

基于改进 YOLOv5s 的车载人员 安全带行为检测

焦波¹, 焦良葆^{1,2}, 吴继薇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: 车载人员佩戴安全带行为的检测对于人的生命安全保障具有重要作用; 针对目前车内复杂环境下车载人员佩戴安全带检测精度不高的问题, 提出一种基于改进的 YOLOv5s 车载人员佩戴安全带的检测方法; 该检测方法将 YOLOv5s 作为基础网络, 在此基础上进行改进; 为改善深度模型对特征信息的提取能力, 采用 RFB 模块增大网络的感受野, 并利用 RFB 模块多分支结构获得混合的感受野; 加入 ECA 注意力通道模块, 使得整个网络更加专注特征信息的提取; 将原 YOLOv5s 的损失函数替换为 EIOU, 进一步提高网络对安全带的检测精度; 经过实验结果表明, 改进后网络与原 YOLOv5s 网络相比, 其平均精度均值 mAP 提高了 2.2%, 查准率提升了 5.1%; 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: 安全带; YOLOv5s; 感受野; RFB 模块; 注意力机制; 损失函数

Seatbelt Behavior Detection of Vehicle Occupants Based on Improved YOLOv5s

JIAO Bo¹, JIAO Liangbao^{1,2}, WU Jiwei¹, 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: The detection of seatbelt wearing behavior of vehicle-borne personnel plays an important role in ensuring human life safety. Aiming at the low detection accuracy of seatbelt worn by vehicle occupants in complex environments, an improved detection method based on YOLOv5s is proposed. The detection method takes YOLOv5s as the basic network and improves on it. In order to improve the ability of the depth model to extract feature information, the receptive field of the network is expanded by using the receptive field RFB module, and the hybrid receptive field is obtained by using the multi-branch structure of the RFB module. Adding the efficient channel attention (ECA) modules to make the entire network more focused on extracting the feature information. The loss function of the original YOLOv5s is replaced by the EIOU to further improve the detection accuracy of the safety belt. The experimental results show that compared with the original YOLOv5s network, the mean average precision (mAP) of the improved network is increased by 2.2%, and the precision by 5.1%. The improved network has a good enhancement effect, which shows the effectiveness of the method.

Keywords: seatbelt; YOLOv5s; receptive field; RFB module; attention mechanism; loss function

0 引言

近年来, 随着国家综合实力的不断提高, 各种类型的机动车作为主要的交通工具已经慢慢成为人们的必要选择, 但是各类机动车的过度使用也对人们的安全有着极大的隐患^[1]。据中国交通事故工作组调查结果数据显示: 在 542 起车辆正面碰撞事故中, 系安全带可以使车内人员的死亡率降低 71%; 在 487 起侧面碰撞事故中, 正确佩戴安全带可以使车内人员的死亡率降低 67%; 在 114 起车辆翻滚事故中, 正确佩戴安全带可以使车内人员的死亡概率降低

78%^[2]。因此, 正确佩戴安全带对车内人员的人身安全具有重要意义。

现实生活中许多人的安全意识不高, 经常忘记佩戴汽车安全带。为解决这一问题, 经常使用的方案有两种: 一种方式是车内管理人员往往通过人工方式强调要佩戴安全带, 这种方式不仅效率低, 而且实时性差; 另一种安全带警示的主要方式为车内安装安全带报警装置, 确认方式为卡口正确插入则解除连续报警声音, 但是现在很多车内人员在车辆行驶过程中存在着有意躲避传感器提示的行为。

收稿日期: 2023-05-11; 修回日期: 2023-06-15。

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: 焦波(1998-), 男, 硕士研究生。

焦良葆(1972-), 男, 博士, 教授, 硕士生导师。

引用格式: 焦波, 焦良葆, 吴继薇, 等. 基于改进 YOLOv5s 的车载人员安全带行为检测[J]. 计算机测量与控制, 2024, 32(4): 22-28, 53.

随着计算机视觉的快速发展, 相关技术凭借检测效率高、人工成本低等优势有效弥补了传统人工检测和基于传感器检测方法的不足, 通过计算机视觉技术对车载摄像头或者交通监控视频进行驾驶员安全带检测已经成为必然的研究趋势^[3]。而近年来基于深度学习的汽车驾驶员安全带检测算法能够实现更高的检测精度, 能够提取到驾驶员安全带更深层次的特征, 这在一定程度上解决了传统目标检测精度不高和鲁棒性差的问题^[4]。基于深度学习的目标检测算法在步骤上包括双阶段 (Two-stage) 和单阶段 (One-stage) 两种目标检测框架^[5]。双阶段目标检测是在处理过程中先确定出样本的候选框, 接着对样本通过卷积神经网络 CNN 分类, 其主要以 R-CNN、Fast R-CNN、Faster R-CNN 网络为代表; 而单阶段目标检测是在处理时不产生候选框, 直接在一定回归分析基础上实现目标检测, 其主要以 YOLO 和 SSD (Single Shot Multibox Detector) 等基于回归的目标检测框架。

