**Modeling and Simulation** 建模与仿真, 2023, 12(3), 1840-1849
Published Online May 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.123170

Hans汉斯

# 基于轻量化改进型YOLOv5的车辆检测方法

**郭　雨**

上海理工大学机械工程学院，上海

收稿日期：2023年2月23日；录用日期：2023年5月1日；发布日期：2023年5月8日

## 摘　要

**本文建立了一种基于计算机视觉的检测前方车辆模型，实现对前方目标车辆的实时检测。本文提出一种基于YOLOv5算法的轻量化车辆检测模型。首先，搭建了YOLOv5检测网络，针对误检漏检问题，使用ConvNeXt重建YOLOv5的主干网络来进行特征提取，以提升网络的细粒度特征融合能力，提高检测精度；然后，为有效解决图像尺度特征变换较大问题其次，在主干网络中引入坐标注意力机制引入提高了主干特征提取效率，进一步提升了算法的特征提取能力；其次，对模型进行剪枝操作，使模型更加轻量化。实验结果表明，改进YOLOv5算法平均精度均值达到97.41%，较原始算法提升5.3%，剪枝后模型检测速率达到175 f/s，较原速率提升了69.4%，证明了该算法可以满足对车辆的实时检测要求。**

## 关键词

**车辆检测，YOLOv5，ConvNeXt，CA注意力，轻量化**

# Vehicle Detection Method Based on Lightweight Improved YOLOv5

**Yu Guo**

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 23rd, 2023; accepted: May 1st, 2023; published: May 8th, 2023

## Abstract

**In this paper, a vehicle detection model based on computer vision is established to realize real-time detection of the target vehicle in front. This paper proposes a lightweight vehicle detection model based on YOLOv5 algorithm. First, the YOLOv5 detection network is built. For the small target model in the image, the backbone network of YOLOv5 is reconstructed by ConvNeXt to extract features, so as to improve the fine-grained feature fusion ability of the network and improve the detection accuracy; Secondly, in order to effectively solve the problem of image scale feature transformation, the introduction of coordinate attention mechanism in the backbone network**

文章引用: 郭雨. 基于轻量化改进型 YOLOv5 的车辆检测方法[J]. 建模与仿真, 2023, 12(3): 1840-1849.
DOI: 10.12677/mos.2023.123170

**improves the efficiency of backbone feature extraction and further improves the feature extraction ability of the algorithm; Thirdly, prune the model to make it more lightweight. The experimental results show that the average accuracy of the improved YOLOv5 algorithm reaches 97.41%, which is 5.3% higher than the original algorithm. The model detection rate after pruning reaches 175 f/s, which is 69.4% higher than the original rate. It is proved that the algorithm can meet the requirements of real-time vehicle detection.**

## Keywords

**Vehicle Detection, YOLOv5, ConvNeXt, CA Attention, Lightweight**



## 1. 引言

随着互联网、5G、计算机视觉等技术的不断进步，无人驾驶的关键技术正在得到越来越多的研究和发展。特别是在多目标检测和识别技术方面，针对无人驾驶场景下的车辆检测，取得了重要的突破。在无人驾驶汽车的交通安全方面，高效精确的车辆目标检测技术具有重要的研究价值和实用意义。因此，对于车辆目标检测技术的研究仍然具有长远的发展前景。

通常传统的车辆检测算法是通过学习汽车的局部特征，例如车牌、车灯等进行车辆的识别[1]。然而，这些基于手工特征的方法在实现车辆检测时通常需要满足一定的环境约束条件，同时也无法充分学习车辆的整体特征。早期基于视频的检测技术首先需要利用摄像机采集视频，然后人工检测、识别，这些需要大量的人力资源，无法实时、准确地预测下一时刻的交通状况。如今朱英凯等人[2]为了克服这些限制，近年来已经提出了基于深度学习的车辆检测算法。在光线较暗的环境下，一些研究者提出通过捕捉车辆尾灯特征的方法来进行车辆的识别；而 Chen 等人[3]提出了一种根据车底阴影特征来获取车辆假设信息的方法。相比于传统的手工特征方法，这些基于深度学习的方法不仅可以减少对环境的依赖，而且能够更好地学习车辆的整体特征，从而提高车辆检测的准确性和鲁棒性。

