

基于改进 YOLOv5s 的机坪特种车辆检测算法研究

诸葛晶昌, 李 想

(中国民航大学 电子信息与自动化学院, 天津 300300)

摘要: 机坪特种车辆作为航班保障服务的重要一环, 其种类多, 形状各异; 目前已有的车辆检测算法在识别机坪特种车辆检测精度不高, 在遮挡时无法检测; 针对于此问题, 提出了一种基于改进 YOLOv5s 的机坪特种车辆检测算法; 为了在机坪特种车辆检测中快速、准确地定位感兴趣区域, 在主干网络中融合协同注意力机制; 考虑到机坪监控场景下特种车辆尺度差别较大的情况, 为了能够增强对不同尺度特种车辆的检测能力, 提出了四尺度特征检测网络结构; 为了提高检测网络多尺度特征融合能力, 结合加权双向特征金字塔结构对网络的 Neck 部分进行改进; 将改进后的算法在自建的机坪特种车辆数据集上进行训练、测试, 实验结果表明, 与 YOLOv5s 相比, 改进后算法的精确度提升了 1.6%, 召回率提升了 3.5%, 平均精度 mAP_{0.5} 和 mAP_{0.5:0.95} 分别有 2.3% 和 3.3% 的提升。

关键词: 机坪特种车辆; 协同注意力机制; 四尺度特征检测; 加权双向特征金字塔; 特征融合

Research on Apron Special Vehicle Detection Algorithm Based on Improved YOLOv5s

ZHUGE Jingchang, LI Xiang

(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Apron special vehicles are taken as an important part of flight guarantee service, which have the features of various types and different shapes. existing vehicle detection algorithms have the low detection accuracy of identifying special vehicles on the apron and cannot detect when obscured. Aiming at this problem, a special vehicle detection algorithm based on improved YOLOv5s is proposed. To locate the region of interest quickly and accurately in the detection of special vehicles on the apron, a coordinate attention mechanism is integrated into the backbone network. Considering that the scale of special vehicles varies greatly in the apron monitoring scene, a four-scale feature detection network structure is proposed to enhance the detection ability of special vehicles with different scales. To improve the multi-scale feature fusion capability of the detection network, the neck part of the network is improved by combining with the weighted bidirectional feature pyramid structure. The improved algorithm is trained and tested on the special vehicle dataset of the self-built apron. The experimental results show that compared with YOLOv5s, the precision of the proposed algorithm is improved by 1.6%, the recall by 3.5%, and the average precision mAP_{0.5} and mAP_{0.5:0.95} by 2.3% and 3.3%, respectively.

Keywords: special vehicles on the apron; coordinate attention mechanism; four-scale feature detection; weighted bidirectional feature pyramid; feature fusion

0 引言

机坪特种车辆检测是从计算机视觉技术出发, 使用目标检测算法实现图像或视频序列中机坪特种车辆的检测。目前, 大多数的航班保障任务依赖人工对特种车辆进行判别、记录。该方式消耗大量的人力且数据的随意性较大, 不利于提升航班保障效率。因此, 针对机坪特种车辆提出一种车辆细粒度分类的算法具有重要意义。

传统算法对于物体的检测通常包括区域选取、特征提取及特征分类这 3 个阶段。传统算法通常使用滑动窗口算法, 在得到待测目标位置后, 一般使用人工精心设计的提取器进行特征提取, 如尺度不变特征变换^[1]和类 Haar 特

征^[2]等。最后, 对提取到的特征进行分类得到最终的检测结果。由于传统算法是人工设计的提取器, 鲁棒性较差^[3], 因此无法达到实际应用的要求。

随着人工智能的发展和计算机算力的提升, 基于深度学习的目标检测算法逐渐成为主流。深度学习的目标检测算法又可以分为基于区域建议的二阶检测算法和基于回归的一阶检测算法。

基于区域建议的二阶检测算法由 Girshick 等人^[4]提出, 之后发展出 Fast R-CNN^[5]目标检测方法, 而 Fast R-CNN 由于候选区域提取的计算量过大, 严重影响了检测的速度。Ren 等人^[6]在此基础上进行改进, 提出了 Faster R-CNN 目

收稿日期: 2022-10-11; 修回日期: 2022-11-07。

基金项目: 国家重点研发专项(2018YFB1601200); 中国民航大学中央高校基本科研业务费专项资金(3122019047)。

作者简介: 诸葛晶昌(1981-), 男, 天津人, 博士, 副教授, 硕士研究生导师, 主要从事计算机视觉、光学检测方向的研究。

引用格式: 诸葛晶昌, 李 想. 基于改进 YOLOv5s 的机坪特种车辆检测算法研究[J]. 计算机测量与控制, 2023, 31(6): 27-33, 39.

