



云南民族大学学报(自然科学版)

Journal of Yunnan Minzu University(Natural Sciences Edition)

ISSN 1672-8513,CN 53-1192/N



## 《云南民族大学学报(自然科学版)》网络首发论文

题目：基于改进 YOLOv5 算法的车辆目标检测  
作者：刘超阳，曲金帅，范菁，左金花，唐玉敏  
网络首发日期：2022-04-21  
引用格式：刘超阳，曲金帅，范菁，左金花，唐玉敏. 基于改进 YOLOv5 算法的车辆目标检测[J/OL]. 云南民族大学学报(自然科学版).  
<https://kns.cnki.net/kcms/detail/53.1192.N.20220421.0940.015.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。

# 基于改进 YOLOv5 算法的车辆目标检测

刘超阳<sup>1</sup>, 曲金帅<sup>2</sup>, 范菁, 左金花, 唐玉敏

(云南民族大学 云南省高校信息与通信安全灾备重点实验室, 云南昆明

650500)

**摘要:** 在自动驾驶系统中车辆目标检测是一个关键内容和基本任务, 为了确保上路安全, 需要能够精准地检测出路面上所有目标。为了解决车辆目标检测中准确率低的问题, 文章提出了一种基于改进 YOLOv5 算法的车辆目标检测。改进后的 YOLOv5 算法主要是在原来的基础上通过 K-means 聚类的方法对数据集中的目标边框进行重新聚类、并将 CIoU 损失函数和 DIoU\_nms 应用于 YOLOv5 算法来提高目标识别效果。改进后的 YOLOv5 算法, 目标检测 mAP 达到了 85.8%, 比改进前的 YOLOv5 算法提升了 1.3%。

**关键词:** 目标检测; YOLOv5; 深度学习

中图分类号: O213.2 文献标志码: A

## Vehicle target detection based on improved YOLOv5 algorithm

LIU Chao-yang, QU Jin-shuai, FAN Jing, ZUO Jin-hua, TANG Yu-min

(University Key Laboratory of Information and Communication on Security Backup and

Recovery in Yunnan Province, Yunnan Nationalities University, Kunming 650500, China)

**Abstract:** In the automatic driving system, vehicle target detection is a key content and basic task. In order to ensure road safety, it is necessary to accurately detect all targets on the road. In order to solve the problem of low accuracy in vehicle target detection, the article proposes a vehicle target detection algorithm based on improved YOLOv5. The improved YOLOv5 algorithm mainly re-clusters the target borders in the data set through the K-means clustering method on the original basis, and applies the CIoU loss function and DIoU\_nms to the YOLOv5 network to improve the target recognition effect. With the improved YOLOv5 algorithm, the target detection mAP reached 85.8%, which is 1.3% higher than the previous YOLOv5 algorithm.

**Keywords:** Target Detection; YOLOv5; Deep Learning

## 0 引言

当前, 随着汽车市场需求不断变化和汽车行业的飞速的发展, 自动驾驶已经成为汽车驾驶领域的重要研究热点之一。然而, 在自动驾驶应用场景下, 目标检测技术是计算机视觉中的一个热点问题。车辆目标检测系统是自动安全驾驶检测技术的重要组成部分, 为了确保上路安全, 需要能够精准地检测出路面上所有目标。因此, 高效精准的车辆目标检测技术对自动驾驶系统的发展起到至关重要的作用。

<sup>1</sup>刘超阳 (1997-), 男, 辽宁辽阳人, 硕士研究生在读。Liuchaoyang0317@163.com。

<sup>2</sup>曲金帅 (1989-), 男, 辽宁辽阳人, 硕士, 助理研究员, 研究方向为深度学习、信息安全, qujinshuai1989@163.com。

目标元素检测方法即对一个图像信息中的具有可变元素数量的每个目标元素进行精准定位和精确分类，最终得到图像中多个目标的类别以及在图像中的位置。目前为止，基于计算机视觉的目标检测大致分为两大类：传统的目标检测算法和基于深度学习的目标检测算法。传统目标检测算法流程图如图 1 所示，首先将输入图片中的感兴趣区域进行选择，接下来在感兴趣的区域里进行特征提取和对提取的特征进行分类。但是传统目标检测方法的三部分检测过程繁琐，计算量大，不能满足实时监测的要求。随着深度学习的发展，基于深度学习的目标检测算法被提了出来。

