

2-15-2024

## Research on Vehicle Detection Method Based on Improved YOLOX-s

Xiliu Zhang

*College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China, zxl13606146329@163.com*

Xiaoling Zhang

*College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China, 23175809@qq.com*

Minjun He

*College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact [xtfzxb@126.com](mailto:xtfzxb@126.com).

---

## Research on Vehicle Detection Method Based on Improved YOLOX-s

### Abstract

**Abstract:** A improved vehicle detection model based on multi-scale feature fusion of YOLOX network is proposed to solve the problem of missing and false detection of small vehicle targets. Ghost-cross stage partial(CSP) based on the depth separable convolution is designed to replace part of cross stage partial in network to speed up the speed of detection. The max pooling mode of model is improved to Softpool mode, and coordinate attention mechanism is introduced to enhance the feature expression of target to be detected and to optimize the problem of target missing detection. Focal Loss is selected as the confidence loss function of model to increase the weight of inaccurate classification samples and improve the prediction ability of the model for small targets. The experimental results show that the average accuracy of the improved algorithm is improved to 74.96%, and the speed is up to 73 frames per second, which can better meet the requirements of real-time vehicle target detection.

### Keywords

YOLOX, multi-scale feature fusion, vehicle detection model, Softpool, coordinate attention, Focal Loss

### Recommended Citation

Zhang Xiliu, Zhang Xiaoling, He Minjun. Research on Vehicle Detection Method Based on Improved YOLOX-s[J]. Journal of System Simulation, 2024, 36(2): 487-496.

# 基于改进 YOLOX-s 的车辆检测方法研究

张稀柳, 张晓玲\*, 何敏军

(江苏理工学院 电气信息工程学院, 常州 213001)

**摘要:** 为缓解车辆小目标漏检及误检问题, 提出一种基于 YOLOX 网络的多尺度特征融合的改进车辆检测模型。设计基于深度可分离卷积的 *Ghost-CSP (cross stage partial)*, 替换网络的部分跨阶段局部结构, 加快检测速度; 将模型的最大池化方式改进为 *Softpool* 方式, 并引入坐标注意力机制, 增强待检测目标的特征表达, 优化目标漏检问题; 选用 *Focal Loss* 作为模型置信度损失函数以增加分类不准确样本的权重, 提高模型对小目标的预测能力。实验结果表明: 改进算法平均准确率提高到 74.96%, 速度达到 73 帧/s, 在满足实时性要求下可以更好地完成车辆目标检测要求。

**关键词:** YOLOX; 多尺度特征融合; 车辆检测模型; *Softpool*; 坐标注意力; *Focal Loss*

中图分类号: TP391

文献标志码: A

文章编号: 1004-731X(2024)02-0487-10

DOI: 10.16182/j.issn1004731x.joss.22-1106

**引用格式:** 张稀柳, 张晓玲, 何敏军. 基于改进 YOLOX-s 的车辆检测方法研究[J]. 系统仿真学报, 2024, 36(2): 487-496.

**Reference format:** Zhang Xiliu, Zhang Xiaoling, He Minjun. Research on Vehicle Detection Method Based on Improved YOLOX-s[J]. Journal of System Simulation, 2024, 36(2): 487-496.

## Research on Vehicle Detection Method Based on Improved YOLOX-s

Zhang Xiliu, Zhang Xiaoling\*, He Minjun

(College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China)

**Abstract:** An improved vehicle detection model based on multi-scale feature fusion of YOLOX network is proposed to solve the problem of missing and false detection of small vehicle targets. *Ghost-cross stage partial(CSP)* based on the depth separable convolution is designed to replace part of cross stage partial in network to speed up the speed of detection. The max pooling mode of model is improved to *Softpool* mode, and coordinate attention mechanism is introduced to enhance the feature expression of target to be detected and to optimize the problem of target missing detection. *Focal Loss* is selected as the confidence loss function of model to increase the weight of inaccurate classification samples and improve the prediction ability of the model for small targets. The experimental results show that the average accuracy of the improved algorithm is improved to 74.96%, and the speed is up to 73 frames per second, which can better meet the requirements of real-time vehicle target detection.

**Keywords:** YOLOX; multi-scale feature fusion; vehicle detection model; *Softpool*; coordinate attention; *Focal Loss*

收稿日期: 2022-09-20

修回日期: 2022-12-08

基金项目: 2021 年江苏理工学院研究生实践创新项目(XSJXC21\_57)

第一作者: 张稀柳(1998-), 女, 硕士生, 研究方向为机器视觉、图像处理。E-mail: zx113606146329@163.com

通讯作者: 张晓玲(1974-), 女, 副教授, 博士, 研究方向为光电检测和机器人视觉导航。E-mail: 23175809@qq.com

## 0 引言

随着互联网、5G、计算机视觉等技术的发展，无人驾驶关键技术的研究持续增多，无人驾驶场景中的多目标检测和识别技术不断被突破。其中，车辆是自动驾驶场景下目标检测的最主要类别之一。为了确保无人驾驶汽车的交通安全，高效精确的车辆目标检测技术有着重要的研究价值和实际意义。

传统的车辆检测算法一般通过学习汽车的车牌、车灯等局部特征进行识别<sup>[1]</sup>。例如，为了实现在光线较暗的地方识别车辆，Chen 等<sup>[2]</sup>提出通过捕捉车辆尾灯特征的方法；朱英凯等<sup>[3]</sup>提出根据车底阴影特征得到车辆假设信息。这些基于手工特征的方法通常需要具备一定的环境约束条件，并且对于车辆的整体学习不足。

