

基于改进 YOLOv3 的交通标志检测

王 卜¹, 何 扬²

(1. 四川大学计算机学院, 成都 610065;
2. 四川大学视觉合成图形图像技术重点学科实验室, 成都 610065)

摘 要: 针对交通标志检测小目标数量多、定位困难及检测精度低等问题, 本文提出一种基于改进 YOLOv3 的交通标志检测算法. 首先, 在网络结构中引入空间金字塔池化模块对 3 个尺度的预测特征图进行分块池化操作, 提取出相同维度的输出, 解决多尺度预测中可能出现的信息丢失和尺度不统一问题; 然后, 加入 FI 模块对 3 个尺度特征图进行信息融合, 将浅层大特征图中包含的小目标信息添加到深层小特征图中, 从而提高小目标检测精度. 针对交通标志数据集特点, 使用基于 GIoU 改进的 TIoU 作为边界框损失函数替换 MSE 函数, 使得边界框回归更加准确; 最后, 通过 k-means++ 算法对 TT100K 交通标志数据集进行聚类分析, 重新生成尺寸更小的候选框. 实验结果表明, 本文算法与原始 YOLOv3 算法相比 mAP 提升 11.1%, 且检测每张图片耗时仅增加 6.6 ms, 仍符合实时检测要求. 与其他先进算法相比, 本文算法具有更好的检测精度和检测速度.

关键词: 小目标检测; 交通标志; YOLOv3; 信息融合; TIoU

中图分类号: TP39 **文献标识码:** A **DOI:** 10.19907/j.0490-6756.2022.012004

Traffic sign detection based on improved YOLOv3

WANG Bu¹, HE Yang²

(1. College of Computer Science, Sichuan University, Chengdu 610065, China;
2. State Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China)

Abstract: Aiming at the problems of traffic sign detection such as large number of small targets, difficulty in positioning and low detection accuracy, a traffic sign detection algorithm based on YOLOv3 is proposed. First, to solve the problem of information loss and scale disunity that may occur in the multi-scale prediction, the spatial pyramid pooling module is introduced into the network structure to perform the block pooling operation on three prediction feature maps with different scales, and the output of the same dimension is extracted,. Secondly, in order to improve the accuracy of small target detection, the FI module is added to carry out information fusion of the three scale feature maps, and the small target information contained in the shallow large feature map is added to the deep small feature map. According to the characteristics of traffic sign dataset, the MSE function is replaced by the TioU improved based on GIoU as the bounding box loss function, which makes the boundary box regression more accurate. Finally, k-means++ algorithm is used to perform clustering analysis on the TT100K traffic sign dataset, new anchors are generated with smaller size. Experimental results show that the proposed algorithm im-

收稿日期: 2021-03-02
基金项目: 国家重点研究与发展计划基金(2018YFC0830300); 国家自然科学基金(61571312)
作者简介: 王卜(1995-), 男, 硕士研究生, 主要研究方向为计算机视觉. E-mail: wangbu@stu.scu.edu.cn
通讯作者: 何扬. E-mail: heyang@stu.scu.edu.cn

proves mAP by 11.1% compared with the original YOLOv3 algorithm, and the detection time of each image only increases by 6.6ms, which still meets the real-time detection requirements. Compared with other advanced algorithms, the proposed algorithm has better detection accuracy and detection speed.

Keywords: Small object detection; Traffic sign; YOLOv3; Information fusion; TIoU

1 引言

近年来,随着深度学习的快速发展和不断进步,目标检测作为计算机视觉领域的一个基本问题,被广泛应用于自动驾驶、行人检测、人脸识别及无人机等领域,是理论和应用结合的研究热点. 目标检测的主要目标是在图像或图像序列中精确定位出各种目标的类别和位置信息^[1]. 在车辆自动驾驶任务中,要保证汽车能够在马路上安全地行驶,必须快速并且准确地识别出前方的交通标志牌,在车辆通过该指示牌前给予车辆正确的指令. 在检测速度不变的情况下,越早开始检测前方交通标志,在车辆到达标志牌前留给车辆执行指令的时间也越长,可以增加容错性,进而提高车辆自动驾驶安全性. 因此尽早开始检测前方交通标志,也意味着当前车辆离前方标志的距离较远,此时交通标志在整个图像中只占据较小的像素,可以算作小目标^[2]. 微软公司的 MS COCO 数据集^[3]中将小目标进行了绝对尺度上的定义,即数据集中尺寸小于等于 32×32 个像素值的目标. 而小目标检测领域中存在缺少大规模数据集支持、自身分辨率低、像素占比少、复杂环境中容易被干扰等问题. 因此,在进行交通标志检测时,我们需要专门的数据和特定的算法.

随着卷积神经网络^[4]研究深入,现在的目标检测算法在许多通用数据集上,包括 PASCAL-VOC^[5]、ImageNet^[6]、MS COCO^[7]、OpenImage^[8],以及行人检测、车辆检测等领域都取得了良好的精度和较快的识别速度,但是它们检测结果中的小目标检测效果往往不够理想. 小目标检测任务在目标检测领域仍然是一项具有挑战性的任务^[9],当前热门的目标检测算法主要分为两大类:一种是以 R-CNN^[10]、SPP-Net^[11]、Fast R-CNN^[12]、Faster R-CNN^[13]、R-FCN^[14]等为代表的基于候选区域的两阶段(Two Stage)目标检测算法. 另一种是以 YOLO^[15]、SSD^[16]等为代表的基于回归的单阶段(One Stage)目标检测算法. 两阶段算法需要通过边界框搜索算法^[17]或者选择性搜索算法^[18]生成一系列候选区域,因此尽管其检测精度较高,但是

检测速度很慢,远不能满足实时检测的需求. 而单阶段算法因为省略了生成候选框的步骤,检测速度大大提升,但是对小目标的检测精度较差,不能满足真实场景应用中对准确率的要求. 为了解决 YOLO 算法对小目标检测不理想的问题,Redmon 等又提出了新的 YOLOv3^[19]算法,引入特征融合实现多尺度预测,不同尺度的特征图分别对大、中、小目标进行检测,有效提高了小目标检测效果.