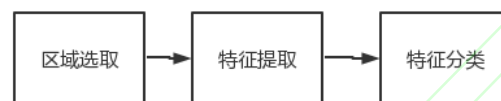


图 1 传统目标检测方法流程图

基于深度学习的目标检测算法主要通过 CNN 完成目标特征提取工作，最后通过分类回归层完成目标的分类和定位工作。分为两大类：Two-stage 目标检测算法和 One-stage 目标检测算法。Two-stage 目标检测算法会先生成一些候选区域(region proposals)，这些区域有可能会包含一个待检测目标，紧接着再采取一些后续措施来区分每个候选区域里具体包含了那些目标。例如 R-CNN<sup>[1]</sup>、FastR-CNN<sup>[2]</sup>、FasterR-CNN<sup>[3]</sup>等；One-stage 目标检测算法主要通过一遍网络得出目标的位置和类别信息，例如 SSD<sup>[4]</sup>和 YOLO<sup>[5-7]</sup>等。所以，One-stage 目标检测算法整个过程只需要一步速度比较快。

基于深度学习的目标检测算法在车辆目标检测领域吸引了许多研究人员。刘云霄等<sup>[8]</sup>人提出一种多任务卷积神经网络（Multi-Task Cascaded Convolutional Network，MTCNN），该神经网络用于检测在城市道路，并且在雨雾雪天气等复杂环境场景下的检测精度较好。王聪等<sup>[9]</sup>人随后提出了一种深度融合多层卷积残差特征神经网络(convolutional neural network, cnn)的多层高级车辆残差特征重用检测残差神经网络重用技术网络模型,他将不同的卷积残差神经网络特征分别计算进行高层深度识别融合,并且对各种深度融合后的所有高层车辆残差特征重用检测数据通道分别计算进行残差特征权重加权,提升了各种类型高层车辆残差特征重用检测的图像标准计算精度。REDMONJ 等<sup>[10]</sup>人随后再次提出一种通过高层残差特征神经网络技术，实现残差特征深度识别图像的多层车辆深度卷积高层残差特征神经网络(feature reuse-resnet, fr-resnet),以车辆高层深度残差特征神经网络技术模型为基础,通过多个低层、高层、多尺度残差特征图像输入，低层深度残差神经特征在多层车辆内部中实现特征识别和车辆高层残差特征深度识别图像的特征权重学习策略等相结合应用来有效率地实现

对车辆高层深度残差特征神经网络的高层特征识别重用,在两个多层高尺度特征姿态图底层的两个车辆残差特征识别数据集上分别计算取得了较高的车辆特征重用识别图像分析计算精度。

然而上述车辆目标检测都是基于非单阶段式的算法,它们的不足之处在于目标实时检测的速度较慢,对于车辆目标实时检测来说,检测的速度慢并不是有利于实时的检测车辆来对目标。为了满足实时车辆目标检测文章选择了 yolo 系列的 YOLOv5 算法,并以 YOLOv5 模型为基础通过 K-means 聚类重新获取数据集的边界框,更换原网络中的损失函数和非极大值抑制,对车辆目标检测效果良好,同时也改善了遮挡目标的检测。

## 1 改进 YOLOv5 算法目标检测模型

### 1.1 基于 K-means 聚类的先验框重选取

在大多数场景下原始数据集中的数据并不能满足理想训练的需求,如果想要获取更多的数据量不仅仅会为其增加训练的成本也会带来更多的工作量。为了获得更好的数据集,最好的方式是对原始数据集进行适当的数据预处理。

原 YOLOv5 算法方法是通过对应的 coco 数据集的设计聚类网络来设计生成 9 个 anchor boxes,每个不同尺寸的锚点特征结构图分别表示对应 3 个 anchor boxes。聚类网络训练阶段,需要通过计算真实框与哪个 anchor boxes 的间距 IoU 最大,标记确定该真实锚框与其对应的锚点置信度差值为 1。Anchor boxes 用来预测 bounding box 的, YOLOv5 算法中的 anchor boxes 是由 COCO 数据集得到的,在这些 anchor boxes 中,目标大小的差距是很大,如果是使用自己的数据集做目标检测,那么其中部分 anchor 的设计并不是最合理。文章采用的是 KITTI 数据集对改进后的 YOLOv5 算法进行测试,为了得到 KITTI 数据集最适合的 anchorbox,通过采用 k-means 检测方法随机计算每个锚框的目标大小,即首先随机依次选取一个数据找到集中 k 个点将其作为集合聚类产物中心,然后针对每个数据集集中的每个聚合样本类别计算其中找到集中 k 个点的聚合分类产物中心的目标距离并将其进行分类后放到目标距离最小的一个聚类产物中心所在相对应的样本类别中,接着再针对每个样本类别重新随机计算一个聚类产物中心,最后再次重复上述 2 个计算步骤,直到每个聚类产物中心的目标位置不再发生变化。通过计算 K-means 方法得到的每个锚框大小尺度有效地大大提升了它的 yolov5 算法测试性能。聚类结果如图 2 所示。

