

# 基于 YOLOX 的车辆检测<sup>①</sup>



赵帅豪

(西安石油大学 电子工程学院, 西安 710065)  
通信作者: 赵帅豪, E-mail: zhaoshuaihao01@163.com

**摘要:** 车辆检测是智能交通系统重要的一个研究方向. 针对监控视角下的车辆检测问题, 提出了一种改进 YOLOX 算法的车辆检测方法. 使用网络深度更小的 YOLOX\_S 模型, 对网络结构改进. 使用 GHOST 深度可分离卷积模块代替部分传统卷积, 在保证模型检测精度的同时减少模型参数; 将 CBAM 注意力模块融合到特征提取网络中, 并添加特征增强结构, 加强特征提取网络获得的特征图语义信息, 增强提取网络对目标的检测能力; 通过使用 CIoU\_loss 优化损失函数, 提高模型边界框的定位精度. 测试实验结果表明, 改进后的网络识别准确率提升了 2.01%, 达到 95.45%, 证明了改进方法的可行性.

**关键词:** 目标检测; YOLOX; 注意力机制; 深度分离卷积; 特征融合

引用格式: 赵帅豪. 基于 YOLOX 的车辆检测. 计算机系统应用, 2022, 31(12): 195-202. <http://www.c-s-a.org.cn/1003-3254/8806.html>

## Vehicle Detection Based on YOLOX

ZHAO Shuai-Hao

(School of Electronic Engineering, Xi'an Shiyou University, Xi'an 710065, China)

**Abstract:** Vehicle detection is an important research direction for intelligent transportation systems. In terms of vehicle detection from the monitoring perspective, a vehicle detection method based on an improved YOLOX algorithm is proposed. The YOLOX\_S model with a smaller network depth is used to improve the network structure. The GHOST depthwise separable convolution module is adopted to replace some traditional convolutions, and model parameters are reduced with the model detection accuracy ensured. The CBAM attention module is integrated into a feature extraction network, and a feature enhancement structure is added to enhance the semantic information of feature maps obtained by the network and strengthen the ability of the network in detecting targets. By using the CIoU\_loss to optimize the loss function, this study finds that the positioning accuracy of the bounding box of the model is improved. The test results show that the detection accuracy of the improved network is increased by 2.01%, reaching 95.45%, which proves the feasibility of the improved method.

**Key words:** target detection; YOLOX; attention mechanism; depthwise separable convolution; feature fusion

## 1 引言

随着智能交通系统 (intelligent traffic system) 的发展和人工智能的出现<sup>[1]</sup>, 实时检测道路上正在行驶的车辆变得尤其重要, 车辆检测也是目标检测中的重要对象之一. 目前我国已经建立多个无人驾驶示范区, 大力

研究无人驾驶系统. 通过车辆检测功能可以得到道路上行驶车辆的种类、数量信息, 在无人车的车速控制、安全驾驶、路径规划等方面起到关键作用, 这也使得车辆检测有更高的准确度要求<sup>[2]</sup>, 尤其是在雨雪恶劣天气下的车辆检测任务.

<sup>①</sup> 基金项目: 西安石油大学研究生创新与实践能力的培养计划 (YCS21213212)

收稿时间: 2022-03-15; 修改时间: 2022-04-12; 采用时间: 2022-04-20; csa 在线出版时间: 2022-09-14

目标检测算法领域内研究方向分为传统目标检测算法和基于深度学习的目标检测算法<sup>[3]</sup>。传统目标检测算法主要包括区域选择、提取特征和分类3个部分,过程中需要大量人工设计提取特征,面对多样性的目标时效果较差。

基于深度学习的目标检测算法无需人工设计提取特征,泛化能力强,鲁棒性好检测速度更快,包括两阶段(two stage)目标检测算法和单阶段(one stage)目标检测算法。2013年, Girshick 提出了具有里程碑意义的两阶段目标检测算法 R-CNN<sup>[4]</sup>,没有采用滑动窗口技术,使用基于启发式的区域提取方法选择候选框,重复合并两个可能性最大的小区域,直到将图像合并为一个区域后输出候选框。然后引入 CNN 操作提取候选区域特征,将提取特征融合输出并用 SVM 进行分类,最后通过边界回归框获得精确的区域信息。R-CNN 提高了物体的检测准确度,但提取特征过程花费时间较长、数据量大因此效率较低。