目前, YOLO 模型是当前安全带检测研究的主流方向。基于单阶段检测的车载人员安全带佩戴检测方法在检测速度上更具实时性, 但易受复杂背景和遮挡的干扰, 造成检测精度下降。殷远齐等人^[6]将 CBAM^[7]注意力机制的方法引入到 YOLOv4 算法中, 大大提升了目标车辆检测的准确率。牛洪超等人^[8]提出一种基于改进 YOLOv5 算法的多层特征融合算法 M-YOLO 用于自动驾驶下的目标识别。田坤等人^[9]采用深度学习多层次网络训练和模型级联的方式, 基于 YOLO v3 快速定位汽车主驾驶区域, 使用极限学习机作为分类器, 提高了驾驶员安全带区域分类的准确性和系统的鲁棒性。Hosseini 等人^[10]使用 YOLOv5s 模型将 Res-Net34 模型与空间金字塔池 (SPP)、时间金字塔池 (TPP) 和幂均值变换 (PMT) 层相组合的方式, 检测汽车乘员和驾驶员的安全带违规行为。

为了有效提高对于车载人员佩戴安全带行为的检测精度以及检测的准确率, 本文以 YOLOv5s 6.0 版本的 YOLOv5s

(You Only Look Once v5s) 进行改进。通过在原骨干网络中嵌入 ECA 模块, 该模块能在众多复杂环境中选择出当前更加关注的信息, 提升对于输入特征图的通道特征的加强; 由于车载内多物件的特殊环境, 为更好对安全带进一步的特征提取是非常有必要的, 于是引入可以增强感受野的 RFB 模块, 进一步提升网络的特征融合能力; 更换损失函数为 EIoU, 使网络处理预测框与真实框更加完善, 提高检测目标框的准确性。

1 YOLOv5 网络介绍

YOLOv5 是一种单阶段目标检测算法, 相比较之前一代, 它汲取了许多优秀网络结构的优点, 精度更高速度也更快, 已经能够做到实时性检测目标^[11]。根据网络的不同深度和宽度、它可以分成 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 共 4 个版本。在这些模型中, YOLOv5s 的网络结构相较于其他 3 个版本更为简洁, 计算资源消耗最少, 同时也更容易移植到其他平台。YOLOv5s 模型主要包括图片输入端、Backbone 网络、Neck 网络及输出层等 4 部分。YOLOv5s 6.0 版本网络结构如图 1 所示。

YOLOv5 的输入端采用了与 YOLOv4 相同的 Mosaic 数据强化方式, 该方式将 4 张不同的图片通过自由排布、缩放、裁剪进行拼接, 由此丰富图片。这种方式下不仅可以减少 GPU, 而且可以提升该网络的训练速度和识别准确率。其次, 提出一种自适应锚框计算, 基于初始锚框基础上, 计算输出框与预测框之间的差值, 再反馈给网络。对比 YOLOv3、YOLOv4, 都是通过独自运行程序计算初始锚框的。但在 YOLOv5 中, 算法能自己推理不同训练集中的最佳锚框值, 这是因为将初始锚框功能融入代码中。最后提出自适应图片缩放方法, 数据集集中的图片其尺寸大概率不尽相同的, YOLO 算法在训练时需要统一图片规格大小。但在此过程中, 容易造成图片排布不恰当, 产生大量黑边, 就会导致推理过多不必要的信息, 影响运行效率。进而

