

基于改进YOLOv5s的交通信号灯检测算法

黄寅杰, 王福元, 肖海宁, 王仲楼

盐城工学院机械工程学院, 江苏 盐城

收稿日期: 2023年10月23日; 录用日期: 2023年11月22日; 发布日期: 2023年11月29日

摘要

针对交通信号灯检测在背景环境复杂、小目标检测困难和实时性要求高的问题, 基于YOLOv5s提出一种YL-YOLOv5s方法用于交通信号灯检测。首先, 将Neck特征融合网络中残差结构替换为多层次跨通道DenseNet连接方式, 以增强特征提取能力与减少梯度消失问题; 然后, 将DenseNet中的卷积操作替换为倒残差结构MobileNetV2, 以加快推理速度并降低内存占用; 最后在模块中嵌入通道注意力机制ECA (Efficient Channel Attention), 有效建立了通道之间的相关性, 从而提高了交通信号灯检测的准确率。实验结果表明, 在法国巴黎信号灯LaRa和博世小型交通信号灯BSTLD上进行测试, 与YOLOv5s算法相比, mAP@0.5:0.95、mAP@0.5等关键指标都有提升, 在LaRa数据上mAP@0.5:0.95提升8.5%, 并且模型计算量减少了23.2%, 说明YL-YOLOv5s模型能够有效提升复杂环境下小目标的检测精度, 且轻量化效果明显。

关键词

目标检测, 小尺寸目标, YOLOv5s, 交通信号灯

Improved YOLOv5s-Based Traffic Signal Detection Algorithm

Yinjie Huang, Fuyuan Wang, Haining Xiao, Zhonglou Wang

School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

Received: Oct. 23rd, 2023; accepted: Nov. 22nd, 2023; published: Nov. 29th, 2023

Abstract

In order to address the challenges of complex background environments, difficult detection of small targets, and high real-time requirements in traffic signal detection, this paper proposes a YL-YOLOv5s network based on the YOLOv5s architecture. Firstly, the residual structures in the Neck feature fusion network are replaced with a multi-level cross-channel DenseNet connection to

enhance feature extraction capabilities and reduce the problem of gradient vanishing. Next, the convolution operations in DenseNet are replaced with inverted residual structures from MobileNetV2 to accelerate inference speed and reduce memory usage. Finally, an Efficient Channel Attention (ECA) mechanism is embedded in the modules to effectively establish correlations between channels, thereby improving the accuracy of traffic signal detection. Experimental results show that when tested on the LaRa signal lights in Paris, France, and the Bosch Small Traffic Lights Dataset (BSTLD), the proposed YL-YOLOv5s model achieves improvements in key indicators such as mAP@0.5:0.95 and mAP@0.5 compared to the YOLOv5s algorithm. Specifically, it achieves an 8.5% increase in mAP@0.5:0.95 on the LaRa dataset, while reducing the computational cost by 23.2%. This demonstrates that the YL-YOLOv5s model effectively improves the detection accuracy of small targets in complex environments and exhibits significant lightweight effects.

Keywords

Object Detection, Small-Sized Targets, YOLOv5s, Traffic Signal Lights

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着科学发展的不断进步,智能交通不断发展,辅助驾驶技术成为研究学者重要的研究内容之一。据数据显示,色盲、色弱人群占比达到8%,城市环境中车辆拥堵导致的交通事故发生率高达30%。交通信号灯识别技术的提高可以帮助视力障碍人群更安全地驾驶汽车,同时有效减少行车过程中的交通拥堵问题。准确实时识别检测交通信号灯可以有效解决无人驾驶的安全问题。因此,研究交通信号灯的识别技术不仅具有重要意义,也是未来智能交通研究的热点。

近年来,利用深度学习的目标检测技术得到了广泛的运用。目标检测算法主要分为两大类:基于两阶段(Two-Stage)的目标检测算法,这类算法先通过区域提取(Region Proposal)阶段生成候选目标框(Region of Interest, ROI),然后对这些候选框进行分类和精确定位。典型的两阶段算法包括 Faster R-CNN [1]、R-CNN [2]、Fast R-CNN [3]、Mask R-CNN [4]等。基于单阶段(One-Stage)的目标检测算法直接在图像上进行密集的检测预测,无需显式的区域提取步骤。这类算法直接在图像上进行密集的检测预测,无需显式的区域提取步骤。它们通过在不同位置和尺度上应用密集的锚点(Anchors)或默认框(Default Boxes)来预测目标的位置和类别。典型的单阶段算法包括 YOLO (You Only Look Once) [5]、SSD (Single Shot MultiBox Detector) [6]、RetinaNet [7]等。目前,国内外研究人员对已有的现有算法进行改进,以提高对交通信号灯小目标检测的精度和速度。文献[8]提出一种基于图像增强的交通信号灯识别方法,通过 HSV 颜色空间中对颜色通道进行分离并增强 V 通道分量再合并通道提取交通信号灯的颜色,与传统算法比较识别准确率提升了 1.05%,并且有良好的实时性。文献[9]提出了基于 Matlab 的红绿灯识别系统,图片经过灰度化、图像锐化、滤波降噪等预处理,使用直方图阈值分割法有效分割红绿灯,通过二值化图像的矩阵进行红绿灯的定位,基于 HSV 色彩空间识别出红绿灯的颜色状态,通过样本检测与分析,该模型在普通条件下的红绿灯识别率达 80%。文献[10]提出改进的 Yolov4 的交通信号灯倒计时数字检测与识别,通过将 CSPResNet50 代替主干网络 CSPDarkNet53,增加卷积感受野,提高特征提取能力。文献[11]提出了一种改进的 YOLOX 网络来解决多目标情况下的小目标漏检率高、模型特征融合不充分问题,在骨干网络引入特征提取更加细粒度的