深度神经网络可以挖掘目标的本质特征，学习目标的全局信息，具有很强的鲁棒性和记忆能力，因此，许多学者利用神经网络完成目标检测任务。目前，基于深度学习的目标检测算法分为两大类，分别是以 R-CNN、Fast R-CNN<sup>[4]</sup>为代表的生成区域建议框后再运用神经网络分类的两阶段模型和以 YOLO<sup>[5]</sup>、SSD<sup>[6]</sup>为代表的基于卷积神经网络回归思想的一阶段模型。前者算法检测精度高，但检测效率低。然而，Redmon 等<sup>[7]</sup>设计了单阶段算法 YOLO，只需利用一个卷积神经网络直接获得整张图片中的目标类别和位置，这使得算法运算速度加快，但会导致检测精度有所降低。随着一阶段检测模型研究的深入，YOLO 算法也不断被改进和优化，从而衍生出 YOLO9000<sup>[8]</sup>、YOLOv3<sup>[9]</sup>和 YOLOv4<sup>[10]</sup>等 YOLO 系列算法。将 YOLO 应用于车辆目标检测任务的研究也有很多，文献[11-13]分别使用改进的 YOLOv3-v4 模型用于车辆目标检测。

本文通过优化基于 YOLOX-s 的网络架构，缓解自动驾驶场景下车辆检测中环境复杂，物体遮挡重叠干扰以及小目标检测精度较低等问题。

(1) YOLOX 算法忽略了浅层网络的较高分辨率、强定位信息的优点，一定程度上影响了全局特征学习和网络特征提取，因此，将浅层网络与路径聚合网络(Path Aggregation Network, PANet)中感受野较大的深层网络相融合，提高网络的特征学习能力；

(2) YOLOX 的卷积操作比 YOLOv5 多，造成模型参数量也较多，必将影响网络推理速度，因此，利用深度可分离卷积设计 Ghost-CSP(cross stage partial)，缩减了原主干网络中的跨阶段局部(CSP)的参数量，加快模型推理时间；

(3) YOLOX 模型的空间金字塔池化(spatial pyramid pooling, SPP)模块中最大池化方式只保留最强烈的特征部分，忽略其他激活，导致细微特征丢失，从而影响小目标检测，因此，使用 SoftPool 池化改进，保留更多的细粒度特征信息；

(4) 在主干网络后引入坐标注意力机制，增加感兴趣区域学习，提高检测效果；

(5) YOLOX 中所使用的损失函数并未解决目标检测算法中正负样本不均衡问题，因此，通过优化网络损失函数，提高模型对小目标的预测能力。

## 1 YOLOX

在 YOLOv3、YOLOv4 和 YOLOv5 中都存在需要手工设定样本正负，导致模型泛化能力较差的问题。为了解决此类问题，旷视科技研究院 Base Detection 组提出了 YOLOX<sup>[14]</sup>。YOLOX 设计思路：以 YOLOv3 为原始模型，在其架构上增加解耦头(Decoupled Head)、数据增强、Anchor Free，以及 SimOTA 部件。YOLOX 的性能超过 YOLOv3，较 YOLOv5 也有一定优势，其检测速度及检测效果都取得了不错的成绩。

YOLOX 在网络输入端的数据增强中采用了 Mosaic 和 Mixup，对于小目标的检测效果有一定提升；在主干网络部分采用了跨阶段局部网络；在中间网络结构中，依旧使用了特征金字塔网络

(feature pyramid network, FPN)来进行多尺度输出。在网络预测头部分, YOLO系列都采用了耦合的检测头, 即将分类、回归的检测任务同时处理。YOLOX采用了解耦头结构, 将特征平行分成两路卷积特征, 分别处理分类和回归任务, 不仅加快了网络收敛速度, 还提高了目标定位的精确性和分类的准确度。除此之外, YOLOX在预测框和真实框进行关联中采用了比Anchor based参数量少的Anchor free锚框机制和简化版的SimOTA标签分配策略。这两个方法的利用使YOLOX的训练时间更短, 不需要额外的优化参数步骤, 同时也保证了在绝大多数数据集下精度基本不受影响。YOLOX具有多个版本的网络结构, 可供不同要求的检测任务选择, 其中, YOLOX-s<sup>[15]</sup>是较轻量化的版本, 具有代表性且方便部署, 因此, 选择在其基础上进行改进以更好地解决车辆目标检测问题。

## 2 模型改进

### 2.1 主干网络浅层特征层与PANet网络融合

浅层网络提取的特征图具有更多的像素点信息和细粒度信息, 但是语义性更低, 噪声更多。深层网络提取的特征图感受野较大, 包含更多的

语义信息, 但是对细节感知能力较差<sup>[16]</sup>。原YOLOX-s网络利用主干网络中的“feat2”“feat3”“feat4”浅层特征层与颈部网络进行连接, 但小目标在图中本身尺寸较小, 并且经过模型的下采样处理后, 小目标特征感受野不断被放大, 那么能被利用的特征会越来越少。因此, “feat1”浅层特征层中的信息也应该被利用。

