# 复杂环境下基于多策略YOLOv5s的无人机 小目标检测

罗 茜<sup>1</sup>,陆荣灿<sup>2</sup>,王宇泽<sup>1</sup>,吴晓梅<sup>2</sup>,葛嘉玄<sup>2</sup>,王 芳<sup>1\*</sup>

<sup>1</sup>燕山大学理学院,河北 秦皇岛 <sup>2</sup>燕山大学电气工程学院,河北 秦皇岛

收稿日期: 2025年3月4日; 录用日期: 2025年5月15日; 发布日期: 2025年5月26日

# 摘 要

针对复杂场景下无人机航拍小目标检测时特征提取的主观性和局限性,本文提出了三种改进策略: 1)为了提升无人机对不同尺度目标的检测能力,将多头自注意力机制(MHSA)融入到YOLOv5s骨干网络的最后一层; 2)为了增强特征信息的利用,构建了BiFPN特征融合网络; 3)将SimAM模块集成到YOLOv5s模型中,以提高语义与位置信息的匹配。通过将上述三种改进策略两两组合,构建了三种多策略YOLOv5s检测模型:第一种是多头自注意力机制(MHSA)与BiFPN特征融合网络的结合;第二种是多头自注意力机制(MHSA)与SimAM注意力机制的结合;第三种是SimAM注意力机制与BiFPN特征融合网络的结合。在VisDrone2019数据集上的对比实验结果表明,第二种多策略模型在检测效果上优于其他两种模型,其平均精度(mAP)提升至38.9%,比原模型提高了4.8%。

#### 关键词

无人机航拍,YOLOv5s,多头自注意力机制(MSHA),BiFPN网络,SimAM注意力机制

# Small Target Detection of UAVs in Complex Environments Based on Multi-Strategy YOLOv5s

Qian Luo<sup>1</sup>, Rongcan Lu<sup>2</sup>, Yuze Wang<sup>1</sup>, Xiaomei Wu<sup>2</sup>, Jiaxuan Ge<sup>2</sup>, Fang Wang<sup>1\*</sup>

<sup>1</sup>School of Science, Yanshan University, Qinhuangdao Hebei

<sup>2</sup>School of Electrical Engineering, Yanshan University, Qinhuangdao Hebei

Received: Mar. 4<sup>th</sup>, 2025; accepted: May 15<sup>th</sup>, 2025; published: May 26<sup>th</sup>, 2025

\*通讯作者。

文章引用: 罗茜, 陆荣灿, 王宇泽, 吴晓梅, 葛嘉玄, 王芳. 复杂环境下基于多策略 YOLOv5s 的无人机小目标检测[J]. 人工智能与机器人研究, 2025, 14(3): 590-604. DOI: 10.12677/airr.2025.143058

#### **Abstract**

Aiming at the subjectivity and limitations of feature extraction in small target detection of UAV aerial photography in complex scenarios, this paper proposes three improvement strategies: 1) To enhance the detection capability of UAVs for targets of different scales, the Multi-Head Self-Attention mechanism (MHSA) is integrated into the last layer of the YOLOv5s backbone network; 2) To strengthen the utilization of feature information, a Bi-directional Feature Pyramid Network (BiFPN) for feature fusion is constructed; 3) The SimAM module is incorporated into the YOLOv5s model to improve the matching of semantic and positional information. By combining the above three improvement strategies in pairs, three multi-strategy YOLOv5s detection models are built: The first model combines the Multi-Head Self-Attention mechanism (MHSA) with the BiFPN feature fusion network; The second model combines the Multi-Head Self-Attention mechanism (MHSA) with the SimAM attention mechanism; The third model combines the SimAM attention mechanism with the BiFPN feature fusion network. Comparative experiments on the VisDrone2019 dataset show that the second multi-strategy model outperforms the other two models in detection performance, which improves the mean Average Precision (mAP) to 38.9%, a 4.8% increase compared to the original model.

# **Keywords**

UAV Aerial Photography, YOLOv5s, Multi-Head Self-Attention Mechanism (MHSA), BiFPN Network, SimAM Attention Mechanism

