

改进 YOLOv5s 的交通多目标检测方法^{*}

单慧琳^{1,2} 吕宗奎¹ 付相为¹ 王 煜¹ 张培琰² 孙佳琪²

(1. 南京信息工程大学电子信息工程学院 南京 210044; 2. 无锡学院电子信息工程学院 无锡 214105)

摘要:为了提高交通目标检测的精度和效率,提出一种改进 YOLOv5s 的交通场景多目标检测方法,在 YOLOv5s 的主干网络中引入高效的层聚合网络结构来提高模型学习目标特征的能力,引入了通道注意力和空间注意力结合的卷积注意力模块(CBAM)机制,进一步提高网络模型的特征提取能力,通过采用 α -IoU 作为边界框回归损失函数,提高了边界框回归精度。实验结果表明,改进的目标检测模型相较于 YOLOv5s 原模型在检测精度上提升了 2.4%,模型参数量和模型大小分别降低了 20.9%和 19.1%。实现了在不同时间段准确且高效的检测交通场景的多种目标,保证了实时检测的应用需求。

关键词:目标检测;YOLOv5 算法;注意力机制;损失函数

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

Traffic multi-target detection method of YOLOv5s is improved

Shan Huilin^{1,2} Lv Zongkui¹ Fu Xiangwei¹ Wang Yu¹ Zhang Peiyan² Sun Jiaqi²

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China)

Abstract: In order to improve the accuracy and efficiency of traffic intersection target detection, this paper proposes an improved traffic scene target detection model based on YOLOv5s. An efficient layer aggregation network structure is introduced into the backbone network of YOLOv5s to improve the ability to learn target features. The attention mechanism CBAM of channel attention and spatial attention is introduced to further improve the feature extraction ability of the network model. The α -IoU is used as the bounding box regression loss function to improve the bounding box regression accuracy. The experimental results show that compared with the original YOLOv5s model, the improved object detection model proposed in this paper has the detection accuracy increased by 2.4% and the model parameter number and model size reduced by 20.9% and 19.1% respectively. It realizes all kinds of targets of accurate and efficient detection of traffic intersection scenes in different time periods, and ensures the application requirements of real-time detection.

Keywords: object detection; YOLOv5s; attention mechanism; loss function

0 引言

随着城市化进程进一步加深,城市人口激增,随之而来的城市道路交通问题愈发突出,大量的通行车辆带来了严重的交通拥堵问题,尤其是早晚高峰时段,给人们的上下班通勤带来极大的不便。同时,交通拥堵还可能潜在地导致车辆碰撞等交通事故。作为道路路口的指挥系统,交通信号灯的优化成了缓解交通拥堵的重要环节,优化交

通信号的一个关键步骤是长时间收集道路使用者的信息,包括机动车道上行驶的车辆、人行道上的行人、非机动车道上的自行车、电动车等目标。与此同时,对交通十字路口场景中的各种车辆和行人目标检测,也是实现智能驾驶、智能交通等应用场景的基础^[1]。道路目标检测算法的性能影响着无人驾驶的安全性、舒适性和硬件成本,这就要求检测系统具有高准确性和实时性。因此,基于计算机视觉技术的道路路口情况实时状况检测对于交通监控系

收稿日期:2023-02-06

^{*} 基金项目:国家自然科学基金(62071240,62106111)项目资助

统来说意义重大。

基于深度学习的目标检测常用算法主要分为两类。一类是以基于区域的卷积神经网络(region based convolutional neural network, R-CNN)为代表的双阶段^[2](two-stage)检测方法,即先经过网络模型获取可能含有待检目标的候选框,再通过神经网络进行分类和回归获取更加准确的检测框。two-stage 检测算法的检测精度往往较高,但由于训练过程较为复杂,需要花费较高的时间成本,一般很难应用在实际生活场景中进行实时检测。另一类是以 YOLO(you only look once)算法为代表的一阶段(one-stage)检测方法,该方法为端到端的形式,即采用一个网络,输入需要检测的图片数据,即可获得输出的物体,包括物体的位置及种类。这类算法具有更加轻量化的模型、更小的参数量以及能够快速检测等优点,可以应用在多种场景进行实时检测,实用性更强。尽管 two-stage 算法具有一定的准确率,但考虑到其有限的实时性,很少有研究将其应用在交通监测任务中,相反, YOLO 体系结构因具有较快的速度而得到广泛使用。

