

## 基于 SKNet 改进 YOLOv5s 的无人机对道路小目标的检测

周秦汉<sup>1</sup>, 贾 杰<sup>1</sup>, 陈 昊<sup>1</sup>, 张长箭<sup>1</sup>, 吕国云<sup>2</sup>

(1. 南昌航空大学 信息工程学院, 南昌 330063; 2. 西北工业大学 电子信息学院, 西安 710072)

[摘 要] 针对无人机航拍图像中出现的道路小目标检测精度较低的问题, 提出以融合 SKNet 卷积核注意力机制与 YOLOv5s 的目标检测模型, 提高检测模型对地面小目标特征信息提取识别能力。在此改进基础上, 基于 Visual Studio Code 配置的 Pytorch 深度学习开发环境, 对 SKNet + YOLOv5s 的性能进行测试试验。结果表明: 以 VisDrone2019 作为数据集训练时, 相较于几种常规注意力机制的改进方法, 如 SENet + YOLOv5s、CBAM + YOLOv5s, SKNet + YOLOv5s 的检测精度有所提升。

[关键词] SKNet; YOLOv5s; 目标检测; 注意力机制; 深度学习; 性能测试

[中图分类号] TP3

[文献标志码] A

doi: 10.3969/j.issn.2096-8566.2023.04.005

[文章编号] 2096-8566(2023)04-0039-07

## Improving YOLOv5s Based on SKNet for Unmanned Aerial Vehicle Detection of Small Road Targets

ZHOU Qin-han<sup>1</sup>, JIA Jie<sup>1</sup>, CHEN Hao<sup>1</sup>, ZHANG Zhang-jian<sup>1</sup>, LYU Guo-yun<sup>2</sup>

(1. School of Civil and Architectural Engineering, Nanchang Hangkong University, Nanchang 330063, China;

2. School of Electronic Information, Northwestern Polytechnical University, Xi'an 710072, China)

**Abstract:** Aiming at addressing the issue of low detection accuracy in detecting small road targets in drone aerial images, this work proposes a target detection model based on the fusion of SKNet convolutional kernel attention mechanism and YOLOv5s, which has enhanced ability to extract and recognize feature information of small ground targets. Based on this improvement, a Python deep learning development environment configured with Visual Studio Code was used to conduct performance testing experiments on SKNet+YOLOv5s. The experimental results show that when using VisDrone2019 as the dataset for training, compared to several conventional attention mechanisms such as SENet+YOLOv5s and CBAM+YOLOv5s, the detection accuracy of SKNet+YOLOv5s is higher.

**Key words:** SKNet; YOLOv5s; target detection; attention mechanism; deep learning; performance testing experiments

### 引 言

无人机上手简单, 具有灵活便捷的起降和飞行能力, 能够在多种多样的场地环境中替代人工完成

指定任务<sup>[1]</sup>。并且随着软硬件技术的发展其成本逐渐降低, 无人机在军民领域的应用前景也更加广泛。在搭载机载摄像头后, 无人机独特广阔的视角, 能够为工作人员提供地面摄像机难以捕捉到的

[收稿日期] 2023-10-19 [修回日期] 2023-11-15

[基金项目] 江西省引进培养创新创业高层次人才“千人计划”(jxsq2020102038); 江西省重大科技研发专项《通用型应急救援智能空中机器人系统》(20214ABC28W002)

[通讯作者] 贾 杰 (1972—), 男, 博士, 教授。主要研究方向: 飞行器总体, 导航制导与控制。

画面和场景<sup>[2-4]</sup>。而这些拥有极高价值的航拍图像数据,可以进一步用于地面目标检测与追踪。在实际无人机视觉下,航拍图像通常会涵盖大量密集的地面小目标。但尺度差异、环境亮度、遮挡的存在等因素会在不同程度上影响各类别目标的检测精度,导致缺检漏检等问题。