本文以 YOLOv3 为基础提出一种新的改进算法应用于交通标志检测中,为了进一步提高算法的检测精度,对网络模型和损失函数进行优化,通过一系列对比实验表明,改进后的算法精度更高,小目标检测效果提升明显.

2 YOLOv3 算法

为了提高 YOLO 的小目标检测效果,Joseph 等在 YOLOv1 和 YOLO9000^[20]的基础上做了进一步改进. 他们用 Darknet-53 代替 Darknet-19 成为新的特征提取骨干网络,同时参考特征金字塔网络^[21](FPN),在预测网络中引入基于多尺度预测的特征金字塔结构,利用网络中三个不同位置、不同大小的特征图对大、中、小三种目标进行检测. Darknet-53 由若干个 DBL 模块和残差结构组成,残差结构的引入可以让网络结构设置得更深,从而具有更强的特征提取能力. DBL 模块由卷积层、批归一化处理(Batch Normalization)^[22]和 Leaky ReLU 激活函数三部分组成,可以有效解决梯度消失、梯度爆炸以及过拟合等不利于训练的问题.

YOLOv3 首先通过 Darknet-53 提取输入图像特征,当提取到最深层特征比如特征图大小为 13×13 时,将此特征图送入预测层输出的同时进行上采样操作,与另一个尺度为 26×26 的特征图进行融合. 重复该操作可以让参与预测的三个特征图中都包含浅层低级特征(如物体的边缘、颜色等)和深层高级特征(如物体的语义信息等),有利于提高预测精度. 其中浅层特征图感受野较小,用于预测小目标;深层特征图感受野较大,用于预测大目标. YOLOv3 引入多尺度预测的特征金字塔结构,将预测层分为 13×13 、 26×26 、 52×52 三个不同尺

度,在大、中、小目标的检测上都取得了不错的成果, YOLOv3 的网络结构如图 1 所示。

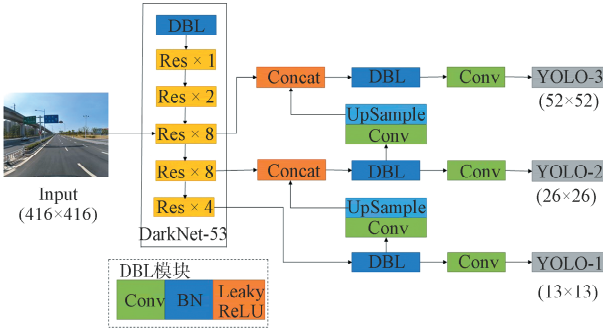


图 1 YOLOv3 网络结构

Fig. 1 YOLOv3 network architecture

YOLOv3 对 YOLOv1 和 YOLO9000 的损失函数做了进一步优化,它在边界框回归、置信度、分类三个部分都做了改进,让当前的目标检测器能够取得更好的识别精度和定位精度。YOLOv3 的损失函数如下所示。

$$lbox = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} \cdot (2 - w_i \times h_i) \cdot [(\hat{w}_i - \tilde{w}_i)^2 + (h_i - \hat{h}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} \cdot [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \quad (1)$$

$$lcls = \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c \in \text{class}} BCE(\hat{p}_i, p_i) \quad (2)$$

$$lobj = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} BCE(\hat{c}_i, c_i) + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} BCE(\hat{c}_i, c_i) \quad (3)$$

$$Loss = lbox + lcls + lobj \quad (4)$$

其中, S 表示网络输出层网格数量; S^2 表示网络输出的特征图大小,即 13×13 、 26×26 、 52×52 ; B 表示初始候选框 Anchors 的数目; 1_{ij}^{obj} 表示如果第 i 个网格的第 j 个候选框负责当前这个 object, 则 1_{ij}^{obj} 的值为 1, 反之就为 0; 1_{ij}^{noobj} 与 1_{ij}^{obj} 的表示含义相反。YOLOv3 在前两个版本的边界框损失函数基础上添加因式 $(2 - w_i \times h_i)$ 并且去掉了 w 和 h 的根号; 在置信度和分类损失上采用 Logistic 回归, 损失函数从原来的 MSE 改为交叉熵损失函数 BCE, BCE 的计算公式如下。

$$BCE(\hat{c}_i, c_i) = -\hat{c}_i \times \log(c_i) - (1 - \hat{c}_i) \times \log(1 - c_i) \quad (5)$$

3 改进的 YOLOv3 算法

YOLOv3 算法在大多数通用数据集中都取得不错的成绩,但是这些数据集中大型目标数据占绝大部分,小目标数据量较少,原模型对小目标检测效果并不理想。而在真实应用场景中,为了保证车辆与驾乘人员安全,必须快速且正确地完成交通标志识别任务。因此当车辆前方一出现交通标志时就开始检测,在车辆到达标志前尽快完成从而留出足够时间让汽车执行操作指令。而检测开始时汽车距离前方交通标志较远,且交通标志本身较小在整幅图像中只占据很小的部分,所以交通标志检测任务属于小目标检测范畴。本文通过改进 YOLOv3 网络结构,优化边界框损失函数,最后通过 k-means++ 聚类算法重新生成候选框,提出一个能更好完成交通标志检测任务的改进算法。

3.1 改进 YOLOv3 网络结构

在原始算法中深层特征图负责检测大目标,中层特征图负责检测中型目标,但是在交通标志检测任务中大多检测是小目标检测。当输入图片通过特征提取网络得到尺度为 13×13 和 26×26 的特征图时,原图片中的小目标也会缩小到几个像素之内,丢失大量信息,因此原模型中这两个预测层的检测效果并不理想。为了改善两个中大目标预测层的检测效果,同时进一步提高尺度为 52×52 负责小目标的预测层精度,本文在 3 个特征图输入预测层之前添加一个 FI (Fusion Information) 模块,结构如图 2 虚线框所示。

FI 模块将大特征图下采样,小特征图上采样,然后与自身 Concat 结合,保留自己主要信息的同时添加一些辅助信息。其中 Concat 操作是指将两个大小相同的特征图在通道维度上结合起来,如 $13 \times 13 \times 256$ 的特征图与 $13 \times 13 \times 256$ 的特征图经过 Concat 操作后结合成一个 $13 \times 13 \times 512$ 的特征图。该模块使用最大化 (Maxpool) 对图像下采样,可以提取出特征响应最大、最强烈的部分传递待融合特征图,从而增强信息。

对于尺度为 13×13 的深层特征图,FI 模块会把尺度为 52×52 和 26×26 的特征图下采样到 13×13 ,然后再与其 Concat 结合,融入浅层特征图中更多关于小目标的信息,从而提升小目标定位精度;对于尺度为 26×26 的特征图,FI 模块会将 52×52 的特征图下采样以及 13×13 的特征图上采样后再 Concat 结合;对于尺度为 52×52 的小目标检测层,

13×13 和 26×26 的特征图经 FI 模块上采样到 52×52 尺度后再进行融合,可以添加一些目标的高级特征提高识别精度. 经过实验表明,FI 模块的加入在增加少量模型复杂度和计算参数数量的前提下,极大提高了交通标志检测效果.