与此同时,深度学习^[3]和计算机视觉技术有了极大的进展,作为计算机视觉领域的核心任务之一,目标检测的方法愈发丰富,技术日趋成熟。在此背景下,得益于其低廉的成本和多种多样应用支持而被广泛的应用于交通检测场景。研究者们在过去的几年中针对道路交通场景的目标检测已经提出了许多优秀的算法。Dong 等^[4]针对 YOLOv5 算法存在计算量大、检测率低的问题,提出了一种改进的轻量级 YOLOv5 车辆检测方法。该方法在 YOLOv5 颈部网络中引入 C3Ghost 和 Ghost 模块,减少特征通道融合过程中的浮点运算,增强特征表达性能。在 YOLOv5 主干网络中引入卷积块注意力模块优化特征融合,提高算法的检测精度。将 CIoU_Loss 作为边界框回归损失函数,加快边界框回归速率,提高算法的定位精度。Zhang 等^[5]为了克服传统 YOLOv3 算法对小型车辆目标检测精度低的问题,提出了在 YOLOv3 传统 3 个卷积层的基础上增加了一个更大的卷积层,并使用 K-means++ 聚类算法再次得到 12 个锚框的方法。该方法中,新增加的 104×104 特征尺度比 YOLOv3 原有的特征尺度更适合小目标检测,利用 K-means++ 算法更容易得到全局最优锚点,从而提高检测精度。Xue 等^[6]为了提高夜间行人检测的准确性和实时性,提出一种基于多模态注意力结合 YOLO 的实时行人检测方法。通过构建基于压缩 Darknet53 框架的多模态特征提取模块来适配夜间行人检测,保证检测效率,并从两种模态中提取特征,然后通过模态加权融合模块进行融合。刘素行等^[7]针对复杂交通场景下的目标检测准确率和实时性要求,以交通场景中的汽车、行人目标,提出了一种基于 YOLOv3 的交通场景目标检测方法。通过 K-means 方法对 YOLOv3 重聚类出适于交通场景的锚框,基于新的锚框对 YOLOv3 进行重训练,并在

交通数据集上检验了目标检测效果。黄凯文等^[8]在 YOLOv4-Tiny 算法中引入深度可分离卷积,缩减模型的计算量,增加检测分支,采用多尺度特征融合结构增强对小目标的检测效果。通过改进的全局纹理注意力模块,加强网络对全局特征的提取能力。

本文考虑到 YOLO 体系结构在交通检测系统中的适用性,因此采用高效、准确的 YOLOv5s 作为基础进行研究,设计了一种改进的 YOLOv5s 模型。根据交通场景图像的特点,为了使网络模型更高效且保证较高的检测精度,通过一种高效的层聚合网络,加强了对特征的提取和融合,提升模型的检测效率;为了进一步提升模型的性能,在深层网络中添加了注意力机制实现对重点特征信息的关注,优化特征的提取;引入 α -IoU 损失函数,提升边界框(bounding box, bbox)回归精度,增强模型的鲁棒性。

1 YOLOv5s 原算法

在 YOLO 神经网络之前, R-CNN^[9]、Fast-RCNN^[10]等经典目标检测算法的思路为先由目标聚类提取感兴趣区域(region of interest, ROI),即提取可能带有物体的区域,尽可能保证每个区域有且仅有一个物体,然后对区域中的物体种类进行识别。上述步骤分由两个神经网络处理,一个提取 ROI,另一个识别目标。相较于 R-CNN 系列的经典目标检测算法, YOLO 将目标检测转化成一个回归(regression)问题,即给定输入图像,在图像上直接获得 bbox 以及要检测目标的信息。

