

DOI:10.19651/j.cnki.emt.2211375

# 基于改进 YOLOv5m 的轻量化车脸检测方法\*

贾玮迪 余鹏飞 余国豪 李海燕 李红松

(云南大学信息学院 昆明 650091)

**摘要:** 为解决车检站车辆检测中需要对车辆前照灯快速准确定位,同时防止车辆代检的问题,建立了一个车脸检测数据集 Car-Data。针对车检站场景中车辆检测问题,提出了一种基于 YOLOv5m 的轻量化车脸检测方法。首先,将原网络的卷积块替换为改进型跨阶段深度可分离卷积块,以减少网络整体的参数量和计算量。其次,提出增强感受野的空间金字塔扩张卷积模块代替 YOLOv5m 的主干提取网络中的空间金字塔池化模块,从而提升网络的目标检测精度。最后,在颈部特征增强网络中修改上采样方法,并提出上下层特征融合模块,以减少特征信息的损失。在 Car-Data 数据集上进行的实验结果表明,改进后的算法相较于原 YOLOv5m 模型大小减少了 48%,每秒检测帧数提高了约 10 帧,且平均检测精度仍提升了 2.02%。因此该改进算法可以满足车检站车辆检测场景中车脸检测的需求。

**关键词:** 深度学习;车脸检测;轻量化网络;感受野;特征增强

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.60

## Lightweight car front detection method based on improved YOLOv5m

Jia Weidi Yu Pengfei Yu Guohao Li Haiyan Li Hongsong

(College of Information, Yunnan University, Kunming 650091, China)

**Abstract:** Aiming at the problem of fast and accurate positioning of car headlights at car inspection stations and preventing car from replacing inspection, a car front detection dataset Car-Data is established. To solve the problems of car detection in the vehicle inspection station scene, a lightweight car front detection algorithm based on YOLOv5m is proposed. First, the convolution block of the original network is replaced by an improved cross stage depth separable convolution block to reduce the parameters and computation of the network as a whole. Then, the spatial pyramid pooling module in the backbone extraction network of YOLOv5m is replaced with the spatial pyramid dilated convolution module of the enhanced receptive field, thereby improving the object detection accuracy of the network. Finally, the upsampling method is modified in the neck feature enhancement network, and an upper and lower layer feature fusion module is proposed to reduce the loss of feature information. The experimental results on the Car-Data show that compared with the original YOLOv5m, the size of the improved algorithm is reduced by 48%, the number of detection frames per second is increased by about 10 frames, and the detection accuracy is still improved by 2.02 percentage points. Therefore, the improved algorithm can meet the needs of car front detection in the car detection scene of the car inspection station.

**Keywords:** deep learning; car front detection; lightweight network; receptive field; feature enhancement

## 0 引言

随着社会化水平和科学技术水平的不断提高,汽车数量迅速增长,目前已成为人们最重要的交通工具之一,故车辆安全显得尤为重要。车辆前照灯在车辆安全行驶中发挥着重要的作用<sup>[1]</sup>,对前照灯检测是车辆年检中消除车辆安全隐患的重要环节之一。传统的车检灯光检测仪是通过追

踪光强来移动支架对准前照灯进行灯况检测,这个对准过程耗时久、效率低。因此本研究团队提出一种利用相机成像,然后基于深度学习实现对车灯的快速准确定位的方法,从而提高车灯检测仪的效率。同时,利用抓拍到的车辆正面(车脸)图像检测出车标、车辆颜色等车辆信息,可以进一步防止代检作弊违规现象的发生。

深度学习是机器学习近几年较为热门的一个分支,目

收稿日期:2022-09-13

\* 基金项目:国家自然科学基金(62066046)项目资助

前基于深度学习的目标检测算法主要分为两类<sup>[2]</sup>：一类是基于候选区域的两阶段算法。该算法具有较高的检测精度，但检测速度较慢，难以满足目标检测的实时性要求，如 R-CNN<sup>[3]</sup>、Fast R-CNN<sup>[4]</sup> 和 Faster R-CNN<sup>[5]</sup> 等。另一类是基于回归方法的一阶段算法，该算法的检测速度更快，但检测精度比两阶段算法低，如 SSD<sup>[6]</sup> 和 YOLO<sup>[7]</sup> 等。