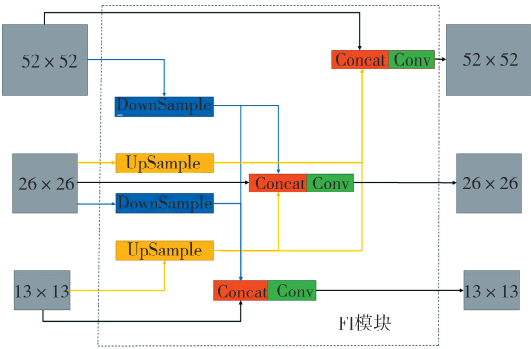


图 2 FI 模块结构
Fig. 2 FI module structure

考虑到特征融合时自身尺度特征图包含的信息应该作为主要信息,另外两个尺度特征图信息作为辅助信息进行补充,因此在融合前引入空间金字塔池化^[23] (SPP)对特征图分块池化提取多尺度特征,生成 3 个包含主要信息的待融合特征图. 然后选择 Darknet-53 特征提取过程的特征图分别上下采样作为辅助信息,再与待融合特征图进行信息融合. FI 模块在没有增加网络深度和模型复杂度的前提下,融合多尺度特征增加网络宽度,从而增强信息提高目标检测精度. 以尺度 26×26 的特征图为例,其特征图选择与融合过程如图 3 所示.

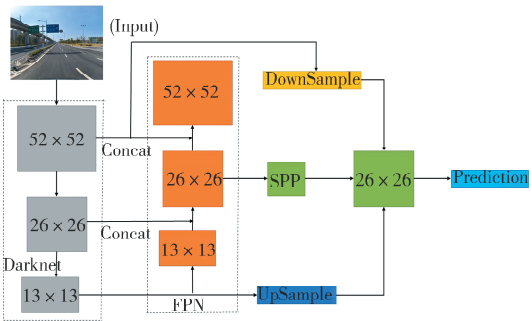


图 3 特征图选择与融合(26×26)
Fig. 3 Feature map selection and fusion (26×26)

在图 3 中,灰色矩形为 Darknet 特征提取阶段的基础特征图,橙色矩形为 FPN 阶段的多尺度特征图,绿色矩形为经过空间金字塔池化后的待融合特征图(主要信息). 在网络预测阶段,对于尺度 26×26 的检测层,首先选择 Darknet 阶段中 52×52 和 13×13 大小的特征图(灰色)分别下、上采样,然后

和 SPP 池化后的自身特征图(绿色)进行 Concat 结合,实现特征融合信息增强后再预测目标.

其中,空间金字塔池化(SPP)对不同大小特征图通过 3 个尺寸的分块网格进行池化操作,每个块提取一个特征作为一个维度从而输出相同大小的特征图,可以解决图片在多尺度预测时可能存在的信息丢失和尺度不统一问题. 为了保留更多小目标信息,故在 FI 模块之前加入空间金字塔池化层. 当池化网格大小尽可能接近特征图大小时,可以实现局部特征和全局特征的融合,并且因为最小特征图尺寸是 13×13,所以把池化网格的大小依次设置为 5,9,13. 多尺度图像经 SPP 池化后输出维度统一且信息充足的特征图,再通过 FI 模块进行信息融合,使小目标信息尽可能多得用于预测中,有效提高改进模型的小目标检测精度. 改进后的 YOLOv3 网络结构如图 4 所示.

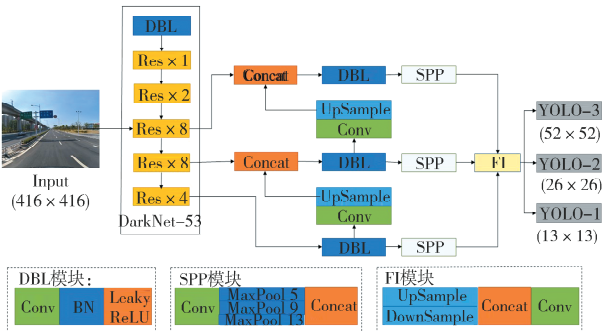


图 4 改进的 YOLOv3 网络结构
Fig. 4 Improved YOLOv3 network architecture

3.2 边界框回归损失函数优化

当前大多数目标检测算法包括 YOLOv3 都使用交并比(IoU)作为度量标准,但原始 YOLOv3 的边界框回归损失函数采用的是均方误差(MSE)函数. 但是 MSE 只能表示两个框之间的距离关系,不能表示它们的相交情况,如图 5 所示,其中红色矩形为目标框,绿色矩形为预测框. 从图 5 可以明显看到在相同 MSE Loss 值的情况下,IoU 值越大的预测框质量越好. IoU 具有尺度不变性的特点,不仅可以衡量目标之间的距离,还能表示目标的重叠程度,因此它作为回归损失函数比 MSE 能更好体现边界框的质量.

当目标框和预测框相交时,不同大小的 IoU 值可以很好地体现预测效果的好坏程度,IoU 值越大预测精度越高. 然而当目标框和预测框不相交时,IoU 的值永远是 0,无法体现二者的距离远近,但是通过图 6 可以发现,IoU 为 0 时,其中红色矩

形是目标框,绿色矩形是预测框,两框距离较小的图片预测效果更好,左侧图片预测效果明显优于右侧图片.同时若 IoU 作为损失函数,当它为 0 时,会导致优化损失函数过程中梯度也为 0,从而无法优化.

