

# 多通道交互注意与轮廓增强的红外无人机检测

聂苏珍<sup>1</sup>, 曹杰<sup>2</sup>, 郝群<sup>2\*</sup>, 庄须叶<sup>1\*</sup>

(1. 山东理工大学机械工程学院, 山东淄博 255000;

2. 北京理工大学光电学院, 北京 100081)

**摘要:** 无人机因其小巧、轻便、灵活的特点, 在农业、物流、救援、赈灾等方面有着广泛的应用。然而如果使用不当或管理不善, 不仅会造成个人隐私泄露、财产损失, 还可能对公共安全甚至军事安全构成威胁。因此, 实时准确地对空域内的无人机进行检测与预警具有重要作用。对此, 提出了一种用于红外无人机检测的多通道交互注意力和边缘轮廓增强的方法(MCIAECE)。首先, 通过构建多通道交互注意力机制模块和边缘轮廓增强模块组成的双通道对红外图像的浅层和深层特征进行提取, 经过注意力机制可以增强目标特征而边缘轮廓增强则可以获得更多细节信息。然后使用多级特征融合模块将所提取的各层特征进行融合增强, 从而获得检测结果。实验结果表明, 在三个数据集上用多通道交互注意力和边缘轮廓增强的方法都能够达到较好的效果。其中在NUDT-SIRST红外数据集上效果最佳, 检测概率和交并比分别为98.83%和85.11%, 与基线网络相比提高了1.95%和6.88%, 与其他方法相比, 在目标的边缘轮廓还原方面效果显著。

**关键词:** 多通道交互注意力; 边缘轮廓增强; 多级特征融合; 红外无人机检测

中图分类号: TH741

文献标识码: A

## Infrared UAV detection based on multi-channel interactive attention mechanism and edge contour enhancement

NIE Su-Zhen<sup>1</sup>, CAO Jie<sup>2</sup>, HAO Qun<sup>1\*</sup>, ZHUANG Xu-Ye<sup>2\*</sup>,

(1. School of Mechanical Engineering, Shandong University of Technology, Zibo 255000, China;

2. School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** UAVs have a wide range of applications in agriculture, logistics, rescue and disaster relief because of their compactness, lightness and flexibility. However, if they are used improperly or mismanaged, they may not only cause personal privacy leakage and property loss, but also pose a threat to public safety and even military security. Therefore, real-time and accurate detection and warning of UAVs in the airspace plays an important role. In this regard, a multi-channel interactive attention and edge contour enhancement (MCIAECE) method for infrared UAV detection is proposed. Firstly, the shallow and deep features of the infrared image are extracted by a dual-channel consisting of a multi-channel interactive attention mechanism module and an edge contour enhancement module, after which the attention mechanism enhances the target features while the edge contour enhancement obtains more detailed information. Then the extracted features of each layer are fused and enhanced using the multilevel feature fusion module to obtain the detection results. The experimental results show that better results can be achieved with multi-channel interactive attention and edge contour enhancement on all three datasets. Among them, the best results are obtained on the NUDT-SIRST infrared dataset, with the detection probability and intersection over union of 98.83% and 85.11% respectively, and the effect is significant in the edge contour restoration of the target compared with other methods.

**Key words:** multi-channel interactive attention, edge contour enhancement, multi-level feature fusion, Infrared drone detection

收稿日期: 2024-09-09, 修回日期: 2024-11-16

Received date: 2024-09-09, revised date: 2024-11-16

基金项目: 山东省自然科学基金(ZR2021MF042); 山东省泰山学者项目(TSQN201909108)

Foundation items: Supported by the Natural Science Foundation of Shandong Province (ZR2021MF042); Taishan Scholars Program of Shandong Province (TSQN201909108)

作者简介(Biography): 聂苏珍(1999—), 女(汉), 江西省宜春人, 硕士, 主要研究领域为红外小目标检测. E-mail: szhen0203@163.com

\* 通讯作者(Corresponding authors): E-mail: qhao@bit.edu.cn; zxye@sdut.edu.cn

## 引言

随着科技的进步,无人机为人们的日常生活带来了极大的便利。小型无人机具有体积小、便于携带、效率高等特点,在民用和军用领域都变得越来越普遍<sup>[1]</sup>。但是无人机“黑飞”“乱飞”会对社会公共安全和军事安全带来威胁。如无人机闯入居民隐私区域进行拍摄,飞入机场“净空区”对飞机航行造成干扰甚至拍摄窃取国家军事机密等事件层出不穷。因此,快速而又精准的检测无人机并对其进行严格管控显得尤为重要。目前,常用的检测技术通常包括基于可见光图像探测、声波探测和雷达探测等手段<sup>[2]</sup>。由于无人机体积小、速度慢,且主要采用非金属材料制造,雷达的反射截面小,因此使用雷达探测器检测无人机效果不佳。而使用声波和电磁信号检测无人机稳定性差,使其在信号干扰较多的城市场景中检测能力大幅下降。使用可见光来进行检测在大多数情况下效果较好,也是目前比较常用的方式,但是在夜晚、雾霾等可视度差的环境中检测具有局限性<sup>[3]</sup>。而红外探测器具有成像距离远、灵敏度高、隐蔽性好、抗干扰能力强、全天候探测等优点,因此采用红外探测器来对无人机进行检测在实际应用中具有独特的优势。红外图像检测在军事侦察、农业预警和民用监测等方面有着广泛的应用。与可见光相比,红外图像中的目标缺乏颜色纹理等特征且与背景对比度低,因此在红外拍摄的图像中检测无人机具有一定困难。目前,目标检测网络主要以 Faster-RCNN<sup>[4]</sup>、YOLO<sup>[5]</sup>、SSD<sup>[6]</sup>等为代表。由于它们在各种数据集上都有良好的检测效果,因此许多检测无人机的网络都是在这些基础检测网络结构上进行改进的。田紫薇等<sup>[7]</sup>提出一种在 YOLOv5 框架上添加通道空间注意力机制的微型无人机检测方法。通过优化锚点框的大小、嵌入卷积块注意力模块(Convolutional Block Attention Module, CBAM)和优化损失函数(Complete Intersection over Union, CIoU),提高了原有算法在复杂背景下对无人机的检测性能。梁晓等<sup>[8]</sup>通过引入注意力机制,加强对目标区域的表征,提高图像的空间信息量。张灵灵等<sup>[9]</sup>融合高层语义信息和浅层细节信息,并引入通道注意力机制,加强了目标在通道层次的特征表达能力。然而,上述工作都是基于可见光摄像机采集的图像进行训练和测试,对红外图像中的无人机目标检测效果并不好。随着红外小目标探测技术的不断发展,社会各界对于解决这