标检测算法, 相较 Fast R-CNN 算法而言在速度上优势明显。

王林等人^[7]将 Faster R-CNN 应用于车辆检测, 但对于小型目标车辆的检测效果并不理想。Yang 等人^[8]基于 Faster R-CNN 算法采用小区域放大检测的策略, 进行道路车辆检测, 但在实际检测场景下该算法计算量大实时性较差, 且容易受到遮挡的影响。

YOLO 系列算法是具有代表性的一阶目标检测算法。最初的两代 YOLO 算法在检测精度和速度两方面并没有较为亮眼的表现, 直到采用 Darknet 作为主干网络的 YOLOv3^[9]诞生才真正做到了检测精度与检测速度较好的平衡。在 YOLOv3 之后提出的 YOLOv4^[10]以及 YOLOv5^[11]又在此基础上进行了大规模的改进, 成为了现阶段性能突出的目标检测算法。

马睿等人^[12]使用 YOLOv4 算法实现水位的自动识别, 叶树芬等人^[13]针对电力线和杆塔应用场景使用深度可分离卷积技术降低 YOLOv5 的模型计算量, 并改进 NMS 算法提升检测算法性能。在车辆检测研究方面, 王银等人^[14]在 YOLOv4 基础上使用 MobileNetv2 深度可分离卷积模块代替传统卷积, 并将 CBAM 注意力模块融合到特征提取网络中, 解决了传统车辆检测算法检测精度低, 小尺度目标识别效果差的问题。郭宇阳等人^[15]针对路侧交通监控场景和智能交通管控需要, 借鉴 GhostNet 结构, 在 YOLOv4 基础上提出了轻量化车辆检测算法, 在检测速度上有显著优势。

目前大部分的车辆检测算法研究面向交通监控场景,

主要划分汽车、卡车、公交车三类, 针对特定场景下的车辆细粒度分类研究较少。对此, 本文提出了一种基于改进 YOLOv5s 的机坪特种车辆检测方法, 主要有以下三个贡献:

1) 将位置信息融入通道注意力中, 使检测算法将注意力集中于感兴趣区域, 扩大机坪特种车辆特征权重覆盖的范围, 对特种车辆全局特征的把握能力更强。

2) 在三尺度特征检测网络的基础上, 提出了一种四尺度特征检测网络, 对尺度差异较大的机坪特种车辆具有更好的检测效果。

3) 针对多尺度特征融合中各输入对于最终输出贡献不同的问题, 结合四尺度检测网络提出了一种双向加权融合结构。各节点根据输入分辨率的不同学习匹配相应的权重, 使机坪特种车辆多尺度特征融合的结果更合理。

1 改进的 YOLOv5s 网络

YOLOv5s 由 Input 输入、Backbone 主干网络、Neck 层及 Detect 输出四部分构成。本文在 YOLOv5s 中融合 CA 注意力机制 (CA, coordinate attention)^[16], 增强网络对不同特种车辆特征信息的获取能力; 网络 Neck 部分增加第四层检测尺度, 结合加权双向特征金字塔 (BiFPN)^[17]结构对路径聚合网络 PANet^[18] (PANet, path aggregation network) 结构进行改进, 加强网络对目标特征更高层次的融合及网络对不同尺度特种车辆的检测能力, 改进后的网络结构如图 1 所示。