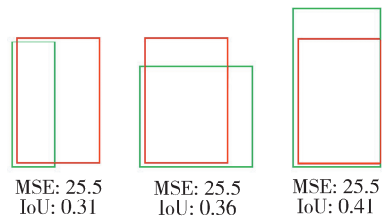


图 5 同一 MSE 对应不同 IoU
Fig. 5 Same MSE corresponds to different IoUs



图 6 不同预测效果(IoU=0)
Fig. 6 Different forecast results (IoU=0)

GIoU^[24]的提出解决了 IoU 不能优化无重叠目标框的问题. IoU 的计算公式如下.

$$IoU=\frac{|A\cap B|}{|A\cup B|}\tag{6}$$

在式(6)中,A 和 B 表示两个区域. GIoU 引入一个正好包含 A 和 B 两个区域的最小区域 C,这样当两者不相交时,它们之间的距离关系可以通过 C 的面积反映,如图 7 所示. 其中,黑色矩形代表 A 区域,绿色矩形代表 B 区域,红色矩形代表 C 区域,灰色阴影部分表示 A 和 B 的相交区域.

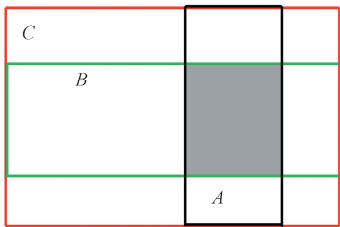


图 7 GIoU 示意图
Fig. 7 GIoU diagram

GIoU 计算公式如下:

$$GIoU=IoU-\frac{|C\backslash(A\cup B)|}{|C|}\tag{7}$$

在式(7)中, $C\backslash(A\cup B)$ 表示 C 的面积减去 $A\cup B$ 的面积. 对于任意两个区域 A 和 B,首先找出一个可以包含它们的最小区域 C,然后通过计算 $C\backslash(A\cup B)$ 面积和 C 面积的比值,最后用 A 与 B 的 IoU 值减去这个比值即可得到 GIoU 的值,此时边界框回归损失函数公式如下.

$$lbox=1-GIoU\tag{8}$$

但文献[25]指出当目标框完全包含预测框时,GIoU 会退化成 IoU 从而不能完整反映出两框的实际情况,如图 8 所示,其中红色矩形是目标框,绿色矩形是预测框. 当 A 完全包含 B 时,C 的面积等于 A 的面积并且不会变化,而 $A\cup B$ 的面积也等于 A 的面积,因此 $C\backslash(A\cup B)$ 的值为 0,GIoU 的值等于 IoU 的值. 从图 8 可知,中间预测效果最优,但 GIoU 无法反映出这种情况下的优劣性.

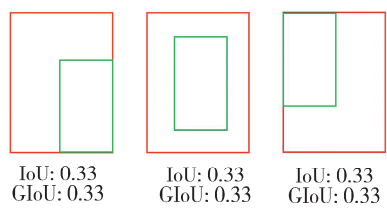


图 8 重叠时 GIoU 与 IoU 的值
Fig. 8 The value of GIoU and IoU when overlapped

本文通过研究交通标志数据集特点,发现数据集中大多数交通标志都会和目标框形成两个共同交点,如图 9 所示. 因为大部分交通标志形状是圆形或者三角形,它们和预测框的共同交点就是框上下边界的中点,即图 9 中箭头指向的点.



图 9 交通标志的共同交点
Fig. 9 Common intersection of traffic signs

因此,本文在 GIoU 基础上把上下边界中点的距离作为考虑因素,把它们距离之和与最小包围框 C 对角线长度的平方的比值作为新的惩罚项,提出一个针对交通标志数据集优化的边界框损失函数 TIoU,它的惩罚项与 TIoU 计算公式如下.

$$P_{TIoU}=\frac{\rho^2(d_1,d_1^g)+\rho^2(d_2,d_2^g)}{c^2}\tag{9}$$

$$TIoU=IoU-P_{TIoU}\tag{10}$$

式(9)中, ρ 代表两个点的欧式距离; d_1 与 d_1^g 代表真

实框和预测框上边界中点; d_2 与 d_2^g 代表真实框和预测框下边界中点; c 代表最小包围框 C 的对角线距离.此时边界框损失函数的公式如下.

$$lbox=1-TIoU$$

(11)

为了证明 TIoU 的有效性,参考 DIoU 论文中的模拟实验,模拟不同距离、尺寸的边界框回归情况.本次实验选择 7 个不同尺度的目标框,并且目标框中点固定在坐标(10,10),一共包含 5000×7×7 个分布均匀的框.然后指定损失函数,见式(12),通过梯度下降来模拟边界框回归过程.

$$B_i^t=B_i^{t-1}+\eta(2-IoU_i^{t-1})\nabla B_i^{t-1}$$

(12)

其中, B_i^t 表示迭代第 t 轮时的预测框; ∇B_i^{t-1} 表示 B_i 在 $t-1$ 轮的损失梯度; η 表示步数,并且通过乘以 $(2-IoU_i^{t-1})$ 来加快收敛.然后使用 L1 范数对边界框回归性能进行评价.本文对 IoU、GIoU、DIoU 以及提出的 TIoU 进行对比实验,当迭代轮次达到 200 轮时停止迭代,此时各损失函数的回归误差曲线如图 10 所示.

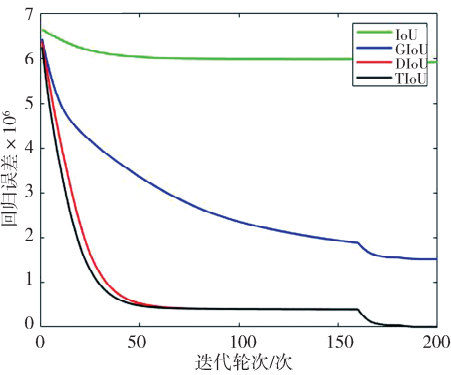


图 10 损失函数回归误差曲线图
Fig. 10 Loss function regression error curve

从图 10 可知,TIoU 在相同迭代轮次下收敛速度,回归效果均明显优于 GIoU,甚至比基于中心点回归的 DIoU 更好,由此可知 TIoU 的有效性.

