基于改进 YOLOv5 的红外车辆检测方法

张学志[1]，赵红东[1,2*]，刘伟娜[1]，赵一鸣[1]，关　松[2]

(1. 河北工业大学 电子信息工程学院，天津 300401；
2. 电磁空间安全全国重点实验室，天津 300308)

摘　要：红外图像可在低照度、恶劣天气等条件下工作，红外车辆检测技术旨在使用红外传感器来监测道路上的车辆，实现对车辆数量、车速等信息的收集与分析，该技术不仅可应用于路面车辆，还可应用于铁路、机场、港口等场景，为交通运输行业的安全和便捷提供了有效的技术支持。然而，由于红外图像成像原理的局限和外部环境的干扰，通常导致红外图像成像质量不理想，红外车辆检测仍然存在许多问题。文中提出了一种改进的 YOLOv5 模型，在 YOLOv5 的主干部分引入了混合注意力机制，使模型能够更好地关注研究者感兴趣的区域，抑制图像噪声的干扰。此外，在 BiFPN 基础上提出了一种改进的 Z-BiFPN 特征融合结构，融合更多的浅层信息，提高浅层信息利用率，并增加一个四分之一下采样的小目标检测层，同时将 YOLOv5 的检测头替换为解耦头来提升模型的检测能力。在自建的七类红外车辆数据集 INFrared-417 上进行了实验，验证了算法的有效可行性。与原始 YOLOv5 相比，mAP 从 81.1% 提升到了 85.3%。

关键词：红外车辆；　目标检测；　注意力机制；　特征融合；　YOLOv5

中图分类号：TP391.4　　**文献标志码：**A　　**DOI：**10.3788/IRLA20230245

0　引　言

红外车辆检测技术是一种基于红外技术的非接触式车辆检测技术，通过红外传感器对路面上的车辆进行实时监测和识别。红外车辆检测具有高效、准确、无人值守等优点，在交通中的作用尤为重要。

YOLO 系列算法是一种基于单阶段 (one-stage) 的目标检测算法，并且在速度和准确率上具有优势。YOLOv3[1] 在前代基础上使用了特征金字塔结构 (Feature Pyramid Network, FPN)，增加了更多的先验框，进一步提升了准确率。YOLOv4[2] 是目标检测领域中先进的方法之一，通过改进网络结构并采用多种优化策略取得了更好的性能表现。而 YOLOv5 具有高精度、快速推理和易部署等特点，被广泛应用于工业界。

基于深度学习的目标检测算法可分为两阶段算法和一阶段算法。两阶段算法通过 Region Proposal Network (RPN) 结构生成一系列候选框，然后对这些候选框进行分类与位置回归，如 R-CNN[3]、Fast R-CNN[4]、Faster R-CNN[5]；而一阶段算法不需要产生候选框，直接生成类别概率和坐标等信息，如 SSD[6]、YOLO 系列[7−8] 等。两阶段算法检测精度较高，但速度较慢，难以完成实时检测任务。一阶段算法单次检测直接得到最终结果，但精度稍低于两阶段检测算法。

通常情况下，红外图像的分辨率比可见光图像低。这主要是由于红外光的波长较长，相同大小的探测器下包含的像素数量减少。同时，由于红外光的波长长、传输距离远，在大气中传输时会衰减，这也导致了红外图像的对比度相对较低。因此，红外小目标的边缘轮廓通常更难以辨别。为此，Li 等人提出了 YOLO-FIRI，将注意力机制加入残差块中，并改进 CSP 结构，改善网络对鲁棒性特征的学习[9]。Zhou 等人设计了自适应特征提取模块，并在 FPN 中引入注意力机制 (Coordinate Attention, CA)，提高了行人红外图像的检测精度[10]。Bai 等人提出了 CBP-Net，设计了交叉连接的双向金字塔与区域特征增强模块，提高

收稿日期：2023−04−23；　**修订日期：**2023−06−01

基金项目：天津市科技计划项目 (21YDTPJC00050)；电磁空间安全全国重点实验室基金项目 (2021JCJQLB055008)

作者简介：张学志，男，硕士生，主要从事计算机视觉、深度学习方面的研究。

导师(通讯作者)简介：赵红东，男，教授，博士，主要从事光电信息处理、图像处理、半导体器件方面的研究。

了模型对红外弱小目标的检测性能[11]。Lv 等人在 FPN 中使用了双注意力机制，并使用门控聚合路径 (Deep Aggregation-Gating Pathway, DAGP) 增强模型对小目标的检测能力[12]。Du 等人在 YOLOv4 基础上进行改进，引入难例挖掘模块 (Hard Example Mining Module)，提高了模型对遮挡车辆的识别能力[13]。Long 等人通过在 YOLOv5 中添加 CBAM 注意力机制，并使用扩张卷积，解决了在复杂背景干扰下红外舰船识别率低的问题[14]。

文中针对红外车辆检测的实际需要，提出了一种改进的 YOLOv5 检测方法。在主干施加混合注意力机制以优化提取到的特征。优化模型 Neck 部分的特征融合方式，从而更高效地利用提取到的特征。使用解耦头的同时，增加了一个小目标检测层，提高模型对小目标车辆的捕获能力。为了提高模型的泛化能力，使用了由笔者自行采集、华南理工大学的 SCUT_FIR_Pedestrian_Dataset[15] 以及东京大学的 MULTISPECTRAL DATASET[16] 三部分构成的红外车辆数据集 INFrared-417 来训练笔者的模型。