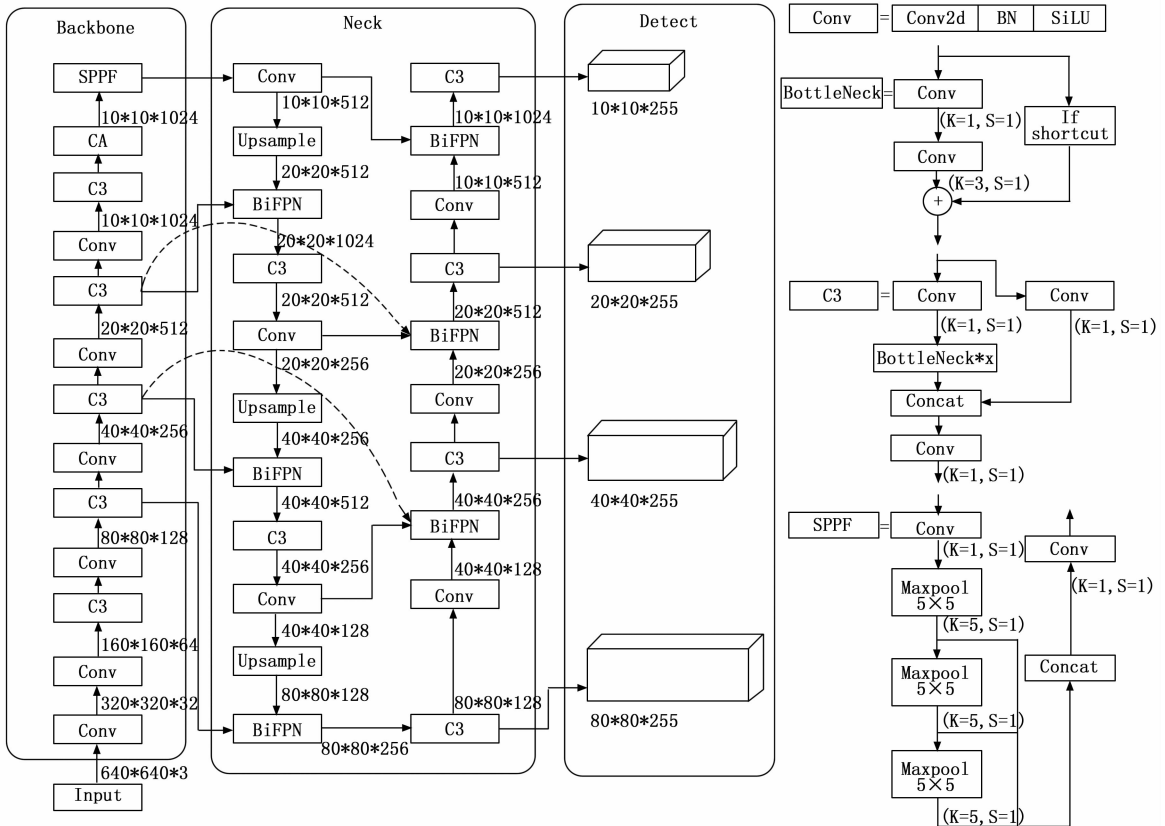


图 1 改进的 YOLOv5s 网络结构

1.1 融合 CA 注意力机制

机坪特种车辆检测相较城市交通场景下汽车、卡车、公交车三类车辆的检测来说, 车辆的种类更多, 难度更大, 需要检测算法对特种车辆特征有更高的分辨力。注意力机制的核心是让卷积神经网络更多的关注图像中重要的部分, 而不是对图像中所有物体都进行关注。

传统的注意力机制包括 SENet^[19]、ECA^[20]、CBAM^[21] 等。CA 注意力机制相较于传统的通道注意力, 将位置信息融入注意力中, 利用获取的位置信息更快的定位感兴趣位置, 在几乎不额外消耗计算资源的同时对于算法性能的提升效果更好。

CA 注意力机制将两个通道注意力分解为两个一维特征编码, 分别沿两个空间方向进行特征聚合。CA 注意力机制结构如图 2 所示。

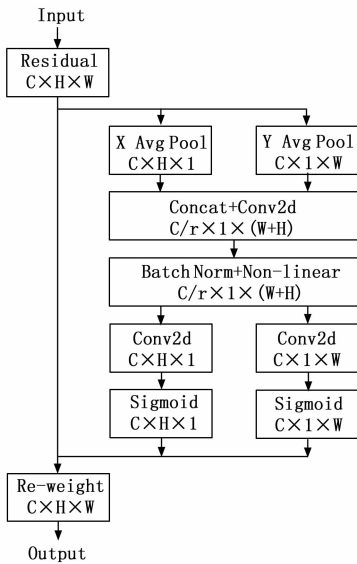


图 2 CA 注意力机制结构

首先, 是 CA 注意力机制的嵌入, CA 注意力机制分别使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的卷积核沿着水平坐标和垂直坐标对每一个通道进行编码, 其中高度为 h 的 c 通道输出表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

同理, 宽度为 w 的 c 通道输出表示为:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

上述两种变换分别沿两个空间方向聚合特征, 得到一对方向感知的特征图。这两种转换允许注意力模块捕捉到沿着一个空间方向的长期依赖关系, 并保存沿着另一个空间方向的精确位置信息, 这有助于网络更准确地定位感兴趣的目标。

通过这两种变换后生成 CA 注意力。在转换过程中, CA 注意力机制先将之前生成的两个特征图进行级联, 用一个 1×1 的卷积进行 F_1 变换, 表示为:

$$f = \delta(F_1([Z^h, Z^w])) \quad (3)$$

其中: 生成的 $f \in \mathbf{R}^{C/r \times (H+W)}$ 是空间信息在水平方向和垂直方向上的中间特征图, r 表示下采样比例。

接着, 沿着空间维度将 f 分为两个单独的张量 $f^h \in \mathbf{R}^{C/r \times H}$ 和 $f^w \in \mathbf{R}^{C/r \times W}$, 经过两个 1×1 卷积 F_h 和 F_w 将两者变换到与输入相同的通道数, 上述过程可以表示为:

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

其中: σ 表示 sigmoid 激活函数。

最后, 对 g^h 和 g^w 进行拓展, 得到注意力权重。CA 注意力机制的最终输出可以表示为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

YOLOv5s 网络各部分对输入特征的提取作用有所不同, Backbone 网络对输入图像进行初始阶段的特征提取, SPPF 模块将提取的初始阶段特征图转化为特征向量输出, Neck 部分对输入提取深层语义特征图, 并将 Backbone 网络提取的浅层语义特征图与深层语义特征图进行融合。因此, 在特征提取的不同阶段融合 CA 注意力机制效果会有一些的差别。

在 YOLOv5s 网络中融合注意力机制的方式通常有 4 种, 如图 3 所示, 分别为: 融入 Backbone 网络的每个 C3 模块后 (a)、单独融入 Backbone 网络的 SPPF 模块前 (b1) 或后 (b2)、融入网络 Neck 的每个 C3 模块后 (c)、将注意力机制融合进网络的 Backbone (d) 或 Neck (e) 部分的每个 C3 模块中。除此以外, 也可以将上述方式进行组合。

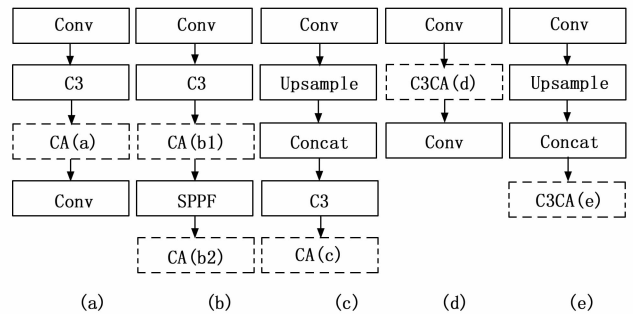


图 3 注意力机制融入网络方式

其中, C3CA 模块是把 CA 注意力机制融入 C3 模块, 具体实现方式如图 4 所示, 给 C3 模块赋予一层注意力。C3CA 模块相较单独融合 CA 注意力机制来说, 不改变网络的总层数, 且参数量更低。

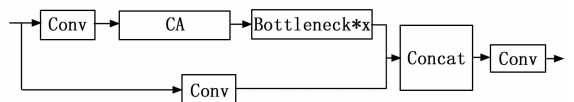


图 4 CA 融入 C3 模块的方式

1.2 提出四尺度特征检测网络

YOLOv5s 网络分别经过 8 倍、16 倍、32 倍下采样输

出 P_3 、 P_4 、 P_5 三个尺度的特征图。三个尺度的特征图分别适用于检测小、中、大三类目标。

考虑到机坪监控视角下特种车辆形态差异较大，以中、中大型目标为主的情况，在原始输出 P_3 、 P_4 、 P_5 三个分别为 80×80 、 40×40 、 20×20 尺度特征图的基础上增加 P_6 检测尺度，输出 10×10 的特征图。如图 5 所示，改进的四尺度的特征检测网络对于大、中、小型目标的检测精度均有一定程度的提升。

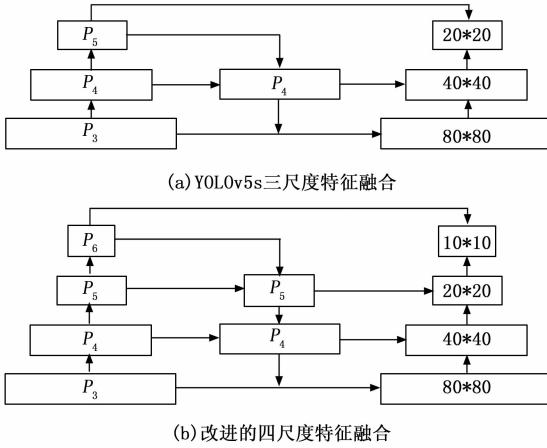


图 5 改进的多尺度特征融合结构

1.3 改进多尺度特征融合结构

YOLOv5s 的 Neck 部分采用的 PANet 结构是在特征金字塔^[22] (FPN, feature pyramid networks) 的基础上增加了一条自底部向上的通道，虽然在一定程度上提升了网络对不同尺度特征融合的能力，但这种融合方式只是简单的相加。由于机坪应用场景对特种车辆检测精度的要求更高，因此，本文针对 PANet 结构进行改进。考虑到不同通道的输入对于最终输出贡献不同的问题，结合加权双向特征金字塔结构，改进 Neck 中通道连接结构，各通道节点采用加权融合的方式，改进的特征金字塔结构如图 6 所示。