在深度学习理论高速发展的趋势下,基于深度学习的目标检测算法,对于提高机器视觉检测精度有着不可估量的意义。目标检测算法可以分为单阶段算法和双阶段算法。相比于双阶段算法,单阶段算法如 YOLO、SSD 等,不需要生成目标候选区,只需要使用卷积神经网络进行特征提取<sup>[5-6]</sup>,因此具有较高的检测速度和实时性。在 YOLO 基础上,Shen 等<sup>[7]</sup>提出了一种 ASFF 改进 YOLO 检测方法,增强了自适应空间特征融合,抑制了冗余特征。罗茜等<sup>[8]</sup>提出了一种 YOLO 与 Deep-SORT 联合优化检测器的方法,利用时空注意力模块的残差网络作为特征提取网络,加强了微特征感知以及抗干扰能力。

本文对标准 YOLOv5s 检测模型加以改进,在其 Backbone 端融合 SKNet 卷积核注意力机制,加强网络结构对道路小目标的特征感知与提取能力,以 VisDrone2019 作为数据集,对改进后的 YOLOv5s 检测模型进行训练,为小目标检测精度的优化提供了一种改进方向。

## 1 模型构建

### 1.1 YOLOv5s 简介

YOLOv5 是由 Ultralytics 团队提出的一种基于 Deep CNN 的端到端单阶段目标检测算法<sup>[9-10]</sup>,根据不同的深度和特征图宽度可分为 5 种: YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 与 YOLOv5x<sup>[11]</sup>。其中,YOLOv5s 是深度和特征图宽度都相对较小的模型,大小仅为 14.1 MB,且在 5 种模型中拥有较快的检测效率,推理时间相对较短<sup>[12]</sup>,所以综合上述特点,更加轻量化的 YOLOv5s 非常适合应用于移动设备平台。本文即是以 YOLOv5s 作为改进对象应用于无人机航拍图像中