1 算法原理

1.1 YOLOv5

YOLOv5 有 Backbone、Neck、Head 三个部分和四个版本 s、m、l、x，选择最轻量的 YOLOv5s 以便于后期部署。对于车辆红外图像分辨率差、对比度低、成像模糊和小目标检测困难等问题，提出了一种改进的 YOLOv5 网络模型，如图 1 所示，其中 SPPF 模块采用的卷积核尺寸为 5×5、9×9、13×13。C3 模块主要用于

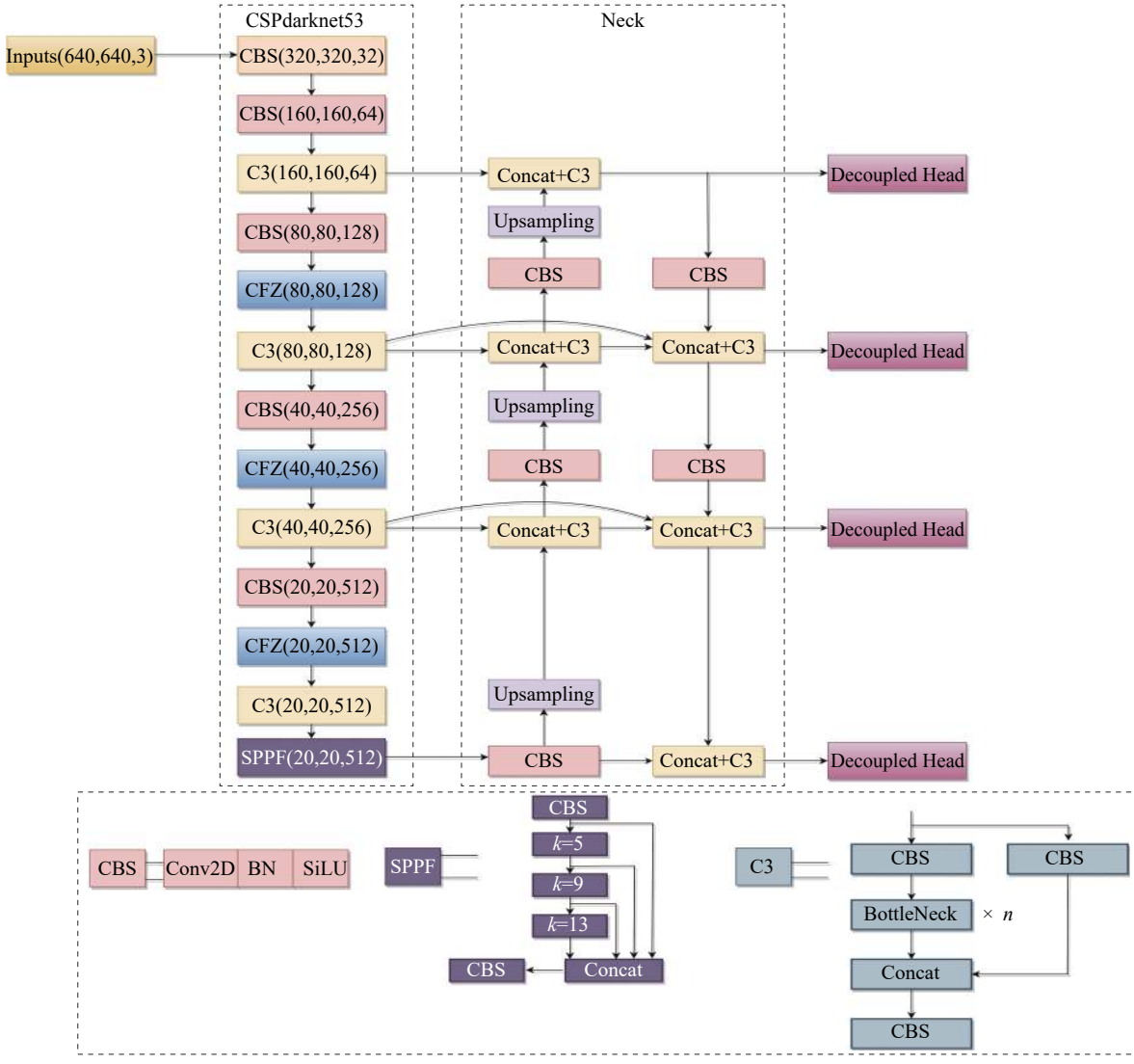


图 1 改进后的 YOLOv5 网络结构图

Fig.1 Improved YOLOv5 network architecture diagram

学习残差特征，压缩模型参数和计算量。

1.2 混合注意力机制

为了抑制车辆红外图像中的噪声，文中在主干网络中引入了通道和空间混合的注意力机制模块 CFG，如图 2 所示。

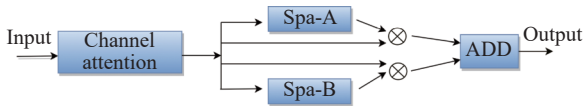


图 2 CFG 结构图

Fig.2 Structure of CFG

Squeeze-and-Excitation (SE) 注意力忽略了位置信息的重要性[17]，Convolutional Block Attention Module (CBAM)[18] 能够自适应地对卷积神经网络中的特征图像素进行加权，增强特征表示，但在捕捉长距离依赖关系方面较差。CA 是发表于 CVPR2021 的一种轻量型注意力机制[19]，通过嵌入位置信息到通道注意力，避免在二维全局池化中位置信息的损失，还可以捕获长距离的依赖关系，故采用其作为笔者的通道注意力机制。CA 结构如图 3 所示，其分为信息嵌入与生成注意力两个步骤。在信息嵌入部分，对于输入 X，在水平坐标方向与垂直坐标方向分别使用尺寸为 (H, 1) 或 (1, W) 的池化核对通道进行编码，对于高度为 h 的第 c 通道的输出可以表示为：

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

对宽度为 w 的第 c 通道的输出可以表示为：

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

通过这两种变换聚合特征，得到一对方向感知的特征图。相较于 SE 产生的单一特征向量，这种转换可以让注意力模块捕捉到沿着一个空间方向的长期依赖关系，并保存沿另一个空间方向的精确位置信息，有利于模型更精确地定位目标。

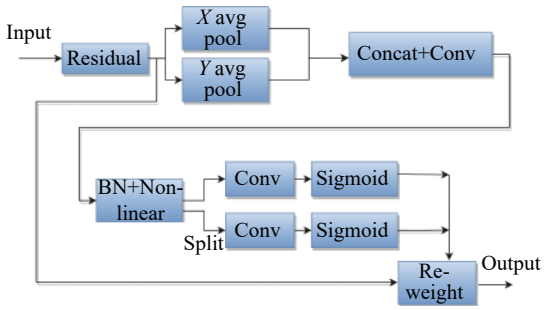


图 3 CA 结构图

Fig.3 Structure of CA