由于一阶段算法结构简单且实时性好，更容易应用于复杂多变的实际环境，因此国内外很多研究人员都在进行一阶段车辆相关检测算法的研究工作。例如：李国进等<sup>[8]</sup> 在 SSD 算法的基础上引入基于空洞卷积的改进 Inception 模块和多尺度融合均衡化网络，在不增加大量参数的同时提升了算法对目标车辆的特征提取能力，但是该算法在出现重叠小目标的时候仍会出现漏检现象。Piao 等<sup>[9]</sup> 提出以轻量级网络 PeleeNet 为主干结合旋转边界框的 SSD 算法，加快了网络的推理速度，但是其模型误检率高于原 SSD 算法。张宝朋等<sup>[10]</sup> 采用 Shufflenet v2 代替 YOLOv4 的主干网络并使用深度可分离卷积代替普通卷积，明显降低了模型大小，提高了网络推理速度，但是牺牲了检测精度。Alsanabani 等<sup>[11]</sup> 通过改进 YOLOv4-tiny 的卷积块减少浮点运算量，提高了实时检测性能但是模型精度没有提升。上述的车辆检测算法<sup>[8-11]</sup> 主要针对智能交通应用领域，该领域与车检站车辆检测应用领域区别较大。

目前，查阅文献后尚未发现有前人研究过使用机器学习或深度学习来解决基于图像的车检站车辆检测问题。因此，为满足车辆检测实际应用的需求，本文提出了基于改进 YOLOv5 模型的车脸检测算法，以解决车检站车辆检测的问题。YOLOv5<sup>[12]</sup> 模型作为一个同时兼顾准确性和实时性的一阶段网络，相比于 YOLOv4<sup>[13]</sup>，YOLOv5 有多个不同大小的结构，定制化程度较高。为满足实际应用实时性的需要，本文对 YOLOv5 进行改进，首先，使用更加轻量的卷积块代替原来网络中的大量卷积块，适当降低网络的参数量和计算量；其次提出空间金字塔扩张卷积模块，减少参数量的同时增加网络的感受野；最后，更换原网络中部分上采样方法，同时提出一个上下层特征融合模块，以减少不同尺度特征的冲突。这些改进方法在自建数据集上进行训练和验证，结果表明改进方法有效地减少了模型的参数量同时又兼顾了模型的性能。

## 1 YOLOv5 算法简介

YOLOv5 和之前的 YOLO 系列算法的不同之处在于其根据模型大小分为 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 4 种结构。这 4 种结构在公开数据集 COCO 上的测试结果如表 1 所示，模型参数越多，运算量越大，检测精度也越高。由于本文需要对汽车正面进行实时性检测，既要选取更加轻量化的结构来降低硬件成本，又要兼顾检测性能。由表 1 的各项数据可知，YOLOv5s 和 YOLOv5m 的结构更具轻量化，但两者相比，YOLOv5m 能够保持较高

的检测精度，故本文选取 YOLOv5m 作为基准模型。YOLOv5 主要由输入端 (Input)、主干特征提取网络 (Backbone)、颈部特征融合网络 (Neck)、头部检测网络 (Head) 组成，其结构图由图 1 所示。

表 1 YOLOv5 4 种结构测试结果对比

模型	参数量/M	检测精度/%	浮点运算量/B
YOLOv5s	7.5	56.2	13.2
YOLOv5m	21.8	63.2	39.4
YOLOv5l	47.8	66.5	88.1
YOLOv5x	89.0	67.7	166.4

主干特征提取网络由 Focus、CSPDarknet 和金字塔池化 (spatial pyramid pooling, SPP)<sup>[14]</sup> 模块组成，Focus 模块是 YOLOv5 为了减少计算量和提升速度而提出的一种对图片进行切片的操作。图 1 中的 Conv(3, 2) 代表着卷积核为 3，步长为 2 的二维卷积、归一化和激活函数的组合，其中 YOLOv5 将 SiLU 作为激活函数，数学表达式为式 (1) 所示。

$$\text{SiLU}(x) = x \cdot \frac{1}{1 + e^{-x}} \quad (1)$$

YOLOv5 延续使用了 YOLOv4 所采用的跨级部分网络 (cross stage partial network, CSP)<sup>[15]</sup> 结构。SPP 模块旨在提升模型的感受野，对最后一个特征层进行多尺度最大池化，实现不同尺度的特征融合。

颈部特征融合网络采用特征金字塔 (feature pyramid networks, FPN)<sup>[16]</sup> 和路径聚合网络 (path aggregation network, PAN)<sup>[17]</sup> 相结合的组合。这两个结构的结合实现了自顶向下和自下而上的多尺度双向特征融合，从而减少了语义信息的流失，进一步提高了网络的性能。