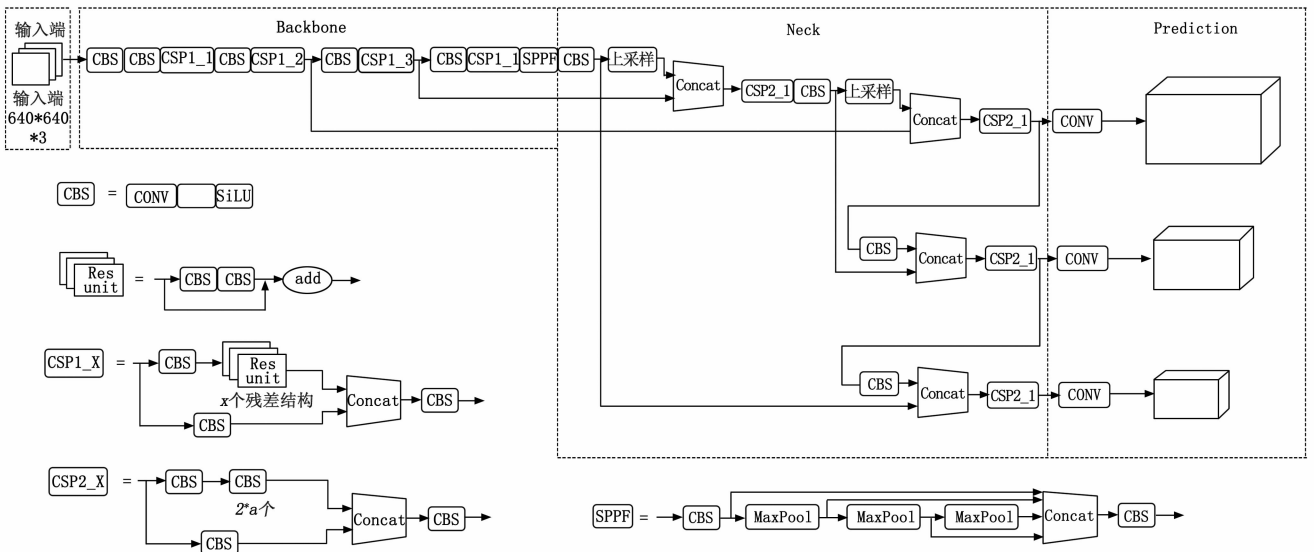


图 1 YOLOv5 6.0 版本网络结构图

YOLOv5 为解决这一问题, 添加了自适应缩放方法^[12]。

Backbone 骨干网络的主要作用就是提取特征, 并不断缩小特征图。Backbone 中的主要结构有 Conv 模块、C3 模块、SPPF 模块。Conv 模块由一个 Conv2d、一个 BatchNorm2d 和激活函数 (SiLU) 构成。CSP 结构也被称为 C3 模块, 其有两种设计结构, 一种是多用于 Backbone 主干网络的 CSP_X1 结构, 特征图进入 CSP_X1 结构后, 将会分成两路, 一条支路先通过 CBS, 再经过 x 个残差结构进行卷积, 另一条支路仅进行 CBS, 然后两个支路进行 concat, 最后再进行一次 CBS。而另一种是 CSP_X2 结构, 它与 CSP_X1 结构唯一不同的是将残差组件换成了 CBS, 且 CSP_X2 多用于 Neck 网络中。YOLOv5s 6.0 中的空间金字塔模块由 5.0 版本的 SPP 改进为 SPPF, SPPF 在输出相同的情况下速度更快。SPPF 结构中, 数据先经过 CBS, 将 3 个 kernel 为 5×5 的 MaxPool 做串行计算, 这一点与 SPP 中将 3 个并行的 MaxPool 区别开来, 再进行 concat 融合, 最后进行 CBS 的特征提取。

Neck 网络中包含 FPN+PAN^[13-14] 结构, 在接受来自 Backbone 网络的下采样测试图像后, FPN 层通过上采样的方式自顶向下传达语义信息, 对整个金字塔结构进行了增强, 同时 PAN 层又通过下采样由下而上上传递目标定位信息, 进行目标的多尺度融合, 提升目标特征的多样性及鲁棒性; Prediction 的输出层主要对真实框以及预测框进行处理, 推理采用 Bounding box 的损失函数, 并用非极大值抑制 (NMS) 对检测的目标框进行筛选, 增强对目标的检测能力。

2 改进策略

2.1 RFB 模块

特征融合对改善神经网络的性能具有重要意义。低层特征图具有较高的分辨率、较丰富的细节, 但也含有较多的噪声^[15]。目标检测更加青睐高层抽象特征去预测目标, 这就使得在本文车载环境中, 由于乘车人员的物品较多, 会极大干扰传统的 YOLO 算法去检测特定目标, 导致在多物品的复杂环境下, 安全带检测受到极大阻碍。而对不同深度的特征层进行多尺度融合, 不仅可以保持底层更丰富的信息, 还可以凸显抽象的特征, 从而有效地提升网络对物体的探测能力。因此, 本文在骨干网络后端添加 RFB 模块扩大感受野, 增强网络的特征表达能力。

2.1.1 感受野原理

其原理是通过将输入图像与特征图上的对应点相关联, 形成的区域的像素通过一层一层的卷积进行计算得出。它的目的在于将最终的特征层与输入图像相关联, 去估计每一层的抽象层次。首先, 感受野使用最终的特征层, 其至少要涵盖输入图片的整个面积。不然, 就会使得图片的部分信息缺失, 这样的结果就会导致用局部信息去预测整体; 然而, 感受野所关注的输入图片像素密集度并不是均匀分布的, 集中于中心区域的图像的像素点越密集, 而远离中心区域图片的像素点逐渐弱化, 甚至消失。文献 [16] 对

最终结果有较大影响的区域被称为有效感受野。在实时图像中识别检测目标所占区域小, 网络中小区域感受野并不能覆盖全部有效面积, 这样就会影响有效感受野的发挥, 对待检测目标特征的能力就会削弱。因此, 增强网络对于目标特征的检测能力, 扩大模型的有效感受野是必要的。

