

基于改进 YOLOv5s 汽车驾舱遗忘物检测

吴继薇¹, 焦良葆^{1,2}, 焦波¹, 祝阳¹, 高阳¹

(1. 南京工程学院 人工智能产业技术研究院, 南京 211167;

2. 江苏省智能感知技术与装备工程研究中心, 南京 211167)

摘要: 针对目前汽车驾舱内遗忘物检测精度不高、速度慢和检测效果差的问题, 提出一种基于改进的 YOLOv5s 汽车驾舱遗忘物的检测方法; 该检测方法将 YOLOv5s 作为基础网络, 在此基础上进行改进; 首先本实验将 SE 注意力模块添加到 Backbone 网络中, 加强模型对通道信息的关注提升目标检测性能; 其次改进空间金字塔池化模块, 将原有的 SPPF 模块改进为 SPPCSPC 模块, 使得该网络更加关注待检测目标的特征; 最后同时引入 GSConv 层, 能够缓解 DSC (深度可分离卷积) 的缺陷, 并充分利用 DSC 的优势, 在小目标检测方面取得明显的提升效果, 既保证了语义信息又平衡了模型的准确性, 也提升了检测速度; 通过训练结果说明, 改进后的网络与原 YOLOv5s 网络相比, 其平均精度均值 mAP 提高了 2%, 查准率提升了 3.5%; 改进后的网络具有良好的提升效果, 表明了该方法的有效性。

关键词: 遗忘物; YOLOv5s; 注意力机制; SE; GSConv; SPPCSPC

Automotive Cockpit Oblivion Detection Based on Improved YOLOv5s

WU Jiwei¹, JIAO Liangbao^{1,2}, JIAO Bo¹, ZHU Yang¹, GAO Yang¹

(1. AI Industrial Technology Research Institute, Nanjing Institute of Technology, Nanjing 211167, China;

2. Jiangsu intelligent perception technology and equipment Engineering Research Center, Nanjing 211167, China)

Abstract: Aimed at the low detection accuracy, slow speed and poor detection of amnesia in current automobile cockpits, an improved YOLOv5s automotive cockpit oblivion detection method forgetting is proposed. The detection method uses YOLOv5s as a basic network and improves on this basis. Firstly, the SE attention module is added to a backbone network to strengthen the model's attention to the channel information and improve the target detection performance. Secondly, the space pyramid pool module is improved, and the original SPPF module is improved to the SPPCSPC module, which makes the network pay more attention to the characteristics of the target to be detected. Finally, the GSConv layer is simultaneously introduced to alleviate the detects of depth wise separable convolution (DSC) and fully utilize the advantages of the DSC to achieve significant results in detecting small targets. It ensures the semantic information, balances the accuracy of the model, and improves the detection speed. Experimental results show that compared with the original YOLOv5s network, the average mean average precision (mAP) of the improved network is increased by 2%, and the detection accuracy by 3.5%. The improved network has a good improvement effect, which shows the effectiveness of the method.

Keywords: oblivion; YOLOv5s; attention mechanism; SE; GSConv; SPPCSPC

0 引言

现如今深度学习发展越来越快速, 目标检测已经成为新的热潮, 目前的目标算法主要分为两类: 一类是以 R-CNN 算法^[3]、Fast R-CNN 算法^[4]、Faster R-CNN 算法^[5]和 Mask R-CNN 算法^[6]为主要的双阶段算法; 另一类是以 SSD 算法^[7]和 YOLO 系列算法为代表的单阶段算法。

YOLO (You Only Look Once) 是一种用于实时目标检测的深度学习算法, 它的主要优点在于能够实时处理图像和视频, 并同时检测多个不同类别的物体。YOLO 设计的

目标之一是实现实时性。它可以在处理图像或视频流时高效地检测目标, 因此在自动驾驶、监控系统、无人机等应用中非常有用。由于 YOLO 只使用一个神经网络, 而不像其他方法需要多个阶段的处理, 这使得 YOLO 的设计相对简单, 易于训练和部署。YOLO 不仅能够检测输入图像的不同尺度上检测目标, 而且可以检测多个不同的物体类别, 这意味着 YOLO 不但检测不同尺寸的目标, 例如从小物体到大物体, 而且在交通标志、动物、人、车辆等的检测中也广泛运用着。在训练过程中, YOLO 不仅检测目标, 还

收稿日期: 2023-09-03; 修回日期: 2023-10-16。

基金项目: 江苏省自然科学基金资助项目(BK20201042); 江苏省政策引导类计划项目(SZ-SQ2020007)。

作者简介: 吴继薇(1998-), 女, 硕士。

焦良葆(1972-), 男, 教授, 硕士生导师。

引用格式: 吴继薇, 焦良葆, 焦波, 等. 基于改进 YOLOv5s 汽车驾舱遗忘物检测[J]. 计算机测量与控制, 2024, 32(9): 27-35.

回归出目标的边界框，即目标的位置和尺寸信息。为了去除重叠的检测框，YOLO 还使用非极大值抑制 (NMS) 来筛选最终的检测结果。YOLO 算法在各种领域中都有广泛的运用，例如自动驾驶，YOLO 可以用于自动驾驶系统中检测道路上的车辆、行人、交通标志和障碍物，以帮助车辆做出决策。例如视频监控，监控摄像头中使用 YOLO 可以实时检测并跟踪人员、入侵者或异常事件，从而提高安全性。例如医疗影像分析，在医疗领域，YOLO 可用于医学影像中的病灶检测和器官分割。例如物体跟踪，YOLO 还可以用于目标跟踪，通过不断更新检测目标的位置，以实现目标的实时跟踪