Res2Block 模块，同时嵌入了自注意力机制，增加隐性小目标的区域特征，减少漏检率。

虽然上述算法对于交通信号灯检测性能上取得了一些提升，但还是存在问题。首先模型中过多的参数，会使模型降低了推理速度并且很难优化到移动设备上使用。其次，交通信号灯周边又有较多的干扰物，模型也容易发生误检。最后，模型在检测精度和速度上很难取得一个平衡。

鉴于上述描述，YOLOv5s [12]作为一种基于单阶段的目标检测算法，在交通信号灯识别任务中具有实时性、准确性和特征提取能力强的优势，并且 YOLOv5s 在工业界有大量的使用案例，其稳定性和可靠性都有保障，另一方面，该模型后续需要部署到移动设备，YOLOv5s 在移动设备的部署上有一套完善的工具和充分的先验经验可以借鉴，所以选用 YOLOv5s 进行改进，提高对交通信号灯识别的精度和速度。设计出一种基于多层次跨通道[13]并结合倒残差结构的 C3 模块，替换特征提取网络中的模型，进而减少计算参数与加快网络收敛速度。同时，通过插入优化过后的 ECA 注意力模块[14]，加强网络通道上的相关性，提高对小目物体检测的精度。为了验证改进后 YL-YOLOv5s 算法的有效性，通过两个公开交通信号灯 LaRa 和 BSTLD 数据集[15]对模型进行验证，并做了大量对比实验证明了改进算法的有效性。

2. YOLOv5s 算法介绍

YOLOv5 是一种用于目标检测的深度学习算法，由 Ultralytics 团队开发，它建立在 YOLOv3 的基础上通过一系列的改进和优化，提供了更高的检测性能和更快的推理速度。目前 YOLOv5 算法有五个版本分别是 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，每个版本的网络结构相同，不同的是网络的深度和宽度。考虑到实际的需求，该文检测的任务对检测精度和实时性要求较高，所以选择计算量和参数第二小的 YOLOv5s 作为改进模型，具体的网络结构如图 1 所示。

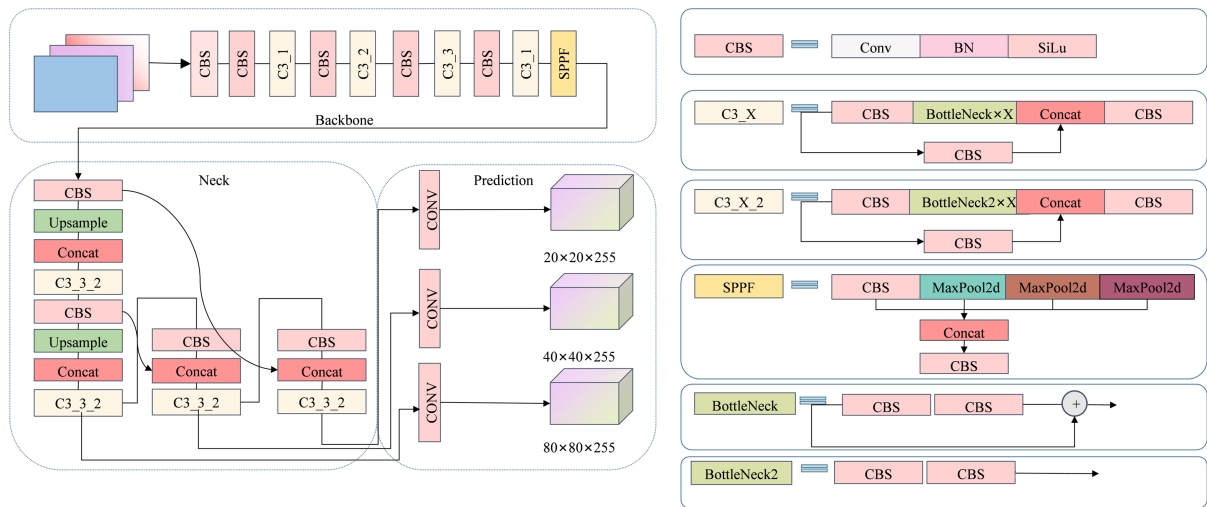


Figure 1. YOLOv5s network structure
图 1. YOLOv5s 网络结构

YOLOv5s 网络结构[16]主要由三个部分组成分别是主干网络 Backbone、颈部网络 Neck 和预测网络 Prediction [17]。主干网络采用的是 CSPDarknet 结构[18]，其是一种基于 Darknet 架构的改进版本，由 CBS、C3_X 和 SPPF(Spatial Pyramid Pooling Fusion) [19]构建。CBS 由卷积、BN(Batch Normalization)层和 SiLu 激活函数组成，在主干网络中对特征图进行下采样处理。C3_X 结构由多个 CBS 模块和一个 BottleNeck 模块组成，通过多个残差模块对特征图进行细节特征的提取。SPPF 模块是一种通过空间池化实现多尺度特征融合的方法，通过对特征图进行多次最大池化，最后将所有池化层的特征图连接在一起，得到目标

更多的语义信息。颈部网络 Neck 采用了 FPN (Feature Pyramid Networks) [20]与 PAN (Path Aggregation Networks)结构, 加强了模型的特征提取能力。颈部网络中所采用的 C3_X_2 特征提取模块与主干网络中有所区别, 其中内部结构 BottleNeck2 结构不采用残差结构, 如图 1 中结构所示。预测网络 Prediction 中, 通过对三个 1×1 卷积对不同的特征图进行处理, 就得到了三个不同的检测尺寸 (20×20 , 40×40 , 80×80), 每个分支都会输出一个边界框位置和目标类别概率。

