</sup>。

2)采用加权双向特征金字塔网络(Bi-directional feature pyramid network, Bi-FPN)<sup>[14]</sup>,其使用了有效的跨尺度连接和加权特征融合,能够有效结合深层和浅层的特征信息,改善交通标志小目标的检测性能。

3)采用 K 均值聚类算法(K-means++)取代默认窗口的随机生成机制,将几个聚类中心点的窗口大小分别作为各个特征层的窗口的基准大小。有效避免因原始默认窗口太大但交通标志较小而无法匹配的问题,从而改善检测效率。

### 1 原始 SSD 算法

SSD 算法的全称为单激发多框探测器,其实现了端到端的目标检测,并引入了锚框(anchor)的概念。作为 one-stage 目标检测算法的一种,省去了 two-stage 算法的区域搜索阶段,可以直接产生物体的类别概率以及位置坐标值,并且采用了多尺度特征预测的思想,在检测精度和速度方面都有明显的优势。

SSD 模型的整体结构如图 1 所示,其以 VGG16 网络作为主干特征提取网络,移除其最后一层全连接层 FC8,将 FC6 与 FC7 更换为卷积层 Conv6 与 Conv7,并在 Conv7 后添加多个额外的卷积块 Conv8、Conv9、Conv10、Conv11。SSD 算法主要通过在网络结构中进行卷积、池化和下采样操作,使每个用于检测的特征层都获得不同通道数的特征信息,在将这些不同尺度的特征信息送入分类和回归网络用于模型预测。原论文中使用 Conv4\_3、Conv7 检测小目标,使用更深层的 Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2 检测中等目标和大目标,尽可能保证将图像中的所有目标都检测出来。本文沿用原论文中的方法进行对比和消融实验。SSD 在不同尺度的特征图上设置多种宽高比的 anchor,即通过在不同尺度特征图的每个单元上,生成不同宽高比的 anchor,并以此来获得训练过程中目标物体感受野的大小,从而降低训练的难度。最后,将从图像上得到的所有 anchor 的定位坐标信息和识别目标概率信息进行分类以及回归预测,再通过非极大值抑制计算去除不需要的 anchor,进而确定最终测量目标的类别置信度及其定位坐标,从而完成了多尺度特征检测。

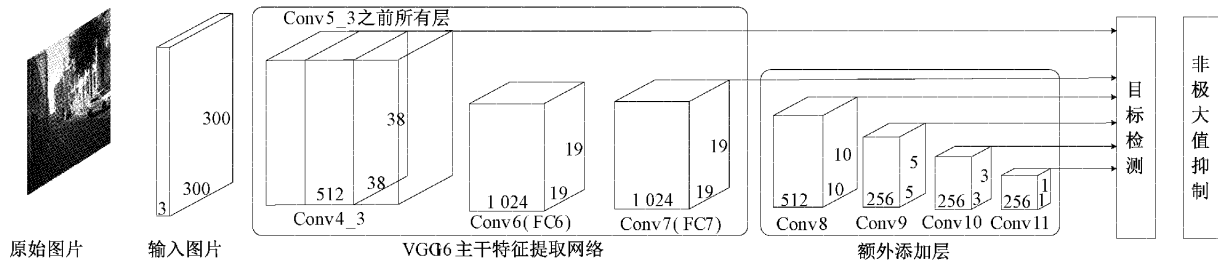


图 1 SSD 网络结构

## 2 改进 SSD 算法

### 2.1 改进特征提取网络

#### 1) 主干特征提取网络

近年来,深度学习越来越受到广大学者的重视,尤其是在智能交通领域,因为它几乎可以拟合各种类型的目标函数,并达到较好的拟合效果,而深度学习模型基本都是在进行非线性变换。更深的网络就意味着网络可以学习到更加复杂的变换,拟合更复杂的特征输入。但是原始 SSD 模型在 VGG16 的基础上通过增加网络层数改善特征提取能力时会受到梯度不稳定因素的影响,故本文使用在残差网络(residual network, ResNet)<sup>[13]</sup>基础上改进的 ResNest 网络。ResNet 引入了跳跃连接,可将任意浅层的输入特征传递到深层,避免了由于网络过深引起的梯度消失问题和退化问题。但 ResNet 缺少跨通道间的交互作用,不能很好的进行特定的图像分类任务。ResNest 中引入 Split-Attention 模块,使用分组卷积以及通道注意力机制,将通道进行分组,对每组运用注意力机制,同时保留了残差结构,增强了弱目标特征的强表征能力。其中 ResNest block 的结构如图 2 所示。