2.1.2 RFB 模块构成

RFB (Receptive Field Block) 是一种受到人类视觉系统中的感受野 (RFs) 结构启发而创造的深度神经网络模块^[17]。它的网络结构中存在一种单向传递的主干神经网络, 该网络结构的感受野与其深度变化成正比。但是, 由于这种神经网络是单向传递的, 所以它是利用多个卷积核串行去获得感受野, 这就会导致得到的感受野较集中。由于共用卷积所导致采样中心也比较固定, 从而影响 CNN 缺乏感受野和特征多样性。于是, 就有人提出采取 Inception^[18] 中的结构。Inception 结构采用通道拼合的方法获得混合的感受野, 它原理是采用多条分支路线, 并在每条支路上设置不同规格的卷积核, 将结果进行拼接。但是, 在这种情况下, 不同的感受野却仍有着较相似的采样中心, 但是混合不同感受野应该有不同的离心率。在人类视觉皮层中, 人群感受野 (PRF, population receptive field) 的大小与视网膜的离心率呈现正相关, 它有助于突出靠近中心区域的重要性, 并提高对小空间变化的不敏感性。根据 Inception 的结构基础, RFB 模块模拟人眼不同方位的视觉感受, 采用带不同离心率的空洞卷积去扩大视野范围。具体来说, 首先在每个支路中使用 1×1 的卷积层构成的瓶颈结构, 以减少特征图中通道数量, 再加上 $n \times n$ 的卷积层。其次将 5×5 卷积层替换为两个叠加的 3×3 卷积层, 以减少参数和更深层次的非线性层。出于同样的原因, 使用 $1 \times n$ 和 $n \times 1$ 的卷积层来代替原来的 $n \times n$ 卷积层。其具体结构如图 2 所示。

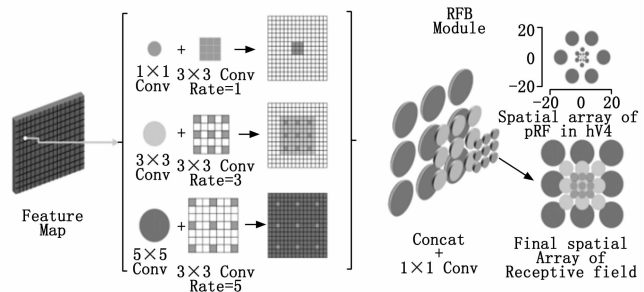


图 2 RFB 模块感受野混合方法示意图

基于 Inception 结构基础, RFB 模块利用其多分支结构进行通道拼合去获得不同大小的感受野, 以得到最佳的特征融合, 这就是 RFB 模块的核心基础。RFB 模块通过设置不同的空洞率去进一步扩大视野面积。如图 3 所示, 当空洞率 $Rate=1$ 时, 空洞卷积就是常规卷积。根据设置不同的空洞率和结合多层卷积叠加效应, 采样区域和感受野迅速扩大, 即为空洞卷积混合感受野。但空洞卷积采样方法决定了其获得的元素是断断续续的, 这样就会使得图像信息

的相关特征缺失, 叫做网格效应。丢失了局部的信息特征, 就会影响上层元素周围信息的获取。

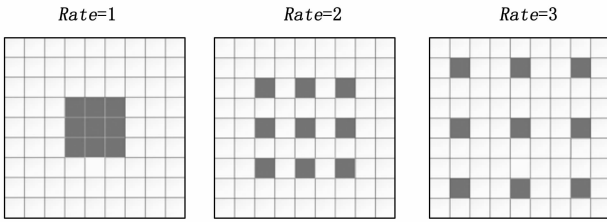


图 3 不同空洞率的空洞卷积

当设置的空洞率不恰当, 或者使用单通道串行卷积核的空洞卷积, 就会使得输入图片中的检测样本信息丢失或没有参与计算, 这就会导致在车内原本复杂的环境, 安全带的检测效果更加受限, 阻碍神经网络的检测。Luo 等人^[16]就提出一种方案, 利用 Inception 的多分支结构, 若能够改变每组卷积前后不同的位置, 使其能够存在依赖关系并且交错并行, 则网格效应的弊端就可以得到解决。为了对抗在空洞卷积下的空洞效应, 其最直接的思想就是在 RFB 模块中设置可以使得采样点交错组合的空洞率, 比如设置为 [1, 2, 5]、[1, 3, 5] 这样的空洞率, 使其不会重复采样, 并最大可能获取图像全部特征信息, 这样就会提高神经网络在复杂环境下的检测能力, 增强网络特征融合。RFB 模块结构如图 4 所示。

