

DOI: 10.20189/j.cnki.CN/61-1527/E.202602005

稀疏空间特征与线性注意力融合的对空 红外小目标检测

陈光晨¹, 张 枫¹, 张印辉^{1*}, 何自芬¹, 杨小冈², 卢瑞涛²

(1. 昆明理工大学 机电工程学院, 昆明 650500;

2. 火箭军工程大学, 西安 710025)

摘 要: 针对空中红外小目标探测存在目标距离远、热辐射衰减严重及环境信息复杂等问题, 提出稀疏空间特征与线性注意力融合的对空红外小目标检测 (aerial infrared small target detection, AISTD) 模型。首先, 为解决因特征图尺寸降低而导致目标细节信息丢失的问题, 设计了稀疏空间特征提取模块和对称式挤压激励特征提取模块来加强模型对红外小尺度目标的特征提取能力; 然后, 针对热辐射信号易受环境干扰的问题, 将二维图像转换为一维序列后输入到多头线性注意力模块, 并对其进行旋转位置编码, 从多个注意力子空间中获取红外目标细节纹理和轮廓特征信息, 以增强模型在复杂场景下对红外小目标的关注度; 最后, 建立对空红外小目标数据集并对模型进行训练和测试, 以验证模型的有效性。结果表明: 所提方法在对空红外小目标数据集上的 mAP75 达到 90.3%, mAP50-95 达到 74.7%, 模型推理时间缩短至 3.9 ms, 相比现有的 YOLOv11 和 YOLOv13 等目标检测网络, AISTD 能够更快速准确地识别出空中红外小目标。

关键词: 红外空中目标; 稀疏空间特征; 对称式挤压激励; 线性注意力

中图分类号: P237

文献标志码: A

开放科学标识码(OSID):

文章编号: 2097-2741(2026)02-0055-11



听语音
与作者互动
聊科研

红外成像具有全天候、全天时等特点, 因而被广泛应用于目标探测、安防系统以及遥感领域^[1]等。然而, 由于成像距离远导致目标在红外图像中占比较少, 同时受限于按物体热辐射强弱进行成像的特点, 使得红外图像分辨率较低且目标细节信息丢失严重。此外, 红外成像过程易受周围建筑物、云层以及树木等环境因素干扰, 导致目标淹没于嘈杂背景中, 出现图像对比度低、信噪比低^[2]的情况。

传统红外小目标探测通常采用滤波算法, 如自适应频率域巴特沃斯高通滤波^[3]、中值滤波^[4]以及高帽滤波^[5]等, 这些方法都是通过将前景目标

与背景进行分离, 从而得到被检测目标, 计算量较小且实时性好, 但在复杂背景下则会出现误检率高、鲁棒性差等问题。为提高复杂背景下对空红外小目标检测精度, 相关学者提出基于低秩稀疏的检测方法, 如红外块图像 (infrared patch image, IPI) 模型^[6]、重权重红外块图像 (reweight infrared patch image, RIPI) 模型^[7]以及张量核范数的部分和^[8]等, 能够较好地抑制复杂的红外背景噪声, 提升模型对于小尺度目标的检测能力, 但该方法计算量较大, 计算时间长, 难以满足实时性的要求。近年来, 受到人类视觉机制的启发, 拉普拉斯-高斯滤波器^[9]、高斯差分滤波器^[10]以及

收稿日期: 2024-11-25

修回日期: 2025-09-12

录用日期: 2025-10-08

基金项目: 国家自然科学基金 (62561032, 62061022, 62171206); 装备智能运用教育部重点实验室开放基金 (AAIE-2023-0203)

第一作者: 陈光晨 (1997—), 男, 博士研究生, 主要从事图像处理、机器视觉等方面研究。E-mail: guangchen-chen@stu.kust.edu.cn

***通信作者:** 张印辉 (1977—), 男, 教授, 主要从事图像处理、机器视觉及机器智能等方面研究。E-mail: zhangyinhuai@kust.edu.cn

引用格式: 陈光晨, 张枫, 张印辉, 等. 稀疏空间特征与线性注意力融合的对空红外小目标检测[J]. 火箭军工程大学学报, 2026, 40 (2): 55-65. CHEN Guangchen, ZHANG Feng, ZHANG Yinhuai, et al. Sparse spatial feature and linear attention fusion-based aerial infrared small target detection[J]. Journal of Rocket Force University of Engineering, 2026, 40 (2): 55-65.

基于熵的自适应窗口技术^[11]等方法通过对目标特征的显著性程度进行阈值分割,然后利用分割结果对目标进行检测,有效提高了网络对于小尺度红外目标的检测能力,但此种方法采用滑动窗口机制检测显著区域,会使得窗口特征块之间语义关联度降低,进而引起漏检率升高。

当前,随着深度学习在各领域中的广泛应用,基于深度学习的卷积神经网络为红外小目标检测提供了新的解决方案。相比传统目标检测算法,卷积神经网络能自适应提取目标特征信息,具有鲁棒性好、特征表达能力强等优点。常见目标检测算法有快速区域卷积神经网络(fast region-based convolutional neural network, Fast R-CNN)、更快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)^[12]、You Only Look Once (YOLO) 以及单次多框检测器(single shot multi-box detector, SSD)^[13]等,但上述方法大多仅适用于可见光高分辨率图像识别,而红外小目标存在目标尺度占比小、对比度低等问题,因此将其直接应用于红外小目标检测领域难以取得良好效果。为提高模型对红外目标检测精度,鲁棒红外小目标检测网络(robust infrared small target detection network, RISTDnet)^[14]和多层多尺度特征融合网络(multi-layer and multi-scale feature fusion network, MMFF-Net)^[15]通过在网络中引入自适应感受野融合模块和多尺度特征融合主干结构,将不同尺度特征信息进行多特征融合,从而减少了网络在特征提取过程中的信息丢失。除设计多尺度特征融合结构外,红外小目标 U 形检测网络^[16](infrared small-target detection U-Net, ISTDU-Net)在编解码 U 形网络结构的基础上分别引入降采样、全连接和感知损失、结构损失等,来解决小目标在特征提取过程中背景纹理特征的损失问题,以提高小目标特征权重表征能力。此外,注意力机制能够提高网络在复杂场景中对小尺度目标的关注能力,如双向注意力聚集模块和由粗到精的内部注意感知网络,通过将低级信息与高级信息进行融合,使其能够在复杂背景下对目标进行精确检测。