对输入为 $C\times H\times W$ 的特征图，在经过信息嵌入变换后，得到尺寸 $C\times H\times 1$ 和 $C\times 1\times W$ 的特征图，将其进行变换后进行 concat 操作，可表示为：

$$f=\delta\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{3}$$

之后沿着空间维度进行 split 操作，分别利用 1×1 卷积升维，结合 Sigmoid 激活函数得到最后的注意力向量。

空间注意力如图 4 所示。Spa-A 模块将输入特征图 $C\times H\times$W 在 channel 维度上分别经过最大池化与平均池化进行压缩，然后进行 concat 操作得到 $2\times H\times W$ 的特征图，再通过 7×7 的卷积核捕获空间 $W\times H$ 的注意力，最后通过 Sigmoid 激活函数生成 $1\times H\times W$ 的注意力权值。对于输入为 $C\times H\times W$ 的特征图，Spa-B 模块先使用 7×7 卷积核降低通道数，得到 $C/4\times H\times W$ 的特征图，之后再通过一个 7×7 卷积核增加通道数，还原特征图尺寸为 $C\times H\times W$。Spa-A 和 Spa-B 相辅相成，Spa-A 模块通过最大池化与平均池化可以提取较为全面的特征，但两种池化均会造成信息损失。因此，笔者的 Spa-B 模块去掉池化操作以保留较为完整的特征映射。对两个模块的输出进行 Add 操作，得到相应的空间注意力权重。

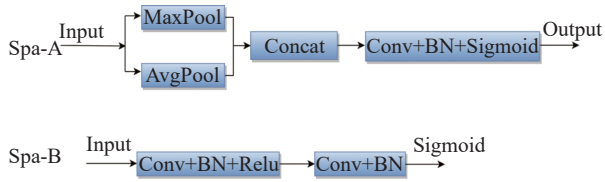


图 4 Spatial Attention 结构图

Fig.4 Structure of Spatial Attention

CFG 模块兼顾了通道信息与空间信息，首先使用通道注意力对各个通道分配权重，然后通过空间注意力对位置信息加权，得到调整后的特征。

1.3 改进的多尺度融合结构

一般的神经网络通常会进行多次下采样操作并在最后一层直接进行预测，这样丢失了部分信息，不利于对小目标的检测。FPN 增加一条自小尺寸特征图向上的路径，在一定程度上缓和了信息丢失问题，如图 5(a) 所示。PANet 在 FPN 基础上增加了一条自大尺寸特征图向下的路径，通过增加分支来增加信息

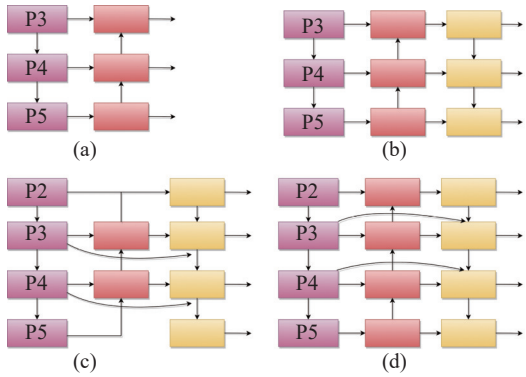


图 5 (a) FPN 结构, 增加了一条自小尺寸特征图向上的路径; (b) PANet 结构, 在 FPN 基础上增加了一条自大尺寸向下的 路径; (c) BiFPN 结构; (d) Z-BiFPN

