国防科技大学学报
Journal of National University of Defense Technology
ISSN 1001-2486,CN 43-1067/T

《国防科技大学学报》网络首发论文

题目：多尺度学习的红外无人机目标检测算法
作者：左震，袁书东，李灿，黄泓赫
收稿日期：2024-09-28
网络首发日期：2025-09-11
引用格式：左震，袁书东，李灿，黄泓赫. 多尺度学习的红外无人机目标检测算法[J/OL]. 国防科技大学学报. https://link.cnki.net/urlid/43.1067.t.20250910.1447.002

doi: 10.11887//j.issn.1001-2486.24090041

多尺度学习的红外无人机目标检测算法

左震，袁书东*，李灿，黄泓赫

（国防科技大学 智能科学学院，湖南 长沙 410073）

摘　要：针对无人机目标体积小、在图像中所占像素少、纹理细节信息弱、算法难以有效提取红外无人机目标特征导致检测精度较低等问题，提出多尺度学习的目标检测算法。通过在模型的颈部网络中构造多尺度特征融合结构，引入多尺度特征学习模块，将深层网络和浅层网络的特征进行级联，获取目标在多个尺度上的特征，丰富特征图的语义信息和特征信息，显著提高了算法对小型无人机目标检测精度。在训练过程中使用 SIoU 代替 CIoU 损失函数，使网络模型在训练过程中损失最小化，提高了回归精度。实验结果表明，与其他红外小目标、主流检测算法相比，本文所提方法能有效提高无人机目标的检测精度，在实际应用中可以满足探测无人机目标的检测精度需求。

关键词：红外图像；探测无人机；多尺度学习

中图分类号：TP391.4　　**文献标识码：**A

Multi-scale learning algorithm for infrared UAV target detection

Zuo Zhen , Yuan Shudong, Li Can , Huang Honghe*

(College of Intelligent Sciences, National University of Defense Technology, Changsha 410073, China)

Abstract: The issues of small UAV (unmanned aerial vehicle) target size, limited pixel coverage in images, weak texture detail information, and the difficulty in effectively extracting infrared UAV target features, which lead to low detection accuracy, were addressed by proposing a multi-scale learning-based target detection algorithm. A multi-scale feature fusion structure was constructed in the neck network of the model, and a multi-scale feature learning module was introduced. Features from both deep and shallow networks were cascaded to capture target features at multiple scales, enriching the semantic and feature information of the feature map, which significantly improved the detection accuracy of small UAV targets. During training, SIoU was used in place of CIoU loss, minimizing the network model's loss and enhancing regression accuracy. Experimental results demonstrate that, compared to other infrared small target detection algorithms and mainstream methods, the proposed approach effectively improves the detection accuracy of UAV targets and meet the detection accuracy requirements for UAV target detection in practical applications.

Keywords: Infrared imagery; detecting drones; multi-scale learning

* **收稿日期：**2024-09-28

基金项目：国家自然科学基金资助项目（52101377）

第一作者：左震（1982—），男，安徽安庆人，副研究员，博士，硕士生导师， E-mail：z.zuo@nudt.edu.cn;

* **通信作者：**袁书东（1999—），男，湖南常宁人，博士，E-mail：yuanshudong21@nudt.edu.cn

引用本文：左震，袁书东，李灿，等. 多尺度学习的红外无人机目标检测算法[J]. 国防科技大学学报

Citation： ZUO Z, YUAN S D, LI C, et al. Multi-scale learning algorithm for infrared UAV target detection[J]. Journal of National University of Defense Technology

目前，黑飞无人机对公共安全构成了重大威胁。因此，及时发现这些无人机并采取有效的应对措施势在必行。红外相机适用于各种复杂环境，可以全天候工作。基于红外图像的低空无人机目标探测具有很大的应用前景和发展潜力。但红外图像中的无人机目标有其独特的“弱”和“小”的特点，且由于红外成像设备受传感器灵敏度、背景温度、红外噪声等因素的影响，成像质量较差。上述问题导致算法对红外无人机目标特征提取困难，检测精度较低。当前提升红外小目标检测精度的方法主要集中在多尺度融合方面。

在多尺度融合方面，ZHAI 等人介绍了一种用于探测低空无人机目标的增强型方法，该方法保留了原有 YOLOv3 架构的基本框架，但通过多尺度预测方法对其进行了改进，以增强对小型目标物体的检测能力[1]。WANG 等人为解决小目标检测问题，提出了相似性融合模块，利用相似性选择性地融合多尺度特征，有效避免了小尺度特征被大尺度特征覆盖，导致小目标检测不到[2]。SUNKARA 等人提出了由两阶段特征学习管道和廉价的线性变换组成的 YOGA 算法，只使用传统卷积神经网络所需的一半卷积滤波器来学习特征映射。但是，两阶段特征学习流水线会占用大量计算资源，不便于轻量级操作[3]。XU 等人针对现有检测算法无法有效区分和融合多种特征的问题，提出了一种用于检测任务的密集多尺度特征学习网络，从而有效提取图像中的目标信息，提高检测效果[4]。YANG 等人依托自下而上的多尺度特征融合网络，横向联立建立一个自上而下的路径，通过横向联立建立多尺度高级语义特征的综合层次结构[5]。

在此基础上，WANG 等人提出了 PANet 网络，引入了一条自下而上的路径，实现了层间特征共享，以促进高级特征与来自低级特征的足够细节的整合[6]。GHIASI 等人利用神经架构搜索，构建了 NAS-FPN 网络，采用空间搜索策略在特征层内建立跨层连接从而实现可扩展的特征信息[7]。ZHAO 等人构建了一个多层级的提取多层次和多尺度特征的 M2Det 网络、为跨层特征融合提供了便利[8]。YU 等人提出了多尺度局部对比度学习和双线性特征金字塔网络，在网络训练过程中学习局部对比度特征，以充分提取多尺度目标特征[9]。针对目标纹理特征差、对比度低等问题，李等人基于 YOLOv4 架构提出了一种融合通道注意力机制的多尺度红外目标检测模型。该模型通过降低主干特征提取网络深度，减少了模型参数，提高了计算资源的利用率[10]。针对目标存在形态多变以及特征过少等问题，张等人对 YOLO 网络结构进行重构，搭建多尺度网络，增加目标检测层，提高对小目标的检测能力[11]。