Glenn Jocher 在 YOLOv4 的基础上添加了一些改进思路提出了一种高效的 YOLOv5 算法。该算法相较于之前的 YOLO 系列,在速度和精度上都获得极大的性能提升, YOLOv5 在输入端增加了 Mosaic 数据增强、自适应锚框^[11]计算以及自适应图片缩放;在骨干网络中添加了 Focus 和 CSP 结构;在特征融合部分该算法添加了 FPN+PAN 结构;在输出端改进了损失函数 CIoU Loss 等,该模型在 COCO 数据集上的平均精度和检测速度分别提高了 10% 和 12%。 YOLOv5 具有 4 种网络模型, 4 种模型的区别主要在于网络中子模块的深度(depth_multiple)和宽度(width_multiple)不同,通过这两个参数调整骨干网络的深度和宽度,并将其划分为 4 个数量级的模型, YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x。 YOLOv5 对象检测的主要过程为先将图片划分为 $N \times N$ 个单元格,如果目标物体的中心坐标落在某个单元格内部,则该网格负责检测该目标,每个单元格有 P 个边界框。每一个边界框有 5 种信息参数,包括边界框的中心坐标 (x, y) 、宽和高 (w, h) 以及置信度(confidence)。每张图片输入模型的值为一个 $N \times N \times (P \times 5 + C)$ 维向量(其中 C 为类别数),模型一共预测 $N \times N \times P$ 个边界框。 YOLOv5 算法将置信度阈值设定为 0.25,先筛选置信度小于 0.25 的预测边界框,随后利用非极大值抑制算法(non-maximum suppression, NMS)过滤同一个目标的多个预测框,保留最优预测框。

2 算法改进

2.1 主干网络改进

本文提出在YOLOv5的主干网络中引入一种新型高效的层聚合网络(newefficient layer aggregation networks, NELAN),如图1所示。对于高效网络架构的设计,一般主要考虑的因素是网络的参数量、计算量和

计算密度。从内存访问成本的角度上来看,神经网络的推理速度受到网络的分支结构、通道数以及输入输出的影响^[12]。具有代表性的高效聚合网络结构有CSPVoVNet^[13],CSPVoVNet不仅考虑了参数量等因素,还额外考虑了梯度路径的影响,以使不同层的权重学习到更多不同的特征,也提升了模型的推理速度和精度。

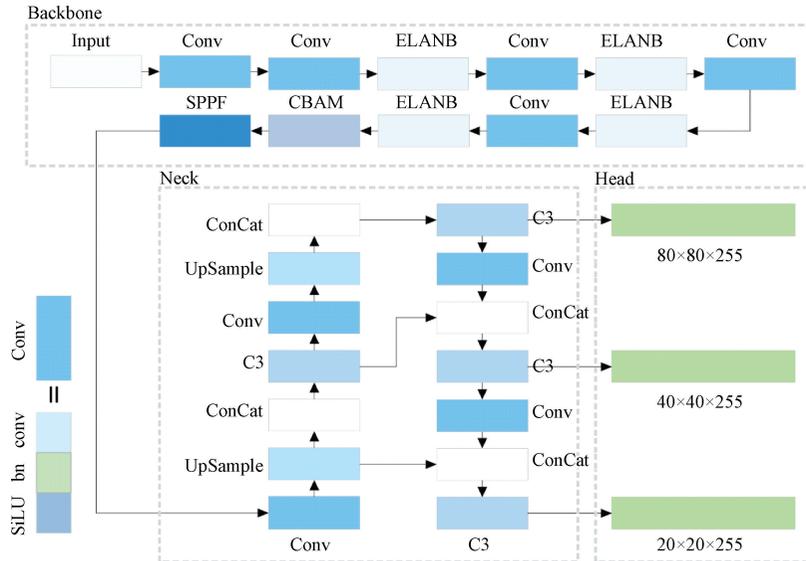


图1 YOLOv5-NELAN 网络结构

本文提出的 NELAN 本质上是一个级联结构^[14],为了使网络的性能更加高效,通过控制最长的梯度路径和更深层次的网络来进行有效的学习图像特征并收敛。在大规模的 NELAN 中,不论梯度路径长度和计算块的堆叠数量是多少,都达到了一个稳定的状态。但是如果无限地堆叠计算块,则可能破坏这种稳定状态,导致参数利用率下降。NELAN 的结构由一系列的卷积模块组成,其级联分支是一个前向传播的过程。如图2所示,输入特征经过两个并行的卷积模块得到输出 x_1 和 x_2 ,接着 x_2 作为两个串行卷积模块的输入而得到输出 x_3 ,随后 x_3 作为另外两个串行卷积模块的输入得到输出 x_4 ,最后将4个输出进行 concat 操作得到该模块的最终输出。