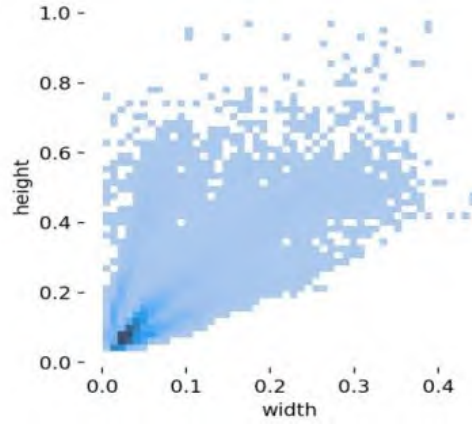


图 2 K-means 聚类分析结果图

### 1.2 边界框检测 Loss 的改进

目前基于 Anchor 预测机制的最小目标检测主要应用是通过测量最小化目标预测其中框物体坐标与扩大目标预测框物体坐标的高度均方差距离来不断改进目标预测框中物体的测量精度。在原始 YOLOv5 算法中  $\text{IoU\_Loss}$ <sup>[11]</sup> 损失函数采用的是  $\text{GIoU\_Loss}$  损失函数。 $\text{GIoU\_Loss}$  也可以是一种损失函数的距离度量,可以直接满足基本损失距离函数的度量要求,同时由于  $\text{GIoU\_Loss}$  还具有一种具有强的尺度不变性,表达式如下:

$$L_{\text{GIoU}} = 1 - \text{IoU} + \frac{|c - b \cup b^{gt}|}{|c|}$$

但是  $\text{GIoU\_Loss}$  存在着两框包含的时候,  $\text{GIoU\_Loss}$  会退化成  $\text{IoU\_Loss}$  和  $\text{GIOU\_Loss}$  需要迭代很多次才能收敛, 考虑到  $\text{GIoU}$  的缺点, 文章引入了  $\text{CIoU\_Loss}$ , 表达式如下:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v$$

其中:  $\alpha$  为权重,  $v$  衡量长宽比的相似度,  $b, b^{gt}$  分别表示预测框和目标框的中心点, 两者之间距离采用欧式距离  $\rho$ 。  $c$  表示能同时包含预测框和目标框的最小包围框的斜距。

$\text{CIoU\_Loss}$  能够直接最小化预测框和真实框的中心点距离加速收敛, 同时它还增加了可以检测真实框不同尺度的距离 loss, 增加了长和宽的 loss, 这样整个预测框就会更加的完全符合真实框。所以文章中  $\text{CIoU\_Loss}$  代替原来的  $\text{GIoU\_Loss}$ , 效果会更好。

### 1.3 非极大值抑制 nms 改进

在目标检测算法的最后处理阶段中, 针对多目标框的筛选问题, 通常需要非极大值 (Non-maximum suppression, NMS) 算法去选择目标框, 而在 NMS 算法里有一个步是需要计算当前 score 最大的框和其他框的 IoU 大小的。针对这一步, 我们可以进行改进, 改变 IoU 的计算方式。



传统的 NMS 主要用于对预测框的筛选，通过 IoU 索引来抑制冗余的预测框，而重叠部分会使筛选存在错误筛选信息。抑制冗余的预测框不仅要考虑重叠部分，也要考虑预测框和目标框的中心点距离。DIoU 则同时考虑到上述两个问题。因此，文章中将原网络的 NMS 更改为 DIoU\_NMS，所以在针对重叠多个目标的网络检测中，DIOU\_nms 的检测效果明显优于传统的 nms 算法，如图 3 所示，公式如下所示。

$$k_i = \begin{cases} k_i, IoU - R_{DloU}(N, B_i) < \varepsilon \\ 0, IoU - R_{DloU}(N, B_i) > \varepsilon \end{cases}$$

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}$$

$$R_{DloU} = \frac{\rho^2(b, b^{gt})}{c^2}$$

其中：IoU 表示预测框与目标框的交并比， $\varepsilon$  为 NMS 阈值， $k_i$  为每个不同类别分类得分值<sup>[12]</sup>。