Redmon 等人提出了单阶段目标检测 YOLO 算法<sup>[5]</sup>。与 R-CNN 系列算法中分别提取目标定位和目标分类不同,而是使用回归思想同时完成分类和定位任务,并且没有预选框结构检测速度更快,但精度相比于 R-CNN 系列并没有优势。YOLO 算法为学者们提供了新的思想。Liu 等人借鉴了 YOLO 的回归思想和 Faster R-CNN 的 Anchor 机制提出了 SSD 目标检测模型<sup>[6]</sup>。SSD 算法提取了不同尺度的特征图进行检测,并采用卷积核预测一系列边界框的类别分数和偏移量,使得网络在保证检测高准确度的同时,提高了检测效率。YOLO 作者将 YOLO 算法改进后提出了 YOLOv2 算法<sup>[7]</sup>,YOLOv2 用检测数据集数据学习检测目标物体的准确位置,学习分类数据集数据提高检测物体种类,可检测对象种类达到 9 000 种,并加入了多尺度图像训练,使网络在检测不同输入尺寸的图片时都有一个较高的准确率。同时使用锚点框来预测边界框,大幅提升了召回率。之后作者再改进后提出了 YOLOv3<sup>[8]</sup>。加入了多尺度预测特征金字塔(FPN)、使用加入残差结构的 Darknet-53 网络和二分类交叉熵代替 Softmax 分类器,提高了检测精度和检测较小目标的能力,并且大幅提高了检测速度。

随后 Bochkovskiy 等人提出了 YOLOv4 算法<sup>[9]</sup>,YOLOv4 加入了大量改进措施,与 YOLOv3 比较,在输入端加入了 CutMix 和 Mosaic 数据增强,DropBlock 正

则化和 Class label smoothing,对 Darknet-53 改进得到 CSPDarknet53 主干网络、Neck 部分添加 SPP 和 PAN 结构。YOLOv4 与以往的目标检测相比,检测速度不落后而识别准确度方面有大幅提升。

旷视科技在 2021 年根据 YOLO 系列算法的经验提出了最新的 YOLOX 算法<sup>[10]</sup>,选择 Anchor-free 结构,无须手动设计 anchor,集成了 Focus、分离式的解耦头结构、数据增强、标签分配等机制,最终使 YOLO 算法的检测性能获得了大幅提升。

本文提出一种改进的 YOLOX-S 模型。基于 YOLOX 模型,通过 GHOST 操作代替部分卷积操作生成特征图,以实现减小网络参数量;在 Neck 结构中添加 CBAM 注意力机制;通过不同感受野提取一张图片的信息实现强化特征图信息;使用 CIoU 作为边界框回归的损失函数。对不同道路情况监控视角下的车辆进行识别。

## 2 YOLOX 原理

YOLOX 算法是由旷视科技发布的单阶段目标检测算法,结构如图 1 所示,分为输入部分、Backbone 特征提取、Neck 特征融合、Prediction 四部分。

### 2.1 输入端

输入端包括 Mosaic 数据增强、Mixup 数据增强、Focus 结构。Mosaic 数据增强每次读取 4 张图片,然后分别对 4 张图片进行缩放、翻转等操作,最后组合在一张图片上,达到丰富图片的背景;MixUp 是在 Mosaic 基础上改变的增强策略,MixUp 把 2 张图片按一定的融合系数融合在一起,功能和 Mosaic 一样;Focus 把一张图片中每隔一个像素拿一个值,切片获得 4 个独立的特征层,然后堆叠使通道扩充至 4 倍。

### 2.2 Backbone

Backbone 是 YOLOX 的主要结构,YOLOX 所使用的主干特征提取网络为 CSPDarknet,它具有多个重要特点:残差网络(Residual)缓解深度神经网络中梯度消失的问题;CSPBlock 可以提高卷积网络的计算能力和学习能力并减少计算量;使用 SiLU 激活函数,SiLU 是 Sigmoid 和 ReLU 的改进版,具有平滑、无上界由下界和非单调的特性,在深层网络中效果比 ReLU 更好。

### 2.3 Neck

Neck 特征融合结构的核心是特征金字塔(feature pyramid networks,FPN)和路径聚合(path aggregation

networks, PAN) 结构<sup>[11,12]</sup>. 随着卷积层数的加深, 高层检测大目标的类别特征丰富; 而低层检测大目标的位置和小目标的类别特征更优, 常规的FPN层与低层向

上层的PAN进行结合, 融合高低层的类别特征和位置特征信息, 并与主干网络的检测层特征融合, 增强网络特征提取能力.