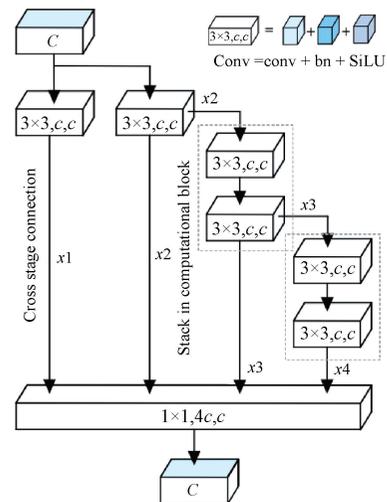


图2 NELAN 网络模型

另外,在改变级联模块深度后,宽度也会随之改变(图3),这就导致后续网络的输入宽度发生改变,进而引起后面输入 channel 和输出 channel 发生改变,这就会导致模型占用更多的硬件资源。因此,需要先根据级联模块深度的变化计算出相应模块输出 channel 的变化,即模块宽度变化,然后根据变化比例执行剩余的传输层。

2.2 添加注意力机制

本文在改进的YOLOv5s中引入卷积注意力模块^[15](convolutional block attention modul, CBAM),该模块的功能是让网络知道重点关注哪些目标信息,相应地突出表现重要特征,而抑制不太重要的特征;CBAM是一个轻量级模

块,它包括通道注意力和空间注意力两个独立子模块,分别关注重要特征信息和目标位置信息,使特征覆盖了待检测物体的更多部位^[16],有助于更好地进行特征提取。

如图4所示,通道注意力模块的操作流程为输入的特征图采用全局平均池化操作(AvgPool)和全局最大池化操作(MaxPool),获得的特征图通过共享网络(MLP),随后再次进行特征提取,得到最大池化通道和全局平均池化

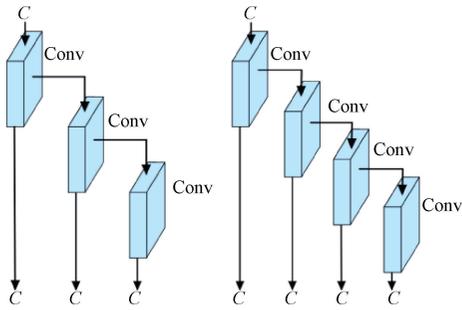


图3 级联结构

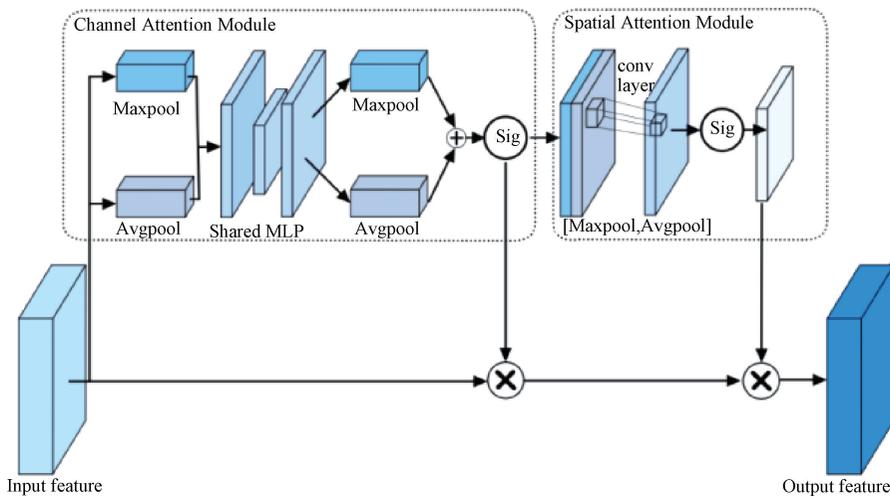


图4 CBAM网络模型

优势。经过长时间的改进与发展, IoU loss 从 GIoU、DIoU 到 CIoU^[18], 每一种损失函数都有其各自的特点。在 YOLOv5s 中采用了 CIoU 作为损失函数, 如下式所示:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \beta\nu \quad (1)$$

$$\beta = \frac{\nu}{(1 - IoU) + \nu} \quad (2)$$

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

式中: ρ 表示 b 和 b^{gt} 之间的欧氏距离用; c 表示 b 和 b^{gt} 之间最小外接长方形对角线的长度用; ν 为 b 和 b^{gt} 之间长宽比的相似性; w^{gt} 、 h^{gt} 表示真实检测框的宽和高; w 和 h 表示预测检测框的宽和高; β 用于平衡比例参数。