利用多尺度融合的方式主要有两个方面的显著优势：(1)能够在不增加额外计算成本的前提下提升网络性能，(2)能够生成包含高分辨率信息的特征图。

在上述研究的基础上，本文创新工作如下。(1)在模型的骨干网络中引入三维无参注意力机制减弱背景干扰。在模型的颈部网络构造了多尺度融合结构，通过加入多尺度特征学习模块，融合目标的多尺度特征，提升算法对目标全局特征信息提取能力；(2)优化损失函数加快算法的收敛速度；(3)针对当前特定场景中红外无人机目标数据集匮乏的问题，制作了一个含有多场景的单帧红外图像目标检测数据集。

1 红外目标检测难点

红外目标与干扰源具有相似的特征，且在整幅图像中目标的形状或轮廓依然表现为弱纹理特征，且在图像中面积占比小，难以直观有效地判断该目标是否为无人机目标。

目前检测红外小目标是一个非常具有挑战性的问题，导致目标检测方法对于红外小目标的检测性能不佳主要有以下四个因素：

(1) 卷积神经网络中的卷积步长较大。在卷积过程中，特征图尺寸不断缩小，而卷积步长比红外小目标尺寸大，导致小目标特征很难传递到深层网络。

(2) 当前目标检测数据集中的样本分布状况并不理想。其中小目标的样本数量相对较少，而大目标和小目标之间存在显著的尺寸差异，导致检测算法难以适应目标的尺寸变化。

(3) 先验框的超参数设置欠优。在目标检测过程中，红外小目标的尺寸往往与设定的先验框尺寸相差较大，只有小部分先验框与标注的真实框能够重叠，算法的检测效果不佳。

(4) 交并比阈值的选择不固定。红外小目标的候选边界框与真实框之间的交并比较小，交并比阈值的大小会影响训练的效果，从而影响检测性能。

2 多尺度学习的目标检测算法

2.1 算法总体结构

本文提出了多尺度学习的目标检测算法 IRSDD-YOLOv8，该算法的网络模型总体框架如图 1 所示，在传统 YOLOv8n-seg 架构的基础上，

在颈部网络构建多尺度特征融合结构，引入多尺度特征学习模块，以提升特征提取能力和检测性能。IRSDD-YOLOv8 算法模型总体结构如下。

(1) 输入层接收图像并进行预处理。使用 YOLOv8n-seg 的 Conv 和 C2f 模块作为骨干网络，从输入图像中提取目标特征。C2f 模块通过残差结构和跨阶段部分网络来提高特征提取能力。此外，针对无人机目标纹理信息弱等问题，在骨干网络中引入三维无参注意力机制，增强算法对无人机目标特征的提取能力并减弱背景干扰。

(2) 在颈部网络中，借鉴特征金字塔、自上而下和自下而上的路径聚合结构，构建多尺度特征融合结构，引入多尺度特征学习模块。在颈部网络中分别与来自骨干网络的第 2 层、第 4 层、第 6 层的输出特征图级联，以融合来自不同尺度的特征，这样既能得到包含对象类别的高层次抽象特征，又能保留浅层特征图的低层次的边缘纹理等细节信息，有助于提升红外无人机小目标的检测性能。

(3) 基于 YOLOv8n-seg 的原有预测头，增加一个微小目标预测头(PH_1)，使得网络能够更好地提取不同尺度和复杂背景下的目标特征并预测每个网格单元的边界框、类别概率和置信度得分。

(4) 优化损失函数，有助于更快地学习到正确的边界框位置，加速训练过程中的收敛速度。本文提出的多尺度学习的目标检测算法具有以下突出优势。(a) 多尺度检测。IRSDD-YOLOv8 算法借鉴了特征金字塔、自上而下和自下而上的路径聚合结构，能够有效处理不同尺度的目标，减小特征图池化的速度，使得小目标的特征能够更好地传递到深层网络。(b) 高效检测。在保留 YOLOv8n-seg 高效推理速度的同时，增强了特征提取能力，使得 IRSDD-YOLOv8 算法在精度和速度之间达到了较好的平衡。

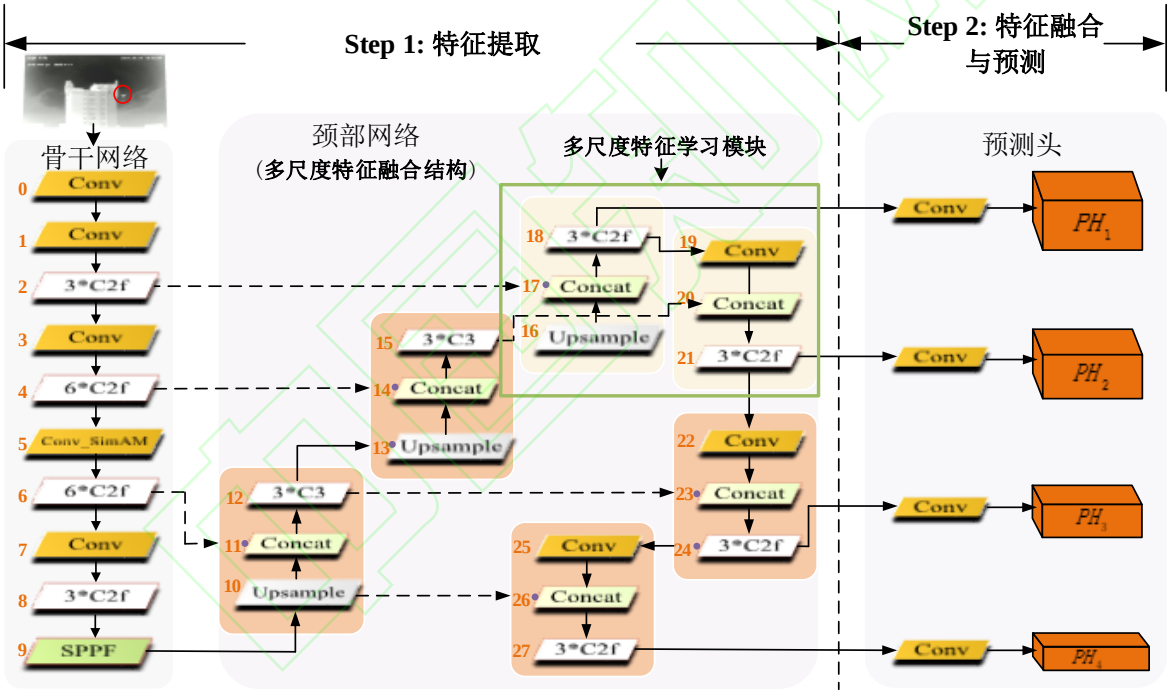


图 1　IRSDD-YOLOv8 网络结构
Fig.1 IRSDD-YOLOv8 network structure

2.2　多尺度特征融合结构

典型的特征融合算法的网络结构如图 2 所示，白色方块部分代表特征融合结构，绿色代表微小目标检测器，橙色代表小目标检测器，黄色代表中中型目标检测器，蓝色代表大目标检测器。在图 2(a) 中，特征金字塔(feature pyramid networks, FPN) 构建了一个自上而下的特征网络、在不同层次的特征图之间添加横向连接，以更好地利用低层次特征中包含的高频细节，从而形成一个包含多尺度信息的特征网络。在图 2(b) 中，自上而下和自下而上的路径聚合结构网络(path aggregation network, PANet)在编码器和解码器之间加入了路径聚合模块，以聚合不同尺度的特征，生成多尺度的特征图。在图 2(c)中，YOLOv8n-seg 借鉴了 PANet 的思想，来提高物体检测的准确性和效率。