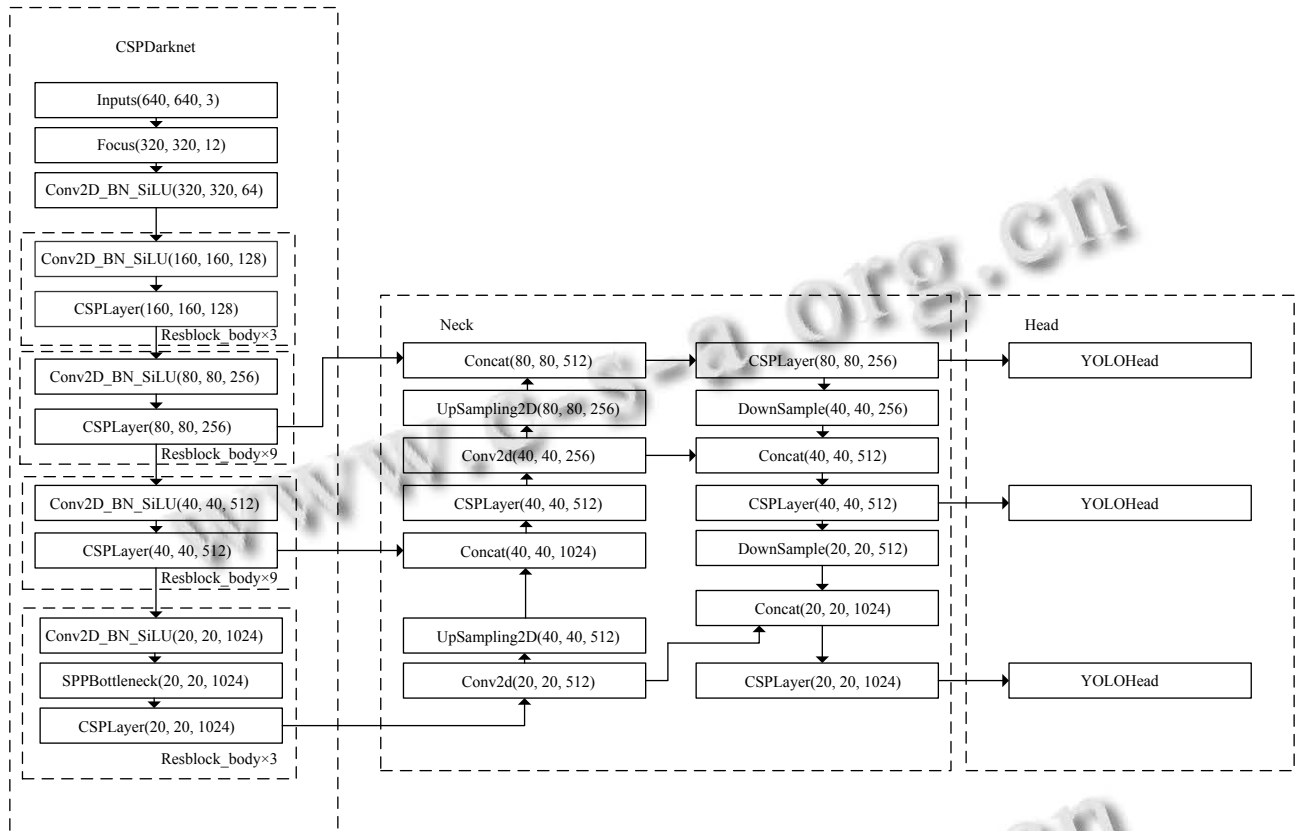


图1 YOLOX 结构

### 2.4 Prediction

Predicton 是 YOLOX 网络的分类器和回归器. 通过 Neck 得到 3 个有效特征层, 分别针对大、中、小型目标进行识别. 每一个特征层可以看作是特征点的集合, 每个特征点都有位置参数和通道数. Prediction 判断特征点是否有物体与其对应. Prediction 主要由 Decoupled Head、Anchor Free、SimOTA、LOSS 构成.

YOLOX 的 YOLOHead 与之前 YOLO 算法的 YOLOHead 不同, 此前版本的解耦分类和回归在一个卷积内实现. 在 YOLOX 中的 YOLOHead 解耦分为两部分分别实现, 如图 2, 第 1 部分是 class\_output, 是对目标框的类别预测的分数; 第 2 部分是 obj\_output 和 reg\_output, 分别是判断目标框内是否由物体和目标框的坐标信息. 最后整合在一起得到特征信息.

YOLOX 中使用 Anchor-free 结构. 在此前的 YOLO

算法中使用的是 Anchor-Based 结构, 参数量只有 Anchor-Based 的 1/3.