头部预测网络是对不同尺寸目标进行分类和锚框的回归操作，实现对目标位置和类别的检测。

## 2 改进方法

### 2.1 卷积块的改进

由于需要对人脸进行实时性检测，对设备的性能要求较高，为了减少计算量，提高模型的实时性能，本文采用深度可分离卷积对 Bottleneck 进行一系列改进。

目前较为主流的轻量化网络如 Mobilenet v2<sup>[18]</sup> 等采用深度可分离卷积代替常规卷积操作，相对于常规卷积，其参数量更少、运算成本更低。深度可分离卷积将一个卷积运算分解为两步进行，即深度卷积 (depthwise convolution, dwConv) 与点卷积 (pointwise convolution, pwConv)，该过程如图 2 所示。但深度卷积的一个卷积核只负责一个通道，这会缺少通道间特征的关联，因此后续进行点卷积可实现通道间特征的融合。

对于一个通道数为  $C_m$ 、尺寸为  $H \times W$  的输入特征图，

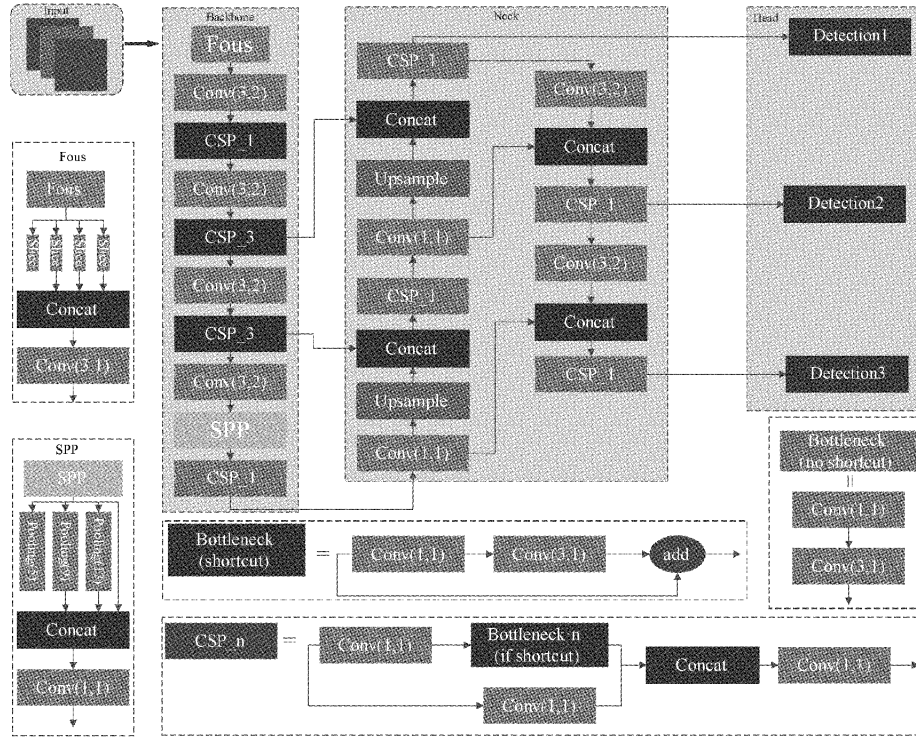


图1 YOLOv5网络结构

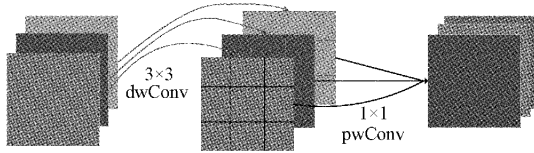


图2 深度可分离卷积示意图

经过常规  $3 \times 3$  的卷积核进行卷积,得到通道数为  $C_{out}$ 、尺寸为  $H \times W$  的输出特征图。常规卷积的计算量为:

$$P_{Conv} = H \times W \times C_{in} \times C_{out} \times 3 \times 3 \quad (2)$$

深度可分离卷积先经过  $3 \times 3$  的深度卷积再经过  $1 \times 1$  的逐点卷积,逐通道卷积的计算量为式(3),逐点卷积的计算量为式(4):

$$P_{dwConv} = H \times W \times C_{in} \times 3 \times 3 \quad (3)$$

$$P_{pwConv} = H \times W \times C_{in} \times C_{out} \times 1 \times 1 \quad (4)$$

综上,以  $C_{out} = 32$  为例,可得到深度可分离卷积和常规卷积的计算量之比为:

$$\frac{P_{pwConv} + P_{dwConv}}{P_{Conv}} = \frac{H \times W \times (C_{in} \times 3 \times 3 + C_{in} \times C_{out} \times 1 \times 1)}{H \times W \times C_{in} \times C_{out} \times 3 \times 3} = \frac{1}{C_{out}} + \frac{1}{3 \times 3} = \frac{1}{32} + \frac{1}{9} \approx \frac{1}{7} \quad (5)$$