尽管上述方法在红外小目标检测领域取得一定进展,然而面对远距离、高速飞行的飞机,目标往往会在短时间内横跨建筑、云层以及树林等多个应用场景,导致其成像背景极为复杂,对网络响应速度提出更高要求。同时,飞行目标通常是由远到近再到远的飞行过程,相比其他小尺度

目标,飞行目标尺度变化大、形状变化多,给红外检测带来了更大挑战。本文通过稀疏空间特征提取模块将图像空间信息转换为通道信息,而不使用池化或卷积抽象特征空间,以保留更多的目标细节信息。针对红外小目标占比小、信息损失大的问题,设计对称式挤压激励特征提取模块来挤压与扩张通道信息,使得网络能够自适应地提取关键信息,获取更为丰富的目标边缘和形状特征。针对红外图像背景复杂、对比度低的问题,本文引入多头线性注意力机制,通过线性结构、旋转位置编码和多头融合来捕捉全局细节纹理和轮廓特征信息,更好地帮助网络获取上下文信息,从而提高对小目标的关注和表征能力。

1 对空红外小目标检测网络设计

本文提出一种兼顾检测精度与实时性的对空红外小目标检测(aerial infrared small target detection, AISTD)网络,其结构如图 1 所示。首先,将 YOLOv8s 主干网络^[17]中的普通卷积块替换为稀疏空间特征提取(sparse space feature extraction, SSFE)模块进行下采样,使得红外小尺度目标在特征提取过程中尽可能多地保留目标形状和边缘细节信息;其次,将 C2f 模块替换为对称式挤压激励特征提取(symmetrical squeezing excitation feature extraction, SSEFE)模块,以更好提取小尺度目标特征信息,上述 SSFE、SSEFE 以及快速金字塔池化(spatial pyramid pooling fast, SPPF)模块共同组成 AISTD 的特征提取网络;随后,在特征融合结构中嵌入多头线性注意力(multihead linear attention, MHLA)模块,以减少目标周围复杂环境干扰,加强网络对于小尺度目标的关注能力;最后,利用解耦头(decoupled head)解耦检测得到的特征信息,并生成最终的可视化红外小目标检测结果。图 1 中, C 为通道数, G 为组数, H 和 W 为特征图的高和宽, k 为卷积核大小, B 为输入图像数量。

1.1 稀疏空间特征提取模块

主干网络在目标特征提取过程中多次下采样,以减少特征图尺度,从而简化模型计算参数,但特征空间的升维过程会导致目标细节特征信息损失,若仍旧采用卷积模块进行下采样操作,会导致孱弱特征信息进一步丢失。因此,本文设计 SSFE 模块,将空间信息稀疏化处理并在通道维度上拼接,此时空间信息转化为通道信息,使得特

征图在尺度减少的同时降低目标细节信息损失, 其结构如图2所示。首先, 对特征图上的像素点间隔取值再进行拼接, 在同一平面上拼接相同间隔的像素点, 在维度上拼接不同间隔的像素点, 最终拼接得到的特征图长宽缩放为原来的二分之一, 通道维度扩张为原来的4倍; 然后, 通过混洗模块^[18]将特征图通道信息进行通道混洗; 最后, 通过卷积核为3、步长为1的卷积模块(Conv)进行特征融合。图2中, s 为步长, p 为填充参数。

1.2 对称式挤压激励特征提取模块

红外相机能够对空中小尺度目标进行全天候

成像, 但受限于红外成像分辨率差、对比度低以及信噪比低等问题, 极易丢失小尺度目标形状和边缘细节特征。为增强有限特征信息提取能力, 本文设计对称式挤压激励特征提取模块。首先, 利用卷积核为3、步长为1的卷积模块将输入通道压缩至原来的二分之一, 接着利用卷积核为1、步长为1的卷积模块将通道数恢复至原通道数, 通过对特征图通道的挤压和激励来自适应调节图片的细节特征信息; 然后, 将输入的特征信息与挤压激励后的特征进行融合, 通过卷积核为1、步长为1的卷积模块将之前融合的特征信息进行混合; 最