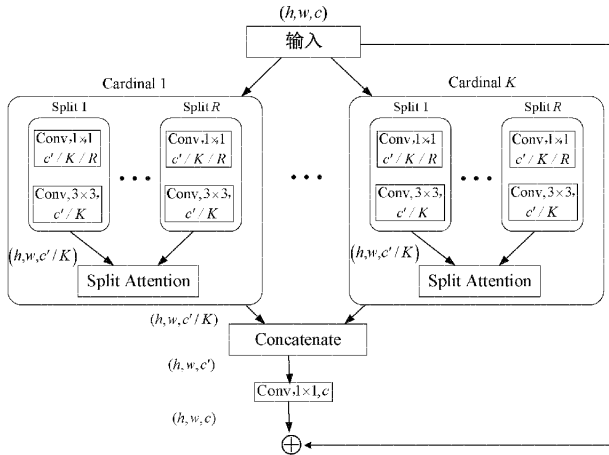


图 2 ResNest block 结构

ResNest block 由特征映射模块以及分散注意力模块组成,将尺度为  $(h, w, c)$  的输入特征图沿着通道维度分成  $K \times R$  个特征图小组,其中  $K$  和  $R$  分别表示特征图组的个数和在每个特征图组内分组的数量,  $c'$  为中间卷积层的特征图通道数。在每一个特征图组中,分别对  $R$  个小组进行  $1 \times 1 \text{Conv} + 3 \times 3 \text{Conv}$ , 并将该特征组内所有小组的输出结果都输入到分散注意力模块 Split Attention 中处理,最后通过残差结构将  $K$  个特征图组的输出经过 Concatenate 和  $1 \times 1$  的卷积后与输入特征相加得到 ResNest block 的输出。

Split Attention 模块作为 ResNest block 的核心模块,借鉴了 SKNet(Selective Kernel Network, SKNet)<sup>[16]</sup>的思想,其基本结构如图 3 所示。

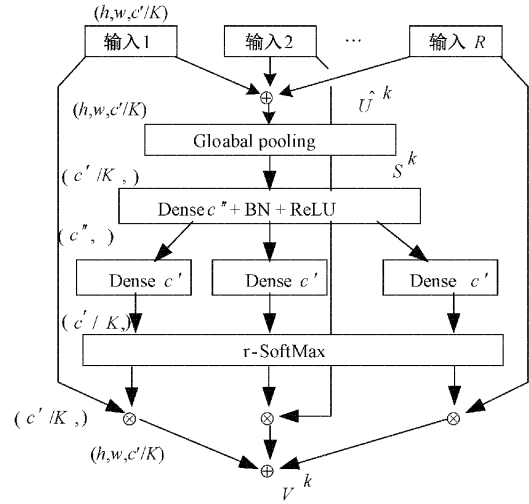


图 3 Split Attention 模块图

该模块为整个 ResNest 结构添加了通道注意力机制,其中  $c''$  为中间卷积层的特征图通道数。首先将生成的  $R$  个特征图小组通过式(1)进行求和得到第  $k$  个分支的输出特征:

$$\hat{U}^k = \sum_{j=R(k-1)+1}^{Rk} U_j (j \in 1, 2, \dots, R; k \in 1, 2, \dots, K) \quad (1)$$

其中,  $R$  表示每个基数组内的分组数量,  $U_j$  表示 Split Attention 模块的第  $j$  个输入,  $k$  表示第  $k$  个 Split Attention 模块。

将融合后的特征图经过全局平均池化得到基于通道的全局上下文信息,如式(2)所示。  $H$  和  $W$  分别表示特征图的高和宽;  $U_c$  对应式(1)中得到的结果。

$$S_c^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c^k(i, j) \quad (2)$$

之后将上述所得结果  $S^k$  经过全连接层以及 BN + ReLU 的计算,并通过分类器(Softmax)计算出每个子组的权重,如式(3)所示。其中  $G_i^c$  是由两个全连接层加 ReLU 激活函数构成的注意力权重函数。

$$a_i^k(c) = \begin{cases} \frac{\exp(G_i^c(s^k))}{\sum_{j=0}^{Rk} \exp(G_j^c(s^k))}, & R > 1 \\ \frac{1}{1 + \exp(-G_i^c(s^k))}, & R = 1 \end{cases} \quad (3)$$