Fig.5 (a) FPN structure, adds an upward path from small-sized feature map; (b) PANet structure, adds a downward path from large-sized feature map based on FPN; (c) BiFPN structure; (d) Z-BiFPN

流动, 如图 5(b) 所示。BiFPN 使用了高效的双向跨尺度连接, 以在模型参数量不会增加太多的同时, 融合更多特征并减少信息损失, 如图 5(c) 所示。文中提出了一种 Z-BiFPN 结构, 如图 5(d) 所示: 将 P3、P4 特征图进行跨尺度连接, 与 BiFPN 不同, 笔者的跨尺度连接分别在与特征图 P3、P4 同等大小的特征图进行了聚合, 而 BiFPN 将跨尺度连接引向了比 P3、P4 小一倍的特征图。旨在通过此种形式提高浅层信息利用率及网络对小目标车辆的检测能力。

1.4 增强型 YOLO Head

YOLOv5 有三个尺度的耦合检测头, 当输入图片大小为 640 pixel×640 pixel 时, 三个检测头的输入特征图尺寸分别为 80 pixel×80 pixel、40 pixel×40 pixel、20 pixel×20 pixel。在 YOLO 系列版本中, YOLOX[20] 已将传统的 YOLO Head 更换为 Decoupled Head。笔者对 YOLOv5 检测头做出如下改进:

1) 添加一个四分之一下采样的小目标检测层以更好地捕捉像素较小的车辆等目标, 同时有效减少误检和漏检的情况。

2) 将 YOLOv5 的检测头替换为 Decoupled Head。如图 6 所示, 它有三个输出, 分别为 cls, reg 和 obj。cls 用于预测目标框的类别和输出分数, reg 返回目标框的坐标信息进行回归预测, obj 则用于判断目标框中是否有物体。传统检测头分类与回归在同一个头部结构完成, 造成权重共享的问题, 可能会导致模型对其中一个任务表现不足。考虑到分类和回归任务所关注的内容不同, Decoupled Head 将其分离, 采用不同分支完成, 从而可以分别调整两个任务的权重, 提高模型收敛速度及性能。

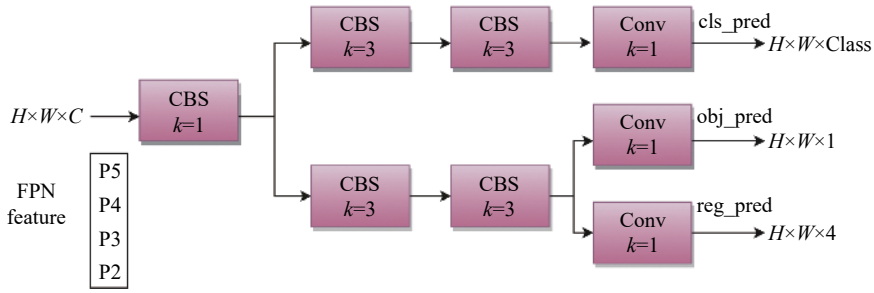


图 6 Decoupled Head 结构示意图

Fig.6 Diagram of Decoupled Head architecture

2 实验结果与分析

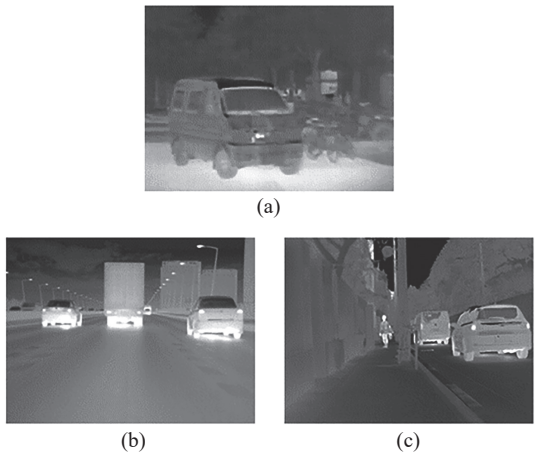
2.1 实验环境配置

实验硬件所用 CPU 为 12th Gen Intel(R) Core(TM) i9-12900 K@3.2 GHz, GPU 为 NVIDIA GeForce RTX-4090, 24 GB RAM。软件环境为 CUDA11.1, torch 版本 1.8.1, python 版本为 3.8, 操作系统为 Ubuntu 22.04.2。实验中 batch size 为 32, 运行 300 个 epoch, 初始学习率为 0.01, 动量为 0.937, 采用余弦退火策略降低学习率, 权重衰减系数为 0.000 5。

2.2 红外数据集 INFrared-417

因现有公开红外车辆数据集较少, 且数据来源单一, 对车辆种类划分不够细致, 不足以达到现代交通要求。文中所使用数据集为 INFrared-417 自行制作数据集, 共包含公共汽车 (bus)、货车 (truck)、轿车 (car)、面包车 (van)、行人 (person)、自行车 (bicycle)、

电动车及摩托车 (elecmot) 七类，其来源分别为自行采集与已有红外数据集，比例约为 1.3∶1，如图 7 所示。对原始图像进行水平翻转后，通过逐一筛选去除部分极端图像，剩余共 2186 张图像。随机选取训练集、验证集、测试集比例约为 8∶1∶1。