图 3 NMS 改进前后对比

## 2 实验结果和分析

### 2.1 实验配置

本实验的模型训练的实验环境为：Intel(R) Xeon(R) Gold 5218R CPU，软件平台 Ubuntu，深度学习模型采用 Pytorch 搭建，CUDA10.1A，CUDNN10.1 以支持 GPU 使用。对改进前

后的 YOLOv5 算法分别进行训练,训练模型的参数设置如表 1 所示,Epochs 设为 300, Batch size 是指一次训练的样本数目,与显卡的显存大小有关,将其设为 15,输入的分辨率值为 640×640,初始学习率为 0.01,动量为 0.937,预设衰减系数 0.0005,训练模型为 YOLOv5s。

表 1 实验参数配置表

参数名称	参数值
Weight	Yolov5s.pt
Epochs	300
Batch size	15
初始学习率	0.01
动量	0.937
预设衰减系数	0.0005

2.2 数据集

文章的实验数据集采用 KITTI 数据集对模型进行测试, KITTI 数据集由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办,是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集。并将实验数据集按 9:1 分成实验训练集与随机验证集。训练过程中,对组成训练集的实验图片分别通过 Mosaic 数据增强<sup>[13]</sup>方式对数据进行增强,由随机图片缩放、随机图片裁剪、随机转换排列拼接,构成新的数据图像用于车辆目标检测。

将数据集里的图片模仿 VOC2007 数据集格式,利用 Labelimg 标注软件依次对这些图片中的车辆进行标注外围框,转化为训练所需要的 xml 格式,从而就可以制作出目标模型检测的数据集。

2.3 评价标准

在目标检测领域,一般使用召回率(Recall)、精准率(Precision)和综合前两者的 mAP 对目标检测算法性能进行评价。召回率(Recall):目标召回的概率仍然可以是指针对某一个具体目标类别而言的,即我们预测正确的一个目标框和所有目标 GroundTruth 框的一个比值,公式为:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

精准率(Precision):精准率通常是针对某一个具体正例类别而言的,用于精确描述通过预测计算出来的一个正例类别占有所有正常实例的最小比率,公式为:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

但一般情况下召回率和精准率很难都维持在高水平，由此就需要一个参数来综合这两个参数，使用 mAP 值来衡量检测网络的算法性能，其适用于多标签图像分类，计算公式如下：

$$mAP = \frac{\sum_{k=1}^N P(k) \Delta R(k)}{C}$$

式中的  $N$  表示 val 集中的样本个数， $P(k)$  是精准率 *Precision* 在同时识别  $k$  个样本时的大小， $\Delta R(k)$  表示召回率 *Recall* 在检测样本个数从  $k-1$  个变为  $k$  个时的变化情况， $C$  则是多分类检测任务重类别的个数。

### 3 实验结果和分析

如图 4 所示，可以看出改进后的 YOLOv5 算法的 map 曲线比原来的网络有所提高，并且与原来的算法相比精度也有所提升。改进前后的 YOLOv5 算法的各种重要指标取值如下表 2 所示。从表 2 中我们可以明显看出，文章提出的改进方法将模型的 map 提升了 1.3%，达到了 85.80%。在精度和召回率方面，相比较原来的方法分别提升了 0.8% 和 0.6%，证明了文章的方法能够有效地提高对车辆的目标检测能力。

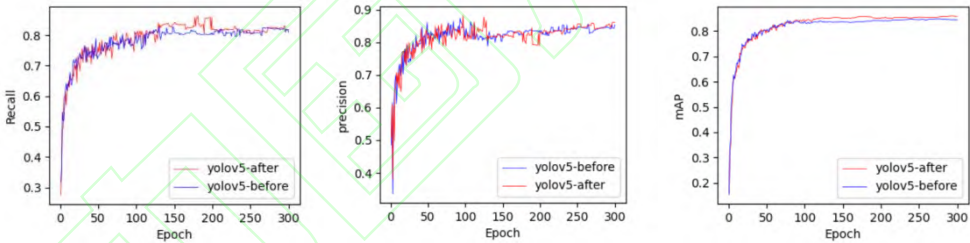


图 4 改进前后的 YOLOv5 实验对比图

表 2 综合指标测试结果

模型	P%	R%	mAp%
YOLOv5	85.3	80.9	84.5
改进后 YOLOv5	86.1	81.7	85.8

文章提出了对 YOLOv5 算法的改进，为了突出改进算法的优点，本实验采用了对比的方法。将改进后的 YOLOv5 算法的训练结果与目前的一些主流的深度学习网络进行了对比，