在YOLOv5的卷积中基本都满足  $C_{out} \geq 32$ ,由式(5)可知,深度可分离卷积虽然比常规卷积多了一个步骤,但是计算量大大减小。因此将YOLOv5m卷积块中的常规卷积替换为深度可分离卷积,可显著降低模型整体的计算量。

考虑到原YOLOv5网络中的Bottleneck有两种结构,本文也根据是否有Shortcut结构设计了两种结构的dwBottleneck,如图3所示。第1种结构首先将输入的特征按通道均分割为两个分支,其中一个分支不做改变实现特征复用,另一个分支经过一组深度可分离卷积,卷积之后将两个分支拼接(concat)起来,然后再通过通道混洗(channel shuffle)<sup>[19]</sup>以达到不增加计算量的情况下,使两个分支的通道信息得到关联。另一个结构直接经过一组深度可分离卷积,输入和输出通道相等。本文将两种结构分别嵌入原网络CSP\_n模块对应的位置生成dwCSP\_n,模型整体的参数量和计算量大幅减少,实现了模型轻量化的目的。

### 2.2 空间金字塔扩张卷积

YOLO里的SPP模块旨在提升模型的感受野,其对最后一个特征层进行多尺度最大池化,其池化核为大中小3种不同尺寸:  $13 \times 13, 9 \times 9, 5 \times 5$ ,然后将3个池化的输出和模块原输入进行拼接操作得到最终的结果。但是池化操作常常会丢掉特征的位置信息,这对目标定位产生了一定的影响。为了在保留位置信息情况下得到更大的感受野,本文提出空间金字塔扩张卷积(spatial pyramid dilation convolution, SPDC)。SPDC采用混合扩张卷积的思想<sup>[20]</sup>,将多个池化操作替换为扩张卷积,有效地增加了模型的感受野。扩张卷积是卷积核与固定间隔的像素点进行相应的乘法运算,扩张卷积感受野  $N$  的计算公式为:

$$N = n + (n - 1)(r - 1) \quad (6)$$



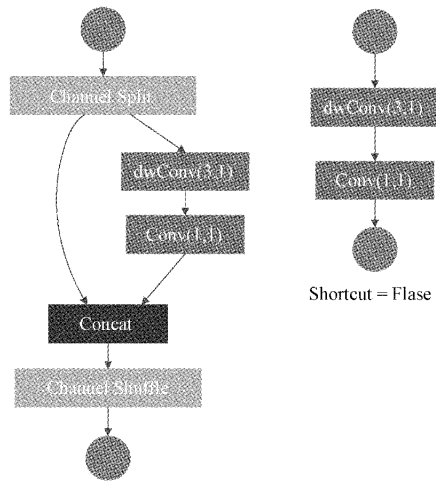


图 3 dwBottleneck 结构

由式(6)可得扩张率为 2、3、4、5 的扩张卷积的感受野为 5、7、9、11。当空洞率过大时,虽然感受野变大但是参与运算的像素点过于稀疏,容易错失关键信息,产生网格效应。根据混合扩张卷积思想,通过设置多个扩张率(扩张率不能有大于 1 的公约数)的扩张卷积串联叠加可以增加感受野的同时避免网格效应<sup>[21]</sup>。受以上思想启发,本文设计了不同尺寸的扩张卷积来代替原池化操作,SPDC 的主要结构由  $1 \times 1$  卷积和多尺度扩张卷积组成,如图 4 所示。

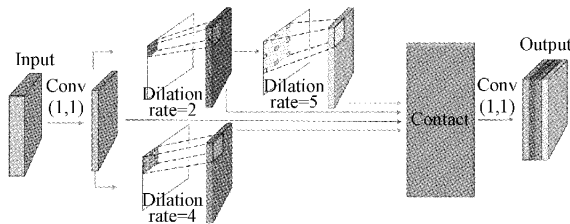


图 4 SPDC 结构图

首先采用一个  $1 \times 1$  卷积减少特征通道数,其次经过两个分支:第 1 个分支根据混合扩张卷积的思想,采用扩张率为 2 和 5 的扩张卷积串联叠加的方式能够获得小尺度特征信息和更加全局尺度的上下文特征信息;第 2 个分支选取扩张率为 4 的扩张卷积得到中尺度的特征信息,通过使用不同扩张率的卷积能够提高模型对不同尺度目标的适应能力;最后将产生的特征图都拼接起来并通过  $1 \times 1$  卷积保证输入输出通道保持一致。为了进一步减少参数量和计算量,SPDC 中间层的通道数设为 SPP 中间层的一半。SPDC 不改变特征图的尺寸,减少了参数量,同时又增强了网络对多尺度信息的提取能力,实现了在不产生额外计算开销的前提下提升了网络的整体性能。

