



中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 基于改进YOLOv3的车辆检测算法

陈文玉, 赵怀慈, 刘鹏飞, 房建, 孙晖

引用本文:

陈文玉, 赵怀慈, 刘鹏飞, 房建, 孙晖. 基于改进YOLOv3的车辆检测算法[J]. 控制与决策, 2024, 39(4): 1151–1159.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1044>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 多目标小尺度车辆目标检测方法

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

#### 改进YOLOv2的端到端自然场景中文字符检测

End-to-end Chinese character detection in natural scene based on improved YOLOv2

控制与决策. 2021, 36(10): 2483–2489 <https://doi.org/10.13195/j.kzyjc.2020.0270>

#### 基于MobileNet的多目标跟踪深度学习算法

Deep learning algorithm based on MobileNet for multi-target tracking

控制与决策. 2021, 36(8): 1991–1996 <https://doi.org/10.13195/j.kzyjc.2019.1424>

#### Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

#### 基于生成对抗网络学习被遮挡特征的目标检测方法

Object detection via learning occluded features based on generative adversarial networks

控制与决策. 2021, 36(5): 1199–1205 <https://doi.org/10.13195/j.kzyjc.2019.1319>

# 基于改进YOLOv3的车辆检测算法

陈文玉<sup>1,2,3,4</sup>, 赵怀慈<sup>1,2,3†</sup>, 刘鹏飞<sup>1,2,3</sup>, 房建<sup>1,2,3</sup>, 孙晖<sup>1,2,3,4</sup>

(1. 中国科学院光电信息处理重点实验室, 沈阳 110016; 2. 中国科学院沈阳自动化研究所, 沈阳 110016;  
3. 中国科学院机器人与智能制造创新研究院, 沈阳 110169; 4. 中国科学院大学, 北京 100049)

**摘要:** 交通场景下的车辆检测问题存在小目标多、目标遮挡严重等情况, 鉴于此, 提出一种基于改进YOLOv3的车辆检测算法. 由于小目标仅包含较少的像素, 特征不明显, 算法在空间金字塔结构中融入软池化操作, 搭建Soft-SPP结构将多重感受野融合, 通过软池化操作最大程度地保留细节, 有效提取小目标特征; 引入坐标注意力机制, 在调整每个通道特征分配权重的同时能够捕捉具有精确位置信息的远程依赖关系; 提出一种新的损失函数KIoU Loss作为边界框损失函数, 同时考虑边界框的关键点与长宽比使之回归更加准确. 实验结果表明, 改进后的算法在自动驾驶KITTI数据集上平均精度达到94.69%, 相比原始YOLOv3算法精度提升4.13%, 且检测速度仅下降3.16 frame·s<sup>-1</sup>, 在保持检测速度的情况下能够明显提升检测精度.

**关键词:** 车辆检测; 深度学习; YOLOv3; 坐标注意力; Soft-SPP; KIoU Loss

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1044

引用格式: 陈文玉, 赵怀慈, 刘鹏飞, 等. 基于改进YOLOv3的车辆检测算法[J]. 控制与决策, 2024, 39(4): 1151-1159.

## Vehicle detection algorithm based on improved YOLOv3

CHEN Wen-yu<sup>1,2,3,4</sup>, ZHAO Huai-ci<sup>1,2,3†</sup>, LIU Peng-fei<sup>1,2,3</sup>, FANG Jian<sup>1,2,3</sup>, SUN Hui<sup>1,2,3,4</sup>

(1. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China; 2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 3. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China; 4. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Aiming at the problems of vehicle detection in the traffic scene such as a large number of small targets and severe target occlusion, a single-stage target detection algorithm based on the improved YOLOv3 is proposed. Since the small target only contains fewer pixels and features are not obvious, this algorithm builds a Soft-SPP structure based on the idea of spatial pyramid pooling, which integrates multiple receptive fields and adopts soft-pooling operation to retain details to the maximum extent and avoids information loss. The coordinate attention mechanism is introduced to capture the remote dependence with accurate location information. and adjust the weight assigned to each channel feature to make the network better learn important information. A loss function KIoU Loss based on key points and aspect ratio is proposed as the boundary box loss function, which makes the boundary box regression more accurate. The experimental results show that the mAP of the improved algorithm on the autopilot KITTI data set is 94.69%, which is 4.13% higher than that of the original YOLOv3 algorithm, and the detection speed is only reduced by 3.16 frame·s<sup>-1</sup>, which significantly improves the detection accuracy while maintaining the detection speed.

**Keywords:** vehicle detection; deep learning; YOLOv3; coordinate attention; soft-SPP; KIoU Loss

## 0 引言

随着经济的发展,近年来我国各城市车辆数量迅速攀升,除了造成交通拥堵外,每年用于车流量统计、车辆规范管理以及车辆停泊调度等方面的成本也随之增加. 因此在真实场景下对车辆进行检测获取相关信息,有助于城市交通工作的开展,对停车场泊车调度系统的也有现实意义. 近年来,针对真实交通场

景中不同尺度和类型的车辆进行准确实时地检测,成为计算机视觉领域内的一项重要研究内容<sup>[1]</sup>.

传统的车辆检测算法主要针对静态背景,目标运动速度过慢时目标提取困难且计算量大,目标的大小、形态发生改变时识别准确率低. 随着深度学习的研究深入,目前的目标检测算法具有很强的泛化能力,不仅可以自动学习目标深层语义特征,还可以

收稿日期: 2022-06-14; 录用日期: 2022-11-10.

责任编辑: 张国山.

†通讯作者. E-mail: hczhao@sia.cn.