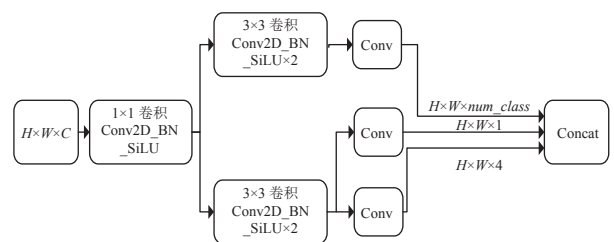


图2 YOLOHead

SimOTA 结构对特征点进行筛选, 得到正样本的候选框. 定义一个 Cost 成本函数来代表每个真实框和描框的关系. 筛选出中心点落在真实框的范围内或者中心点在以真实框中心点为基准设置的正方形范围内的描框, 筛选出的描框作为候选框. 计算每个真实框和

对应描框的位置重合度,将重合度最高的10个候选框与真实框的IoU加起来求得每个真实框的k值,代表每个真实框有k个特征点与之对应.计算每个真实框和当前描框的种类预测准确度.通过位置重合度和种类预测准确度计算Cost成本函数,将Cost最低的k个点作为真实框的正样本.每个目标框和标签分配的正样本计算Loss.

$$C_{ij} = L_{ij}^{cls\_loss} + \gamma \times L_{ij}^{reg\_loss} \quad (1)$$

### 3 改进的YOLOX

#### 3.1 上采样算法改进

在PAN结构中,使用最邻近插值法完成上采样操作,最邻近插值法计算速度快,但是采样效果不是很好,为了减少检测特征在采样过程中的损失,本文选择使用双线性插值法进行上采样.双线性插值法计算4个点作为参考点,增强上采样效果.

#### 3.2 引入CBAM注意力机制

注意力机制(attention mechanism)是机器学习中常用的数据处理方法,已经广泛应用在自然语言处理、语音识别和图像识别等机器学习任务中.在检测网络中加入注意力机制,可以获取有用的信息,使检测网络将更多注意力分配到需要关注的地方<sup>[13]</sup>.注意力机制分为通道注意力机制、空间注意力机制和混合域注意力机制.本文使用的CBAM属于第3种.

如图3,卷积注意力机制(convolutional block attention module, CBAM)<sup>[14,15]</sup>通过通道注意力模块(channel attention module, CAM)和空间注意力模块(spatial attention module, SAM)来分别在通道和空间维度上学习关注什么信息、在哪个位置关注信息,重点关注重要特征并抑制不必要特征.

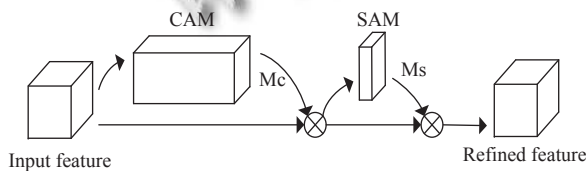


图3 CBAM结构

给定一个特征图  $F \in R^{C \times H \times W}$ , CBAM 首先从  $F$  中推出一维通道注意力  $M_c$ ,  $F$  与  $M_c$  相乘运算后推出二维空间注意力  $M_s$ ,  $F$  再次相乘得到输出特征图,输出特征图与输入特征图  $F$  拥有相同的维度.计算公式为:

$$F' = M_c(F) \otimes F \quad (2)$$

$$F'' = M_s(F') \otimes F' \quad (3)$$

为了聚集空间信息和不同对象的特征线索以获得更精确的通道注意力,同时采用平均池化法和最大池化法对输入特征图的空间维度进行压缩.池化向量经过多层感知器和一个隐含层,MLP输出的特征向量进行逐元素加法运算并通过Sigmoid激活后得到通道注意力  $M_c$ .计算方式为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (4)$$

空间注意力目标是在哪个位置的信息部分,是通道注意力的补充.得到的通道注意力  $M_c$  沿通道轴上应用平均池化和最大池化操作,生成的有效特征向量经过卷积操作和Sigmoid激活之后得到空间注意力  $M_s$ .计算公式为:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (5)$$

本文将注意力机制添加到特征融合结构中.

#### 3.3 改进损失函数

YOLOX网络模型中边界框回归损失函数选择为IoU loss 或者 GIoU loss. GIoU loss 定义如下:

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \quad (6)$$

GIoU loss 使用相交尺度衡量方式,  $C$  是包含  $A$  和  $B$  的最小矩形框,  $IoU=0$  时,  $A$  和  $B$  距离很远,  $GIoU$  趋于-1,解决了边界框不重合时造成损失函数不可导,和两个预测框大小相同时  $IoU$  也相同,出现  $IoU$  loss 不能正确反映相交情况的问题.但是如果出现多个大小相同的预测框在目标框内部时,每个最小矩形框  $C$  相同,  $A$  与  $B$  并集也相同,此时  $GIoU$  loss 不能正确反映预测框的相交情况.

