



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 改进 YOLOv3 的交通标志检测方法研究
作者: 邓天民, 周臻浩, 方芳, 王琳
网络首发日期: 2020-07-14
引用格式: 邓天民, 周臻浩, 方芳, 王琳. 改进 YOLOv3 的交通标志检测方法研究. 计算机工程与应用.
<https://kns.cnki.net/kcms/detail/11.2127.TP.20200713.1716.022.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

改进 YOLOv3 的交通标志检测方法研究

邓天民, 周臻浩, 方 芳, 王 琳

重庆交通大学 交通运输学院, 重庆 400074

摘 要: 针对我国自动驾驶的辅助识别交通标志误差率大、检测速度慢、需人工参与等问题, 提出一种基于改进 YOLOv3 的交通标志检测识别方法。首先通过改进 Darknet53 网络结构来减少网络迭代过程中前向推理计算, 提升网络迭代速度。其次, 引入目标检测的直接评价指标 GIoU 指导定位任务来提高检测精度。最后使用 k-means++ 聚类算法获取 anchor 尺寸并匹配到对应的特征层。实验结果表明, 提出的方法相较于原始 YOLOv3 在标准数据集 Lisa 上的平均精度提升了 8%, 检测速度达到了 76.9f/s; 在自制数据集 CQ-data 上平均精度可达 94.8%, 与传统识别以及其他算法相比, 不仅具有更好的实时性、准确性, 对各种环境变化具有更好的鲁棒性, 而且可以识别多种交通标志的类型。

关键词: 交通标志检测; YOLOv3; GIoU; 维度聚类

文献标志码: A 中图分类号: TP394.41 doi: 10.3778/j.issn.1002-8331.2004-0043

邓天民, 周臻浩, 方芳, 等. 改进 YOLOv3 的交通标志检测方法研究. 计算机工程与应用

DENG Tianmin, ZHOU Zhenhao, FANG Fang, et al. Research on Improved YOLOv3 traffic sign detection method. Computer Engineering and Applications

Research on Improved YOLOv3 traffic sign detection method

DENG Tianmin, ZHOU Zhenhao, FANG Fang, WANG Lin

College of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China

Abstract: Aiming at the problems of the large error rate, slow detection speed, and manual participation required for assisted identification of traffic signs in autonomous driving in China, an improved YOLOv3 traffic sign detection and recognition method is proposed. Firstly, the network structure of Darknet53 is improved to reduce the forward inference calculation during network iteration and increase the network iteration speed. Secondly, the direct evaluation index GIoU of target detection is introduced to guide the positioning task to improve the detection accuracy. Finally, the k-means ++ clustering algorithm is used to obtain the anchor size and match the corresponding feature layer. The experimental results show that compared with the original YOLOv3, the proposed method has an average accuracy improvement of 8% on the standard data set Lisa, and the detection speed reaches 76.9f / s. The average accuracy on the homemade data set CQ-data can reach 94.8%. Compared with traditional recognition and other algorithms, it not only has better real-time performance and accuracy, it is more robust to various environmental changes, but also can recognize multiple types of traffic signs.

Key words: traffic sign detection; YOLOv3; GIoU; dimension clustering

基金项目: 国家自然科学基金 (No.51678099); 重庆市科学技术委员会科技人才培养计划项目(No.CSTC2013KJRC-QNRC0148)。

作者简介: 邓天民 (1979-), 男, 副教授, 博士, 主要研究方向: 交通大数据、自动驾驶、交通控制, E-mail: dtianmin@cqjtu.edu.cn;

周臻浩 (1995-), 男, 硕士研究生, 主要研究方向: 人工智能、交通信息与控制。

1 引言

交通标志检测与识别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中,随着计算机性能、人工智能和目标检测技术快速的发展,端到端的智能化检测是自动驾驶领域的一个全新的途径。因此,本文研究基于 YOLOv3 的交通标志牌检测与识别方法对于提高自动驾驶公路交通规范具有重要意义。

传统机器学习应用在交通标志检测的方法主要是基于颜色、形状或者两者相结合的方法。Li 等^[1]将图像从 RGB 颜色空间转换到高斯颜色空间,然后使用 K-means 对颜色进行聚类并提取出交通标志;Kaplan 等^[2]提出使用形状检测算法来检测圆形交通标志。这些方法主要是利用交通标志颜色与形状的特殊特性进行特征的提取,通过依据分类器来实现分类。但是存在检测速度慢、检测精度差、抗环境干扰能力差等问题,难以在实际应用中实现精准的识别效果。

近年来,随着深度学习在图像分类、目标检测、图像语义分割等领域取得了一系列突破性的研究成果。John 等^[3]使用 CNN 提取特征并检测道路交通标志,绘制出包含交通灯位置的特征图。Girshick 等^[4]提出了一种基于快速区域的卷积网络方法(Fast R-CNN),使用深度卷积网络有效地对目标进行分类。为了执行快速准确的交通标志检测和识别,Zhu 等^[5]采用了整体嵌套的边缘检测网络。Tian 等^[6]提出了采用多尺度 CNN 方法,融合了不同网络层的特征信息来识别交通标志,并将检测问题看作回归问题,使用深度卷积神经网络作为回归预测其输入图像中目标。总体而言,这些方法中使用的主要方法是首先提取感兴趣区域的特征,然后使用分类器定位位置。因此,这些方法统称为两阶段目标检测。它们具有较高的检测精度和较强的泛化能力,但是确定感兴趣区域需要大量的计算,这就导致实现实时检测的能力很差。

为了提高检测效率,Redmon 等^[7]提出了将目标检测作为回归问题的 YOLOv1 (You Only Look Once) 网络,该网络是一个采用深度卷积神经网络

的全新实时目标检测方法,它直接利用网格对目标位置进行回归。与当前的卷积神经网络相比,YOLO 网络显著提高了检测速度,并在很多领域得到了广泛的应用。与 R-CNN 相比,YOLOv1 网络以较低的准确性和较低的召回率来定位目标物体。为了进一步提高检测精度,YOLOv3^[8]使用了增量改进技术,大大提高了小目标的检测精度。