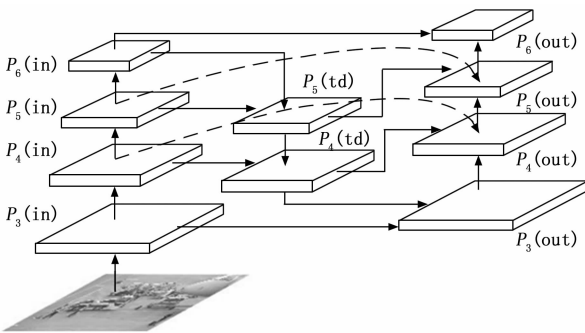


图 6 改进的特征金字塔结构

改进的特征金字塔结构是双向跨尺度连接与快速归一化融合两者的结合，其中快速归一化融合相较于 Softmax 融合来说，精度相近却有更快的速度。

加权双向特征金字塔的快速归一化融合可以表示为：

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} * I_i \quad (7)$$

其中： ω_i 表示特征 I_i 对应的一个可学习权重且 $\omega_i \geq 0$ ，为避免数值不稳定 $\epsilon = 0.0001$ ，其余的权重在归一化后都介于 $0 \sim 1$ 之间。以加权双向特征金字塔输出 P_5 (out) 为例，可以表示为：

$$P_5(td) = Conv \left(\frac{\omega_1 * P_5(in) + \omega_2 * Resize(P_6(in))}{\omega_1 + \omega_2 + \epsilon} \right) \quad (8)$$

$$P_5(out) = Conv \left(\frac{\omega'_1 * P_5(in) + \omega'_2 * P_5(td) + \omega'_3 * Resize(P_4(out))}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon} \right) \quad (9)$$

其中： $P_5(in)$ 表示 P_5 特征层的输入节点， $P_5(td)$ 表示 P_5 特征层的第一个输出节点， $P_5(out)$ 表示 P_5 特征层的第二个输出节点， $Conv$ 表示卷积操作， $Resize$ 表示上采样或下采样操作。

改进后的特征金字塔结构在四尺度特征检测网络的基础上增加了 $P_4(in)$ 到 $P_4(out)$ 及 $P_5(in)$ 到 $P_5(out)$ 两条通道，在不占用更多计算资源的前提下，扩大了多尺度特征融合范围。改进后各节点可以根据输入特征尺度的分辨率不同，学习匹配相应的权重，使网络在多尺度特征融合过程中得到更合理的特征权重。

2 实验结果与分析

2.1 实验准备

鉴于通用数据集无法为机坪特种车辆检测提供数据支持，本文在实验前期构建了机坪特种车辆数据集。针对机场旅客航班保障工作中较为重要且暂无替代方式的车辆，包括飞机牵引车、行李拖车、客梯车、加油车、食品车及摆渡车共六类。数据包含白天、夜晚不同时间段及远近不同视角，示例如图 7 所示。其中，部分图片来自机坪视频数据筛选提取，又从网络上搜集了部分特种车辆图片进行补充，六类车辆共计 5 022 张图片。所有数据均采用 Labelimg 进行标注，其中 4 472 张作为训练集，500 张作为测试集，50 张作为验证集。

使用 k-means 算法对自建的机坪特种车辆数据集进行重新聚类，结果如图 8 所示。聚类得到 12 个锚点框 [29, 52], [87, 54], [57, 106], [80, 103], [73, 197], [112, 133], [140, 176], [93, 272], [229, 155], [198, 266], [353, 355], [553, 433]，将其按先后顺序分为 4 组，分别适用于小、中、中大、大型四类机坪特种车辆的检测。

本文实验均在 Windows10 操作系统下进行，CPU 为 Intel (R) Core (TM) i7-10700@2.90 GHz，内存为 64 G。

GPU 型号为 Nvidia GeForce RTX3080，实验仿真使用 PyTorch 深度学习框架，开发环境为 Python3.7，CUDA 版本为 11.0。

实验中输入图像的分辨率统一设置为 640×640 ，网络



图 7 数据集示例

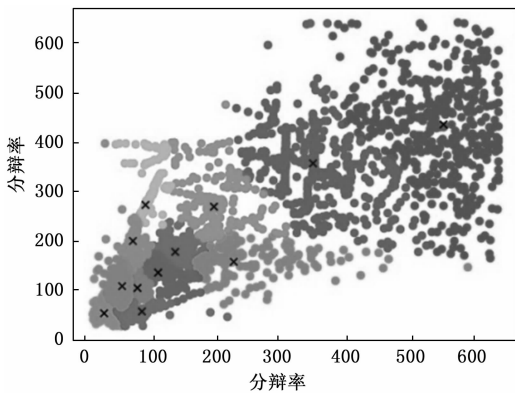


图 8 聚类结果

的初始学习率为 0.01, 动量因子为 0.937, 训练 epochs 为 300。

2.2 评价指标

为了对改进后的算法性能做出评估, 本文选取了精确度 (P, Precision)、召回率 (R, Recall)、平均精度 mAP0.5、mAP0.5: 0.95、参数量 Params 及浮点计算量 GFLOPs 作为评价指标。

