

# 基于 Darknet23 和特征融合的交通标志检测方法\*

杜婷婷, 钟国韵, 江金懋, 任维民

(东华理工大学 信息工程学院, 江西南昌 330013)

**摘要:** 道路交通标志检测是智能交通的重要环节之一, 针对交通标志检测存在背景复杂、目标较小、检测速度慢等问题, 选取工业界青睐的 YOLOv3 模型提出一种改进的检测方法。利用双向特征金字塔结构实现图像低、中、高层特征语义信息的双向融合, 提升低层预测目标的分类和高层预测目标的定位能力; 将原模型的主干特征提取网络进行改进, 提出 Darknet23 网络, 以提高网络的提取能力和减少计算量; 根据目标形状的特点, 使用 K-means 聚类算法得到用于训练合适的锚点框, 并在边框回归中引入灵活性更强的  $\mathcal{L}_{\alpha-\text{CIoU}}$  损失函数, 使网络朝着预测框与真实框重叠度较高的方向去优化。实验结果表明, 该方法在 CCTSDB 数据集上 mAP@0.75 达到 86.10%、mAP@0.5:0.05:0.95 达到 70.017%, 相比原网络分别提升 10.17% 和 5.656%, 参数量减少 3 622 091, 速度提升 8.27 f/s, 且优于 SSD 和 Faster RCNN 等主流的检测网络。

**关键词:** 交通标志检测; 双向特征金字塔; Darknet23 网络; K-means 聚类; 损失函数

中图分类号: TP391

文献标志码: A

DOI: 10.16157/j.issn.0258-7998.222900

中文引用格式: 杜婷婷, 钟国韵, 江金懋, 等. 基于 Darknet23 和特征融合的交通标志检测方法[J]. 电子技术应用, 2023, 49(1): 14-19.

英文引用格式: Du Tingting, Zhong Guoyun, Jiang Jinmao, et al. Traffic sign's detection method based on Darknet23 and feature fusion[J]. Application of Electronic Technique, 2023, 49(1): 14-19.

## Traffic sign's detection method based on Darknet23 and feature fusion

Du Tingting, Zhong Guoyun, Jiang Jinmao, Ren Weimin

(School of Information Engineering, East China University of Technology, Nanchang 330013, China)

**Abstract:** Road traffic sign's detection is one of the important links of intelligent transportation. A detection method based on the improved YOLOv3 model by the industry is proposed for the problems of complex background, small targets and slow detection speed in traffic sign detection. The method used a bidirectional feature pyramid structure to achieve bidirectional fusion of semantic information of low, middle and high level features of images to improve the classification of low-level prediction targets and the localization of high-level prediction targets. The main feature extraction network of the original model is improved, and the Darknet23 network is proposed to improve the extraction ability of the network and reduce the computational burden. According to the characteristics of the target shape, the K-means clustering algorithm for training the appropriate anchor frames and a more flexible  $\mathcal{L}_{\alpha-\text{CIoU}}$  loss function is introduced into the bounding box regression to make the network optimize towards a higher degree of overlap between the prediction boxes and the ground-truth boxes. The experimental results show that the method reaches 86.10% mAP@0.75 and 70.017% mAP@0.5:0.05:0.95 on the CCTSDB dataset, which are 10.17% and 5.656% higher than the original network, the number of parameters is reduced by 3 622 091 and the speed is improved 8.27 f/s, which is better than mainstream detection networks such as SSD and Faster RCNN.

**Key words:** traffic sign's detection; bidirectional feature pyramid; Darknet23 network; K-means clustering; loss function

## 0 引言

随着全球汽车数量的逐渐增加, 智能交通系统(Intelligent Traffic System, ITS)迎来了新的发展机遇。交

通标志检测与识别技术作为在 ITS 中的一种关键技术应用, 它可以辅助驾驶员准确、高效地识别道路交通标志, 有效减轻驾驶疲劳, 从而保障安全驾驶。因此, 在真实路况下准确并及时地检测出交通标志, 对无人驾驶、高

\* 基金项目: 国家自然科学基金项目(62162002)