### 2.3 FPN 的改进

FPN 中的上采样方法是最近邻上采样算法,该算法取距离它最近的输入像素值作为变换后的像素值,优点在于操作简单和速度快,但也明显丢失了原特征的局部信息,

不能很好地用来捕捉车前脸检测所需要的丰富语义信息。为此本文引入轻量高效的 Pixshuffle 方法<sup>[22]</sup>来改进上采样操作。Pixshuffle 方法的主要作用是增大特征图的空间分辨率即特征图的尺寸大小,其方法原理如图 5 所示。

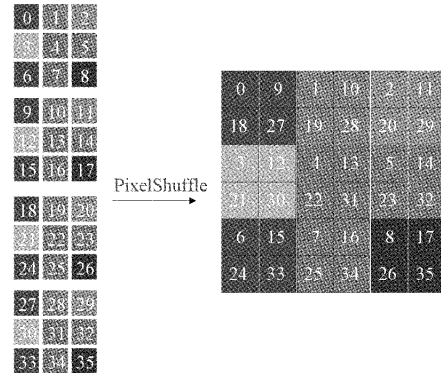


图 5 Pixshuffle 结构图

图 5 所示的操作过程是将在不同特征图中处于相同空间位置的数值在重排列的时候放在相邻位置,即 4 个  $3 \times 3$  的特征图经过 Pixshuffle 操作变成 1 个  $6 \times 6$  的特征图,实现了没有丢失像素点的上采样过程。这一上采样过程使得原特征图的空间相对位置信息在一定程度上得以保留,同时又压缩了通道数。

在 FPN 结构中,不同尺度的特征图由于感受野不同,语义信息也存在着一定的差异,如果将不同尺度的特征直接进行融合反而会使多尺度特征的表达能力减弱。因此,本文提出上下层特征融合(upper and lower layer feature fusion, UFM)模块来使模型更好地聚合上下层特征。为了抑制不同层的冲突信息,本文在上下层特征融合模块中引入了注意力机制。SENet (squeeze-and-excitation network)<sup>[23]</sup>通过学习自动获取每个特征通道的重要程度,利用得到的重要程度来对特征通道加权,但它们通常会忽略特征的位置信息。CBAM(convolutional block attention module)<sup>[24]</sup>是一种常见的混合域注意力机制,通过将通道注意力模块和空间注意力模块串联,在通道上进行全局池化来引入位置信息,但这种方式只能捕获局部信息,无法获取长范围依赖的信息。本文提出基于混合域注意力的上下层特征融合模块,结构如图 6 所示。

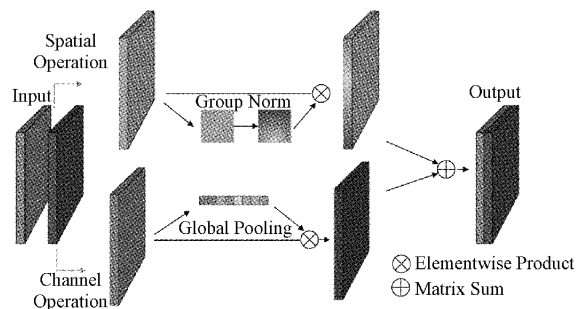


图 6 上下层特征融合模块结构

首先,将输入的上下层特征图按照通道融合,然后分别通过两个注意力分支模块。一个分支  $X_1$  是利用通道间的关系生成通道注意力,而另一个分支  $X_2$  利用空间特征关系生成空间注意力,从而使上下层特征在通道和空间维度上更有效地融合。

通道分支部分的改进措施叙述如下:

在之前的注意力方法中,全局池化常用于通道注意力,其将全局空间信息都压缩进一个通道描述符,通道分支  $X_c$  表示为式(7):

$$X_c = F_{GP}(X_1) = \sigma\left(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_1(i, j)\right) \quad (7)$$

式(7)表示,对  $X_1$  沿着通道维度进行,全局平均池化即对每个通道的特征进行加权求和,将每个通道的权值代替该通道的信息,然后通过一个 Hsigmoid 函数得到每一个通道所对应的通道注意力权重。Hsigmoid 函数是对 sigmoid 函数的近似,相比于 sigmod 的指数运算可以节省大量的计算开销。Hsigmoid 函数的公式为式(8):

$$\sigma(x) = \begin{cases} 1, & x > 3 \\ x/6 + 0.5, & 3 \geq x \geq -3 \\ 0, & x < -3 \end{cases} \quad (8)$$

空间分支部分的改进措施叙述如下:

在之前的注意力方法中,空间注意力常使用卷积层来关注空间信息,但会导致计算量大增,不符合本文选择更加轻量化的初衷,因此本文采用 GroupNorm 来获得空间维度的特征信息,GroupNorm 是按照通道划分组,并计算每一组之内的均值和方差,以进行归一化,然后同样用 Hsigmoid 函数进行加权。空间分支  $X_s$  表示为式(9):

$$X_s = F_{GN}(X_2) = \sigma(GNorm(X_2)) \quad (9)$$

最后将通道分支和空间分支的特征进行有效融合。那么该模块的总输出为  $X_o$ :

$$X_o = X_c + X_s \quad (10)$$

融合了 Pixshuffle 和 UFM 的 PU-FPN 实现了上下层特征根据自适应特征权重的有效融合,从而获得更加重要的特征、抑制无用特征信息,有助于整个模型更好地进行识别。

针对上文所介绍的改进模块,首先将原卷积块 CSP\_n 替换为 dwCSP\_n,然后将 SPP 替换为 SPDC,并放在 Backbone 的最后以进一步增大模型感受野,最后将原 FPN 替换为 PU-FPN,这些改进的模块加入到 YOLOv5 检测网络中,得到了改进的 YOLOv5,结构如图 7 所示。

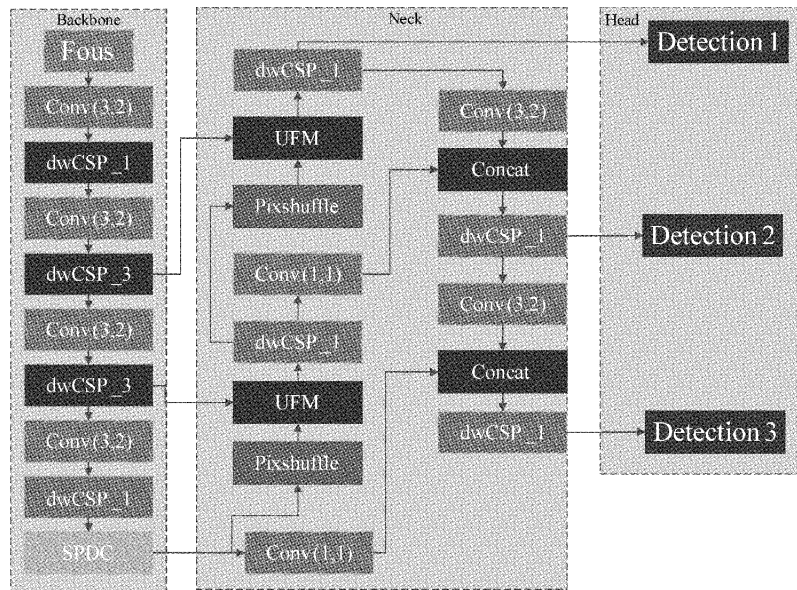


图 7 改进的 YOLOv5 结构

### 3 实验分析与结果

#### 3.1 实验环境与数据集

本文算法是基于 Windows 10 操作系统,采用 Python3.6、PyTorch 1.9.0 深度学习框架完成,硬件环境为 CPU: Intel(R)Core(TM) i7-11700,内存:16 GB,GPU: RTX 3060 12 GB 显卡。

由于车脸检测需要较为清晰的车辆正面信息,本文建立了一个车脸数据集 Car-Data,如图 8 所示。该数据集图片来源

于各大汽车媒体网站和现实生活,清晰地记录了汽车的颜色、车灯、车标等车辆特征。该数据集共有图片 9 500 张,采用 labelImg 软件人工标注了实验所需要的真实框的大小、位置以及类别信息,包含了车灯、9 种汽车颜色、37 种汽车品牌标志,共计 47 个类别的标注框。本文所进行的实验均采用 Car-Data 数据集,其中训练集、测试集和验证集按照 7 : 2 : 1 的比例对数据集所有图片进行划分,各分组数据严格独立。

#### 3.2 实验参数及性能指标

本文算法采用输入尺寸为  $416 \times 416$  的图片进行训练。



图 8 自建 Car-Data 数据集部分图片

整个训练过程共有 150 个 epoch, Batch size 设置为 16, 初始学习率为 0.001, 采用 Adam 函数优化参数, 动量因子 0.937, 采用 Mosaic 数据增强。