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/



Open Access

# 1. 引言

无人机航拍图像目标检测在交通安全、建筑安全、农业安全等领域运用很广泛,提高无人机航拍图 像目标检测的准确度不仅能为各方面的应用提供便利,也能为社会安全以及社会发展助力。无人机目标 检测技术可分为传统目标检测技术和基于深度学习的目标检测技术。随着深度学习的发展,将深度学习 融合到数字图像处理领域得到了大力的发展。主要的目标检测算法有: R-CNN 算法[1]、SSD 算法[2]、 RetinaNet 算法[3]、Mask R-CNN 算法[4]以及 YOLO 系列算法。R-CNN 和 Faste R-CNN [5]等两阶段检测 算法精度高,但是对于候选框需要进行一定程度的修正,增加了时间成本。相对于两阶段检测算法,SSD [6]和 YOLO 等移动端算法无需产生过多的边框,可以直接预测出边界框,节省内存,同时保证精度。贾 世娜[7]对 Head 端和 Neck 端进行了优化,在 Head 端增加 P2 检测头,并引入 CBMA 注意力机制,优化 了 YOLOv5s 的特征融合网络。郝紫霄等[8]利用 GhostNet 代替 DarkNet-53 作为主干网络,并在主干网络 内添加注意力机制模块,实现了网络结构的轻量化,从而更快速地检测无人机航拍的小目标。李华清[9] 提出了基于 SSD 目标检测算法的 F-SSD 改进算法,并进一步将特征增强的操作添加进网络中,得到 EF-SSD 网络,提高了对于小目标航拍的精度和速度。郭君斌等[10]提出了基于检测算法的先验框优化,结合 K-means 聚类方法更新先验框的值,提高了检测召回率,得到了更好的训练效果。席光泽等[11]引入 ShuffleNetv2 轻量化模块和 Hard-swish 激活函数,使模型参数量减少的同时,进一步增加检测精度。谢忠坚 等[12]采用 GhostNetV2 网络替换 C3 模块,引入解耦全连接注意力机制,使 GhostNetV2 能够同时聚合局 部和远距离信息,并增加 ECA 通道注意力机制,从而提高特征融合的质量,进一步提升检测精度。宁涛

等[13]提出了一种级联通道 - 空间的混合注意力机制,用于获取特征之间的通道关系,并挖掘特征图的上下文信息,从而实现目标图像关键信息的提取,提升小目标的检测精度。此外,他们提出了稠密残差金字塔池化,以提高精确率和召回率,更好地保留目标区域的全局特征。吕佳铭等[14]将骨干网络优化为Rep ViT-m1 结构,并引入残差块以增强图像中低层特征的捕捉,实现对细节信息的多层次特征提取。此外,在头部网络中引入 Dynamic Head 模块,提高了多尺度信息的利用效率。

本文在 YOLOv5s 模型的基础上,利用三种改进策略,提出了三种改进的检测方案,使训练出的模型能够更好地适应航拍图像目标较小、背景复杂、特征提取困难的情形。本文提出的三种改进的检测策略为:

- 1) 利用 Bottleneck Transformer (BoT)模块替换原模型骨干网络 CSPDarknet 的最后一层 C3 卷积模块。在 CSPDarknet 的最后三个 C3 卷积模块中,用多头自注意力机制(MHSA)替换空间卷积,解决无人机目标检测中图像目标尺度差异大和小目标检测率低的问题。
- 2) 在原模型中加入小目标检测层和加权双向金字塔,将两者进行融合替代原本的特征融合网络 PANet,构成 BiFPN 特征融合网络结构,可以用来解决尺度不一、特征信息利用不充分的问题。
- 3) 在对航拍图像进行目标检测的过程中,由于背景复杂且无人机视角的特殊性,易出现语义信息与位置信息的矛盾性问题,导致模型难以提取有效的特征对目标进行检测,为解决这个问题,在模型三个检测头的最后一层添加无参数注意力(SimAM)模块,通过优化权重计算提升模型的特征检测能力。

基于上述三种改进策略,我们匹配出三种多策略 YOLOv5s 模型改进方案: ① 改进方案一: 多头自注意力机制(MHSA)和 BiFPN 特征融合网络的结合。② 改进方案二: 多头自注意力机制(MHSA)和 SimAM 注意力机制的结合。③ 改进方案三: SimAM 注意力机制和 BiFPN 特征融合网络的结合。通过对比实验结果对三种不同的改进方案进行分析评估。

#### 2. YOLOv5s 模型

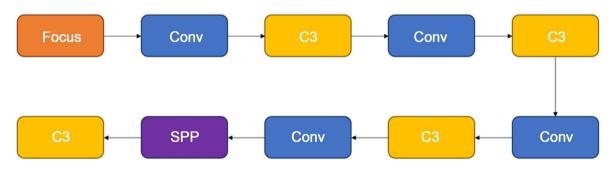


Figure 1. Backbone network structure of YOLOv5s 图 1. YOLOv5s 主干网络结构

本文使用的 YOLOv5s 模型的主干网络结构简图如图 1 所示,主要包括 Conv 模块、C3 模块和 SPP 模块。每部分的简要介绍如下所述: Conv 模块主要由卷积层、BN 层和激活函数组成。用于提取输入特征中的局部空间信息的卷积层,一般由多个卷积核组成,每个卷积核对应一个输出通道。BN 层是卷积之后加入的归一化层,用于规范神经网络中的特征值分布,可以加速训练过程,提高模型的泛化能力,减轻对初始化的依赖性。每个卷积层后面跟随一个 LeakyReLU 激活函数,用于引入非线性,从而增强模型对复杂特征的表达能力。C3 模块:每一个 C3 模块由三个 Conv 块构成。其中第一个 Conv 块的步幅为 2,可以将特征图的尺寸减半,第二个 Conv 块和第三个 Conv 块的步幅为 1。C3 模块中的卷积层(Conv 层)一般采用 3×3 和 1×1 的卷积核。每个卷积层之间加入了批量归一化(BN)层和 LeakyReLU 激活函数,提高