最后将 Split Attention 的输入分别与  $a_i^k(c)$  相乘,为每个子组赋予权重,并将各个子组的特征进行相加,得到该基数组的加权融合特征  $V^k \in R^{H \times W \times C}$ ,其中,  $V^k$  的具体计算如式(4)所示,  $U_{R(k-1)+i}$  表示第  $(k-1)+i$  个子组的特征图。

$$V_c^k = \sum_{i=1}^R a_i^k(c) U_{R(k-1)+i} \quad (4)$$

#### 2) 额外添加层

在自然背景下的交通标志多以小目标的形式出现,本文除使用 ResNest 网络替代主干特征提取网络 VGG 以增强弱目标特征的强表征能力外,还使用 RFB-s 模块有效提升浅层目标特征的感受野,使浅层能够获得更为有效的信息,同时使用 RFB 模块替换原始 SSD 模型的额外添加层的前两层,1x1Conv+3x3Conv 的卷积块替代最后三层,使得改进后的网络模型在没有增加参数数量的同时大幅提升了性能。

RFB 模块的整体结构借鉴了 Inception<sup>[17]</sup> 网络的多分支思想,通过残差结构与不同尺度的多分支卷积块模拟多尺度的感受野。与 Inception 网络不同的是 RFB 模块在各个分支中加入不同扩张率的空洞卷积,因为其能有效的获取小目标特征的有效信息,所以在目标检测方面有较好的应用。空洞卷积与常规卷积的不同在于前者在实现时通过扩张率这个参数来调整卷积过程中感受野的大小,且感受野大小变化与扩张率成正比。图 4(a)、(b) 分别展示了两者在相同大小的卷积核下感受野的大小示意图。

由图 4 可知,在特征图大小与卷积核大小均相同的情况下,空洞卷积在卷积核大小为 3×3,膨胀系数为 2 时可获得与大小为 5×5 的卷积核相同的感受野,而相同情况

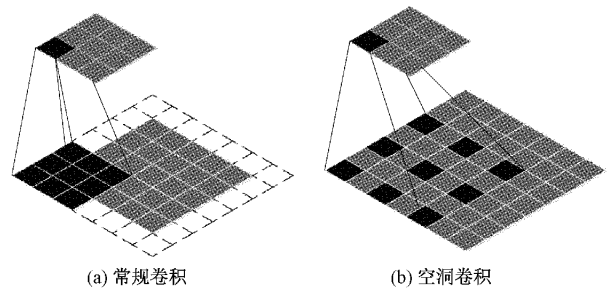


图 4 常规卷积与空洞卷积示意图

下常规卷积只能获得卷积核大小为 3×3 的感受野,说明在参数数量相同的情况下空洞卷积较常规卷积具有更大的感受野。空洞卷积在保证不损失图像信息以及输出特征图与常规卷积相同的情况下,保留了更多的图像特征并输出了较大范围的特征信息。在卷积神经网络中感受野的大小决定了是否能获取更多的图像信息,并生成更清晰的特征图。对于交通标志来说,其在图像中的位置往往是以小目标的形式出现,所以需要更大的感受野来获取信息,以便于对目标的分类识别。综合了 Inception 网络以及空洞卷积的 RFB 与 RFB-s 模块分别如图 5(a)、(b) 所示。

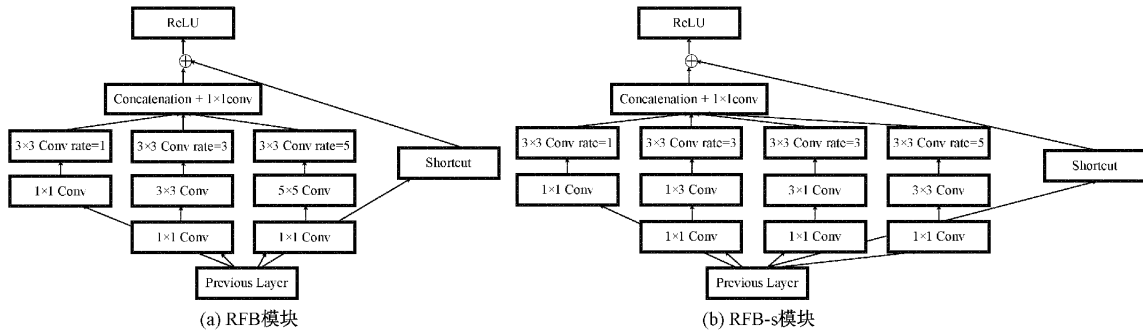


图 5 RFB 模块与 RFB-s 模块示意图

同时在增加 RFB 模块的将额外添加层的最后三层由原始的 1×1Conv+3×3Conv 替换为 1×1Conv+BN+ReLU+3×3Conv+BN+ReLU 的卷积块,在不增加参数数量的同时,提升网络的检测性能。