(a) (b) (c)

图 7 (a) 自行采集数据集实例；(b) SCUT_FIR_Pedestrian_Dataset 实例；(c) MULTISPECTRAL DATASET 实例

Fig.7 (a) An example of a self-collected dataset; (b) An example of SCUT_FIR_Pedestrian_Dataset; (c) An example of MULTISPECTRAL DATASET

使用 FLIR 公司的 SCOUT TK 来构建数据集，其波长范围为 7.5~13.5 μm，分辨率为 160 pixel×120 pixel，图像采集显示屏输出分辨率为 320 pixel×240 pixel，视场为 20°×16°。拍摄的主要场景为天津市和河北省的城区及郊区夜间，数据的采集方式主要包括图片和视频。

采用的已有数据集源自华南理工大学的 SCUT_FIR_Pedestrian_Dataset[3] 以及东京大学的 MULTISPECTRAL DATASET[4]。SCUT_FIR_Pedestrian_Dataset 由安装在汽车上的单眼 FIR 相机采集，空间分辨率为 384 pixel×288 pixel，焦距 13 mm，视场 28°×21°，波长范围为 8~14 μm，采集场景为广州的市中心、郊区、校园和高速公路。通过逐帧剪辑并去除大量重复图像，最终筛选出了 366 张包含车辆的图像，这些图像的清晰度相对较高，但小型车辆数量较多，部分图片车流密度较大。MULTISPECTRAL DATASET 主要拍摄于东京大学校园，共有四组数据，分别为可见光图像、远红外图像、中红外图像及近红外图像。每组有 7521 帧，由 RGB2、FIR3、MIR4 和 NIR5 相机拍摄而得，这里选择视野范围较大、成像较为完整的远红外图像组。MULTISPECTRAL DATASET 共包含 bike、car、car_stop、color_cone、person 五种类别，经过筛选后选取 85 帧图像加入笔者的数据集。

将以上三组数据混合，统一尺寸为 320 pixel×240 pixel，组成笔者的 INFrared-417 数据集进行标注。不同来源的数据具有目标大小不同、拍摄角度多变、场景多样、清晰度高低不同等特点，使模型能够更全面地学习数据特征，提高模型的泛化能力。

2.3 评估指标

文中采用两个主要指标对模型进行评估，分别为平均精度 (Average Precision, *AP*)、均值平均精度 (mean Average Precision, *mAP*)。同时，还有常见指标如准确率 (Precision, *P*)、召回率 (Recall, *R*) 的表达式分别如公式 (4)、(5) 所示：

$$P=\frac{TP}{TP+FP} \tag{4}$$

$$R=\frac{TP}{TP+FN} \tag{5}$$

式中：*TP*、*TN*、*FP*、*FN* 分别代表真阳性、真阴性、假阳性、假阴性。

AP、*mAP*表达式分别如公式 (6)、(7) 所示：

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{6}$$

$$mAP=\frac{\sum_{i=0}^{n} AP(i)}{c} \tag{7}$$

式中：*AP* 代表以Precision 为纵坐标、Recall 为横坐标构成的曲线下的面积；*mAP* 为所有类别的 *AP* 值的平均值，文中取 *IOU*=0.5 时的 *mAP* 值作为最终指标；*c* 为类别数量。

2.4 消融实验

为了进一步验证所提方法的有效性，设计了五组实验来分析不同的改进方法。采用 *AP*、*mAP* 以及 *P*、*R* 作为实验评估指标，如表 1 所示。

从表 1 可以看出，引入 CFG 混合注意力机制，模型可以更好地关注红外图像中的行人和自行车，从而较大提升了 person 类和 bicycle 类的检测精度，模型的准确率和召回率分别提升了 2.9% 和 1.6%，*mAP* 提

升了 0.5%。接着增加小目标检测层，提高了模型对小目标的检测能力，bus、truck、car、van 类精度均有所提升，通过优化 Neck 部分的特征重聚方式，*mAP* 提升了 0.6%。在此基础上将检测头更换为 Decoupled Head，准确率和召回率分别提升至 88.2% 和 77.4%，*mAP* 达 85.3%，van 类、person 类以及 bicycle 类的 *AP* 显著提升，如图 8 所示。总之，与基线 YOLOv5 相比，笔者的模型实现了更好的检测准确率，如图 9 所示。

表 1 不同改进方法的实验结果

Tab.1 Experimental results of different improvement methods

	a	*b*	*c*	*d*	*e*
YOLOv5s	√	√	√	√	√
CFG		√	√	√	√
Four Head			√	√	√
Z-BiFPN				√	√
Decoupled Head					√
AP-bus	88.5%	85.3%	87.1%	86.0%	89.4%
AP-truck	81.0%	81.2%	82.9%	81.0%	85.4%
AP-car	89.6%	88.7%	89.1%	89.6%	90.3%
AP-van	78.3%	76.4%	77.8%	79.8%	82.6%
AP-person	79.3%	82.6%	81.0%	79.7%	83.5%
AP-bicycle	72.0%	76.6%	75.7%	79.5%	86.2%
AP-elecmot	79.0%	80.1%	80.2%	82.5%	79.7%
P	86.5%	89.4%	85.8%	86.9%	88.2%
R	73.8%	75.4%	74.7%	76.7%	77.4%
mAP	**81.1%**	**81.6%**	**82.0%**	**82.6%**	**85.3%**