综上,本文针对交通标志易受光照、实际道路场景复杂等特征影响导致识别准确率和识别速率不高的问题,采用了 Darknet 框架结合改进 YOLOv3 算法的交通标志检测方法。最后,通过大量实验对比改进算法和原始算法在复杂交通标志情况下的检测识别率和鲁棒性的差异,验证了该算法的有效实用性。

2 交通标志检测模型

网络整体框架选用整图训练的方式将检测问题看做了回归问题,可以更好地区分目标和背景区域;并且接受不同大小尺寸输入的图片,由 Darknet 网络提取图像的特征,然后将特征图输入到后续的多尺度分类检测网络中,最后 YOLOv3 输出 3 个不同尺度的特征,可以对不同大小的目标进行检测,有助于实现交通标志目标的检测要求。

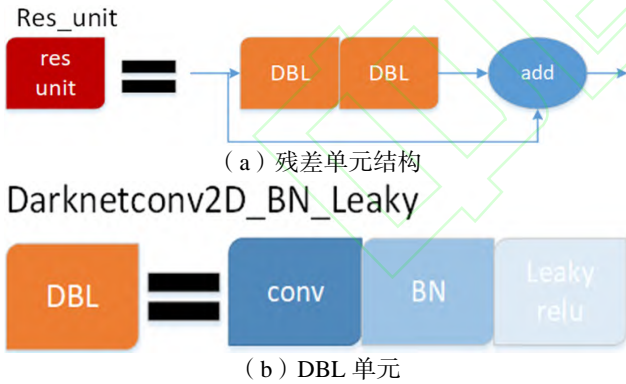
2.1 Darknet-53 特征提取网络结构

在交通标志图像特征提取方面,YOLOv2^[9]采用了 Darknet-19 网络结构,由于直筒型网络结构层数过多会造成训练过程中梯度爆炸等问题,因此 YOLOv2 所采用的 Darknet 网络架构只有 19 层。YOLOv3 采用了称之为 Darknet-53 的网络结构(含有 53 个卷积层)结合了 YOLOv2 的基础特征提取器并借鉴了残差网络的思想^[10],如图 1 所示。

Darknet-53 由 5 个残差块组成,每个残差块由残差单元组成,残差单元中由 DBL 单元组成,两个 DBL 单元在残差块中进行残差操作,如图 2(a)所示。其中 DBL 单元包含了卷积层(convolutional)、批归一化(Batch Normalization)和 leaky Relu 激活函数,如图 2(b)所示,通过引入残差结构,可以增加网络架构的深度并防止训练的过程梯度爆炸和消失等问题。

	类型	滤波器	尺寸	输出
1×	卷积层	32	3×3	416×416
	卷积层	64	3×3/2	208×208
	卷积层	32	1×1	
	卷积层	64	3×3	
	残差层			208×208
2×	卷积层	128	3×3/2	104×104
	卷积层	64	1×1	
	卷积层	128	3×3	
	残差层			104×104
8×	卷积层	256	3×3/2	52×52
	卷积层	128	1×1	
	卷积层	256	3×3	
	残差层			52×52
8×	卷积层	512	3×3/2	26×26
	卷积层	256	1×1	
	卷积层	512	3×3	
	残差层			26×26
4×	卷积层	1024	3×3/2	13×13
	卷积层	512	1×1	
	卷积层	1024	3×3	
	残差层			13×13
平均池化			Global	
连接数			1000	
Softmax				

图1 Darknet-53 网络结构



(b) DBL 单元

图2 结构单元

本文对 Darknet53 网络进行了模型简化。YOLO 网络中引入了 BN^[11] (Batch Normalization) 层, 通过对每个卷积层后添加 BN 层, 可以在网络训练时期加速网络收敛以及防止过拟合, 并且可以简化调参, 使网络更加稳定。虽然 BN 层在训练时起到了归一化作用, 但是在网络前向推理时增加了运算, 影响了模型的性能, 且占用了更多的内存或者显存空间, 从而降低了网络模型的前向推理速度。

卷积层与 BN 层都是线性变换, 提供了合并的可能性。在 YOLOv3 中, BN 计算过程如下:

$$x_{out} = \frac{\gamma(x_{conv} - \mu)}{\sqrt{\delta^2 + 0.0000001}} + \beta \quad (1)$$

$$x_{conv} = \sum_{i=0}^n (x_i \cdot w_i) \quad (2)$$

其中 γ 为缩放因子, μ 为均值, δ^2 为方差, β 为偏置, x_{out} 为 BN 计算结果, x_{conv} 为卷积计算结果。

合并卷积层与 BN 层, 式 (2) 代入到式 (1) 得:

$$x_{out} = \frac{\gamma(\sum_{i=0}^n (x_i * w_i) - \mu)}{\sqrt{\delta^2 + 0.0000001}} + \beta, \text{ 即}$$

$$x_{out} = \sum_{i=0}^n (x_i * \frac{\gamma * w_i}{\sqrt{\delta^2 + 0.0000001}}) - \frac{\gamma * \mu}{\sqrt{\delta^2 + 0.0000001}} + \beta$$

合并后权值参数变为:

$$w_i' = \frac{\gamma * w_i}{\sqrt{\delta^2 + 0.0000001}}$$

$$\text{偏置变为: } \beta' = \beta - \frac{\gamma * \mu}{\sqrt{\delta^2 + 0.0000001}}$$

$$\text{即合并后的计算变为: } x_{out} = \sum_{i=0}^n (x_i * w_i') + \beta' \quad (3)$$

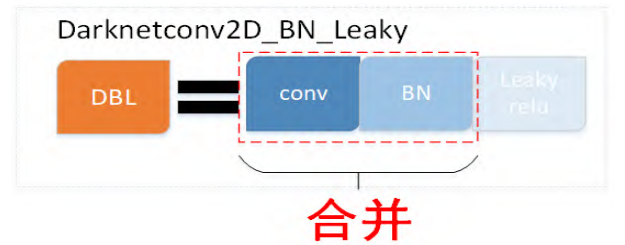


图3 合并 BN 层和 CONV 层

卷积层与 BN 层合并后可以与卷积层共用 Blob 数据, 从而可以减少内存占用, 有利于速度提升。

如上图 3 所示, 输入图片经过卷积得到特征图后, 然后输入到 BN 层中进行归一化处理, 再经过 Leaky-Relu 函数激活之后, 进入网络的下一层。本文将 DBL 单元结构中的批归一化层和卷积层合并为一层, 通过实验可知该方法提高了模型的前向推理速度, 从而可以提高目标检测的速率。