在YOLOX-s主干网络中“feat1”浅层特征层是模型经过切片操作、一步卷积操作和一步CSP操作后得出的特征层, 包括更多的未处理和模糊的特征信息, 感受野也较小, 但是也包含更多的全局特征和细粒度信息。将这层网络与路径聚合网络模块的深度特征层相互融合, 使模型将浅层的强定位信息和边缘特征传递到深层语义信息中, 增加更多的梯度信息, 可以提高目标检测效果。改进后的YOLOX-s网络结构如图1所示。在无人驾驶过程中, 远处车辆在相机实时捕捉中往往存在目标太小且模糊的特征, 为了更加精确地预测目标坐标信息, 判断远处车辆位置, 需要将模型网络的浅层定位信息与深层特征信息相结合。因此, 在YOLOX-s模型的基础上利用多尺度特征融合提高目标定位和识别准确度尤为重要。

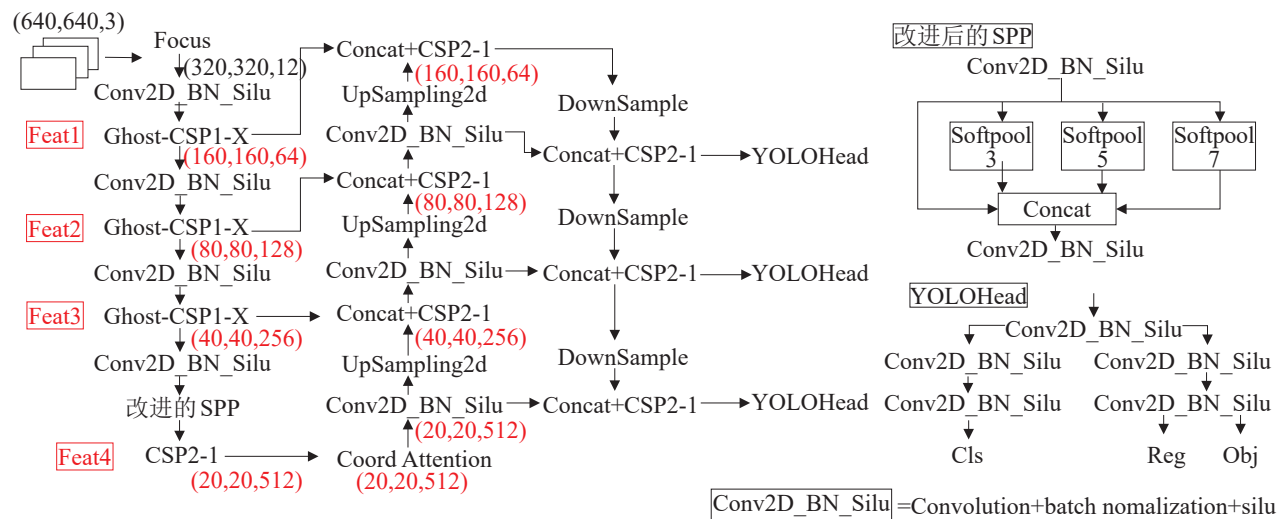


图1 改进后的YOLOX-s结构  
Fig. 1 Structure of improved YOLOX-s network

<http://www.china-simulation.com>



## 2.2 CSP 结构优化

在原始模型的 YOLOX 主干网络中, 卷积操作较多, 计算耗时, 而且小目标所具有的特征信息会随着卷积操作而减少或者消失, 从而使小目标的检测速度更慢, 检测难度更大。为此, 借鉴了 GhostNet<sup>[17]</sup> 中的 Ghost 模块, 其主要操作是先利用普通卷积计算生成部分内在的特征映射, 然后再进行逐层线性卷积, 从而增强特征, 增加信道, 最后将它们组合在一起得到更多的特征图, 具体结构如图 2 所示。其中, 图中  $\Phi$  是卷积线性运算。本文采用由 2 个 Ghost 模块组合成的 Ghost Bottleneck 优化 YOLOX-s 主干网络中的 CSP 结构, Ghost Bottleneck 的结构如图 3 所示。

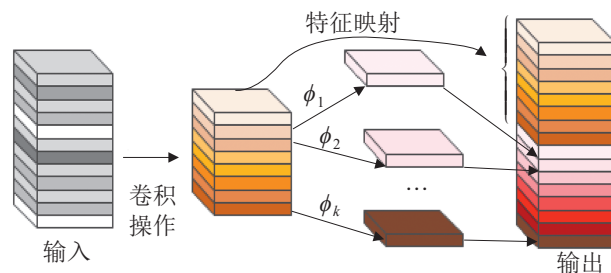


图 2 Ghost modules 结构  
Fig. 2 Structure of Ghost modules

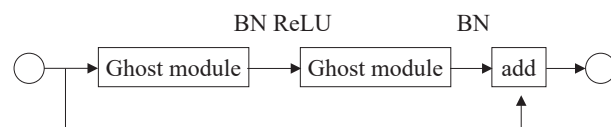


图 3 Ghost Bottleneck 结构  
Fig. 3 Structure of Ghost Bottleneck

Ghost Bottleneck 中第 1 个 Ghost 模块用作扩展层, 增加通道数, 第 2 个 Ghost 模块的作用是为匹配网络直径结构缩减输出通道数, 并利用直径结构连接 2 个模块的信息传输<sup>[18]</sup>。通过该结构方式不仅能减少模型参数量和计算量, 还能优化网络的信息传输, 提高模型的检测效率。

如图 4 所示, 将 YOLOX-s 主干网络的原 CSP 模块中的残差结构( $x$  个 Bottleneck)替换成 Ghost Bottleneck, 并去除了原模块分支上的卷积层, 将 Ghost-CSP