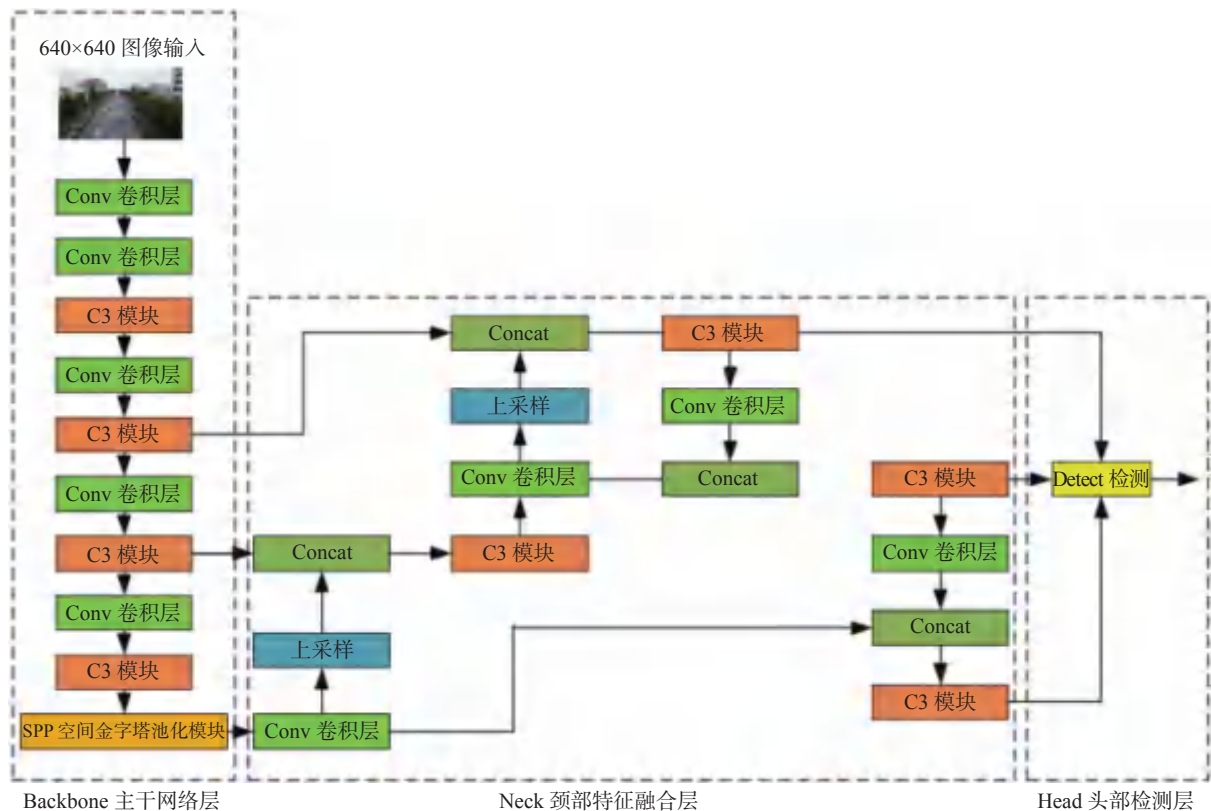


图 1 YOLOv5s 网络结构图

地面小目标的检测。

YOLOv5s 网络主要包含了 3 个结构: Backbone 主干网络层、Neck 颈部特征融合层以及 Head 头部目标检测层(图 1)。3 个结构的功能如下。

1) 主干网络层(Backbone): 该结构包括 C3 模块和 SPP 模块。其中 C3 模块用于提升网络对输入图像的特征信息提取能力。而 SPP 模块则是将输入的特征图像进行空间金字塔池化, 并将生成尺度大小各异特征图像组合形成新的特征向量, 使得特征提取鲁棒性更强。

2) 颈部特征融合层(Neck): 该结构包括 FPN 特征金字塔模块和路径聚合网络 PANet 模块。FPN 模块将语义信息通过上采样由深层传递给底层, PANet 模块则是将定位信息通过下采样由底层传递给深层<sup>[13]</sup>; 加强了信息的优化融合, 进而能够生成具有多尺度信息的特征图。

3) 头部目标检测层(Head): 头部目标检测层采用 3 个卷积层进行检测, 当输入图像大小为  $640 \times 640$  时, 输出  $20 \times 20$ 、 $40 \times 40$ 、 $80 \times 80$  3 个尺度的特征图用于检测不同大小的目标, 最后返回预测框位置坐标、置信度和类别概率值等信息。

## 1.2 融合 SKNet 注意力机制改进 YOLOv5s

SKNet(Selective Kernel Networks)是 Li 等<sup>[14]</sup>提出的一种由多个 SK(Selective Kernel)单元组成的视觉注意力机制。由于 SK 单元可以融合不同大小卷积核适应不同尺度目标得到的特征图, 所以使得 SKNet 网络获得能够自适应调节感受野的能

力<sup>[15]</sup>, 进而使得目标特征信息提取更加精确。SKNet 对图像中不同目标相异特征信息的提取主要包括 3 个步骤(图 2)。

1) 分离(Split): 将输入的特征图像  $X \in \mathbb{R}^{H \times W \times C}$  分离(其中  $\mathbb{R}$ 、 $H$ 、 $W$ 、 $C$  分别为特征图像集、特征图高、宽和通道数), 在分别进行  $3 \times 3$  卷积与  $5 \times 5$  的空洞卷积后, 得到 2 个不同的特征图  $\tilde{U}$  与  $\hat{U}$ 。

2) 融合(Fuse): 将步骤 1) 得到的  $\tilde{U}$  与  $\hat{U}$  融合, 即  $\tilde{U} + \hat{U} = U$ , 进而将  $U$  全局平均池化得到统计各通道特征信息的特征向量  $S$ , 特征向量  $S$  被输入至全连接层进行降维后得到向量  $Z$ 。

3) 选择(Select): 向量  $Z$  经过全连接拓展后得到 2 个向量, 在其通道维度上由 softmax 函数求得两者对应的注意力机制权重向量  $a$  和  $b$ 。将  $a$  和  $b$  分别与  $\tilde{U}$  和  $\hat{U}$  相乘加权, 对输出的特征图结果融合即得到特征信息加强后的向量  $V$ 。