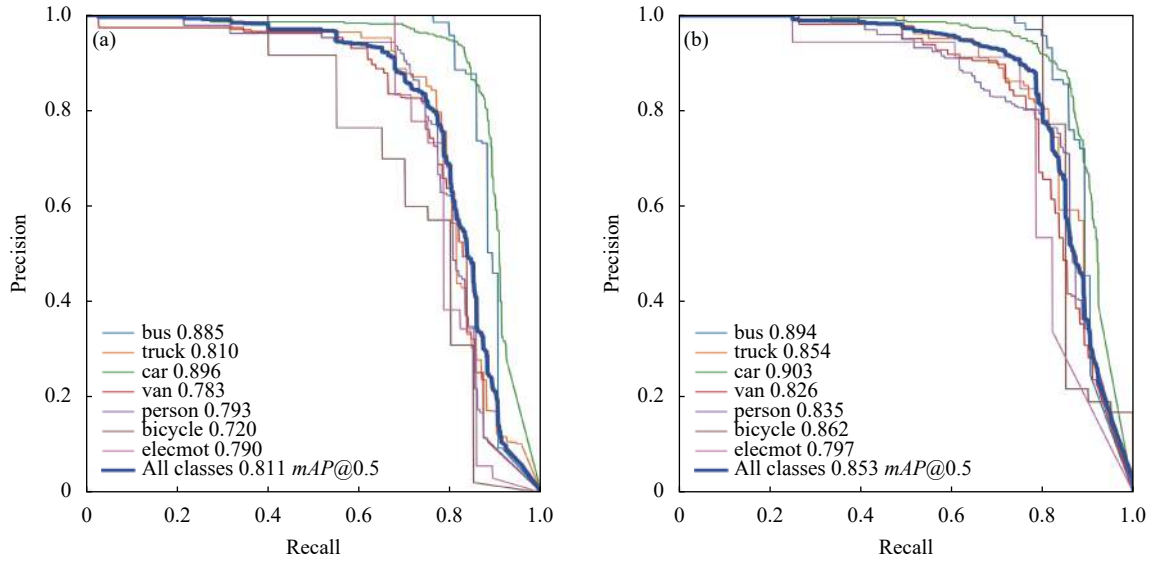


图 8 PR 曲线。(a) YOLOv5; (b) 改进后的 YOLOv5

Fig.8 PR curve. (a) YOLOv5; (b) Improved YOLOv5

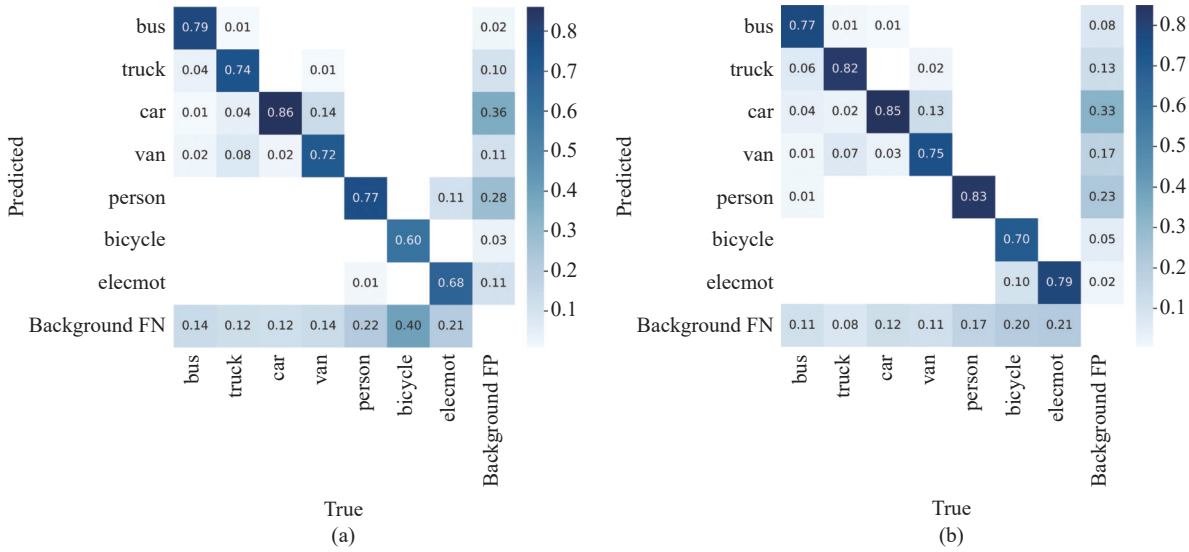


图 9 混淆矩阵。(a) YOLOv5; (b) 改进后的 YOLOv5

Fig.9 Confusion matrices. (a) YOLOv5; (b) Improved YOLOv5

2.5 对比实验

笔者使用当前主流目标检测算法 SSD、YOLOv3、YOLOR、YOLOv7-tiny、YOLOX 来评估模型的性能，在 INFrared-417 数据集上进行了对比实验，结果如表 2 所示，Parameters 为模型总参数量，Weight 为模型生成的权重文件大小。可以看到文中的方法精度远高于其他方法，与基线相比，在参数量增加 6.6 M 的情况下，精度提高了 4.2%。SSD 对大目标有着较好的检测性能，因此对 truck 类检测精度最高，但由于其多次下采样得到的特征图分辨率较低，很难准确检测小目标。YOLOv3、YOLOR-W6 的精度分别比文中的方法低 12.3%、8.7%，且其参数量及模型权重远远高于文中的方法。YOLOv7-tiny、YOLOX的参数量与模型权重虽然比文中的方法小，但精度低了 9.4%、1.8%。总而言之，文中的方法在这些模型中具有最好的检测性能。