模块的输出特征图与另一个分支的输出特征图直接相连, 进一步有效减少了模块的参数数量。

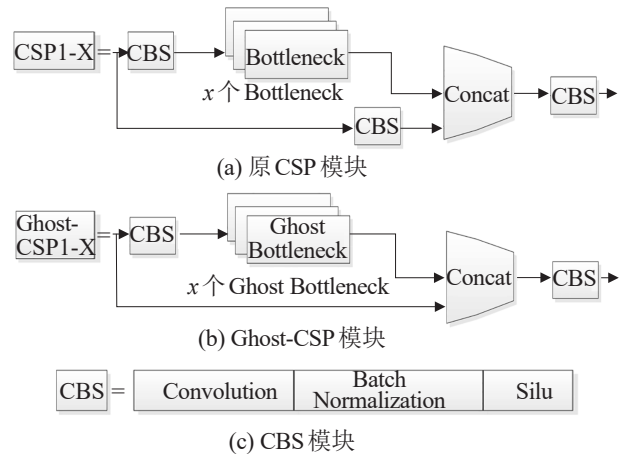


图 4 Ghost-CSP 结构  
Fig. 4 Structure of Ghost-CSP

## 2.3 SPP 结构

在加深卷积神经网络的网络层数后, 其更深层次特征图的信息就越抽象, 导致识别小目标的精度就越低。YOLOv4 系列之后, YOLO 网络就引入了 SPP 模块<sup>[19]</sup>, 成功融合了多尺度局部特征和全局特征, 增强了最终特征图的表达能力, 有效地改善了检测小目标的准确性。在 YOLOX 中, SPP 模块主要由并行的 4 个池化分支构成, 分别是内核尺寸为 5、9、13 的最大池化和 1 个跳跃连接。其中, 最大池化(Max Pooling)是选图像区域最大值作为该区域池化后的值, 从而保留特征图响应最强烈的部分, 但是这类池化容易丢失大部分激活, 易损失重要检测信息导致漏检。因此, 本文提出用 SoftPool<sup>[20]</sup>池化替换 Max Pooling, 从而保留更多细粒度特征信息, 改进对比如图 5 所示。

由于 SoftPool 在权重与相应的激活值一起用作非线性变换, 如式(1)所示, 最终实现较高的激活比较低的激活占更多的主导地位。因为大多数池化操作都是在高维的特征空间中执行的, 突出显示具有更大效果的激活相比简单地选择最大值是一种更平衡的方法。由图 6 可以看出最大池化和 SoftPool 两者输出的不同和最后的效果对比。

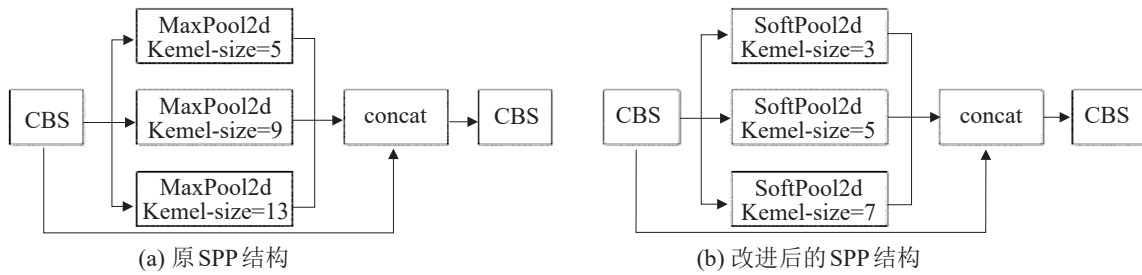


图5 原SPP结构和改进后SPP结构对比

Fig. 5 Comparison of original SPP module and improved SPP module

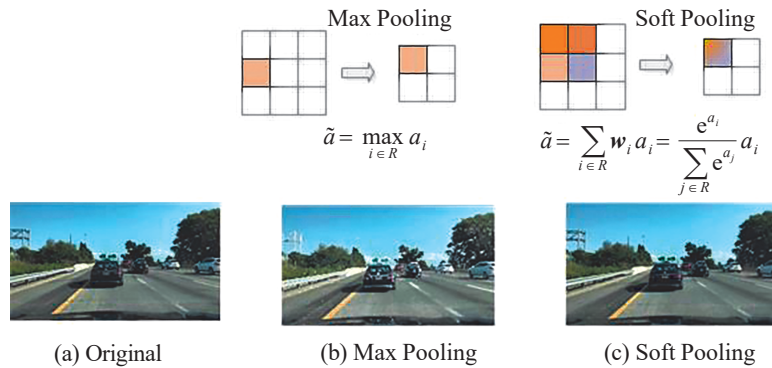


图6 最大池化和SoftPool池化对比

Fig. 6 Comparison of the Max Pooling and the SoftPool

SoftPool以自然指数为基础, 结合Softmax加权方法来保留输入的基本属性, 同时放大更大强度的特征激活。具体来说, SoftPool为内核区域 $R$ 内每个激活因子 $a_i$ 赋予不同的权重 $w_i$ 后, 再对 $R$ 内所有激活因子进行加权求和。其中,  $w_i$ 为该激活因子的指数 $e$ 与区域内的所有激活因子的自然指数之和的比值。因此, SoftPool也是可微的, 即在反向传播时, 局部内核区域内的所有激活因子将至少被分配一个最小梯度值<sup>[21]</sup>。

$$\tilde{a} = \sum_{i \in R} w_i a_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} a_i \quad (1)$$