其中, mAP0.5 表示 IoU 阈值为 0.5 时所有目标类别的平均检测精度, mAP0.5: 0.95 表示以步长 0.05, 计算 IoU 从 0.5 到 0.95 的 10 个 IoU 阈值下的平均检测精度, 通常情况下 IoU 阈值越高对于算法的回归能力要求更高; Params 表示参数量, 用于计算内存消耗; GFLOPs 表示每秒 10 亿次浮点计算, 是衡量训练复杂程度的重要指标。精确率、召回率以及平均精度的计算式如下所示:

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

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

$$mAP = \frac{\sum AP}{N_c} \quad (13)$$

其中: TP 表示把正类预测为正类, FP 表示把负类预测为正类, FN 表示把正类预测为负类, N_c 表示类别数。

2.3 不同 YOLOv5 模型对比实验

YOLOv5 根据不同网络深度和宽度, 划分出 n、s、m、l、x 五种模型, 模型规模依次增大, 检测精度也有所差异。其中, YOLOv5n 网络深度和宽度分别为 0.33 和 0.25; YOLOv5s 网络深度和宽度分别为 0.33 和 0.50; YOLOv5m 网络深度和宽度分别为 0.67 和 0.75; YOLOv5l 网络深度和宽度分别为 1.0 和 1.0; YOLOv5x 网络深度和宽度分别为 1.33 和 1.25。

本文依据 5 种模型在机坪特种车辆数据集上的实验结果选择最为适合的模型进行改进, 选择的依据主要是模型精度、参数量及计算量的平衡。

网络参数量的增大、模型计算量的提升会在一定程度上降低网络检测速度, 训练所消耗的时间也有所增加。如表 1 所示, YOLOv5n 和 YOLOv5s 是五种模型中相对较小的两种模型, 在网络训练阶段二者所消耗的时间较为接近, 远少于其余三种模型所需时间。YOLOv5x 由于参数量和计算量均最大, 因而训练所需的时间也最长。

表 1 不同深度和宽度的 YOLOv5 模型对比实验

模型	mAP0.5/%	Params/106	GFLOPs/109	大小/M
YOLOv5n	89.4	1.77	4.2	3.8
YOLOv5s	93.0	7.04	15.9	13.8
YOLOv5m	93.1	20.9	48.0	40.3
YOLOv5l	93.6	46.17	108.0	88.6
YOLOv5x	93.7	86.25	204.3	165.0

网络规模最小的 YOLOv5n 在实验中由于网络深度和宽度的限制, 精度与其余 4 种模型有着较为明显的差距, YOLOv5s 在参数量和计算量仅高于 YOLOv5n 的情况下, 精度与网络规模更大的 YOLOv5m 很接近, mAP0.5 仅低 0.1%, 参数量却减少了 66.3%, 计算量减少了 66.9%。在与 YOLOv5 系列更大的模型对比中, YOLOv5s 的 mAP0.5 仅低于 YOLOv5l 模型 0.6%, 而参数量和计算量分别只有 YOLOv5l 的 15.2% 和 14.7%, 与网络模型最大的 YOLOv5x 相比 mAP0.5 也仅低了 0.7%。

结合实验结果, 在考虑算法部署难易程度外, 还应杜绝资源消耗与算法性能提升不匹配的情况。因此, 本文选择网络参数量与计算量较低但精度表现尚佳的 YOLOv5s 作为改进的基础。

2.4 注意力融合位置及不同注意力对比实验

在 YOLOv5s 网络的不同位置融合 CA 注意力机制的效果有较大差异, 本文针对图 3 所述的 YOLOv5s 网络融合注意力机制的几个位置进行实验。

由表 2 的实验结果可以看出, CA 注意力机制并非在所

有位置都能提升网络的检测性能。位置 d, 即对 Backbone 网络的 C3 模块融合 CA 注意力机制的效果最差; 位置 a, 即在 Backbone 网络的每个 C3 模块后单独加一层 CA 注意力机制的效果也较差。

表 2 CA 注意力机制不同位置对比实验

种类	位置	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%
/	/	96.9	87.6	93.0	87.1
CA	a	97.5	85.6	92.4	86.5
CA	b2	97.5	88.2	93.8	88.9
CA	c	98.5	88.6	94.2	89.2
CA	d	97.8	84.0	91.7	85.7
CA	e	96.9	88.4	93.8	88.3
CA	b1+c	97.9	88.5	94.0	89.2
CA	b2+c	97.4	87.9	93.6	88.5
CA	b1+d	97.8	84.1	91.6	83.9
CA	b2+d	98.0	85.0	92.3	84.9
CA	b1+e	97.0	87.8	93.6	88.0
CA	b2+e	98.2	89.2	94.3	88.5
CA	b1	98.0	89.3	94.3	89.6