要对模型的准确性进行评价时, 首先计算出准确率 (precision, P) 和召回率 (recall, R) 指标, 以 Precision 为横坐标、Recall 为纵坐标所围成的曲线的面积即为单类别的平均精度 (average precision, AP), 所有类别的 AP 值的平均数就是所有种类的平均精度 (mean AP, mAP), 上述指标的具体计算公式为式(11)~(14)。

$$P = Precision = \frac{TP}{TP + FP} \tag{11}$$

$$R = Recall = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 P(R) dR \tag{13}$$

$$mAP = \frac{\sum_{i=1}^N AP(i)}{N} \tag{14}$$

式(11)~(12)中, 若图像此处为物体且算法预测是物体则为 TP(true positive); 若图像此处为背景且算法预测是背景则为 TN(true negative); 若图像此处为物体且算法预测是背景则为 FN(false negative); 若图像此处为背景且算法预测是物体则为 FP(false positive)。式(14)中平均精度 mAP 是将 N 个类别的 AP 值求平均得到的。同时, 使用带下标 s、m、l 的 AP 分别表示在小、中和大尺度目标上的精度。

对于模型的轻量化性能, 使用卷积乘法的浮点运算量 (Flops) 作为对模型的计算复杂度的评价指标; 使用模型的权重 (Weight) 大小作为模型大小的评价指标; 使用模型每秒处理图片的张数 (frames per second, FPS) 作为模型实时性的评价指标。

### 3.3 消融实验

本文以 YOLOv5m 为基础进行了一系列改进, 为了验证各个改进模块的有效性, 本文在相同的实验条件下采用 Car-Data 数据集进行了消融实验, 实验内容如表 2 所示。由表 2 可以看出共进行了 5 组实验, 第 1 组实验使用原始 YOLOv5m 进行实验, 之后的 4 组实验在此基础上进行改进; 第 2 组实验使用 dwCSP\_n 代替原算法的卷积块, 模型权重大小和计算复杂度都下降了约 50%, 实时性能 FPS 提高了约 13 帧, 但 mAP 只下降了 0.22%, 证明了 dwCSP\_n 结构可以有效实现轻量化; 第 3 组在第 2 组的基础上将 SPP 替换为 SPDC 模块, 在牺牲了少许 FPS 的情况下实现了 mAP 的大幅提升, mAP 达到 94.06%; 第 4 组在第 2 组的基础上引入 PU-FPN, mAP 得到进一步地提升, 达到 94.19%; 第 5 组的最终算法相较于原 YOLOv5m 算法, 模型权重大小下降了近一半, 而 mAP 提升了 2.02%, 模型的实时性能也提升了 10 帧左右。

表 2 消融实验结果对比

YOLOv5m	dwCSP_N	SPDC	PU-FPN	权重大小/M	运算量/G	mAP/%	实时帧数
✓				81.3	10.77	92.99	60.88
✓	✓			45.0	5.15	92.77	74.02
✓	✓	✓		42.2	5.03	94.06	73.85
✓	✓		✓	44.7	5.05	94.19	72.24
✓	✓	✓	✓	<b>41.9</b>	<b>4.93</b>	<b>95.01</b>	70.76

每个模块的改进对模型整体的表现都有不同程度的提升, 3 个模块的有效组合对最终的检测精度和实时性能提升效果最优, 该算法证明了在大幅度减少模型大小的情况下也能维持较好的网络精度同时也具有良好的实时性能。为了更加直观的说明改进算法对原算法的提升效果, 我们选取部分图片分别用两个算法进行对比检测, 对比结果如图 9 所示。

由图 9 的 A 组结果可知本文改进算法可以降低错检, 由 B 组结果可知改进算法可以减少漏检, 由 C 组结果可知

改进算法可提升检测精度。原 YOLOv5m 算法对于车标等小目标检测出现错误情况, 而改进算法增强了小目标的语义信息, 能够更好地检测出图中的小尺寸目标。因此, 从可视化的角度说明了改进后的算法比原算法在车脸检测上效果更优。

### 3.4 与其他算法对比

为了比较本文算法与其他目标检测算法的性能, 本文在相同的实验条件下, 包括相同的实验环境及训练策略等, 采用 Car-Data 数据集进行了对比实验。该实验选取





图 9 改进算法检测结果对比

mAP,模型大小、计算复杂度、实时性 FPS 和大中小目标精度( $AP_l$ 、 $AP_m$ 、 $AP_s$ )这几种性能指标进行对比,实验结果如表 3 所示。

对比表 3 中的 5 种网络模型实验结果可知,改进的 YOLOv5m 相较于二阶段网络 Faster R-CNN、一阶段网络 SSD 和 YOLOv4 等热门网络模型,FPS 分别提升了 58.21、33.74 和 29.07,mAP 分别提升了 9.34%、13.67% 和 1.38%, $AP_m$  和  $AP_s$  也有大幅提升,说明空间金字塔扩张卷积和上下层特征融合模块更加有助于多尺度特征的有效提取,有效地提升了网络对车脸信息的检测能力。改进的