本文在多尺度学习的目标检测算法 IRSDD-YOLOv8 的颈部结构采用了一种新颖的多尺度特征融合结构，该结构如图 2(d) 所示，

融合步骤主要如下。

(1) 使用卷积神经网络作为骨干网络来提取图像中的特征。随着网络层数加深，特征图的空间尺寸会逐渐减小，深层特征图更倾向于包含对象类别的高层次抽象特征，而浅层特征图则保留了更多的低层次的边缘纹理等细节信息。

(2) 将深层特征图进行上采样，与相应尺度的浅层特征图进行拼接，帮助恢复因下采样丢失的空间信息，并且将深层的语义信息传递给浅层。

(3) 将自顶向下路径中的特征图与来自主干网络的同尺度特征图相结合，帮助每个尺度都获得既有语义信息又有空间信息的特征图。通过此方式，可构建出一个多尺度的特征网络，每个层级的特征图都可用于后续不同尺度目标检测任务。

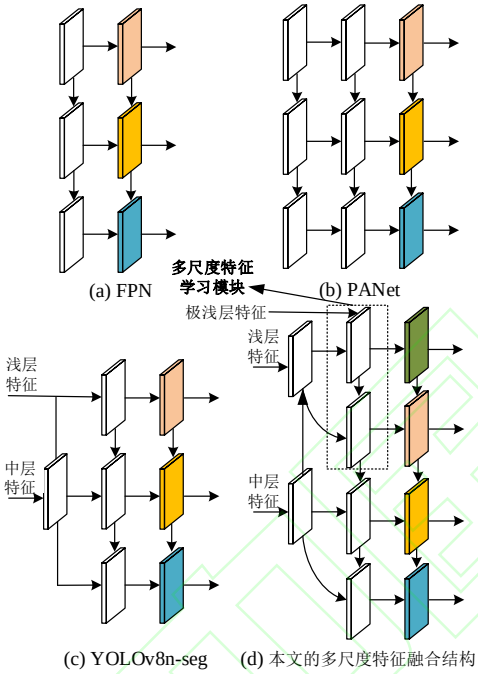


图 2 特征融合结构的比较

Fig.2 Comparison of feature fusion structures

2.3 多尺度特征学习模块

浅层特征网络保留了空间位置信息，但却无法充分提取目标的语义信息；深度特征网络因为对输入图像进行了多次卷积和池化操作来提取图像特征而失去了目标的空间细节信息。若模型未能充分利用来自多个特征层的信息，则无法有效学习目标的特征。为此，本文通过在 YOLOv8n-seg 网络模型中添加多尺度特征学习模块，融合浅层网络得到的空间纹理信息与中层网络得到的语义信息，从而保留更多可供网络模型学习的目标特征，避免无人机目标被深层网络卷积后导致的目标特征丢失。图 3 展示了多尺度特征学习模块的工作过程。

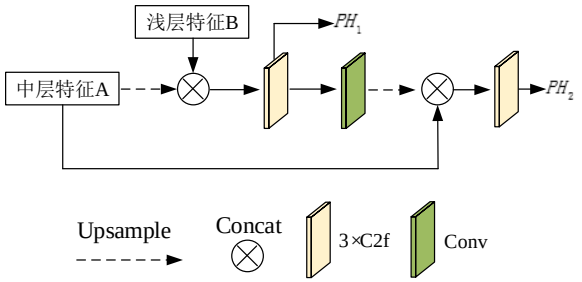


图 3 多尺度特征学习模块

Fig.3 Multi-scale feature learning module

多尺度特征学习模块融合目标全局信息的方法如式(1)-(2)所示：

$$H_1 = \mathrm{C2f\ module}(\delta(A)\otimes B) \tag{1}$$

$$H_2 = \mathrm{C2f\ module}(\mathrm{A}\otimes \mathrm{Conv}(H_1)) \tag{2}$$

其中，$A\in\mathbb{R}^{C\times H\times W}$ 和 $B\in\mathbb{R}^{C\times H\times W}$ 分别表示来自于中层网络和浅层网络的特征图。δ 表示上采样，$\otimes$ 表示对不同的输入进行级联。对中层语义特征图 A 进行上采样，与对应的浅层特征图 B 级联经过 C2f 模块处理后得到 H_1，送入微小目标预测头 PH_1，再对进行 H_1 进行卷积和上采样，与中层语义特征图 A 级联，再经过 C2f 模块处理得到输出 H_2，将得到特征图 H_2 送入小目标预测头 PH_2，融合分类后解耦出图像中物体的类别和位置信息。

2.4 损失函数优化

YOLOv8n-seg 网络模型中的损失函数由 $L_{\text{class loss}}$、$L_{\text{box loss}}$ 组成，$L_{\text{class loss}}$ 和 $L_{\text{box loss}}$ 分别是分类损失与边界框损失。分类损失采用 BCE With Logits Loss 函数进行计算，而边界框损失函数通过 DFL Loss， CIoU Loss 进行计算。其中，CIoU 损失计算公式如下：

$$L_{\text{box IoU}} = 1 - I_{\text{IoU}} + \frac{p^2(b,b_{\text{gt}})}{c^2} + \alpha\nu \tag{3}$$

在式(3)中，b 表示预测框，b_{gt} 表示真实框。c 表示能够同时包含预测框和真实框的最小闭包区域的对角线距离，p 表示两个矩形框之间的欧氏距离，α 为平衡参数，ν 用来衡量框的长宽比是否一致。

$$\begin{cases} \nu = \dfrac{4}{\pi^2}(\arctan\dfrac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan\dfrac{w}{h})^2 \\ \alpha = \dfrac{\nu}{(1-I_{\text{IoU}})+\nu} \end{cases} \tag{4}$$

$$R = \frac{p^2(b,b_{\text{gt}})}{c^2} + \alpha\nu \tag{5}$$

从式(5)可以看出，当预测框与真实框的长宽

比一样大时，v 取 0，此时长宽比的惩罚项 R 并没有起到作用，CIoU 损失函数不能稳定工作[12]。为解决上述问题，本文通过引入新的损失函数 SIoU 优化 YOLOv8n-seg 的边界框损失函数。SIoU 引入了角度损失、距离损失、形状损失和 IoU 损失，考虑了预测框与目标框之间的匹配性，重新定义了距离损失，有效降低了回归的自由度[13]。