3.3 k-means++聚类分析

YOLOv3 使用 k-means 算法对 COCO 数据集聚类分析,给出了 9 个固定大小的初始候选框(Anchors).因 COCO 数据集中小目标数量较少,生成的 Anchors 尺寸偏大,会影响小目标交通标志检测的速度和精度,且设置不同的随机种子会得到完全不同的结果,因此本文使用改进后的 k-means++^[26]算法对 TT100K 数据集进行聚类分析,重新生成有利于交通标志检测的初始候选框.

k-means++算法首先随机选取一个初始聚类

中心,然后选择数据中与第一个聚类中心距离最远的 $k-1$ 个点作为剩余的初始聚类中心,如此迭代直至选出 k 个聚类中心. k-means++ 算法相比 YOLOv3 使用的 k-means 算法更符合直觉,即聚类中心离得越远越好.因存在多种方法选取聚类中心个数 k ,为获得符合当前数据集的初始候选框,本文采用平均重叠度(AvgIoU)作为聚类分析度量,选取 AvgIoU 值较大且平稳时的聚类中心个数.平均重叠度的计算公式如下.

$$AvgIoU=\frac{\sum_{i=1}^n\max(IoU(B,C))}{n}$$

(13)

其中, B 为边界盒(groud truth),值为 groud truth 的长和宽; C 为簇的中心,值为 k ; n 为总的边界盒数量.首先求出一个边界盒与 k 个先验框(anchors)的交并比(IoU)值,然后取最大值,依次求出 n 个边界盒的 IoU 值后再求均值,即可得到平均重叠度 AvgIoU. 本文设置 k 值从 1 到 12 分别对 TT100K 数据集进行聚类分析,结果如图 11 所示.

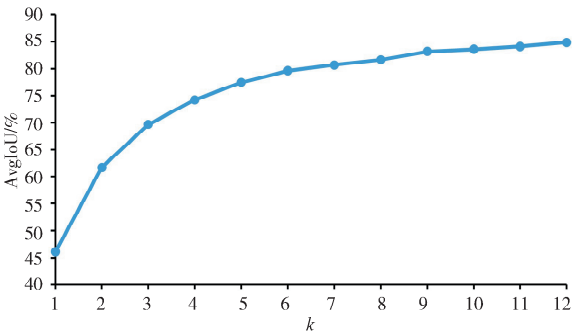


图 11 聚类分析结果
Fig. 11 Clustering analysis

AvgIoU 的值随着 k 增大而逐渐增大,且当 k 趋于 9 时,曲线也随之平稳. YOLOv3 有 3 个尺度的预测层,且聚类中心为 9 时 AvgIoU 的值也较高,因此本文选择 $k=9$,即初始候选框数量也为 9. 通过 k-means++ 聚类分析可以得出输入图片为 416×416 时对应的 9 个候选框尺寸为[4×4, 5×6, 5×11, 8×8, 10×11, 12×22, 15×16, 22×23, 37×37],用新生成的 9 个尺寸更小的初始框代替原来的候选框还可以减少算法计算量.

当聚类中心个数为 9 时, k-means++ 对 TT100K 数据集聚类分析结果如图 12 所示. 其中五角星表示聚类中心,横坐标是初始框的宽度(w),纵坐标是初始框的高度(h),两者都相对图片宽高取偏移量,其值介于 0 到 1 之间.

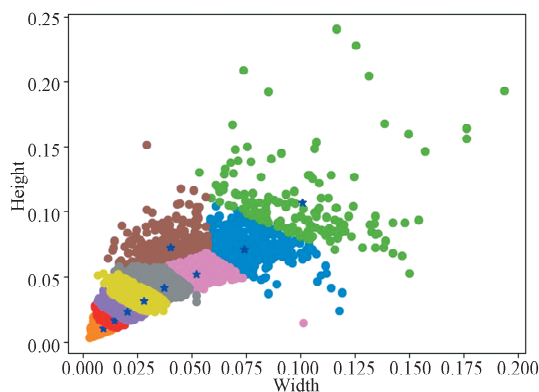


图 12 Anchors 聚类分布

Fig. 12 Anchors clustering distribution

4 实验与分析

4.1 数据集准备

本文使用 TT100K 交通标志数据集, 该数据集提供 10 万张高分辨率图像, 包含 3 万个交通标志实例, 这些图片涵盖了不同光照和天气条件的情形, 如图 13 所示。



图 13 不同照度和天气下的交通标志

Fig. 13 Traffic signs under different illuminance and weather

由于该数据集中各类别实例数量不平衡, 部分类别数据量少不利于网络训练, 且部分交通标志在真实场景中并不常见, 因此参考文献[27]后去掉部分类别, 选择其中的 45 个类别供网络训练和测试, 图 14 显示了其中常见的 20 个类别。该数据集虽然包含高达 10 万张图片, 但是其中大多数图片没有相应标注信息, 或者只含有类别标签用于图像分类任务, 去掉这些无法使用的图片后数据集提供了 9176 张包含完整标注信息的图像进行试验, 其中训练集包含 6105 张图像, 测试集包含 3071 张图像。实验过程中可以通过数据增强手段如翻转, 裁剪等来进一步加大数据量, 保证网络有足够的数据量进行训练。同时使用迁移学习的手段来增强检测器的

泛化能力^[28], 首先, 在 COCO 数据集上训练特征提取网络 DarkNet-53, 然后使用样本较少的交通标志数据集训练预测网络, 微调网络参数, 最终的实验结果也很好地证明近万张图片的数据量可以训练一个优秀的目标检测器。