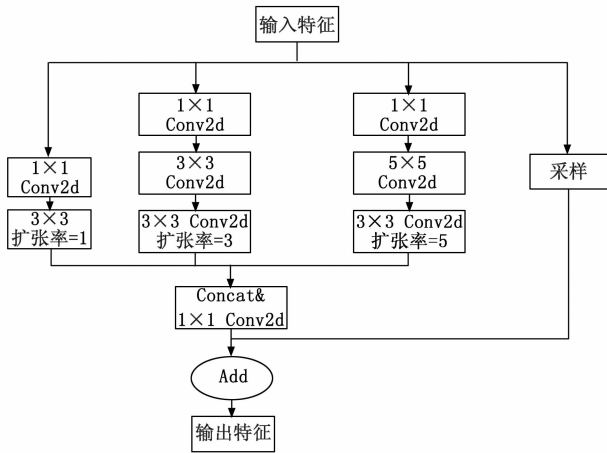


图 4 RFB 模块内部结构图

2.2 添加 ECA 注意力机制

譬如客运汽车内这种小目标、多复杂背景下来检测安全带佩戴情况, 排除背景复杂信息干扰、聚焦有效特征的能力较为重要。ECA (Efficient Channel Attention)^[19] 模块是在 SENet^[20] 的基础上改进而来, 在跨通道信息交互时, 避免了其降低维度, 进一步提升了通道对待检测目标的关注能力。ECA 模块的工作原理如图 5 所示。

在 ECA 模型接受 $1 \times 1 \times C$ 大小的输入特征时, 首先对输入特征进行全局平均池化, 得到所有不需要进行降维的特征; 在此基础上, 采用一维快速卷积算法, 通过对输入

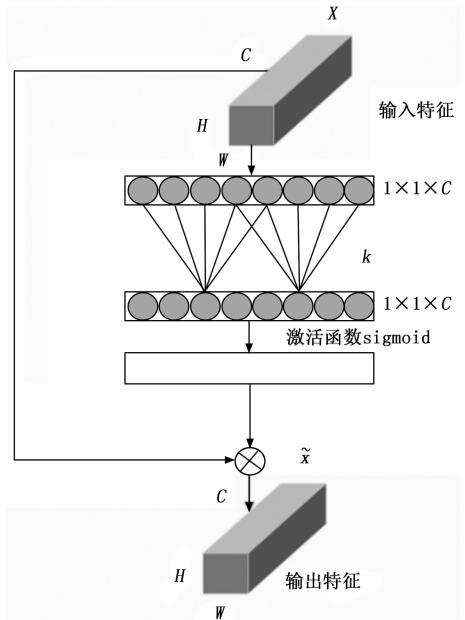


图 5 ECA 工作原理图

通道数目 C 的自适应函数进行计算, 得到区域间的跨通道通信覆盖范围; 然后, 利用 sigmoid 函数计算各个通道中所占的比例, 并将这些输入特征和通道的权重相结合, 最终获得通道注意力的特征。

k 值在不同的网络结构中一般采取自适应计算来选取其参数。 k 和 C 之间的映射关系为:

$$C = \varphi(k) = 2^{(\gamma \times k - b)} \quad (1)$$

式中, k 值代表卷积核大小与 C 值输入通道数成正比, 且通道维数一般为 2 的倍数。

根据式 (1) 中的映射关系可以求得一维卷积核 k 大小的自适应函数, 这样形成的非线性映射关系会使高位通道之间的交流强于低位通道间的交流。其计算过程如式 (2):

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor \text{ odd} \quad (2)$$

式中, k 值代表卷积核大小, C 表示通道数, odd 表示 k 只能取奇数, γ 和 b 设置为 2 和 1, 用于改变 C 和卷积核大小之间的比例。

根据此公式可以看出, ECA 在 SENet 的基础上进行了充分的改进和拓展, 通过添加很少的参数和计算量, 在提高计算速度的同时, 特征提取效果上也获得了不错的增益效果, 有效实现了一种较高效率的跨通道交流方式。由此, 本文将此模块添加至 YOLOv5s 网络中的骨干部分 CSP 层的输出位置, 如图 6 所示。以此增加算法聚焦有效特征的能力, 最终提高对车载人员佩戴安全带行为的检测效果。

2.3 修改损失函数

损失函数被用于评估一个模型的预测值与实际值的偏差。它是一个非负实值函数。损失函数越小, 模型的鲁棒性越好。在深度学习目标检测中, 原 YOLOv5s 损失函数使用的是 CIoU, 它考虑了边界框回归的重叠面积、中心点距

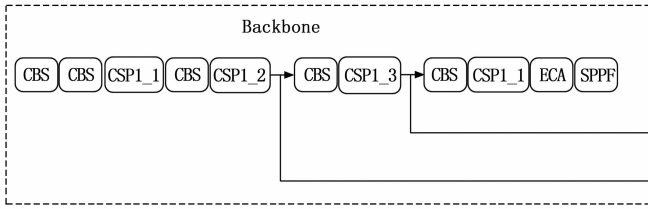


图 6 嵌入 ECA 模块

离、纵横比。其计算方法如式 (3):

$$L_{CI\text{OU}} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha \quad (3)$$

$$\alpha = \frac{\nu}{(1 - IOU) + \nu} \quad (4)$$

$$\nu = \frac{4}{\pi} (\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h})^2 \quad (5)$$