2.1. C3_X_2 残差结构改进

YOLOv5s 颈部网络中运用了 4 组 C3_X_2 结构, 用于对输入图片进行卷积操作并提取相对应的特征信息。C3_X_2 结构中包含了 3 个标准卷积、1 个 BottleNeck2 模块和连接层, 可以有效提取输入图片的高层次特征, 增加网络检测的准确性和鲁棒性。但是该结构由于输入特征图通道数较大, 通常会需要大量的参数进行运算, 并且网络收敛速度也较慢, C3_X_2 网络结构如图 2 所示。

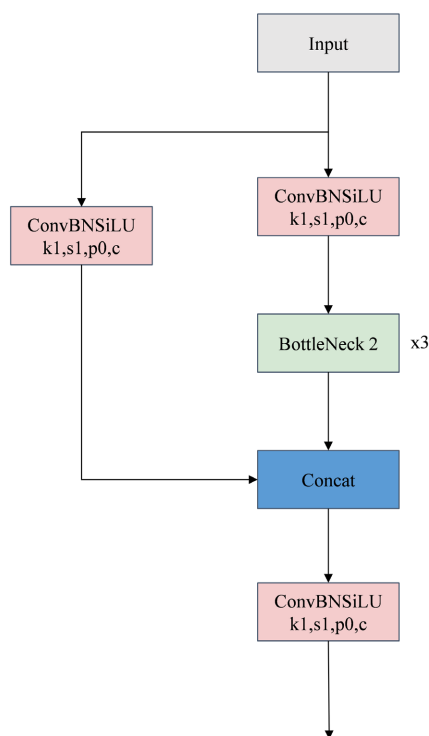


Figure 2. C3_X_2 network structure

图 2. C3_X_2 网络结构

为了更好地解决网络发散和减少参数量的问题, 设计了一种基于 DenseNet 结构的 C3_D_2 模块。在 DenseNet 网络中, 通过在层之间引入密集连接来改善信息流动, 这有助于缓解梯度消失的问题。密集链接允许高效地利用上层信息, 使得网络能够在不同深度上重用特征, 提高了参数效率并增强了特征传播。相比较于残差连接, DenseNet 在更少的参数情况下实现了最先进的性能。密集连接使特征重用成为可能, 减少了学习特征的冗余性, 使得网络更加紧凑。其核心表达式如式(1)所示:

$$x_k = H_k ([x_0, x_1, \dots, x_{k-1}]) \quad (1)$$

其中: x_k 表示第 k 层网络的输出; $[x_0, x_1, \dots, x_{k-1}]$ 第 0 层到第 $k-1$ 层网络提取地特征合并; H_k 表示为卷积、归一化和 ReLu 激活函数层地组合方式。Densenet 结构如下图 3 所示。

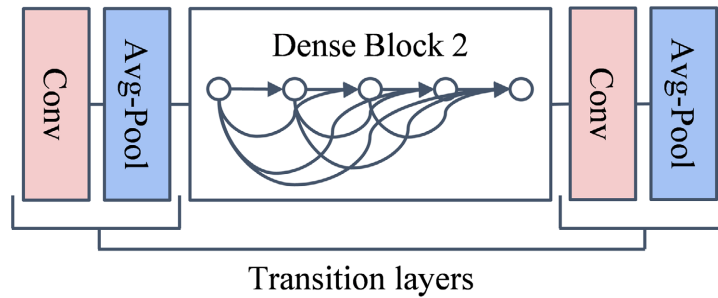


Figure 3. Densent structure diagram

图 3. Densenet 结构图

C3_D_2 结构中采用了 1×1 卷积和 3×3 卷积进行特征提取，不同之处在于，该模块将 BottleNeck2 替换成了 3 组 MobileNet [21] 进行特征提取。通过跨通道密集连接，使各个模块紧密连接，并且降低了输入通道的维度。这种设计更好地解决了深度学习网络模型退化问题并降低了计算参数[22]。C3_D_2 网络结构如图 4 所示。

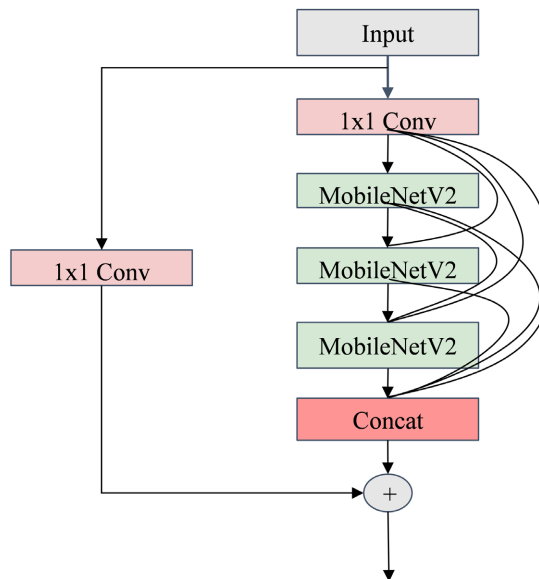


Figure 4. C3_D_2 structure diagram

图 4. C3_D_2 网络结构

首先，特征图经过右侧 1×1 卷积处理，将特征通道降低为原来的 $1/4$ 大小，再经过第一个 MobileNet 模块进行特征处理，卷积后得到的特征图信息与输入信息进行特征融合 Concat 操作，再送入第二个 MobileNet 模块，按照上面的流程完成三次 MobileNet 特征提取。最后，经过卷积处理的四个不同层级的特征图被通过 Concat 操作合并，以恢复特征图的通道数到原来的大小，并与经过左侧 1×1 卷积处理的图像进行特征相加，从而得到最终的特征图。相比较原模型，引入后的 MobileNet 结构通过跨通道结构，对不同的特征图进行特征融合，能提高对小目标信号灯检测的效果，因为参数的降低也能够提高网络的运行速度。