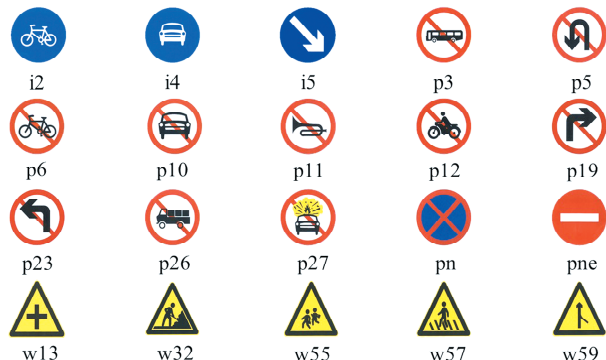


图 14 部分交通标志类别

Fig. 14 Some traffic sign categories

4.2 实验参数与评价指标

4.2.1 实验参数 实验使用一张显存大小是 11 GB 的 NVIDIA Geforce 1080Ti 显卡训练和测试所有模型。在训练过程中, 因显存大小限制将 batch size 设置为 8, 权值衰减 (Decay) 设置为 0.0005, 动量衰减 (Momentum) 设置为 0.9, 初始学习率 (Learning rate) 设置为 0.1, 最大迭代次数设置成 300 轮。训练初期较大的学习率可以加速收敛, 但是训练后期学习率过大会导致结果越过最优值无法拟合, 因此每隔 75 轮当前学习率在前一个学习率的基础上衰减 10 倍。

4.2.2 评价指标 第一个评价指标是准确率-召回率 (P-R) 曲线, 以准确率 (Precision) 作为纵坐标轴, 以召回率 (Recall) 作为横坐标轴, 计算式如下。

$$Precision = \frac{TP}{(TP + FP)} = \frac{TP}{Detections} \quad (14)$$

$$Recall = \frac{TP}{(TP + FN)} = \frac{TP}{GroundTruths} \quad (15)$$

其中, $Detections$ 表示检测框总数; $Ground Truths$ 表示所有真实框 (GT) 的数量; TP (True Positive) 表示 IoU 大于特定阈值 (本文阈值为 0.5) 的检测框数量, 且同一个 GT 只计算一次; FP (False Positive) 表示 IoU 小于阈值或者检测到同一个 GT 的多余检测框数量; FN (False Negative) 表示没有检测到的 GT 数量。通过置信度对所有样本降序排列, 再逐个将样本作为正例计算当前 P 、 R 值即可绘制出 P - R 曲线。

第二个评价指标是平均检测精度 (mAP), 计

算每个类别的 AP 值再求平均就能得到 mAP 的值,每个类别的 AP 值则是当前类别对应的 P-R 曲线包含的面积.第三个指标是检测一张图片的耗时(ms/张),值越小越好.

4.3 实验结果分析

4.3.1 实验前置条件 本文首先对原始 YOLOv3 算法使用 k-means 根据 COCO 数据集生成的候选框和使用 k-means++ 针对 TT100K 数据集生成的候选框分别训练和测试,图片输入尺寸为 608×608,实验结果见表 1.

表 1 聚类算法对比实验结果

Tab. 1 Experimental results of clustering algorithm

模型	聚类算法	mAP	耗时/(ms/张)
YOLOv3	k-means	0.481	22.6
	k-means++	0.709	20.6

从表 1 可知,YOLOv3 使用原始候选框时交通标志检测效果并不理想,mAP 只有 48.1%,精度过低无法满足真实应用要求,容易检测错误或者漏检导致发生安全事故.而使用 k-means++ 针对交通标志数据集生成的尺寸更小的候选框,在 YOLOv3 算法不做任何改动的情况下,mAP 可以达到 70.9%,预测精度大大提高,能满足交通标志检测任务要求.同时由于新初始框的尺寸更小,可以更好地覆盖数据集中的小目标,有效消除了初始框带来的误差并降低了模型计算量,因此在处理每张图片耗时上也减少了 2.0 ms,用时低至 20.6 ms,充分满足实时检测需求.

为更好地检测小目标,后面所有实验均使用新初始框,同时将使用新框的检测结果作为原 YOLOv3 的实验结果与其他模型进行对比实验和分析.

4.3.2 对比实验结果与分析 在相同实验条件下,对原 YOLOv3 算法,损失函数为 GIoU 的 YOLOv3 算法(YOLOv3-G),损失函数为 TIoU 的 YOLOv3 算法(YOLOv3-T),在网络结构中加入 SPP 模块的 YOLOv3 算法(YOLOv3-S),在网络结构中加入 FI 模块的 YOLOv3 算法(YOLOv3-F)以及本文提出的算法(Ours)分别进行训练.在输入图片尺寸为 608×608 的条件下,分别对 6 个模型进行测试,实验结果见表 2.

由实验结果可知,YOLOv3-G 模型 mAP 提升 1.4%且依然保持较快的检测速度,其图片检测耗时只有 20.8 ms. YOLOv3-T 模型 mAP 提升明显

达到 74.1%,但因为新增上下边界中点的距离增大了计算量,耗时比 GIoU 作为损失函数时增加 0.2 ms,微小计算量和耗时的增加并不影响实时检测效果且检测精度提升明显,因此表明 TIoU 相对 GIoU 在加快收敛的同时也能更好地表示目标框与预测框的重叠情况,有利于提高目标检测精度. YOLOv3-S 模型的 mAP 值是 74.5%,可以看出,加入空间金字塔池化模块能解决多尺度预测的信息丢失和尺度不统一问题,但 3 个尺度的分块池化操作增加了计算量,处理每张图片耗时也提高到了 26.3 ms. 而 FI 模块通过下采样大特征图和上采样小特征图再进行 Concat 结合,使 3 个特征度均包含有效的小目标信息,让原特征图中可能损失的信息从另外两个尺度提供的辅助信息中得到补充,从而极大提高模型预测精度,因此 YOLOv3-F 模型的预测精度高达 76.2%. 但是多次上下采样操作也进一步增大了计算量,耗时继续增加达到 26.9 ms. 最后,融合所有改进点的 Ours 模型的 mAP 取得最大值 82.0%,相比原始模型提高 11.1%,并且每张图片耗时仅为 27.2 ms,符合交通标志检测任务高精度、实时性的要求.

表 2 对比实验结果

Tab. 2 Comparative experimental results

模型	mAP	耗时/(ms/张)
YOLOv3	0.709	20.6
YOLOv3-G	0.723	20.8
YOLOv3-T	0.741	21.0
YOLOv3-S	0.745	26.3
YOLOv3-F	0.762	26.9
Ours	0.82	27.2

由于现实中部分交通标志图像尺寸较小可能导致检测器精度降低,因此修改输入图像尺寸为 416×416 后再次进行测试以验证改进算法的有效性,实验结果如表 3 所示.