模型的稳定性和泛化性能。第一个 Conv 模块的步幅为 2,那么它会减半特征图的尺寸,可以使网络更加 关注物体的全局信息,提高特征提取的效果,第二个和第三个 Conv 模块的步幅为 1,目的是保证图片的 空间分辨率,更好地对物体的局部信息进行保留。SPP 模块: SPP 模块为池化模块,在该模块中,首先对 特征图实施不同的池化操作,得到一组大小不一致的特征图,将图片连接、降维,得到固定大小的特征 向量。主要由三个步骤组成:池化:将输入的特征图分别进行不同大小的池化操作,以获得一组不同大 小的特征图。连接:将不同大小的特征图连接在一起。全连接:通过全连接层将连接后的特征向量降维, 得到固定大小的特征向量。

# 3. 多策略 YOLOv5s 检测模型

本文通过应用多头自注意力机制(MHSA)的引入、SimAM (Simple Attention Module)注意力机制的添加,以及 BiFPN (Bi-directional Feature Pyramid Network)特征融合网络的整合,改进 YOLOv5s 模型。具体地,通过将上述三种改进策略两两组合,构建了三种多策略 YOLOv5s 检测模型。通过对比分析每个改进方案的性能,得出三种改进方案中相对较优的组合方案。

## 3.1. 三种改进策略

## 1) 融入多头自注意力机制的主干网络的改进

YOLOv5s 模型在处理无人机图像时,对于目标尺度差异大和小目标的情况,可能会出现检测率低的问题。因此,我们将主干网络与多头自注意力机制相结合,以提高对小目标的检测性能。MHSA 由两个模块构成,如图 2 所示,分别是: ① 自注意力模块,由 q,k,v三个矩阵的运算组成, q,k,v三个矩阵都由原特征矩阵变换而来,所以本质上来说是对自身的运算。② 多头注意力机制,是单头注意力机制的进化版,把每次的 attention 运算分组(头)进行,能够从多个维度提炼特征信息。

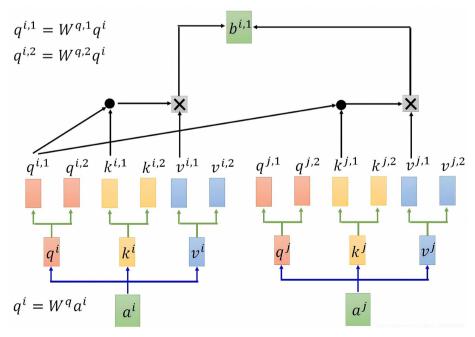


Figure 2. Schematic diagram of the multi-head self-attention mechanism (illustrated with two heads as an example) 图 2. 多头自注意力机制原理图(以两个头为例)

结合多头自注意力机制的具体改进策略是:利用 BoT 模块代替原模型骨干网络 CSPDarknet 的最后

一层 C3 卷积模块; MHSA 在 CSPDarknet 的最后三个 C3 卷积模块替换空间卷积,如图 3 所示。MHSA 代替空间卷积后,将一头线性变换分成了多个头,实现了对数据进行多视角的关注和处理。每个头执行自注意力操作且都计算出一个注意力矩阵,该矩阵在整个序列中对不同位置的词进行加权求和,以得到每个位置的表示。然后将所有头的输出(每个视角的注意力权重)拼接在一起作为最终的表示,由此提高了得到的位置信息的精确度。通过改善网络,有效地提高了检测率,使模型更好地适应图像目标尺度差异大以及是小目标的情况。

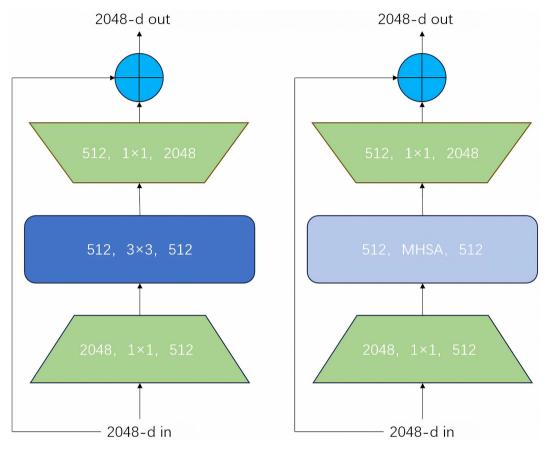


Figure 3. Comparison diagram of spatial convolution structures before and after replacement with MHSA 图 3. MHSA 替换前后空间卷积结构对比图

#### 2) 特征融合网络的改进

针对在目标检测时出现检测目标尺度不一和特征信息利用不充分的问题,利用小目标检测层与加权 双向金字塔(BiFPN)代替 YOLOv5s 的特征融合网络 PANet。加权双向金字塔来源于特征金字塔,是常用 的目标检测算法中的模块,它通过自下而上以及自上而下的方式对不同层级的特征进行融合,从而可以 表述多尺度的特征进行,以增加代码的健壮性,使得其能够在不同大小的目标上进行更加准确的检测。