3 数据集构建与实验设置

3.1 数据集构建

3.1.1 公开数据集

当前针对低空弱小目标检测应用的公开数据集有第一届 CVPR(IEEE conference on computer vision and pattern recognition)反无数据集、和地/空背景下红外图像数据集[14]。第一届 CVPR 反无数据集由 160 个视频序列组成，涵盖了大量的无人机飞行的素材，但该数据集只提供了无人机在图像中的坐标与目标框大小，未对目标位置区域进行像素级的标注，并未细分无人机的飞行场景。IRSTD-1K 数据集是一个用于红外小目标检测的公开数据集，该数据集包含 1001 张红外图像，提供了图像中每个目标框的宽度、高度等信息；地/空背景下红外图像数据集是以固定翼无人机为探测对象，在不同背景下对该固定翼无人机进行数据采集与整理而成。

3.1.2 自制数据集

针对当前无人机目标探测领域中数据集不足的情况，本文挑选了公开的地/空背景下红外图像数据集与第一届 CVPR 反无数据集的部分数据用于制作自制数据集。对处于不同场景的目标进行分类，并进行了像素级的标注。自制的单帧红外图像目标检测数据集(single-frame infrared image object detection dataset, SIDD)包含 4737 张 640×512 像素的红外图像。

为了尽可能地模拟无人机的真实入侵场景，本文在 SIDD 数据集中划分了为四个场景，以探究不同背景对低空红外无人机目标检测精度的影响。表 1 展示了 SIDD 数据集集中不同场景训练集和测试集中的图像数量。

表 1 SIDD 训练集和测试集中的图像数量

Tab.1 Number of images in the SIDD training and test sets

模拟场景	训练集	测试集	总计
城市	874	219	1093
山地	1720	431	2151
海面	570	143	713
天空	624	156	780
总计	3788	949	4737

在对红外无人机目标检测的研究中，我们分析了不同场景下目标在图像中的尺度分布。通过对城市场景、山地场景、海面场景以及天空场景中目标区域与整个图像面积的占比进行了统计，目标面积大多占图像面积极低。

3.2 实验设置

本文的实验平台的软硬件配置如表 2 所示。所有检测算法在数据集上均被训练 50 个周期。在训练过程中，每次迭代处理 2 张图像，设置模型的初始学习率为 0.0025，设置权重衰减为 0.005，选择 AdaGrad 作为训练优化器。

表 2 软硬件环境配置

Tab.2 Hardware and software environment configuration

项目	配置
硬件配置	GPU：NVIDIA GeForce RTX3060 CPU：AMD Ryzen 7 5800H
软件配置	系统：Ubuntu18.04 版本：CUDA11.1

3.3 算法评价指标

本文使用平均交并比、归一化交并比、平均精度、ROC 曲线、每秒检测图像张数(Frames Per Second，FPS)等指标评估不同算法的检测性能。

交并比：指预测的目标掩码与目标真实区域的并集与交集之比，其计算方式如式(6)所示：

$$I_{\mathrm{IoU}}=\frac{S_{\mathrm{overlop}}}{S_{\mathrm{union}}} \tag{6}$$

平均交并比：平均交并比是对所有类别的交并比取平均值。定义如下：

$$\mu_{\mathrm{IoU}}=\frac{1}{N}\sum_{i=1}^{N}(I_{\mathrm{IoU}})_i \tag{7}$$

归一化交并比：作为红外小目标探测模型和数据驱动方法之间更平衡的指标。定义如下:

$$n_{\mathrm{IoU}}=\frac{1}{N}\sum_{i}^{N}\frac{T_p[i]}{T_p[i]+F_p[i]+F_n[i]} \tag{8}$$

其中，T_p 是指被模型预测为正且确实是正样本的数量。F_p 是指被模型预测为正，但实际是负样本的数量。F_n 是指被模型预测为负，但实际上是正样本的数量。

通过计算精度-召回曲线下方的面积得到 m_{AP} 值来评估检测精度。计算方式如式(9)所示：

$$m_{\mathrm{AP}}=\int P(R)\mathrm{d}R \tag{9}$$

其中，P 指精度，R 指召回率。P 和 R 的计算方式如下：

$$\begin{aligned} P&=\frac{T_p}{T_p+F_p}\\ R&=\frac{T_p}{T_p+F_n}\end{aligned} \tag{10}$$

平均精度： 使用 m_{AP}@0.5:0.95 评估 IoU 阈值为 {0.5,0.55, ⋯⋯， 0.95} 时的平均值， 使用 m_{AP}@0.5 评估 IoU 阈值为 0.5 时的取值。

ROC 曲线：横轴表示假阳性率。纵轴表示真阳性率。

每秒检测图像张数：每秒处理帧数。

4 实验结果与分析

4.1 对比实验

4.1.1 与典型红外小目标检测方法对比结果

红外无人机目标属于经典的红外小目标，本文通过对比传统的红外小目标检测算法 Top-hat[15]、MPCM[16]，基于深度学习的算法 UIUnet[17]、DNANet[18]、ISTDUNet[19]，评估 IRSDD-YOLOv8 算法的性能，明确 IRSDD-YOLOv8 是否在准确性、鲁棒性上有显著改进。

表 3 是 IRSDD-YOLOv8 算法与其他红外小目标检测算法在数据集中的检测结果。在城市场景中， IRSDD-YOLOv8 的 μ_{IoU} 和 n_{IoU} 指标分别达到了 82.4%和 83.1%的良好性能，，达到了业内的先进水平。在山地场景中，IRSDD-YOLOv8 的 μ_{IoU} 和 n_{IoU} 指标分别达到了 0.752 和 0.770 的良好性能，远高于传统的红外小目标检测算法和其他基于深度学习的算法，与原算法 YOLOv8n-seg 相比， μ_{IoU} 和 n_{IoU} 指标分别提高了 2.4%和 2.8%。