## 2. YOLOv5 算法原理

在目标检测中 YOLOv5 是一种单阶段目标检测算法，YOLOv5 因运算速度快，识别正确率较高等原因，实际应用范围较广。YOLOv5 相比于 YOLOv4 新增加了 Focus 模块，Mosaic 数据增强，CSPNet (cross stage partial network，CSPNet，YOLOv5 的检测策略为：将输入的图像分为若干网格，包含检测目标的网格负责预测目标位置，最终输出与真实框契合度最高的预测框。YOLOv5 模型包括输入端、Backbone、Neck、Head 四部分。其图 1 是 YOLOv5s 的模型结构图。

### 2.1. 输入端(Input)

在输入端会对图像数据进行 Mosaic 数据增强，自适应锚框计算和自适应图像填充。Mosaic 数据增强是参考了 Cutmix 数据增强方法，是 CutMix 数据增强方法的改进版。Mosaic 数据增强利用了四张图片，对四张图片进行拼接，每一张图片都有其对应的预测框，将四张图片拼接之后就获得一张新的图片，同时也获得这张图片对应的预测框，然后将这样一张新的图片传入到神经网络当中去学习，相当于一下子传入四张图片进行学习，大大提高了效率，也有效解决模型训练中小目标难以被检测的问题。

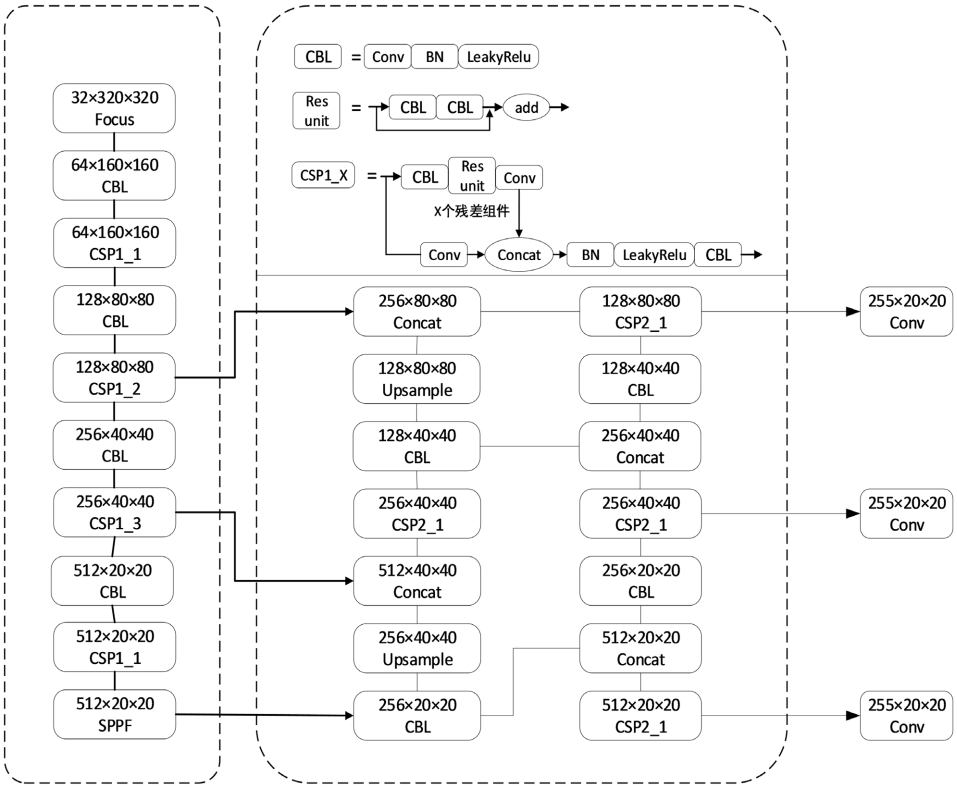


**Figure 1.** YOLOv5s structural model
**图 1.** YOLOv5s 结构模型

在 YOLOV5 算法之中，针对不同的数据集，都需要设定特定长宽的锚框，因此设定初始锚点框是比较关键的一环。首先在网络训练阶段，模型在初始锚点框的基础上输出对应的预测框，进而和真实框 Groundtruth 进行比对，并执行反向更新操作，从而迭代网络参数，每次训练时根据数据集情况自适应的计算出最佳的锚框。