其中: α 是权重函数, ν 用来度量宽高比的一致性 $\rho^2(b, b^{gt})$, c 代表能同时覆盖真实框与预测框重合区域的对角线距离。

但是通过其公式中的 V 反映的纵横比的差异, 而不是宽高分别与置信度的真实差异, 所以有时候会阻碍模型有效地优化相似性。针对这一问题, 2021 年中国科学院大学张等人^[21]在损失函数 CIOU 的基础上将纵横比拆开, 提出了 EIOU Loss, 并且加入 Focal 聚焦优质的锚框。EIOU 是在 CIOU 的基础上将纵横比的影响因子拆开分别计算目标框和锚框的长和宽, 该损失函数包含 3 个部分: 重叠损失、中心距离损失, 宽高损失, 前两部分延续 CIOU 中的方法, 但是宽高损失直接使目标框与锚框的宽度和高度之差最小, 使得收敛速度更快^[22]。其定义的公式如 (6) 所示:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{C_{\omega}^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (6)$$

其中: C_{ω}^2 、 C_h^2 分别为能同时覆盖预测框与真实框的最小闭包区域的高和宽, ω 、 h 分别为实际框的宽和高, ω^{gt} 、 h^{gt} 分别为预测框的宽和高, b, b^{gt} 分别为实际框与预测框的中心点, L_{EIOU} 为预测框与真实框的重叠损失, L_{dis} 为预测框与真实框的中心点距离损失, L_{asp} 为预测框与真实框的宽和高损失。

EIOU 损失的前两部分延续了 CIOU 的方法, 而宽高损失直接使预测框与真实框的宽度和高度之差最小, 使得收敛速度更快。基于此本文将原网络的 CIOU Loss 替换为 EIOU Loss, 使得网络在处理模型的预测值与真实值更加精准, 提高待检测目标框的精确性。

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 1 202 张样图, 均是从 ImageNet 数据集内仔细挑选, ImageNet 数据集最初是拥有超过 100 万张图像的数据集, 是由斯坦福大学的李飞飞教授带领创建。它是目前世界上图像识别最大的数据库, 具

有较高质量的不同情况的车载人员佩戴安全带的数据集。数据集中的图片环境包含了从不同方位对车载人员佩戴安全带进行拍摄、对不同车型内车载人员佩戴安全带处于不同复杂环境程度进行拍摄画面, 以及对车载人员数量不同进行拍摄的画面。尽管图片数量并不是很多, 但图片种类仅有安全带这一类, 在训练车载人员佩戴安全带模型时该数据集更加突出和高效。准备好的数据集通过 Labelme 软件进行标注, 操作界面如图 7 所示。



图 7 Labelme 软件操作界面

标注的标签命名为 seatbelt, 将图片路径保存在 images 文件夹内, 标注完的 xml 文件保存在 labels 目录下, 然后使用代码脚本将 xml 文件转换成 YOLO 可以运行的 txt 文件。最后将 images 内的图片与 txt 文件目录下的数据集对应按照 6 : 2 : 2 的比例划分为训练集、验证集和测试集, 其对应数量分别为 720、242、240 进行划分。

3.2 实验环境的搭建及流程

本实验调用 GPU 在 Pytorch 上进行训练, 训练过程所处环境及硬件配置如表 1 所示。

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

本实验模型设置参数为: 输入图片大小为 640×640 , 训练轮数 (epochs) 为 300 轮, 批次大小 (batch-size) 为 48, 交并比阈值为 (IOU) 为 0.5, 初始学习率设置为 0.01, 其参数均在不添加预训练权重下进行训练。

具体流程如图 8 所示。1) 根据所选取数据集进行标注处理, 并按照比例划分好训练集、测试集和验证集; 2) 采用 Python 编程语言, 利用 Pytorch 深度学习框架构建并改进 YOLOv5s 模型; 3) 选取合适的模型对其进行训练; 4) 根据所添加模块训练好的权重文件在测试集上面进行模型效果测试; 5) 并将最终的模型训练结果与其他多组模型进行横向对比实验。

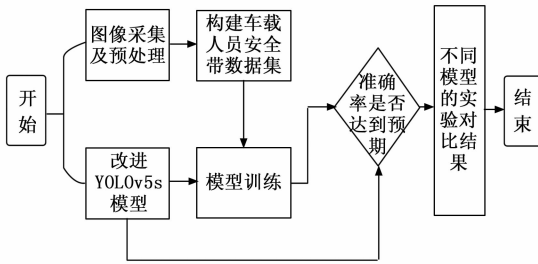


图 8 试验流程图

3.3 评价指标

实验采用的评价指标为: 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) 这 4 个指标作为评价模型的指标, 其中 P 、 R 、 mAP 公式如下:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$mAP = \frac{\sum_{i=1}^K \sum_{r_1=1}^{r_i-1} (r_{i+1} - r_i) \rho_{\text{interp}}(r_i + 1)}{K} \quad (9)$$