为了提高汽车内遗忘物的检测精度，本文使用 YOLOv5s 6.0 版本的 YOLOv5s (You Only Look Once v5s) 对模型进行改进。通过在原骨干网络上添加了 SE 注意力模块，增强了模型对通道信息的关注，提高了目标检测的性能；其次改进空间金字塔池化模块，将原有的 SPPF 模块改进为 SP-PCSPC 模块，通过增加一点计算量来进一步提升检测模型的精度；最后同时引入 GSCONV 层，在保证语义信息的同时平衡模型准确性与检测速度。

1 YOLOv5 算法介绍

YOLOv5 是在 YOLOv4 的基础上进行了优化和改进，以实现更高效的目标检测算法。YOLOv5 算法根据子模块的深度和宽度设计了 4 种不同的模型，分别命名为 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x，这些模型

在检测精度和模型大小方面逐步提高。而且不同深度和不同宽度的模型能够适应不同的数据集需求。在本实验中，由于数据较少且出于计算效率考虑，我们选择了 YOLOv5s 模型作为基础模型进行研究。YOLOv5s 的网络结构主要包括 4 个部分：输入端、Backbone 骨干网络、Neck 网络和输出端。YOLOv5s 6.0 版本的模型结构如图 1 所示。

YOLOv5 的输入端使用了 Mosaic 数据增强技术，这种方法将四张图像以随机缩放、裁剪和排列的方式进行组合，这对于检测小目标非常有效。这不仅增加了数据集的多样性，还降低了 GPU 的负担。此外，在 YOLO 算法中，数据集的不同初始的锚框的尺寸也不一样，因此在数据集训练过程中网络将输出初始锚框的值并生成预测框，将预测框与真实框对比计算两者间的差异，最后利用该信息来反向更新从而迭代网络参数。在 YOLOv3、YOLOv4 中，不同数据集的原始数据的锚框单独用程序运行，但是 YOLOv5 将此功能镶嵌代码内，并且在后面的训练中自动计算不同训练集的最佳锚框值。若是觉得计算的锚框效果不如意可以将自动计算功能关闭。目前的一些目标检测算法中，不同的图片有不同的大小。因此常见的方法是将原始图片经过图像处理达成标准尺寸的图片，置于检测网络。但是缩放填充后两端的黑边大小尺寸都不一样，容易造成信息出错，影响训练推理的速度。因此，对 YOLOv5 中的代码进行改进，能够使得原始的图片能自适应的添加最少的黑边，该方法称为自适应图片缩放。