在同一网络中集成不同任务。

Faster RCNN<sup>[2]</sup>是两阶段目标检测模型中的典型代表,将特征提取、区域生成和边界框回归等步骤集成到一个网络中,使网络综合性能有较大提高。Luo等<sup>[3]</sup>在Faster R-CNN的基础上结合神经结构搜索和特征丰富处理实现多尺度交通车辆检测,能够有效提取多层特征并且结合上下文丰富车辆信息,但是算法检测速度较差。Nguyen<sup>[4]</sup>采用MobileNet网络代替Faster R-CNN中的VGG网络,通过软非极大值抑制算法和上下文感知池化,有效解决了车辆间严重遮挡的情况。Zhou等<sup>[5]</sup>使用RoIAligns代替RoI池化,并将Faster R-CNN与对抗网络相结合,提高了网络遮挡目标鲁棒性。

单阶段检测算法不需要提取候选区域,而是基于回归的思想直接获得定位和分类信息,在检测速度与精度之间做了很好的平衡。孔方方等<sup>[6]</sup>以YOLOv3为基础,将4个检测尺度的特征进行融合,并且使用K-means对目标框进行聚类分析,有效提升了全景交通监控中的车辆检测精度。王永顺等<sup>[7]</sup>将空洞卷积引入YOLOv3算法的下采样层,提高特征图的分辨率,加强对小目标的检测效果。宋世奇等<sup>[8]</sup>在SSD网络框架加入自适应的低通滤波层以减小混叠效应对小目标特征提取的干扰,但是SSD网络中先验框的大小和形状不能直接通过学习获得,而是需要手工设置,非常依赖经验。

在算法检测速度、精度和部署难度等多方面进行权衡后,选择整体性能较好的YOLO系列算法,在此基础上进行车辆检测算法的研究和改进。目前,YOLO系列算法已发展出若干版本,其中YOLOv3<sup>[9]</sup>算法在业界应用范围较广,因此选择该版本作为基础,提出一种针对车辆检测的改进算法。首先借鉴特征金字塔结构(FPN),使用SoftPool软池化方法搭建空间金字塔池化(soft spatial pyramid pooling, Soft-SPP)模块将多重感受野融合;然后引入坐标注意力机制<sup>[10]</sup>,将位置信息嵌入通道注意力中,增强感兴趣的目标的表示引导网络更好地学习网络特征;接着针对目标边界框的关键点和长宽比提出新的损失函数KIoU Loss,进一步优化边界框回归损失;最后使用K-means算法在KITTI数据集中进行聚类获得更加适合本数据集的锚框(anchor)。通过对比实验表明,改进后的算法精度更高,小目标检测效果提升明显。

## 1 YOLOv3算法

YOLOv3网络使用Darknet-53作为backbone提取输入图像特征,Darknet-53通过卷积核步长控制

张量的维度,实现尺寸降维、通道升维的目的。这样的方式一定程度上增强了模型的特征提取能力,使Darknet-53对所占图像比例较小的物体更加敏感和准确,有助于后续预测和定位<sup>[11]</sup>。网络经过Darknet-53模型后获得一些特征图,仿照FPN结构,对于13×13大小的特征图,将其送入预测层的同时进行上采样,然后与26×26大小的特征图进行拼接。注意到,YOLOv3中的FPN结构与FPN网络有些许不同,前者在深度方向将两个特征图拼接,后者则是按照维度将两个特征图求和。重复特征拼接操作,将尺寸为26×26的特征图与尺寸为52×52的特征图进行拼接。浅层特征中包含边缘、颜色等低级特征,而深层特征中包含高级语义信息<sup>[12]</sup>,以上操作将浅层特征与深层特征进行有效融合,丰富了特征所含的信息并考虑了目标的不同尺寸,有助于提升检测效果。

## 2 改进的YOLOv3算法

YOLOv3算法在大多数通用数据集中均取得了较好的成绩,是商业届最广泛的网络模型之一。针对交通场景下的车辆目标,合理改进模型使之更适用于车辆检测有利于定位和识别准确率的提高。

### 2.1 K-means聚类分析

YOLOv3借鉴Faster RCNN中基于锚框(anchor box)的思想,使用一组宽高固定的初始候选框,避免模型训练时盲目搜索,有助于模型快速收敛。K-means算法的思想是依据欧氏距离对给定的样本进行分类,将样本集划分为 $k$ 个簇,使簇内的点尽量紧密地连在一起,而簇间的距离尽量大<sup>[11]</sup>。本文使用平均重叠度(AvgIoU)作为数据之间相似度的量度,有

$$\text{AvgIoU} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} I_{\text{IoU}}(B, C)}{n} \quad (1)$$

其中: $B$ 为聚类样本,即数据集真实框的宽和高; $C$ 为簇中心; $n_k$ 为属于第 $k$ 个聚类样本个数; $n$ 为样本总个数; $k$ 为簇的个数; $I_{\text{IoU}}(B, C)$ 为簇的中心框与聚类框的交并比。

本文选取 $k = 1 \sim 12$ 分别对KITTI数据集中的样本进行聚类分析,结果如图1所示。根据聚类结果可知,随着 $k$ 值的增大,AvgIoU值随之增大,并且当 $k > 9$ 时曲线趋于平稳。可以知道,YOLOv3有大小不同的3个预测尺度,所以选取 $k = 9$ 。输入尺寸为512×512时得到的9个新的锚框的具体尺寸为[78×151, 126×244, 52×95, 40×57, 28×75, 27×39, 19×52, 15×37, 10×27]。

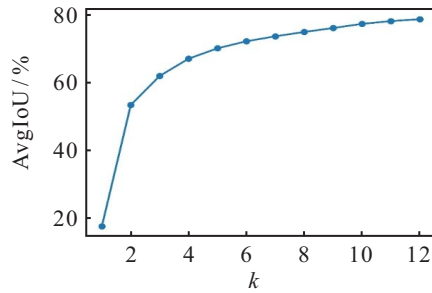


图1 K-means聚类结果分析

## 2.2 改进YOLOv3网络结构

### 2.2.1 Soft-SPP结构

空间金字塔池化 (spatial pyramid pooling, SPP) 是目标检测算法中增加感受野的重要措施之一<sup>[13]</sup>. 深层特征包含有高级的语义信息, 而浅层特征对空间细节信息的表征能力较强. 经典的空间金字塔池化模块对输入特征进行不同的池化操作, 在池化后的特征中抽取固定的维度进行拼接, 因此最后得到特征的维度是固定不变的, 这可以解决信息丢失和尺度不统一的问题<sup>[14]</sup>.