传统的特征金字塔只通过上采样和下采样的操作进行特征融合,没有跳跃连接。BiFPN 引入的跳跃连接,不限于自上向下或者自底向上的路径,可以传递不同层级之间的特征,同时很好地保存高层级和低层级的语义特征信息。BiFPN 将删除节点的输入特征图和下一级特征图使用加权融合,形成新的 BiFPN 结构,如图 4 所示,对不同特征的信息进行学习,进行归一化加权,以解决尺度不一特征信息利用不充分的问题。

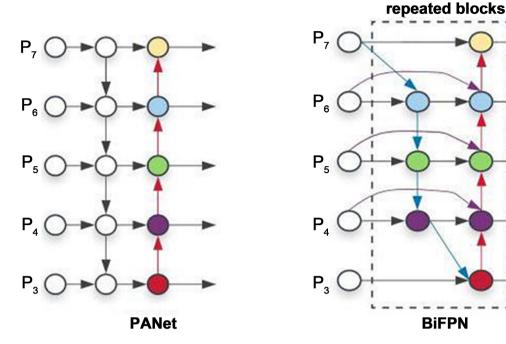


Figure 4. PANet network and BiFPN network 图 4. PANet 网络和 BiFPN 网络

# 3) 注意力模块的改进

现有的注意力机制通过加权的方式突出重要的特征,这种加权过程通常沿着通道维度或空间维度进行,以区分不同通道或空间位置的重要性:① 通道注意力: 1D 注意力,它对不同通道进行加权,以突出重要的通道特征,而对所有空间位置同等对待;② 空间注意力: 2D 注意力,它对不同空间位置进行加权,以突出重要的空间区域,而对所有通道同等对待。SimAM 注意力机制是一种无参数的注意力机制,它通过优化能量函数来计算每个位置的注意力权重,从而增强模型对重要特征的关注。因此,为了实现SimAM 注意力,需要用能量函数对每个神经元的重要性进行评估,如文献[15]为每个神经元定义了如式(1)的能量函数:

$$e_{t}(w_{t}, b_{t}, y, x_{i}) = \frac{1}{M-1} \sum_{i=1}^{M-1} \left(-1 - \left(w_{t} x_{i} + b_{t}\right)\right)^{2} + \left(1 - \left(w_{t} x_{i} + b_{t}\right)\right)^{2} + \lambda w_{t}^{2}$$

$$(1)$$

因此所得的最小能量为:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{2}$$

当能量越低,神经元与周围神经元的差异性越大。因此,每个神经元的重要性可以通过  $\frac{1}{e_{t}^{*}}$  获得。通过对单个神经元进行操作,将线性可分离性集成到一个端到端的框架中,从而实现高精度的特征提取。

SimAM 注意力机制与通道注意力机制和空间注意力机制相比,可以在不增加任何参数量的前提下,直接在网络层中推理出三维的注意力权重,通过充分利用三维数据的信息,使模型更好地捕捉到目标的特征,降低了周围环境对检测目标提取的影响,提高了目标检测时语义信息与位置信息的匹配程度。因此,为了解决被检测目标背景复杂、被遮挡、相似形状物体干扰等因素导致的语义信息与空间信息之间的矛盾,在模型的三个检测头的最后一层中添加了无参数注意力模块。原模型三个检测头的最后一层结

595

构如图 5 所示, 改进后的结构如图 6 所示。



Figure 5. Schematic diagram of the final layer of the three detection heads in the YOLOv5s model 图 5. YOLOv5s 模型的三个检测头的最后一层的结构图

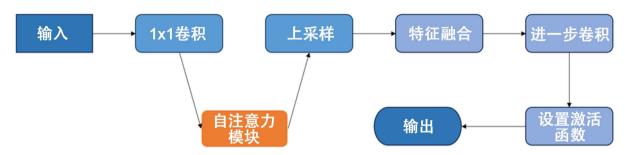


Figure 6. Schematic diagram of the final layer of the detection head after incorporating the SimAM attention mechanism 图 6. 添加 SimAM 注意力机制之后的检测头最后一层的结构简图

#### 3.2. 三种多策略 YOLOv5s 方案

为了进一步地提高 YOLOv5s 的性能,本节探讨将上一节的三种改进方法两两融合,分析它们的优势。

多策略改进方案一: MHSA 和 BiFPN 特征融合网络的结合。

该改进方案相较于传统模型具有以下优点:

1) 增强特征表征:

传统 YOLOv5 模型在特征提取和融合方面依赖于卷积神经网络。通过集成 MHSA,模型能够通过自注意力机制动态地调整不同区域的重要性,从而在全局范围内捕捉到更丰富的语义信息。这种增强的特征表征能力有助于提升模型对于复杂场景的检测性能,尤其是在小目标检测和遮挡情况下。

2) 优化特征融合机制:

BiFPN 提供了一种更为有效的特征融合策略,通过上下文信息的高级整合,它能够生成更加全面和细致的特征表示。与传统 YOLOv5 中的特征金字塔网络(FPN)相比,BiFPN 的双向融合不仅考虑了自顶向下的信息流动,还考虑了自底向上的信息反馈,这在很大程度上提高了模型对于不同尺度和分辨率下特征的综合利用能力。

多策略改进方案二: MHSA 和 SimAM 注意力机制的结合。

该改进方案相较于传统模型具有以下优点:

1) 增强的表征和交互信息:

MHSA 通过其多头的架构,能够在不同子空间中并行地关注输入特征的不同方面,从而提供了更为复杂的特征交互和表征。而 SimAM 的引入进一步强化了每个特征的重要性,使得模型在处理复杂场景时,能够更加精确地检测关键特征。

2) 效率与表征的平衡

SimAM 模块的设计理念符合神经科学理论,其通过计算能量函数来确定神经元的重要程度,无需额外的参数,从而保持了模型的参数效率。这一特点与 MHSA 的结合,不仅提升了模型的表达能力,而且维持了模型的计算效率。

多策略改进方案三: BiFPN 特征融合网络和 SimAM 注意力机制的结合。

#### 1) 综合效果

将 SimAM 与 BiFPN 结合,不仅提升了模型对于关键特征的敏感度,也增强了特征融合的全面性。 这有助于在目标检测任务中,特别是在小目标检测和复杂场景处理上,提高模型的准确性和鲁棒性。

三种改进方案各有其侧重点和优点,体现为:

方案一:通过结合多头自注意力机制 MHSA 和 BiFPN 特征融合网络,该方案在特征融合和上下文信息整合上具有显著优势。与其它两种方案相比,方案一在处理多尺度特征信息时,展现了更高的效率和准确性,能够更全面地捕捉不同层级的特征依赖,从而提高了模型对于复杂场景下目标检测的能力,尤其在处理小目标和远距离目标时表现突出。

方案二:融合了多头自注意力机制 MHSA 和 SimAM 注意力机制的方案,在注意力分配和特征表征方面具有一定的长处。相较于其他两种方案,方案二更注重于关键特征的高效提取和显著性的增强。 SimAM 的引入使得模型能够更精确地调整特征图中的注意力权重,从而增强了模型对目标特征的学习和检测能力,尤其在噪声环境下和目标遮挡情况下具有更高的鲁棒性。

方案三: SimAM 注意力机制与 BiFPN 特征融合网络的结合,相较于前两种方案,可以在不增加额外计算负担的前提下,提高特征图的表征能力和模型的检测精度。该方案在优化特征融合策略上具有一定的优势,通过 SimAM 对 BiFPN 的增强,模型能够更好地利用不同层次的特征信息,实现对目标边界和形状的更精确描绘,进而提高了模型的定位精度和分类准确性。

为了更好地比较三种多策略方案,下一节将它们应用到无人机目标检测中,对比分析它们的检测效果。

# 4. 对比实验及结果分析

实验环境使用 Python 搭建,操作系统为 Windows 11, GPU 型号为 AMD R7,运行内存为 32 GB,编程语言为 Python,编译器为 Pycharm, CUDA 版本为 12.3。使用 PyTorch 深度学习框架训练目标检测模型。初始学习率设置为 0.01,迭代次数为 300,其他为默认设置。

本文采用 VisDrone2019 进行算法验证。VisDrone 数据集由天津大学机器学习与数据挖掘实验室的 AISKYEYE 团队发布,共包含 8599 张图片。数据集来自中国 14 座不同城市的不同环境,包括城市与乡村的多种天气和光照条件。检测目标多样,稀疏环境与密集环境都具备,是目前中国无人机遥感图像中数据量最大、类型最全面、环境最复杂、覆盖范围最广的数据集之一。

VisDrone2019 数据集共包括 10 类航拍检测目标,由 7019 张训练集、548 张预测集与 1610 张测试集组成,大约有 540,000 个标注信息。10 个类别分别为 pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus 与 motor。

#### 4.1. 评价指标

对实验结果的评价指标采用精确度(Precision)、召回率(Recall)、平均精度均值(mean Average Precision, mAP)。

Precision (精确度): 以预测结果为判断依据,预测为正例的样本中预测正确的比例。预测为正例的结果分两种,要么实际是正例 TP,要么实际是负例 FP,计算公式如(3)式所示:

$$Precision = \frac{TP}{TP + FP}$$
 (3)

Recall(召回率): 以实际样本为判断依据,实际为正例的样本中,被预测正确的正例占总实际正例样

本的比例。实际为正例的样本中,要么在预测中预测正确 TP,要么在预测中预测错误 FN,计算公式如 (4)式所示:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

AP(平均精度): 绘图将计算的图像 Recall 为横轴, 计算的图像 Precision 为纵轴围成的面积是平均精度, 计算公式如(5)式所示:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) \times p(r_{i+1})$$
 (5)

mAP(平均精度均值): YOLOv5s 在所有类别的平均精度的均值,其中N表示被检测目标的类别总数,AP为单类别目标的平均精度,计算公式如(6)式所示:

$$mAP = \frac{\sum_{i=1}^{n-1} AP_i}{N}$$
 (6)