一应用领域的挑战表现出了浓厚的兴趣,也最终促进了该技术在资源受限环境下的实际应用<sup>[10]</sup>。

因此,针对红外目标检测也提出了许多方法,如林再平等<sup>[11]</sup>通过在编码层和解码层之间进行高频多尺度特征交互,并在编码器瓶颈处级联轻量型混合注意力模块,进一步增强网络深层的目标特征。姚迎乐等<sup>[12]</sup>构建红外过采样扫描模型,对图像背景以及噪声进行滤除,并设计特征融合模块和孪生网络提高图像特征表示能力。湛海云等<sup>[13]</sup>提出了以 YOLOX 为主干网络的红外目标检测算法,并引入改进的路径特征融合模块和空间通道混合注意力机制加强特征提取。陈广秋等<sup>[14]</sup>将红外与可见光进行融合得到显著的红外目标,提取多尺度特征后利用注意力残差密集融合网络对多个尺度特征进行融合,获取更多目标信息。刘婷等<sup>[15]</sup>提出了一种带因子先验的非凸张量 TD 模型,结合因子先验和 logdet 函数,优化了红外目标检测。同时,对小的因子矩阵进行奇异值分解(Singular Value Decomposition, SVD)计算降低计算复杂度。最后采用组稀疏正则化增强背景杂波抑制,并通过交替方向乘法(Alternation Direction Method of Multipliers, ADMM)高效求解,展现出其在复杂场景下的优越性能。但是,在红外无人机图像检测中,这些算法除了要能够对无人机目标定位准确外还需要能够利用有限资源易于部署。因此李博扬等<sup>[16]</sup>首次提出了一种基于单点监督的 SIRST 检测方法,通过聚类从单点标签恢复每个目标的像素掩码。为应对目标模糊和背景杂波,在聚类中引入随机性,添加噪声并平均结果以获得可靠的伪掩码,从而减少对大量标注数据的依赖,降低了训练成本。此外李博扬等<sup>[17]</sup>还提出了一种简单而有效的对称渐进递减混合精度量化(SPMix-Q)方法,以在低比特量化下实现高性能分割。因此,根据红外图像中无人机的特征及实际情况,本文基于 U-Net<sup>[18]</sup>网络进行改进,将目标检测视为目标分割问题。同时引入一种多通道交互注意力模块,通过把输入通道分成多个部分来进行不同的处理,使得模型能够更加全面和细致地捕捉到不同维度上的特征信息,在不增加网络模型复杂度的同时使网络能够聚焦于有用的目标信息,并抑制复杂背景等无用信息,然后使用边缘轮廓增强模块,得到丰富特征图多尺度边缘轮廓和深层语义信息,最后送入多尺度特征融合模块与多层特征级联进一步提高对目标的检测准确率。

## 1 原理

### 1.1 U-Net基本原理

U-Net 因其结构简单、易于训练等优点成为人们进行语义分割任务广泛使用的算法之一。它主要由编码器和解码器组成且为对称分布。同时引入复制和裁剪跳跃连接来进行特征传递与融合。

U-Net 详细网络结构图如图 1 所示。从图 1 可以看到,Unet 网络的结构比较简单,左侧编码器分支主要由卷积块和最大池化层组成。卷积块包含两层卷积核大小为 3 的卷积层,它们能够对图像进行特征提取。然后通过最大池化层进行下采样,得到尺寸大小减半,通道数加倍的特征图。右侧解码器分支与左侧编码器相同,一样由四个卷积块组成。首先通过反卷积把特征图的大小还原为原来的两倍,同时通道数变为原来二分之一,再和左侧编码器卷积层输出的特征图进行合并。由于左侧编码器和还原后的右侧特征图的大小并不一致,因此需要将左侧卷积层的特征图复制裁剪为和右侧特征图的大小一致再进行特征拼接。

U-Net 最开始用于语义分割任务,但由于其简单高效的 U 形结构,越来越多的人在目标检测当中使用它,同时也启发了后面很多算法。如果仅用 U-Net 网络对红外无人机数据集进行检测,存在以下几个问题:

(1)网络结构简单,因此对复杂图像检测效果不够好;

(2)图像越大需要的池化层越多,而特征在多层池化后容易产生特征不明显、细节丢失等问题,并且需要较大的内存和计算资源。

(3)在进行图像分割时,可能会出现一些边界模糊或者分割错误的情况。

### 1.2 本文方法

为了让 UNet 网络适用于红外图像中无人机的目标检测,在原网络中增加多通道注意力和边缘轮廓增强来对图像进行特征提取。首先,利用多通道交互注意力模块来获得不同深度特征图,使每层特征图之间具有上下文信息。然后利用边缘轮廓增强模块提取每层特征的细节信息,从而达到增强目标边缘轮廓的目的,最后将经过多通道交互注意力之后的特征图和经过边缘轮廓增强模块的特征图一同输入到多级特征融合模块中,得到融合增强后的特征图。网络结构如图 2 所示,其中 ECE(Edge Contour Enhancement)为提出的一种边缘轮廓增强模块,MCIA (Multi-Channel Interactive Attention)为提出的一种多通道交互注意力模块,MLFF(Multi-Level Feature Fusion)为多级特征融合模块。

#### 1.2.1 多通道交互注意力模块

仅使用卷积对图像进行局部特征提取,容易丢失目标细节信息。在深度学习中,特征图的关联性对于模型的性能至关重要,尤其是在复杂的背景下进行红外图像目标检测时。特征图的目标与背景关联性不足会导致模型在特征提取过程中,不能有效地捕捉到目标与其背景之间的空间关系和上下