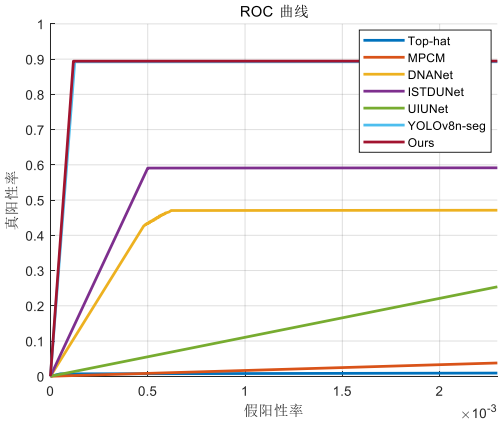
IRSDD-YOLOv8 算法在海面场景中的 μ_{IoU} 和 n_{IoU} 值分别为 66.5% 和 65.3%，虽然 IRSDD-YOLOv8 算法在该场景下的 μ_{IoU} 和 n_{IoU} 值相较于其他场景较低，原因在于该场景下的目标检测条件最为苛刻，不仅背景复杂，且面积占比最小，但本文所提算法相较于原算法指标依然有一定提升。IRSDD-YOLOv8 算法在天空场景中的 μ_{IoU} 和 n_{IoU} 值则达到了 83.9%和 80.4%，检测性能优于其他红外小目标检测算法。

表 3　在不同场景下红外小目标算法的检测结果

Tab.3 Detection results of infrared small target algorithms in different scenarios s

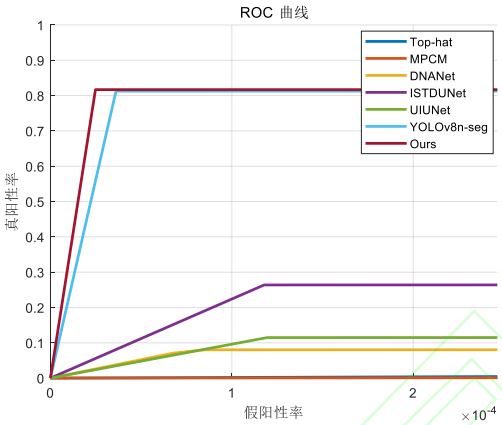
算法评价指标	检测算法	城市场景	山地场景	海面场景	天空场景
μ_{IoU}	Top-hat	0.010	0.006	0.110	0.097
	MPCM	0.027	0.001	0.004	0.034
	UIUnet	0.136	0.125	0.288	0.504
	DNANet	0.364	0.062	0.114	0.464
	ISTDUNet	0.608	0.367	0.480	0.586
	YOLOv8n-seg	0.812	0.728	0.649	0.837
	IRSDD-YOLOv8	**0.824**	**0.752**	**0.665**	**0.839**
n_{IoU}	Top-hat	0.010	0.013	0.134	0.172
	MPCM	0.028	0.001	0.005	0.043
	UIUnet	0.322	0.200	0.402	0.595
	DNANet	0.366	0.062	0.114	0.468
	ISTDUNet	0.571	0.336	0.475	0.573
	YOLOv8n-seg	0.820	0.742	0.639	0.804
	IRSDD-YOLOv8	**0.831**	**0.770**	**0.653**	**0.806**

图 4 为不同算法的 ROC 曲线比较结果，本文提出的 IRSDD-YOLOv8 算法达到了最佳性能，在同等的虚警率条件下，检测成功率更高，优于其他经典的红外小目标检测算法。值得注意的是，传统的检测算法的性能传在很大程度上取决于先验假设，它们无法适应复杂背景的变化，在 ROC 方面表现不佳。



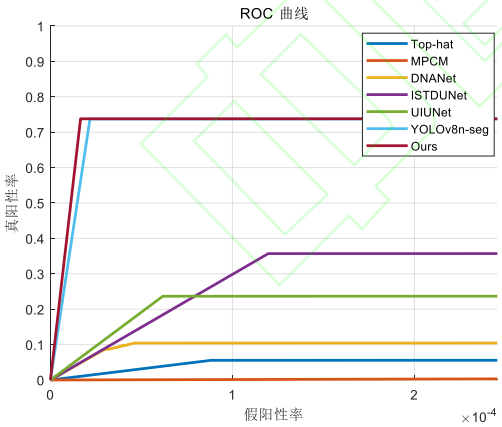
(a) 城市场景

(a) City scene



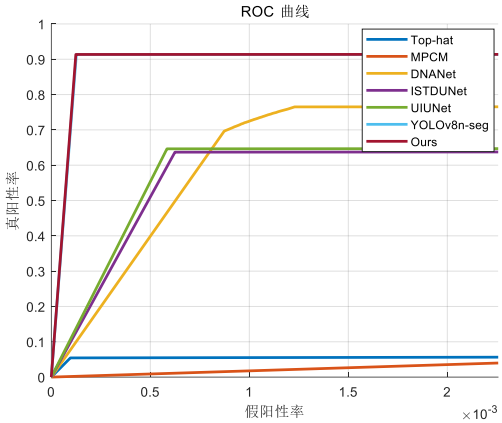
(b) 山地场景

(b) Mountain scene



(c) 海面场景

(c) Sea scene



(d) 天空场景

(d) Sky scene

图 4 IRSDD-YOLOv8 和其他红外小目标检测算法的 ROC 结果

Fig.4 ROC results for IRSDD-YOLOv8 and other infrared small target detection algorithms

4.1.2 与主流 SOTA 检测方法对比结果

为了验证本文提出的 IRSDD-YOLOv8 算法的检测性能，本文在自制数据集上对所提算法与 BoxInst[20]、CondInst[21]、Yolact++[22]、SOLOv2[23]、Mask-Rcnn[24]、YOLOv5、YOLOv7[25]、YOLOv8 等 SOTA(State of the art)方法进行比较。为了保证实验的完整性，对算法在自制的 SIDD 数据集的不同场景的检测结果进行了定量和定性比较。

（1）定量结果

如表 4 所示，在城市场景下，IRSDD-YOLOv8 算法的 m_{AP}@0.5:0.95 和 m_{AP}@0.5 值分别达到 51.1%和 93.9%，m_{AP}@0.5 值与其他先进算法相比，取得了排名第一的结果，比原算法 YOLOv8n-seg 分别高出 1.2%，尽管提升比例较低，但在高精度状态下，模型已经接近其性能瓶颈，进一步提升的空间变得非常有限，且 mAP@0.5 值达到了 93.9%。

在山地场景下，目标占比面积小，导致原算法 YOLOv8n-seg 的 m_{AP}@0.5:0.95 和 m_{AP}@0.5 值分别只有 25.3%和 66.4%。但是通过本文的改进，算法精度分别达到了 30.9% 和 78.5%，比原算法 YOLOv8n-seg 分别高出 5.6% 和 12.1%。

在海面场景下，IRSDD-YOLOv8 算法的 m_{AP}@0.5:0.95 和 m_{AP}@0.5 分别达到 35.2% 和 90.2%，相较于原算法分别提升了 1.5%和 7.3%。通过山地场景和海面场景的检测结果可知，算法在检测精度低的场景中精度提升较明显。