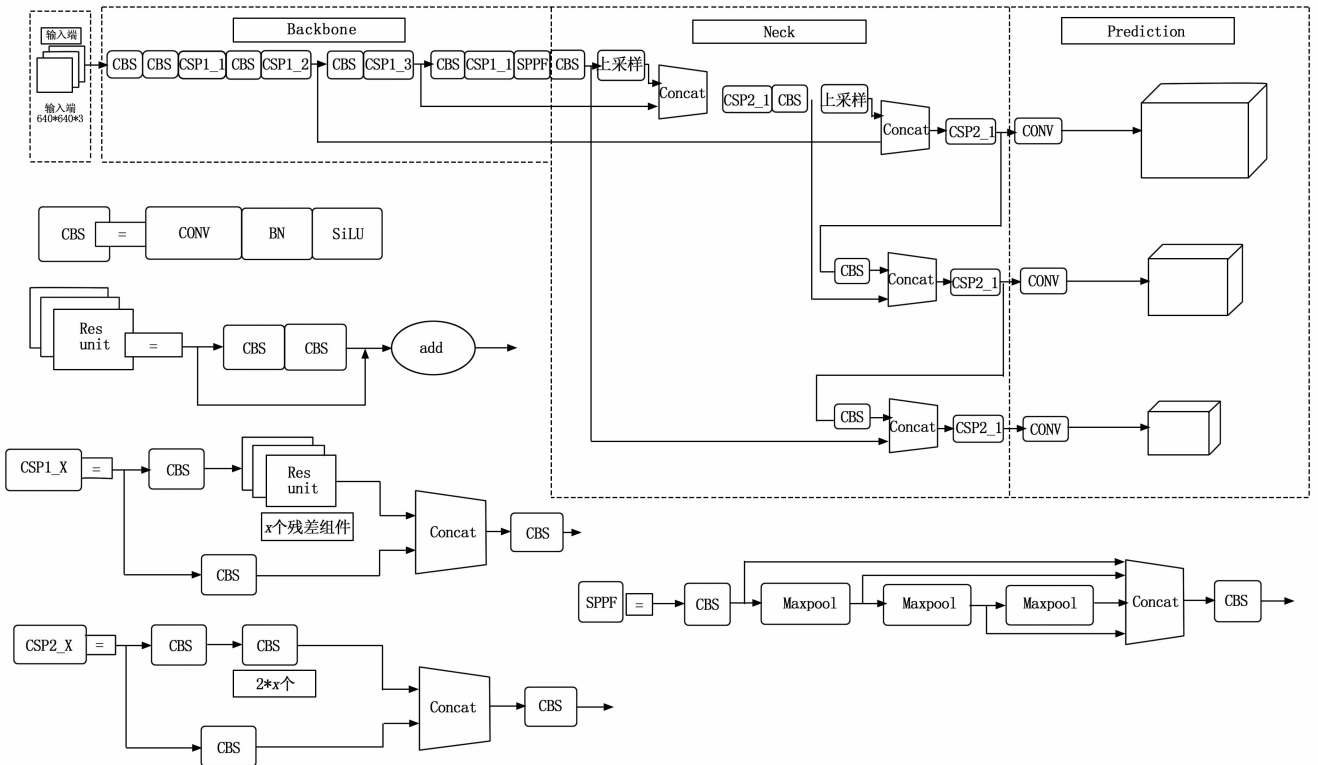


图 1 YOLOv5 6.0 版本网络结构图

Backbone 骨干网络主要的作用是提取图片中的信息供后面的网络训练使用, Backbone 中的主要结构有 Conv 模块、C3 模块、SPPF 模块。相比较于早期的版本, YOLOv5s 6.0 把网络中的第一层 Focus 模块替换成一个卷积层, 两者理论上是等价的, 但在某些 GPU 设备中新替换的卷积层更加高效。C3 模块即 CSP 结构, 细分为两种结构: 一种是 CSP_X1 结构, 这种结构主要使用在骨干网络, 而模型特征图在进入 CSP_X1 结构后又分成了两条支路, 一条是直接通过 CSP 结构, 另一条是在通过 CSP 结构之后再通过 x 个残差进行卷积, 上述两条路通过 concat 最终形成 C3 模块; 另外一种 CSP_X2 结构主要用于 Neck 网络中, 它是将 CSP_X1 结构中的 x 个残差卷积替换成 CBS 结构。YOLOv5s 6.0 版本中空间金字塔将原来的 SPP 模块替换成了 SPPF 模块, 采用多个小尺寸池化核级联代替原来模块中单个大尺寸池化核, 实现了在保持原有的单元性能的同时, 也提升了运行的速度。

目前的 Neck 网络是包含 FPN+PAN 结构的, 首先由 FPN (feature pyramid network) 层通过上采样的方式自顶向下来传递特征图的语义特征, 再由 PAN (path aggregation network) 层通过下采样的方式自底向上传递定位特征, 以此将语义特征和定位特征进行融合, 来改善模型的特征融合能力及提高目标特征的鲁棒性。在 Head 输出部分, 利用 Neck 网络传输不同尺度的特征图像, 采用非大极值抑制技术, 将目标区域筛选出来并将其划分为 3 个预测层, 实现了对目标区域的预测, 以此提高对目标区域的检测能力。

首先, 智能驾舱环境比较复杂且经常会有物体遗忘在驾舱内, 影响驾舱安全, 我们可以通过加入注意力机制 SE (Squeeze and Excitation) 来加强模型的关注度, 提高检测精度。其次, 改进空间金字塔池化模块, 将原有的 SPP 模块改进为 SPPCSPC 模块, 通过增加一点计算量来进一步提升检测模型的精度。最后, 计算机不能通过简单无休止的增加模型参数的数量来构建模型, 因此如何平衡模型计算成本和检测精度成为轻量话网络的主要目标。深度可分离卷积 DSC (depth-wise separable convolution) 能够有效果的提升检测速度, 但是不可以充分利用不同通道的空间位置特征信息。本文将标准卷积 SC (Standard Convolution) 替换成 GSConv, 模型参数量下降。在此又引入 Gsbottleneck 搭建的 VoV-GSCSP 轻量级网络, 与 GSConv 组成纤细颈部网络 (Slimneck), 极大地平衡精度和速度的双向提升, 使网络减轻模型重量的同时精度也得到很好的提升。

2 改进策略

2.1 SE 模块

为了加强模型对通道信息的关注, 在主干网络引入注意力机制 SE (Squeeze and Excitation) 模块。SE 模块通过 Squeeze 和 Excitation 两个模块进行学习, 从而获取每个通道的重要程度, 然后使用得到的重要程度来提升特征, 并

抑制当前任务不重要的特征, 因此为后面的注意力机制建立了良好的基础。SE 模块的优势在于, 它不会通过破坏原来的主体结构为代价来改善网络的性能, 而是在原来的基础上能够进行新的工作。

首先, 经压缩操作对空间的维度进行压缩, 将每个特征图全局平均池化并压缩成一个实数, 如此就可以将感受野扩展到全局范围。这类实数可由下列公式得到:

$$z_k = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_k(i, j) \quad k = 1, 2, \dots, C$$

式中, u 为卷积后的特征图, C 为 u 通道数, $W \times H$ 为 u 的空间维度。

其次, 激励操作获取压缩后的实数, 使用两个 FC (Fully Connected) 全连接层增加模块的非线性。先通过第一个全连接层对模块进行降维, 再经过 ReLU 激活后通过第二个连接层对模块进行升维, 最后使用激活函数 sigmoid。过程如下所示:

$$s = \sigma[W_2 \delta(W_1 z)]$$

式中, δ 是非线性激活函数 ReLU, W_1 和 W_2 分别为两个全连接层的参数, σ 为 sigmoid 激活函数。

最后, 使用通道权重相乘将 SE 模块计算出来的权重值分别和原特征通道相乘得到新标定的特征:

$$\bar{x}_k = s_k \cdot u_k \quad k = 1, 2, \dots, C$$

文章在原始的网络模型的骨干网络的基础上, 引入了 SE 注意力模块, 如图 2 所示, 构建基于通道之间的特征映射, 提高有用信息在特征层中所占的比重, 从而提高车辆遗忘物的识别精度, 降低模型检测时的误检概率。

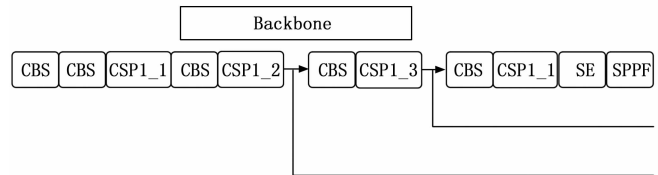


图 2 嵌入 SE 模块

2.2 空间金字塔池化的改进

2.2.1 SPP

YOLO 算法从 v3 开始借鉴空间金字塔池化的思想, 在 SPP 模块的帮助下完成了局部特征和全局特征的 Feature map 级别的融合, 增强 Feature map 的表达力, 使得算法的检测精度优于原来算法。如图 3 所示, SPP 模块由 4 个分支组成: 1) 第一是输入到输出; 2) 第二是池化核为 5×5 的最大池化; 3) 第三是池化核为 9×9 的最大池化; 4) 第四是池化核为 13×13 的最大池化。这些池化的步距都为 1, 即池化之前需要进行 padding 填充, padding 值从第二分支到第四分支分别为: 2、4、6, 最后池化所得到的特征图的尺寸大小和深度都没有变化。