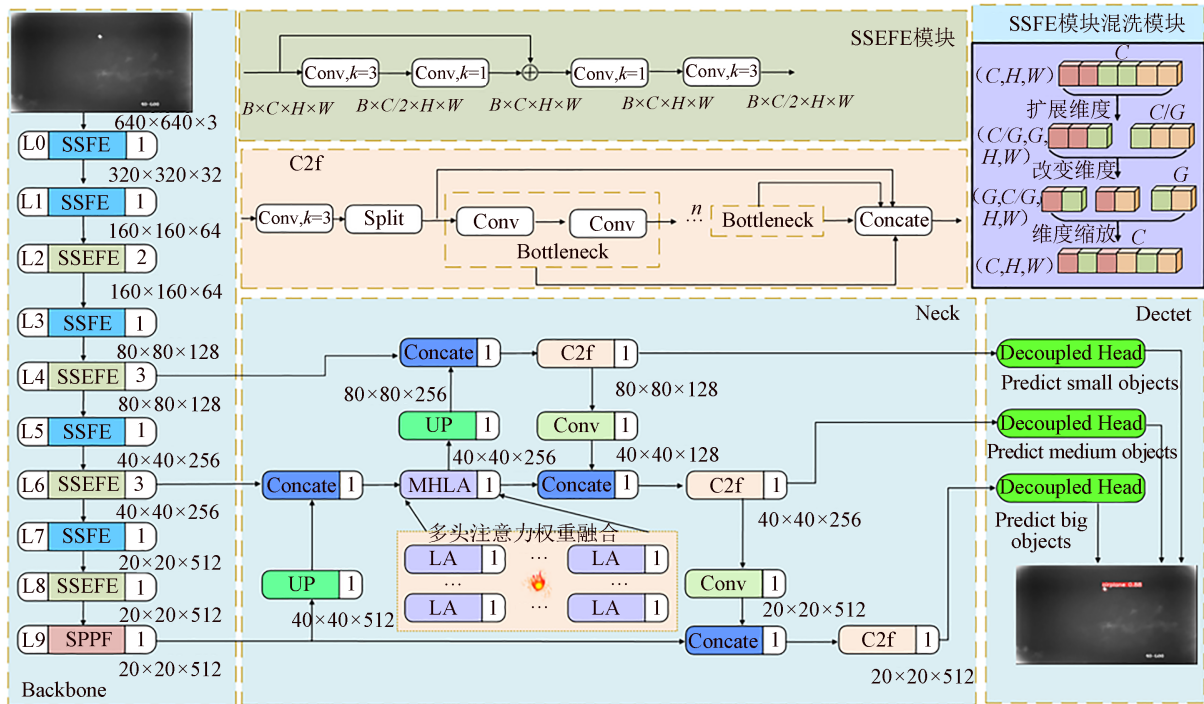


图1 AISTD 架构

Fig. 1 Architecture of the AISTD

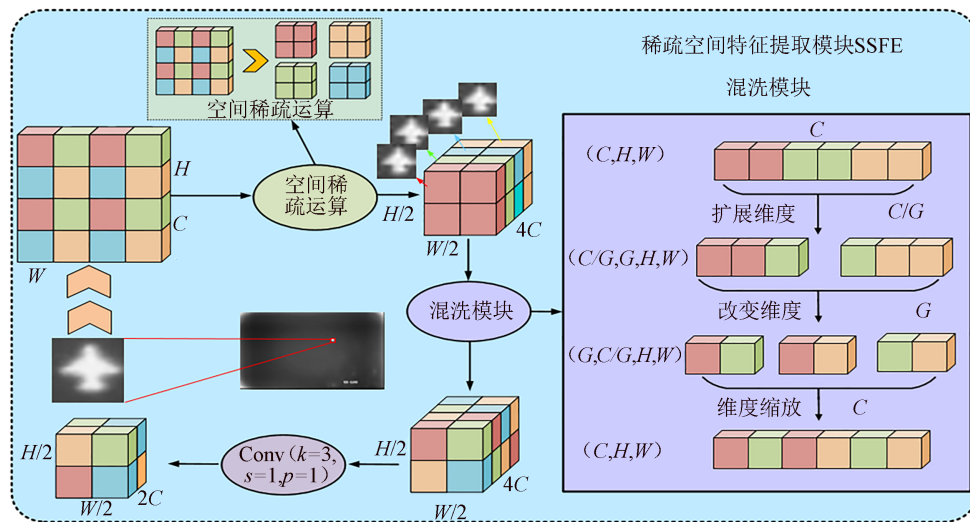


图2 稀疏空间特征提取模块

Fig. 2 Sparse space feature extraction module

后,通过卷积核为3、步长为1的卷积模块细致全面地提取特征图中的细节特征信息,降低特征提取过程中的特征损耗。

如图1所示,对比本文 SSEFE 模块与传统 C2f 模块,从结构特征与计算效率来看,C2f 模块随着 n 值增大将会导致计算量显著增加。而 SSEFE 模块摒弃复杂并行设计,采用仅包含4层卷积的固定串行结构,在主干网络中实现浅层特征的直接传递,有助保持微小目标的边缘与纹理信息,同时具备明显计算效率优势。针对红外小目标检测任务,SSEFE 模块的轻量化设计更适合维持高分辨率特征图,有利于实现亚像素级的小目标定位,适用于实时空中红外监测场景。

1.3 多头线性注意力模块

由于应用场景和成像原理的特殊性,对空红外小目标探测面临双重挑战:一是成像过程易丢失细节,导致目标成像模糊;二是热辐射信号常受背景干扰,进一步增加了探测难度。为应对上述问题,本文引入多头线性注意力模块。该模块为满足机载平台的快速检测需求,将二维图像转换为一维序列以加速数据处理,再经过并行工作的多个独立的注意力头,从不同维度捕捉目标的上下文特征,从而增强了模型对复杂干扰背景下微弱目标的关注能力。同时,该模块集成线性注意力与旋转位置编码,不仅能够有效降低长序列建模的计算复杂度,还可以通过显式编码位置信息来弥补红外图像因目标模糊而导致的结构信息

缺失,从而共同提升模型在复杂场景下的鲁棒性与准确性。多头线性注意力模块整体结构如图3所示,其中每个注意力头都需要查询(Query)、键(Key)和值(Value)向量,分别为 Q 、 K 和 V 。