2.2 基于 YOLOv3 的交通标志检测模型

YOLOv2 中引入一种称为 passthrough layer 的方法在特征图中保留一些细节信息，为了加强 YOLO 算法对小目标检测的精度。在 YOLOv3 中采用类似 FPN 的上采样 (upsample) [12] 和尺度融合方法 [13]，将上采样后的高语义特征与浅层的定位细节特征进行融合。如图 4 所示，卷积网络在 79 层后，经过下方几个卷积层得到一种尺度的检测结果。相比输入图像，这里用于检测的特征图有 32 倍的下采样，特征图就是 13×13 了。由于下采样倍数高，这里特征图的感受野比较大，因此适合检测图像中尺寸较大的目标。为了实现细粒度的检测，第 79 层的特征图又开始作上采样，然后与第 61 层特征图融合，这样得到第 91 层较细粒度的特征图，同

样经过几个卷积层后得到相对输入图像 16 倍下采样的特征图。最后，第 91 层特征图再次上采样，并与第 36 层特征图融合，最后得到相对输入图像 8 倍下采样的特征图。它的感受野最小，适合检测小尺寸的目标。此方法融合了 3 个尺度 (13×13 、 26×26 、 52×52)，通道数为 3，就是说每个 box 负责对三个 anchor box 进行回归，取其中的一个作为最终检测结果，共对 9 个 anchor box 进行回归，所以对于一张输入图像，最后的输出尺寸为 $1 * (3 * (13 * 13 + 26 * 26 + 52 * 52)) * (5 + m)$ ，m 代表检测类别数。在多个尺度的融合特征图上分别做检测，最终对小目标的检测效果较前几代的 YOLO 有了明显的提升。

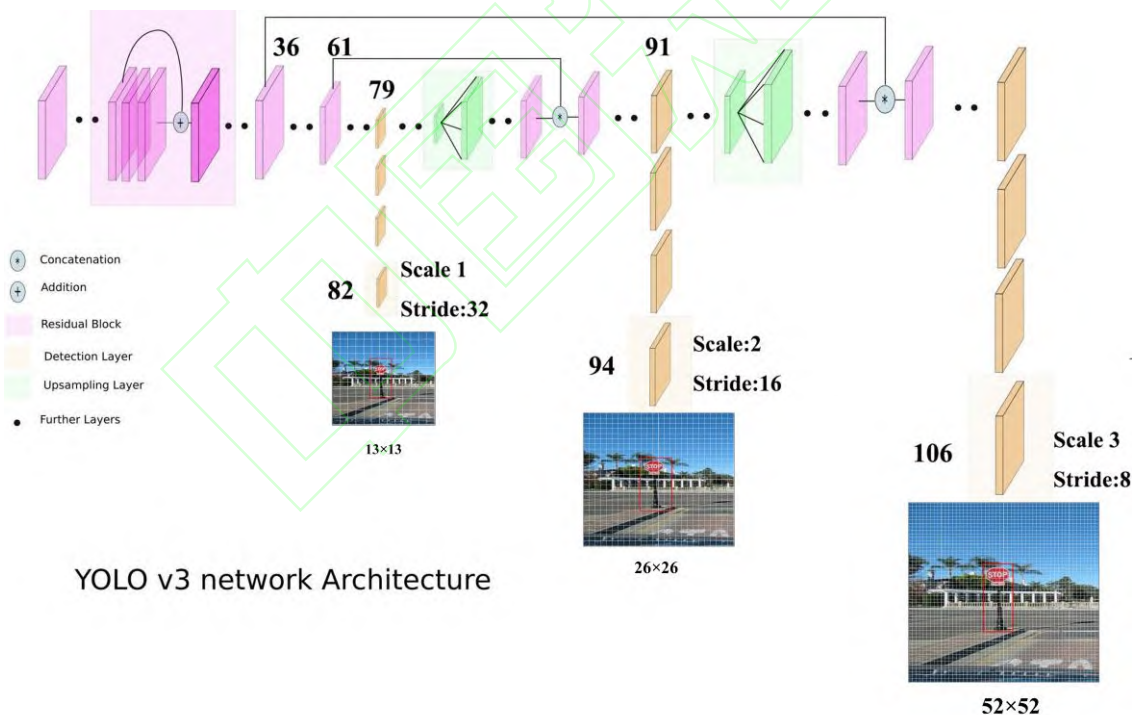


图 4 YOLOv3 多尺度融合

2.2.1 目标边界框的预测

YOLOv3 的设计理念遵循端到端的训练和实时检测，可将图片作为输入，直接输出目标位置和对应该位置的置信度评分。首先通过特征提取网络对输入图像提取特征，得到一定大小的特征图，然后将输入图像分成 $s \times s$ 个单元格，如果真实框中某个目

标的中心坐标落在哪个单元格中，那么就由该单元格来预测该目标。每个单元格都会预测 3 个固定数量的边界框。每一个边界框获得 1 个偏移坐标 (t_x, t_y)、1 个真实宽高坐标 (t_w, t_h) 和一个置信度评分 C 。