针对无人机航拍图像中存在大量小目标的特点, 在 YOLOv5s 主干网络层(Backbone)中引入了 SKNet 卷积核注意力机制(网络结构如图 3 所示)。在 Backbone、Neck、Head 层等总体组成结构上, 改进 YOLOv5s 与原 YOLOv5s 基本保持一致, 但在 Backbone 层的 C3 模块与 SPP 模块输出端增加了 SK 单元。3 种尺度的特征图像通过 C3 模块与 SPP 模块增强特征融合后, 分别作为 SK 单元的输入, 由 SK 单元进一步加强特征信息的获取, 同时降低背景环境中干扰信息对检测的影响。于是, 主干网络层(Backbone)输出的 3 种不同尺度的目

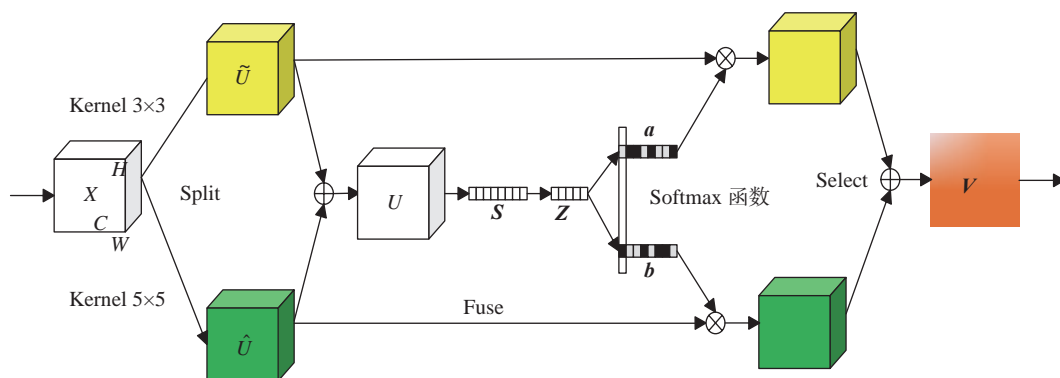


图 2 SKNet 网络结构图

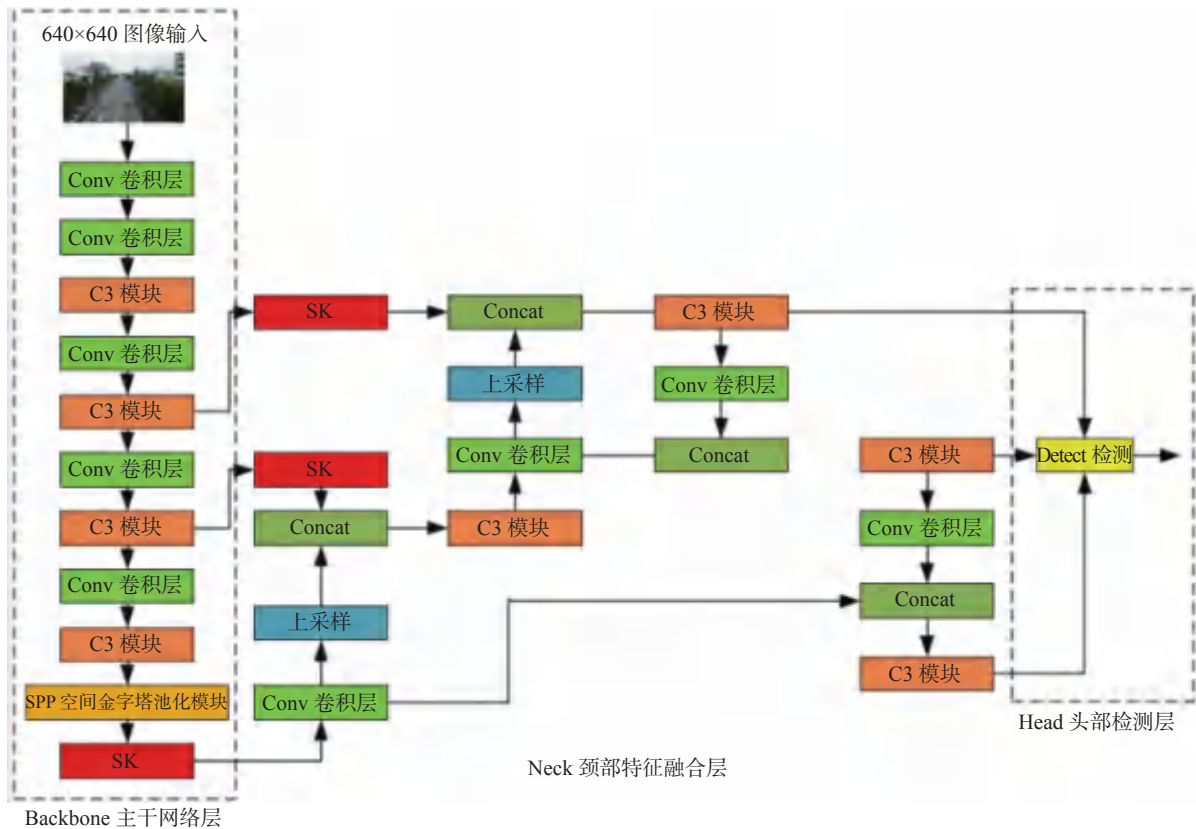


图 3 SKNet + YOLOv5s 网络结构图

标特征信息经过 SK 单元加强特征提取后,送入颈部特征融合层(Neck),接着由头部目标检测层(Head)对 3 种不同尺度大小的特征图进行检测,返回预测框位置坐标、置信度和类别概率值等信息。

## 2 目标检测试验

### 2.1 数据集选择

采用 VisDrone2019 作为基础数据集进行试验。该数据集由天津大学机器学习和数据挖掘实验室 AISKEYEYE 团队采用高清航拍无人机以斜俯视角拍摄得到,分辨率最高达到了  $2000 \times 1500$ 。VisDrone2019 验证集包含了 10 209 张静态照片以及 288 个视频片段,以 10 个检测对象作为数据集标签,并标注了边界框。其中,训练集共有 6471 张航拍照片,验证集共有 548 张航拍图像。

VisDrone2019 数据量能够满足 YOLOv5s 及其改进模型的训练要求。此外,由于 VisDrone2019 在图像尺度、方向上存在多样性,各数据集标签分