如图3所示,首先将二维输入张量信号转换为一维特征序列,再利用卷积核大小分别为1和3的连续一维卷积块进行滤波处理,采用层归一化后输入到2个线性变换层,其中右侧输出经过卷积核为1的一维卷积块后送入线性注意力模块。在线性注意力模块中,输入经过线性变换后通过索引切片和指数线性单元激活函数 ELU 计算查询向量 $Q = \text{elu}(\mathbf{x}\mathbf{W}_q) + 1$,键向量 $K = \text{elu}(\mathbf{x}\mathbf{W}_k) + 1$,并将初始注意力输入作为值向量 $V = \mathbf{x}\mathbf{W}_v$ 。其中, $\text{elu}(\cdot)$ 为指数线性单元激活函数, $\mathbf{x}$ 为原始输入的特征图, $\mathbf{W}_q$ 、 $\mathbf{W}_k$ 和 $\mathbf{W}_v$ 为可学习的权重矩阵。通过旋转位置编码,为查询向量 Q 和键向量 K 乘上旋转矩阵 R_θ^d 得到 Q^r 和 K^r ,编码过程为

$$\begin{cases} Q^r = R_\theta^d Q \\ K^r = R_\theta^d K \end{cases} \quad (1)$$

$$R_\theta^d = \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 & \cdots & 0 & 0 \\ \sin \theta_1 & \cos \theta_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cos \theta_{d/2} & -\sin \theta_{d/2} \\ 0 & 0 & \cdots & \sin \theta_{d/2} & \cos \theta_{d/2} \end{bmatrix} \quad (2)$$

式中: $\theta_i = 10^{\frac{-8i}{d}}$, d 为符合特征图大小的偶数; Q^r 和 K^r 分别为通过旋转位置编码的查询向量 Q 和

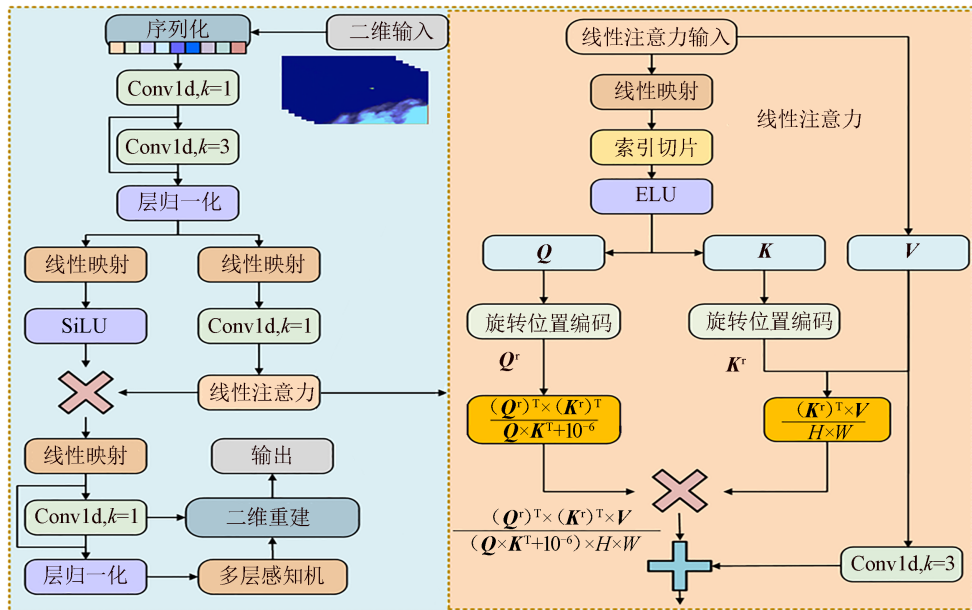
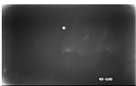


图3 多头线性注意力模块

Fig. 3 Multi-head linear attention module

表 1 不同场景的统计数据
Table 1 Statistics for different scenes

场景	目标数量	图像数量/张	示例
无干扰	单机	557	
	多机	86	
建筑干扰	单机	1 025	
云层干扰	单机	399	
树木干扰	单机	87	
其他干扰	单机	195	

2.3 评价指标

本文采用的评价指标为 mAP75 和 mAP50-95。其中, mAP75 代表多个预测框与标注框交并比(intersection over union, IoU) 阈值取 0.75 时所有测试图片的平均精度; mAP50-95 代表 IoU 阈值取 0.50、0.55、0.60 一直到 0.95 时的平均精度, 数学表达式为

$$P = \frac{T}{T + F} \quad (5)$$

$$\text{mAP} = \frac{\sum_{i=0}^K \text{AP}_i}{K} \quad (6)$$

$$\text{mAP50-95} = \frac{\text{mAP50} + \text{mAP55} + \dots + \text{mAP95}}{10} \quad (7)$$

式中: P 为预测精度; T 为预测正确的正例; F 为预测错误的正例; AP_i 为各类场景中不同置信度情况下使用准确率和召回率构建的曲线所围成的面积; K 为数据集中的类别数。

3 实验结果分析

3.1 主干网络对比

主干网络是模型进行特征提取的重要组成部分, 为探究不同主干对于模型检测精度的影响程度, 分别选取 GhostNet^[20]、ShuffleNetv2^[21]、MobileNetv3^[22] 以及 PP-LCNet^[23] 4 个主干网络进行对

比, 实验结果如表 2 所示。将 SSFE 和 SSEFE 作为主干的模型检测精度明显优于其他主干网络, 相比于 GhostNet 主干网络, 模型的 mAP75 增长了 0.4 个百分点, mAP50-95 增长了 0.6 个百分点。ShuffleNetv2 采用深度可分离卷积模块, 能够将通道数按照分组进行计算, 并减少模型计算参数, 可以显著提高推理速度, 但同时也会导致通道间组别信息关联缺失, 从而引起精度下降, 模型的 mAP75 相比 GhostNet 下降了 0.4 个百分点, mAP50-95 下降了 0.7 个百分点, 推理时间缩短 15.6%。MobileNetv3 也采用深度可分离卷积模块组成特征提取网络, 采用新的激活函数并引入注意力模块来提升网络对目标的关注能力, 同时提升模型推理速度, 由于缺少通道间的组别关联, 相比 ShuffleNetv2 的 mAP75 下降 0.6 个百分点, mAP50-95 下降 2 个百分点, 但推理时间降低至 3.7 ms。PP-LCNet 为所有对比网络中推理速度最快的主干网络, 其推理时间仅为 2.7 ms, 这得益于其引入了推理加速策略, 但对空中小目标的检测精度明显下降, 模型的 mAP75 仅为 87.9%, mAP50-95 仅为 72.6%。