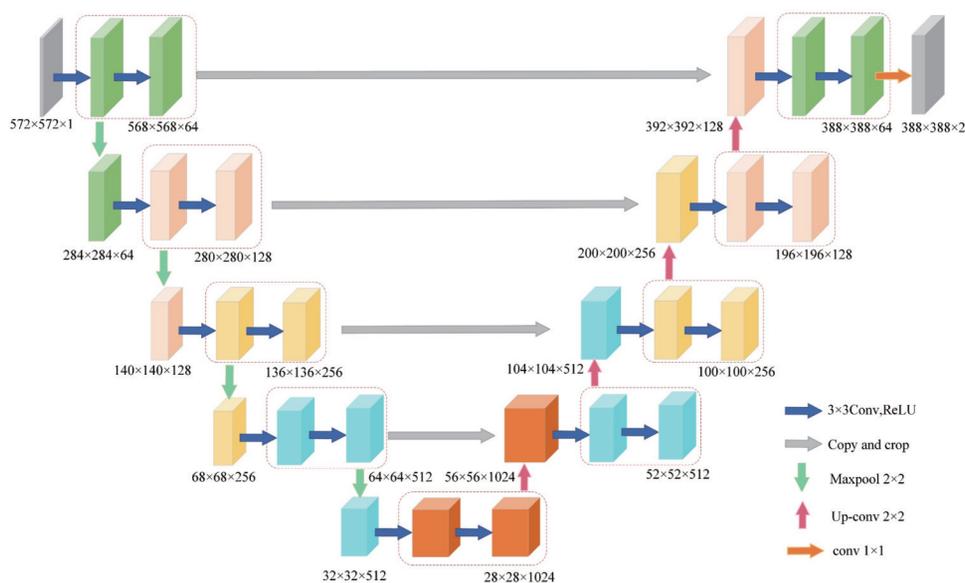


图 1 UNet网络结构

Fig. 1 The structure of UNet

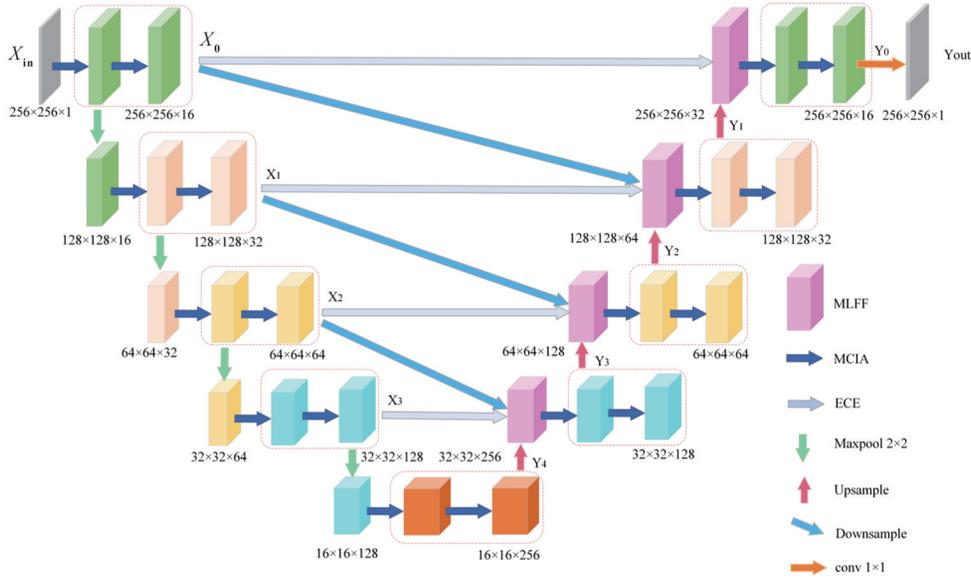


图2 MCIAECE网络结构图

Fig. 2 The structure of the MCIAECE network

文信息,导致模型难以区分目标和背景,从而影响检测的准确性。因此,为了让网络能够自动学习特征图中目标像素的重要性和与背景的相关性,引入了一种多通道注意力机制,如图3所示。图中输入特征表示为  $F_{in} \in \mathbb{R}^{H \times W \times C}$ , 经过两个卷积块获得局部特征  $F_1 \in \mathbb{R}^{H \times W \times C}$ 。随后对  $F_1$  进行分组操作,在通道维度上分成  $g$  组,从而使空间语义特征在每个特征组内得到良好分布。此时得到子特征图  $F_2 \in \mathbb{R}^{H \times W \times C}$ , 通道数变为  $C_g$  (表示为  $F_1$  通道数的  $g$  分之一)。为捕获跨维度的交互作用和建立维度间的依赖关系,提高模型对长距离依赖关系的建模能力,增强特征提取的效果。对  $F_2$  进行多分支交互操作,其中分支一首先用最大池化和平均池化对  $F_2$  进行并行处理,然后将得到的特征图相加后用 Sigmoid 函数激活再与原  $F_2$  相乘,从而得到第一个空间注意力图。分支二则对  $F_2$  进行  $3 \times 3$  卷积操作后导出第二幅空间注意力图。然后让两个空间注意力图都进行两种操作,一种是不操作直接输出,另一种则进行 Softmax 函数和平均池化操作。最终得到四个特征图  $F_{11} \in \mathbb{R}^{H \times W \times C}$ ,  $F_{12} \in \mathbb{R}^{H \times W \times C}$ ,  $F_{21} \in \mathbb{R}^{H \times W \times C}$ ,  $F_{22} \in \mathbb{R}^{H \times W \times C}$ 。其中  $F_{11}$ ,  $F_{12}$  两个特征图专注于捕捉目标的细节特征,它们通过较小的卷积核(如  $1 \times 1$  卷积)来提取目标的边缘和形状信息,这些信息对于识别目标至关重要。而  $F_{21}$  和  $F_{22}$  这两个特征图则更关注背景信息,通过较大的卷积核(如  $3 \times 3$  卷积)来捕捉背景的上下文信息。四个特征图的设计使其能够在保持模型效率的同时,提供足够的信息来增

强目标与背景之间的关联性。最后,将每组内两两生成的  $F_{12}$ ,  $F_{21}$  和  $F_{11}$ ,  $F_{22}$  进行跨通道交叉相乘后再相加得到聚合权重值。让其经 Sigmoid 函数后再与原  $F_2$  逐像素相乘得到每个像素的重要程度和相关性。再与经过  $1 \times 1$  卷积进行通道升维后的输入特征图相加得到最终的特征图  $F_{out}$ , 详细的计算过程如公式(1)所示。

$$F_{11} = \text{GN}\left(F_2 \times \text{sigmd}\left(\max(F_2) + \text{avg}(F_2)\right)\right), (1)$$

$$F_{12} = \text{soft}\left(\text{avgpl}(F_{11})\right), (2)$$

$$F_{21} = \text{Conv}_{3 \times 3}(F_2), (3)$$

$$F_{22} = \text{soft}\left(\text{avg}(F_{21})\right), (4)$$

$$F_{out} = \text{Conv}_{1 \times 1}\left(F_2 \odot \left(\text{sigmd}(F_{11} \otimes F_{22} + F_{12} \otimes F_{21})\right)\right), (5)$$

其中,  $\max$  和  $\text{avg}$  表示最大池化层和平均池化层,  $\text{sigmd}$  和  $\text{soft}$  分别表示 Sigmoid 和 Softmax 函数,  $\text{GN}$  表示组归一化操作。多通道交互注意力模块通过捕获像素级关系,重点突出所有像素的全局上下文信息,从而得到整个精确的空间位置信息。

### 1.2.2 边缘细节增强模块

由于 U-Net 左右两边对应的卷积层输出的特征图大小并不一致,需要将左侧输出的特征图裁剪为与右侧特征图相同大小后再进行拼接。这在一定程度上会使得特征图细节丢失,或重要像素被裁剪造成缺失,使得分割目标边缘轮廓模糊。为了获得边缘轮廓清晰的特征图,本文提出边缘细节增强模