### 2.2 多尺度特征融合

通过上述对主干特征提取网络的改进,模型的检测性能相较于原始网络有了明显的提升。但是改进后的网络依然没有将各个特征层之间的信息有效的联系起来,故本文引入了 Bi-FPN 加权双向特征金字塔网络来有效结合深层和浅层的特征信息,使每一层的输出都包含了多尺度的上下文特征信息,提升预测的准确性。

在早期的目标检测中多使用特征金字塔网络(feature pyramid network, FPN)<sup>[18]</sup> 进行自上而下的特征融合,有效的将深层的特征信息融合到浅层中。根据这一想法,路径聚合网络(path aggregation network, PANet)<sup>[19]</sup> 在其基

础上添加了一条自下而上的路径聚合网络,使特征网络的深层与浅层信息进行有效的交互。但在以上网络进行特征融合时,会将具有不同分辨率的特征都调整到相同的分辨率,在对其加和。然而根据实验表明,不同尺度的输入特征对经过特征融合网络处理后的输出特征的影响是不同的,所以本文使用 Bi-FPN 网络来解决上述问题,该网络通过引入一个可学习的权重来学习不同输入特征的重要性。Bi-FPN 采用快速归一化融合的方式来学习权重,如式(5)所示。

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} \cdot I_i \quad (5)$$

其中,  $\omega_i$  为可学习的权重,且  $\omega_i \geq 0$ , 归一化后的取值范围在 0~1, 该方式与基于 Softmax 的融合方式相比具有相近的学习能力与准确率,但在 GPU 上的运行效率要比 Softmax 高 30%。本文所使用的 Bi-FPN 特征融合的结果



构如图 6 所示。

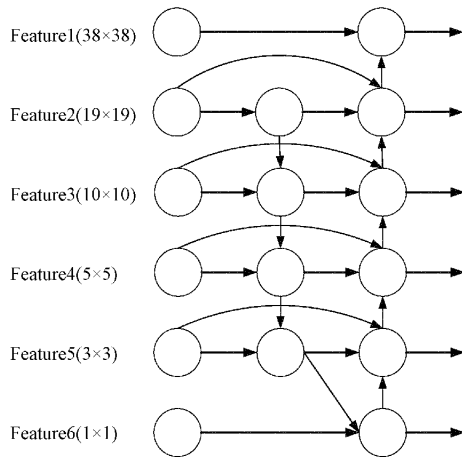


图 6 Bi-FPN 特征融合结构

由图 6 可知,  $Feature^*$  可由  $F_1$  表示, 左侧圆圈表示为  $F_5^{in}$ , 中间表示为  $F_5^{td}$ , 右侧表示为  $F_5^{out}$ 。  $F_5^{out}$  由  $F_5^{in}$ ,  $F_5^{td}$  以及  $F_4^{out}$  得到, 具体如式(6)。同时该结构为提高检测效率, 采用了经过优化的双向跨尺度连接结构, 在 PANet 的基础上先取消有一条输入边的结点, 因为这些结点对整合不同特征的特征网络没有实际效果。然后在同层次的输入节点和输出节点之间增加了跳跃结点, 由于他们在相同层, 在没有添加过多额外成本的同时, 整合了更多的特征数据。最后整个结构中采用了深度可分离卷积<sup>[20]</sup>实现特征融合, 并在每次卷积后添加批量归一化以及激活函数, 在不增加参数的同时提升了网络的整体性能。

$$F_5^{td} = Conv\left(\frac{\omega_1 \cdot F_5^{in} + \omega_2 \cdot upsample(P_6^{in})}{\omega_1 + \omega_2 + \epsilon}\right)$$

$$F_5^{out} = Conv\left(\frac{\omega'_1 \cdot F_5^{in} + \omega'_2 \cdot F_5^{td} + \omega'_3 \cdot downsample(F_4^{out})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon}\right)$$