受此启发, 本文构建一种新的空间金字塔池化模块, 称为Soft-SPP结构, 如图2所示. 该模块由3个池化层构成, 为了实现局部特征和全局特征的有效融合, 3个池化层的窗口大小各不相同. 为了使最大的池化窗口尽可能接近需要池化的特征图的大小, 窗口最大尺寸设计为13×13. 其他两个窗口尺寸的设计目前有两种方案: 一是将窗口尺寸设计为等差数列使窗口尺寸分布均匀, 则其余两个窗口可依次递减4, 设置为9×9和5×5; 二是将窗口尺寸设计为  $size_{pool} = size_{max} n_i$ , 其中  $size_{max} = 13$ ,  $n_i$  分别设置为1、2、4, 则其余两个窗口大小设置为7×7和4×4. 本文对以上两种方法分别进行实验, 结果表明第1种方法的检测精度更高, 因此本文算法将池化层的窗口大小设计为13×13, 9×9, 5×5.

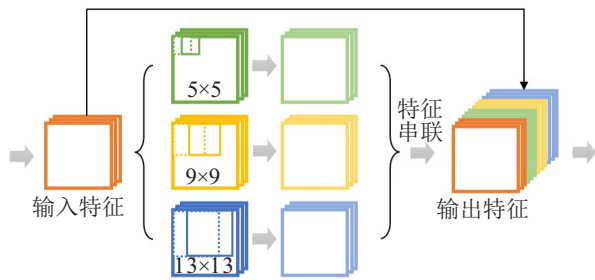


图2 Soft-SPP模块

在卷积神经网络中采用池化层实现降采样, 一方面能使模型获得更大范围的特征, 另一方面能明显降低运算量和参数量, 避免出现过拟合现象. 目前, 卷

积神经网络所采用的池化操作通常为最大池化或平均池化, 这两种方法运算快速但在保留重要信息方面仍有改进空间. Stergiou等<sup>[15]</sup>提出一种新的池化操作SoftPool, 使用softmax加权在很大程度上保留了描述性特征, 同时保持了计算效率和内存效率, 因此所提出Soft-SPP模块中池化层采用SoftPool池化方式.

SoftPool具体运算过程如下. 对于特征图 $a$ 中的局部区域 $R$ , 通过下式对每个像素点 $a_i$ 计算一个权重:

$$w_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} \quad (2)$$

其中:  $w_i$  为对应像素点权值,  $j$  为该像素点邻域 $R$ 中的元素. 最后, 池化的输出值用求得的权值 $w_i$ 与相应位置像素点 $a_i$ 加权并求和得到, 有

$$\tilde{a} = \sum_{i \in R} w_i \cdot a_i \quad (3)$$

其中:  $\tilde{a}$  为池化后的输出值,  $\cdot$  表示点乘.

较小的车辆目标所包含信息量少而噪声多, 经过多次卷积操作后易丢失信息, 如果不能选择合适的池化方式将进一步加重上述情形. 图3展示了各种池化方式的对比效果, 可以看出采用MaxPool后出现了明显的“马赛克”效果, 部分像素点变白, 图中的小目标车辆甚至完全变白. AvgPool对细节的保留则相对较差, 减少了邻域的像素点导致“模糊”程度更为严重. 而SoftPool在降低信息冗余的基础上能够最大程度地保留相关细节. 由于特征图中某个像素对应原图中的一片区域, 如果在池化时丢失某些像素, 则相当于损失了原图中某片区域的信息, 因此本文采用SoftPool搭建空间金字塔池化模块.

### 2.2.2 注意力机制

人类可以在复杂背景中轻松地关注到显著区域, 在计算机视觉中模拟人类的视觉注意力机制, 对输入的图像特征进行动态权重调整能够引导神经网络更加关注重要区域. 本文引入一种全新的注意力机制Coordinate Attention<sup>[10]</sup>, 以下简称CA Module. CA Module通过精确的位置信息对通道关系和长程依赖进行编码, 使网络能够以较小的计算成本关注大的重要区域. CA Module包括坐标信息嵌入和坐标注意力生成两个步骤. 具体而言, 对于输入的特征张量 $X$ , 沿水平坐标使用尺寸为 $(H, 1)$ 的池化核为每个通道的特征进行编码, 相应地, 沿着垂直坐标方向使用尺寸为 $(1, W)$ 的池化核. 因此, 高度为 $h$ 的第 $c$ 个通道的输出可以表示为