YOLOv5 在 6.0 版本之后对 SPP 结构进行了一些改进, 将原来的 SPP 模块改为了 SPPF, SPPF 是基于 SPP 结构所

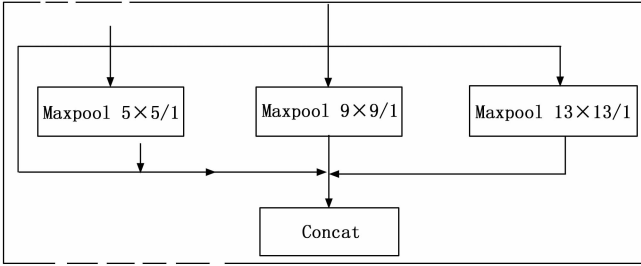


图 3 SPP 结构

改进而提出的，只是在结构上略有一些差异，将 SPP 改为 SPPF 之后，模型的计算量变小了很多，模型的检测速度得到了一定的提升。SPPF 的结构图如图 4 所示。相较于 SPP 中并行通过多个不同大小的 Maxpool 层，SPPF 的结构是将 3 个 5x5 的 Maxpool 层进行串行，相比于 SPP 的并行，SPPF 的串行对计算结果没有任何影响，例如在 SPPF 中就是将 SPP 中的 9x9 的 MaxPool 层、13x13 的 MaxPool 层改为两个和 3 个 5x5 的 MaxPool 层。

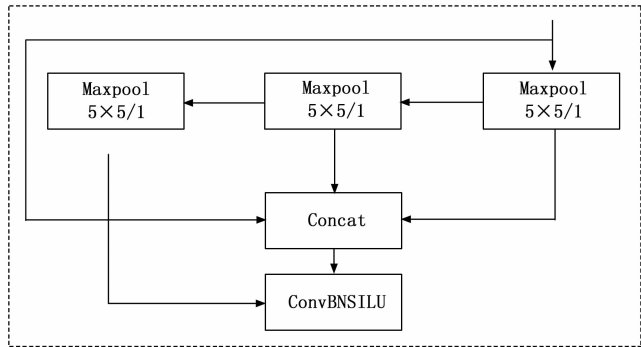


图 4 SPPF 结构

2.2.2 SPPCSPC

SPPCSPC (Spatial Pyramid Pooling, Cross Stage Partial Channel) 包含了两个部分：空间金字塔池化 SPP (Spatial Pyramid Pooling) 和跨阶段部分通道连接 CSPC (Cross Stage Partial Channel)。SPPCSPC 本质上还是借鉴的 SPP-

Net 思想所设计的，跟 YOLOv4 和 YOLOv5 中的 SPP 模块所不同的是需要三次卷积之后再行最大池化，在连接操作之后，还需要进行两次卷积操作，之后再与经过一次卷积的特征图进行拼接，SPPCSPC 的结构图与添加的位置示意图如图 5~6 所示。

从图 5 可以看出，相比于 YOLOv5 中 SPPF 的最大池化窗口大小都为 (5x5)，在 SPPCSPC 结构中最大池化窗口大小调整为 (5x5)、(9x9)、(13x13)、(1x1)，从而扩大感受野，更加便于区别大目标和小目标。在对比 SPPCSPC 与 SPPF 模块的性能差异之后，为了提升基于 YOLOv5 所设计的道路目标检测算法的精度，本研究采用 SPPCSPC 模块来替代原 YOLOv5 算法中的 SPPF 模块。

SPP 部分会将特征图分成不同尺度的子区域进行池化，从而得到固定尺寸的特征图，这样可以减少后续计算量。而 CSPC 部分则会将特征图分成两部分，一部分进行卷积操作，另一部分则直接输出，这样可以减少卷积计算的通道数，从而进一步减少计算量。

2.3 Neck 部分引入 GSConv_slimneck 范式

对于驾舱环境内遗留物体的检测是困难的且工作环境是复杂的，在各种不利于检测的环境下，需要考虑到网络的轻量化，也要保证达到足够的精度。深度可分离卷积 (DSC, depthwise separable convolution) 是一种轻量化卷积形式。它将完整的卷积分成两个部分：深度卷积 (DW Conv, depth wise convolution) 和逐点卷积 (Pointwise Convolution)。使用 DSC 操作可以减少参数，DSC 与 SC 的目的相同，都可以提取图像中的特征信息，但是 DSC 降低了参数和计算的复杂程度。但是，在计算的过程中，输入图像的通道信息是分开的，这就是为什么 DSC 的特征融合能力比 SC 的差的原因。为了使得 DSC 的输出尽可能接近 SC，选择使用 GSConv (Ghost-shuffle Convolution) 来改善这一缺点。与 DSC 卷积方式不同的是，GSConv 中使用 shuffle 将 SC 密集卷积操作生成的信息与 DSC 生成的信息融合在一起，这样的话使得 SC 的信息完全融合在 DSC 的输出，在不同的通道上均匀的交流特征信息。