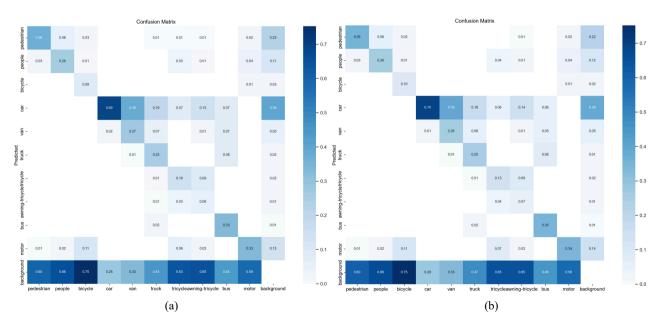
## 4.2. 数据结果分析

#### 1) 混淆矩阵

应用三种方案进行无人机目标检测的训练检验,得到混淆矩阵数据,混淆矩阵的对角线元素体现每个类别的准确率,矩阵的非对角线元素数据分析得知模型在不同类别之间的混淆程度。如图 7 所示,图 (a)为改进前模型的混淆矩阵图,图(b)、图(c)、图(d)图分别为多策略方案一、二、三的混淆矩阵图。由图可得方案一、二、三的各个类别检测精度都有提升。方案一中,车类的准确率为 70%,提升率最高,自行车类非错判率最高为 75%。方案二中,车类检测的准确率提升了 73%,精确度的提高最显著,自行车类非错判率最高为 85%。方案三中,车类的准确率提升最高为 69%,自行车类非错判率最高为 77%。

#### 2) 精确度和召回率

在图像目标检测实验中,精确度被视为评估模型性能的关键指标之一。图 8 的(a)、(b)、(c)、(d)分别为 YOLOv5s 基础模型和多策略方案一、二、三下的精确度对比图,可以看出方案二的置信度和精确度明显提高,各检测类的平均水平达到了 96.1%。方案一和方案三分别提高到了 94.6%和 95.8%。



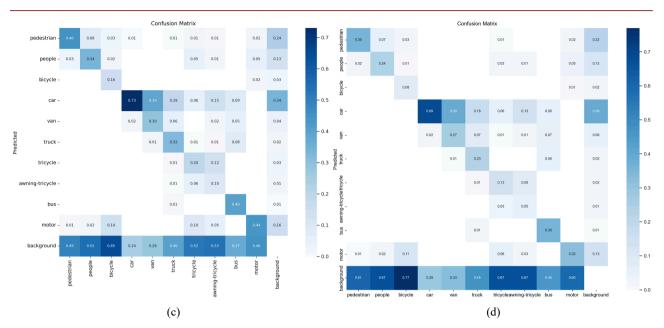


Figure 7. Comparison diagram of confusion matrices of the improved model 图 7. 改进模型的混淆矩阵对比图

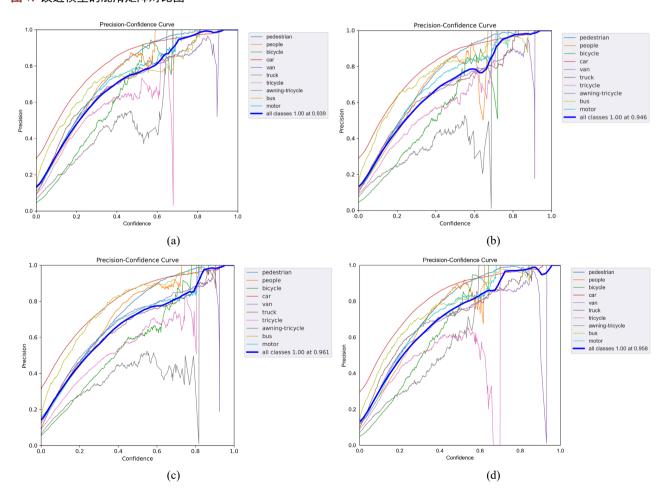


Figure 8. Comparison diagram of precision and recall 图 8. 精确度和召回率对比图

召回率体现了当置信度设置为某一数值的时候,对各个类别查全的概率进行评估,数值越高则类别检测越全面。如图 9 的(a)、(b)、(c)、(d)分别为 YOLOv5s 基础模型和多策略方案一、二、三下的召回率对比图,可以看出,在相同的置信水平下,方案二的召回率得到了提升,各检测类召回率的平均水平达到了 58%,方案一和方案三的召回率没有提高,保持在 54%。