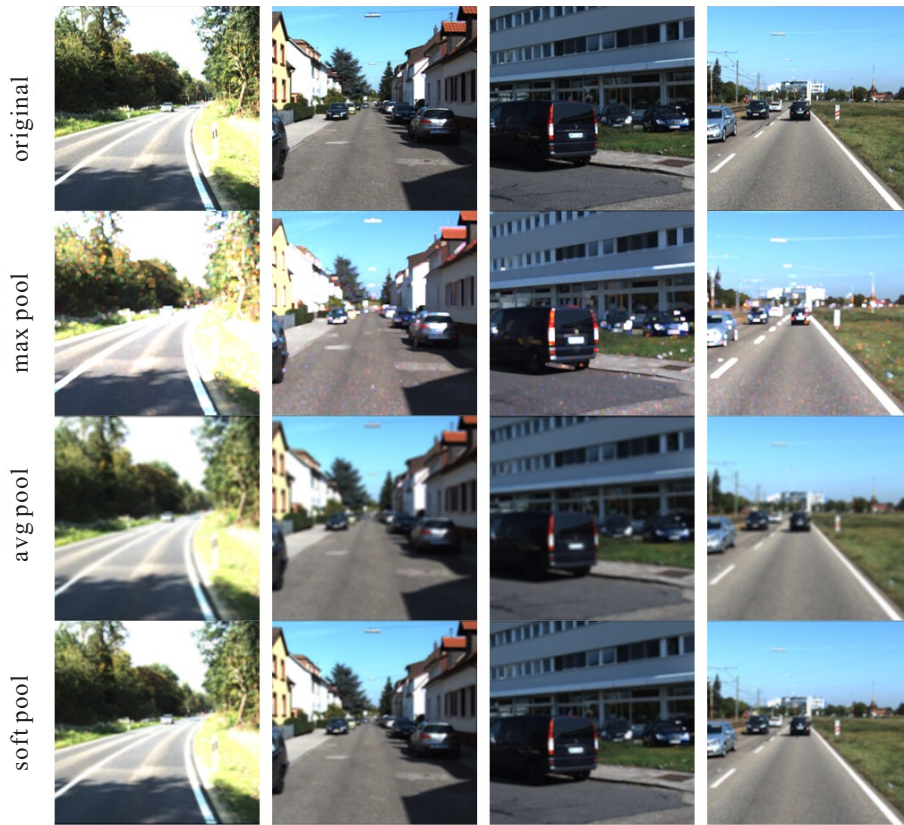


图3 几种池化方式对比

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i). \quad (4)$$

宽度为  $w$  的第  $c$  个通道的输出表示为

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w). \quad (5)$$

对于式(3)和(4)生成的聚合特征图,首先沿维度进行串联操作,然后使用一个共享的  $1 \times 1$  卷积进行变换  $F_1$ ,上述过程表示为

$$f = \delta(F_1([z^h, z^w])). \quad (6)$$

其中:  $\delta$  为非线性激活函数;  $f \in \mathbf{R}^C/r \times (H + W)$  为经过水平方向和竖直方向编码的中间特征图,  $r$  为控制模块大小的下采样率.

将  $f$  沿空间维度拆分为两个单独的张量  $f^h \in \mathbf{R}^{C/r \times H}$  和  $f^w \in \mathbf{R}^{C/r \times W}$ , 利用两个  $1 \times 1$  卷积  $F_h$  和  $F_w$  将特征图  $f^h$  和  $f^w$  变换到与输入  $X$  同样的通道数, 得到如下结果:

$$g^h = \sigma(F_h(f^h)), \quad (7)$$

$$g^w = \sigma(F_w(f^w)). \quad (8)$$

最后对  $g^h$  和  $g^w$  进行拓展, 作为注意力权重, CA Module 的最终输出可以表述为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (9)$$

改进后的网络结构如图4所示, 其中虚线表示部分卷积层省略未画出.

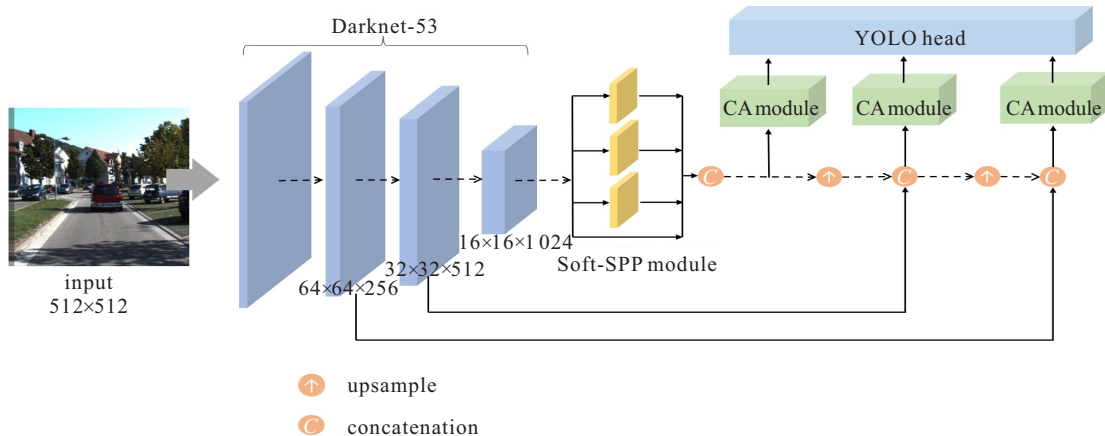


图4 改进后的网络结构

### 2.3 改进边界框回归损失函数

目标检测任务的损失函数一般由分类损失、置信度损失和边界框回归损失3部分构成. 原始YOLOv3采用均方误差(MSE)函数作为边界框回归的损失函数, 无法区分候选框与真实框之间不同的包含情况, 并且不具有尺度不变性<sup>[16]</sup>. 因此, 当前大多数目标检测算法包括YOLOv3后续版本, 均使用交并比(IoU)<sup>[17]</sup>作为度量标准. IoU表示预测框和真实框的交集与并集之比, 其损失计算公式为

$$L_{IoU} = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}. \quad (10)$$

其中:  $B$ 为预测候选框,  $B^{gt}$ 为真实框,  $\cap$ 表示两者相交的面积,  $\cup$ 表示两者并集的面积.

IoU具有尺度不变性的特点, 可以更全面地表征真实框与预测框之间的关系. 但是, 当候选框与真实框没有任何重叠时没有梯度回传, 无法继续进行学习训练. 另外, IoU仅采用面积的交并比而忽略了其他因素, 因此会出现IoU相同但两个目标框的位置关系却不同的情形, 这导致卷积神经网络无法确定下一步的优化方向.