YOLOv5m 模型通过引入深度可分离卷积,在模型参数量大幅减少的情况下整体 mAP 优于表 3 中的其他 4 种模型,但是对大目标检测精度  $AP_l$  低于 YOLOv4 和原 YOLOv5m。经分析,Car-Data 数据集中车标和车灯等中小目标所占比重较大,而过深的网络和过多的卷积操作可能不适合检测细密的中小目标特征,所以改进的 YOLOv5m 牺牲了大目标检测精度换来了车脸检测整体 mAP 的提升。综合以上实验结果表明,改进的 YOLOv5m 在车脸检测任务中确实有不错的性能,这样一个既有良好性能又相对轻量的网络模型,更适合在车检场景中进行车脸的检测识别。

表 3 不同网络模型性能对比

模型	输入尺寸	权重大小/M	运算量/G	mAP/%	实时帧数	$AP_s$ /%	$AP_m$ /%	$AP_l$ /%
Faster R-CNN	416×416	109.8	185.20	85.67	12.55	0.380	0.427	0.548
SSD	416×416	114.0	59.61	81.34	37.02	0.334	0.373	0.486
YOLOv4	416×416	244.9	30.17	93.63	41.69	0.498	0.551	0.710
YOLOv5m	416×416	81.3	10.77	92.99	60.88	0.500	0.544	0.707
本文	416×416	<b>41.9</b>	<b>4.93</b>	<b>95.01</b>	<b>70.76</b>	<b>0.539</b>	<b>0.574</b>	<b>0.700</b>

## 4 结 论

为了提升车灯检测仪的检测效率同时检测出关键的车辆信息防止车辆代检,本文建立了多品牌车灯检测数据集 Car-Data,并提出一种改进 YOLOv5m 的轻量化车灯检测方法。实验结果表明本文算法相较于原 YOLOv5m 算法大幅度降低模型大小的同时又在平均精度上有明显提高,实现了准确度和实时性的兼顾,能够满足实际车灯检测应用的需要。

实验的数据集还需进一步丰富,之后的研究会原数据集上进一步增加车辆的品牌和颜色种类以增强模型的泛化能力。下一步将在嵌入式设备上对网络结构进行改进,从而使网络进一步降低模型大小的同时具有较优的准确度和较好的实时性,使其更易在实际环境部署使用。

### 参考文献

- [1] 李慧,白鹏飞,李世晓,等. 基于车灯特征的夜间车辆检测方法[J]. 电子测量技术, 2020, 43(14): 89-95.
- [2] 周晓彦,王珂,李凌燕. 基于深度学习的目标检测算法综述[J]. 电子测量技术, 2017, 40(11): 89-93.
- [3] GIRSHICK R, DONALD J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [6] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [8] 李国进,胡洁,艾娇燕. 基于改进 SSD 算法的车辆检测[J]. 计算机工程, 2022, 48(1): 266-274.
- [9] PIAO Z, ZHAO B, TANG L, et al. VDetector: An effective and efficient neural network for vehicle detection in aerial image [C]. 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), IEEE, 2019: 1-4.
- [10] 张宝朋,康谦泽,李佳萌,等. 轻量化的 YOLOv4 目标检测算法[J]. 计算机工程, 2022, 48(8): 206-214.
- [11] ALSANABANI A A, SAEED S A, AL-MKHLAFI M, et al. A low cost and real time vehicle detection using enhanced YOLOv4-Tiny [C]. 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), IEEE, 2021: 372-377.
- [12] 张上,王恒涛,冉秀康. 基于 YOLOv5 的轻量化交通标志检测方法[J]. 电子测量技术, 2022, 45(8): 129-135.
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [14] HE K, ZHANG X, REN S, 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.
- [15] 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.
- [16] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [17] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [18] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [19] 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.
- [20] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J]. ArXiv Preprint, 2015, ArXiv: 1511.07122.
- [21] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation [C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018: 1451-1460.
- [22] TALAB M A, AWANG S, NAJIM S A M. Super-low resolution face recognition using integrated efficient sub-



- pixel convolutional neural network (ESPCN) and convolutional neural network (CNN) [C]. 2019 IEEE International Conference on Automatic Control and Intelligent Systems(I2CACIS), IEEE, 2019: 331-335.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [24] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV),

2018: 3-19.

### 作者简介

贾玮迪,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail: wdjia@mail.ynu.edu.cn

余鹏飞(通信作者),副教授,主要研究方向为模式识别、深度学习。

E-mail: pfyu@ynu.edu.cn

余国豪,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail: ygh3702@163.com