除了将SPP结构中最大池化替换成SoftPool外, 实验中还将3个内核尺寸修改为3、5、7, 从而以较小的内核保留更多的特征信息。为保证输出的特征图尺寸一致, 在SoftPool2d()方法的基础上增加了填充功能, 使SoftPool可以在多尺寸变化下保持平衡。

## 2.4 增加CA机制

注意力模块可以作为一种提高网络特征表达能力的计算单元, 通常用来告诉模型需要关注的感兴趣内容和位置, 目前常被应用于深度神经网络。然而, 常用的SE(squeeze-and-excitation)注意力机制只考虑通过建模通道关系来重新衡量每个通道的重要性, 而忽略了位置信息。CBAM(convolutional block attention module)通过在通道上进行全局池化来引入局部的位置信息, 但是无法捕获特征图上的长期依赖。

CA机制<sup>[22]</sup>是一种高效的注意力机制, 它将全局池化改造为2个一维向量的编码操作, 分别沿2个空间方向聚合特征, 具体结构如图7所示。这样, 坐标注意力可以同时从2个空间方向获取长距离依赖关系和精确的位置信息。最后, 将生成的特征图分别编码为一对方向感知和位置敏感的注意力图<sup>[22]</sup>, 进而有助于网络更准确地定位感兴趣的目标。

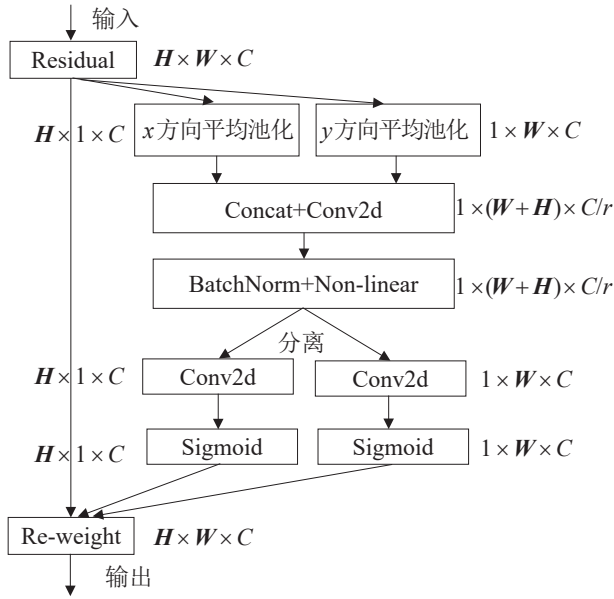


图7 坐标注意力的结构  
Fig. 7 Structure of CA

CA机制首先进行Coordinate信息嵌入，具体使用 $(H, 1)$ 和 $(1, W)$ 的池化核分别沿着水平坐标和垂直坐标对每个通道进行编码。因此，高度为 $h$ 和宽度为 $w$ 的第 $c$ 个通道的输出可以分别表示为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq H} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (3)$$

上述2个变换可以产生2个独立方法感知注意力特征图。接着，对一对特征图进行拼接(concatenate)操作，并使用 $1 \times 1$ 卷积 $F_1$ 生成中间特征图 $f \in \mathbb{R}^{C/r \times 1 \times (H \times w)}$ ， $r$ 代表通道下采样的比例系数，具体公式为

$$f = \delta(F_1([z^h, z^w])) \quad (4)$$

式中： $[..]$ 为沿空间维度的拼接操作； $\delta$ 为非线性激活函数。

然后，沿着空间维数将 $f$ 拆分为2个单独的张量 $f^h \in \mathbb{R}^{C/r \times H}$ 和 $f^w \in \mathbb{R}^{C/r \times W}$ ，再利用另外2个 $1 \times 1$ 卷积变换 $F_h$ 和 $F_w$ 以及 $\sigma$ 变换(sigmoid激活函数)得到2个独立空间方向的注意力权值 $g^h$ 、 $g^w$ 。最后，对其进行扩展并作用于输入端，得到最后的输出 $y_c$ 。具体公式为

$$g^h = \sigma(F_h(f^h)) \quad (5)$$

$$g^w = \sigma(F_w(f^w)) \quad (6)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

## 2.5 损失函数优化

在原来的YOLOX网络模型中，经历了第三部分的SimOTA标签分配之后，就将目标框与正样本预测框进行对比，计算两者误差，即Loss函数。YOLOX-s的损失函数包括3类：检测框位置损失 $l_{loU}$ 、类别损失 $l_{cls}$ 、置信度损失 $l_{obj}$ 。

计算检测框位置：

$$I_{GIoU} = 1 - \left( \frac{|A \cap B|}{|A \cup B|} - \frac{|C - (A \cup B)|}{|C|} \right) \quad (8)$$

式中： $A$ 为预测框； $B$ 为真实框； $C$ 为同时包含真实框与预测框的最小面积。

然而，置信度损失包含图像的前景和背景信息。为了提高小目标的检测效果，本文用Focal Loss<sup>[23]</sup>方法替换原来的二元交叉熵损失函数BCEWithLogitsLoss。BCEWithLogitsLoss将sigmoid与BCELoss<sup>[24]</sup>合成一步，计算公式与BCELoss无差别，定义为

$$l_{BCE} = -y \cdot \log_e(p(y)) - (1-y) \cdot \log_e(1-p(y)) = \begin{cases} -\log_e(p(y)), & y=1 \\ -\log_e(1-p(y)), & y=0 \end{cases} \quad (9)$$