表 2 不同主干网络的对比实验结果

Table 2 Results of comparative experiments of different backbone networks

主干网络	mAP75/%	mAP50-95/%	运行时间/ms
SSFE+SSEFE	89.2	74.3	4.2
GhostNet	88.8	73.7	4.5
ShuffleNetv2	88.4	73.0	3.8
MobileNetv3	87.8	71.0	3.7
PP-LCNet	87.9	72.6	2.7

为直观展示本文主干网络的特征提取效果, 将原始热图与利用本文主干网络处理后的热图进行对比, 其可视化结果如图 5 所示。从图 5 (a) 中可以看到, 原始热图特征更为分散, 并且在第二张图中背景干扰尤为明显, 而引入 SSFE+SSEFE 主干提取网络后, 三张热图均能够清晰展示前景目标轮廓特征。

3.2 不同检测模型对比

为验证本文所提模型对空中红外小目标检测的有效性, 将其与 YOLOv3^[24]、YOLOv5^[25]、YOLOv7^[26] 以及 YOLOv8^[27] 等进行对比, 引入多种评价指标对模型进行全面评估, 实验结果如表 3 所示。其中, 粗体表示每列中的最优值。

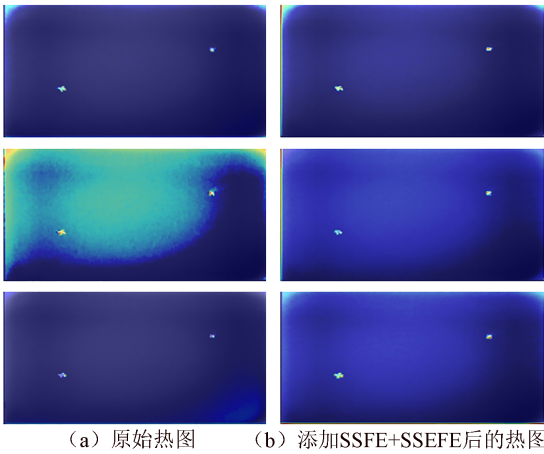


图 5 主干网络处理前后的热图可视化结果

Fig. 5 Results of heatmap visualization of backbone network pre- and post-procession

表 3 不同网络对比实验				
Table 3 Results of comparative experiments of different networks				
模型	mAP75/ %	mAP50- 95/%	运算时 间/ms	参数 量/Mb
YOLOv3-Tiny	82.4	67.6	1.3	8.7
YOLOv5s-6.0	86.6	72.0	3.6	7.0
YOLOv5s-5.0	80.9	68.0	3.7	7.1
YOLOv7-Tiny	84.7	68.0	1.9	6.0
YOLOv8s	86.8	73.1	5.1	11.1
YOLOv8m	88.1	74.1	10.2	25.8
YOLOv11s	89.5	74.5	2.3	9.4
YOLOv12s	88.5	74.3	2.9	9.1
YOLOv13n	88.0	73.8	2.6	2.5
YOLOv13s	87.7	74.1	3.4	9.0
AISTD	90.3	74.7	3.9	13.8

注: 粗体数字表示本列最优值。

由表3可得, 本文所提出的网络模型对于空中红外小尺度目标检测精度最高, 模型的mAP75为90.3%, mAP50-95为74.7%, 相比于次优方法YOLOv11s, 所提模型的mAP75提高0.8个百分点, mAP50-95提高0.2个百分点。与基准网络YOLOv8s相比, 所提模型的mAP75提高3.5个百分点, mAP50-95提高1.6个百分点。此外, 本文模型与YOLOv13等网络模型进行对比, 均有不同程度的提高, 表明向基础网络中加入稀疏空间特征提取模块、对称式挤压激励特征提取模块以及多头线性注意力模块后, 使得本文网络模型相比其他对比模型有着更高的检测精度, 并能够在检测精度与运行速度之间达到平衡。

为定性分析本文模型的目标检测效果, 选取YOLOv5s-6.0、YOLOv5s-5.0、YOLOv7-Tiny、YOLOv8s以及YOLOv11s网络模型与本文AISTD网络模型进行对比, 并选取4种具有挑战性的干扰场景(分别为云层干扰场景、其他干扰场景、建筑干扰场景和树木干扰场景)进行分析, 检测结果如图6所示。其中, 为更清晰对比检测框的拟合情况以及置信度, 在右上角红色框中展示了局部放大信息。

由图6可知, 从整体检测效果出发, 所有参与对比的网络都能准确识别出空中飞机, 但并非每个网络在不同场景中都有较好的检测效果。在云层干扰场景中, 目标尺度较小, 所有参与对比的网络都能较好地检测出空中飞行目标。其中, 表现最好的为AISTD模型, 其小目标检测的置信度达到0.91, 而置信度同为0.91的YOLOv7-Tiny模型将云层中的噪声误检为小目标。在其他干扰场景中, 由于目标尺度较小和建筑物热辐射较大, 在图像上呈现大片建筑物亮斑干扰目标识别, 导致YOLOv5s和YOLOv8s精度下降幅度较大, 本文模型在该场景下的目标检测置信度为0.76, 与YOLOv11s模型仅相差0.01, 表明本文模型和YOLOv11s的抗干扰能力要明显强于其他对比模型。在建筑和树木干扰场景中, 由于干扰物与远处小目标红外特征相似, 使得网络对目标的识别难度提升, 但本文模型仍然具有较好的检测效果, 置信度分别为0.81和0.77, 与YOLOv11s的抗干扰能力相近。综上所述, AISTD网络在检测精度、推理速度以及不同场景的抗干扰能力方面, 相比其他对比网络, 整体表现较好。