在 YOLOv1 版本中, x, y, w, h 是直接预测物体框的中心坐标和预测框的宽度和高度，预测值的微小

变化都会被放大到整个图像的任何范围,导致坐标波动较大,预测不准确。YOLOv2 和 YOLOv3 对其进行了改进,其公式为:

$$b_x = \sigma(t_x) + c_x \quad (4)$$

$$b_y = \sigma(t_y) + c_y \quad (5)$$

$$b_w = p_w e^{t_w} \quad (6)$$

$$b_h = p_h e^{t_h} \quad (7)$$

其中, t_x, t_y 为预测的坐标偏移值, $\sigma(t_x), \sigma(t_y)$ 用 sigmoid 运算将 t_x, t_y 压缩到 $[0, 1]$ 区间内,可以有效确保目标中心处于执行预测的网格单元中,防止偏移过多。 t_w, t_h 为尺度缩放。 c_x, c_y 为 grid cell 坐标,也就是距离左上角顶点的偏移量; p_w, p_h 代表 cell 对应先验框 (anchor box) 的宽高,计算出 bounding box 的位置,如图 5 所示。 $confidence = pr(Object) \times IoU_{pred}^{truth}$, 通过对 confidence 这一项设定阈值,过滤掉低分的预测框,然后对剩下的预测框执行非极大抑值 (Non Maximum Suppression, NMS)^[14] 处理,得到网络的最终预测。

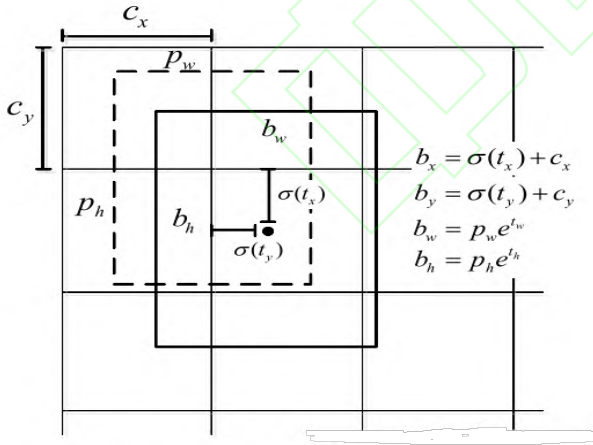


图 5 bounding box 坐标转换公式

2.2.2 目标定位损失函数的改进

当前很多目标检测算法的坐标误差损失函数都使用预测框与真实框角点坐标距离的 1-范数 Loss 或者 2-范数 Loss,但是评测的时候却又采用 IoU (Intersection over Union,)来判断是否检测到目标。显然二者有一个分歧,即 loss 低不代表 IoU 就一定小。由下图 6 可知,其中红色框为真实框,黑色框为

预测框。第一排,所有物体的 2-范数 loss 都一样,但是第三个 IoU 显然要高于第一个,主观来看也是第三个检测结果要好。第二排也是类似,1-范数都一样,但是 IoU 却不同。

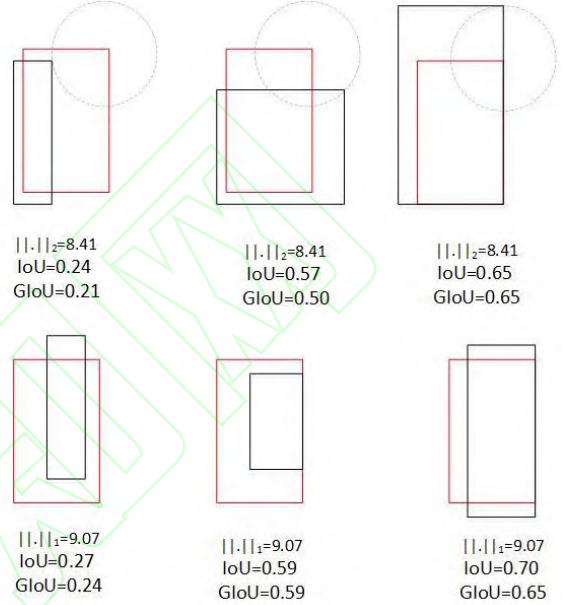
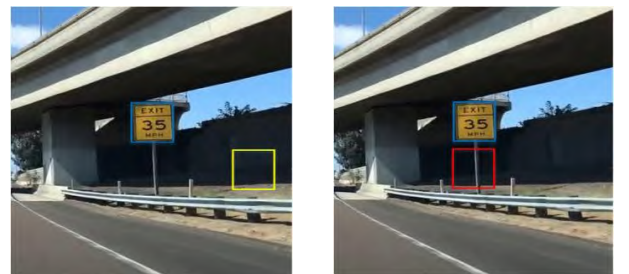


图 6 相同 1-范数或 2-范数对应的 IoU 和 GIoU

当 L1 或 L2 范数都相同的时候,发现 IoU 和 GIoU 的值差别都很大,这表明使用 L 范数来度量边界框的距离是不合适的。针对这个问题会直接采用 IoU 当作模型的坐标误差损失函数,但是 IoU 作为损失函数存在两个问题: 1. 当预测框和真实框之间没有重合时,不管两者之间的距离有多远对应的 IoU 值都为 0,导致优化损失函数时梯度也为 0,意味着无法优化,无法进行学习训练,见图 7。2. IoU 无法辨别不同方式的对齐,比如方向不一致等。如下图,场景 A 和场景 B 的 IoU 值都为 0,但是显然场景 B 的预测效果较 A 更佳,因为两个边界框的距离更近(L 范数更小)。



场景 A

场景 B

图 7 IoU 为 0 的两种不同情况的回归

IoU 是目标检测中使用最广的评价指标,然而优化回归 bounding box 参数的距离损失并不等价于最大化 IoU 指标。对于轴对齐的 2D bounding box, IoU 可直接用作坐标误差回归损失,但是 IoU 无法优化无重叠的 bounding box,所以本文使用一种泛化版的 GIoU^[15] (Generalized Intersection over Union) 来代替 IoU 作为坐标误差损失函数,该改进方法加快模型拟合预测框与真实框的位置关系,并提高模型的预测精度。