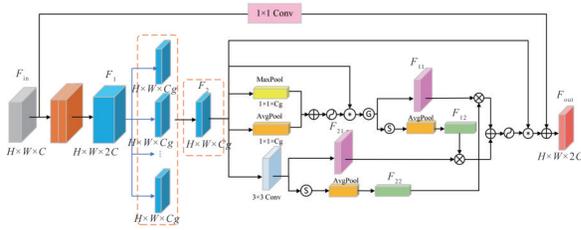


图3 MCIA 模块结构图

Fig. 3 The structure of the MCIA module

块代替 UNet 中的复制裁剪跳跃连接。其详细结构如图 4 所示。首先在通道维度上对特征图  $X_i \in \mathbb{R}^{H \times W \times C}$  进行分割,共分为四部分。为了获得更加丰富的特征信息,对每个部分的特征图使用不同扩张速率的空洞卷积。这是因为相较于普通卷积,空洞卷积采用的是具有空洞的卷积核,因此可以得到不同大小感受野的特征图,从而能够对物体和背景之间的差异进行更详细的建模,增强其检测能力。然后再将得到的所有特征图进行融合,最后得到边缘轮廓增强的特征图  $X'_i \in \mathbb{R}^{H \times W \times C}$ ,详细计算如公式(6)所示。

$$X'_i = \text{concat}(\text{pw}(D_i(X_i))) \quad (6)$$

其中,concat 表示通道维度的特征拼接,pw 表示  $1 \times 1$  逐点卷积, $D_i$  表示卷积速率为  $i(i=1, 2, 3, 4)$  的空洞卷积。边缘特征增强模块通过将通道分割后交错排列来增强多尺度特征的多样性,同时使用逐点卷积进行组间和跨组信息融合,实现轻量高效的效果。

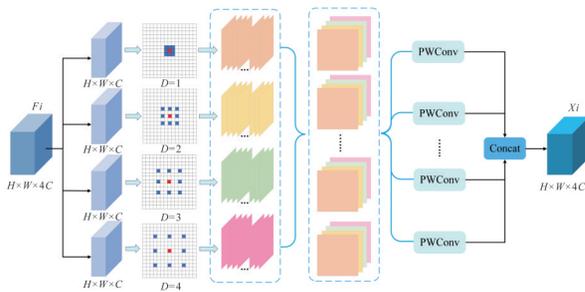


图4 ECE 模块结构图

Fig. 4 The structure of the ECE module

### 1.2.3 多级特征融合模块

卷积神经网络的浅层图像具有更高的分辨率和更多的纹理细节;深层图像则分辨率低,但蕴含更加丰富的语义信息。然而,深层特征和浅层特征之间的相互关联性并没有充分的体现。因此,如图 5 所示,本文引入多级特征融合模块,将第  $i$  层( $i=0,$

1, 2, 3) 特征  $X_i \in \mathbb{R}^{H \times W \times C}$  和相应的下层特征  $Y_{i+1} \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 2C}$  与上层特征  $X_{i-1} \in \mathbb{R}^{2H \times 2W \times \frac{C}{2}}$  进行融合,从而得到增强的目标,抑制背景噪声。详细计算如公式(7)所示。

$$Y_i = \text{MCIA}(\text{concat}[\text{down}(X_{i-1}), X_i, \text{up}(Y_{i+1})]) \quad (7)$$

其中 MCIA 表示多通道交互注意力模块,down 表示下采样操作,up 表示上采样操作,concat 表示特征拼接。通过上采样操作和下采样操作得到与当前层相同大小的特征图,从而进行特征拼接得到增强表示后的特征图。

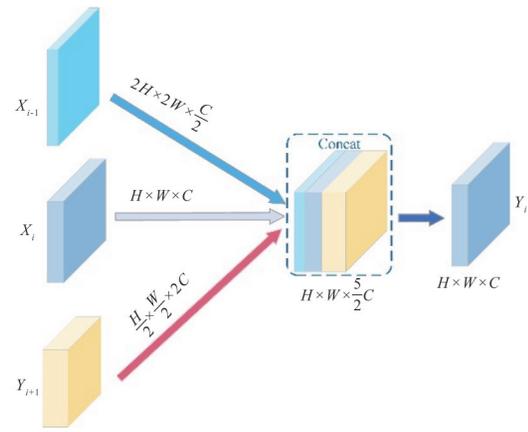


图5 多级特征融合模块结构图

Fig. 5 The structure of the MLFF module

## 2 实验分析

### 2.1 实验细节

#### 2.1.1 数据集

虽然本文方法主要针对红外无人机目标检测,但在多种目标场景的数据集上进行训练验证能够证明模型的泛化性能。因此选择在 NUDT-SIRST 数据集<sup>[9]</sup>上训练,其中包含了 1327 张多目标和多场景的红外图像。并将数据集随机分成两部分,其中 50% 为训练集,另外 50% 的数据为测试集。

#### 2.1.2 实验环境

实验是在 NVIDIA GTX 1660 GPU 上使用 PyTorch 完成的。为验证 MCIAECE、在红外图像中对无人机检测的有效性,在相同运行环境及红外数据集下,将 MCIAECE 模型与其他多个红外检测方法如 ALC-Net<sup>[19]</sup>,RDIAN<sup>[20]</sup>,ACM-Net<sup>[21]</sup>,MTU-Net<sup>[22]</sup>,WPCM<sup>[23]</sup>,RIPT<sup>[24]</sup>,IPI<sup>[25]</sup>,PSTNN<sup>[26]</sup>和 Top-Hat<sup>[27]</sup>进行对比实验。在训练过程中,使用 FocalIoU<sup>[18]</sup>作为损失函数,Adagrad 作为优化器。且批量大小为 4、初始学习率为 0.05,而权重衰减则设置为  $1 \times 10^{-4}$ 。

### 2.1.3 评价指标

本文采用了 IoU、Pd 以及 Fa 作为评价指标,这些指标能够全面地反映模型的性能,包括模型对目标的检测能力、准确性以及在不同情况下的鲁棒性。以下是这些指标的详细解释:

1) 检测概率 (Pd): 表示正确预测的目标数量  $P_c$  和总目标数  $P_a$  的比率。反映了模型的检测能力,即模型是否能够准确地识别出目标。Pd 值越高,说明模型的检测能力越强。Pd 定义如下:

$$Pd = \frac{P_c}{P_a} \quad (8)$$

2) 虚警率 (Fa): 表示错误检测的目标像素数  $F_f$  和总目标像素数  $F_a$  之间的比率。反映了模型在避免误检方面的表现。Fa 值越低,说明模型的鲁棒性越好,虚警越少。Fa 定义如下:

$$Fa = \frac{F_f}{F_a} \quad (9)$$

3) 交并比 (IoU): 评估的是算法的形状描述能力,表示为检测目标与真值之间的交集和并集的比值, IoU 值越高,表示预测结果与真实情况越接近,通常 IoU 值达到 0.5 以上被认为是有效的检测。IoU 定义如下:

$$IoU = \frac{Inter}{Union} \quad (10)$$

其中 Inter 和 Union 分别表示检测目标与真值的交集和并集。

## 2.2 实验结果分析

### 2.2.1 与其他方法对比

为了验证本文所提出 MCIAECE 红外无人机检测算法的有效性,与现有的一些经典算法进行比较,结果如表 1 所示。由于传统算法大多依赖于手工选取特征,不能够根据背景变化自适应提取特征,在处理一些背景复杂的场景时性能受限。而那些仅仅基于 CNN 的算法,只关注卷积自身部分的特征,缺乏对图像全局特征的提取,导致难以进行准确的预测;此外,这些算法通常在面对复杂多变的噪声背景时,其学习判别的能力较弱。这种局限性导致它们难以有效识别和区分真实目标与干扰因素,从而极易将噪声误认为目标或者遗漏与背景相似的目标。而与这些方法相比,MCIAECE 网络在大多数的评价指标上都表现最佳,且与基线网络相比在 NUDT-SIRST 数据集上检测概率与交并比分别提高了近 2% 和 7%,而虚警率则降低了  $2.4 \times 10^{-6}$ 。这说明了本文提出的 MCIAECE 能够适应对比度不

明显等问题,且能较好地检测并获得目标轮廓,有利于对目标进行区分。

表 1 不同的方法在 NUDT-SIRST/NUAA-SIRST/IRSTD-1k 数据集上获得的 IoU、Pd、Fa 值

Table 1 IoU, Pd, Fa values obtained by different methods on NUDT-SIRST dataset

Model	NUDT(Tr=50%)	NUAA(Tr=50%)	IRSTD-1k (Tr=50%)
	Pd/ Fa/ IoU	Pd/ Fa/ IoU	Pd/ Fa/ IoU
Top-Hat <sup>[27]</sup>	78.41/166.7/ 20.72	79.84/1012/ 7.143	75.53/1346/ 8.74
IPI <sup>[25]</sup>	74.49/41.23/ 17.76	85.55/11.47/ 25.67	80.75/16.68/ 24.98
RIPT <sup>[24]</sup>	91.85/344.3/ 29.44	79.08/22.61/ 11.05	77.47/28.41/ 14.33
MPCM <sup>[23]</sup>	84.32/356.8/ 27.28	83.27/17.74/ 12.35	69.73/29.47/ 11.68
PSTNN <sup>[26]</sup>	66.13/44.17/ 22.40	77.95/29.11/ 14.85	22.40/74.15/ 54.37
ACM <sup>[21]</sup>	95.68/9.34/ 68.28	92.93/3.45/ 72.46	90.35/12.42/ 60.47
MTU- Net <sup>[22]</sup>	97.35/3.89/ 83.83	98.55/1.30/ 73.12	91.52/1.71/ 63.12
RDIAN <sup>[20]</sup>	97.98/8.49/ 78.23	98.23/1.45/ 69.71	89.06/1.34/ 62.21
ALC-Net <sup>[19]</sup>	96.51/9.26/ 81.43	92.18/37.23/ 67.84	84.36/62.12/ 60.25
MCIAECE- Net	98.83/2.09/ 85.11	98.09/1.21/ 69.89	91.64/1.08/ 61.16

为了进一步直观地看到方法的对比效果,如图 6 所示,为不同方法地可视化结果。从图中可以看到,传统方法不能够很好的区分背景与目标,当目标与背景对比度不明显时,容易出现大量虚警也就是图中黄色虚线标明的区域。而深度学习的方法相较于传统方法在这一点上有着明显区别,但大部分仍然不能够清晰地检测出目标的轮廓。但是,即使是在一些目标并不明显的红外图像中,如图 6 中 1,4 行所示,本文所提出的 MCIAECE 依旧能够对其进行精准的判断,且检测出的无人机轮廓跟真实无人机轮廓更加相似。这是因为本文所采用的边缘细节增强操作对输入图像进行的跳跃连接,能得到更多有用信息避免信息的损失;而多通道交互注意力模块则通过分别对局部和全局建模,使得网络能够提取更多细节信息并建立全局上下文联系,减少出现漏检以及误检情况。