受文献[16]启发, 本文设计KIoU Loss作为边界框回归损失函数, 能够更好地优化边界框的位置信息. 本文认为, 边界框的角点在回归过程中有关键作用, 因此边界框左上和右下两个关键点之间的距离纳入度量范围, 将两个关键点的距离之和与最小包围框 $A_c$ 对角线长度平方的比值作为一个惩罚项. 同时考虑边界框三要素之一的长宽比作为另一个惩罚项, 以保证使两框的“形状”尽可能接近. KIoU计算公式如下:

$$KIoU = IoU - \left( \frac{\rho^2(k_1, k_1^{gt}) + \rho^2(k_2, k_2^{gt})}{c^2} + \alpha v \right); \quad (11)$$

$$\alpha = \frac{v}{(1 - IoU) + v}; \quad (12)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2. \quad (13)$$

其中:  $\rho$ 为两点之间的欧氏距离;  $k_1$ 、 $k_2$ 分别为预测框的左上角点和右下角点;  $k_1^{gt}$ 、 $k_2^{gt}$ 分别为真实框的左上角点和右下角点;  $C$ 为最小包围框 $A_c$ 的对角线长度;  $v$ 用来度量长宽比的相似性, 使得预测框与真实框的形状趋于相似;  $\alpha$ 为权重函数. 为了使边界框回归更加准确、收敛更加迅速, 训练初始阶段可以给 $\alpha$ 乘以适当的系数, 使边界框首先拥有与真实框相似的形状, 然后通过关键点的惩罚项使两框逐渐靠近重合, 最终边界框损失函数设计为

$$L_{box} = 1 - KIoU. \quad (14)$$

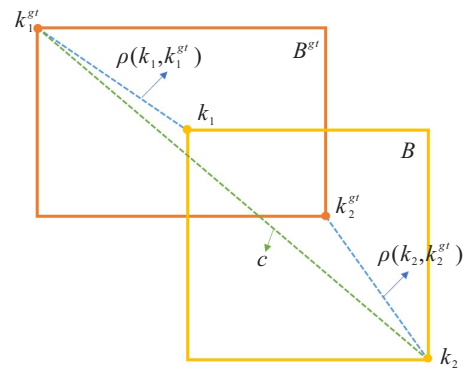


图5 KIoU角点惩罚项示意图

## 3 实验与分析

### 3.1 KITTI数据集

本文面向真实场景下的车辆检测, 选取KITTI数据集<sup>[18]</sup>中的2D图像作为本文车辆检测模型的训练数据. 将该数据集的7481张训练图像作为实验数据, 并根据需求预处理数据集原有的标签信息, 仅保留Car、Van、Truck这3个类别的标签, 处理后的数据按照7:3的比例划分为训练集和验证集.

### 3.2 实验配置与训练

实验使用的操作系统为Ubuntu16.04, GPU为Nvidia RTX 2080Ti, 显存为11G. 实验采用Pytorch 1.5.0框架对模型进行搭建、训练和测试, Python版本为3.7, CUDA版本为10.1. 在训练阶段, 根据显存大小将batchsize设置为8, 每个批次中的输入图像尺寸均固定为512×512. 动量配置为0.937, 权重衰减配置为0.0005, 初始学习率为10<sup>-3</sup>. 实验发现, 当程序运行到60代之后, 损失值出现震荡不再下降, 因此在第60代将学习率设置为原来的0.1以实现损失值继续小范围下降, 达到更好的拟合效果. 图6为训练过程中的损失函数收敛曲线, 可以看出, 在训练次数达到45000次时损失函数收敛曲线趋于平缓.

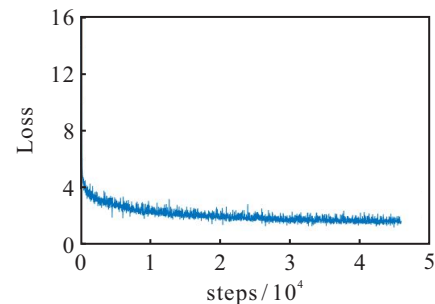


图6 训练迭代次数与损失值的关系

### 3.3 实验结果与分析

本文使用目标检测任务中常用的指标P-R曲线、mAP(mean average precision)和FPS(frames per

second)对所提出模型进行评价.在绘制P-R曲线时,首先通过真正例(TP,true positive)、真反例(TN,true negative)、假正例(FP,false positive)、假反例(FN,false negative)计算准确率Precision和召回率Recall,有

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \quad (15)$$

针对某一类别,以召回率为横轴,以准确率为纵轴绘制P-R曲线,曲线所包含的面积即为该类别的AP,本文所使用数据集包括Car、Van、Truck三类,mAP是对这3种类别的AP值求平均所得.AP值代表模型对某一类目标的检测效果,mAP代表对所有类别的检测效果,值越大表示检测效果越好.实验设置IoU阈值为0.5以检测预测框与真实框的交并比,并以此划分样本.

### 3.3.1 实验前置条件

本节对Soft-SPP模块中池化窗口不同的尺寸方案进行消融实验,两种方案的设计在2.2.1节已详细描述,此处不再赘述,实验结果见表1.

表1 池化窗口的尺寸选择

模型	mAP/%	FPS/(frame·s <sup>-1</sup> )
YOLOv3+Soft-SPP(13, 9, 5)	91.47	48.31
YOLOv3+Soft-SPP(13, 7, 4)	91.44	48.30

由实验结果可知,两种方案在精度和速度这两方面相差不大,第1种方案有微弱的优势,因此在所

提出的改进网络中,Soft-SPP模块的池化尺寸设计为13×13,9×9和5×5.