式中： $y$ 为二元标签0(分类为负样本)或者1(分类为正样本)； $p(y)$ 为输出属于 $y$ 标签的概率。二元交叉熵损失函数对于所有的样本都采用同样的损失度量方式，对于分类不准确的样本，例如，难以预测的小目标样本，BCELoss损失函数没有使模型很好的提高对小目标的预测精度。Focal Loss的具体形式为

$$l_{foc} = \begin{cases} -(1-p(y))^\gamma \cdot \log_e(p(y)), & y=1 \\ -p(y)^\gamma \cdot \log_e(1-p(y)), & y=0 \end{cases} \quad (10)$$

也可将Focal Loss表达式统一为一个表达式：

$$l_{foc} = -(1-p_t)^\gamma \cdot \log_e(p_t) \quad (11)$$

$$p_t = \begin{cases} p(y), & y=1 \\ 1-p(y), & \text{otherwise} \end{cases}$$

式中： $p_t$ 为与ground truth即目标 $y$ 的接近程度， $p_t$ 越大说明越接近类别 $y$ ，即分类越准确。对比式(9)(11)，Focal Loss相比二元交叉熵函数多了1个



modulating factor 即  $(1-p_t)^\gamma$ 。Focal Loss 使用 1 个可调节因子  $\gamma$  来控制模型对分类不准确样本的损失权重, 当预测结果  $p(y)$  趋近于 1 时, 在经过 modulating factor 的计算后, 损失函数结果会比较小; 当预测结果  $p(y)$  比较小, 则相对而言损失结果比较大。因此, Focal Loss 增加了分类不准确样本在损失函数中的权重, 使得损失函数倾向于预测值偏小的小目标, 从而提高模型对小目标的预测能力。

### 3 实验方法和结果分析

#### 3.1 实验环境

本实验操作系统为 Ubuntu18.04, CPU 为 Intel (R) Xeon(R) Silver 4110 CPU, 显卡为 RTX2080 Ti, 显存为 11 GB, Python3.8, CUDA 为 11.1 版本, 使用 PyTorch1.8.1 深度学习框架。

#### 3.2 实验数据集

本实验使用伯克利发布的开放式驾驶数据集 BDD100k。此数据集包含了 10 万张图片, 其中, GT 框标签共有 10 个类别, 分别为 Bus、Light、Sign、Person、Bike、Truck、Motor、Car、Train、Rider。由于实验主要对车辆目标进行检测, 因此, 只检测其中 3 类, 分别是 Car、Bus、Truck。训练集、测试集和验证集的比例为 7:2:1。除此之外, BDD100k 的标签是由 Scalabel 生成的 JSON 格式, 所以还需进行数据预处理, 即将标签转换为 coco 格式。

#### 3.3 评价指标

为衡量模型在测试集上的性能, 从不同方面评价所提方法对车辆目标检测的有效性、实时性, 主要选取了单个目标类别检测精度(AP)、平均精确率均值(mAP)和检测速度(FPS)作为评价指标。

$$P = P_t / (P_t + P_f) \quad (12)$$

式中:  $P$  为准确率, 代表在预测为正样本中实际为正样本的概率;  $P_t$  为被正确分类的正样本;  $P_f$  为被错误分类的正样本。

$$R = P_t / (P_t + N_f) \quad (13)$$

式中:  $R$  为召回率, 代表在实际为正样本中预测为正样本的概率;  $N_f$  为被错误分类的负样本。

$$P_a = \int_0^{\text{recall}} PdR \quad (14)$$

式中:  $P_a$  为检测精度, 是计算某一类别 P-R 曲线与坐标轴的面积。

$$P_{ma} = \frac{1}{C} \sum_{i=1}^C P_a(i) \quad (15)$$

式中:  $P_{ma}$  为平均精确率均值;  $C$  为类别数;  $P_a(i)$  为第  $i$  个类别的精确率。

#### 3.4 实验结果与分析

模型训练参数设置: 图像尺寸调节为 640 像素 × 640 像素; 单次训练样本数量设定为 10, Epoch 设置为 100 轮。学习率调整方式为余弦退火衰减 (cosine annealing scheduler), 初始学习率为 0.001, 动量为 0.93。模型训练 Loss 曲线如图 8 所示。

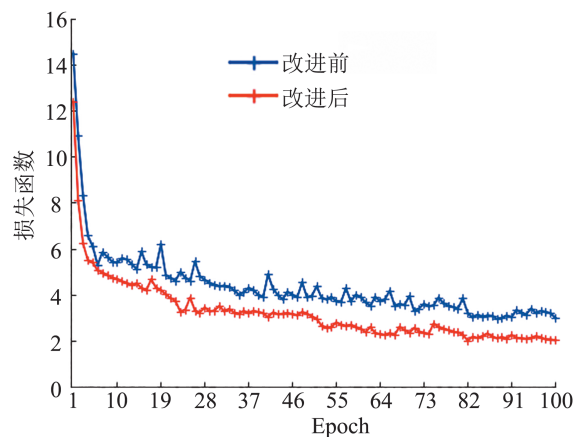


图8 网络模型损失值的对比

Fig. 8 Comparison of loss values for network models

由图 8 的损失曲线可得, 随着模型训练次数不断增加, 模型的 Loss 值逐步下降。本文算法模型的 Loss 曲线相比原模型的 Loss 曲线收敛速度更快、损失值更低, 表明本文提出的改进策略及参数设置合理, 对提升模型识别精度有效。