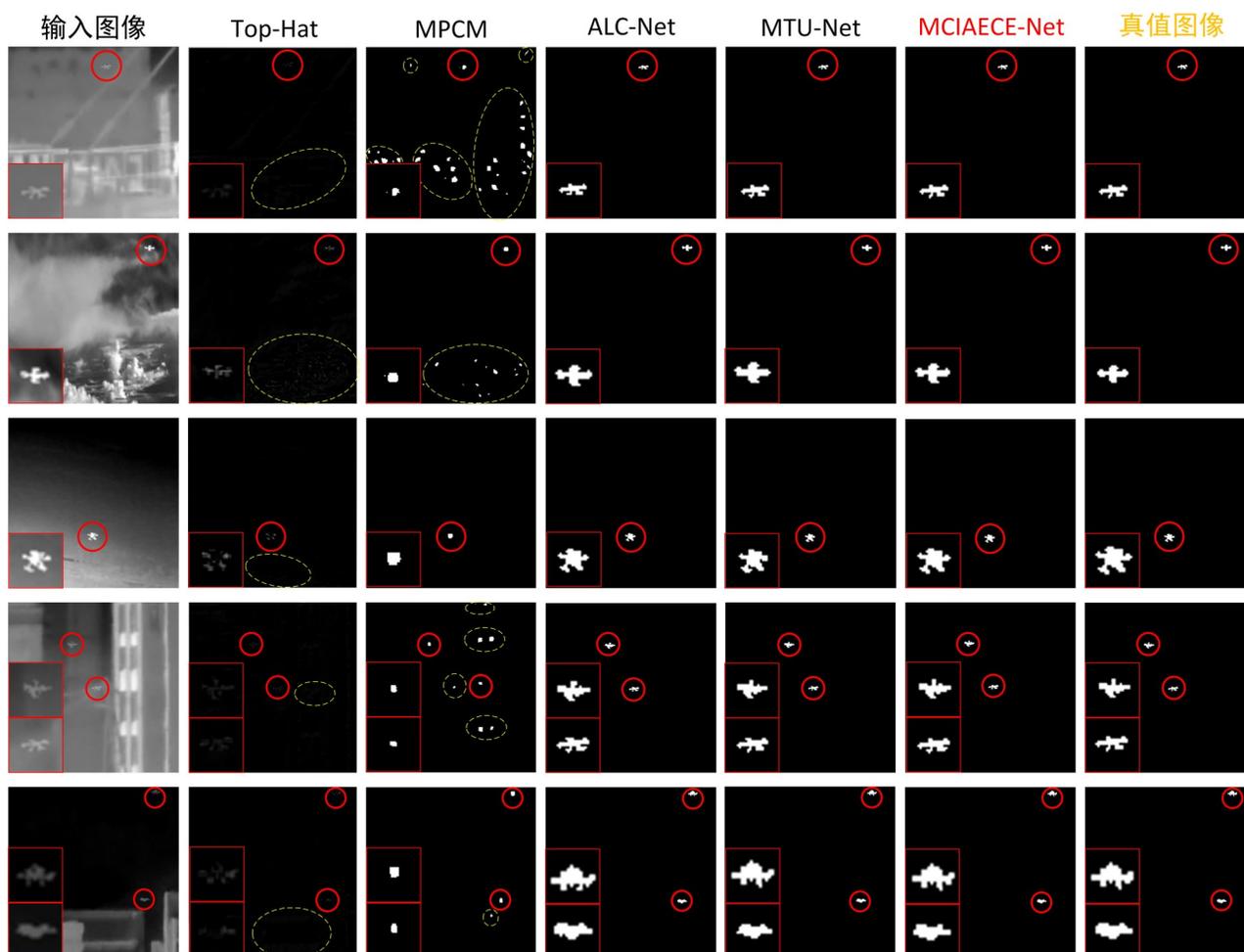


图6 不同的检测方法获得的定性结果。为了更好地可视化,目标区域在左下角被放大。正确检测到的目标、虚警区域分别用红色、黄色圆圈显示

Fig. 6 Qualitative results of different detection methods. For better visualisation, the target area is enlarged in the lower left corner. Correctly detected targets, false alarm regions are shown with red and yellow circles, respectively

在深度学习模型中,除了模型的精度之外,模型的参数量、FLOPs(浮点运算次数)和FPS也是衡量模型性能的重要指标,决定了模型在实际应用中的可行性和效率。模型参数量反映了模型的复杂度,其大小直接影响显存的使用量,GFLOPs(每秒十亿次浮点运算数)则反映了模型对处理器运算能力的需求,FPS指的是模型处理视频流或实时图像数据时,每秒能够处理的帧数。一般来说,模型参数量较少可能导致模型的代表能力受限,从而影响模型的性能,但同时可能带来较低 GFLOPs 和较高的 FPS,使得模型在资源受限的设备上更易于部署。相反,参数量较多的模型可能具有更强的代表能力和更高的 GFLOPs,但可能会牺牲一定的计算效率,导致 FPS 下降。为了进行公平比较,本文在 GPU 上测试了四种深度学习方法在相同条件下(如 256×

256 大小的图像)的参数、GFLOPs 和 FPS。结果如表 3 所示,本文提出的方法在参数数量和 GFLOPs 方面处于中等水平,这为模型的计算效率和性能提供了一个平衡点。具体来说,该方法在保持了适中的参数量和 GFLOPs 的同时,实现了较高的 FPS,这表明它不仅能够有效地处理复杂的任务,而且能够在资源受限的环境中实现实时处理。这种平衡对于实际应用来说是非常有价值的,使得模型既能够在高性能计算环境中发挥作用,也能够移动设备或嵌入式系统中得到有效应用。

### 2.2.2 消融实验

通过消融实验可以分析并得到网络中三个主要的模块对整体性能的影响。

(1)多通道交互注意力。本文通过将常用的 Resnet 残差块和其他注意力机制模块对多通道交互

表2 NUDT-SIRST数据集上的MCIA、ECEM和MLFF消融研究

Table 2 Ablation Study of MCIA, ECEM and MFF on the NUDT-SIRST Dataset

Baseline	MCIA	ECE	MLFF	Pd	Fa	IoU
√	×	×	×	96.88	4.49	78.23
√	√	×	×	96.93	3.3	81.77
√	×	√	×	97.56	4.14	80.38
√	×	×	√	98.20	2.88	80.04
√	×	√	√	98.51	2.29	80.59
√	√	×	√	97.56	2.65	83.27
√	√	√	×	97.09	4.31	80.83
√	√	√	√	98.83	2.09	85.11

表3 不同方法的性能比较结果

Table 3 Performance comparison of different methods

Method	Parameters/MB	GFLOPs	FPS
ACM	1.52	0.55	36.14
ALC-Net	0.52	1.48	29.49
MTU-Net	12.75	6.22	110.01
RDIAN	0.22	3.72	100
MCIAECE	3.88	3.68	47.96

注意力进行替换从而实现对该部分有效性的验证,实验结果如表2和表3所示。发现只替换多通道交互注意力模块时,如表2中第1,2行的结果比较可知,网络性能在各指标都有提升,Pd提高了0.05%,而虚警率则减少了 $1.19 \times 10^{-6}$ 。主要提升的还是IoU,提高了3.54%。这表明多通道交互注意力能够更关注特征图中有用的信息,且能够较好地维持原始图像的信息,提升网络对目标轮廓的描述能力。这在表3中第1行和第4行的数据也可得出,如果在本文提出的网络中移除多通道交互注意力模块,网络模型的IoU会降低4.52%,而其他两个指标所受影响较小。而在表3中可以发现多通道交互注意力模块与其他注意力模块相比能够达到最佳的效果,进一步说明多通道交互注意力模块,能够获取全局上下文信息,显著提升模型性能。

表4 MCIA与其他注意力机制比较结果

Table 4 Comparison of MCIA with Other Attention Mechanisms

Method	$P_d$	$F_a$	IoU
w/o MCIA	98.51	2.29	80.59
CBAM	98.64	3.53	82.45
SE	98.23	4.66	81.14
DCFNet	98.83	2.09	85.11

(2)边缘轮廓增强。本文将边缘轮廓增强模块替换为跳跃连接来验证该模块对目标边缘轮廓增强的有效性。从表2中的第1,3,6,8行实验结果比较可以看出,在基线网络中增加边缘轮廓增强模块能够在Pd提升0.68%,Fa减少 $0.35 \times 10^{-6}$ 的同时IoU提高2.15%,这也说明了该模块能够通过不同大小的空洞卷积有效提取目标的特征,从而获取更细节的目标轮廓并得到高IoU的检测结果。而在本文提出的模型中移除边缘增强模块IoU同样显著降低了1.84%。而其他两个指标相差并不大,这也证明边缘轮廓增强模块能够对目标轮廓进行深层提取获得更精确的检测结果。