(6)

Bi-FPN 采用双向跨尺度连接以及加权特征融合, 丰富了各个特征层的语义信息, 使目标检测算法具有更强的检测性能。

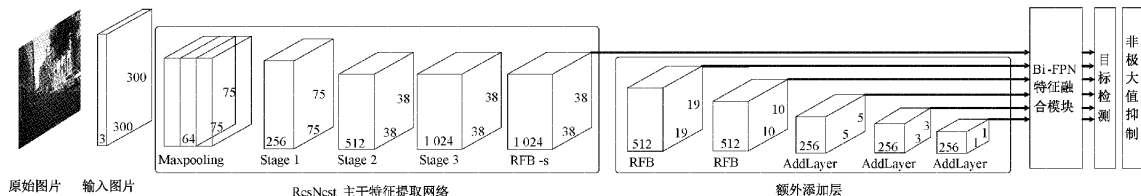


图 7 改进 SSD 模型结构

### 3 实验

为了检验本文所给出方法的合理性, 在中国交通标志数据集(CCTSDB)上对本文方法进行消融研究, 并与其他前沿方法在该数据集上的性能进行比较分析。

### 2.3 K-means++算法优化默认框参数

原始 SSD 模型采用在不同尺度的特征图上生成数量不同, 大小不同的 anchor 来检测目标, 其基准大小  $s_k$  的计算公式如式(7)所示。

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m - 1}(k - 1), k \in [1, m]$$

(7)

其中,  $m$  所代表的是共有几个尺度的特征层,  $s_{min}$  和  $s_{max}$  分别代表默认窗口在被检测图片上所占据的最小和最大比例。SSD 算法默认的 anchor 的宽高比为  $a_r \in \{1, 2, 3, 1/2, 1/3\}$ , 宽、高的计算公式为:  $w = s_k \times \sqrt{a_r}$ ,  $h = \frac{s_k}{\sqrt{a_r}}$ 。

但在现实中,  $s_{min}$  和  $s_{max}$  的数值必须反复试验才能得出, 实质上是盲目进行的, 且原始 SSD 算法在不同尺度特征图的默认框尺寸都较大, 但交通标志在图片中多以小目标的形式存在, 导致默认框太大而交通标志太小无法匹配, 对检测效率没有实质的提升。本文将通过 K-means++算法对训练集中的交通标志进行聚类来得到默认框的标准尺寸, 并以此使检测标志的标准尺寸成为一个先验知识, 通过聚类获取并进行对默认框标准尺寸的正确选取。

K-means++算法的具体实现为将本文所使用的数据集制作成数据样本, 从该样本集合中随机选择一个点作为此类样本的中心, 计算全部剩余样本与已聚类样本的距离, 分别按距离远近将剩余样本分类, 在从这些距离中计算出剩余样本成为聚类中心点的概率, 保证各个聚类中心点之间的距离最远并选出聚类中心, 重复上述步骤, 直到  $K$  个中心全部找到为止。选取默认框聚类所得默认框的基准大小分别为 [6.22 15.34]、[8.90 22.38]、[12.90 30.93]、[19.35 41.87]、[34.49 58.16]、[60.21 83.29]。

### 2.4 改进 SSD 网络模型

通过以上对原始 SSD 模型的改进, 得到本文最终模型如图 7 所示。本文以 ResNest 作为基准网络, 额外添加两层 RFB 模块以及三层  $1 \times 1 + 3 \times 3$  的卷积块, 最后将生成的特征图通过 Bi-FPN 模块进行特征融合。

### 3.1 实验平台

本次实验所用的操作系统为 Windows10, 硬件环境为 Intel i7 7700K 处理器、16 G 内存和 NVIDIA GeForce GTX3080 12 G 显卡, 软件环境为 Anaconda3、Cuda11.0、CUDNN7.6.3、Python3.7.6、Pytorch-GPU1.7.0。

3.2 实验数据集

中国交通标志数据集 CCTSDB 包含来自城市道路以及高速公路的包含不同自然场景的 15 734 张照片,其分辨率在  $1\ 000 \times 350 \sim 1\ 024 \times 768$ , 数据类别为蓝色的 mandatory 标志、红色的 prohibitory 标志以及黄色的 warning 标志。由于这 3 类标志在数量上不平衡,故本文选取在不同场景下相似度低的 10 000 张照片作为数据集,并使这 3 类标志的数量相对平衡,其中 8 000 张照片作为训练集,剩余的 2 000 张照片作为测试集对本文的算法进行评估。