CIoU 在 DIoU 的基础上增加了检测框的尺度损失, 增加了预测框和真实框之间的长宽比的一致性考量, 也考虑了边界框回归的重叠面积、纵横比以及中心点距离等问题, 使得预测框更加符合真实框。但是通过其公式中的 ν 反映的是纵横比的差异, 而不是宽高分别与其置信度的真实差异, 所以有时会阻碍模型有效的优化相似性, 同时没有考虑难易样本的平衡问题。

α -IoU 同样作为 IoU loss 的拓展形式有着不俗的性

能, 能够对 CIoU 的加速收敛起到改进作用, 可以增大梯度加速收敛。当 α 的取值大于 1 时可以通过提升 high IoU^[19] 目标的损失值和梯度, 从而提升 bbox 的回归效果, 当 α 取值为 3 时, 多数情况下都有较好的效果。 α -IoU 在多种数据集上都展现了比现有一般形式的 IoU loss 更好的效果, 通过调节 α 可以实现不同水平的 bbox 回归精度, 尤其是对小数据集和具有噪声的 bbox 具有更好的鲁棒性。另外, α -IoU 损失函数更有利于轻量化模型^[20], 当在计算资源有限的场景中训练轻模型时(如移动设备、自动驾驶车辆和机器人等), α -IoU 损失函数更具优势。

通道。将两者叠加成一维矢量求经过 Sigmoid 激活函数传递, 最终获得通道注意力特征。

空间注意力模块操作流程为通道注意力模块的输出特征图作为本模块的输入特征图, 首先进行 AvgPool 操作和 MaxPool 操作获得的特征图, 再经过 7×7 的卷积和 Sigmoid 激活操作得到空间注意力特征。

2.3 Alpha-IoU 损失函数

对于目标检测任务而言, 在 bbox 的回归问题^[17]中, 损失函数主要分为两种, 分别是 $l_n - norm$ 和 IoU loss, 由于 IoU loss 对 bbox 尺度不敏感, 使其在检测任务中更具

因此, 本文提出的 YOLOv5s_NELAN 模型中提出采用 α -IoU 损失函数替代原模型中的 CIoU 损失函数, 选择 $\alpha = 3$, 引入 α -IoU 损失后检测效果优于现有基于 CIoU 的损失, 且不会引入额外的参数和训练时间。 α -IoU 损失函数如下:

$$L_{\alpha-IoU} = 1 - IoU^\alpha + \frac{\rho^{2\alpha}(b, b^{gt})}{c^{2\alpha}} + (\beta\nu)^\alpha \quad (4)$$

3 实验结果与分析

3.1 实验平台

硬件和软件平台参数如表 1 所示。

表1 硬件平台和软件平台的参数配置

配置	型号
CPU	Intel i9-12900KF
GPU	NVIDIA GeForce RTX 3090
操作系统	Windows 10
运行内存	32 G
显存	24 G
语言及框架	Anconda3 + Python3.8 + Pytorch1.12.1

3.2 实验数据采集与处理

本文使用的实验数据为自主收集,并使用vott软件进行手动标注的交通路口数据集,标注后的标签为YOLO格式。该数据集共有2311张由摄像头在不同时间段拍摄的交通场景图像,共包含行人、汽车、自行车、卡车4个类别,训练时采用Mosaic数据增强方法^[21],使数据内容更加丰富。由于该数据集中包含部分夜间拍摄图像,因此存在局部过暗、对比度差异以及色彩信息不足等问题,本文采取在部分原始图像的基础上对图像进行图像亮度增强,并额外生成300张增强后的图片,一定程度上提高网络模型的鲁棒性。本数据集按照8:2的比例划分训练集和测试集。如图5所示,图5(a)和(c)经过图像增强后得到图5(b)和(d)。

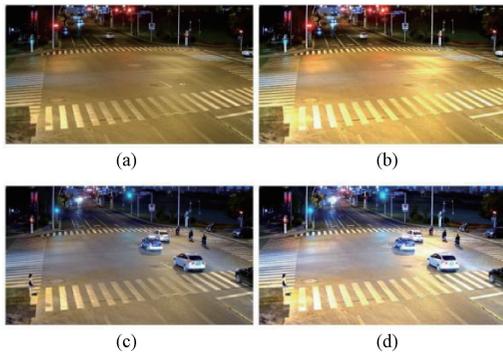


图5 图像增强对比

3.3 实验训练设计

本文模型在训练输入的图像尺寸设置为 640×640 ,训练轮数 $epochs = 300$,batchsize大小设置为16,一次性加载16张图片进入内存。初始学习率设置为0.001,衰减系数为0.0005,动量参数为0.937。如图6所示,训练损失和验证损在大于50轮附近开始收敛,且两者之间的差值很小,当到达280轮时两者趋于相等,由此可以判断该网络模型性能良好。