表 2 不同目标检测算法对比

Tab.2 Comparison of different object detection algorithms

Models	SSD	YOLOv3	YOLOv5	YOLOR-W6	YOLOv7-tiny	YOLOX	Ours
AP-bus	76.4%	85.9%	88.5%	81.7%	85.7%	87.6%	89.4%
AP-truck	88.0%	83.8%	81.0%	82.3%	82.4%	84.2%	85.4%
AP-car	68.7%	83.3%	89.6%	90.1%	90.8%	90.3%	90.3%
AP-van	63.2%	71.9%	78.3%	80.3%	79.2%	82.1%	82.6%
AP-person	35.8%	70.1%	79.3%	76.9%	75.1%	81.5%	83.5%
AP-bicycle	41.9%	50.2%	72.0%	44.7%	53.0%	78.3%	86.2%
AP-elecmot	47.3%	65.6%	79.0%	80.5%	65.3%	80.7%	79.7%
mAP	60.2%	73.0%	81.1%	76.6%	75.9%	83.5%	85.3%
Parameters	24.4×10^6	61.6×10^6	7.0×10^6	79.3×10^6	6.0×10^6	8.9×10^6	10.4×10^6
Weight/MB	93.7	235.2	13.7	151.8	11.7	17.3	20.3

2.6 结果分析

为了进一步直观显示改进模型的优势，在图 10 中展示了几个场景的检测对比结果。从第一、二行图中可以看出，由于对目标特征学习不充分，原始 YOLOv5 错将一辆汽车识别为面包车。从第三、四行图对比看出，原始 YOLOv5 对小目标学习能力不足，导致漏检了第三行图中的汽车以及左下角汽车里的人，并将第四行图中的路边建筑检测为行人。而文中

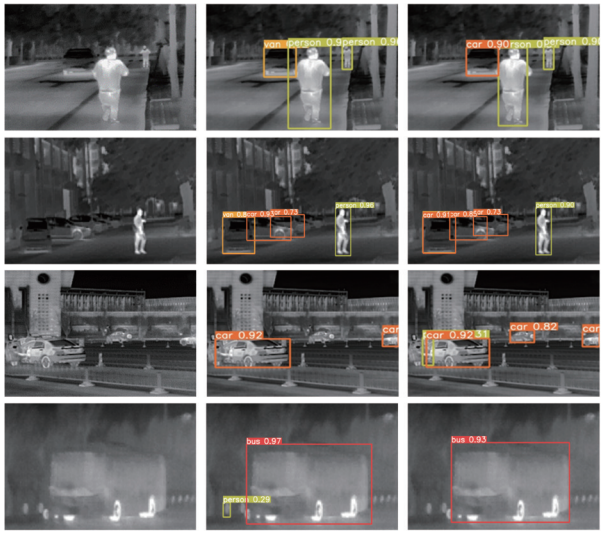


图 10 检测结果对比。(a) 原始图像; (b) YOLOv5; (c) 改进后的 YOLOv5

Fig.10 Comparison of detection results. (a) Original image; (b) YOLOv5; (c) Improved YOLOv5

改进后的模型对目标特征学习的更为科学、充分, 同时对小目标的检测能力有所提高。

3 结 论

在该研究中基于 YOLOv5 提出了一种改进的红外车辆检测模型。通过引入混合注意力来使模型能够更好地关注图像中的车辆区域, 增强模型提取特征的能力。在模型颈部使用改进后的 Z-BiFPN, 充分利用提取到的特征, 并将其进行高效地融合。同时将检测头更换为更先进的 Decoupled Head 以提高检测能力, 并增加一个小目标检测层去捕获小目标。在模型参数量小幅度增加的同时, *mAP* 值从 81.1% 提高至 85.3%, 准确率提高了 1.7%, 召回率提高了 3.6%。后续将会在嵌入式设备上对模型进行部署。

参考文献:

[1] Zhang X X, Zhu X. An efficient and scene-adaptive algorithm for vehicle detection in aerial images using an improved YOLOv3 framework [J]. ***ISPRS International Journal of Geo-information***, 2019, 8(11): 483.

[2] Zhu Q F, Zheng H F, Wang Y B, et al. Study on the evaluation method of sound phase cloud maps based on an improved YOLOv4 algorithm [J]. ***Sensors***, 2020, 20(15): 4314.

[3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580-587.

[4] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV), 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]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2017, 39(6): 1137-1149.

[6] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector [C]//Computer Vision-ECCV 2016, 2016, 9905: 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] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017: 6517-6525.