### 3.3.2 对比实验结果与分析

为了验证所提出算法的有效性,将本文算法与代表性的目标检测算法在KITTI数据集上进行比较.表2展示了本文方法与经典的目标检测算法的检测性能对比.其中:Fast RCNN<sup>[19]</sup>使用选择性搜索生成感兴趣区域,Faster RCNN<sup>[2]</sup>提出了RPN网络,但两者均使用了全连接层,因此速度较慢,与实时检测的要求还有一定距离.SSD 512<sup>[20]</sup>网络的主要设计思想是使用更加密集的锚框和特征金字塔,以改善YOLOv1<sup>[21]</sup>版本较为粗糙的锚框设计,检测精度达到79.63%,同时检测速度可以达到实时检测的要求.所提出算法mAP为94.69%,在精度方面优于所有其他方法,相比YOLOv3算法提升了4.13%;检测速度相比YOLOv3算法下降3.16 frame·s<sup>-1</sup>,这是由于改进网络结构时增加的模块提高了计算复杂度,但是仍然可以达到实时检测的要求.

目前,已有研究提出了较多基于YOLOv3的改进算法,表3显示了本文方法与其他文献方法的对比结果.其他文献多是将现有模块与YOLOv3进行直接结合,而本文则针对具体问题对现有模块进行改进,并且提出了新的基于关键点的损失函数,因此在检测精度方面更有优势,检测速度虽然稍有逊色,但仍然满足实时检测的需要.

表2 不同目标检测算法在KITTI数据集上的检测结果

模型	输入尺寸	FPS/(frame·s <sup>-1</sup> )	mAP/%	Car/%	Van/%	Truck/%
Fast RCNN	600×-	0.63	62.90	65.35	57.94	65.41
Faster RCNN	600×-	11.82	76.98	80.21	73.28	77.45
SSD512	512×512	29.56	79.63	84.67	69.94	84.28
YOLOv3	512×512	49.54	90.56	92.59	86.10	92.99
ours	512×512	46.38	94.69	94.71	93.51	95.85

表3 不同文献检测算法结果对比

方法	mAP/%	提升/%	FPS/(frame·s <sup>-1</sup> )	提升/(frame·s <sup>-1</sup> )
文献[22]	90.08	0.52	76.04	53.31
文献[23]	93.06	1.25	28	-32.62
文献[24]	93.61	2.22	45.48	-6.42
ours	94.69	4.13	46.38	-3.16

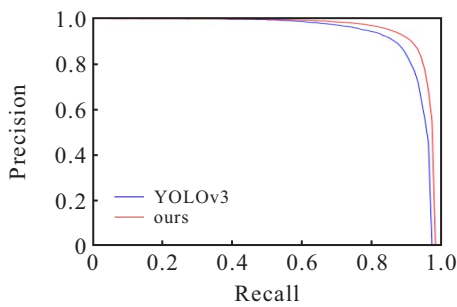
为了进一步验证本文算法的有效性,对本文数据集所使用的3个类别在置信度阈值为0.5的情况下分别测试其准确率和召回率,结果显示在表4中.绘制本文算法和YOLOv3算法的P-R对比曲线如图7所示.可以看出,经过本文改进,3个类别的准确率分别

提高了2.12%、7.41%、2.86%,召回率分别提高了0.78%、5.01%、1.09%,表明Soft-SPP模块能够在最大程度保留细节的前提下对特征进行融合,提高了特征的丰富程度,从而提高检测效果.同时,CA Module能够捕捉具有精确位置信息的远程依赖关系,有助于

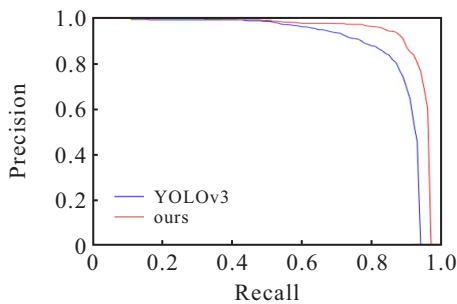
网络更准确地定位目标而不遗失,尤其是在目标密集的情况下有效避免漏检.在改进的边界框回归损失函数KIoU Loss的作用下,网络能够更加精准地找全该类别所有目标,不仅提高了召回率也提升了算法的总体检测精度.

表4 不同类别的准确率和召回率对比

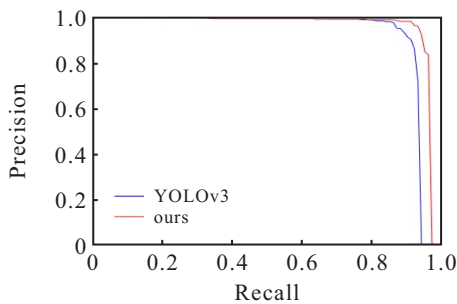
类别	模型	AP <sub>0.5</sub> /%	AR <sub>0.5</sub> %
Car	YOLOv3	92.59	96.32
	ours	94.71	97.10
Van	YOLOv3	86.10	91.61
	ours	93.51	96.62
Truck	YOLOv3	92.99	96.01
	ours	95.85	97.10



(a) Car类P-R曲线



(b) Van类P-R曲线



(c) Truck类P-R曲线

图7 不同类别的P-R曲线

本文提出的车辆检测算法共包含4处改进,为验证其有效性,使用KITTI数据集进行消融实验,在YOLOv3算法的基础上递进增加不同的改进方法,实验结果如表5所示.为了验证本文算法针对小目标检测的有效性,沿用coco评价指标的AP<sub>small</sub>表征在阈值为0.5~0.95的区间内本文算法对小目标的平均检

测精度,其中小目标是指分辨率小于32×32像素的目标.