对于预测框 A 与真实框 B,可以计算出两者的最小凸集(包含 A、B 的最小包围框) C。通过最小凸集,可以计算 GIoU。IoU 和 GIoU 的计算公式如下:

$$IoU = \frac{A \cap B}{A \cup B} \quad (8)$$

$$GIoU = IoU - \frac{|C \setminus (A \cup B)|}{|C|} \quad (9)$$

公式(9)中分式部分分子为最小凸集中不包含 A 和 B 的部分。GIoU 与 IoU 相似,GIoU 作为一个距离度量指标,其坐标误差损失函数 L_{GIoU} 计算公式如(10)所示:

$$L_{GIoU} = 1 - GIoU \quad (10)$$

GIoU 有以下性质: 1、与 IoU 类似,GIoU 也可以作为一个距离度量,loss 可以用 $L_{GIoU}=1-GIoU$ 来计算。2、同原始 IoU 类似,GIoU 对物体的大小不敏感。3、GIoU 是 IoU 的下界,GIoU 总是小于等于 IoU,对于 IoU,有 $0 \leq IoU \leq 1$,GIoU 则是 $-1 \leq GIoU \leq 1$ 。在 A 和 B 形状完全重合时,有 $GIoU=IoU=1$;当 A 和 B 不重合度越高(离的越远),GIoU 越趋近于-1。4、由于 GIoU 引入了包含 A,B 两个形状的 C,所以当 A,B 不重合时,依然可以进行优化。保留了 IoU 的原始性质同时弱化了它的缺点。

2.2.3 Anchor 参数设定

Anchor(先验框)机制起初是由 Faster-RCNN 提出,后来被广泛用于一阶段及二阶段目标检测框架,如 SSD、YOLO、RetinaNet 等。Anchor 主要用来解决目标检测任务中尺度及宽高比例变化范围过大的问题,模型训练的时候真实标签被编码为相对于 Anchor 的偏移量,模型测试时根据模型预测的相对于 Anchor 的偏移量来解码检测结果。通过

Anchor 机制,将尺度和宽高比例的空间划分为相应的几个子空间,降低了目标检测任务的难度,使模型更容易学习。但是在 Faster R-CNN 和 SSD 中,先验框的维度(长和宽)都是手动设定的,带有一定的主观性,如果一开始选取的先验框维度比较合适,那么模型更容易学习,从而做出更好的预测。因此,YOLOv3 中采用 k-means 聚类算法对数据集(COCO)中 80 类目标的标注框做了聚类分析。本文针对 k-means 聚类方法在初始聚类点的选择上存在较大的随机性的问题,使用了随机性更小的 k-means++^[16] 聚类算法来代替原始聚类算法对本文数据集的标注框进行聚类计算,可以有效降低原始算法在初始聚类点所造成的聚类偏差。得到 Anchor 较好尺寸并匹配到对应的特征图,从而可以提高检测准确度和召回率。

3 实验与分析

本文实验在计算机环境配置为 Ubuntu16.04、CUDA(Compute Unified Device Architecture)10.0、CUDNN(CUDA Deep Neural Network library)7.0。硬件配置为 Intel-CPU-i7-8750H 处理器,显卡为 Nvidia GeForce GTX 1050Ti,调用 GPU 进行加速训练。

3.1 交通标志数据集

本文采用的数据集为 Lisa(加州大学圣地亚哥分校数据集)交通标志数据集和 CQ-data(重庆交通标志数据集)见图 8 和图 9。Lisa 数据集中图片拍摄角度以及分类更规范,因此作为本文算法的主要研究数据集。CQ-data 数据集为个人制作数据集,图片的获取来由是百度实景街道截取,该数据集验证本文改进算法的实用性。Lisa 数据集共包含 20 类交通标志,选取了数据集中的 4706 张图片。在 Lisa 数据集中选取其中数量最多的 4 类交通标志作为检测识别对象,分别为 stop、pedestriancrossing、signalahead、speedlimit,共计 5373 个标注。CQ-data 数据集共制作了 1000 张图片。在 CQ-data 数据集中,

选取了 5 类交通标志作为检测识别对象，分别为：禁止鸣喇叭、禁止停车、禁止向左转弯、限速、其它禁令标志，共计 1500 个标注。将标注信息生成与之对应的 xml 格式的目标区域位置的信息，编写 python 程序将 xml 格式的目标区域位置的信息统一转化为 txt 格式，以便让 YOLOv3 读取图片标注信息位置。最后按照 VOC2007 数据集格式进行了整理，分别按 3:1 比例将两个数据集中的图片分为训练集和测试集两类。



图 8 Lisa 交通标志数据集



图 9 重庆交通标志数据集

3.2 Anchor 参数预设

由于 YOLOv3 最原始选定 anchor 尺度和宽高比

是基于 COCO 数据集（80 个分类）确定的，但是本文的 Lisa 数据集应用目标是交通路面上的交通标志牌，原先的设置 anchor box 尺寸不适合本文研究的环境中，基于此，本文使用 Kmeans++ 聚类算法对交通标志数据集的真实标注框进行聚类操作，真实标注框长宽映射到模型输入大小 size 下的聚类结果如图 10 所示。

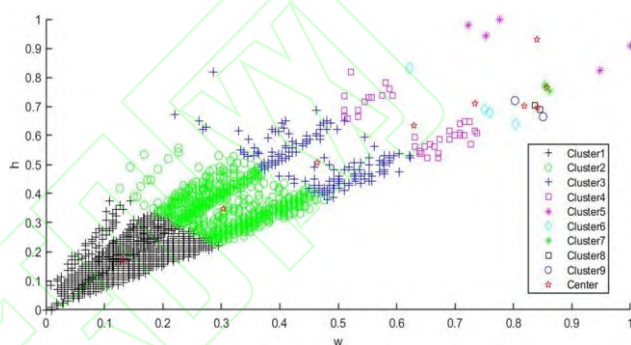


图 10 Grounding truth box 长和宽聚类散点图

由代码实现上来看，anchor 尺寸是相对于预测的特征图大小（13*13），因此真实标注框长宽映射到模型后输出的 anchor 尺寸大小为：(12.41,23.50)、(15.96,31.73)、(21.11,34.88)、(21.74,48.11)、(28.83,42.13)、(29.57, 63.04)、(37.08, 52.16)、(44.06, 75.64)、(64.64, 103.84)，使用自己的数据集进行维度聚类，得出结果和原始的参数相比，自己的聚类结果更为集中，宽高比更符合交通标志数据集的特点，因此使用 k-means++ 算法聚类得到的 anchor 参数来代替原始的参数进行训练和测试。可以使 anchor 框更容易拟合真实目标，从而可以降低模型训练的难度。