[9] Li S, Li Y, Li Y, et al. YOLO-FIRI: improved YOLOv5 for infrared image object detection [J]. ***IEEE Access***, 2021, 9: 141861-141875.

[10] Zhou L, Gao S, Wang S, et al. IPD-Net: infrared pedestrian detection network via adaptive feature extraction and coordinate information fusion [J]. ***Sensors***, 2022, 22(22): 8966.

[11] Bai Y, Li R, Gou S, et al. Cross-connected bidirectional pyramid network for infrared small-dim target detection [J]. ***IEEE Geoscience and Remote Sensing Letters***, 2022, 19: 7506405.

[12] Lv G, Dong L, Liang J, et al. Novel asymmetric pyramid aggregation network for infrared dim and small target detection [J]. ***Remote Sensing***, 2022, 14(22): 5643.

[13] Du S, Zhang P, Zhang B, et al. Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4 [J]. ***IEEE Access***, 2021, 9: 25671-25680.

[14] Long Y, Jin D, Wu Z, et al. Accurate identification of infrared ship in island-shore background based on visual attention [C]//2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2022: 800-806.

[15] Xu Z, Zhuang J, Liu Q, et al. Benchmarking a large-scale FIR dataset for on-road pedestrian detection [J]. ***Infrared Physics & Technology***, 2019, 96: 199-208.

[16] Karasawa T, Watanabe K, Ha Q, et al. Multispectral object detection for autonomous vehicles [C]//Proceedings of The Thematic Workshops of ACM Multimedia 2017 (Thematic Workshops' 17), 2017: 35-43.

[17] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks

[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7132-7141.

[18] Woo S, Park J, Lee J-Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018, PT VII, 2018, 11211: 3-19.

[19] Hou Q, Zhou D, Feng J, et al. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, 2021: 13708-13717.

[20] Song J Y, Zhao Y, Song W L, et al. Fisheye image detection of trees using improved YOLOX for tree height estimation [J]. ***Sensors***, 2022, 22(10): 3636.

An infrared vehicle detection method based on improved YOLOv5

Zhang Xuezhi[1], Zhao Hongdong[1,2*], Liu Weina[1], Zhao Yiming[1], Guan Song[2]

(1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China;
2. National Key Laboratory of Electromagnetic Space Security, Tianjin 300308, China)

Abstract:

Objective Infrared image technology is capable of working in low-light and adverse weather conditions. Infrared vehicle detection technology is designed to use infrared sensors to monitor vehicles on roads, enabling the collection and analysis of information related to vehicle quantity and speed, which can be used to achieve traffic management and safety control. This technology can be applied not only to road vehicles, but also to rail transport, airports, and ports, providing effective technical support for the safety and convenience of the transportation industries. However, infrared vehicle detection still faces many challenges due to the low resolution, low contrast, and blurred edges of small targets in infrared images. Traditional hand-crafted image feature extraction methods are not adaptable nor robust, require substantial prior knowledge and have low efficiency. Therefore, this paper aims to explore deep learning-based vehicle detection models, which plays an important role in traffic regulation.

Methods YOLOv5 is a one-stage object detection algorithm that is characterized by its lightweight design, ease of deployment, and high accuracy, making it widely used in industrial applications. In this paper, a CFG mixed attention mechanism (Fig.2) is introduced into the model backbone to help the model better locate the vehicle area in the image and improve its feature extraction ability, due to the low resolution of infrared images. In the feature fusion part, an improved Z-BiFPN structure (Fig.5) is proposed to incorporate more information in the shallow fusion, thereby improving the utilization of shallow information. A small object detection layer is added, and the Decoupled Head (Fig.6) is used to separate classification and regression, improving the model's ability to detect small target vehicles.

Results and Discussions In order to improve the model's generalization ability, an infrared image dataset INFrared-417 (Fig.7) consisting of seven categories of bus, truck, car, van, person, bicycle and elecmot, was constructed by collecting data and combining existing infrared datasets. The main evaluation metrics used were AP (Average Precision) and mAP (mean Average Precision), with P (Precision) and R (Recall) as secondary metrics for the experiments. The ablation experiment results (Tab.1) confirmed the effectiveness and feasibility of the proposed improvement methods, with mAP improving by 4.0%, and AP significantly improving for the van, person, and bicycle categories, while P increased by 1.7% and R increased by 3.6%. In addition, the comparison results (Fig.10) demonstrated that the improved model reduced false alarm and missed detection rates, while improving the detection of small targets. The comparison experiment results (Tab.2) also showed that the

proposed improved model had excellent performance in terms of detection accuracy and model parameter count.

Conclusions This paper proposes an improved infrared vehicle detection algorithm. By introducing the mixed attention mechanism, the model is able to better focus on the vehicle region in the image and enhance its feature extraction ability. The improved Z-BiFPN is used in the model neck to efficiently integrate context information. At the same time, the detection head is replaced with a more advanced Decoupled Head to improve the detection ability, and a small object detection layer is added to improve the ability to capture small targets. It is hoped that this model can be applied in traffic control.

Key words: infrared vehicle; object detection; attention mechanism; feature fusion; YOLOv5

Funding projects: Tianjin Science and Technology Project (21YDTPJC00050); National Key Laboratory of Electromagnetic Space Security Fund Project (2021JCJQLB055008)