布密度不均,特别是在小目标密集情况下容易出现混淆的问题,这对模型检测精度提升带来了诸多挑战。因此,以 VisDrone2019 作为数据集,能够有效反映 SKNet 改进模型与其他改进算法针对地面小目标检测性能的差距。

### 2.2 评估指标

进一步对无人机目标检测模型进行评估,主要采用以下参数作为精度性能指标。

1) 准确率  $P$ (precision rate): 表示预测结果(包括正确预测与错误预测的目标数)中能够正确预测地面道路目标的概率,计算公式为:

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

式中:  $TP$  为正确预测的目标数,  $FP$  为错误预测的目标数。

2) 召回率  $R$ (recall rate): 表示预测结果(包括正确预测与未预测的目标数)中能够正确预测地面道路目标的概率,计算公式为:



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

式中:  $TP$  为正确预测的目标数,  $FN$  为未预测的目标数。

3)  $mAP_{0.5}$ : 表示  $IOU$  阈值为 0.5 时, 所检测的所有道路目标类别的平均  $AP$  值。计算公式为:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (3)$$

其中:  $AP$  在数值上为 P-R 曲线与坐标轴围成的面积。 $IOU$ (intersection over union) 表示预测框与真实框的交并比, 计算公式为:

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

式中:  $A$  为预测框,  $B$  为真实框。

### 2.3 试验环境

使用的试验运行环境如表 1 所示。

### 2.4 模型检测精度性能试验

在标准 YOLOv5s 的基础上, 本文将 SKNet 注意力模块添加到 YOLOv5s 的 Backbone 端。此外, 以增加融合 SENet 注意力模块的 YOLOv5s 作为