3.3 模型训练

训练阶段更改 yolov3 配置文件来对图片进行数据增强预处理以此来提高模型的泛化能力，包括旋转图片角度、调整图片饱和度、调整图片曝光量、调整图片色调、裁剪图片大小、调整图片色度扭曲、调整图片噪声来生成更多的训练样本。之后将增强后的图片 resize 到 416×416 送入到网络模型中进行模型的训练。同时在训练阶段，为了防止 Anchor 与真实框没有 IoU 大于 0.5 的情况，本实验设置当

出现这种情况的时将与真实框 IoU 最高的 Anchor 作为对应类别的正样本参与训练。

该实验以 Darknet53 为基础网络框架，以 YOLOv3 为检测算法框架，基于迁移学习的方法，使用预训练模型作为基础特征提取网络，整个训练过程中共进行 20000 次迭代，其中动量和权重衰减分别配置为 0.9 和 0.0005，批量大小设置为 64，初始学习率为 0.001，并在迭代次数达到 16000 和 18000 时，学习率分别为 0.0001 和 0.00001。由图 11 可以看出，到后期的训练损失值基本低于 0.030，迭代至 17500 次左右，loss 值在 0.030 上下浮动，即达到理想效果。训练过程中平均 Avg IoU 和 Avg GIoU 如图 12 和 13 所示，可以看出 GIoU 整体训练相较于 IoU 更加密集，性能更优。训练过程中精度和召回率如图 14 所示。

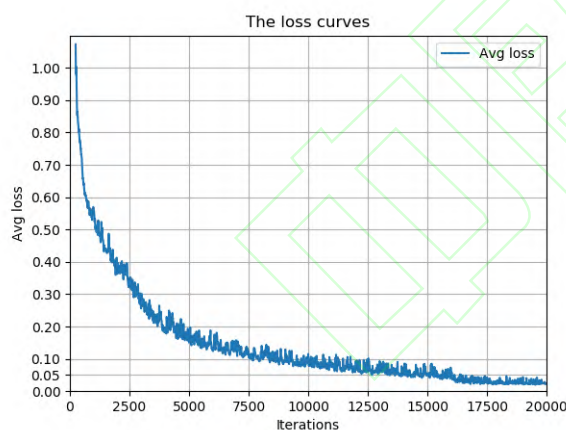


图 11 平均损失变化曲线

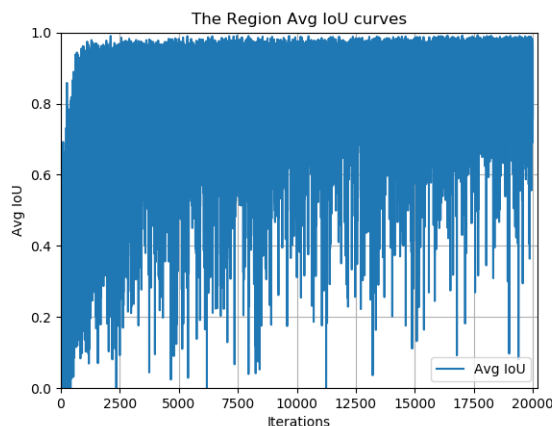


图 12 平均 IoU 变化曲线

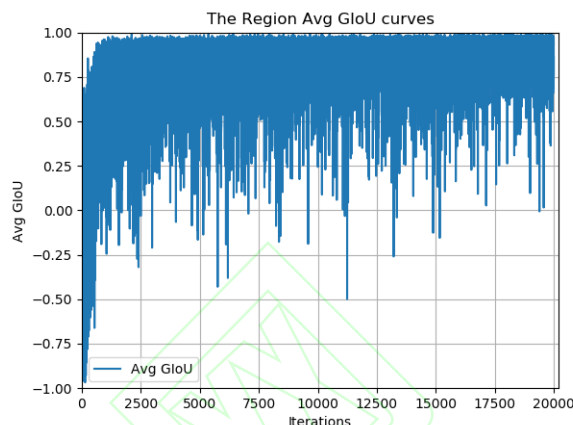


图 13 平均 GIoU 变化曲线

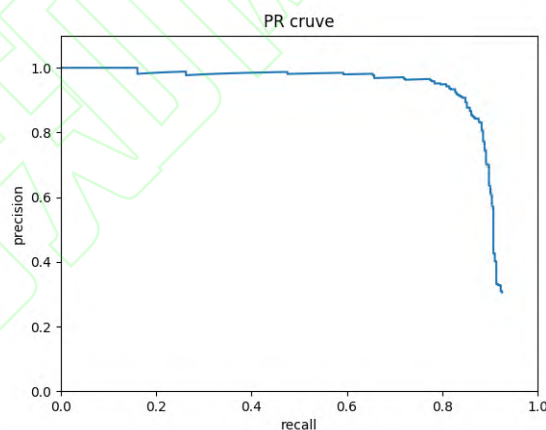


图 14 precision 与 recall 变化曲线

3.4 实验结果与分析

该实验分别使用 Loss-IoU 和 Loss-GIoU 对数据集进行训练，以 IoU 和 GIoU 为度量评价方式，分别得到权重文件 IoU 和权重文件 GIoU，然后设定 Threshold（阈值）为 0.75 分别用 Loss-IoU 和 Loss-GIoU 对这两个权重文件的性能进行测试统计。结果如下表，由表 1 可知，使用 Loss-GIoU 作为回归损失来训练数据集相比与使用原来的 Loss-IoU 作为回归损失，性能得到了提高。