3.3 损失函数以及评价指标

1) 损失函数

本文使用的损失函数为分类损失与回归损失的加权和,如式(8)所示。

$$L(f, c, l, g) = \frac{1}{N} (L_{conf}(f, c) + \alpha L_{loc}(f, l, g)) \quad (8)$$

其中,  $f$  表示预测框与目标框的匹配情况,  $c$  表示类的置信度,  $l$  表示预测框,  $g$  表示真实框,  $N$  表示匹配的默认框的个数,  $\alpha$  为权重系数。第一项分类损失  $L_{conf}(f, c)$  有 Softmax 求得,第二项回归损失  $L_{loc}(f, l, g)$  由 smooth L1 函数所得。 $L_{conf}(f, c)$  可由式(9)所得。

$$L_{conf}(f, c) = - \sum_{i \in Pos} f_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad (9)$$

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

式中:  $Pos$  表示交通标志正类别的数量,也即预测结果与实际标签相匹配的数量,  $Neg$  表示交通标志负类别的数量,也即预测结果与实际标签不一致的数量。 $f_{ij}^p$  表示类别为  $p$  的第  $i$  个默认框与第  $j$  个目标框的匹配情况,为 1 表示匹配成功,为 0 表示匹配失败。 $\hat{c}_i^p$  为表示第  $i$  个候选框属于类别  $p$  的置信度,  $p$  为 1 时表示交通标志,为 0 时表示背景。 $L_{loc}(f, l, g)$  可由式(10)所得。

$$L_{loc}(f, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} f_{ij}^p smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (10)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases}$$

$$\hat{g}_j^w = (g_j^{cx} - d_j^{cx}) / d_j^w \hat{g}_j^{cy} = (g_j^{cy} - d_j^{cy}) / d_j^h$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_j^w}\right) \hat{g}_j^h = \log\left(\frac{g_j^h}{d_j^h}\right)$$

式中:  $d$  表示真实框的参数,  $g$  表示预测框的参数,  $\hat{g}$  表示预测的偏移量。 $(cx, cy)$  是相对于中心点的偏移量。

2) 评价指标

本文使用精确率(Precision)和召回率(Recall)以及由两者分别作为纵横坐标所围成的面积(average precision,

AP)来评价各个类别的准确率,并通过检测一张图片的时间来评价模型的检测速率。模型检测的准确率与 AP 的取值成正比,通过计算所有类 AP 的平均值(mean average precision, mAP)来评价模型整体的准确率。其中 Precision、Recall、AP 以及 mAP 可用式(11)表示。

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

$$Recall(R) = \frac{TP}{TP + FN} \quad (11)$$

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

$$mAP = \frac{1}{C} \sum_{i=1}^C (AP)_i$$

式中:  $TP$  表示阳性样本个数(预测结果为正类与实际标注相同),  $FP$  表示负阳性样本个数(预测结果为正类与实际标注不同),  $FN$  表示假阴性样本个数(预测结果为负类与实际标注相同), Precision 则表示预测正确的正类图片的个数占总的正类预测个数的比例, Recall 表示预测正确的正类图片的个数占总的预测个数的比例。 $C$  表示数据集中的总类别数。

3.4 实验结果与分析

1) 消融实验结果与分析

为了验证本文的改进方法均对原始 SSD 算法有性能提升,在设置同等学习率以及训练批次的条件下,使用本文划分的测试集进行消融实验。图 8 展示了各个改进模型在测试集上的 Loss 曲线变化,由图可知,模型在 160 轮之后趋于收敛,改进模型前后的 Loss 分别保持在 1.980 和 1.110 附近。