表 1 试验运行环境配置

配置	参数或版本
操作系统	Windows10
CPU处理器	Intel(R) Core (TM) i7-9750H CPU @ 2.60 GHz
GPU处理器	NVIDIA GeForce RTX2080
编译环境	Visual Studio Code 2019
CUDA	CUDA 12.2
Torch版本	Pytorch11.0

对照, 对比评价 3 种 YOLOv5s 目标检测模型的性能。

在试验中, 3 种算法模型均以 VisDrone2019 训练集中的静态图像作为训练集数据。其中, 每张训练集数据图像在尺度、方向、检测对象分布、检测对象被遮挡度以及背景光线亮度上各有差异。模型深度与宽度均分别设置为 0.33、0.5, 每次训练输入图片数 batch-size 调整为 16, 初始学习率设为 0.001, 图像数据大小为 64 640, 迭代次数设定为 300。试验效果如图 4~图 7 所示。

由 3 种地面目标检测模型准确率  $P$  随着迭代次数 Epochs 的变化(图 4)可知, 相比于添加 SENet

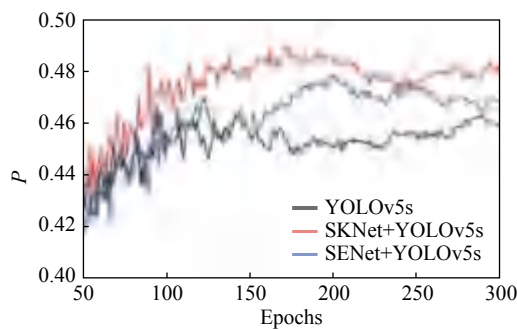


图 4 3 种地面目标检测模型的准确率  $P$  随着 Epochs 的变化以及 Epochs 为 300 时的  $P$  值

训练模型	Epochs 为 300 时 $P / \%$
YOLOv5s	45.89
SKNet+YOLOv5s	48.04
SENet+YOLOv5s	46.79

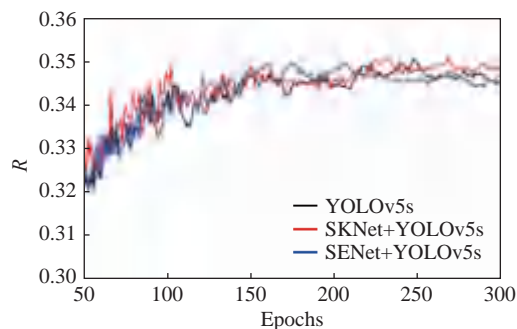


图 5 3 种地面目标检测模型的召回率  $R$  随着 Epochs 的变化以及 Epochs 为 300 时的  $R$  值

训练模型	Epochs 为 300 时 $R / \%$
YOLOv5s	34.54
SKNet+YOLOv5s	34.81
SENet+YOLOv5s	34.69