级辅助驾驶系统具有重要的现实意义。而在驾驶过程中,获取的交通标志图像存在背景复杂、目标小和光照天气变化等问题,导致交通标志的误检或漏检。

针对上述问题,近年来相关专家学者提出了许多不同的检测方法,主要可以分为两类:传统交通标志检测方法和基于深度学习的交通标志检测方法。传统的检测方法分为基于颜色空间<sup>[1]</sup>、基于形状特征<sup>[2]</sup>和基于多特征融合<sup>[3]</sup>,这3类方法往往面临诸如遮挡、褪色和环境等各种因素的影响而造成检测精度下降的问题。在这些方法的基础上,相关研究人员进行了深入的研究和优化,提出了基于深度学习的检测方法,分为基于锚框和不基于锚框两种。目前主流的算法基于锚框的,具体分为两阶段(Two-Stage)和单阶段(One-Stage)两种。前者的典型代表是Zuo Z<sup>[4]</sup>提出的Faster RCNN和徐国整<sup>[5]</sup>提出的改进Cascade R-CNN方法,这类算法首先在特征层上通过区域生成网络(Region Proposal Network, RPN)提取可能存在交通标志的预选框,再用对其进行检测,虽然可以大幅度提高检测精度,但由于计算量过大,检测速度较慢;后者的典型代表有Redmon J<sup>[6-8]</sup>提出的YOLO系列和孙超<sup>[9]</sup>提出的改进SSD方法,此类方法是直接对输入图像做回归任务,得到目标框的左上角、右下角坐标和类别信息,在输出层上一次性完成预测,虽然检测速快,适合做实时检测任务,但因为中间不需要额外的映射,在检测精度上会略有欠缺。张建民<sup>[10]</sup>等提出了融合注意力机制的Cascaded R-CNN算法,在CCTSDB数据集上mAP@0.5达到99.7%,但模型的平均检测速度仅为7.6 f/s;鲍敬源<sup>[11]</sup>等提出了Strong Tiny-YOLOv3算法,虽然在速度上达到33.78 f/s,但在mAP@0.5上只有85.56%。因此,本文提出一种兼检测精度高和速度快的交通标志检测方法。

## 1 YOLOv3 网络

Redmon J<sup>[8]</sup>提出的YOLOv3是单阶段检测算法YOLO系列的第3个版本,其检测精度和速度都有较好的表现,所以一直深受学术界和工业界的青睐。模型由主干特征提取网络、特征利用和预测头部组成。

Darknet53网络作为主干特征提取网络,其本质就是一系列卷积的过程,不断进行下采样,把输入图片的高宽不断压缩,通道数不断扩张。其重要特点是使用多个跳跃连接的残差块Resblock\_body来减少参数和降低训练的难度。在特征利用方面,模型使用特征金字塔(Feature Pyramid Network, FPN)结构。最后,将这3个增强的特征层传入预测头中,采取分治的思想进行分类和回归,对于大的目标在最高层(19,19,1 024)检测,中等的目标在中间层(38,38,512)检测,小的目标在最低层(76,76,256)检测以获得最终的预测结果。

## 2 基于Darknet23和特征融合的交通标志检测方法

针对交通标志检测上的不足,本文提出一种基于Darknet23和特征融合的交通标志检测方法,改进后的模型架构如图1所示。

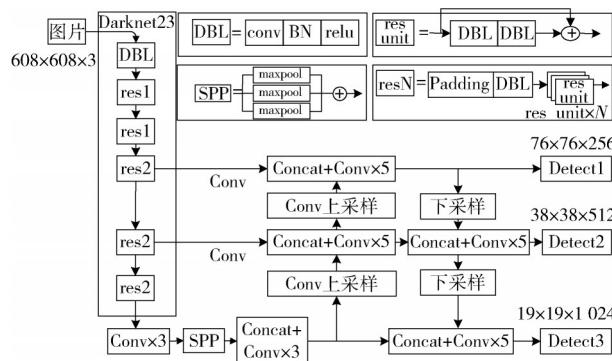


图1 改进的方法结构

### 2.1 Darknet23 网络

YOLOv3主干特征提取网络的一重要特点是残差块中的残差边可有效避免反向传播中出现梯度爆炸的问题,但在进行特征提取时,大量的卷积等操作也会引起特征语义信息的丢失,影响检测的精确度。