在 b1、b2、c、e 四个位置添加 CA 注意力机制对原始网络有一定的效果, 在四个有效位置基础上进行组合实验, 注意力机制组合实验结果可以看出 b2 结合 e 的嵌入方式和单独在位置 c 进行注意力机制嵌入的方式效果是最好的。然而, 比较 P、R、mAP 各项指标可见, 上述两种注意力机制组合的提升效果还是不如在 b1 位置单独融合 CA 注意力机制。

综上, 在 YOLOv5s 网络的 b1 位置, 即 SPPF 模块前单独融合 CA 注意力机制的实验效果最好。由此可见, 在初始阶段特征提取结束, 将提取的特征图转化为特征向量输出前, CA 注意力机制能够最大程度上地提升网络感受野, 提升算法对特种车辆特征的检测能力。

为了验证 CA 注意力机制与其他注意力机制相比对原始网络性能提升的优劣, 本文对实验效果最好的 b1 位置, 融合 SE、CBAM、ECA 三种注意力机制进行对比实验, 对比实验结果如表 3 所示。

表 3 不同注意力机制对比实验

种类	位置	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%
SE	b1	97.6	86.4	92.8	86.9
CBAM	b1	98.3	85.8	92.6	87.4
ECA	b1	98.1	88.9	94.2	89.7
CA	b1	98.0	89.3	94.3	89.6

从表 3 中可以看出, 并非几种注意力机制都能够提升网络的检测能力, SE、CBAM 两种注意力机制只能对精确率有轻微的提升作用, 其他几项如召回率和平均精度均低于原始实验数据。

ECA 注意力机制与 CA 注意力机制的提升效果较为接

近, 但考虑到 CA 注意力机制在召回率上的优势更为明显, 因此, 四种注意力机制中效果最优的还是 CA 注意力。

为了评估算法在融合 CA 注意力机制前后对特种车辆特征检测能力的差异, 本文将改进后的算法与原始算法进行特征权重热力图对比。

由图 9 可以看出, 原始网络对于机坪特种车辆的特征把握能力较弱, 特征权重所占比例较低, 而加入 CA 注意力后的算法对于机坪特种车辆全局特征的把握能力更好, 权重覆盖范围也更大。

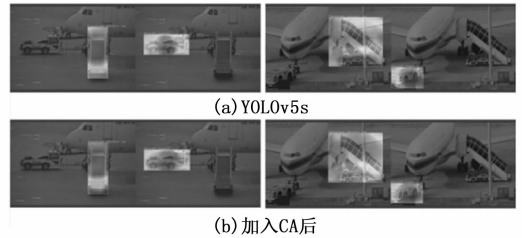


图 9 加入 CA 前后热力图对比

2.5 不同算法对比实验

为了验证本文改进后的算法与其他检测算法性能的优劣, 本文选取经典算法 SSD^[23]、Faster-RCNN、YOLOv3、YOLOv4-tiny^[24]以及 YOLOv4 算法作为对照。此外, 本文分别使用轻量化网络 ShuffleNetv2^[25]、MobileNetv3^[26]替换 YOLOv5 原始主干网络, 得到轻量化的 YOLOv5-ShuffleNetv2、YOLOv5-MobileNetv3 两种检测算法。使用上述 7 种算法和本文算法进行对比, 实验结果如表 4 所示。

表 4 不同目标检测算法对比实验

算法	mAP0.5/%	Params/106	GFLOPs/109
SSD	90.6	26.29	62.8
Faster-RCNN	91.5	137.10	370.4
YOLOv3	93.7	61.55	155.0
YOLOv4-tiny	82.6	6.06	16.4
YOLOv4	90.4	63.96	141.5
YOLOv5-ShuffleNetv2	86.5	3.81	8.0
YOLOv5-MobileNetv3	90.4	3.56	6.3
Ours	95.3	12.58	16.6

从表 4 中可以看出: 本文改进后的算法在参数量和模型复杂程度较低的情况下, 取得了最高的 mAP0.5。与参数量更低, 模型复杂程度更低的 YOLOv5-ShuffleNetv2、YOLOv5-MobileNetv3 及 YOLOv4-tiny 三种算法相比, 本文算法精度优势明显; 与参数量更大, 模型复杂程度更高的 SSD、Faster-RCNN、YOLOv3、YOLOv4 算法相比, 本文算法依旧具有优势。本文改进算法的 mAP0.5 高出对比算法中性能最优的 YOLOv3 算法 1.6%, 而参数量却比 YOLOv3 算法少了 79.6%, 模型复杂度低了 89.3%。