本文采用 CIoU loss 作为边界框的回归损失函数,解决了  $GIoU$  退化为  $IoU$  的问题<sup>[16]</sup>. CIoU loss 定义如下:

$$\begin{cases} CIoU = 1 - IoU + \frac{\rho^2(b_1 + b_2)}{c^2} + \alpha v \\ \alpha = \frac{v}{(1 - IoU) + v} \\ v = \frac{4}{\pi^2} \left( \arctan \frac{w^{b_1}}{h^{b_1}} - \arctan \frac{w^{b_2}}{h^{b_2}} \right)^2 \end{cases} \quad (7)$$

其中,  $b_1$  为预测框中心点,  $b_2$  为目标框中心点,  $c$  是覆盖两个框的最小方框的对角线距离.  $CIoU$  loss 中考虑了预测框与目标框中心点的欧式距离, 预测框在目标框的内部时, 预测框位置不同, 其中心点位置也不同,  $CIoU$  loss 可以计算到.  $w^{b_1}, h^{b_1}, w^{b_2}, h^{b_2}$  是预测框和目标框的宽高, 当两个框中心点也重合的时候, 通过  $v$  也考虑到了预测框和目标框的长宽比, 使定位框时更加精准, 提高准确度.

### 3.4 GHOST 模块代替卷积

如图 4, GHOST 模块的功能是对其中一个特征图通过更轻量化的操作生成另外的特征图, 可以认为两个相似的特征图是彼此的 GHOST<sup>[17]</sup>. 具体操作是, 将普通卷积层分为两部分. 第 1 部分仍是普通的卷积层, 但是减少原本设置的卷积的通道数量, 生成输入特征层的特征浓缩图. 第 2 部分对这些特征浓缩图进行深度可分离卷积生成更多特征图. 在不更改输出特征图大小的情况下, 与使用普通卷积神经网络生成所有特征图的运算量相比, GHOST 模块中运算所需的参数总数和计算复杂度大幅降低.

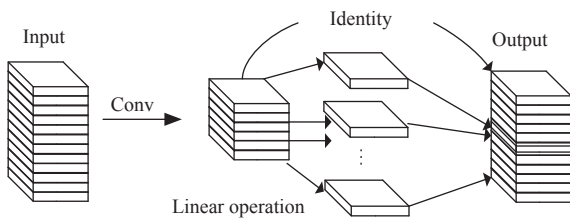


图 4 GHOST 结构

使用原始卷积神经网络时, 给定输入数据  $X \in R^{L \times H \times W}$ ,  $L$  表示输入通道个数,  $H$  和  $W$  是输入数据的高度和宽度; 输出数据  $Y \in R^{M \times H' \times W'}$ ,  $M$  是输出通道个数,  $H'$  和  $W'$  是输出数据的高度和宽度;  $f \in R^{L \times k \times k \times M}$  是这一层的卷积核,  $k \times k$  是卷积核的大小. 用于生成特征图的卷积运算表示为 (为表示方便, 省略偏差项):

$$Y = f \times X$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1L} \\ f_{21} & f_{22} & \cdots & f_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{ML} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_L \end{bmatrix} \quad (8)$$

GHOST 模块操作是, 首先通过一次卷积生成  $m$  个

原始特征图  $Y' \in R^{m \times h' \times w'}$ ,  $m \leq M$ .  $f' \in R^{L \times k \times k \times m}$  是使用的卷积核, 其他超参数和原始普通卷积保持一致. 计算公式为:

$$Y' = f' \times X \quad (9)$$

原始特征图能提供大部分的特征信息, 对原始特征图应用运算得到 GHOST 特征图. 计算公式为:

$$y''_{ij} = \phi_{ij}(y'_i), \forall i = 1, 2, \dots, m, j = 1, 2, \dots, s \quad (10)$$

其中,  $y'_i$  是  $Y'$  中的第  $i$  个原始特征图,  $\Phi_{ij}$  是对第  $i$  个原始特征图进行第  $j$  个线性运算, 得到第  $j$  个 GHOST 特征图  $y''_{ij}$ . 通过操作后得到  $m \times s$  个特征图, 表示为:

$$Y''_{ij} = [y_{i1}, y_{i2}, \dots, y_{ims}] \quad (11)$$

在本文中, GHOST 模块添加到 Bottleneck 模块中代替  $1 \times 1$  和  $3 \times 3$  普通卷积,  $m = M/2, j = 1, \Phi_i$  是  $m$  通道上  $3 \times 3$  卷积核, 通过  $m$  个原始特征图得到  $m$  个 GHOST 特征图, 最后把原始特征图和 GHOST 特征图拼接得到输出特征图.

### 3.5 Enhance feature

Enhance feature 结构利用多分支扩张卷积增加模型的感受野 (reception field), 学习每个特征图中的不同接收域, 从而提高多尺度目标检测识别的精度<sup>[18]</sup>. 结构如图 5.