3.3 消融实验

为探究各模块对模型检测精度的影响, 在YOLOv8s检测模型基础上, 依次加入SSFE模块、SSEFE模块和MHLA模块, 分别组成YoI、YoIS以及AISTD网络模型。各模块的消融实验结果如表4所示。

表 4 各模块的消融实验结果						
Table 4 Results of ablation experiments of different modules						
模型	SSFE	SSEFE	MHLA	mAP75/ %	mAP50- 95/%	运行 时间/ms
YOLOv8s				86.8	73.1	5.1
YoI	√			87.8	73.1	4.1
YoIS	√	√		89.2	74.3	4.2
AISTD	√	√	√	90.3	74.7	3.9

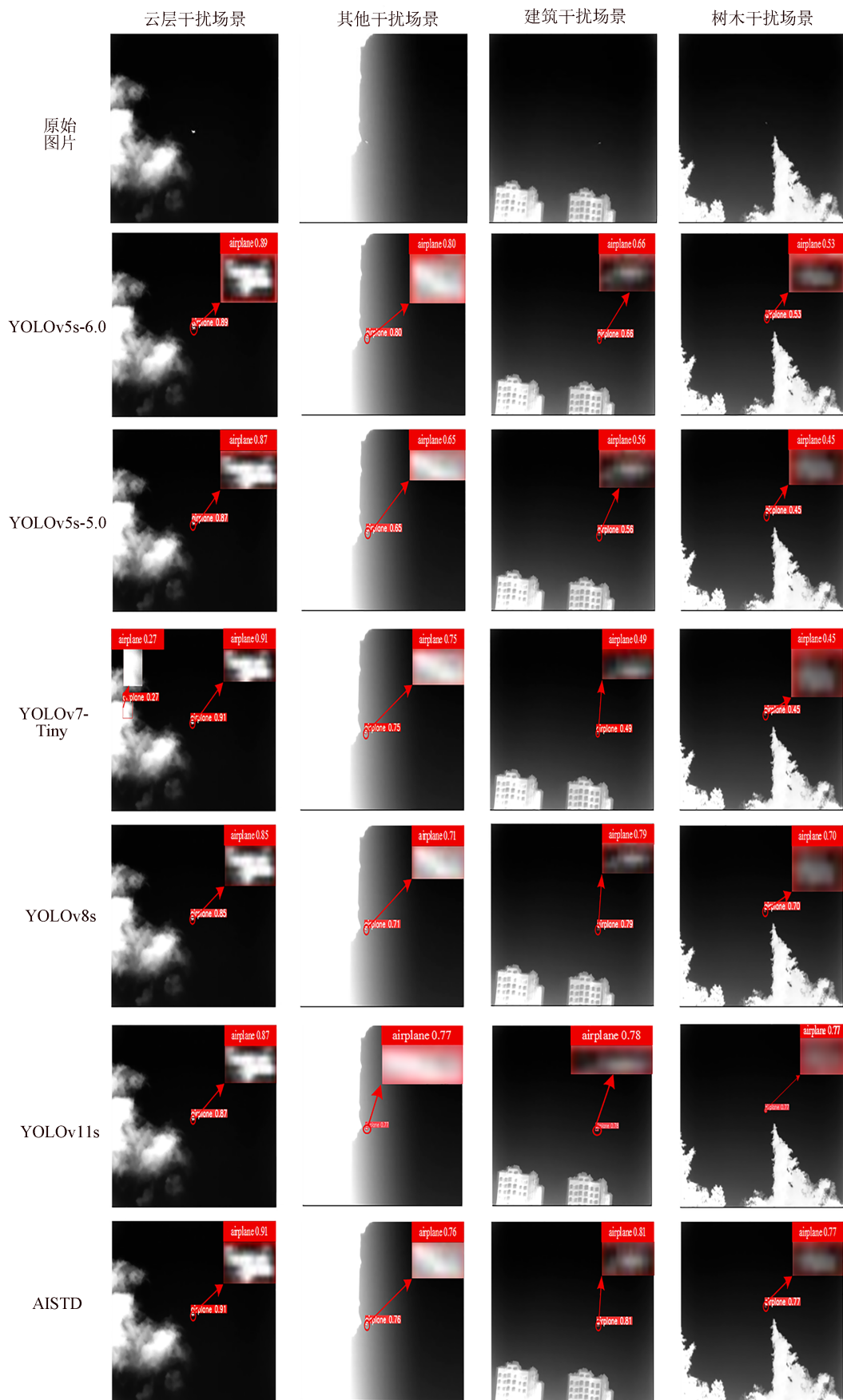


图 6 不同网络的检测可视化结果

Fig. 6 Visualized detection results of different networks

从表4结果分析可得,在初始网络检测模型中加入各个模块后都能提升其对红外小尺度目标的检测精度。首先,向基准网络中加入SSF模块组成的YoI模型能够提高网络对于小尺度空中红外目标的检测能力,模型的mAP75相比YOLOv8s模型增加了1个百分点,同时模型的推理时间缩短了19.6%。当向YoI中加入SSEFE模块后,进一步提高了网络对于小尺度目标特征提取能力,mAP75相比初始YOLOv8s网络模型增长了2.4个百分点,mAP50-95增长了1.2个百分点。最后,当引入全部模块后,AISTD模型的mAP75相比于YOLOv8s增长了3.5个百分点,mAP50-95增长1.6个百分点,同时由于MHLA模块能够显著降低模型参数量,使得AISTD模型的推理速度进一步缩短至3.9 ms。