在目标检测中，输入图片尺寸的大小不相同，根据实验结果，输入网络的尺寸统一缩放到同一个尺寸时，检测效果会更好。在 YOLOv5 中，并不是把图片单单的缩放到同一个尺寸，因为很有可能就造成了图片信息的丢失，而是保证整体图片变换比例一致的情况下，那么就可以有效地利用感受野的信息。

## 2.2. 主干网络(Backbone)

主干网络，用来做特征提取的网络，代表网络的一部分，一般用于前端提取图片信息，生成特征图 feature map，供后面的网络使用。通常用 VGG 或者 Resnet，YOLOv5 所使用的主干特征提取网络为 CSPDarknet，输入的图片首先会在 CSPDarknet 里面进行特征提取，提取到的特征可以被称作特征层，是输入图片的特征集合。

在 YOLOv5 中主要使用 CSP 模块和 SPPF 模块，其中使用 CSP 模块能够提高准确率，其内部的残差块使用了跳跃连接，缓解了在深度神经网络中增加深度带来的梯度消失问题。使用 SPPF 模块能对同一个特征图进行多尺度特征提取，有利于提升模型的精度。

YOLOv5 中设计了两种 CSP 结构，以 Yolov5s 网络为例，CSP1_X 结构应用于 Backbone 主干网络，另 SP2_X 结构则应用于 Neck 中，而 CSP 结构的深度是由配置文件里的类别数 n 和 depth_multiple 参数决定，CSP 结构的宽度 width_multiple 参数决定。

SPPF (Spatial Pyramid Pooling-Fast, SPPF)模块是基于空间金字塔池化(Spatial Pyramid Pooling, SPP) [4]，SPP 模块是 He 等人在 2015 年提出，SPP 模块有效避免了对图像区域裁剪、缩放操作导致的图像失真等问题，解决了卷积神经网络对图相关重复特征提取的问题，大大提高了产生候选框的速度，且节省了计算成本，而 SPPF 模块速度很快，效率更高，如图 2 所示。

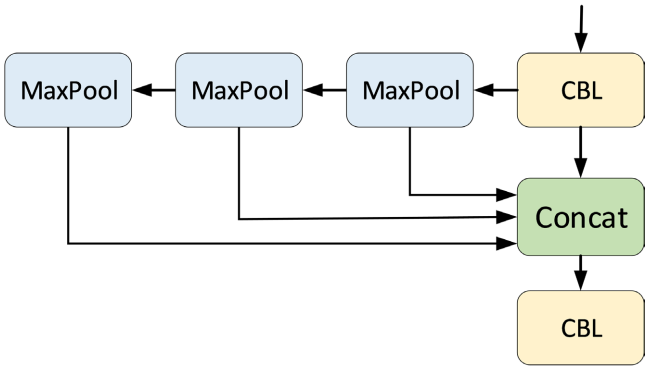


**Figure 2.** SPPF structure diagram
**图 2.** SPPF 结构图

## 2.3. 颈部网络(Neck)

YOLOv5 在 Neck 中采用特征金字塔(Feature Pyramid Network, FPN) [5]与路径聚合结构(Path Aggregation Network, PAN) [6]相结合的方法，加强网络特征融合的能力。FPN 是自顶向下的，将高层的特征信息通过上采样的方式进行传递融合，得到进行预测的特征图，PAN 是自底向上的特征金字塔，通过下采样的方式进行传递融合，传达强定位特征。两者结合会提升模型对不同形状大小物体的识别能力。

## 2.4. 输出端(Head)

在输出端中 YOLOv5 采用 $L_{CIoU}$ 做 Bounding box 损失函数，回归定位损失应该考虑三种几何参数：重叠面积、中心点距离、长宽比。而 *CIoU* 够同时考虑这三者，如图 3 所示。可以加速训练过程中目标检测框回归速度，提高边界框的定位精度。*CIoU* 定义如下：

$$CIoU = IoU - \frac{\rho^2\left(b,b^{gt}\right)}{c^2} - \alpha v \tag{1}$$

其中，$\rho^2\left(b,b^{gt}\right)$ 分别代表了预测框与真实框 Groundtruth 的中心点的欧式距离，$c$ 代表能够同时包含预测框和真实框的最小闭包区域的对角线距离，$\alpha$ 是权重函数，$v$ 用来度量长宽比的相似性，$\alpha$ 和 $v$ 的公式如下：

$$\alpha = \frac{v}{1 - IoU + v} \tag{2}$$

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

*CIoU* 的损失函数公式如下：

$$L_{CIoU} = 1 - CIoU \tag{4}$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b,b^{gt}\right)}{c^2} + \alpha v \tag{5}$$