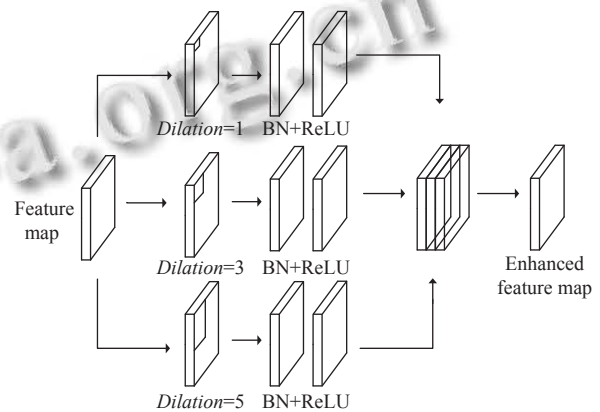


图 5 Enhance feature 结构

多分支卷积层通过扩展卷积为输入特征图提供不同大小的接收域, 然后将接收的信息融合相加, 得到丰富特征信息的特征图. 多分支卷积层由扩张卷积层、BN 层和 ReLU 激活层组成. 3 个并行分支的扩张卷积卷积核为  $3 \times 3$ , 扩张速率分别为 1、3、5. 具体网络结构如图 6.

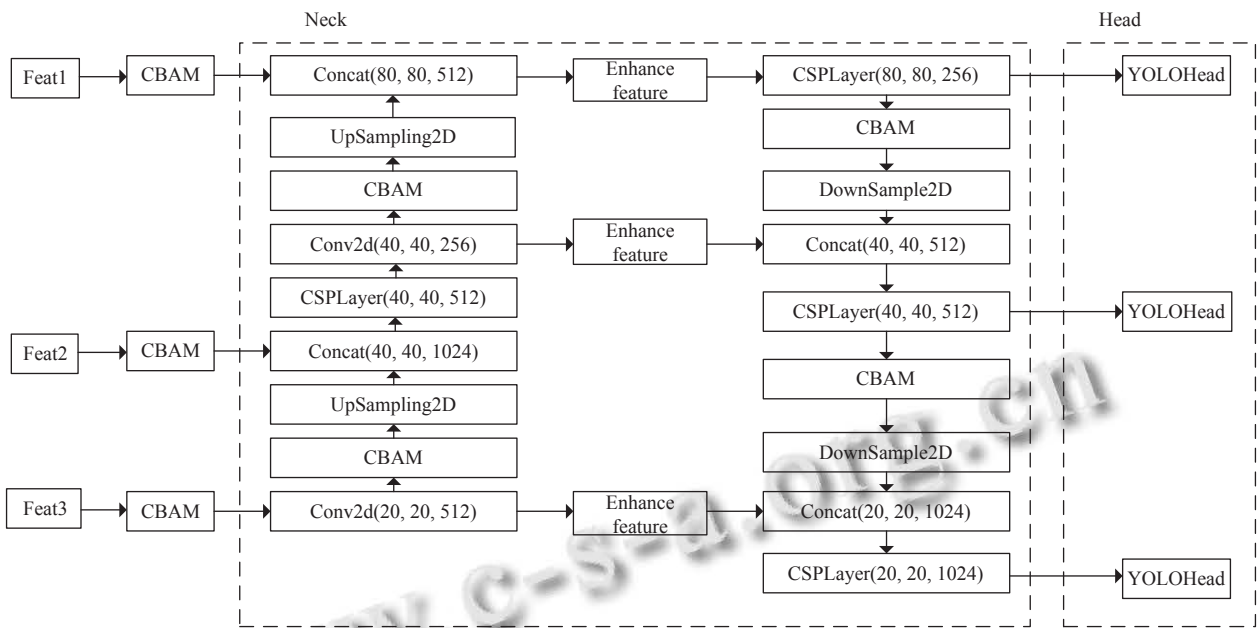


图6 改进后的特征网络结构

## 4 实验

### 4.1 实验平台

本文实验环境是 Ubuntu 18.04.5 系统, 64 GB 内存, CPU: E5-1650 V3 @3.5 GHz, RTX2080TI 显卡, 在 PyTorch 1.8.1、CUDA 10.2 下测试.

### 4.2 数据集

本文的数据来源于西安市内多条道路, 在道路不同位置高点使用高清摄像头拍摄获得, 模拟多种位置摄像机的拍摄角度. 数据集中的图片经过数据增强处理, 模拟恶劣天气下的车辆检测情况, 增强后的天气效果如图7所示.



图7 数据集样本

数据图片整理后使用 labelingm 对图片中的车辆进行标注, 标注 car 标签 15395 个, bus 标签 1432 个. 标

注内容包含目标物体的类别和目标框的坐标, 标注信息存放在 XML 格式文件中.

### 4.3 训练设置

为了保证实验数据的可靠性, 训练最大轮次 (epoch) 设置为 300 次, 前 50 次为冻结主干网络参数训练, 初始学习率设置为 1E-3, 后 250 次为解冻主干网络训练, 初始学习率设置为 1E-4.