其中: TP (True Positives) 表示检测出正确目标数量; FP (False Positives) 表示检测出错误目标数量; FN (False Negatives) 表示没有被检测出来的 TP 。其中查准率 P 表示正确分类的个数与所有已识别样本个数的比值; 而平均精度值 mAP 较 P 、 R 的不稳定波动性, 更能反映出全局性能的指标。当一个模型处于较好性能时, mAP 曲线应尽可能上突。其中, K 表示类别数量, r_1, r_2, \dots, r_n , 是按照升序排列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

3.4.1 对比试验

为了更加直观地展示出 YOLOv5s 算法在其目标检测方面的优越性, 我们选取了若干个不同的网络架构为其进行对比实验, 将其与传统的目标检测算法, 譬如: SSD、YOLOv3、YOLOv5l、YOLOv5m、YOLOv5x 与本网络进行对比。实验结果如表 2 所示, 可以看出, 在众多网络结构中, YOLOv5s 的平均精度最高, 到达了 80.4%, 尽管 YOLOv5n、YOLOv5l 的平均检测精度与原网络较为接近, 但是其查准率均较低。尽管 YOLOv5x 的查准率最高, 但是

表 2 对比实验结果

算法	查准率/%	查全率/%	$mAP@0.5/\%$	权重/MB
SSD	71	78.6	76.3	93.6
YOLOv3	73.1	76.8	74.6	182
YOLOv5l	79.5	78.4	80.1	97.7
YOLOv5m	78.8	78.6	78.8	45.8
YOLOv5x	84.5	76.1	79.6	210.6
YOLOv5s	82.5	75.9	80.4	14.5
YOLOv5n	81.2	73.6	79.1	7.5

其检测精度远远不如 YOLOv5s, 且权重文件远超过本文所用网络。而 SSD 和 YOLOv3 无论是查准率还是平均检测精度均不如 YOLOv5s。综合下面的 4 个参数直观感受, 对于安全带检测算法, YOLOv5s 作为基础网络为最优选择。

3.4.2 注意力模块实验对比

为了证明该模块的有效性, 该实验将 ECA 注意力模块添加到 Backbone 网络中, 嵌入 SPPF 模块前, CSP1_1 后。并且加入 SE、CBAM、CA、SimAM 等注意力机制与其进行对比实验, 实验结果如表 3 所示。使用原 YOLOv5s 网络会得到较快的检测速度, 但是查准率和 mAP 相对于加入的 ECA 模块分别低 5% 和 1.4%。虽然加入 ECA 模块后, 检测速度较原网络有所减慢, 但相较于其它模块的表现, 例如 CBAM 和 CA 模块, 其查准率、查全率、 mAP 均有所提升。所以, 将 ECA 注意力机制添加到此网络中提升其性能最为显著。

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	82.5	75.9	80.4	400	14.5
YOLOv5s+SE	83.1	73.9	80.7	370	14.5
YOLOv5s+CBAM	83.9	72.7	79.9	238	14.5
YOLOv5s+CA	86	71	79.9	370	14.5
YOLOv5s+SimAM	80.9	74.2	81.5	256	14.5
YOLOv5s+ECA	87.5	74.5	81.8	370	14.5

3.4.3 损失函数实验对比

为了表现 EIOU 损失函数在本场景下的优越性, 本文以 YOLOv5s 为网络基础, 将原网络中的损失函数 CIOU 替换为 EIOU, 并引入了 SIoU 以及 Alpha-IoU 两种损失函数与其进行对比试验, 其余部分均不做任何改变, 实验结果如表 4 所示。使用 Alpha-IoU 损失函数会得到最快的检测速度, 但是其 mAP 值要比 EIOU 低 2.6%。对比 SIoU 损失函数, 其 mAP 、查全率和检测速度也有所提升。所以, 将损失函数更换为 EIOU 对该网络的提升效果最为显著。

表 4 融合不同损失函数对比实验

算法	查准率/%	查全率/%	$mAP@0.5/\%$	FPS	权重/MB
YOLOv5s	82.5	75.9	80.4	400	14.5
YOLOv5s+SIoU	88.4	72.7	81.6	263	14.5
YOLOv5s+Alpha-IoU	80.2	71.5	79.4	410	14.5
YOLOv5s+EIOU	88.0	75.5	82.0	357	14.5

3.4.4 消融实验

对比原网络 YOLOv5s, 本文提出了 3 种改进方式: 将 ECA 注意力模块添加到 Backbone 网络中、将原网络中的损失函数替换为 EIOU 损失函数、将 RFB 模块添加到 Backbone 网络中。然后, 将分别对比单独模块实验和组合模块实验下的不同表现做消融实验, 其结果如表 5 所示。实验结果表明, 在加入注意力通道机制 ECA 模块时, 使其更加关注待检测