因此,为了避免这一现象发生和更加快检测速度,本文借鉴Darknet19特征提取网络结构小巧和性能强悍的优点,在其基础上通过减少Darknet53网络上残差单元数量,提出了Darknet23网络,即使整个主干特征提取网络只包含8个残差块以及23个卷积层(Convolutions, Conv),称为Darknet23,如图2所示。改进后的特征提取网络Darknet23含有5组残差网络,每组分别含有1、1、2、2、2个残差块。

### 2.2 双向金字塔模块

在基于卷积神经网络的交通标志方法中,处于最后一层的特征提取网络会因为感受野太大,丢失前面层的高分辨率和小目标,不利于目标识别,但它包含目标的位置信息准确,可以对目标进行精确的定位;而处于特征提取网络高层次的特征语义信息较多,但它包含目标位置信息较为模糊。

因此为了更好地适应大小目标均有的实时交通标志检测任务,更好地将低、中、高层的语义信息结合起来,本文将尺度为19×19的有效特征层拼接3次卷积和Purkait P<sup>[12]</sup>提出的空间金字塔池化SPP模块进行加强特征提取,结构如图2所示。并且采用如图3所示王林<sup>[13]</sup>提出的双向金字塔结构来进一步加强特征,即直接在P3、P4、P5引入一条自下到上的增强路径,每一层包含了低、中、高层丰富的特征信息和细粒度的信息,并且额外地增强路径也可以减少底层与顶层特征融合信息路

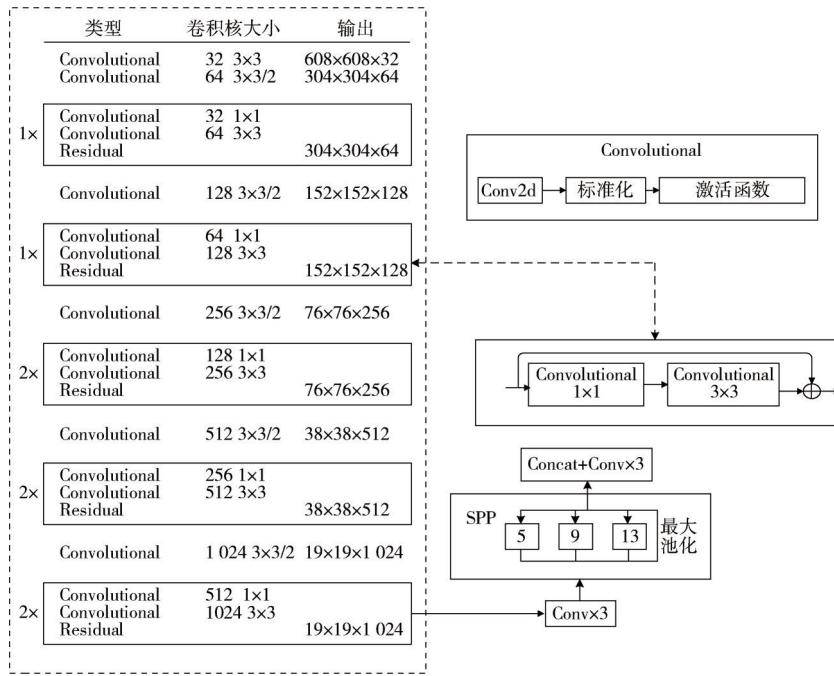


图2 Darknet23 网络

径的计算量。其中在高层和中层、中层和低层做特征融合时,采用可以减少计算时的内存访问成本(Memory Access Cost, MAC)的路径聚合网络(Path Aggregation Network, PAN),其结构如图4所示。