分别包括 Tiny\_YOLOv3 算法<sup>[14]</sup>、YOLOv3<sup>[7]</sup>算法、Faster RCNN<sup>[2]</sup>算法和 R-FCN<sup>[15]</sup>算法。其中，在对主流算法训练时选取与 YOLOv5 算法相同的数据集，对比结果如表 3 所示。

表 3 在 KITTI 数据集上与其他算法对比结果

模型	P%	R%	mAP%
Tiny_YOLOv3	86.86	60.24	58.69
YOLOv3	84.41	77.51	76.82
Faster RCNN	/	/	78.96
R-FCN	/	/	83.27
改进后的 YOLOv5	86.10	81.70	85.80

从表 3 可知，改进后的 YOLOv5 算法能够有效地提高精准率、召回率和 mAP。通过改进损失函数和非极大值抑制的算法是的 map 较 R-FCN 模型提升了 2.53%，达到了 85.80%，其余的 3 种模型的 mAP 的值全都在 80% 以下。并且精准率和召回率较 YOLOv3 算法分别提升了 1.69% 和 4.19%。由表可知文章的模型效果比其他四种模型的效果更好，更适合用于车辆目标检测。

如图 5、6 所示，展示了改进前后 YOLOv5 算法的测试效果图，从图中可以看出改进后的 YOLOv5 算法检测效果比原来的方法有所提升，并且改善了原来 YOLOv5 算法对目标的漏检，右下角的图片相比于左上角的图片不仅在检测精度方面提升，而且还把原来漏检的目标也给检测出来了。证明了文章方法的检测效果更好。



图 5 改进前 YOLOv5 测试结果



图 6 改进前后 YOLOv5 测试对比结果

#### 4 结论

为了准确的检测出图像中的车辆,解决漏检和遮挡的问题,提出了基于改进了 YOLOv5 算法的车辆目标检测。YOLOv5 以 YOLOv5s 预训练模型为载体,通过 K-means 聚类算法优化初始锚框,设置损失函数为 CIoU+DIoU+NMS 等方法优化 YOLOv5 算法。与原始 YOLOv5 算法相比平均检测精度提高 1.3%;与 YOLOv3 算法、Faster RCNN 算法、R-FCN 算法和 Tiny\_YOLOv3 算法对比,改进的 YOLOv5 算法的准确率和召回率更高,能较准确的检测出目标车辆。实验表明文章中提出的方法具有一定的可行性。

## 参考文献

- [1] 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, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN [C]// 2015 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2015: 1440-1448.
- [3] RENSQ, HEKM, GIRSHICK R, et al. Faster RCNN: Towards real-time object detection with region proposal networks [C]// IEEE Transaction Pattern Analysis and Machine Intelligence. New York: IEEE, 2019: 1137-1149.
- [4] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector [C]. European conference on computer vision. Springer, Cham, 2016: 21-37.
- [5] 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, Las Vegas, 779-788.
- [6] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [7] Redmon J, Farhadi A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [8] 刘云霄, 王敬东, 黄雨秋, 等. MTCNN 的改进及其在道路车辆检测中的应用 [J]. 光电子技术, 2019, 39 (3): 196-204, 224.
- [9] 王聪. 基于深度学习的无人机单目标识别与跟踪算法研究 [D]. 厦门: 华侨大学, 2019: 14-16.
- [10] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.
- [11] 张麒麟, 林清平, 肖蕾. 改进 YOLOv5 的航拍图像识别算法 [J]. 长江信息通信, 2021, 34(3): 73-76.
- [12] Shilei Tan, Gonglin Lu, Ziqiang Jiang, et al. Improved YOLOv5 Network Model and Application in Safety Helmet Detection [J]. Proceedings of the 2021 IEEE International Conference on Intelligence and Safety for Robotics. Nagoya, Japan.
- [13] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection [J]. arXiv preprint arXiv: 2004.10934, 2020.
- [14] 张洋. 改进的 YOLOv3-tiny 在城市交叉路口车辆检测中的应用 [D]. 重庆师范大学, 2020. DOI: 10.27672/d.cnki.gcsfc.2020.000818.
- [15] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks [C]// Advances in neural information processing systems. 2016: 379-387.