### 4.4 衡量指标

本文选用  $mAP@0.5$ 、视频帧率 FPS 作为模型评估指标,  $mAP@0.5$  评价指标与混淆矩阵 (confusion matrix)、精确率、召回率有关. 测试结果可划分为 TP、TN、FP、FN. 划分方式如表1. 精确率 (precision, P) 衡量算法错检程度, 定义为所有预测值为正样本的目标中真实值为正样本的样本比例:

$$P = \frac{TP}{TP + FP} \quad (12)$$

表1 衡量指标

预测值	真实值	
	正样本	负样本
正样本	TP	FP
负样本	FN	TN

召回率 (recall, R) 衡量算法漏检程度, 定义为真实值为正样本的目标中预测值为正样本的样本比例:

$$R = \frac{TP}{TP + FN} \quad (13)$$

$AP@0.5$  定义为混淆矩阵 IoU 阈值取 0.5 时, 对其中一类的  $n$  个正样本的  $P$  值求平均计算, 计算公式是:

$$AP = \frac{1}{n} \sum_{i=1}^n P_i \quad (14)$$

$mAP@0.5$  为  $N$  类的  $AP@0.5$  求均值, 计算公式为:

$$mAP = \frac{\sum AP}{N} \quad (15)$$

#### 4.5 结果分析

为了分析本文提出的改进方法对 YOLOX 算法性能的影响, 分 4 次实验验证不同改进方法的效果, 测试结果如表 2 所示, “×”代表在算法模型中没有使用对应的改进方法, “√”代表在算法模型中使用了对应的改进方法. 由表中数据可以看到, 使用 GHOST 结构代替 Bottleneck 中的卷积结构, 可以保证在  $mAP$  有略微提升的同时减小了一定的模型参数量, 也略微提升了 FPS 值. 边界框回归的损失函数改进为  $CIoU\_loss$ , 解决目标框与预测框重合情况下  $GIoU$  退化为  $IoU$  的问题, 提高模型边界框的定位精度. 如图 8 可以看到, 使用

$CIoU\_loss$ , 使得训练前期损失值很快下降, 但在训练 20 轮左右后, 损失值下降趋于稳定, 在训练 50 轮后又缓慢下降, 最终损失值低于未使用  $CIoU\_loss$  的网络模型. 在为网络添加 CBAM 注意力机制时, 实验过多个位置, 例如时将 CBAM 注意力添加到主干网络、SPP 结构中, 多次实验证明在特征融合结构中添加注意力机制效果最好. CBAM 告诉模型分别在通道和空间维度上学习关注什么、在哪里关注, 关注重要特征并抑制不重要特征, 车辆检测任务较依赖位置信息, 所以添加注意力机制能够实现更好的检测效果. 测试数据表明, 添加注意力机制后, 对 bus 较大型的目标检测率有 0.69% 的提升, 而对 car 小目标车辆的提升为 0.12%.

表 2 测试结果

GHOST	CBAM	Enhanced	模型大小 (M)	$AP$ (%)		$mAP$ (%)	FPS
				Bus	Car		
×	×	×	30.6	93.21	93.64	93.44	35.1
√	×	×	29.6	93.37	93.92	93.59	36.4
√	√	×	30.1	94.06	94.04	94.05	35.4
√	√	√	61.2	95.93	94.97	95.45	19.2

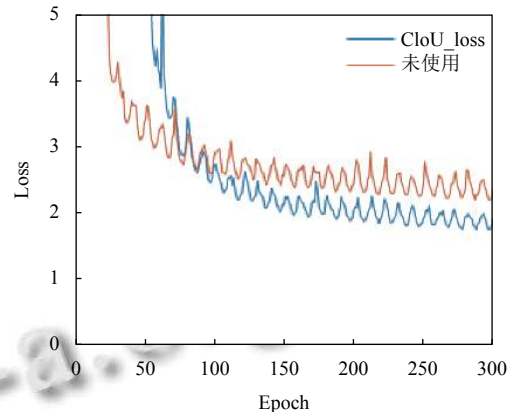
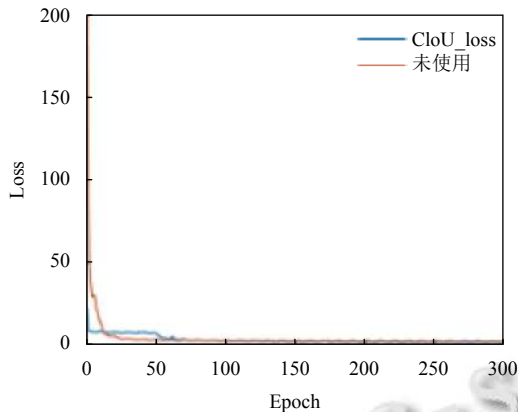


图 8 Loss 曲线

Enhance feature 结构使用不同速率扩张卷积的多分支结构实现更大的感受野, 用来获取多尺度物体信息, 使网络能够学习到每个特征图中不同的接收域信息, 提高了模型特征提取能力和特征融合能力. 测试结果表明, 增加 Enhance feature 结构, 使得 bus 目标检测率提高了 1.87%, car 目标检测率提高了 0.93%, 但是此结构增加了算法模型的参数量, 减慢了一定的运行速度.