在天空场景下，IRSDD-YOLOv8 的 m_{AP}@0.5:0.95 和 m_{AP}@0.5 分别达到 60.6%和 96.5%，相较于 YOLOv8n-seg 算法分别高出 1.5%和 0.8%。由于天空背景比较简单，大多数检测算

法都取得了优异的检测结果，但含有的实时性明显不如本文所提出的方法。

综合上述分析可知，通过多尺度特征融合整合来自不同层的特征图，捕获到目标不同尺度的信息，能够提高算法对目标的检测能力。对于山地和海面这两个检测精度相对较低的场景，在略微牺牲检测实时性的前提下，本文所提出的IRSDD-YOLOv8 算法对精度的提升效果较明显。对于城市和天空这两个检测精度已经较高的场景，本文所使用的模型架构及所提出的多尺度特征学习方法可能在理论上已经达到其极限，模型所能提取和利用的特征已经非常充分，因此检测精度提升比例较低，进一步提升检测精度需要对模型架构作改进，但在实际探测无人机的任务中，所提算法的检测精度已经能够满足检测需求。

表 4　在不同场景下主流检测算法的检测结果

Tab.4 Detection results of mainstream detection algorithms in different scenarios

算法评价指标	检测算法	城市场景	山地场景	海面场景	天空场景
m_{AP}@0.5:0.95	BoxInst	0.197	--	--	0.395
	CondInst	0.565	0.284	0.292	0.673
	Mask-Rcnn	0.629	0.416	0.463	0.711
	Yolact++	0.423	0.177	0.163	0.561
	YOLOv7	0.440	0.269	0.335	0.580
	YOLOv5n-seg	0.473	0.245	0.342	0.572
	YOLOv8n-seg	0.477	0.253	0.337	0.591
	IRSDD-YOLOv8	**0.511**	**0.309**	**0.352**	**0.606**
m_{AP}@0.5	BoxInst	0.538	0.013	--	0.806
	CondInst	0.936	0.731	0.819	0.977
	Mask-Rcnn	0.937	0.749	0.933	0.987
	Yolact++	0.902	0.625	0.445	0.958
	YOLOv7	0.877	0.746	0.930	0.974
	YOLOv5n-seg	0.936	0.689	0.886	0.966
	YOLOv8n-seg	0.927	0.664	0.829	0.957
	IRSDD-YOLOv8	**0.939**	**0.785**	**0.902**	**0.965**
FPS	BoxInst	9.60	10.26	8.97	9.06
	CondInst	9.60	10.35	8.86	8.91
	Mask-Rcnn	3.97	4.04	4.04	4.03
	Yolact++	10.77	11.93	9.42	9.74
	YOLOv7	16.28	20.79	13.82	14.61
	YOLOv5n-seg	35.55	49.73	25.32	26.01
	YOLOv8n-seg	28.07	45.48	22.08	21.44
	IRSDD-YOLOv8	**29.01**	**38.06**	**19.32**	**22.50**

（2）定性实验结果

图 5 和图 6 展示了在 SIDD 数据集的四个不同场景中使用 IRSDD-YOLOv8 和其他方法获得的定性结果。从上到下依次是城市场景、山地场景，海面场景，天空场景，从左到右依次是不同检测方法的检测结果。由于目标在图像中的比例较小，因此以目标掩码图显示检测结果，算法检测正确的结果用红色圈标记，检测错误的结果用黄色圈标记，未检测到目标则不标记。

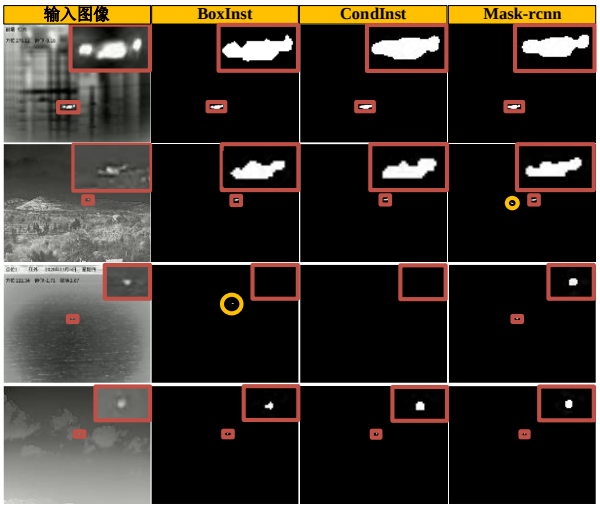


图 5 原始图像及其他算法检测结果

Fig.5 Original image and other algorithmic detection results

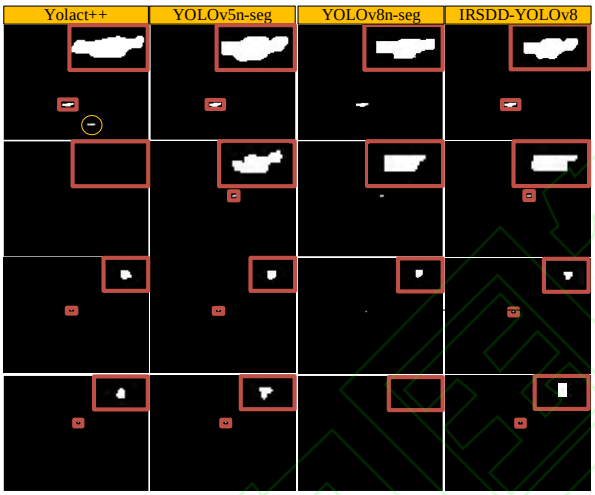


图 6 IRSDD-YOLOv8 及其他算法检测结果

Fig.6 IRSDD-YOLOv8 and other algorithms detection results

在城市场景中 Yolact++算法，存在错误检测目标的情况。在山地场景中，Mask-Rcnn 算法存在错误检测目标的情况，Yolact++算法未能有效检测到目标；在海面场景中，BoxInst 算法存在错误检测目标的情况，CondInst 算法未能有效检测到目标，在天空场景中，YOLOv8n-seg 算法未能有效检测到目标。相比其他算法，本文提出的 IRSDD-YOLOv8 算法在上述场景中都得到了准确的检测结果，表明本文提出的 IRSDD-YOLOv8 算法在复杂背景下对红外无人机目标的检测性能良好。

从定量与定性结果来看，本文所提出的 IRSDD-YOLOv8 算法的检测精度优于 YOLOv8n-seg 算法，IRSDD-YOLOv8 算法对这些场景变化的鲁棒性更强，检测结果更准确，IRSDD-YOLOv8 算法仅增加了少量的计算量就能带来大幅的性能提升，符合实际的检测需求。