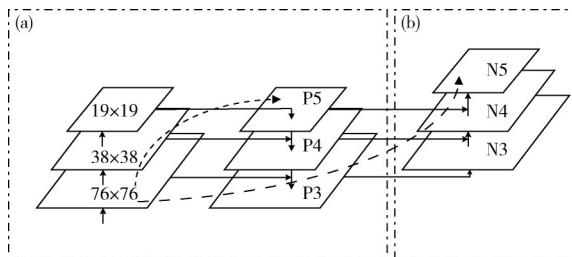


图3 双向特征金字塔结构

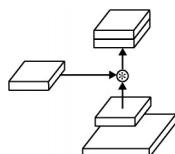


图4 路径聚合网络

### 2.3 $\mathcal{L}_{\alpha-\text{CIOU}}$ 损失函数

在边界框损失函数的早期计算中,广泛使用L1范数或L2范数。在本文中,使用可以更好综合考虑BBox的重叠面积、中心点距离以及BBox长宽比一致性的 $\mathcal{L}_{\text{CIOU}}$ 损失函数的变体,即He J<sup>[14]</sup>等提出的 $\mathcal{L}_{\alpha-\text{CIOU}}$ 损失函数。计算公式见式(1)~式(3):

$$\mathcal{L}_{\alpha-\text{CIOU}} = 1 - \text{IoU}^\alpha + \frac{\rho^{\alpha}(b, b^{\text{gt}})}{c^{2\alpha}} + (\beta v)^\alpha \quad (1)$$

$$\beta = \frac{v}{(1 - \text{IoU}) + v} \quad (2)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

其中, $b$ 和 $b^{\text{gt}}$ 表示预测框和真实框的中心点, $\rho(\cdot)$ 为欧几里得距离, $c$ 为覆盖预测框和真实框的最小封闭方框的对角线长度, $\beta$ 是一个正的权衡考数, $v$ 衡量纵横比的一致性。

该损失函数的一大优点是把功率参数 $\alpha$ 用作旋钮,为检测器提供更大的灵活性,以调整 $\alpha - \text{IOU}$ 损耗,满足不同级别的BBox回归精度。并且 $\mathcal{L}_{\alpha-\text{CIOU}}$ 的实验结果已经证明,当 $\alpha > 1$ 时, $\mathcal{L}_{\alpha-\text{IOU}}$ 损失函数更多地关注IOU数值高的对象来增加BBox的回归精度;当 $0 < \alpha < 1$ 时,它会降低重叠度高对象的权重,从而影响回归精度;当 $\alpha = 3$ ,在大多数情况下表现良好,因此本文方法采用 $\mathcal{L}_{3-\text{CIOU}}$ 损失函数。

## 3 实验

本文实验采用PyTorch深度学习框架在Windows 10操作系统下进行,CPU为i7-11800H@2.30 GHz,GPU为NVIDIA GeForce RTX 3060,CUDA版本为11.0,cuDNN版本为8.0.5,Torch和Torchvision版本分别为1.7.1、0.8.2。改进后损失值Loss(包括训练集损失train loss、验证集损失val loss、平滑训练集损失smooth train loss、平滑验证集损失smooth val loss)随迭代次数Epoch变化曲线如图5所示。

### 3.1 实验数据集

本文采用张建明<sup>[15]</sup>团队公开制作的中国交通标志

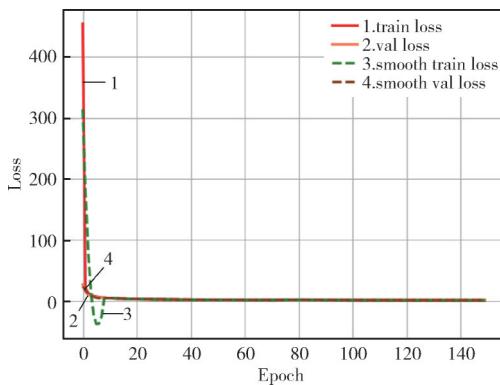


图 5 损失值随迭代次数变化曲线

数据集 (CSUST Chinese Traffic Sign Detection Benchmark, CCTSDB)。该数据集分为警告 (warning)、指示 (mandatory) 和禁止 (prohibitory) 标志 3 个类别。由于 CCTSDB 数据集进行过扩充, 多段行驶的图片相似度过高, 因此本文只选取了前 4 000 张具有代表性的图片, 以 4:1 的比例随机划分为训练集加验证集和测试集两类, 实验结果如图 6 所示。