在 BDD100k 数据集上对各改进策略进行训练和评估, 不同改进策略对基准模型的影响如表 1 所示。实验结果分析可得, 将主干网络与中部网络进

行多尺度融合后，网络特征提取能力增强，提高了网络检测精度；将CSP结构修改成Ghost-CSP后，模型能够在检测效果基本不变的情况下，参数和检测速度都有所提升；将SPP结构中的最大池化替换成SoftPool，且在模型加入坐标注意力机制，网络能更好地识别和定位目标信息，一定程度上改善了小目标漏检情况；对损失函数进行优化后，增加模型对前景正样本目标的学习，检测精度也有一定的提高。

在不同网络模型性能对比中，改进后的YOLOX-s相比其他一阶段检测算法YOLOv3、

YOLOv4和YOLOv5s，平均检测精度分别高出13.42%、9.98%、7.27%，并且在检测速度方面也高于这些算法，具体如表2所示。改进后模型的mAP比原YOLOX-s模型提高了4.02%，达到了74.96%，虽然FPS降低到73帧/s，但相比其他主流大型网络算法依然具备较高的识别率和检测实时性。

为了更直观地了解改进后算法的检测效果，将同一张图像分别在YOLOX-s网络和改进后的网络中进行检测，检测图如图9所示。可以看出改进后的网络对小目标、被遮挡目标的检测效果更好，缓解了漏检和误检情况，精度也有一定的提升。

表1 消融实验  
Table 1 Ablation experiments

Multi-scale fusion	Ghost-CSP	SoftPool	CA	Focal Loss	mAP50/%	FPS/(帧/s)	模型大小/MB
—	—	—	—	—	70.94	77	34.6
√	—	—	—	—	72.63	72	35.5
—	√	—	—	—	71.08	79	30.8
—	—	√	√	—	72.96	73	36.9
—	—	—	—	√	71.11	77	35.0
√	√	√	√	√	74.96	73	34.1

表2 不同网络模型性能的对比  
Table 2 Performance comparison of different network models

Network model	AP50/%			召回/%	mAP/%	FPS/(帧/s)
	Car	Bus	Truck			
YOLOv3	68.32	57.83	58.46	55.26	61.54	27
YOLOv4	72.23	60.78	61.94	59.08	64.98	48
YOLOv5s	77.31	62.67	63.08	61.73	67.69	69
YOLOX-s	83.12	64.43	65.26	63.44	70.94	77
Improved YOLOX-s	88.50	67.71	68.67	66.18	74.96	73

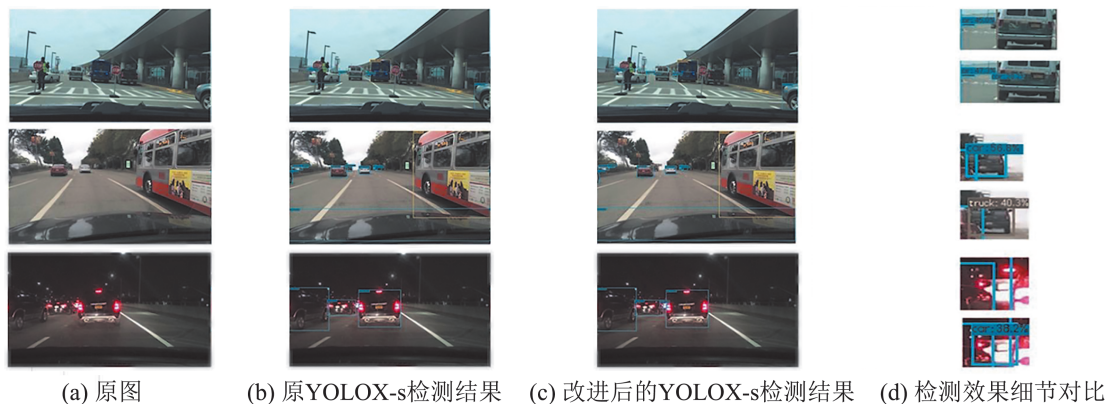


图9 检测效果对比

Fig. 9 Comparison of detection effect

<http://www.china-simulation.com>

## 4 结论

针对深度学习模型在自动驾驶车辆目标检测中存在的不足, 基于YOLOX算法, 提出了多尺度特征融合算法。从特征学习、特征映射和信息传输方式、特征表达能力、样本权重分配的角度, 分别提出多尺度特征融合, 改进CSP结构和SPP结构, 添加Coordinate Attention机制模块, 优化损失函数的改进策略, 提高了对自动驾驶中车辆目标检测效果。实验结果表明: 与原始YOLOX和目前一些主流算法相比, 本文提出的改进算法检测精度和实时检测速度分别达到74.96%和73帧/s。虽然改进策略有效提升了算法的精度, 但是检测速度有所下降, 在未来的工作中, 将继续研究并不断优化改进方法, 使网络具备更优的识别率和实时性。