(3)多级特征融合。与原网络中仅对前一层和当前层特征进行融合不同,本文采用三级特征融合模块。为了验证该模块的有效性,本文通过采用原Unet网络的特征融合方法替换多级融合模块。从表2中的第1,4,7,8行实验结果比较可以看出。只加入多级特征融合模块,Pd、IoU能够显著提升分别为1.32%,1.81%,并且Fa能够减少 $1.39 \times 10^{-6}$ 。而若是在本文提出的网络模型中移除多级特征融合模块,可以发现模型的性能显著下降,Pd、IoU分别减少了1.74%,4.28%,而Fa则增加了 $2.22 \times 10^{-6}$ 。同时评估了多级特征融合模块采用不同层数时对模块的整体影响。实验结果如表5所示,清晰地揭示了层数对模型性能的显著影响。当模型中采用三级特征融合模块,IoU为 $85.11 \times 10^{-6}$ ,与仅采用其中任意两个层的模块相比,分别提升了6%和5%。此外,模型的Pd和Fa也呈现出了一定的改善。因此可以看出,通过增加MLFF的层数,模型能够更有效地捕捉和整合来自不同层次的特征信息。并且通过融合这些特征不仅可以得到高检测概率还能够增强网络对目标的描述能力,获得具有精确定位且轮廓更为完整的检测结果。

表5 MLFF不同层数的性能比较

Table 5 Performance comparison of MLFF

w/o ECE	$X_{i-1}$	$X_i$	$Y_{i+1}$	Pd	Fa	IoU
√	√	×	√	98.51	2.31	79.08
√	×	√	√	97.09	2.63	80.83
√	√	√	√	98.83	2.09	85.11

(4)特征提取可视化。为了更全面地验证设计的模块的有效性,需要进行更深入的分析。为此,在原有消融实验的基础上,引入了Grad CAM工具

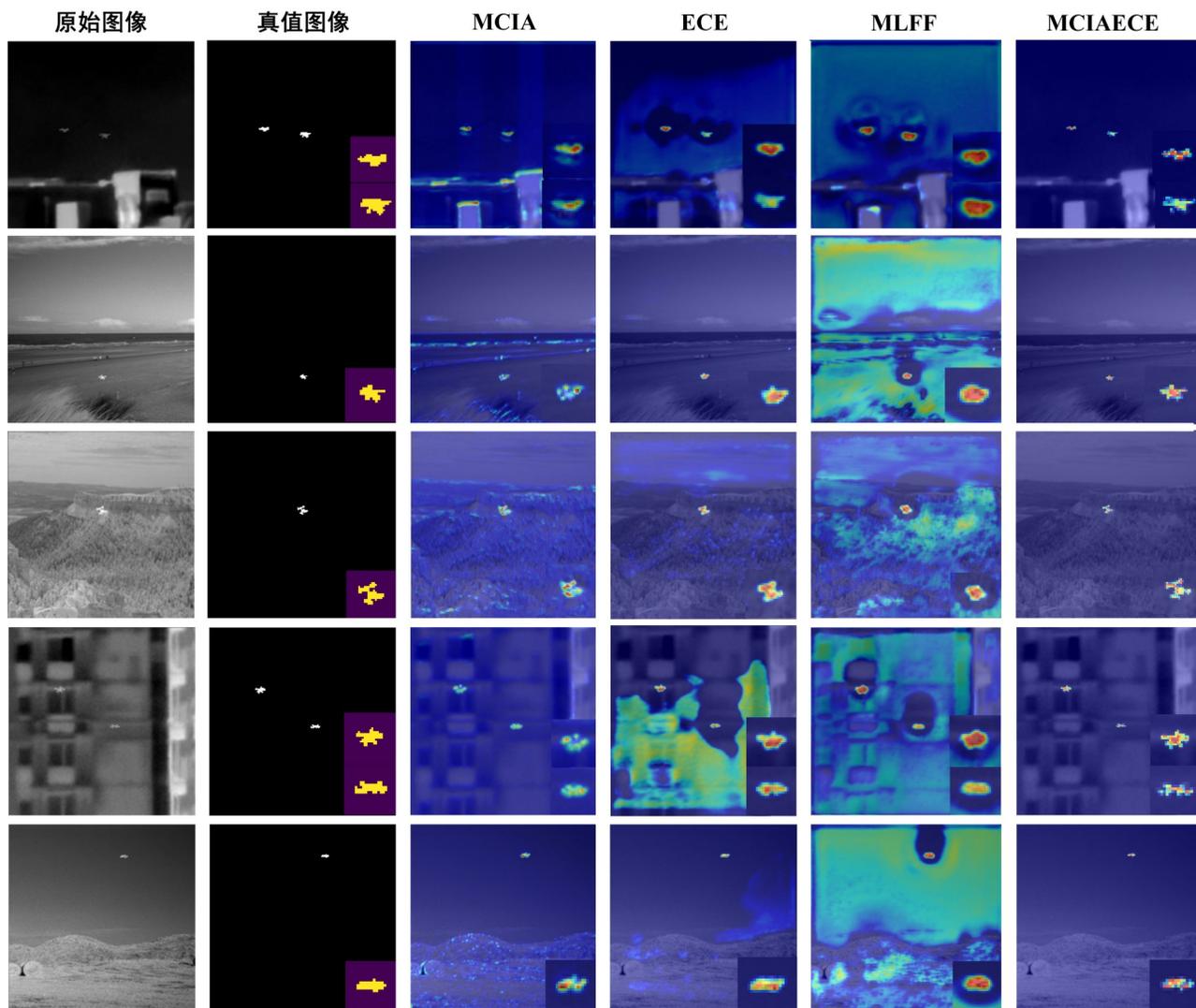


图7 MCIA、ECE、MLFF 三种模块获得的定性结果。为了更好地可视化,目标区域在右下角被放大

Fig. 7 Qualitative results obtained by the three modules MCIA, ECE and MLFF. For better visualization, the target area is enlarged in the lower right corner

来可视化模型的注意力分布。如图7所示,采用可视化方法显示图像的特征层,可以更直观地显示设计结构对特征提取的影响。通过对比添加MCIA、ECE和MLFF模块前后的特征图,可以观察到以下几个显著变化。添加MCIA后,模型的注意力更加集中在目标物体上,背景区域也有部分特征被提取。这一现象表明MCIA能有效地提取目标全局上下文信息,增强了目标与背景的关联性。使得模型可以更好地理解目标所处的环境,从而更加专注于目标特征的学习。而在添加ECE后的特征图中,目标物体的边界变得更加清晰,这表明该模块有助于模型更好地捕捉目标的轮廓信息,从而提高了目标检测的准确性。而加入MLFF模块后,模型的注意