2.2. 采用 MobileNetV2 结构

交通信号灯实时检测对于模型速度和精度都有一定的要求，而且网络模型通常需要部署在嵌入式系统或移动设备上。此外，MobileNetV2 结构通过引入一系列轻量级的操作，设计宗旨在保持较高精度的

同时减少模型的参数量和计算复杂度，如深度可分离卷积 Depthwise Seperable Convolution [23]和线性瓶颈，加快了模型的推理速度。这对于交通信号灯检测任务至关重要，因为需要在实时场景中快速、准确地检测和分类交通信号灯的状态。MobileNetV2 的结构如图 5 所示。

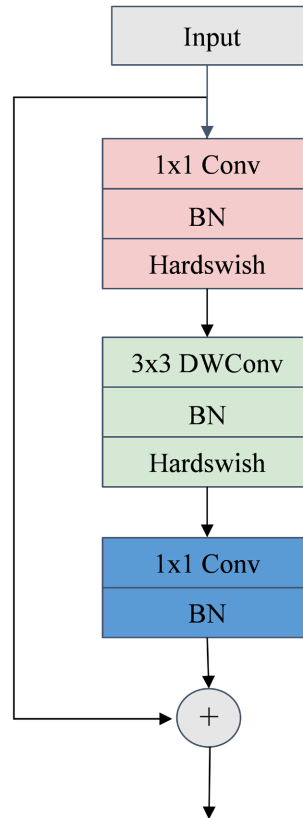


Figure 5. MobileNetV2 structure diagram

图 5. MobileNetV2 网络结构

在 MobileNetV2 结构中[24]，因为网络使用 3×3 深度可分离卷积，所以减少计算参数量。假设输入特征图片大小为 $C \times H \times W$ ，那么网络的总参数量计算公式如(2)所示：

$$1 \times 1 \times C \times 2C + 3 \times 3 \times 2C + 1 \times 1 \times 2C \times C = 4C^2 + 18C \quad (2)$$

使用 BottleNeck2 模块的时候，网络中总参数量计算公式如(3)所示：

$$1 \times 1 \times C \times C + 3 \times 3 \times C \times C^2 = C^2 + 9C^3 \quad (3)$$

上式中因为 C 是大于 0 的整数，可以看出，当特征图输入通道大于 1 的时候，使用 MobileNetV2 结构的网络总参数就会比 BottleNeck2 中正常 3×3 卷积和 1×1 卷积少。

采用 MobileNetV2 结构的 C3_D_2 模块，经实验验证在仅扩展两倍输入特征通道层下，也可以提高较大的性能，并且带来较少的计算量。对交通信号灯检测性能提高的同时，也不会增加模型的计算量，加快了网络检测的运行速度。

2.3. 引入改进注意力机制 ECANet

交通信号灯图像通常属于小目标检测，其在整幅图片中所占比例较小，因此容易受到复杂背景因素的影响。原 YOLOv5s 网络在深层次卷积时容易丢失小目标的特征信息，这不利于小目标检测。为了进一

步增强图像中的通道信息与交通信号灯的关联程度，丰富关键通道的特征表达，引入了能够将通道信息嵌入特征图的 ECANet 注意力机制。该注意力机制模块能够提升了模型的表达能力，通过引入通道注意力机制，可以动态地调整不同通道的重要性，使得模型能够更好地捕捉输入数据的重要特征。增强了模型的感知能力，通道注意力机制可以自适应地学习每个通道的权重，使得模型更关注对当前任务有意义的通道。减少了参数数量和计算复杂度，相比于传统的注意力机制，ECANet 引入了一个全局的池化操作，通过一个简单的线性变换来学习通道的注意力权重。为了能够使注意力机制更好发挥作用，该文对注意力机制进行了以下两种优化，另外，将全局平均池化操作替换为全局最大池化[25]，利用最大值突出感兴趣区域，突出小目标关键特征。此外，注意力机制中的自适应跨通道信息交互卷积被标准的 3×3 卷积所取代，以进一步优化机制。改进后的 ECANet 结构如图 6 所示。

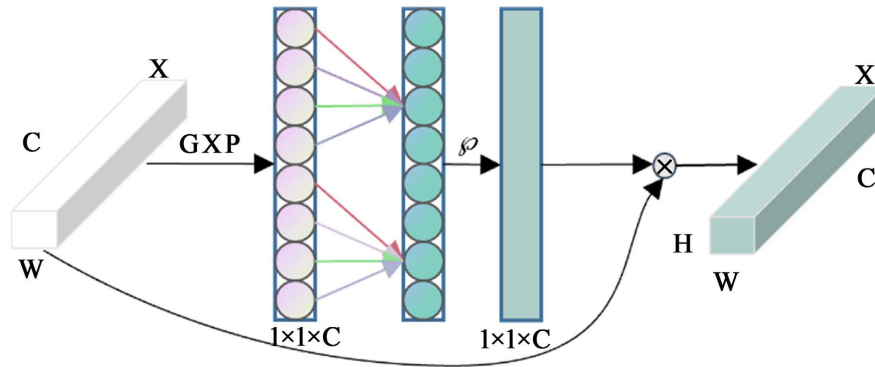


Figure 6. ECANet structure diagram
图 6. ECANet 网络结构

改进后 ECANet 通道注意力计算流程如下：假设输入特征图为 X ，大小为 $W \times H \times C$ ，其中 W 表示宽度， H 表示高度， C 表示通道数。通道维度的全局最大池化可以表现为公式(4)所示：

$$Y_{(c)} = \max_{(i,j)} (X_{(i,j,c)}) \tag{4}$$

其中， $Y_{(c)}$ 表示输出特征图中通道 c 的值， $X_{(i,j,c)}$ 表示输入特征图在位置 (i,j) 和通道 c 上的像素值。则通道注意力模块对通道间权重的计算可以通过式(5)得到。

$$W = \sigma(W_k y) \tag{5}$$

其中， σ 是 Sigmoid 激活函数， W_k 是 ECANet 中计算通道注意力的参数矩阵。 σ 的计算公式(6)， W_k 参数矩阵定义式(7)。

$$\sigma = \frac{1}{1 + e^{-x}} \tag{6}$$

$$W_k = \begin{bmatrix} w^{1,1} & \dots & w^{1,k} & \dots & 0 \\ 0 & w^{2,2} & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & w^{c,c} \end{bmatrix} \tag{7}$$