3.4 评价指标

实验采用目标检测指标^[22]来评估改进的YOLOv5s算法在交通路口场景进行目标检测的性能,衡量算法的两个最重要的性能指标是检测精度和检测速度,对于精度检测,通过对每种类别目标的精度(precision, P)和平均精度均值(mean average precision, mAP)来判断。在目标检测

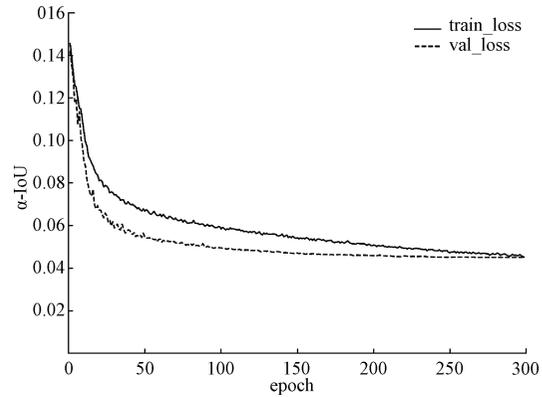


图6 训练和损失曲线

的实际应用中,网络模型不仅需要较高的准确性,其检测速度也要满足实时性,因此衡量网络速度的指标帧率(frames per second, FPS)也作为主要考察的指标之一,帧率定义为模型每秒检测图片数量。另外,由于该检测场景需要考虑到将算法有效部署到硬件设备上,本文也考虑了网络模型的大小。在本文模型中,将样本的IoU大于等于置信度^[23]阈值设定为正例子,样本的IoU小于等于置信度阈值设定为反例,本文将置信度阈值设定为0.25。

P、mAP、FPS指标公式如下:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$AP = \frac{dy}{dx} \int_0^1 P(R) dR \quad (7)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP \quad (8)$$

$$FPS = \frac{FrameNum}{ElapsedTime} \quad (9)$$

式中:P表示检测精度;R表示召回率;TP表示正样本被正确预测的个数;FP为负样本被错误预测的个数;FN正样本被错误预测的个数。N为检测目标类别总数;FrameNum为检测到的图像个数;ElapsedTime为图像处理时间。

3.5 消融实验

本文提出的YOLOv5-NELAN模型对YOLOv5s原模型在特征提取、损失函数、注意力机制等方面做出了改进,为了评估不同模块的改进以及改进模块之间的组合对模型性能的影响,设计了一系列消融实验,使用mAP和帧率作为评价指标,实验结果如表2所示。

由表2可知,本文提出的改进方法中,在YOLOv5s的骨干网络中引入NELAN结构,模型mAP提升了1.1%,帧率降低了4.6fps;原模型加入CBAM注意力机制后,mAP提升了0.3%,帧率变化不明显;改进损失函数为 α -IoU时,mAP提升0.5%,帧率提高了6.8fps。将所

表 2 消融实验

模型	mAP /%	帧率 /fps	参数量 / $\times 10^6$
YOLOv5s	89.6	126.5	7.02
YOLOv5s+NELAN	90.7	121.9	5.53
YOLOv5s+CBAM	89.9	125.3	7.06
YOLOv5s+NELAN+CBAM	91.6	115.2	5.55
YOLOv5s+ α -IoU	91.1	133.3	7.02
YOLOv5s+NELAN+CBAM+ α -IoU	92.0	117.6	5.55

有的改进方法同时加入 YOLOv5s 后,模型整体 mAP 提升了 2.4%,具体为 $mAP = 92.0\%$,帧率降低了约 8.9 fps,

帧率为 117.6 fps。从上述实验结果分析可知,本文提出的不同模块在交通场景目标检测任务中,检测精度都有一定程度的提高,但帧率略有降低,这是由于基于 YOLOv5s 原模型新增模块后,网络相对复杂,对网络的检测速度有一定影响。