4 结论

本文针对空中红外小目标检测面临目标距离远、热辐射衰减大及环境信息复杂等挑战,提出一种基于稀疏空间特征与线性注意力融合的对空红外小目标检测模型,并在本文所构建的红外空中小目标数据集上进行了测试与验证,主要结论如下。

1) 设计稀疏空间特征提取模块,将空间信息稀疏化处理后,采用特征图通道信息混洗,降低细节信息损失,在YOLOv8基准模型中加入此模块后,模型的mAP75和mAP50-95分别达到87.8%和73.1%。

2) 设计对称式挤压激励特征提取模块,通过对特征图通道的挤压和激励来自适应调节图片的细节特征信息,将输入的特征信息与挤压激励后的特征进行融合,从而增强模型对小尺度红外目标的特征感知能力。在YoI模型中加入此模块后,模型的mAP75和mAP50-95分别达到89.2%和74.3%。

3) 利用旋转位置编码的多头线性注意力机制,通过线性特征转化来降低计算复杂度,并通过旋转位置编码和多头注意力来捕获红外目标的细节纹理与轮廓特征,从而显著提高模型对复杂场景中小目标的关注度。在YoIS模型中加入此模块后,模型的mAP75和mAP50-95分别达到90.3%和74.7%。

4) 本文所构建的模型在自建空对空红外小目标数据集上的mAP75达到90.3%,mAP50-95达

到74.7%,推理速度缩短至3.9 ms,相较于初始基准网络YOLOv8s,mAP75提高了3.5个百分点,mAP50-95提高1.6个百分点,推理速度提高23.5%。此外,AISTD还与其他经典目标检测网络进行对比,在不同评估维度上均表现更优。

本文为空中红外小目标检测提供了新的技术方案,同时为基于线性注意力的红外目标检测研究提供了新的思路。本文模型虽然具备较高的目标识别精度,但参数量相应增加约2.7 Mb,与轻量化网络相比,推理效率仍有优化空间。后续,将围绕模型剪枝、知识蒸馏等方法开展模型压缩与提升运算速度,进一步提升其在边缘设备端的部署适配能力。

参考文献

- [1] Zhang Yuting, Li Zhengzhou, Siddique A, et al. Infrared small target detection based on interpretation weighted sparse method[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5001415.
- [2] Wang Xiaozhen, Han Chengshan, Li Jiaqi, et al. Multiscale feature extraction U-Net for infrared dim and small-target detection[J]. Remote Sensing, 2024, 16(4): 643.
- [3] 姚林杰,王耀斌,蒋伟. 巴特沃斯高通滤波器的优化设计与仿真实现[J]. 机电工程技术, 2023, 52(2): 292-296.
Yao Linjie, Wang Yaobin, Jiang Wei. Optimization design and simulation of Butterworth high pass filter [J]. Mechanical & Electrical Engineering Technology, 2023, 52(2): 292-296.
- [4] 黄西莹,张旭辉,黄玉,等. 基于国产异构FPGA的空中红外目标检测跟踪系统[J]. 红外与激光工程, 2024, 53(8): 114-124.
Huang Xiying, Zhang Xuhui, Huang Yu, et al. Domestic heterogeneous FPGA-based aerial infrared target detection and tracking system[J]. Infrared and Laser Engineering, 2024, 53(8): 114-124.
- [5] 贺遵亮,黄营磊. 基于TopHat背景建模的运动小目标检测算法分析[J]. 电子技术, 2022, 51(8): 64-66.
He Zunliang, Huang Yinglei. Analysis of moving small target detection algorithm using TopHat background modeling[J]. Electronic Technology, 2022, 51(8): 64-66.
- [6] Gao Chenqiang, Meng Deyu, Yang Yi, et al. Infrared patch-image model for small target detection in a single image[J]. IEEE Transactions on Image Processing, 2013, 22(12): 4996-5009.
- [7] Dai Yimian, Wu Yiquan. Reweighted infrared patch-