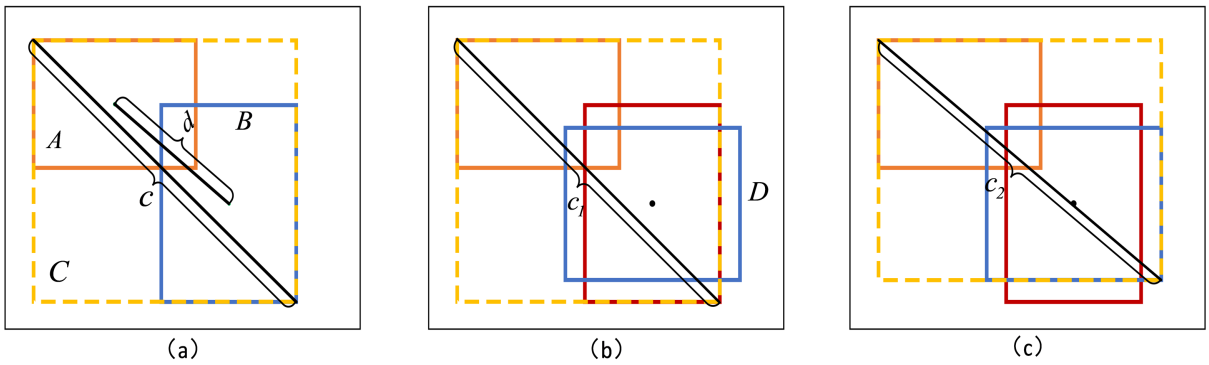


**Figure 3.** CIOU diagram
**图 3.** CIOU 示意图

## 3. 基于改进 YOLOv5 的输电线绝缘子缺陷方法

改进后的 YOLOv5 模型如图 4 所示，在原始 YOLOv5 模型的上采用 ConvNeXt 重建主干网络，用 ConvNeXt 模块替换 CSP 模块，解决原主干网络对小目标提取不足的问题，增强网络对多尺度目标的特征融合能力，最后引入坐标注意力机制 CA 提高了主干特征提取效率。

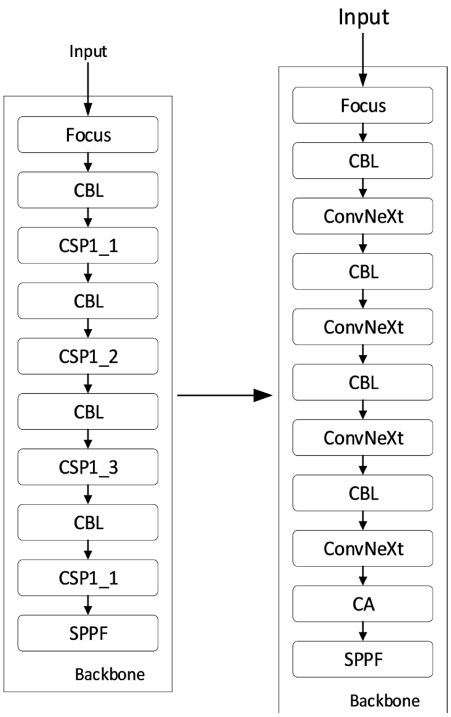


**Figure 4.** Backbone structural diagram
**图 4.** 主干网络结构图

### 3.1. ConvNeXt 网络

在图 5 中，ConvNeXt 网络对原始的 ResNet 模块进行了修改。将 ResNet 模块的堆叠次数从(3, 4, 6, 3)调整为(3, 3, 9, 3)，并通过使用一个卷积核尺寸等于步长的卷积层进行下采样来实现。为了平衡精确率和计算开销，ConvNeXt Block 使用深度可分离卷积，并将通道数调整为 96。此外，ConvNeXt Block 采用了倒瓶颈层设计，使用 3 × 3 深度可分离卷积提取特征后，先使用 1 × 1 卷积升维，将通道数从 96 升到 384，