表 3 小尺寸实验结果

Tab. 3 Small scale experimental results

尺寸	模型	mAP	耗时/(ms/张)
416×416	YOLOv3	0.655	12.3
	Ours	0.723	15.6
608×608	YOLOv3	0.709	20.6
	Ours	0.82	27.2

从表 3 可知,当输入图片尺寸缩小到 416×416 后,原模型和改进的 YOLOv3 模型精度都有所降

低,原算法精度降低 5.4%,而改进算法降低 9.7%后的 mAP 值仍高于原算法在 608×608 时的 mAP 值,具有较高的检测精度,符合交通标志检测要求;且因为输入尺寸减小导致计算量降低,模型检测速度均提升明显,符合实时检测需求.

在此基础上进一步验证改进算法有效性,本文选择与两阶段目标检测算法中的经典算法 Faster R-CNN,单阶段目标检测算法中的经典算法 DSSD^[29]以及刘紫燕等人提出的算法^[30]进行对比实验,结果如表 4 所示.

表 4 不同算法对比实验

Tab. 4 Comparative experiments of different algorithms

尺寸	算法	mAP	耗时/(ms/张)
608×608	YOLOv3	0.709	20.6
	Faster R-CNN	0.715	90.0
	DSSD	0.692	61.7
	文献[30]	0.752	31.9
	Ours	0.82	27.2

由实验结果可知,追求高精度的两阶段算法 Faster R-CNN 在精度上优于 YOLOv3 算法,但是它处理每张图片耗时长达 90 ms,不满足实时检测要求.追求速度的单阶段算法 DSSD 在速度上比 Faster R-CNN 快 28.3 ms,且精度达到了 69.2%,可以完成交通标志检测任务.文献[30]提出的改进算法精度有很大提升达到 75.2%,但是因为引入一个尺度为 108×108 的预测层极大增加了算法计算量,导致耗时增加到了 31.9 ms,依然落后于本文改进算法,实验结果证明了本文算法的有效性.

4.3.3 改进模型实验结果分析 为了进一步证明改进算法效果,从 45 类数据中任意取 3 个类分析改进算法相对原算法的提升之处.固定输入尺寸为

608×608,io 类、pn 类、p12 类以及所有类别的准确率、召回率和 mAP 见表 5.

表 5 不同类别实验结果

Tab. 5 Different categories' results

类别	模型	准确率(P)	召回率(R)	mAP
io	YOLOv3	0.505	0.711	0.658
	Ours	0.553	0.899	0.846
pn	YOLOv3	0.846	0.851	0.881
	Ours	0.55	0.915	0.888
p12	YOLOv3	0.262	0.682	0.35
	Ours	0.407	0.848	0.718
All	YOLOv3	0.445	0.824	0.709
	Ours	0.561	0.867	0.82

从表 5 可知,io 类准确率提升较小,只有 4.8%,p12 类准确率提升 14.5%,而 pn 类准确率不升反降只有 55%.因此表明在网络模型中加入 FI 模块融合特征图信息有效提高了大部分类别的预测效果,如 io 类、p12 类;但对其中部分类别如 pn 类,可能在特征图上下采样和 Concat 结合过程中损失了部分信息,从而造成准确率下降,这是本文算法尚未考虑到的部分,也是未来工作中的改进方向.但是 3 个类别的召回率均有提升,分别为 18.8%、6.4%、16.6%,这表明使用 k-means++ 聚类算法生成的候选框和优化边界框损失函数为 TIoU 后的改进算法比原算法能尽可能多地找全对应类别的所有目标,提高召回率从而提升算法的总体目标检测精度,其中 3 个类别的 P-R 曲线如图 15 所示.表 5 数据显示 3 个类别的 mAP 都分别提高了 18.8%、0.7%、36.8%,并且所有类别(All)的平均识别率、召回率和 mAP 都有不同程度的提升,很好地证明了改进算法的有效性.

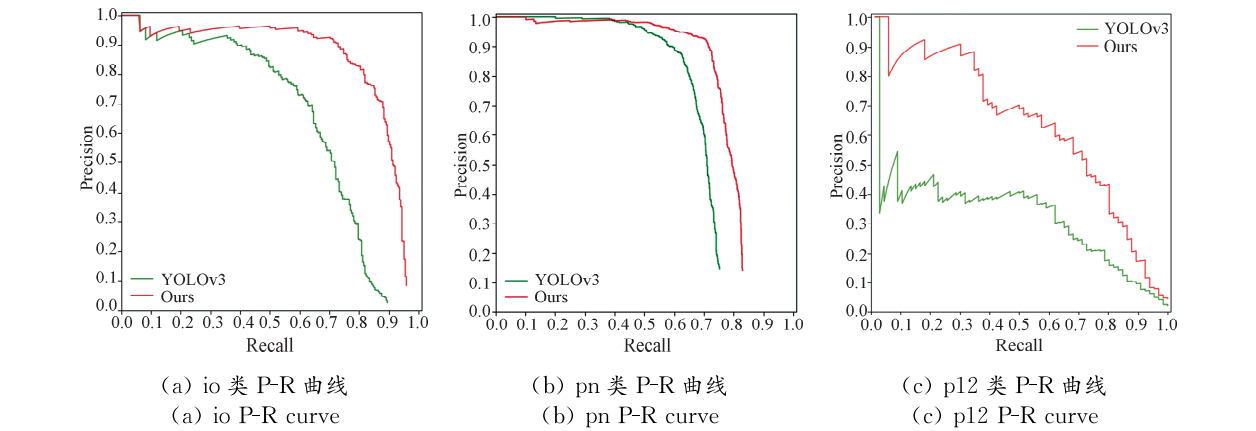


图 15 不同类别的 P-R 曲线
Fig. 15 P-R curves with different categories

4.3.4 图像检测效果对比 从交通标志数据集中随机选择 3 张图片分别用 YOLOv3 和改进算法进行检测,检测结果如图 16 所示。

图 16a 表示 YOLOv3 在正常天气条件下正确检测出较大目标,但漏检了一个近处的 pl40 类小目标,而改进算法正确检测出了图片中所有目标;