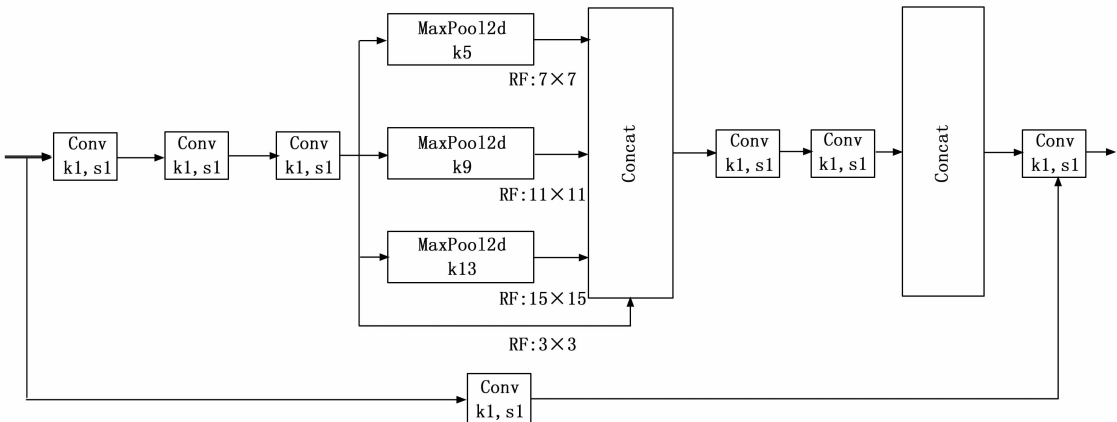


图 5 SPPCSPC 结构

为了使 DSC 的输出尽可能的接近 SC, 使用 SC、DSC、和 shuffle 混合卷积的方式来改善 DSC 的缺点, 该方式称为 GSConv。其结构如图 7 所示。

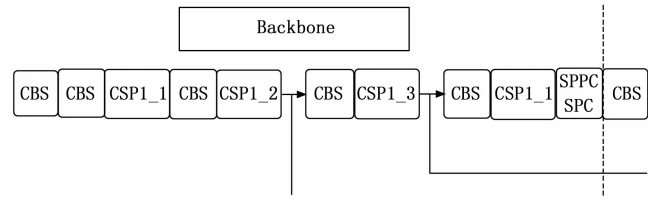


图 6 嵌入 SPPCSPC 结构

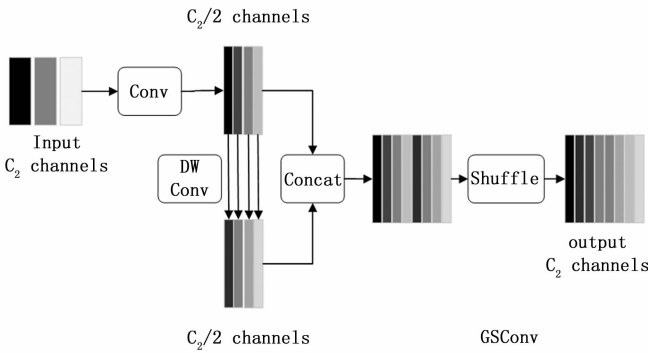


图 7 GSConv 结构图

为了加快预测速度, 卷积神经网络中的输入图片几乎必须在 Backbone 中经历空间信息逐步向通道传输的一个转换过程。而在这个过程中特征图的宽度和高度的压缩与扩展都会造成信息丢失。GSConv 结构中 DSC 层和 shuffle 增强的非线性表达能力尽可能以较低的信息损失来保留转换过程中的通道之间的链接, 但如果将这种结构用于模型推理阶段的每个阶段, 那么将导致模型的网络化程度更深, 越深的网络会减慢数据流动能力, 这将极大地提高模型推

理的时间。在达到 Neck 的时候, 特征图的通道维度达到了最大, 宽度和高度维度达到了最小, 已经不再需要转换。因此, 在 Neck 中加入 GSConv 会减少重复的信息, 最大限度地减小了 DSC 缺陷对模型的负面影响, 有效地利用了 DSC 的优点, 注意力模块效果更好。因此用 GSConv 结构代替 SC 层, 利用轻量级模型结构的特点, 使得模型检测精度明显提高, 成本也逐渐降低。所以加入 GSbottleneck 模块, 使用一次性聚合的方式设计了 VoV-GSCSP 模块。GSbottleneck 模块结构、VoV-GSCSP 模块结构分别如图 8~9 所示。添加的鬼影卷积 GSConv 如图 10 所示。

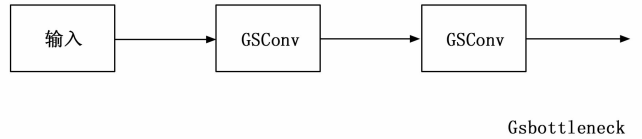


图 8 GSbottleneck 模块结构

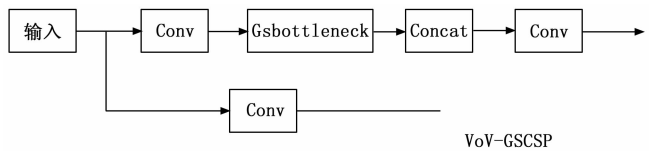


图 9 VoV-GSCSP 模块结构

3 实验结果与分析

3.1 数据集的准备

本实验所使用的数据集包含 1 200 张样图, 均是用林柏视红外摄像头连接电脑拍摄的, 产品规格描述为: $3.0 \mu\text{m} \times 3.0 \mu\text{m}$ 的像素尺寸, 像素面积为 $5\ 808 \mu\text{m} \times 3\ 288 \mu\text{m}$, 曝光时间为 33.33 ms/fps, 镜头焦距为 2.8 MM, 广角无畸变镜头: 对角 120 度、水平 90 度、垂直 51 度, 分辨率为