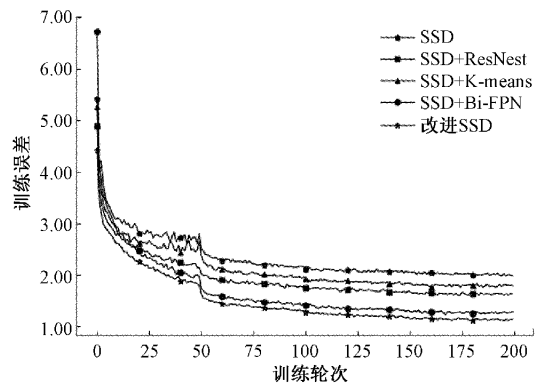


图 8 消融实验的 Loss 曲线

表 1 展示了针对原始 SSD 算法所提出的不同改进对算法的贡献,其中改进 1、2、3 分别表示特征提取网络的改进,添加特征融合模块,使用 K-means + 聚类算法对默认框参数进行优化。由表 1 可知,改进特征提取网络增强了弱目标特征的强表征能力,使模型能够充分学习交通标志的位置和方向信息,可将检测精度提升 5.72%,但由于使用了更深层次的网络,检测速度有所下降;使用 Bi-FPN

特征融合模块将深层与浅层的信息有效的结合,并且该模块使用了较多深度可分离卷积,在 mAP 达到 90.28% 的同时检测速度也有了较大提升;通过 K-means++ 算法对默认框参数进行优化,避免了盲目搜索带来的检测受限问题,使精度提升 6.28%。综合以上 3 种改进得到的改进 SSD 算法比原始 SSD 算法精度提升 9.25%,检测每张图片的时间为 27 ms,在保证精度的同时也能保证实时性。

表 1 不同改进对算法的贡献

算法	AP/%			mAP/ 时间/			
	1	2	3	mandatory	prohibitory warning	%	s
				87.08	83.38	87.77	86.08 0.011
✓				92.00	89.78	93.61	91.80 0.019
SSD	✓			91.65	87.58	91.61	90.28 0.009
		✓		93.92	90.95	92.23	92.37 0.012
	✓	✓	✓	96.12	93.56	96.33	95.33 0.027

2) 与不同交通标志检测算法的对比分析

为了验证本文模型的检测性能,表 2 展示了本文模型在 CCTSDB 数据集上与其他前沿学习方法的比较,其中包括 two-stage 的 Faster R-CNN 以及 one-stage 的 Yolo 系列网络。Faster R-CNN 在检测精度和速度上都低于前沿的 one-stage 算法;Yolov4 算法在检测精度和速度上都优于 Yolov3 算法,检测单张图片仅需要 14 ms;Yolov5-s 轻量级网络在检测精度上优于 Yolov3,略低于 Yolov4,但检测速度较前两者均有提升;原始 SSD 算法在三类交通标

志的检测准确率分别为 87.08%、83.38%、87.77%,在单个类别上均没有取得较好的优势;改进 SSD 算法在三类交通标志的检测准确率分别为 96.12%、93.56%、96.33%,相较于原始 SSD 模型分别提升了 9.04%、10.18%、8.56%,并且对于其他的前沿方法在单个类别上的检测精度均有所提升,检测速率较其他 one-stage 的算法略慢,但高于 two-stage 的 Faster-RCNN,检测单张图片需要 27 ms。

表 2 不同算法的性能对比

算法	AP/%			mAP/ 时间/	
	mandatory	prohibitory	warning	%	s
Faster R-CNN	82.94	85.32	89.31	85.15	0.042
YOLOv3	89.21	89.36	87.18	88.84	0.015
YOLOv4	91.65	91.23	91.08	91.37	0.014
YOLOv5-s	92.04	90.52	89.29	90.61	0.009
SSD	87.08	83.38	87.77	86.08	0.011
改进 SSD	96.12	93.56	96.33	95.33	0.027

此外,将上述几种方法以及改进 SSD 算法在 3 种类别交通标志的精确度 Precision 值以及召回率 Recall 值绘制成 PR 曲线。如图 9 所示,本文所提出的改进 SSD 算法在 3 种交通标志的检测结果中,其 PR 曲线所围成的面积均能包围其他的检测模型,证实了本文所提出的方法可以有效检测出其他算法未能检测到的交通标志,可被用于检测现实交通场景图片中的交通标志。