4.2 消融实验

4.2.1 多尺度特征学习模块的消融实验

(1) 定量结果

在本节中，比较了多尺度特征学习模块中不同数量 C2f 模块的性能，共进行了 6 次实验。表 5 显示了在不同层中设置的不同 n_1 和 n_2 值的 m_{AP} 值。其中， n_1 和 n_2 分别代表在第 18 层和第 21 层的 C2f 数量。

如表 5 所示，从 exp(1)到(6)，多尺度特征学习模块中 C2f 模块的数量从 1 到 3 不等，随着 C2f 模块数量的增加，SIDD 数据集的 m_{AP}@50 值呈上升趋势，在 exp(6)中达到 75.8%，相较于 exp(1)的[n_1, n2]取[1,1]高出了 1.3%。对于 exp(4)至(6)，由于增加了更多的 C2f 模块，算法对目标的检测能力更强。因此，我们将 IRSDD-YOLOv8 网络结构中的第 18、 21 层的 C2f 的模块数量分别取为 3、3。

表 5 在不同层中添加 C2f 数量的实验结果

Tab.5 Experimental results of adding C2f quantities in different layers

实验组别	不同数量的 C2f [n_1, n_2]	m_{AP}@ 0.5:0.95	m_{AP}@ 0.5
exp(1)	[1,1]	0.281	0.745
exp(2)	[1,2]	0.270	0.733
exp(3)	[1,3]	0.276	0.751
exp(4)	[2,1]	0.278	0.742
exp(5)	[3,2]	0.286	0.744
exp(6)	**[3,3]**	**0.282**	**0.758**

(2) 定性结果

图 7 展示了经过多尺度特征学习模块处理后的浅层与中层特征图目标特征提取结果。浅层特征图为模型第 2 层的输出，属于早期卷积层，该层通常提取的是图像的边缘、纹理等低级特征。由于靠近输入层，该层的特征图包含大量的原始图像信息，信息还不够抽象，不足以明确地识别出无人机这样的特定目标。中层特征图为模型第 15 层的输出，该层处于网络的较深位置，能提取到比浅层更复杂的特征，还没有丢失过多细节，但是该层特征图并没有将目标特征与背景区分出来。将浅层特征图与中层特征图送入多尺度特征学习模块后，输出的特征图能够较好地包含无人

机目标的形状、结构特征，使得无人机目标在特征图中的表现更加突出。

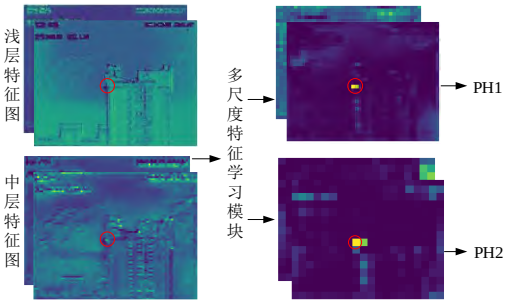


图 7 多尺度特征学习模块的目标特征提取示例

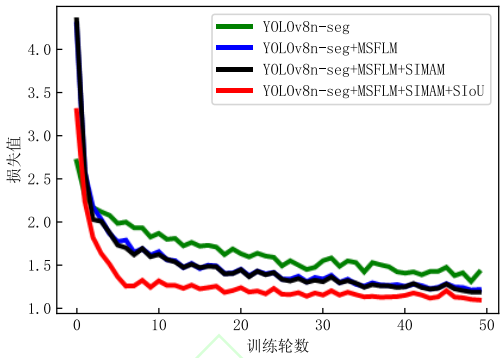
Fig.7 Example of feature extraction from the multi-scale feature learning module

4.2.2 添加不同子模块的消融实验

为了验证多尺度特征学习模块（Multi-scale feature learning module, MSFLM）、三维无参注意力模块(SimAM)和优化损失函数（SIoU）的有效性，我们做了相应的消融实验。为了保证实验结果的可比性，在训练过程中设置相同的迭代轮数和初始学习率，并记录了模型的分割损失变化以及添加不同模块后 PR 曲线的变化。

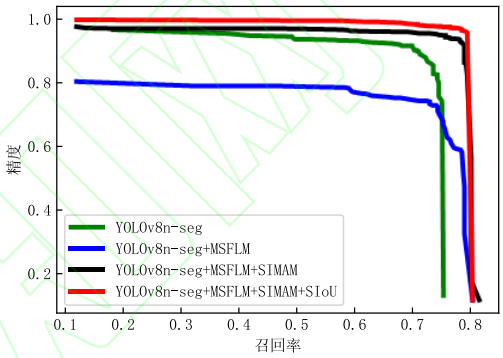
表 6 展示了在 YOLOv8n-seg 算法中添加各个模块的实验结果。由实验结果可知，增加子模块后，模型的检测精度都会出现一定程度的提升，其中，添加多尺度特征学习模块对检测精度的提升最明显，检测精度提升了 9.4%(m_{AP}@0.5)。

如图 8（a）所示，IRSDD-YOLOv8 在模型训练过程中具有更好的收敛性。图 8(b)显示了消融实验中添加不同模块后模型的 PR 曲线。PR 曲线的结果表明，基于多尺度改进的 IRSDD-YOLOv8 算法的 PR 曲线与坐标轴的面积越大，说明性能更优于 YOLOv8n-seg 算法。



(a) 损失曲线

(a) Loss curve



(b) PR 曲线

(b) PR curve

图 8 消融实验的分割损失与 PR 曲线

Fig.8 Segmentation loss and PR curves for ablation experiments

表 6 添加不同子模块的实验结果（山地场景）

Tab.6 Experimental results of adding different submodules (mountain scenario)

实验组别	MSFLM	SimAM	SIoU	m_{AP}@0.5:0.95	m_{AP}@0.5	Paras(M)
消融实验 1				0.253	0.664	9.6
消融实验 2	√			0.282	0.758	11.8
消融实验 3	√	√		0.270	0.763	11.8
消融实验 4	√	√	√	**0.309**	**0.785**	**11.8**

4.3 实物实验

为了验证所提算法在实际环境中探测无人机目标的性能，本文在户外试验场搭建了一套低空无人机目标检测系统，数据采集平台如图 9 所示。使用 Q30TIRM 双光吊舱的红外相机作为红外图像采集器，其成像像元间距为 17um，镜头焦距为 25mm，对应的视场角为 24.6 °×18.5 °，最大变焦倍数为 4。假设探测的无人机目标有效热辐射面积为 200 mm×200 mm，无人机的运动方向与红外相机平面平行，理论上红外相机发现无人机目标的极限距离可达 1452m。