根据公式(7)可知， y_i 的计算权重只考虑其周边邻近 k 个通道。

$$\omega_i = \sigma \left(\sum_{j=1}^k w^j y_i^j \right), y_i^j \in \Omega_i^k \tag{8}$$

公式(8)中, 表示 y_i 周边 k 个临近的集合, 所有通道共享相同的学习参数可以提高网络的性能。注意力机制特征通道权重可以通过内核大小为 3 的一维卷积实现, 如公式(9)所示。

$$\omega = \sigma(C1D_k(y)) \quad (9)$$

公式(9)中, $C1D_k$ 表示卷积核大小为 3 的一维卷积, 使用内核为 3 的卷积可以加快模型的运算速度。

改进后的 ECA 注意力机制分别嵌入到 4 组 C3_D_2 模块所构成的网络架构中。改进的措施旨在特征融合的阶段对特征图进行处理, 从而极大地提升网络在搜索交通信号灯这类微小目标物体时的效能。网络的感知力和定位准确度得到了有益的增强, 从而有效地增进了对这类挑战性目标的探测能力。在 3.4 章节中, 提出了 2 种不同插入方式, 并通过实验讨论了不同插入方式对网络检测性能的影响。

3. 实验结果与分析

3.1. 实验数据集介绍

实验使用了 BSTLD、LaRa 两个交通信号灯数据集进行模型性能验证。LaRa 数据集有 9168 张图片, 包含三种不同类别的标签, 分别是 Green (绿灯)、Yellow (黄灯)、Red (红灯)。由于数据集中有大量重复与模糊的图片, 且数据集中 Yellow (黄灯) 标签类别数量过少, 经过处理后保留了 5340 张图片并添加了 192 张 Yellow (黄灯) 图片数据。BSTLD 数据集有 8334 张图片, 标签类别跟 LaRa 数据一样, 数据集中也添加了 192 张 Yellow (黄灯) 图片数据。上述两个数据集的图片和标签都按 8:1:1 的比例分为训练集、测试集和验证集。

3.2. 实验环境

实验的操作系统为 Windows10, CPU 型号为 AMD R7 5700X, 运行内存 32 GB, GPU 型号 NVIDIA GeForce RTX 3090, 显存大小 24 GB。模型基于 Pytorch1.11.0 深度学习框架, 编程语言使用 Python3.9, 使用 CUDA11.3 进行 GPU 加速。算法输入大小为 640×640 , batch_size 为 32, 模型优化器选用的是随机梯度下降优化器 SGD, 初始训练学习率为 0.01, 权重衰减率为 0.0001, 学习动量为 0.9, 训练迭代次数 epoch 设定为 100 次。

3.3. 评价指标

实验选取以下指标作为评判标准: mAP@0.5, 是一种通过计算不同类别下, 检测算法 IoU 阈值为 0.5 时平均精度(Average-Precision); mAP@0.5:0.95, 是将 IOU 阈值从 0.5 到 0.95, 将其以 0.05 个步长分成若干份, 计算各个 AP 的平均值, AP 可以通过 P-R 曲线面积计算得出; mAP@0.50:0.95 (small), 是数据集中目标像素面积小于 32×32 对应的 mAP 值, 通过这个数值可以反映出模型对小目标物体检测的能力; Frames Per Second 是用于评估模型的实时性能和推理速度, 较高的 FPS 值意味着模型能够更快地处理输入数据, 从而实现更高的实时性能; GFlops, 是指浮点运算数, 可以用来衡量算法模型的复杂度。相关计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (10)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (11)$$

$$A_{AP} = \int_0^1 P(R) dR \quad (12)$$

$$m_{\text{mAP}} = \frac{\sum P_A}{N_c} \tag{13}$$

$$\text{FLOPs} = 2HW(C_{\text{in}}K^2 + 1)C_{\text{out}} \tag{14}$$

上述公式中，TP 代表真阳性(True Positive)，指的是被模型正确地预测为正类的样本数量；FP 代表假阳性(False Positive)，指的是被模型错误地预测为正类的负类样本数量；FN 代表假阴性(False Negative)，指的是被模型错误地预测为负类的正类样本数量； C_{in} 代表卷积输入张量的通道数， C_{out} 代表卷积层输出张量的通道数， K 代表卷积核大小。

3.4. 实验结果和分析

3.4.1. C3_D_2 结构改进实验

为了验证不同特征融合对输出特征图的影响，设计了另外两种不同的特征融合方式进行对比实验以 C3_D2_2 和 C3_D3_2 表示，验证 C3_D_2 结构的有效性，各特征融合方式的结构图如图 7 所示。