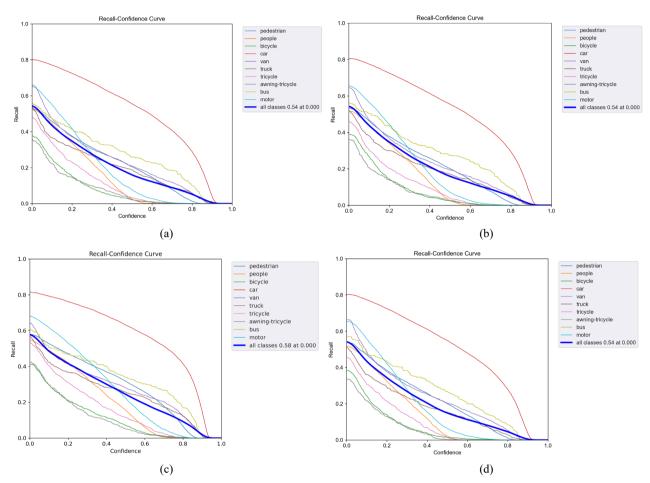
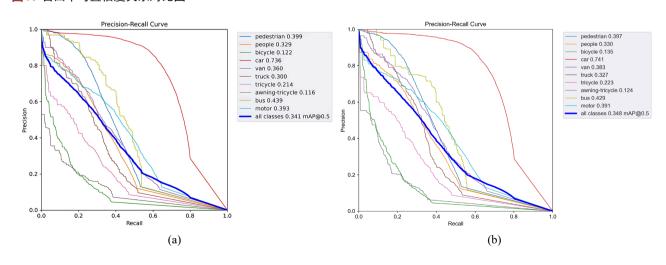


Figure 9. Comparison diagram of the relationship between recall and confidence level 图 9. 召回率与置信度关系对比图



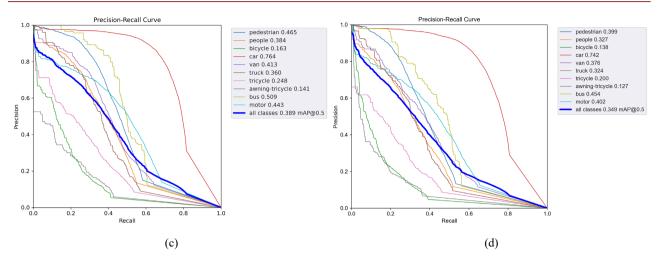


Figure 10. Comparison diagram of average precision 图 10. 平均精度对比图

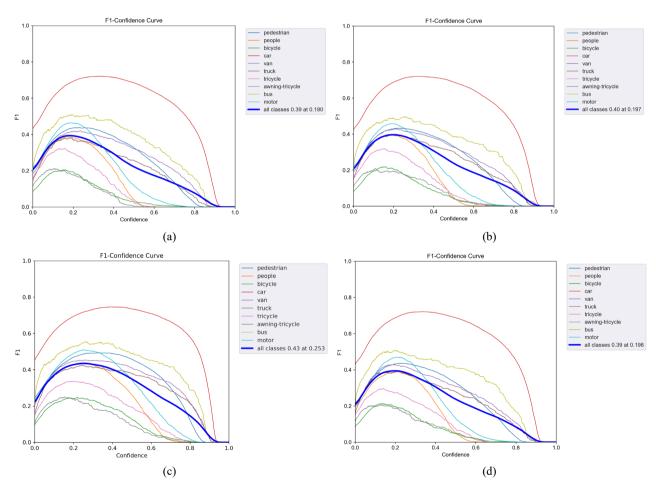


Figure 11. Comparison diagram of F1 scores 图 11. F1 分数对比图

# 3) 平均精度和 F1 得分

平均精度是衡量检测检验效果的重要指标, mAP 综合考虑了模型的精确度和召回率, 是目标检测领

域广泛认可的评价标准,如图 10 的(a)、(b)、(c)、(d)分别为 YOLOv5s 基础模型和多策略方案一、二、三下的平均精度(AP),由图(a)和(c)可以看出,方案二中各检测类 AP 数据都有明显提高,其中 pedestrian 类、bus 类、blcycle 类、van 类和 motor 类的 AP 数据检测效果提升很高,分别提升了 6.6%、7.0%、4.1%、5.3%和 5.0%,mAP 数值也由 34.1%提升到了 38.9%,提升了 4.8%。由图(a)、(b)、(d)可知,方案一和方案三的 mAP 数值分别提升到了 34.8%和 34.9%,分别提升了 0.7%和 0.8%,与方案二相比有明显的差距。

F1 分数是精确率和召回率的调和平均数,对三个创新方案对应的数据进行检验测试,如图 11 的(a)、(b)、(c)、(d)分别为 YOLOv5s 基础模型和多策略方案一、二、三下的 F1 分数。方案二中,各类的平均 F1 得分都有所提升,其中 car 的 F1 分数依然是最高的,进一步接近 0.80。而由图(b)、(d)可知,方案一的 F1 分数没提升,方案三的 F1 分数提升至 0.40,方案二的提升效果较好。