3.6 实验结果分析

不同模型检测效果如图 7 所示,所有模型对大尺寸目标均有较好的检测性能,但本文提出的 YOLOv5-NELAN 模型检测出的精度整体高于其他 3 种模型。其中,YOLOv5-NELAN 模型对遮挡的行人目标具有良好的检测效果,同时对边缘处的小目标的检测能力优于其他模型。不同模型的评价指标对比如表 3 所示,本文模型训练精度如图 8 所示。



图 7 不同模型检测效果

表3 不同模型的评价指标对比

模型	AP/%				mAP/%	参数量 /($\times 10^6$)	帧率 /fps	模型大小 /($\times 10^6$)
	person	car	bicycle	truck				
YOLOv3	81.2	93.1	90.5	91.7	89.1	61.9	74.6	123.5
YOLOv4	83.7	94.1	89.3	93.3	90.1	52.5	96.3	100.6
YOLOv5s	80.8	95.5	89.9	92.1	89.6	7.02	126.5	14.1
本文	85.3	97.6	90.1	95.0	92.0	5.55	117.6	11.4

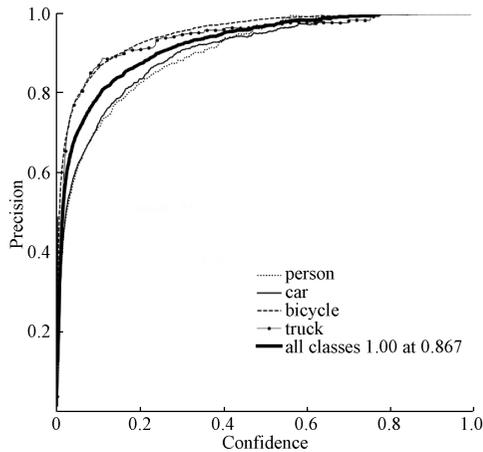


图8 YOLOv5-NELAN 训练精度

由表3可知,在自制交通路口场景的数据集上,本文改进的YOLOv5-NELAN模型比YOLOv3、YOLOv4和YOLOv5s这些模型在平均精度、参数量以及模型大小方面都有着较好的表现。从精确度上来看,YOLOv5-NELAN模型的mAP最高,比YOLOv5s提高2.4%,比YOLOv3和YOLOv4均有不同程度的提升;在模型参数方面,本文模型的参数量达到最小的 5.55×10^6 ,相较于YOLOv5s降低约 1.47×10^6 ;在模型速度方面,本文模型的帧率略低于YOLOv5s,但高于其他模型,满足检测的实时性;在网络模型大小方面,本文模型的大小仅为 11.4×10^6 。

YOLO系列算法的平均精度、参数量和模型大小与改进的YOLOv5-NELAN模型相比都有不同程度降低,这是因为本文引入的NELAN模块,通过控制最长的梯度路径和更深层次的网络来进行有效的学习图像特征并收敛,在模型参数和计算效率等方面都有着很大的改善;添加CBAM注意力机制,能够在网络进行特征融合时关注重要特征信息和目标位置信息,以此提升模型的精确度;引入 α -IoU损失函数,在不增加额外的参数、训练时间前提下,可以增大梯度加速收敛,以此获得mAP的提升。

本文提出的基于改进YOLOv5-NELAN模型的mAP达到了92%,帧率为117.6 fps,参数量为 5.55×10^6 ,模型大小为 11.4×10^6 ,尽管帧率相较于原模型略有下降,但其综合性能优于上述对比算法。

4 结论

本文提出了一种基于改进YOLOv5s的目标检测算

法,并应用到交通路口场景进行车辆及行人检测中。首先,针对自行收集的数据集进行图像亮度增强处理,改善了数据质量,扩充了数据。其次,引入NELAN模型,通过控制最长的梯度路径和深层网络学习图像特征,并将不同组的特征进行组合,以此增强特征学习能力,降低参数量,优化计算效率。通过添加CBAM注意力机制强化重要特征信息,增强了目标所在的区域的显著度。引入 α -IoU损失函数,进一步提升bbox的回归精度,增强了模型的鲁棒性。实验结果表明,本文提出的算法可以有效地提高交通路口场景目标检测的精度,相对YOLOv5s算法提升2.4%。在实际场景中,模型在对远处小目标以及遮挡严重的目标的检测仍有不足,存在误检和漏检现象,因此,进一步改进网络模型的检测性能是未来的主要工作。