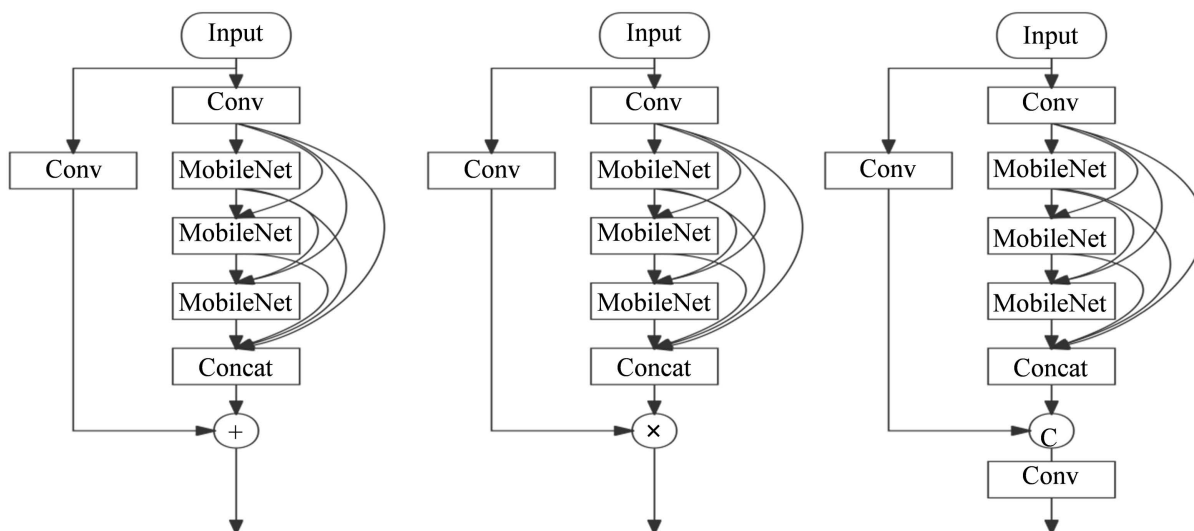


Figure 7. Three feature fusion module
图 7. 三种特征融合模块

三种特征融合方式在 LaRa 和 BSTLD 数据集上的评价指标如表 1 所示(加粗文字表示相同评价指标下的最优结构)。通过对比可以发现，使用 ADD 进行特征融合方式的 C3_D_2 模块在不同的数据集上相比其他两种方法表现最优。在 LaRa 数据集上，与使用 Concat 进行特征融合方式的 C3_D2_2 模块相比，使用 ADD 进行特征融合的 C3_D_2 模块在 GFlops 减少了 0.53 同时，mAP@0.5、mAP@0.50:0.95 和 mAP@0.50:0.95(small)分别提高了 9.2%，5.7%和 6.9%。在 BSTLD 数据集上，使用 ADD 进行特征融合的 C3_D_2 模块，mAP@0.5、mAP@0.50:0.95 和 mAP@0.50:0.95(small)也分别提高了 9.6%，4.3%和 5.4%。而使用了相乘进行特征融合的 C3_D3_2 模块，各项性能指标和计算参数量都不及 C3_D_2 模块。在性能指标上这表明在网络中使用 ADD 进行特征融合不仅可以减少网络的计算量，还能够增强小目标检测的能力。

进一步分析图 8 中三种特征融合方式的 Loss 曲线后发现，C3_D_2 模块在第 5 个 epoch 之后表现出了更快的 Loss 下降速度。这表明该模块在结构和参数上更适合解决小目标检测的问题，并具备更强大的表示能力。它能够更好地拟合训练数据并降低损失，同时也能减少网络模型的训练时间。从而证明 C3_D_2

模块在小目标检测任务中具有较高的性能和效率。经过改进后的模块在其结构和参数方面更为适合解决小目标检测的难题，并且具备更为强大的特征表示能力。这样的模块能够更有效地拟合训练数据，从而降低损失函数的值。与此同时，它还能够显著减少网络模型的训练时间，为模型的训练和优化过程带来明显的加速效果。

Table 1. Performance comparison (%) of three feature fusion methods in YOLOv5s model for object detection
表 1. 三种特征融合方式在 YOLOv5s 模型中检测性能(%)对比

数据集	方法	mAP@0.5 (%)	mAP@0.50:0.95 (%)	mAP@0.50:0.95(small) (%)	FPS	GFlops
LaRa	C3_D_2	90.9	47.4	46.8	52	9.05
	C3_D2_2	81.7	41.7	39.9	49	9.58
	C3_D3_2	80.6	41.0	38.7	50	9.07
BSTLD	C3_D_2	67.5	24.3	20.4	52	9.05
	C3_D2_2	57.9	20.0	19.0	49	9.58
	C3_D3_2	57.1	19.7	18.8	50	9.07

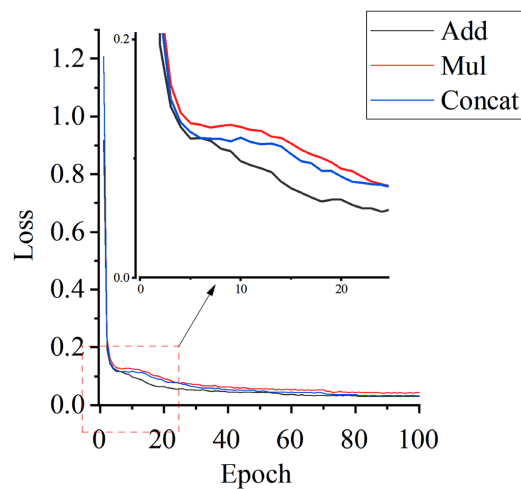
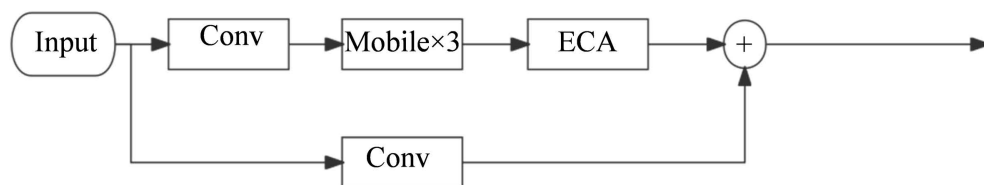