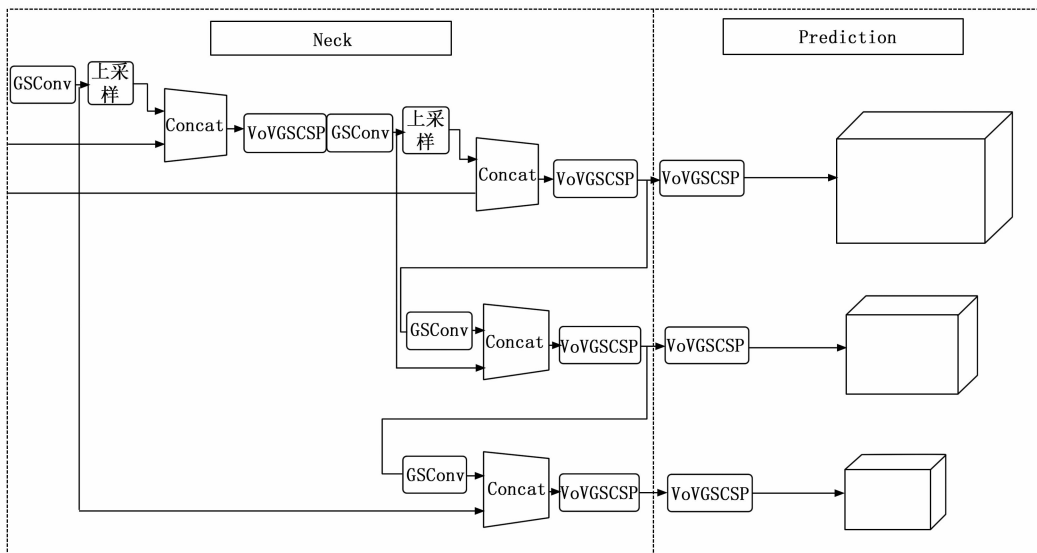


图 10 嵌入 GSConv 卷积

1 920×1 080, 拍摄地点为停车场汽车内和正常阳光露天下的场景。包含可见光下的图片 468 张, 红外图像的图片有 732 张, 比例大约是 3: 5, 具体图示照片如图 11 所示。将遗忘物图片数据集类型分为 3 种: 手机 (phone)、钱包 (wallet) 和钥匙 (key)。并对训练的数据集进行标注。如图 12 所示。

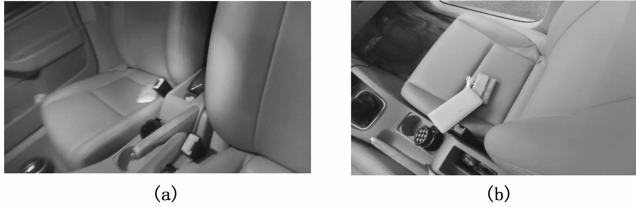


图 11 数据集图片

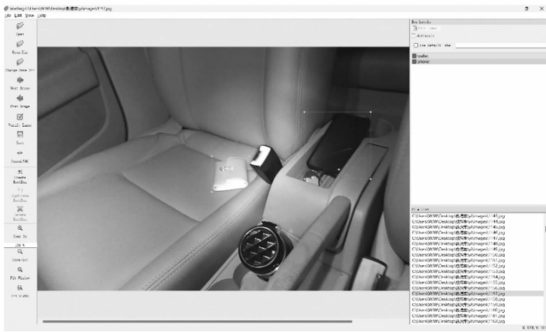


图 12 Labelme 软件操作界面

数据集的标签分别命名为 phone、wallet、key, 将数据集图片存至名为 images 文件夹中, 将标注完的数据集且以 xml 为后缀的文件存至名为 labels 文件中, 通过 python 脚本将以 xml 后缀的文件转化成 txt 文件。最后将 images 和 labels 文件夹中的文件按照 6: 2: 2 的比例划分, 将此文件夹命名为训练集、验证集、测试集。

3.2 实验环境的搭建及流程

实验训练过程中调用 GPU 进行训练, 过程中所用配置如表 1 所示。

训练设备的参数具体设置如下, 初始输入的图片大小为 640 × 640, 训练轮次为 300 轮, 每批次处理 48 张图片。交并阈值为 0.5, 学习率初值为 0.01, 在没有使用预训练权重情况下对其参数进行训练。

表 1 实验环境

名称	配置
操作平台	Ubuntu18.04
CPU	AMD Ryzen Threadripper 3970X
GPU	NVIDIA GeForce RTX 3090 显卡 * 3
显存	24 G
内存	128 G
编译环境	Python

第一步, 标注选取的数据集使用 Labelme, 并按照 3: 1: 1 的比例将标注好的数据集分成训练集、测试集和验证集。第二步, 使用 Python 编程语言, 并借助 Pytorch 深度学习框架改进 YOLOv5s 模型, 进而选择最佳模型对数据集进行训练。第三步, 在测试集上使用已训练好的权重文件进行模型测试; 最后一步, 将训练得到的最终模型结果与其他的模型进行比较。

3.3 评价指标

现如今, 在目标检测区域内训练结果通常使用查准率 (P , precision)、查全率 (R , recall)、平均精度均值 (mAP , mean average precision)、训练所用的权重文件大小 (MB) 四个指标对模型训练的性能进行评价。查准率 P 是指在预测出来的正类样本中, 真实情况下的正类样本所占的比重, 查全率 R 是指在真实情况下的正类样本中, 预测出来的正类样本所占的比重。精确度考虑了对数据集的鉴别真伪的能力, 也就是对假象 (如预测为正, 实际为负) 的假象进行剔除的能力。查全率是指能够容纳一个真实向量的能力, 也就是能够容许一个真实向量与虚假向量有多大的相似性。其中 P 、 R 、 mAP 公式如下所示:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{\sum_{i=1}^K \sum_{r_i=1}^{r_i-1} (r_{i+1} - r_i) \rho_{interp}(r_i + 1)}{K} \quad (3)$$

式中, TP (True positives) 代表检测的正确数量; FP (False positives) 代表检测的错误数量; FN (False negatives) 代表没有被检测出来的正确数量。其中 P 是正确分类的数量与已识别出的样本的数量的比值; mAP 相对于 P 、 R 的不稳定性更加反映了全局性能。 K 代表类别的数量, r_i