图 6 CCTSDB 数据集部分样本检测结果

### 3.2 K-means 聚类

锚框作为初始候选框, 宽度和高度均是固定的, 设计合理的锚框将会直接影响检测精度和速度。而 K-means 聚类算法具有简洁、高效的优点, 并且在进行计算时, 不是直接利用欧氏距离, 而是做归一化限制距离的取值范围为 0~1, 以 1-IOU 等作为衡量, 取消因目标框大小产生的误差, 算法的具体流程为: 首先, 从数据集中随机选择  $m$  个数据点作为初始聚类中心; 其次, 计算每个点到中心点的距离, 将每个数据点分到距离它最近的聚簇中; 然后, 重新计算每个聚簇的中心点; 最后, 当前中心点的位置不再更新或者达到设定的迭代次数则停止, 否则, 继续操作。更改后的先验框为 (92,127), (54,79), (39,57), (27,44), (21,34), (15,43), (15,30), (13,22), (10,20)。

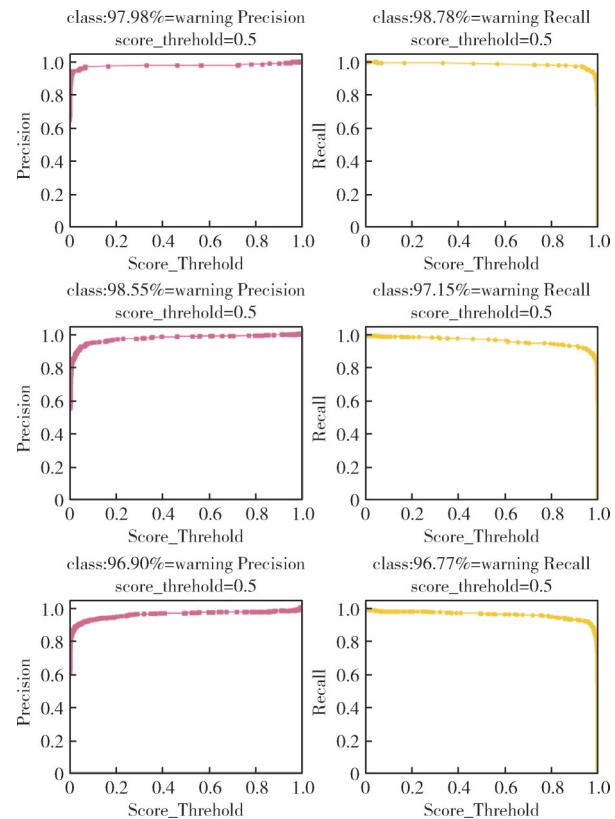
### 3.3 评价指标

交并比 (Intersection Over Union, IOU) 表示某一目标真实框与预测框的交集和并集的比率, 数值越高说明检测框与目标框拟合程度越好, 见式(4):

$$\text{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (4)$$

精确率 (Precision,  $P$ ) 可以直观衡量模型的错检程

度, 召回率 (Recall,  $R$ ) 可以直观衡量模型漏检程度。在 CCTSDB 数据集中, 每一类交通标志不同阈值 (Score\_Threshold) 的  $P$ 、 $R$  如图 7 所示。

图 7 CCTSDB 数据集每一类在不同阈值的  $P$ 、 $R$ 

平均精度均值 (Mean Average Precision, mAP) 表示所有类别精度的平均值, 其中  $m$  为类别数量, 计算见式(5):

$$\text{mAP} = \frac{1}{m} \sum_{i=1}^m \text{AP}_i \quad (5)$$

$\text{mAP}@0.5$  表示当 IOU 取 0.5 时, 计算每一类所有图片的精度均值 (Average Precision, AP), 然后将所有类别求平均, 即  $\text{mAP}@0.5$ , 计算见式(6)。在 CCTSDB 数据集中以  $\text{IOU}=0.5$  为例, 每一类的 AP 和 mAP 值如图 8 所示。

$$\text{mAP}@0.5 = \frac{1}{C} \sum_{k=1}^C \text{AP}@0.5_k \quad (6)$$

$\text{mAP}@0.5:0.05:0.95$  表示  $\text{mAP}@0.5$  以 0.05 的步长增至  $\text{mAP}@0.95$  取所有  $\text{mAP}$  的均值, 可以衡量模型在不同 IOU 阈值下的综合表现。