- tensor model with both nonlocal and local priors for single-frame small target detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(8): 3752-3767.
- [8] Wang Guanghui, Tao Bingjie, Kong Xuan, et al. Infrared small target detection using nonoverlapping patch spatial-temporal tensor factorization with capped nuclear norm regularization[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5001417.
- [9] 梁锐, 魏阳杰. 自适应 B 样条小波函数模糊距离测量方法[J]. *中国图象图形学报*, 2021, 26(3): 503-515.
- Liang Rui, Wei Yangjie. Adaptive distance measurement method with blur of B-spline wavelet function [J]. *Journal of Image and Graphics*, 2021, 26(3): 503-515.
- [10] 曾瀚林, 孟祥勇, 钱惟贤. 高斯差分滤波图像融合方法[J]. *红外与激光工程*, 2020, 49(S1): 168-176.
- Zeng Hanlin, Meng Xiangyong, Qian Weixian. Image fusion algorithm based on DOG filter[J]. *Infrared and Laser Engineering*, 2020, 49(S1): 168-176.
- [11] Deng He, Sun Xianping, Liu Maili, et al. Entropy-based window selection for detecting dim and small infrared targets[J]. *Pattern Recognition*, 2017, 61: 66-77.
- [12] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[EB/OL]. (2016-01-06) [2024-11-29]. <https://arxiv.org/abs/1506.01497>.
- [13] Juhartini, Arwidiyarti D, Desmiwati. Single shot multibox detector (SSD) in object detection: a review [J]. *IJACI: International Journal of Advanced Computing and Informatics*, 2025, 1(2): 118-127.
- [14] Hou Qingyu, Wang Zhipeng, Tan Fanjiao, et al. RISTDnet: robust infrared small target detection network[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 7000805.
- [15] Zhu Ge, Chen Yuhan, Wang Xianquan, et al. MM-FE-NET: multi-layer and multi-scale feature fusion network for low-light infrared image enhancement [J]. *Signal, Image and Video Processing*, 2024, 18(2): 1089-1097.
- [16] Du Xinyue, Cheng Ke, Zhang Jin, et al. Infrared small target detection algorithm based on improved dense nested U-Net network[J]. *Sensors*, 2025, 25(3): 814.
- [17] Mulyana D I, Rowis M A I. Optimization of text mining detection of Tajweed reading laws using the YOLOv8 method on the Qur'an[J]. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, 2022, 14(2): 1089-1110.
- [18] Udeh C P, Chen Luefeng, Du Sheng, et al. Improved ShuffleNet V2 network with attention for speech emotion recognition[J]. *Information Sciences*, 2025, 689: 121488.
- [19] Philip Chen C L, Li Hong, Wei Yantao, et al. A local contrast method for small infrared target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 574-581.
- [20] Tang Jingyi, Xu Bu, Li Jue, et al. Ghost-YOLO-GBH: a lightweight framework for robust small traffic sign detection via GhostNet and bidirectional multi-scale feature fusion[J]. *Eng*, 2025, 6(8): 196.
- [21] Bu Yifan, Liu Hao, Li Hongda, et al. Grading algorithm for Orah sorting line based on improved ShuffleNet V2[J]. *Applied Sciences*, 2025, 15(8): 4483.
- [22] Lyu Weiyi, Xu Kai, Wu Chonglong, et al. Automatic identification method for subaqueous gravity flow facies in core images using convolutional neural networks[J]. *Earth Science Informatics*, 2024, 18(1): 119.
- [23] 胡伟超, 郭宇阳, 张奇, 等. 基于改进 YOLOX 的轻量化交通监控目标检测算法[J]. *计算机工程与应用*, 2024, 60(7): 167-174.
- Hu Weichao, Guo Yuyang, Zhang Qi, et al. Lightweight traffic monitoring object detection algorithm based on improved YOLOX[J]. *Computer Engineering and Applications*, 2024, 60(7): 167-174.
- [24] 冯伟, 薛如翔, 傅志凌, 等. 基于注意力机制和改进 YOLOv3 的红外弱小目标检测[J]. *飞控与探测*, 2023, 6(4): 84-94.
- Feng Wei, Xue Ruxiang, Fu Zhiling, et al. Infrared small target detection based on improved YOLOv3 and attention mechanism[J]. *Flight Control & Detection*, 2023, 6(4): 84-94.
- [25] 邵延华, 张兴平, 张晓强, 等. 联合结构重参数和 YOLOv5 的航拍红外目标检测[J]. *电子科技大学学报*, 2024, 53(3): 382-389.
- Shao Yanhua, Zhang Xingping, Zhang Xiaoqiang, et al. Infrared aerial object detection based on structural reparameterization and YOLOv5[J]. *Journal of University of Electronic Science and Technology of China*, 2024, 53(3): 382-389.
- [26] 许晓阳, 高重阳. 改进 YOLOv7-tiny 的轻量级红外车辆目标检测算法[J]. *计算机工程与应用*, 2024, 60(1): 74-83.

- Xu Xiaoyang, Gao Chongyang. Improved YOLOv7-tiny lightweight infrared vehicle target detection algorithm[J]. Computer Engineering and Applications, 2024, 60(1): 74-83.
- [27] 赵晶钰, 李敏, 陈谢发, 等. 融合注意力机制与全局信息的 SAR 舰船目标检测算法[J]. 火箭军工程大学学报, 2024, 38(4): 40-46.
- Zhao Jingyu, Li Min, Chen Xiefa, et al. SAR ship target detection algorithm combining attention mechanism and global information[J]. Journal of Rocket Force University of Engineering, 2024, 38(4): 40-46.

Sparse Spatial Feature and Linear Attention Fusion-Based Aerial Infrared Small Target Detection

CHEN Guangchen¹, ZHANG Feng¹, ZHANG Yinhui^{1*},
HE Zifen¹, YANG Xiaogang², LU Ruitao²

(1. Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology,
Kunming 650500, China;

2. Rocket Force University of Engineering, Xi'an 710025, China)

ABSTRACT: Aerial infrared small-target detection faces critical challenges including long detection distances, severe thermal radiation attenuation, and complex environmental interference. To address these issues, an aerial infrared small-target detection (AISTD) model that effectively integrates sparse spatial features with linear attention was proposed. First, to alleviate the loss of fine-grained feature information caused by the reduced feature map, a sparse spatial feature extraction module and a symmetrical squeezing excitation feature extraction module were introduced to enhance the model's capability of infrared small-scale target feature extraction. Subsequently, to mitigate environmental interference with thermal radiation signals, the two-dimensional image was converted into a one-dimensional sequence and input into a multi-head linear attention module, with rotary positional encoding applied. The proposed model captured detailed texture and contour information of infrared targets across multiple attention subspaces, thereby improving the sensitivity to small targets in complex aerial scenes. Furthermore, an aerial infrared small-target dataset was constructed to train and test the proposed approach and validate its effectiveness. Extensive experimental results demonstrate that the proposed AISTD achieves an mAP75 of 90.3% and an mAP50-95 of 74.7% on a self-built dataset, while reducing the inference time to 3.9 ms. Compared with existing object detection networks, such as YOLOv11 and YOLOv13, the AISTD model exhibits superior performance in terms of both detection accuracy and computational efficiency.

KEYWORDS: infrared aerial target; sparse spatial features; symmetrical squeezing excitation; linear attention

(编辑: 刘建民)