图 9 数据采集平台

Fig.9 Data Acquisition Platform

在实采数据集上对提出的 IRSDD-YOLOv8 算法进行了实验验证，并与其他性能较好的单阶段目标检测算法进行了对比。实验结果如表 7 和图 10 所示，本文提出的 IRSDD-YOLOv8 算法在实采数据集上的检测精度达到了 94.0%(m_{AP}@0.5)，在三个精度指标上的表现均优于其它单阶段目标检测算法。

表 7 算法在实际数据集上的检测精度对比

Tab.7 Comparison of detection accuracy of algorithms on the actual collection dataset

检测算法	m_{AP}@0.5	m_{AP}@0.5:0.95
YOLOv5n-seg	0.832	0.341
YOLOv8n-seg	0.922	0.373
IRSDD-YOLOv8	0.940	0.425

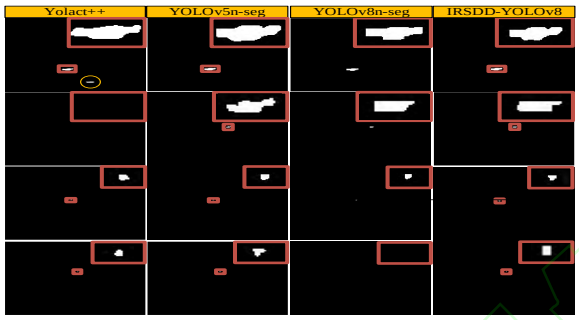


图 10 实采数据集的检测结果示例

Fig.10 Example of test results for the actual collection dataset

5 结论

本文针对在红外图像中检测无人机目标存在尺寸小、纹理特征弱的问题，提出了多尺度学习的目标检测算法 IRSDD-YOLOv8，构建了一个新的 SIDD 数据集，在数据集中划分了四种典型的无人机入侵场景。

对 IRSDD-YOLOv8 算法与经典的红外小目标检测算法、主流检测算法在自制的 SIDD 数据集上进行了大量的实验。实验结果表明，在城市、山地、海面、天空四个场景下，IRSDD-YOLOv8 算法的 μ_{IoU} 值达到了 82.4% 、75.2% 、66.5%、83.9%， 相较于原始 Yolov8n-seg 算法分别高出 1.2% 、2.4%、 1.6% 、0.2%；m_{AP}@0.5 值分别达到了 93.9%、78.5%、90.2%、96.5%，检测精度分别提高了 1.2%、12.1%、7.3%、0.8%。

此外，还搭建了一套低空无人机目标检测系统，采集红外无人机目标数据进行验证，验证结果表明，IRSDD-YOLOv8 算法在实采数据集上的检测精度(m_{AP}@0.5)达到了 94.0%，可以满足实际的探测需求。

参考文献

[1] ZHAI H, ZHANG Y. Target Detection of Low-Altitude UAV Based on Improved YOLOv3 Network[J]. Journal of Robotics, 2022, 2022: 4065734.

[2] WANG M, ZHANG B. Contrastive Learning and Similarity Feature Fusion for UAV Image Target Detection[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.

[3] SUNKARA R, LUO T. YOGA: Deep object detection in the wild with lightweight feature learning and multiscale attention[J]. Pattern Recognition, 2023, 139: 109451.

[4] XU C, ZHANG Q, MEI L Y, et al. Dense Multiscale Feature Learning Transformer Embedding Cross-Shaped Attention for Road Damage Detection[J]. Electronics, 2023, 12(4): 898.

[5] YANG L, ZHONG J, ZHANG Y, et al. An Improving Faster-RCNN With Multi-Attention ResNet for Small Target Detection in Intelligent Autonomous Transport With 6G[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(7): 7717-7725.

[6] WANG K X, LIEW J H, ZOU Y T, et al. PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment[C] //Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2019: 9196-9205.

[7] GHIASI G, LIN T-Y, LE Q V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection[C] //Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019: 7029-7038.

[8] ZHAO Q, SHENG T, WANG Y, et al. M2Det: A Single-Shot Object Detector Based on Multi-Level Feature Pyramid Network[C] //Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, 2019: 9259–9266.

[9] CHUANG Y, LIU Y, WU S, et al. Infrared small target detection based on multiscale local contrast learning networks[J]. Infrared Physics & Technology, 2022, 123: 104107.

[10] 李向荣，孙立辉. 融合注意力机制的多尺度红外目标检测[J].红外技术, 2023, 45(7): 746~754.

[11] 张朝阳，张上，王恒涛等.多尺度下遥感小目标多头注意力检测[J].计算机工程与应用，2023，59(8): 227~238.

[12] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C] //34th AAAI Conference on Artificial Intelligence. New York in America: AAAI Press, 2020.

[13] GEVORGYAN Z. SIoU Loss: More Powerful Learning for Bounding Box Regression [J]. arXiv2022, arXiv: 2205.12740.

[14] 回丙伟，宋志勇，范红旗等.地/空背景下红外图像弱

小飞机目标检测跟踪数据集[J].中国科学数据, 2020, 5(3): 291~302.

[15] BAI X Z, ZHOU F G. Analysis of new top-hat transformation and the application for infrared dim small target detection[J]. Pattern Recognition, 2010, 43(6): 2145-2156.

[16] WEY Y T, YOU X G, Li H. Multiscale patch-based contrast measure for small infrared target detection[J]. Pattern Recognition, 2016, 58: 216-226.

[17] WU X, HONG D F. UIU-Net: U-Net in U-Net for Infrared Small Object Detection[J]. IEEE Transactions on Image Processing, 2023, 32: 364-376.

[18] LI B Y, XIAO C, WANG L G, et al. Dense Nested Attention Network for Infrared Small Target Detection[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 2023, 32: 1745-1758.

[19] HOU Q Y, ZHANG L W, TAN F J, et al. ISTDU-Net: Infrared Small-Target Detection U-Net[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[20] TIAN Z, SHEN C, WANG X, et al. BoxInst: High-Performance Instance Segmentation with Box Annotations [C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 5439–5448.

[21] TIAN Z, SHEN C, CHEN H. Conditional Convolutions for Instance Segmentation[J]. Lecture Notes in Computer Science, 2020, 12346: 282-298.

[22] BOLYA D, ZHOU C, XIAO F, et al. YOLACT++ Better Real-Time Instance Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1108-1121.

[23] WANG X, ZHANG R, KONG T. SOLOv2: Dynamic, Faster and Stronger [J]. arXiv2020, arXiv: 2003.10152.

[24] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 386-397.

[25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors[C]. //2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 7464-7475.