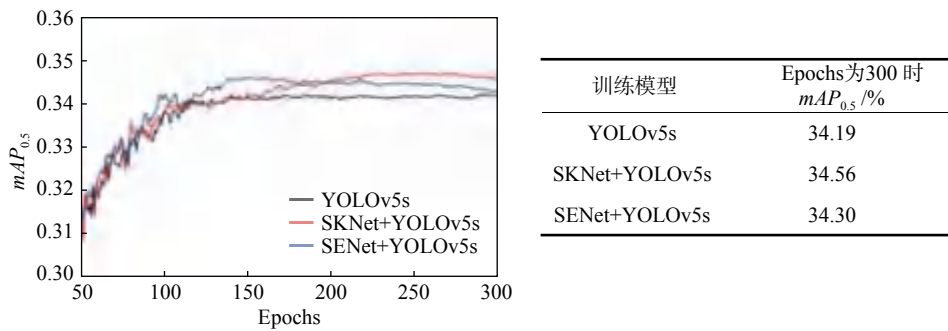


图 6 3 种地面目标检测模型的  $mAP_{0.5}$  随着 Epochs 的变化以及 Epochs 为 300 时的  $mAP_{0.5}$  值

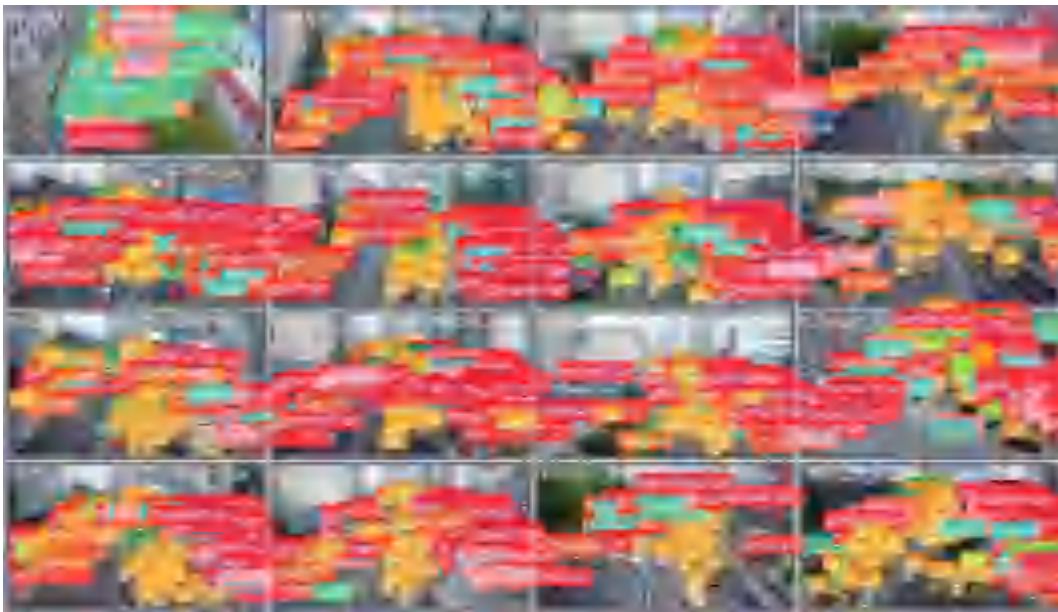


图 7 SKNet + YOLOv5s 改进网络模型对 VisDrone2019 测试集中静态图像的实际检测效果

注意力机制与未添加注意力机制的 YOLOv5s 网络模型,以 VisDrone2019 作为数据集训练得到的融合 SKNet 与 YOLOv5s 的网络模型 SKNet + YOLOv5s 的检测准确率在迭代过程中处于优势,在 Epochs 为 300 时的准确率  $P$  分别提升了 1.25%、2.15%。

由 3 种地面目标检测模型召回率  $R$  随着迭代次数 Epochs 的变化(图 5)可以看出,相比于 SENet + YOLOv5s、YOLOv5s, SKNet + YOLOv5s 的检测召回率在训练大量迭代后均有所提升,Epochs 为 300 时召回率  $R$  分别提升了 0.12%、0.27%。

由 3 种地面目标检测模型  $mAP_{0.5}$  值随着迭代次数 Epochs 的变化(图 6)可以看出,在迭代一定次数后,相较于 SENet + YOLOv5s、YOLOv5s, SKNet + YOLOv5s 的  $mAP_{0.5}$  表现更佳。

表 2 4 种地面目标检测模型对综合目标类别的检测结果

地面小目标检测模型	$mAP_{0.5}/\%$
YOLOv5s	34.19
SKNet + YOLOv5s	34.56
SENet + YOLOv5s	34.30
CBAM + YOLOv5s	34.39

除此之外,本文同样采用 VisDrone2019 数据集,分别对融合 CBAM 注意力机制与融合 SKNet 注意力机制改进的检测模型的  $mAP_{0.5}$  进行比较,再结合前述试验,根据收集的 result.csv 文件得到的结果如表 2 所示。表 2 数据表明: SKNet + YOLOv5s 比 CBAM + YOLOv5s、SENet + YOLOv5s、YOLOv5s 的  $mAP_{0.5}$  分别提高了 0.17%、0.26% 与 0.37%。