目标, 其 mAP 值增长了 1.4%, 查准率也有较大的提高; 将原网络损失函数 CIOU 替换为 EIOU, 使得在车载复杂环境下, 其网络能更加精准地判别待检测目标, 其 mAP 值增长了 1.6%, 查准率增长了 5.5%; 且为了应对车内多物品环境的影响, 进一步提高网络的特征融合能力是有必要的, 故单独引入 RFB 模块时, 该网络的 mAP 值提高了 1.3%, 其查准率提高了 8.9%。当将其模块自由组合也会带来不同效果的提升。最终将 3 种模块均添加到该网络中, 其 mAP 值提高了 2.2%, 查准率提升了 5.1%。为此, 通过消融实验, 证明了该网络在检测目标算法上的有效性。

表 5 消融实验结果

算法	查准率 /%	查全率 /%	$mAP@0.5$ /%	FPS	权重 /MB
YOLOv5s	82.5	75.9	80.4	400	14.5
YOLOv5s+ECA	87.5	74.5	81.8	370	14.5
YOLOv5s+EIOU	88.0	75.5	82.0	357	14.5
YOLOv5s+RFB	91.4	70.6	81.7	385	17.5
YOLOv5s+ECA+EIOU	83.4	71.3	81.3	370	14.5
YOLOv5s+ECA+RFB	87.1	73	81.7	358	17.5
YOLOv5s+ECA+EIOU+RFB	87.6	74.4	82.6	358	17.5

在进行消融实验后, 为了能够更加直观地看出修改后网络相较于原网络的提升点, 本文利用测试集进行前后对比。如图 9 所示, 图 9 (a) 和图 9 (b) 显示了 6 个不同场景改进前后的对比效果。



(a) 原YOLOv5s检测效果 (b) 改进后的YOLOv5s检测效果

图 9 YOLOv5s 网络与改进后的网络效果对比图

从图 9 的检测结果可以看出, 第 1、2 场景中的原网络出现了误检的行为。这是由于场景 1 中左侧门框颜色以及部分轮廓外形与安全带较为相似, 场景 2 中左边人的衣物条纹颜色及形状与安全带极为相似。在添加 ECA 注意力机制后的网络修正了这种误检的行为, 且由于修改 EIOU 损失函数, 边界框的置信度也有所提高; 第 3、4 场景中的原网络出现了漏检的行为。这是由于第 3、4 场景中拍摄视角均在左侧, 这就会导致相比于整车环境下, 原本信息量较少的安全带, 在侧视下的有效特征更加微弱, 就会丢失大部分有效信息。在添加 RFB 模块, 利用它混合空洞卷积来扩大有效感受野的能力, 加上对图片不同深度的特征层进行多尺度融合的有效特性, 增强了网络对安全带特征的融合能力, 使得改进后的网络能够很大程度上识别出这一视角下安全带的有效特征, 修正了网络漏检的行为; 第 5、6 场景中的原网络也存在着漏检行为。第 5 场景中, 由于男子衣物与安全带极为相似, 加上部分安全带信息被手部遮挡, 导致原网络无法识别出安全带有效信息, 这与第 6 场景中的漏检内容较为相似, 使用改进后的网络大大减少了这一漏检行为。

4 结束语

本文针对因车载内物品种类繁多, 从而使得图像中捕捉到的内容较为繁杂, 进而对车内安全带的检测造成了不小的干扰, 就会导致原 YOLOv5s 网络的检测精度不高, 且会存在着漏检误检的问题。因此本文在主干网络中加入更加关注待检测目标的 ECA 模块, 提升了模型的精度; 将原网络损失函数 CIOU 替换为 EIOU, 使得在这样一个多物品的复杂环境下, 算法能够更加精准地判别目标检测框的预测值与真实值; 并且在引入 RFB 模块后, 又极大地提升了原网络的感受野, 不同深度的特征层多尺度融合既能保留较丰富的低层信息, 又能突出抽象特征, 从而有效提高了安全带的目标检测能力。实验结果表明, 将 3 种模块均添加到该网络中, 其 mAP 值提高了 2.2%, 查准率提升了 5.1%, 证明了该算法改进的优越性。但是, 由于不同天气下, 光照强弱的影响、车内人员穿着衣物与安全带颜色相近等环境因素的影响, 会造成误检、漏检的结果, 从而影响检测效果。因此, 本文算法针对上述问题还有很大的改进空间, 后续内容可以针对强光射入、暗光环境、相近颜色衣物等环境下对其检测精度进行研究, 进一步完善此网络的检测效果。

参考文献:

- [1] 邹鹏, 杨凯军, 梁晨. 改进 YOLOv5 的轻量级不规范驾驶行为实时检测 [J]. 计算机工程与应用, 2023, 59 (13): 186-193.
- [2] 王甄延. 基于深度学习的汽车安全带检测 [D]. 兰州: 西北师范大学, 2021.
- [3] 李明熹, 林正奎, 曲毅. 计算机视觉下的车辆目标检测算法综述 [J]. 计算机工程与应用, 2019, 55 (24): 20-28.

(下转第 53 页)