表5 在KITTI数据集上的消融实验

模型	mAP (%)	FPS (frame·s <sup>-1</sup> )	AP <sub>small</sub> (%)
YOLOv3	90.56	49.54	52.36
+KITTI anchors	91.35	49.48	—
+Soft-SPP	92.29	48.32	—
+CA Module	93.42	47.36	—
+KIoU Loss	94.69	46.38	73.12

由表5可知,使用K-means生成的候选框更加贴近KITTI数据集的目标尺寸和形状,在未对YOLOv3算法进行任何改动的情况下精度提升了0.79%,检测速度几乎不变.Soft-SPP模块通过多尺度软池化将局部特征与全局特征进行了有效融合,增大了模型的感受野.小目标本身所包含的像素数量少,在通过卷积操作之后更容易丢失信息,导致网络无法很好地学习相关特征,因此小目标容易被漏检.Soft-SPP模块中所采用的SoftPool池化操作加强了模型对细节特征的提取能力,使网络获取更多小目标的特征信息,提高对小目标的检测率.CA Module不仅获取了通道间信息,还考虑了方向相关的位置信息,这种注意力机制对特征图沿着水平和垂直方向分别进行编码和信息集成,这样处理可以获得一个方向上的长程依赖信息,同时另一个方向的位置信息得以保持,因此即使物体之间存在遮挡情况仍然能够对其进行定位和检测.以上两项改进提升了模型对远距离小目标的敏感程度,容易发现并识别它们;对近距离大目标也有更好的判断,定位精度进一步提高.通过消融实验可以看出,以上两项改进使网络精度分别提升了0.94%和1.13%,但另一方面基于SoftPool池化方法分块池化操作和CA Module中两个一维特征编码过程增加了计算开销,使检测速度降低至47.36 frame·s<sup>-1</sup>.在回归边界框时使用KIoU Loss,充分考虑到了预测框与真值框的交并比、长宽比和角点距离,能够更好地优化边界框的位置信息,使得网络的预测精度在之前的基础上提升了1.27%.但是由于需要计算两个关键角点的欧氏距离和边界框的长宽比,使得耗时有所增加,检测速度方面FPS下降了0.98 frame·s<sup>-1</sup>.综上所述,将以上所有改进点进行融合后mAP为94.69%,相比原始模型提升了4.13%,AP<sub>small</sub>为73.12%,相比原模型提升了20.76%,表明在小目标检测方面提升较大.同时,检测速度仅下降3.16 frame·s<sup>-1</sup>,仍然符合目标检测实时性的要求.

图8展示了不同算法的检测结果对比,未进行改进的YOLOv3算法存在漏检、误检、重复检测等

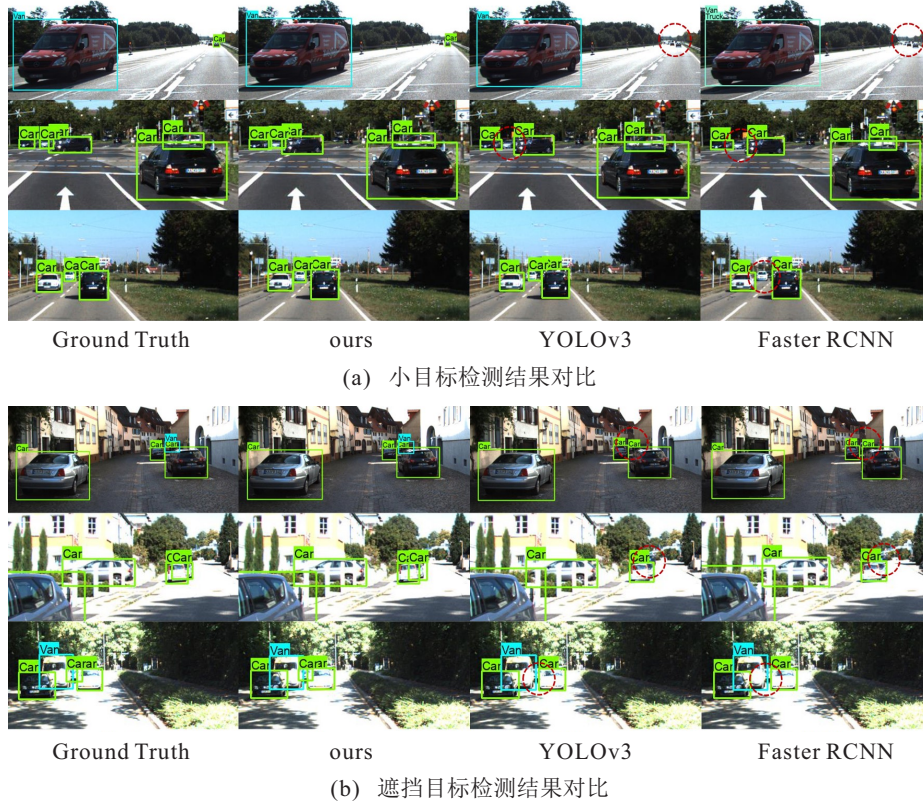


图8 检测效果可视化对比

问题,小目标漏检现象尤为严重.而本文方法通过对YOLOv3算法进行多方面的改进,有效改善了上述问题,达到了更好的检测效果.

#### 4 结论

本文基于YOLOv3算法提出了一种针对车辆检测任务的改进算法,使用改进后的算法能够在检测速度下降很小的情况下提升检测精度.算法设计了Soft-SPP结构,将多重感受野融合,采用软池化操作最大程度地保留细节,减少信息丢失;引入坐标注意力机制,为每个通道的特征分配不同的权重以强调重要信息的同时抑制无关信息;提出一种新的关注目标框的关键角点和长宽比的损失函数KIoU Loss,使目标框位置的回归更加准确,检测效果更好;针对KITTI数据集特点,通过K-means算法对数据集的真值框进行聚类分析,重新生成尺寸更合适的候选框.但本文算法仍有不足,后续研究将在保证精度的基础上搭建轻量化网络以进一步提升检测速度,降低模型大小.