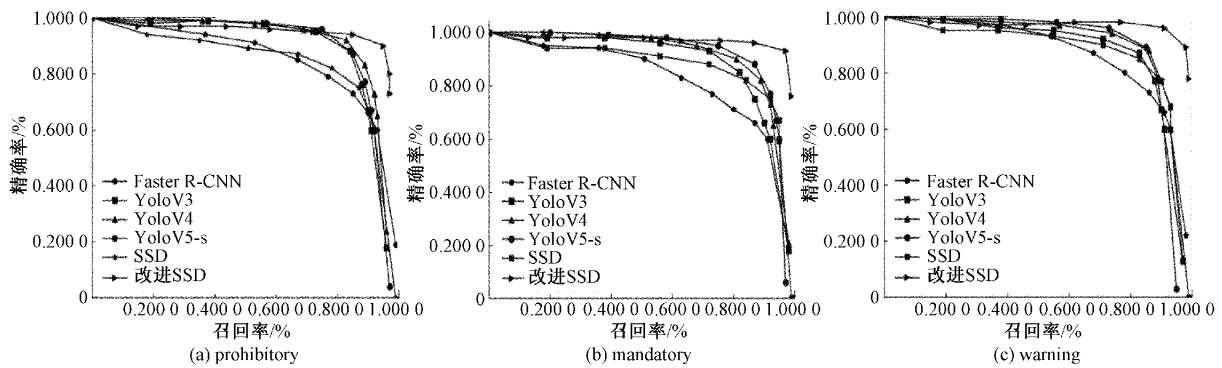


图 9 不同算法的 PR 曲线

4 结 论

针对真实复杂的交通环境下交通标志检测精度低、实时性差等问题,本文对 SSD 算法进行改进。通过使用 ResNest 网络代替原始 VGG 来增强弱目标特征的强表征能力,同时在额外添加层使用 RFB 模块增加目标特征的感受野,使网络能够获得更为有效的信息。在经过特征提取网络提取特征后,使用 Bi-FPN 模块将来自不同特征层的信息

进行有效的融合,使每一层的输出都包含多尺度的上下文特征信息,提升了预测的准确性。最后通过 K-means++ 算法对默认框的尺寸进行优化,解决了原始默认框盲目搜索带来的检测受限问题。在整个网络的搭建过程中还使用了深度可分离卷积以及膨胀卷积等方法,在网络参数没有增加的同时将检测精度提升了 9.25%。在未来的研究中,我们将继续研究如何在保持精度的前提下,减少网络的参数以及模型的复杂度,更好的应用于真实交通场景。

## 参考文献

- [1] GUDIGAR A, CHOKKADI S, RAGHAVENDRA U, et al. Multiple thresholding and subspace based approach for detection and recognition of traffic sign [J]. Multimedia Tools and Applications, 2016, 76 (5): 6973-6991.
- [2] BOUMEDIENE, CUDEL, BASSET, et al. Triangular traffic signs detection based on RSLD algorithm [J]. Machine Vision and Applications, 2013, 24(8): 1721-1732.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA, IEEE, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham, 2016: 21-37.
- [7] 刘鸣瑄,刘惠义. 基于特征融合 SSD 的远距离车辆检测方法[J]. 国外电子测量技术, 2020, 39(2): 28-32.
- [8] 刘琛,王江涛,王明阳. 引入视觉机制的 SSD 网络在摩托车头盔佩戴检测中的应用[J]. 电子测量与仪器学报, 2021, 35(3): 144-151.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [10] 彭继慎,孙礼鑫,王凯,等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报, 2021, 42(10): 161-170.
- [11] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: Split-attention networks [J]. ArXiv Preprint, 2020, ArXiv: 2004. 08955.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]. International Conference on Learning Representations, San Diego: Oxford, 2015: 1-14.
- [13] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385-400.
- [14] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020: 10778-10787.
- [15] 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, Las Vegas; IEEE, 2016: 770-778.
- [16] LI X, WANG W H, HU X L, et al. Selective kernel networks [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA. IEEE, 2019: 510-519.
- [17] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [18] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE. 30th IEEE Conference on Computer Vision and Pattern Recognition, New York; IEEE, 2017: 2117-2125.
- [19] 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, 2018: 8759-8768.
- [20] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017: 1800-1807.
- [21] ZHANG J M, ZOU X, LI D K, et al. CCTSDB 2021: A more comprehensive traffic sign detection benchmark [J]. Human-centric Computing and Information Sciences, 2022, DOI: 10. 22967/HGIS. 2022. 12. 023.

## 作者简介

赵友章(通信作者), 硕士, 主要研究方向为深度学习。

E-mail: 1015045093@qq.com

吕进, 副教授, 博士, 主要研究方向为交通信息工程及控制、人工智能及智能优化与控制、神经网络理论及应用、智能交通系统。

E-mail: lvjin@chd.edu.cn