力更加集中在目标区域,这表明MLFF通过整合不同层次的特征信息,这使得模型能够在保持对目标细节敏感的同时,也能够捕捉到更广泛的上下文信息。这使其不仅能够增强目标区域的特征表示,背景特征也得到了更好的利用。通过上述分析,更加确信MCIA、ECE和MLFF模块能够有效地提升模型对目标与全局上下文的理解,帮助模型更好地理解 and 区分目标与背景。同时还能增强目标的边缘轮廓,从而提高检测的准确性。这些可视化结果不仅验证了模块的设计初衷,也为模型提供了更加直观的解释。

### 3 结论

本文基于U型网络结构提出了一种用于红外无

人机检测的网络。该网络采用多通道注意力增强目标特征,并设计边缘轮廓增强模块来提取多层的细节信息,然后利用多级特征融合模块将来自不同层的不同大小的特征图进行融合,在丰富特征信息的同时也能够起到抑制背景的作用,最终得到检测精度高的结果。与不同方法对比表明,该网络在三个数据集中的多数评价指标上都达到最优,且能够在还原目标轮廓时得到更好的效果。未来工作中,需要增加红外无人机图片数量和类型,以提高模型在复杂环境中的适用性。还需要提出更加轻量化的网络模型,不仅能够保证精度,还能提高模型的检测速度。

## References

- [1] MA Qi, SUN Xiao Jun, ZHANG Yang, et al. Detection and recognition method of low-altitude UAV based on infrared images[J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2020, 40(3): 150-154.  
马旗, 孙晓军, 张杨, 等. 基于红外图像的低空无人机检测识别方法[J]. *弹箭与制导学报*, 2020, 40(3): 150-154.
- [2] CAO Y, WANG Y, SONG H, et al. The unidirectional auxiliary surface sliding mode control for compound high-speed helicopter[C]. 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC), IEEE, 2018: 1-6.
- [3] YE Zhen Hua, LI Hui Hao, WANG Jin Dong, et al. Recent hotspots and innovative trends of infrared photon detectors[J]. *Journal of Infrared and Millimeter Waves*, 2022, 41(1): 001.  
叶振华, 李辉豪, 王进东, 等. 红外光电探测器的前沿热点与变革趋势[J]. *红外与毫米波学报*, 2022, 41(1): 001.
- [4] GIRSHICK R, Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 2015: 1440-1448.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016: 779-788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. 14th European Conference on Computer Vision (ECCV), Springer, 2016: 21-37.
- [7] TIAN Z, HUNG J, YANG Y, et al. The drone detection based on improved YOLOv5[C]. *Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference*. ACM, 2022: 83-89.
- [8] LIANG Xiao, LI Jun, An infrared UAV target detection method based on improved YOLOv7 [J]. *Electronics Optics & Control*, 2023, 30(12): 38-43.  
梁晓, 李俊, 基于改进 YOLOv7 的红外无人机目标检测方法[J]. *光电与控制*, 2023, 30(12): 38-43.
- [9] ZHANG Ling Ling, WANG Peng, LI Xiao Yan, et al. Low-altitude UAV detection method based on optimized SSD[J]. *Computer Engineering and Applications*, 2022, 58(16): 204-212.  
张玲玲, 王鹏, 李晓艳, 等. 基于优化 SSD 的低空无人机检测方法[J]. *计算机工程与应用*, 2022, 58(16): 204-212.
- [10] LI B Y, YING X Y, LI R J, et al. The First Competition on Resource-Limited Infrared Small Target Detection Challenge: Methods and Results. <https://arxiv.org/abs/2408.09615v1>
- [11] LIN Zai Ping, LI Bo Yang, LI Miao, et al. Light-weight infrared small target detection combining cross-scale feature fusion with bottleneck attention module [J]. *Journal of Infrared and Millimeter Waves*, 2022, 41(6): 1102-1112.  
林再平, 李博扬, 李淼, 等. 结合跨尺度特征融合与瓶颈注意力模块的轻量型红外小目标检测网络[J]. *红外与毫米波学报*, 2022, 41(6): 1102-1112.
- [12] YIAO Ying Le, ZHAO Juan. Small object tracking algorithm for infrared oversampled scanning images based on deep learning [J]. *Foreign Electronic Measurement Technology*, 2023, 42(1): 35-40.  
姚迎乐, 赵娟, 基于深度学习的红外过采样扫描图像小目标跟踪算法[J]. *国外电子测量技术*, 2023, 42(1): 35-40.
- [13] CHEN Hai Yun, YU Hong Hao, WANG Hai Chuan, et al. Object detection algorithm of thermal images based improved YOLOX [J]. *Electronic Measurement Technology*, 2022, 45(23): 72-81.  
谌海云, 余鸿皓, 王海川, 等. 基于改进 YOLOX 的红外目标检测算法[J]. *电子测量技术*, 2022, 45(23): 72-81.
- [14] CHEN Guang Qiu, WEN Qi Zhang, YIN Wen Qing, et al. Attentional residual dense connection fusion network for infrared and visible image fusion [J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37(8): 182-193.  
陈广秋, 温奇璋, 尹文卿, 等. 用于红外与可见光图像融合的注意力残差密集融合网络[J]. *电子测量与仪器学报*, 2023, 37(8): 182-193.
- [15] LIU T, YANG J G, LI B Y, et al. Infrared small target detection via nonconvex tensor tucker decomposition with factor prior [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-17.
- [16] LI B Y, WANG Y Q, WANG L G, et al. Monte carlo linear clustering with single-point supervision is enough for infrared small target detection [C]. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: 1009-1019.
- [17] LI B Y, WANG L G, WANG Y Q, et al. Mixed-precision network quantization for infrared small target segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-12.
- [18] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, Springer, 2015: 234-241.
- [19] DAI Y, WU Y, ZHOU F, et al. Attentional local contrast networks for infrared small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021,

- 59(11):9813 - 9824.
- [20] SUN H, BAI J, YANG F, et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-13.
- [21] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection [C]. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2021: 945-958.
- [22] WU T, LI B, LUO Y, et al. MTU-Net: Multi-level TransUNet for space-based infrared tiny ship detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.
- [23] WEI Y, YOU X, LI H. Multiscale patch-based contrast measure for small infrared target detection [J]. *Pattern Recognition*, 2016, 58: 216 - 226.
- [24] DAI Y, WU Y, Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection [J]. *IEEE Journal of Selected. Topics Applied Earth Observations Remote Sensing*, 2017, 10(8): 3752 - 3767.
- [25] GAO C, MENG D, YANG Y, et al. Infrared patch-image model for small target detection in a single image [J]. *IEEE Transactions on Image Process*, 2013, 22 (12) : 4996 - 5009.
- [26] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm [J]. *Remote Sensing*, 2019, 11(4): 382.
- [27] ZHU H, LIU S, DENG L, et al. Infrared small target detection via low-rank tensor completion with Top-Hat regularization [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(2): 1004 - 1016.