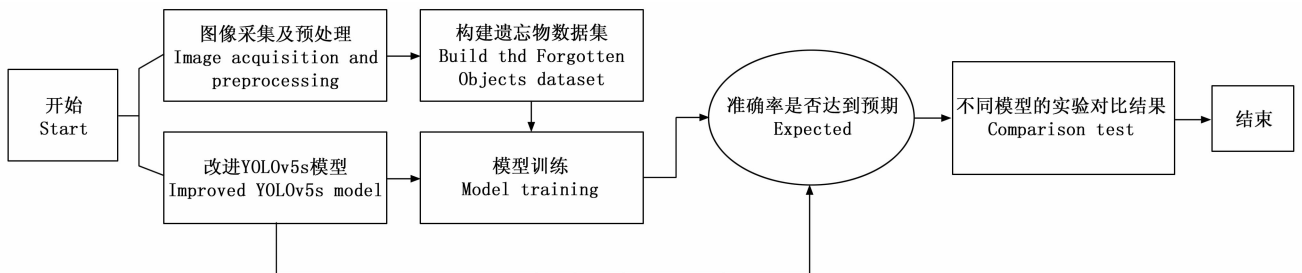


图 13 实验流程图

是按照上升的查准率插值段第一个插值处对应的查全率。

3.4 实验结果及评估分析

1) 对比试验:

表 2 对比实验结果

算法	查准率/%	查全率/%	$mAP@0.5/\%$	权重/MB
SSD	71	74.6	75	93.6
YOLOv3	73.1	76.8	74.6	182
YOLOv5l	78.5	75.1	76.3	97.7
YOLOv5m	78.8	74.3	75.8	45.8
YOLOv5x	84.5	73.2	77.0	210.6
YOLOv5s	89.5	73.7	77.6	14.5

2) 注意力模块实验对比:

本实验将 SE 注意力模块添加到 Backbone 网络中, CSP1_1 之后, 并在添加其他的注意力机制作对比实验, 得到的实验结果如表 3 所示, 从表格中的实验数据可知模型训练的速度明显加快, 查准率和 mAP 值分别提升了 1.5% 和 1.5%。相对于其他注意力模块, 添加了 SE 注意力机制提升性能更加明显。

表 3 融合不同注意力模块对比实验

算法	查准率/%	查全率/%	$mAP@0.5/\%$	FPS	权重/MB
YOLOv5s	89.5	73.7	77.6	133	14.5
YOLOv5s+SE	91	72.6	79.1	143	14.5
YOLOv5s+CBAM	90.9	71.9	78.1	141	14.5
YOLOv5s+CA	88.6	73.9	78.1	167	14.5
YOLOv5s+C3SE	85.7	75.5	77.6	128	14.5
YOLOv5s+ECA	86.1	75.5	78.4	175	14.5

3) 轻量化实验对比:

将 ShuffleNetV2、MobileNetV3 与本文使用的 Slim_neck 进行对比, 将 YOLOv5 特征提取层替换为 ShuffleNetV2 的结构之后, 发现模型检测速度虽然变快了, 但是 mAP 值和查全率分别降低了 5.1% 和 6.7%, 模型权重比也下降了 6.4 MB; 替换为 MobileNetV3 后模型检测速度稍微变快了但是查准率和查全率全部都下降且精度 mAP 值更是下降了 4.1%, 模型权重体积也下降了 7 MB; 而将 GSConv_slimneck 替换掉原来的特征层后, 模型权重体积下降了 0.9 MB, 并且在保证了模型轻量化的同时模型训练的精度 mAP 的值也增加了 3.9%。因此说明了 GSConv+slimneck 能够轻量化的保证 YOLOv5 特征融合的有效性。

表 4 轻量化网络对比表

网络	查准率/%	查全率/%	$mAP@0.5/\%$	FPS	权重/MB
YOLOv5s	89.5	73.7	77.6	133	14.5
ShufflenetV2	89.8	67	72.5	181	8.1
MobileNetV3	86	69.8	73.5	137	7.5
GSConv_slimneck	88.2	75.6	81.5	178	13.6

4) 消融实验:

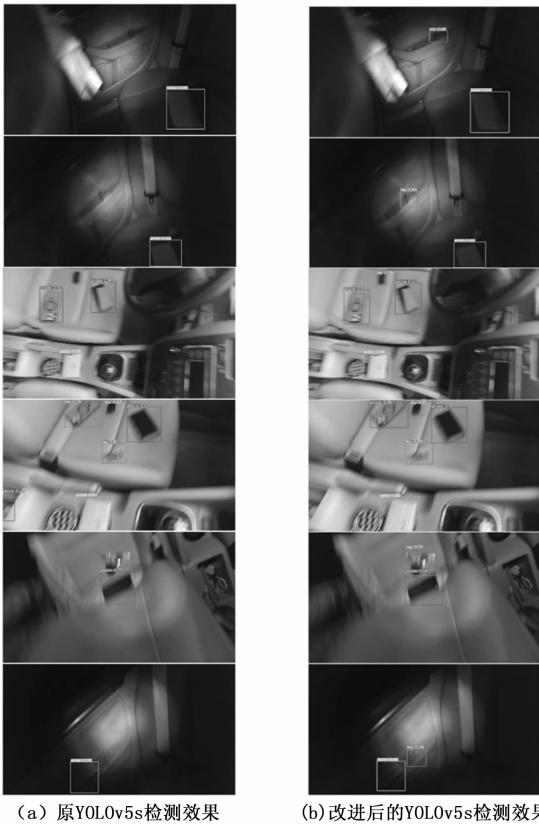
与原网络 YOLOv5s 相比, 实验提出三种改进方式。首先加入注意力机制 SE、将原网络中的空间金字塔 SPPF 替换成 SPPCSPC 并和纤细颈部网络 GSConv_slimneck 互相结合起来。然后, 将单独模块实验组合与单独模块下的效果进行消融实验, 结果如表 5 所示。从表中的实验数据可知, 加入注意力机制 SE 会使得对模型通道信息更加的关注, 其中 mAP 的值增加了 1.5%, 查准率也增加了 1.5%; 将原来的 SPPF 金字塔模块替换为 SPPCSPC 后, 给模型增加了一些计算量并提升了模型的精度, 其中 mAP 的值增加了 1.3%, 查全率也有很大的提升; 添加了 GSConv_slimneck 模块能够显著发挥注意力模块和空间金字塔的应用效果, 提升检测的精度同时使得模型轻量化, 其中 mAP 的值增加了 3.9%, 查全率也有很大的提升, 增加了 1.9%。因此, 从实验过程中可知改进后的网络在检测中取得明显的效果。