再使用 1 × 1 卷积降维，将通道数从 384 降到 96，以减少高维信息的损失。ConvNeXt 通过四个阶段的特征提取，得到了一个大小为 7 × 7 × 768 的特征图，通过全局平均池化和 LN 层，最终通过线性分类器输出。

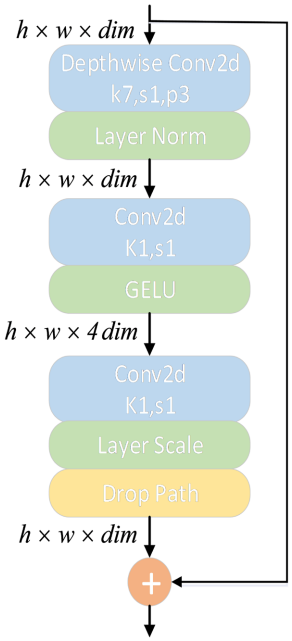


**Figure 5.** ConvNeXt Block structural diagram
**图 5.** ConvNeXt Block 结构图

## 3.2. 注意力模型

CA 注意力机制用于获取图像在宽度和高度方向上的注意力，并对其精确位置信息进行编码。该机制分别在宽度和高度方向上进行全局平均池化，在输入特征图上获得池化后的特征图。将获得的宽度和高度两个方向的特征图拼接在一起，并通过激活函数获得宽度和高度方向上的权值。如图 6 所示。

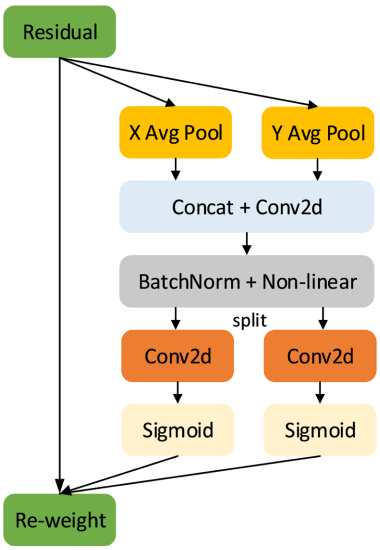


**Figure 6.** CA attention mechanism
**图 6.** CA 注意力机制

### 3.3. 网络模型轻量化

模型剪枝算法采用不同的标准或方法对深度神经网络中的冗余神经元进行剪枝，以最大程度地压缩模型，减少存储需求并提高速度，同时又不会损失模型的精度。

为了实现更快的收敛速度和更好的泛化性能，深度学习网络通常在卷积层后添加批归一化层(BN 层)来处理数据。BN 层通过位移和缩放参数对输入数据进行归一化处理，使得各卷积层输出数据被规范到合理的范围内。BN 层的计算公式如下：

$$y_i = \gamma \cdot \hat{x}_i + \beta \tag{6}$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{7}$$

其中，$x_i$ 和 $y_i$ 分别表示 $BN$ 层的输入和输出，$\mu_B$ 和 $\sigma_B^2$ 分别表示输入样本的均值和方差，$\gamma$ 和 $\beta$ 分别表示位移和缩放参数，缩放参数是参与训练的参数，$B$ 表示当前训练数据的最小批，为预防分母为零，设置很小的非零常数 $\varepsilon$。

在 YOLOv5 神经网络中，激活函数紧接着归一化层进行处理，当归一化层的缩放参数 $\gamma$ 接近于零时，激活函数将通道输入映射到输出端的参数也将更小。这说明该归一化层的输出所对应的通道在网络中的作用非常小，因此可以通过剪枝掉该通道来去除冗余，使网络更加精简。

通常情况下，训练模型时并没有将缩放参数 $\gamma$ 纳入损失函数进行训练，因此在训练结束后得到的 $\gamma$ 值呈正态分布，只有很少一部分接近于零，这样剪枝模型的效果会不尽如人意。为了解决这个问题，可以将 BN 层的 $\gamma$ 参数加入到损失函数中进行训练，筛选出对模型精度影响较小的通道，从而实现更好的剪枝效果。改进后的损失函数如下：

$$L = \sum_{(x,y)} l\left(f\left(x, W\right), y\right) + \lambda \sum_{\gamma \in \Gamma} g\left(\gamma\right) \tag{8}$$