(a) YOLOv3vs Ours



(b) YOLOv3vs Ours



(c) YOLOv3vs Ours

图 16 检测效果对比图

Fig. 16 Comparison pictures of detection effect

5 结 论

本文基于 YOLOv3 提出一种应用于交通标志检测任务中的改进算法,该算法在速度几乎不变的前提下极大提高了检测精度.在网络结构中引入 SPP 模块解决尺度不统一和信息丢失问题,添加 FI 模块进行特征融合,增加预测特征图中的小目标信息,从而提高检测精度;使用 TIoU 作为边界框损失函数替换 MSE 损失函数,不仅能反映目标框和预测框的重叠程度,并且以二者上边框的中点距离作为新的惩罚项加入损失函数中进行优化,加快函数收敛速度,解决训练时梯度为 0 无法优化等问题.最后,使用 k-means++ 算法对数据集进行聚类分析,针对交通标志检测任务中小目标数量较多的情况,重新生成初始候选框,减少计算量和大候选框带来的误差。

本文算法在降低较少速度的前提下,有效提高了 YOLOv3 在交通标志检测中的准确率、召回率和平均精确度,充分满足任务需求.解决融合过程中信息丢失造成部分类别精度下降的问题以及进一步加快检测速度是未来的研究方向。

参考文献:

- [1] 黄健,张钢. 深度卷积神经网络的目标检测算法综述[J]. 计算机工程与应用, 2020, 56: 12.
- [2] 刘洋,战荫伟. 基于深度学习的小目标检测算法综述[J]. 计算机工程与应用, 2021, 51: 37.
- [3] Liu T Y, Marie M, Balgonie S, *et al.* Microsoft coco: common objects in context[C]//Proceedings of

图 16b 表示改进算法能在强光照条件下比原始算法能更好地检测出所有目标;图 16c 表示当两个算法均正确检测出图片中所有目标分类时,改进算法预测分数明显优于原始算法. 综上,实际检测结果表明改进算法比 YOLOv3 更能胜任交通标志的检测任务。

the European Conference on Computer Vision, Switzerland: Springer Cham, 2014.

- [4] 吴华运,任德均,吕义钊,等. 基于改进的 RetinaNet 医药空瓶表面起泡检测[J]. 四川大学学报: 自然科学版, 2020, 57: 1090.
- [5] Everingham M, Van Gool L, Williams C K I, *et al.* The pascal visual object classes(voc) challenge [J]. Int J Comput Vision, 2010, 88: 303.
- [6] Everingham M, Eslami S M A, Van Gool L, *et al.* The pascal visual object classes challenge: A retrospective [J]. Int J Comput Vision, 2015, 111: 98.
- [7] Russakovsky O, Deng J, Su H, *et al.* Imagenet large scale visual recognition challenge[J]. Int J Comput Vision, 2015, 115: 211.
- [8] Kuznetsova H, Rom, Alldrin N, *et al.* The open images dataset V4: unified image classification, object detection, and visual relationship detection at scale [J]. Int J Comput Vision, 2020, 128: 1956.
- [9] 赵亚男,吴黎明,陈琦. 基于多尺度融合 SSD 的小目标检测算法[J]. 计算机工程, 2020, 46: 247.
- [10] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, America: IEEE, 2014.
- [11] He K, Zhang X, Ren S, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE T Pattern Anal, 2014, 37: 1904.
- [12] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015.
- [13] Ren S, He K, Girshick R, *et al.* Faster R-CNN;

- towards real-time object detection with region proposal networks[J]. IEEE T Pattern Anal, 2017, 39: 1137.
- [14] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks [C]//Advances in neural information processing systems. San Francisco, America: Morgan Kaufmann Publishers, 2016.
- [15] 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. Las Vegas, America: IEEE, 2016.
- [16] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector [C]//European conference on computer vision. Switzerland: Springer Cham, 2016.
- [17] Zitnick C L, Dollar P. Edge boxes: locating object proposals from edges [C]//Proceedings of the European Conference on Computer Vision. Switzerland: Springer Cham, 2014.
- [18] Uijlings J R, Gevers T, Sande K E A, *et al.* Selective search for object recognition [J]. Int J Comput Vision, 2013, 104: 154.
- [19] Zhang Y, Sheng Y L, Zhang J. An improved tiny-yolov3 pedestrian detection algorithm [J]. Optik, 2019, 183: 17.
- [20] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii, America: IEEE, 2017.
- [21] Lin T Y, Dollar P, Girshick R, *et al.* Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii, America: IEEE, 2017.
- [22] Ioffe S, Szegedy C. Batchnormalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the International conference on machine learning. Lille, France: PMLR, 2015.
- [23] Li L, Yang Z, Jiao L. High-resolution SAR change detection based on ROI and SPP net [J]. IEEE Access, 2019, 7: 177009.
- [24] Rezatofighi H, Tsoi N, Gwak J Y, *et al.* Generalized Intersection over union: a metric and a loss for bounding box regression [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S. l.]: IEEE, 2019: 658.
- [25] Zheng Z H, Wang P, Liu W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression [C]//Association for the Advance of Artificial Intelligence. New York, America: AAAI, 2020.
- [26] Hamalainen J, Karkkainen T, Rossi T. Improving scalable K-means + + [J]. Algorithms, 2021, 14: 6.
- [27] 张传伟, 李姐姐, 岳向阳, 等. 基于改进 YOLOv2 算法的交通标志检测[J]. 计算机系统应用, 2020, 29: 155.
- [28] 胡恒豪, 王俊峰, 方智阳, 等. 基于深层迁移学习的 DR 胸片肺结核病灶检测[J]. 四川大学学报: 自然科学版, 2020, 57: 459.
- [29] Wang S, Wu L, Wu W. Optical fiber defect detection method based on DSSD network [C]//Proceedings of the 2019 IEEE International Conference on Smart Internet of Things. Tianjin China: IEEE, 2019.
- [30] 刘紫燕, 袁磊, 朱明成, 等. 融合 SPP 和改进 FPN 的 YOLOv3 交通标志检测[J]. 计算机工程与应用, 2021, 57: 164.

引用本文格式:

中 文: 王卜, 何扬. 基于改进 YOLOv3 的交通标志检测[J]. 四川大学学报: 自然科学版, 2022, 59: 012004.

英 文: Wang B, He Y. Traffic sign detection based on improved YOLOv3 [J]. J Sichuan Univ: Nat Sci Ed, 2022, 59: 012004.