## 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]. 北京: 北京交通大学, 2020.  
Sang Zhen. Study on Distance and Velocity Measurement Method of Front Vehicle Based on Monocular Vision[D]. Beijing: Beijing Jiaotong University, 2020.
- [2] Chen Duanyu, Lin Yuhao, Peng Yangjie. Nighttime Brake-light Detection by Nakagami Imaging[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4): 1627-1637.
- [3] 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.  
Zhu Yingkai, Luo Wenguang, Bin Yang. Real-time Detection of Front Moving Vehicles Based on Improved Vehicle Shadow Extraction Algorithm[J]. Application of Electronic Technique, 2018, 44(4): 86-89, 98.
- [4] 陈泽, 叶学义, 钱丁炜, 等. 基于改进Faster R-CNN的小尺度行人检测[J]. 计算机工程, 2020, 46(9): 226-232, 241.  
Chen Ze, Ye Xueyi, Qian Dingwei, et al. Small-scale Pedestrian Detection Based on Improved Faster R-CNN[J]. Computer Engineering, 2020, 46(9): 226-232, 241.
- [5] 茅智慧, 朱佳利, 吴鑫, 等. 基于YOLO的自动驾驶目标检测研究综述[J]. 计算机工程与应用, 2022, 58(15): 68-77.  
Mao Zhihui, Zhu Jiali, Wu Xin, et al. Review of YOLO Based Target Detection for Autonomous Driving[J]. Computer Engineering and Applications, 2022, 58(15): 68-77.
- [6] Gao Xue, Han Siyi, Luo Cong. A Detection and Verification Model Based on SSD and Encoder-decoder Network for Scene Text Detection[J]. IEEE Access, 2019, 7: 71299-71310.
- [7] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-time Object Detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.
- [8] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6517-6525.
- [9] Mao Qichao, Sun Hongmei, Liu Yanbo, et al. Mini-YOLOv3: Real-time Object Detector for Embedded Applications[J]. IEEE Access, 2019, 7: 133529-133538.
- [10] 於积荣, 黄德启, 曾蓉, 等. 基于改进YOLOv4的车型检测算法研究[J]. 激光杂志, 2022, 43(4): 52-59.  
Yu Jirong, Huang Deqi, Zeng Rong, et al. Research on Vehicle Detection Algorithm Based on Improved YOLOv4[J]. Laser Journal, 2022, 43(4): 52-59.
- [11] 金立生, 郭柏苍, 王芳荣, 等. 基于改进YOLOv3的车辆前方动态多目标检测算法[J]. 吉林大学学报(工学版), 2021, 51(4): 1427-1436.  
Jin Lisheng, Guo Baicang, Wang Fangrong, et al. Dynamic Multiple Object Detection Algorithm for Vehicle Forward Based on Improved YOLOv3[J]. Journal of Jilin University(Engineering and Technology Edition), 2021, 51(4): 1427-1436.
- [12] 柳长源, 王琪, 毕晓君. 多目标小尺度车辆目标检测方法[J]. 控制与决策, 2021, 36(11): 2707-2712.  
Liu Changyuan, Wang Qi, Bi Xiaojun. Multi-target and Small-scale Vehicle Target Detection Method[J]. Control and Decision, 2021, 36(11): 2707-2712.
- [13] 赵家琪, 高贵, 黄晓峰, 等. 一种改进的YOLOv4-tiny车辆目标检测方法[J]. 电子产品世界, 2021, 28(9): 39-43.  
Zhao Jiaqi, Gao Gui, Huang Xiaofeng, et al. An Improved YOLOv4-tiny Vehicle Target Detection Method[J]. Electronic Engineering & Product World, 2021, 28(9): 39-43.
- [14] Ge Zheng, Liu Songtao, Wang Feng, et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. (2021-08-06) [2022-03-17]. <https://arxiv.org/abs/2107.08430>.
- [15] Wang Ling, Zhou Hai, Bian Chunjiang, et al. Hardware Acceleration and Implementation of YOLOX-s for On-Orbit FPGA[J]. Electronics, 2022, 11(21): 3473.
- [16] 熊怡梦. 基于标签语义和Transformer的元学习小样本目标检测方法研究[D]. 西安: 西安电子科技大学, 2021.  
Xiong Yimeng. Label Semantics and Transformer for

- Meta Learning Few-shot Object Detection[D]. Xi'an: Xidian University, 2021.
- [17] Mercedes E Paoletti, Juan M Haut, Nuno S Pereira, et al. Ghostnet for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10378-10393.
- [18] 史芳菲. 监控视频下运营车辆司机吸烟行为检测系统的设计与实现[D]. 海口: 海南大学, 2021.  
Shi Fangfei. Design and Implementation of a Smoking Detection System for Operating Vehicle Drivers Under Surveillance Video[D]. Haikou: Hainan University, 2021.
- [19] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [20] Wang Chunyuan, Wu Yang, Wang Yihan, et al. Scene Recognition Using Deep Softpool Capsule Network Based on Residual Diverse Branch Block[J]. Sensors, 2021, 21(16): 5575.
- [21] 曹楨淼, 吉卫喜, 苏璇, 等. 基于扩展跨阶段局部网络的表面缺陷检测[J]. 计算机集成制造系统, 2022, 28(8): 2399-2407.  
Cao Zhenmiao, Ji Weixi, Su Xuan, et al. Surface Defect Detection Based on Scaling Cross-stage Partial Network[J]. Computer Integrated Manufacturing Systems, 2022, 28(8): 2399-2407.
- [22] Xie Chao, Zhu Hongyu, Fei Yeqi. Deep Coordinate Attention Network for Single Image Super-resolution[J]. IET Image Processing, 2021, 16(1): 273-284.
- [23] Liu Zechuan, Wang Song. Broken Corn Detection Based on an Adjusted YOLO with Focal Loss[J]. IEEE Access, 2019, 7: 68281-68289.
- [24] 史彩娟, 涂冬景, 刘靖祎. Re-GAN: 残差生成式对抗网络算法[J]. 中国图象图形学报, 2021, 26(3): 594-604.  
Shi Caijuan, Tu Dongjing, Liu Jingyi. Re-GAN: Residual Generative Adversarial Network Algorithm[J]. Journal of Image and Graphics, 2021, 26(3): 594-604.