表 1 L_{IoU} 和 L_{GIoU} 对权重文件的性能测试

Loss/Evaluation	AP		AP75	
	IoU	GIoU	IoU	GIoU
L_{IoU}	.6558	.6494	.7333	.7224
L_{GIoU}	.6772	.6759	.7729	.7698
Relative improve	3.26%	4.08%	5.40%	6.56%

实验中设置阈值为 0.5 时 4 类交通标志的 AP 值和准确率的均值 (mAP), 表 2 所示。实验中的前向推理时间都是与实验 A 进行对比, 对比结果如图 15 所示。原始的网络检测模型性能较低, 前向推理时间达到了 151s, 由于使用的是原始的 Anchor 尺寸及宽高比, 因此对于“signalAhead”这类较小的交通标志在目标定位方面效果不如其它几类交通标志, 如图 18 (a) 所示。

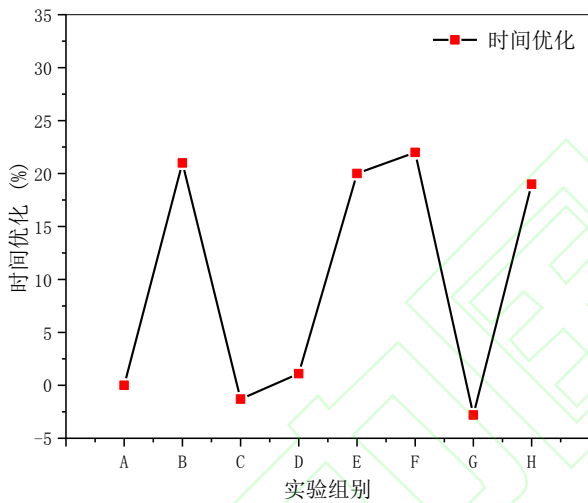


图 15 前向推理时间优化

在网络结构中融合了 BN 层和 Conv 层后, 检测速率以及前向推理速度有 21 个百分点的提高, 但是检测精度基本不变; 在原始模型的基础上加入维度聚类操作后, 对于“signalAhead”这类小目标交通标志 AP 提升较明显, 同时 mAP 有 3 个百分点的提高; 在原始模型的基础上把坐标误差回归的损失函数改进为 GIoU-Loss 后, “Pedestraincrossing”的 AP 值有 8 个百分点的提高; 当同时引入维度聚类 and GIoU-LOSS 的方法后, mAP 相对于只引入 GIoU-LOSS 有 3 个百分点的提高; 同时 mAP 相对

于只加入维度聚类有 6 个百分点的提高, 但是前向推理速度略微下降了 4 个百分点; 当同时改变 3 个模块时, mAP 提升 8.1 个百分点同时前向推理速度也有 19 个百分点的提升, 如图 16 和 17 所示。

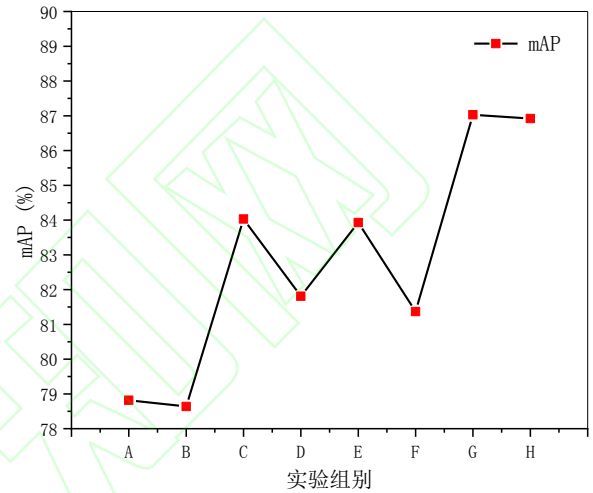


图 16 交通标志检测精度 mAP

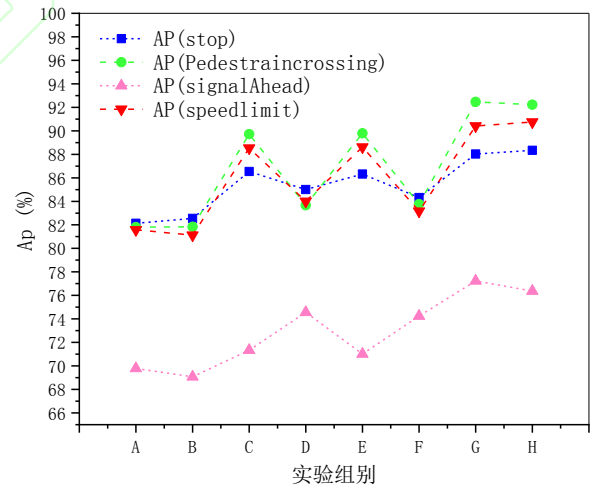


图 17 4 类交通标志检测精度 AP

图 18 (a) (b) 为模型改进前后在 Lisa 数据集测试集的检测效果。改进后的交通标志检测识别效果相较于未改进的在定位预测上更加准确。原始算法中对于复杂路况上容易产生误检, 改进后的算法更有利于实时道路场景中交通标志的检测。图 19 为模型改进后在 CQ-data 数据集测试集的检测效果, 显示了本文提出的改进方法在复杂环境下的检测效果。

表 2 不同模块对模型的性能影响对比

	融合 BN 和 C ONV	L _{GloU}	维度聚类	AP(stop)	AP(Pedestrian crossing)	AP(signalAhe ad)	AP(speedlimi t)	mAP	前向推理时间
实验 A	无	无	无	0.8213	0.8181	0.6978	0.8158	0.7882	
实验 B	有	无	无	0.8255	0.8184	0.6907	0.8112	0.7864	提高 21%
实验 C	无	有	无	0.8654	0.8972	0.7134	0.8853	0.8403	降低 1.3%
实验 D	无	无	有	0.8501	0.8367	0.7456	0.8401	0.8181	提高 1.1%
实验 E	有	有	无	0.8632	0.8979	0.7101	0.8862	0.8393	提高 20%
实验 F	有	无	有	0.8432	0.8377	0.7424	0.8316	0.8137	提高 22%
实验 G	无	有	有	0.8803	0.9246	0.7723	0.9041	0.8703	降低 2.8%
实验 H	有	有	有	0.8834	0.9223	0.7637	0.9075	0.8692	提高 19%