参数量 (Params) 是指模型训练中需要训练的参数总数, 只与定义的网络结构有关, 与前向传播任何操作无关。即定义好了网络结构, 参数就已经决定。

除上述检测精度指标外, 交通标志检测的另一个极其重要的性能指标是速度, 常用的是每秒帧率 (Frame

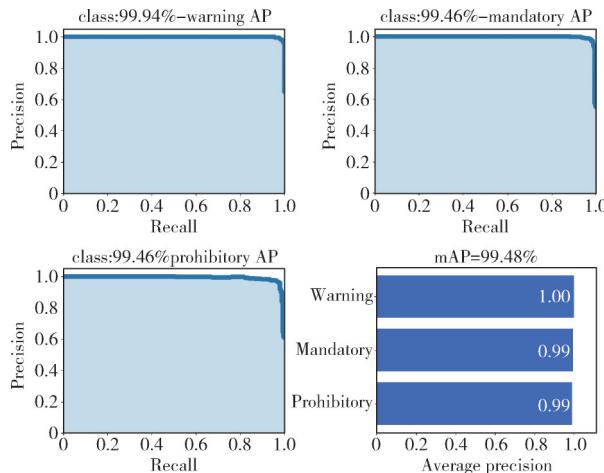


图8 CCTSDB 数据集每一类在 IOU 为 0.5 时的 AP、mAP

Per Second, FPS), 即每秒可处理的图片数。

#### 3.4 消融实验

为了验证本实验各个改进模块对检测算法的影响,本文进行了消融实验,把各个改进模块依次添加进原网络,测试每个改进模块对检测精度和速度的影响,具体对比结果见表1。

通过1、2组实验对比结果可知,改进特征金字塔模块,检测精度优于原始;在检测速度方面,虽然参数量增加了16 381 696,但是产生的影响也不太大,仅降低了1.92 f/s。

其次,将原模型的主干特征提取网络改进,即整个主干特征提取网络只含有5组12个残差网络,每组网络含有1、1、4、4、2个残差块,以及31个卷积层,故称之为Darknet31。实验2、3对比可以看出,原始网络比Darknet31额外多的11个残差块并没有使检测精度取得太大优势。在检测速度上,由于残差块的减少,网络计算量减少,参数量减少17 137 024,改进后的网络检测速度达到35.34 f/s,比实验2快8.26 f/s。因此为了极限缩小网络,本文再次重新设计提取网络,即只含有5组8个残差块的Darknet23网络。实验结果显示,Darknet 23检测效果更好。所以,对于本文的3类交通标志检测任务,选用总体偏优的Darknet23作为特征提取网络,检测精度

和速度更好。

最后,在精度优良和速度不错的实验4上改进损失函数,实验5结果显示,真实框和预测框的误差也在减小,也没有引入额外的参数,检测速度达到37.27 f/s,比原模型速度提升8.27 f/s。

#### 3.5 性能比较

本文方法与主流的SSD、Faster RCNN等模型进行对比实验,各个模型的性能效果如表2所示。

表2 检测网络性能对比

模型	网络	mAP	mAP@	AP(0.5:0.95) /%	FPS/(f/s)
		@0.5 /%	0.75 /%		
SSD512	VGG	94.22	72.42	65.504	41.760
SSD512	MobileNetv2	89.38	62.75	56.422	90.459
Faster RCNN600	VGG	80.72	36.40	42.348	15.657
Faster RCNN600	Resnet50	79.91	35.13	41.690	10.741
YOLOv3	Darknet53	97.95	75.93	64.361	29.00
本文	Darknet23	99.48	86.10	70.017	37.27

#### 4 结论

本文提出了一种基于Darknet23和特征融合的交通标志检测方法。实验结果表明,双向特征金字塔模块可以增强交通标志目标的显著性;Darknet23网络可以增强交通标志的特征表达能力,在一定程度上减少计算量,加快检测速度。改进的损失函数使网络朝着预测框与真实框重叠度较高的方向去优化,提升预测目标的精确度。在未来的研究工作中,将继续增大交通标志数据集的种类研究,找到平衡速度和精确度的最优模型并运用到实际检测与识别之中。