特别地, YOLOv5s 与其他 3 种基于 YOLOv5s 改进的检测模型, 对于单一目标(如汽车)的平均检测准确率结果如表 3 所示。表 3 数据表明: 针对汽车标签, SKNet + YOLOv5s 平均检测准确率  $AP_{0.5}$  达到了 74.2%, 相较于 CBAM + YOLOv5s、SENet + YOLOv5s 以及 YOLOv5s, 分别提升了 0.4%、0.7%、1.4%, 在检测无人机航拍图像中单一地面小目标的精度上更加具有优势。

### 3 结 论

本研究提出了基于改进 YOLOv5s 的无人机地面目标检测模型, 通过将 SKNet 注意力机制与 YOLOv5s 的 Backbone 层融合, 进一步提高了 YOLOv5s 针对航拍图像中道路小目标的特征提取

表 3 4 种地面目标检测对单一目标的检测结果

地面小目标检测模型	对汽车检测平均检测准确率 $AP_{0.5}/\%$
YOLOv5s	72.8
SKNet + YOLOv5s	74.2
SENet + YOLOv5s	73.8
CBAM + YOLOv5s	73.5

能力。通过试验结果分析可知, 采用 VisDrone2019 航拍图像作为数据集, 经过训练后的 SKNet + YOLOv5s 改进网络模型对常见地面小目标(如行人和汽车)的检测能力得到提升, 其中对汽车的平均检测准确率达到 74.2%。与 SENet、CBAM 等常用注意力机制相比, SKNet 对 YOLOv5s 检测精度的提升效果更加优异, 可以实现无人机对道路小目标更加精确的检测。

### 【参考文献】

- [1] 韩玉洁,曹杰,刘琨,等. 基于改进 YOLO 的无人机对地多目标检测 [J]. 电子测量技术,2020,43(21):19-24.
- [2] 张杰. 基于深度学习的行人检测与重识别方法及系统实现 [D]. 济南:山东大学,2021.
- [3] HARIT A,VEDANT K,MINU R. Object detection and classification for autonomous drones [J]. Mathematical biosciences and engineering (IJRTE),2020,8(6):3162-3165.
- [4] HSRIT A,VEDANT K,MINU R. Drone-based autonomous motion planning system for outdoor environment under object detection uncertainty [J]. Remote sensing,2021,13(21):4481.
- [5] 赵浩权. 面向无人机图像的小目标检测方法研究 [D]. 长春:吉林大学,2023.
- [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 and machine intelligence,2017,39(6):1137-1149.
- [7] SHEN S,ZHANG X,YAN W,et al. An improved UAV target detection algorithm based on ASFF-YOLOv5s [J]. Mathematical biosciences and engineering:MBE,2023,20(6):10773-10789.
- [8] 罗茜,赵睿,庄慧珊,等. YOLOv5 与 Deep-SORT 联合优化的无人机多目标跟踪算法 [J]. 信号处理,2022,38(12):2629-2635.
- [9] REDMON J, DIVVALA S, GIRSHICK R,et al. You only look once:Unified, real-time object detection [C]. Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016:779-788.
- [10] 张忠民,吴泽. 基于改进 YOLOv5 的密集行人检测方法 [J]. 应用科技,2023,50(1):33-39.
- [11] 蒋亚军,曹昭辉,丁椒平,等. 基于改进 YOLOv5s 模型纸杯缺陷检测方法 [J]. 包装工程,2023,44(11):249-258.
- [12] 邱天衡,王玲,王鹏,等. 基于改进 YOLO 的目标检测算法研究 [J]. 计算机工程与应用,2022,58(13):63-73.
- [13] 张昊,郑广海,张鑫,等. 改进 YOLOv5 框架的血细胞检测算法 [J]. 计算机系统应用,2023,32(5):123-131.
- [14] LI X,WANG W,HU X,et al. Selective Kernel Networks [C]. 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR).Long Beach,CA,USA:IEEE,2019:510-519.
- [15] 赵梦,于红,李海清,等. 融合 SKNet 与 YOLOv5 深度学习的养殖鱼群检测 [J]. 大连海洋大学学报,2022,37(2):313-319.