(a) 原始

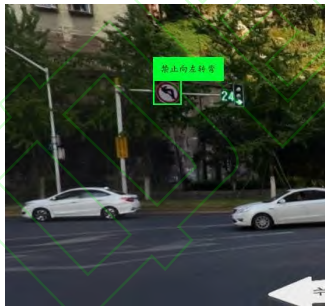


(b) 改进

图 18 改进后算法在 LISA 数据集的检测效果



(a) 暗光照环境



(b) 远距离环境



(c) 多目标物体



(d) 多目标物体

图 19 改进后算法在 CQ-data 数据集的检测效果

为进一步验证本文改进模型对于交通标志检测的有效性，在 Lisa 数据集和 CQ-data 数据集上分别训练了对应的检测模型并与 FasterR-CNN、Yolov2、Yolov3 等主流算法进行了对比。实验结果见下表 3 和 4。

表 3 各模型在 Lisa 数据集上效果对比

算法	骨干网络	mAP	速度 (ms)
Faster-RCNN	VGG16	85.4%	150
Yolov2	Darknet-19	76.3%	11
Yolov3	Darknet-53	80.6%	16
改进 Yolov3	Darknet-53	86.9%	13

表 4 各模型在 CQ-data 数据集上效果对比

算法	骨干网络	mAP	速度 (ms)
FasterR-CNN	VGG16	95.2%	174
Yolov2	Darknet-19	87.4%	14
Yolov3	Darknet-53	90.3%	18
改进 Yolov3	Darknet-53	94.8%	15

从实验结果看，本文的改进模型与 Faster-RCNN 所训练的模型在精度上很近似，但在速度上本文改进方法还是具有显著的优势。在兼顾精度的同时相比 Faster-RCNN 能够达到更快的检测速度，这满足实时交通场景检测的速率要求，与 YOLOv2 相比，虽然检测速度降低了，但是经过 BN 层和 CONV 层的融合已经有了显著的提升。综合来看，本文改进的 Yolov3 模型具有一定的泛化性和鲁棒性。

4 结束语

针对 yolov3 原始算法直接运用于本文的两个交通标志数据集上性能不高的问题，首先将 DBL 单元中 BN 层和 CONV 层合并提高模型前向推理速度，其次引入更加接近检测任务的评价标准 GIoU-Loss 作为代价函数以提高模型的定位能力。最后使用 k-means++ 聚类算法重新确定 anchor 的大小可以使模型更容易收敛来更好适用于本文数据集的检测与识别。实验结果表明，改进后的 yolov3 算法相较于原始算法在 Lisa 数据集和 CQ-data 数据

集上检测平均准确率和检测平均速率有一定的提升。同时本文还有很多地方需要改进,本文算法在识别小目标时,识别精度仍没有达到预期的效果,并且由于工作计算量大消耗大量内存,该模型仍然不适用于实时性工程应用中。如何在提高性能条件的基础上精简网络结构、降低计算量将是未来的主要研究方向。

参考文献:

- [1] Li H, Sun F, Liu L, et al. A novel traffic sign detection method via color segmentation and robust shape matching[J]. *Neurocomputing.*, 169: 77-88.
- [2] Berkaya S K, Gunduz H, Ozsen O, et al. On Circular Traffic Sign Detection and Recognition[J]. *Expert Systems with Applications.* 2015, 48: 67-75.
- [3] John V, Yoneda K, Qi B, et al. Traffic light recognition in varying illumination using deep learning and saliency map[C]. *IEEE*, 2014.
- [4] 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 and Machine Intelligence.* 2017, 39(6): 1137-1149.
- [5] Zhu Y, Zhang C, Zhou D, et al. Traffic sign detection and recognition using fully convolutional network guided proposals[J]. *Neurocomputing.*: S936140309X.
- [6] Tian Y, Gelernter J, Wang X, et al. Traffic Sign Detection Using a Multi-Scale Recurrent Attention Network[J]. *IEEE Transactions on Intelligent Transportation Systems.* 2019, 20(12): 4466-4475.
- [7] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//*Conference on Computer Vision and Pattern Recognition*, 2015: 779-788.
- [8] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[C]//*Conference on Computer Vision and Pattern Recognition*, 2018.
- [9] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]//*IEEE Conference on Computer Vision & Pattern Recognition*, 2017: 6517-6525.
- [10] Tai Y, Yang J, Liu X. Image Super-Resolution via Deep Recursive Residual Network[C]//*IEEE Computer- Vision & Pattern Recognition.* *IEEE*, 2017.
- [11] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//*International Conference on International Conference on Machine Learning.* *JMLR.org*, 2015.
- [12] 白博, 谢刚, 续欣莹. 多尺度级联R-FCN的尾灯检测算法研究[J]. *计算机工程与应用.* 2020, 56(06): 194-200.
- [13] 刘云, 钱美伊, 李辉, 等. 深度学习的多尺度多人目标检测方法研究[J]. *计算机工程与应用.* 2020, 56(06): 172-179.
- [14] Bodla N, Singh B, Chellappa R, et al. Soft-NMS—Improving Object Detection with One Line of Code[C]//*2017 IEEE International Conference on Computer Vision (ICCV).* *IEEE*, 2017.
- [15] Rezatofghi H, Tsoi N, Gwak J, et al. Generalized Intersection over Union: A Metric and a Loss for bounding box regression[C]//*Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: *IEEE*, 2019: 658-666.
- [16] 蒋丽, 薛善良. 优化初始聚类中心及确定K值的K-means算法[J]. *计算机与数字工程.* 2018, 46(01): 21-24.