#### 参考文献

- [1] 王济民,魏怡,周宇,等.基于LeNet-5卷积神经网络和颜色特征的限速标志识别[J].计算机科学,2021,48(S2):345-50.
- [2] 苗丹,卢伟,高娇娇,等.基于聚类与Hough变换的交通标志检测方法[J].计算机系统应用,2019,28(11):213-217.
- [3] MADANI A,YUSOF R.Traffic sign recognition based on

表1 消融实验

组号	特征金字塔	Darknet31	Darknet23	损失函数	mAP	mAP @ 0.5/%	mAP @ 0.60/%	mAP @ 0.65/%	mAP @ 0.70/%	mAP @ 0.75/%	mAP @ 0.80/%	mAP @ 0.85/%	mAP @ 0.90/%	mAP @ 0.95/%	AP(0.5:0.95) /%	参数量	FPS/(f/s)
					0.5/%	%	%	%	%	%	%	%	%	%	%		
1	否	否	否	否	97.95	97.65	97.01	94.60	87.84	75.93	58.32	27.52	6.56	0.23	64.361	61534504	29.00
2	是	否	否	否	99.14	98.97	98.28	95.82	90.08	78.24	56.43	26.65	4.86	0.17	64.864	77916200	27.08
3	是	是	否	否	99.19	98.99	98.11	96.10	90.99	78.28	57.07	29.17	6.53	0.08	65.451	60779176	35.34
4	是	否	是	否	99.21	99.14	98.42	96.69	91.88	82.24	59.58	28.77	6.25	0.16	66.234	57912413	40.01
5	是	否	是	是	99.48	99.29	98.69	96.93	93.13	86.10	68.61	42.97	14.52	0.45	70.017	57912413	37.27



- color, shape, and pictogram classification using support vector machines[J]. Neural Computing and Applications, 2018, 30:2807–2817.
- [4] ZUO Z, KAI Y, QIAO Z, et al. Traffic signs detection based on Faster R-CNN[C]// 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, 2017.
- [5] 徐国整,周越,董斌,等.基于改进 Cascade R-CNN 的交通标志牌识别 [J].传感器与微系统,2021,40(5):142–145,153.
- [6] 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: 779–788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263–7271.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv e-prints arXiv:1804.02767, 2018.
- [9] 孙超,温蜜,景俐娜.改进 SSD 算法在交通标志检测中的应用[J/OL].计算机工程与应用, 1-10[2022-11-28].<http://kns.cnki.net/kcms/detail/11.2127.TP.20210928.0236.006.html>.
- [10] ZHANG J, XIE Z, SUN J, et al. A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection[J]. IEEE Access, 2020, PP(99):1.
- [11] 鲍敬源,薛榕刚.基于 YOLOv3 模型压缩的交通标志实时检测算法 [J].计算机工程与应用, 2020, 56(23): 202–210.
- [12] PURKAIT P, ZHAO C, ZACH C. SPP-Net: deep absolute pose regression with synthetic views[J]. CVPR, arXiv preprint arXiv:1712.03452, 2017.
- [13] 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.
- [14] HE J, ERFANI S, MA X, et al. Alpha-IoU: a family of power intersection over union losses for bounding box regression[J]. Advances in Neural Information Processing Systems, 2021, 34: 20230–20242.
- [15] ZHANG J, HUANG M, JIN X, et al. A real-time Chinese traffic sign detection algorithm based on modified YOLOv2 [J]. Algorithms, 2017, 10(4): 127.

(收稿日期:2022-04-21)

**作者简介:**

杜婷婷(1998-),女,硕士研究生,主要研究方向:计算机视觉。

钟国韵(1979-),通信作者,男,博士,副教授,主要研究方向:计算机视觉和图像音视频处理等,E-mail:gyzhong@ecut.edu.cn。

江金懋(1995-),男,硕士研究生,主要研究方向:视觉障碍感知、单目深度估计、道路分割和小目标检测。



扫码下载电子文档

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所