#### 参考文献(References)

- [1] Li H, Fu K, Yan M L, et al. Vehicle detection in remote sensing images using denoising-based convolutional neural networks[J]. *Remote Sensing Letters*, 2017, 8(3): 262-270.
- [2] Ren S Q, He K M, 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.
- [3] Luo J Q, Fang H S, Shao F M, et al. Multi-scale traffic vehicle detection based on faster R-CNN with NAS optimization and feature enrichment[J]. *Defence Technology*, 2021, 17(4): 1542-1554.
- [4] Nguyen H. Improving faster R-CNN framework for fast vehicle detection[J]. *Mathematical Problems in Engineering*, 2019, 2019: 1-11.
- [5] Zhou T, Li Z, Zhang C. Enhance the recognition ability to Occlusions and small objects with robust faster R-CNN[J]. *International Journal of Machine Learning and Cybernetics*, 2019, 10(11): 3155-3166.
- [6] 孔方方, 宋蓓蓓. 改进YOLOv3的全景交通监控目标检测[J]. *计算机工程与应用*, 2020, 56(8): 20-25.  
(Kong F F, Song B B. Improved YOLOv3 panoramic traffic monitoring target detection[J]. *Computer Engineering and Applications*, 2020, 56(8): 20-25.)
- [7] 王永顺, 贾文杰, 王晨飞, 等. 基于改进YOLOv3的车辆识别方法[J]. *激光与光电子学进展*, 2021, 58(16): 240-247.  
(Wang Y S, Jia W J, Wang C F, et al. Vehicle recognition method based on improved YOLOv3 algorithm[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 240-247.)
- [8] 宋世奇, 李旭, 祝雪芬, 等. 基于改进SSD的航拍城市道路车辆检测方法[J]. *传感器与微系统*, 2021, 40(1): 114-117.  
(Song S Q, Li X, Zhu X F, et al. Urban Road vehicle detection method by aerial photography based

- on improved SSD[J]. *Transducer and Microsystem Technologies*, 2021, 40(1): 114-117.)
- [9] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J/OL]. 2018, arXiv: 1804.02767.
- [10] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, 2021: 13708-13717.
- [11] 顾恭, 徐旭东. 改进YOLOv3的车辆实时检测与信息识别技术[J]. *计算机工程与应用*, 2020, 56(22): 173-184.  
(Gu G, Xu X D. Real-time vehicle detection and information recognition technology based on YOLOv3 improved algorithm[J]. *Computer Engineering and Applications*, 2020, 56(22): 173-184.)
- [12] 邹承明, 薛榕刚. 融合GIoU和Focal loss的YOLOv3目标检测算法[J]. *计算机工程与应用*, 2020, 56(24): 214-222.  
(Zou C M, Xue R G. Improved YOLOv3 object detection algorithm: Combining GIoU and focal loss[J]. *Computer Engineering and Applications*, 2020, 56(24): 214-222.)
- [13] 王林. 基于YOLO的改进目标检测算法研究[D]. 无锡: 江南大学, 2020.  
(Wang L. Research on improved object detection based on YOLO[D]. Wuxi: Jiangnan University, 2020.)
- [14] 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(7): 164-170.  
(Liu Z Y, Yuan L, Zhu M C, et al. YOLOv3 Traffic sign Detection based on SPP and Improved FPN[J]. *Computer Engineering and Applications*, 2021, 57(7): 164-170.)
- [15] Stergiou A, Poppe R, Kalliatakis G. Refining activation downsampling with SoftPool[C]. *IEEE/CVF International Conference on Computer Vision*. Montreal, 2022: 10337-10346.
- [16] 钱惠敏, 陈纬, 马宜龙, 等. 基于ResNet34\_D改进YOLOv3模型的行人检测算法[J]. *控制与决策*, 2022, 37(7): 1713-1720.  
(Qian H M, Chen W, Ma Y L, et al. Pedestrian detection based on developed YOLOv3 with ResNet34\_D[J]. *Control and Decision*, 2022, 37(7): 1713-1720.)
- [17] Yu J H, Jiang Y N, Wang Z Y, et al. UnitBox: An advanced object detection network[C]. *Proceedings of the 24th ACM International Conference on Multimedia*. Amsterdam, 2016: 516-520.
- [18] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, 2012: 3354-3361.
- [19] Girshick R. Fast R-CNN[C]. *IEEE International Conference on Computer Vision*. Santiago, 2016: 1440-1448.
- [20] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[J/OL]. 2015, arXiv: 1512.02325.
- [21] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2016: 779-788.
- [22] 杜金航, 何宁. 基于改进的YOLOv3道路车辆实时检测[J]. *计算机工程与应用*, 2020, 56(11): 26-32.  
(Du J H, He N. Real-time road vehicles detection based on improved YOLOv3[J]. *Computer Engineering and Applications*, 2020, 56(11): 26-32.)
- [23] 李汉冰, 徐春阳, 胡超超. 基于YOLOv3改进的实时车辆检测方法[J]. *激光与光电子学进展*, 2020, 57(10): 332-338.  
(Li H B, Xu C Y, Hu C C. Improved real-time vehicle detection method based on YOLOv3[J]. *Laser & Optoelectronics Progress*, 2020, 57(10): 332-338.)
- [24] 张富凯, 杨峰, 李策. 基于改进YOLOv3的快速车辆检测方法[J]. *计算机工程与应用*, 2019, 55(2): 12-20.  
(Zhang F K, Yang F, Li C. Fast vehicle detection method based on improved YOLOv3[J]. *Computer Engineering and Applications*, 2019, 55(2): 12-20.)

## 作者简介

陈文玉(1997—), 女, 硕士生, 从事目标检测的研究, E-mail: chenwenyu@sia.cn;

赵怀慈(1974—), 男, 研究员, 博士生导师, 从事光学信息感知、光信息与仿真技术等研究, E-mail: hczhao@sia.cn;

刘鹏飞(1988—), 男, 副研究员, 博士, 从事图像处理与图像仿真等研究, E-mail: liupengfei@sia.cn;

房建(1992—), 男, 助理研究员, 博士, 从事智能感知、光信息与仿真技术等研究, E-mail: fangjian@sia.cn;

孙晖(1998—), 男, 硕士生, 从事目标检测、三维重建等研究, E-mail: sunhui@sia.cn.