表 5 消融实验结果

算法	查准率/%	查全率/%	$mAP@0.5/\%$	FPS	权重/MB
YOLOv5s	89.5	73.7	77.6	133	14.5
YOLOv5s+SE	91	72.6	79.1	143	14.5
YOLOv5s+SPPCSPC	88.1	76.3	78.9	131	27.3
YOLOv5s+GSConv	88.2	75.6	81.5	178	13.6
YOLOv5s+GSConv+SPPCSPC	89.3	72.3	78.5	125	27.3
YOLOv5s+SPPCSPC+SE	89.8	72.8	78	125	27.4
YOLOv5s+GSConv+SE	88.2	74.3	77.9	243	13.7
YOLOv5s+GSConv+SPPCSPC+SE	93	72.6	79.6	130	27.4

通过消融实验后, 本文将通过测试集来直观的感受改进后的网络的差距, 进行前后对比。如图 14 效果对比图所示, 两幅图分别选取不同时段的结果进行对比。

第 1、2 场景中出现了目标物品的漏检的情况发生, 这是由于场景 1 和场景 2 中的图片比较模糊且漏检的目标比较小, 加入 SE 注意力机制后的网络能加强对物体特征的提取, 使网络更加关注重要的特征。第 3、4 场景中原网络存在漏检、误检的行为。这是因为两个场景的图片均比较模糊, 目标比较小、位置比较偏, 且图片中场景比较复杂, 在加入 SPPCSPC 空间金字塔模块和 SE 注意力机制后, 提高了特征图的分辨率, 加强了对目标的特征识别, 从而提高了目标检测的精度。第 5、6 场景中原网络也出现了漏检的情况。第 5 个场景中由于图片质量不清晰的原因, 导致原网络无法对遗忘物类别钥匙进行识别, 通过 SPPCSPC 空间金字塔模块提升了图片的分辨率, 获取不同尺度的特征信息, 第 6 个场景中漏检的目标比较小、特征不明显, 图片模糊不清楚, 通过使用 GSConv 模块能够缓解深度可分离卷积的缺陷并能充分利用其优势, 在小目标的检测方面有



(a) 原YOLOv5s检测效果 (b) 改进后的YOLOv5s检测效果

图 14 YOLOv5s 网络与改进后的网络效果对比图

很大提升；在加入 SPPCSPC 空间金字塔模块后提高了特征图的分辨率，提高了目标检测的精度，大大减少了漏检、误检的行为发生。

4 结束语

本文针对驾舱环境比较复杂的情形、内容繁杂，继而车内遗忘物进行检测带来干扰，这样会使得原网络的检测精度低，检测效果性能比较差，在网络训练的过程中出现漏检误检。因此本文首先在主干网络的尾部引入 SE 注意力模块，加强模型对通道信息的关注提升目标检测性能，从而提升模型检测精度；其次改进空间金字塔池化模块，将原有的 SPPF 模块改进为 SPPCSPC 模块，通过增加一点计算量来进一步提升检测模型的精度，在神经网络中进行特征图的池化操作，提取不同尺度的特征信息，提高特征图的分辨率，从而提升目标检测的性能；最后同时引入 GSConv 层，能够缓解 DSC（深度可分离卷积）的缺陷，并充分利用 DSC 的优势，在小目标检测方面取得明显的提升效果，既保证了语义信息又平衡了模型的准确性，也提升了检测速度。经过实验结果表明，改进后的网络与原 YOLOv5s 网络相比，其平均精度均值 mAP 提高了 2%，查准率提升了 3.5%。改进后的网络具有良好的提升效果，表明了该方法的有效性。但是，由于天气、光线强度、驾舱环境因素等影响，从而造成目标检测出现漏检误检，影响检测精度和检测性能。因此，本文算法对这些问题有待提

高，后续可以考虑到这些问题进一步完善，提升网络检测效果。

参考文献：

- [1] 杨祥, 王华彬, 董明刚. 改进 YOLOv5 的交通标志检测算法 [J]. 计算机工程与应用, 2023, 59 (13): 194-204.
- [2] 黄尧. 基于改进 YOLOv5s 的道路目标检测算法研究 [D]. 沈阳: 沈阳师范大学, 2023.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. IEEE Computer Society, 2014: 580-587.
- [4] GIRSHICK R. FAST R-CNN [C] // Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago: IEEE, 2015: 1440-1448.
- [5] REN SQ, HE KM, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [6] HE KM, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (2): 386-397.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD [C] // Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision, Amsterdam: Springer, 2016: 21-37.
- [8] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [9] CHEN Q, GEORGANAS N D, PETRIU E M. Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar [J]. IEEE Transactions on Instrumentation and Measurement, 2008, 57 (8): 1562-1571.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in neural information processing systems, 2012, 60: 84-90.
- [11] 张政超. 改进 YOLOv5 的轻量级带钢表面缺陷检测 [J]. 计算机系统应用, 2023, 32 (6): 278-285.
- [12] 倪黎, 邹卫军. 基于 SE 模块改进 Xception 的动物种类识别 [J]. 导航与控制, 2020, 19 (2): 106-111.
- [13] 王素珍, 赵霖, 邵明伟, 等. 基于改进 YOLOv5 的输电线路绝缘子识别方法 [J]. 电子测量技术, 2022, 45 (21): 181-188.
- [14] 范天浩, 顾寄南, 王文波, 等. 基于改进 YOLOv5s 的轻量化金银花识别方法 [J]. 农业工程学报, 2023, 39 (11): 192-200.
- [15] 游越, 伊力哈木·亚尔买买提. 基于改进 YOLOv5 在电力巡检中的目标检测算法研究 [J]. 高压电器, 2023, 59 (2): 89-96.