Figure 8. Loss curves for three feature fusion methods

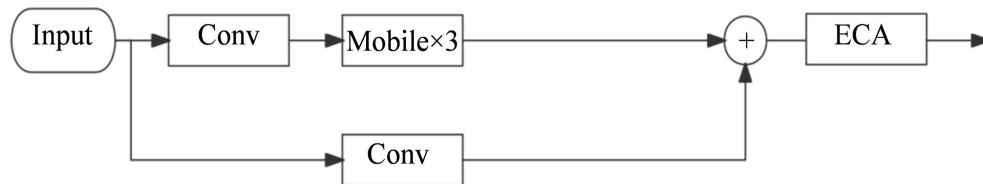
图 8. 三种特征融合方式 Loss 曲线

3.4.2. ECANet 结构改进实验

基于 ECANet 的 C3_D_2 结构能够通过特征图不同通道，有效选择出有用的特征，并且只使用了很少的计算参数。为了研究 ECANet 模块对 C3_D_2 性能上的影响，设计了两组不同的插入方式，如图 9 所示。



(a) 插入方式 A



(b) 插入方式 B

Figure 9. Two integration approaches for the ECANet module**图 9.** ECANet 模块两种插入方式

两组插入方式在 LaRa 和 BSTLD 数据集上的评价指标如表 2 所示。通过对比可以发现, ECANet 模块的插入和池化方式选择的不同, 对网络性能指标都有影响。插入方式对性能有显著影响, 在数据集 LaRa 和 BSTLD 中, 网络如果使用了 A 插入方式, 不管是使用了全局平均池化还是全局最大池化, 检测性能指标都是优于使用 B 插入方式。其中数据集 LaRa 上池化方式相同的情况下都使用了全局平均池化, 但选择了不同的插入方式, 两者 mAP@0.5 的性能最多提升了 29.7%。不同池化方式也能提高网络的性能, 在两种数据集中, 使用了 MaxPool2d 池化方式相对于 AvgPool2d 池化方式表现更好。MaxPool2d 池化方式 mAP@0.5、mAP@0.50:0.95 和 mAP@0.50:0.95(small) 指标上获得了较高的得分。因此, MaxPool2d 池化方式可以更好地保留重要的特征信息, 提高目标检测性能。所以使用插入方式 A 和 MaxPool2d 池化方式的结构可以取得更好的性能。

Table 2. Performance (%) comparison of two integration approaches for ECANet module in YOLOv5 model**表 2.** 两种 ECANet 模块插入方式在 YOLOv5 模型中检测性能(%)对比

数据集	插入方式	池化方式	mAP@0.5 (%)	mAP@0.50:0.95 (%)	mAP@0.50:0.95(small) (%)
LaRa	A	MaxPool2d	92.3	49.0	48.6
		AvgPool2d	89.8	47.4	46.1
	B	MaxPool2d	92.0	48.9	48.3
		AvgPool2d	60.1	32.1	32.2
BSTLD	A	MaxPool2d	71.6	27.4	23.6
		AvgPool2d	67.7	26.5	23.5
	B	MaxPool2d	66.6	25.8	23.2
		AvgPool2d	64.4	25.1	22.2

为了进一步验证不同 ECANet 模块对网络的影响, 分别绘制了采用不同池化方式的 YOLOv5 模型的 Grad-CAM 图像。图像对不同算法在输入图片上进行了可视化处理, 以便更好地了解模型在做出决策时的核心图形区域内容。从图 10 中可以看出, 使用了 ECA + MaxPool 模块进行预测的网络, 在感兴趣区域的范围上相较于使用 ECA + AvgPool 和原始网络更广且更加集中, 网络的感兴趣区域如果更集中于交通信号灯上, 这样就能够提供更多有用的特征信息, 并排除无关区域的干扰。因此使用了最大池化的 ECANet 模块, 对于小目标信号灯检测特征图更为突出, 从而获得更优的检测性能指标。



Figure 10. Grad-CAM images of ECANet with different pooling methods for the same input image
图 10. ECANet 不同池化方式相同输入图像下的 Grad-CAM 图

3.4.3. 网络模型消融实验

通过实验分析,对原网络结构分步骤进行改进,改进后的网络模型进行训练并测试,消融实验结构如表 3 所示。由表 3 可以看出:在采用了 C3_D_2 模块和 ECA 模块替换了原来的 C3 结构,各项性能指标都有较大的提升,其中用于检测小目标物体的 $mAP@0.50:0.95(\text{small})$ 提升了 9.9%,并且模型计算量还

降低了 2.74 个 GFlops，检测速度也提高了 5 个 FPS。由此可见，改进的网络能够较好地实现对交通信号灯的检测与识别，并且满足实时性的要求。

为了进一步验证该文算法对小目标检测的性能优化，在 LaRA 数据集上将该文的方法与 Faster-RCNN、SSD300、YOLO 系列等主流算法进行算法性能对比实验，并对不同方法的目标检测性能进行比较分析。表 4 是不同的实验结果；与两阶段目标检测算法 Faster-RCNN 相比，单阶段检测算法 SSD300 和 YOLO 系列在计算量 GFlops 上更小，检测速度也更快。与 YOLOv7 相比，该文算法帧率 FPS 减少了 13 帧，但是 mAP@0.5 提升了 0.7%，减少了 43.51GFlops，大大减少了算法的计算复杂度。实验数据表明，改进后的 YL-YOLOv5s 算法可以满足交通信号灯小目标检测的准确性和实时性。

Table 3. Performance (%) comparison of ablation experiments in YOLOv5s model for object detection

表 3. 消融实验在 YOLOv5s 模型中检测性能(%)对比

原网络	C3_D_2	ECA	mAP@0.5 (%)	mAP@0.50:0.95 (%)	mAP@0.50:0.95(small) (%)	FPS	GFlops
√			84.2	40.5	38.7	47	11.79
	√		90.9	47.4	46.8	52	9.05
		√	85.2	41.1	39.4	47	11.79
	√	√	92.3	49.0	48.6	52	9.05