2.6 消融实验

为验证本文改进的 CA 注意力机制、四层特征检测网络

以及双向加权特征金字塔结构的有效性, 进行消融实验, 评估各个部分在相同实验条件下对本文检测算法性能的影响。消融实验以原始的 YOLOv5s 网络实验结果作为基准, 实验数据如表 5 所示。

表 5 消融实验

模型	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%
原始网络	96.9	87.6	93.0	87.1
四尺度	98.1	89.3	94.4	89.2
四尺度+CA	98.7	89.9	94.8	90.1
四尺度+双向加权特征金字塔	98.3	90.4	95	90.3
四尺度+CA+双向加权特征金字塔	98.5	91.1	95.3	90.4

由消融实验结果可以看出: 四尺度特征检测和三尺度特征检测相比, 检测效果的提升是全面的。此外, 无论是改进的双向加权特征金字塔结构还是融合 CA 注意力机制都对算法的 Precision、Recall、mAP 有显著的提升。与原始网络相比, Recall 提升 3.5%, mAP0.5:0.95 提升了 3.3%, mAP0.5 有 2.3% 的提升, Precision 虽有轻微波动也提升了 1.6%。

将改进后的算法与 YOLOv5s 在机坪应用场景下的检测结果进行对比, 由图 10 可以看出, YOLOv5s 算法在遮挡较大的场景下无法有效的对特种车辆进行检测, 而改进后的算法对机坪特种车辆的特征检测能力更强, 在面对部分遮挡时算法的鲁棒性更好。总的来说, 改进后的机坪特种车辆检测算法在精度更高的同时, 稳定性也更为出众。



(a) 本文改进算法

(b) YOLOv5s

图 10 实际检测结果对比

3 结束语

本文提出了一种基于改进 YOLOv5s 的目标检测算法, 旨在提升机坪应用背景下对常见航班保障特种车辆的检测能力。

首先, 考虑 YOLOv5 不同模型大小和性能的差异, 基于实验结果选择 YOLOv5s 作为改进算法。之后, 在 YOLOv5s 网络的不同位置融合 CA 注意力机制, 确定融合效果最佳的位置。为了验证 CA 注意力机制性能的优劣, 在效果最佳的位置上分别融合 SE、CBAM、ECA 三种注

意力机制进行对比实验。接着, 提出了一种四尺度特征检测网络, 在原始三尺度检测网络的基础上增加了 10×10 的输出尺度, 强化了网络对不同尺度特种车辆的检测能力, 并结合加权双向特征金字塔结构进行改进, 改进后的算法对不同尺度特征的融合更为合理。此外, 使用 k-means 算法对机坪特种车辆数据集进行聚类, 用聚类得到的锚框数据替换原始数据。在自建的机坪特种车辆数据集上的实验结果表明: 本文提出的机坪特种车辆检测算法在 Precision、Recall、平均精度 mAP 指标上较原始网络均有显著提升。

最后, 需要指出本文在采用四尺度特征检测网络结构后, 虽带来了更优的检测效果, 但网络参数量和模型整体的复杂程度均有一定增加, 与一些轻量化算法相比对设备性能的要求较高。因此, 下一步将考虑如何将检测网络整体轻量化的同时依然保持较高的检测能力, 便于在更低算力设备上的部署。

参考文献:

- [1] 刘淑萍, 刘羽, 於俊, 等. 结合手指检测和 HOG 特征的分层静态手势识别 [J]. 中国图象图形学报, 2015, 20 (6): 781-788.
- [2] 文学志, 方巍, 郑钰辉. 一种基于类 Haar 特征和改进 Ada-Boost 分类器的车辆识别算法 [J]. 电子学报, 2011, 39 (5): 1121-1126.
- [3] 黄超, 胡志军, 徐勇, 等. 基于视觉的车辆异常行为检测综述 [J]. 模式识别与人工智能, 2020, 33 (3): 234-248.
- [4] GIRSHICK R, DONAHUE 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.
- [5] GIRSHICK R. Fast R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39 (6): 1137-1149.
- [7] 王林, 张鹤鹤. Faster R-CNN 模型在车辆检测中的应用 [J]. 计算机应用, 2017, 38 (3): 666-670.
- [8] YANG B, ZHANG Y, CAO J, et al. On road vehicle detection using an improved faster RCNN framework with small-size region up-scaling strategy [C] // Pacific-Rim Symposium on Image and Video Technology, Springer, Cham, 2017: 241-253.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv: 1804.02767v1, 2018.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint, 2020, arXiv: 2004.10934.

(下转第 39 页)