参考文献

- [1] XU Y, ZHAO R. A prediction model of endometrial cancer lesion metastasis under region of interest target detection algorithm [J]. Scientific Programming, 2021, DOI:10.1155/2021/9928842.
- [2] 张索非, 冯焯, 吴晓富. 基于深度卷积神经网络的目标检测算法进展[J]. 南京邮电大学学报(自然科学版), 2019, 39(5): 72-80.
- [3] XU Y, ZHOU X, CHEN S, et al. Deep learning for multiple object tracking: A survey[J]. IET Computer Vision, 2019, 13(4): 355-368.
- [4] DONG X, YAN S, DUAN C. A lightweight vehicles detection network model based on YOLOv5[J]. Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation, 2022, 113: 104914.
- [5] ZHANG C, MENG Q, SUN Z, et al. Research on a road target detection method based on improved YOLOv3[C]. 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics). IEEE, 2020: 654-660.
- [6] XUE Y, JU Z, LI Y, et al. MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detec-

- tion[J]. *Infrared Physics & Technology*, 2021, 118: 103906.
- [7] 刘素行, 吴媛, 张军军. 基于 YOLO v3 的交通场景目标检测方法[J]. *国外电子测量技术*, 2021, 40(2): 116-120.
- [8] 黄凯文, 凌六一, 王成军, 等. 基于改进 YOLO 和 DeepSORT 的实时多目标跟踪算法[J]. *电子测量技术*, 2022, 45(6): 7-13.
- [9] TAO C, HE H, XU F, et al. Stereo priori RCNN based car detection on point level for autonomous driving[J]. *Knowledge-Based Systems*, 2021, 229: 107346.
- [10] CAO D, YANG S. A method based on Faster RCNN network for object detection[J]. *Recent Advances in Computer Science and Communications*, 2022, 15(9): 1239-1244.
- [11] GAO M, DU Y, YANG Y, et al. Adaptive anchor box mechanism to improve the accuracy in the object detection system[J]. *Multimedia Tools and Applications*, 2019, 78: 27383-27402.
- [12] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 116-131.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-yolov4: Scaling cross stage partial network[C]. *Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern Recognition*, 2021: 13029-13038.
- [14] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. *Computer Vision and Pattern Recognition*, 2022.
- [15] WANG G, DING H, YANG Z, et al. TRC-YOLO: A real-time detection method for lightweight targets based on mobile devices[J]. *IET Computer Vision*, 2021, 16(2):126-142.
- [16] 龙乐云, 周腊吾, 刘淑琴, 等. 改进 YOLOv5 算法下的输电线路外破隐患目标检测研究[J]. *电子测量与仪器学报*, 2022, 36(11):245-253.
- [17] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 658-666.
- [18] ZHANG G, DU Z, LU W, et al. Dense pedestrian detection based on YOLO-V4 network reconstruction and CIoU loss optimization[C]. *Journal of Physics: Conference Series. IOP*, 2022, 2171(1): 012019.
- [19] HE J, ERFANI S, MA X, et al. Alpha-IOU: A family of power intersection over union losses for bounding box regression[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 20230-20242.
- [20] ZAIBI A, LADGHAM A, SAKLY A. A lightweight model for traffic sign classification based on enhanced LeNet-5 network [J]. *Journal of Sensors*, 2021, DOI: 10.1155/2021/8870529.
- [21] 吕禾丰, 陆华才. 基于 YOLOv5 算法的交通标志识别技术研究 [J]. *电子测量与仪器学报*, 2021, 35(10): 137-144.
- [22] 韩航迪, 徐亦睿, 孙博, 等. 基于改进 Tiny-YOLOv3 网络的航天电子焊点缺陷主动红外检测研究[J]. *仪器仪表学报*, 2020, 41(11): 42-49.
- [23] SALMAN O, ELHAJJ I H, CHEHAB A, et al. Towards efficient real-time traffic classifier: A confidence measure with ensemble Deep Learning [J]. *Computer Networks*, 2022, 204: 108684.

作者简介

单慧琳, 硕士, 副教授, 主要研究方向为深度学习、图像处理等。

E-mail: shanhuilin@nuist.edu.cn

吕宗奎(通信作者), 硕士研究生, 主要研究方向为图像处理、深度学习等。

E-mail: 1372805287@qq.com