其中，$x_i$ 表示输入，变量 $W$ 表示可训练的权重值，$y$ 表示样本标签值，$\gamma$ 是缩放因子，$g()$ 表示 $L1$ 稀疏正则化，$L$ 是新的损失函数。

## 4. 实验结果与分析

### 4.1. 评价指标

本实验中软件和硬件平台配置参数如下表 1 所示。

**Table 1.** Software and hardware platform configuration parameters
**表 1.** 软件和硬件平台配置参数

| 环境 | 配置 |
|---|---|
| GPU | NVIDIA GeForce RTX3060ti |
| CPU | Intel Core i5-12500H |
| 操作系统 | Windows10 |
| 深度学习框架 | pytorch 1.9.2 |
| CUDA 版本 | cuda 11.2 |

本文实验采用 5 个指标，包括精确率(Precision, P)、类别的平均精度均值(mean average precision, mAP)、

帧率(frames per second, FPS)和召回率(recall, R)、模型大小。

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{11}$$

其中，$TP$ 为正样本预测正确的数量，$FN$ 为负样本预测错误的数量，$FP$ 为正样本预测错误的数量，$TN$ 为负样本预测正确的数量，为第 $i$ 类检测准确率，$N$ 为类别数量。

## 4.2. 实验结果及对比分析

首先，需要用 YOLOv5 训练一个标准的模型。到了权重模型，接下来的步骤是将权重文件输入到模型中，进行稀疏训练，相关的训练参数已经在表 2 中列出。

**Table 2.** Training parameters
**表 2.** 训练参数

| 参数 | 数值 |
|---|---|
| 初始学习率 | 0.001 |
| 终止学习率 | 0.01 |
| 权重衰减系数 | 0.0005 |
| warmup 初始化动量 | 0.75 |
| epoch | 300 |

为了阐述本文中改进的 YOLOv5 算法的优越性，绘制改进后的定分类损失函数和置信度损失函数的图示。其中，分类损失函数是计算锚框与对应的标定分类之间的误差。置信度损失函数则是计算模型置信度的误差损失。

由图 7 可知，改进后的 YOLOv5 算法收敛较快，分类损失函数在 100 轮稳定下来，并收敛与 0.001 左右，而置信度损失函数最终收敛到 0.01 左右，通过以上两组实验可知，改进 YOLOv5 算法的损失函数收敛更快且稳定高效。

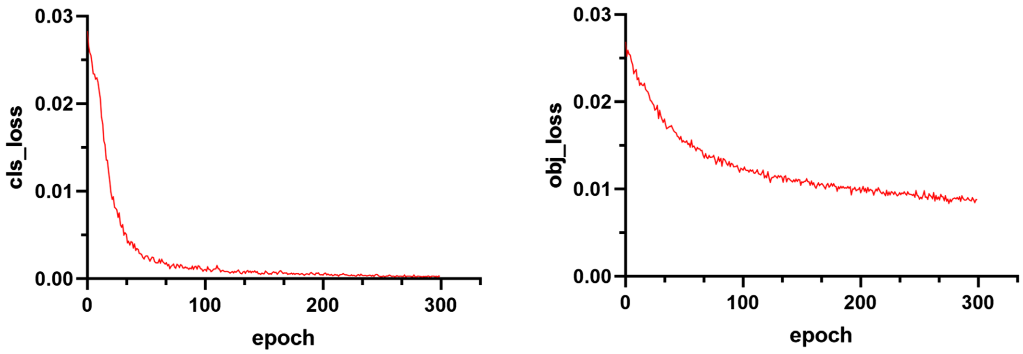


**Figure 7.** YOLOv5 loss function
**图 7.** YOLOv5 损失函数

为了验证改进 YOLOv5 算法的性能，在同样的训练集和验证集下用同样的配置环境与下列 Faster R-CNN、YOLOv4、YOLOX 算法进行比较。

由表 3 数据可知，双阶段检测算法 Faster R-CNN 检测速率不高，只有 22 f/s，不能达到实时检测的要求，而且检测精度相对较低。单阶段目标检测算法，检测精度和速率相对平衡，而改进 YOLOv5 的检测精度最高，达到 97.41%，叫改进前提升了 5.3%，检测速率也最快，达到 175 f/s，提升了 56.3%，可能实现实时检测的要求。如图 8 所示，并没有呈现漏检，错检等现象，且在多目标情况下又能全部检测出，且检测精度较高，有着良好的检测效果。