#### 4.3. 改进效果分析

将三种改进方案在如下三种情形的检测效果进行对比:小目标情况下的检测、目标尺度差异大的情况下的检测以及昏暗条件下的检测。对比检测效果如图 12~14 所示。由图中的数据对比分析可知,方案二在精确度、召回率、平均精度均值以及 F1 分数等多个关键评价指标上均表现最优。MHSA 的全局感知能力与 SimAM 的轻量化特征强化能力的结合,提升了精确度和召回率的综合性能,达到了更好的平衡,使 mAP 从 34.1%提升至 38.9%,F1 得分也明显提高,尤其在车类目标的检测中表现最为突出。相比之下,方案一虽然在精确度上有所提升,但在召回率和综合性能上表现不足;方案三因缺乏全局感知能力而在精确度提升上受限,导致其综合性能提升幅度较小。因此,方案二凭借其在精确度与召回率平衡上的优势,是三种改进方案中相对较优的选择。



Figure 12. Comparison of detection performance for small objects 图 12. 小目标情况下检测效果对比



Figure 13. Comparison of detection performance for objects with large scale variation 图 13. 目标尺度差异大时检测效果对比

我们对方案二的多策略模型在实际场景中的应用进行了进一步检测。下面三组对比图中,左图为原始 YOLOv5s 模型的实验效果,右图为方案二的多策略模型的实验效果。从图 12 中可以明显看出,检测小目标时,对各类检测的数量和精确度方面都有提高。由图 13 可知,当检测目标尺度差异大时,车类和行人类存在检测目标尺度差异大的问题,本文算法与原始模型相比,对行人的检测数量大幅增加,且对车类检测的精确度也有明显提高。通过图 14 中左右两图的对比可知,在昏暗条件下,方案二的多策略模型检测出的类别和数量都增加了,从而降低了漏检、误检的可能性,进而能够更多地检测出行人目标,同时增加对目标的确信度。综上所述,方案二多策略模型相较于原始的 YOLOv5s 模型具有显著优势,检测精度高,对环境干扰的鲁棒性强,并且在复杂环境下对小目标检测的特征提取能力显著提高。



Figure 14. Comparison of detection performance under low-light condition 图 14. 昏暗条件下检测效果对比.

## 5. 结束语

本文对 YOLOv5s 模型进行三种策略的改进,得到三种多策略 YOLOv5s 检测模型,并进行了复杂环境下无人机小目标检测的对比实验。由精确度、召回率和平均精度三个方面进行比较后可知,改进方案二,即融入 MSHA 多头自注意力机制,并利用 BoT 模块取代原模型骨干网络 CSPDarknet 的最后一层 C3 卷积模块;在模型三个检测头的最后一层引入无参数注意力(SimAM)模块,其改进效果方面表现优异,相较于其他方法具有显著优势。其检测精度较高,且对环境干扰表现出较强的鲁棒性。

#### 基金项目

2024 年燕山大学科研项目(2024 年燕山大学省级大学生创新创业训练计划项目)(S202410216065); 2023 年燕山大学创新创业教育教学改革与实践项目(2023xjexcy013)。

# 参考文献

- [1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 24-27 June 2014, 580-587. https://doi.org/10.1109/cvpr.2014.81
- [2] Jisoo, J., Hyojin, P. and Nojun, K. (2017) Enhancement of SSD by Concatenating Feature Maps for Object Detection. arXiv: 1705.09587.
- [3] 张杨, 辛国江, 王鑫, 等. 基于改进的 YOLOv5 网络的舌象检测算法[J]. 计算机技术与发展, 2024, 34(2): 156-162.
- [4] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.322
- [5] Girshick, R. (2015) Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Chile, 13-16 December 2015, 1440-1448. https://doi.org/10.1109/iccv.2015.169
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot Multibox Detector. In:

Proceedings of the European Conference on Computer Vision (ECCV), Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0\_2

- [7] 贾世娜. 基于改进 YOLOv5 的小目标检测算法研究[D]: [硕士学位论文]. 南昌: 南昌大学, 2023.
- [8] 郝紫霄, 王琦, 高尚. 基于 YOLO-v5 算法的航拍图像小目标检测改进算法[J]. 常州大学学报, 2023, 35(6): 45-51.
- [9] 李华清. 基于 SSD 的航拍图像小目标快速检测算法研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2019.
- [10] 郭君斌,于琳,于传强. 改进 YOLOv5s 算法在交通标志检测中的应用[J]. 国防科技大学学报, 2024, 46(6): 123-130.
- [11] 席光泽, 周建平, 许燕, 等. 基于改进 YOLOv5s 的复杂环境下棉花顶芽检测[J]. 中国农机化学报, 2024, 45(12): 275-280.
- [12] 谢忠坚,廖珩宇,文春明,等. 基于改进 YOLOv5s 的蔗节检测方法[J]. 中国农机化学报, 2024, 45(12): 224-229.
- [13] 宁涛, 付世沫, 常青, 等. 基于改进 YOLOv5s 的无人机航拍图像目标检测[J]. 电光与控制, 2024, 31(12): 41-47+63.
- [14] 吕佳铭, 张峰, 罗亚波. 基于改进 YOLOv5s 的烟梗物料目标检测算法[J]. 浙江大学学报(工学版), 2024, 58(12): 2438-2446.
- [15] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18-22 June 2018, 8759-8768. https://doi.org/10.1109/cvpr.2018.00913