图 9 给出了不同改进模型检测效果图, 检测结果中可以看出, 改进后的 YOLOX 检测图中, 目标检测框的位置标记更能准确地包括到整个车辆; 检测到原始 YOLOX 算法漏检到的车辆.

## 5 结论

对于监控视角下的车辆检测问题, 本文基于 YOLOX 模型, 提出一种改进的 YOLOX 网络结构, 采用 GHOST 模型代替部分卷积, 减少一定参数量同时提升  $mAP$ ; 结合 CBAM 注意力模块和特征增强结构对特征融合结构改进, 较大幅度提高了网络性能; 使用  $CIoU\_loss$  作为回归损失函数, 提高回归精度. 实验证明, 改进后的网络模型能实现高精度的检测, 并且检测速度能满足使用, 在智能交通系统中可以提供有效且及时的交通信息. 由于智能交通系统对于传输速度、延迟性等的要求, 下一步改进在保证目

标检测高准确度的基础上,减小网络结构,提高传输速度.

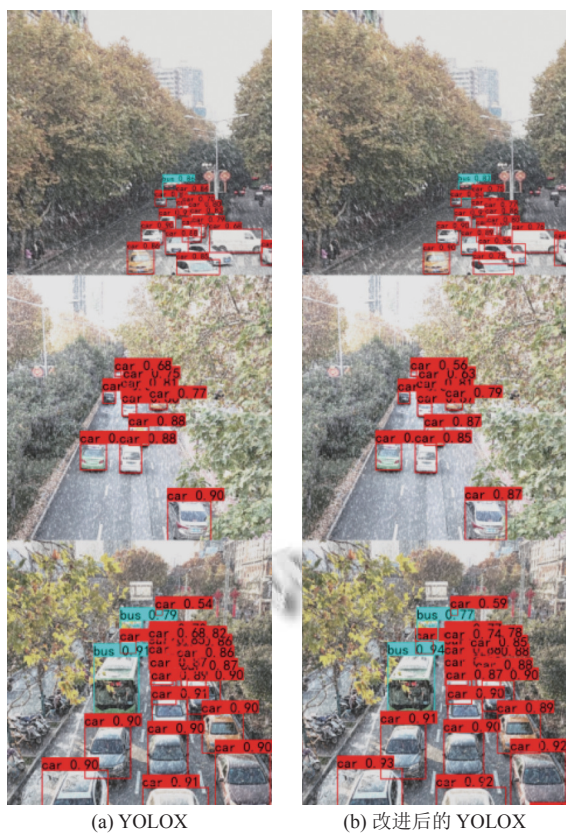


图9 检测结果对比

### 参考文献

- 艾峰. 城市智能交通系统的发展现状与趋势. 智能城市, 2021, 7(20): 141. [doi: 10.19301/j.cnki.zncs.2021.20.065]
- 潘黎琼, 刘翔. 人工智能技术在城市智能交通系统中的应用. 信息记录材料, 2021, 22(10): 136–137. [doi: 10.16009/j.cnki.cn13-1295/tq.2021.10.065]
- 李坤坤, 刘正熙, 熊运余. 基于深度学习的目标检测系统性文献综述. 现代计算机, 2021, (16): 98–102, 117.
- Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934v1, 2020.
- Ge Z, Liu ST, Wang F, *et al.* YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
- Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 936–944.
- Woo S, Hwang S, Kweon IS. StairNet: Top-down semantic aggregation for accurate one shot detection. Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018. 1093–1102.
- Li HC, Xiong PF, An J, *et al.* Pyramid attention network for semantic segmentation. Proceedings of British Machine Vision Conference 2018. Newcastle: BMVA Press, 2018. 285.
- Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
- Zhang ZX, Wang MW. Convolutional neural network with convolutional block attention module for finger vein recognition. arXiv:2202.06673, 2022.
- Zheng ZH, Wang P, Ren DW, *et al.* Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Transactions on Cybernetics, 2022, 52(8): 8574–8586. [doi: 10.1109/TCYB.2021.3095305]
- Han K, Wang YH, Tian Q, *et al.* GhostNet: More features from cheap operations. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1577–1586.
- Wang JF, Chen Y, Gao MY, *et al.* Improved YOLOv5 network for real-time multi-scale traffic sign detection. arXiv:2112.08782, 2021.

(校对责编:牛欣悦)