Table 4. Performance (%) comparison of several detection methods

表 4. 几种方法检测性能(%)对比

算法	mAP@0.5 (%)	FPS	GFlops
Faster-RCNN	87	14	469.96
SSD300	72.5	47	30.59
YOLOv3	83.0	37	32.77
YOLOv4	86.5	43	28.89
YOLOX	89.8	60	13.32
YOLOv7	91.6	65	52.56
YL-YOLOv5s	92.3	52	9.05

4. 结束语

在交通信号灯检测领域，复杂的背景环境和高实时性需求一直是一项具有挑战性的任务，导致了现有目标检测算法存在漏检、误检和较慢的检测速度等问题。本文针对这一问题，提出了 YL-YOLOv5s 的轻量级交通信号灯目标检测算法。该算法引入 C3_D_2 结构以增强 YOLOv5s 网络的特征提取能力，采用 MobileNetV2 替代传统卷积结构以简化卷积过程降低模型参数计算量，以及引入 ECANet 注意力机制以提高小目标物体的检测精度，同时保持计算效率。实验结果表明，YL-YOLOv5s 相较于传统 YOLOv5s 算法，在减少了 23.3% 的参数计算量的同时，mAP@50 提高了 8.5%。尽管该方法在交

通信号灯检测方面取得了显著成果,但仍存在一定程度的漏检问题,同时还存在模型大小进一步优化的潜力。

参考文献

- [1] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [2] Cheng, B., Wei, Y., *et al.* (2018) Revisiting RCNN: On Awakening the Classification Power of Faster RCNN. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, Cham, 453-468. https://doi.org/10.1007/978-3-030-01267-0_28
- [3] Cheng, B., Wei, Y., Shi, H., *et al.* (2018) Revisiting RCNN: On Awakening the Classification Power of Faster RCNN. *Computer Vision—ECCV 2018*, Vol. 11219, Springer, Cham, 473-490. https://doi.org/10.1007/978-3-030-01267-0_28
- [4] He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- [5] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [6] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [7] Lin, T.Y., Goyal, P., Girshick, R., He, K.M. and Dollár, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/ICCV.2017.324>
- [8] Asha, C.S. and Narasimhadhan, A.V. (2018) Vehicle Counting for Traffic Management System Using YOLO and Correlation Filter. 2018 *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, 16-17 March 2018, 1-6. <https://doi.org/10.1109/CONECCT.2018.8482380>
- [9] 刘乔寿, 赵志源, 王均成, 皮胜文. 高性能 YOLOv5: 面向嵌入式平台高性能目标检测算法研究[J]. 电子与信息学报, 2023, 45(6): 2205-2215. <https://doi.org/10.11999/JEIT220413>
- [10] 桂欣悦, 李振伟, 吴晨晨, 李彦玥. 基于 MATLAB 的红绿灯识别系统研究[J]. 电子设计工程, 2020, 28(16): 133-136. <https://doi.org/10.14022/j.issn1674-6236.2020.16.029>
- [11] 李亚东, 马行, 穆春阳. 改进 YOLOX 网络的轴承缺陷小目标检测方法[J]. 计算机工程与应用, 2023, 59(1): 100-107. <https://doi.org/10.3778/j.issn.1002-8331.2206-0100>
- [12] Ultralytics (2022) YOLOv5. <https://github.com/ultralytics/yolov5>
- [13] Huang, G., Liu, S., Van der Maaten, L. and Weinberger, K.Q. (2018) CondenseNet: An Efficient Densenet Using Learned Group Convolutions. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2752-2761. <https://doi.org/10.1109/CVPR.2018.00291>
- [14] Wang, Q.L., Wu, B.G., Zhu, P.F., *et al.* (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 11531-11539. <https://doi.org/10.1109/CVPR42600.2020.01155>
- [15] Sandler, M., Howard, A., Zhu, M., *et al.* (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/CVPR.2018.00474>
- [16] Yan, B., Fan, P., Lei, X., *et al.* (2021) A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing*, **13**, Article 1619. <https://doi.org/10.3390/rs13091619>
- [17] Li, X., Wang, W.H., Hu, X.L. and Yang, J. (2019) Selectivekernel Networks. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 510-519. <https://doi.org/10.1109/CVPR.2019.00060>
- [18] Misra, D. (2019) Mish: A Self Regularized Non-Monotonic Activation Function. arXiv: 1908.08681.
- [19] Tang, H., Liang, S., Yao, D. and Qiao, Y.J. (2023) A Visual Defect Detection for Optics Lens Based on the YOLOv5-C3CA-SPPF Network Model. *Optics Express*, **31**, 2628-2643. <https://doi.org/10.1364/OE.480816>
- [20] Ghiasi, G., Lin, T.Y. and Le, Q.V. (2019) NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 7029-7038. <https://doi.org/10.1109/CVPR.2019.00720>

- [21] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E.H. (2020) Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2011-2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
- [22] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 1800-1807. <https://doi.org/10.1109/CVPR.2017.195>
- [23] Fu, J., Liu, J., Tian, H., *et al.* (2019) Dual Attention Network for Scene Segmentation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 3141-3149. <https://doi.org/10.1109/CVPR.2019.00326>
- [24] Gao, H., Wang, Z. and Ji, S. (2018) Large-Scale Learnable Graph Convolutional Networks. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1416-1424. <https://doi.org/10.1145/3219819.3219947>
- [25] Wang, C.Y., Liao, H.Y.M., Wu, Y.H., *et al.* (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 14-19 June 2020, 1571-1580. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [26] Gao, Z., Xie, J., Wang, Q. and Li, P.H. (2019) Global Second-Order Pooling Convolutional Networks. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 3019-3028. <https://doi.org/10.1109/CVPR.2019.00314>