**Table 3.** Comparison of experimental results of different algorithms
**表 3.** 不同算法实验结果对比

| 模型 | P/% | R/% | mAP 0.5/% | mAP 0.5:0.95/% | FPS |
|---|---|---|---|---|---|
| Faster R-CNN | 88.82 | 86.07 | 89.40 | 52.67 | 22 |
| YOLOv3 | 88.52 | 89.69 | 90.85 | 58.01 | 80 |
| YOLOv4 | 89.87 | 90.32 | 92.19 | 65.84 | 91 |
| YOLOX | 91.63 | 90.89 | 92.57 | 67.65 | 87 |
| YOLOv5 | 91.51 | 92.59 | 93.58 | 70.46 | 112 |
| 改进 YOLOv5 | 95.08 | 94.61 | 97.41 | 73.86 | 175 |

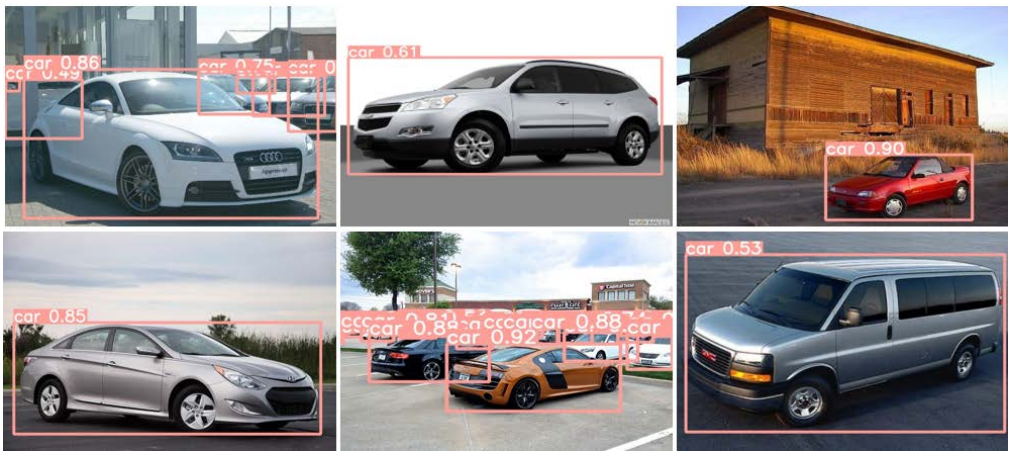


**Figure 8.** YOLOv5 detection rendering
**图 8.** YOLOv5 检测效果图

## 5. 结论

本文提出了一种改进的 YOLOv5 算法，实现了对前方目标车辆的实时检测，解决了目标小，检测速度和精度无法同时满足的问题。根据实验结果显示，相对于原始 YOLOv5 和其他一些主流算法，本文提出改进算法的检测精度达到 97.41%。检测速率达到 175 f/s，较原算法分别提升了 5.3%和 56.3%，且具有较高的检测精度与检测速度，可以完成对车辆的实时检测。

## 参考文献

[1] 桑振. 基于单目视觉的前方车辆测距测速方法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2020.

[2] 朱英凯, 罗文广, 宾洋. 基于改进车底阴影提取算法的前方运动车辆实时检测[J]. 电子技术应用, 2018, 44(4): 86-89+98.

[3] Chen, D.Y., Lin, Y.H. and Peng, Y.J. (2012) Nighttime Brake-Light Detection by Nakagami Imaging. *IEEE Transactions on Intelligent Transportation Systems*, **13**, 1627-1637. https://doi.org/10.1109/TITS.2012.2199983

[4] He, K.M., Zhang, X.Y., Ren, S.Q., *et al.* (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[5] Lin, T.Y., Dollár, P., Girshick, R., *et al.* (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 936-944.

[6] Li, H.C., Xiong, P.F., An, J., *et al.* (2018) Pyramid Attention Network for Semantic Segmentation. *British Machine Vision Conference* 2018, Newcastle, p. 285.