

文章编号: 1672-8785(2025)07-0012-08

一种基于多帧集合预测的面向空中无人机的 红外小目标检测算法

吴 晗* 李嘉巍

(郑州大学计算机与人工智能学院, 河南 郑州 450001)

摘 要: 针对红外小目标检测少像素、弱纹理、无颜色导致的识别难题, 提出了一种面向空中无人机的红外小目标检测算法。基于多帧红外序列, 采用 Res-Net-50 网络逐帧提取深度特征, 结合多尺度可变形注意力机制与三次插值增强小目标细节表征; 设计帧解码器与目标解码器, 通过视频级查询向量生成目标实例的时空掩码。训练时利用匈牙利算法优化分类、掩码及相似性损失; 推断时融合高置信度查询掩码, 并采用帧间差分法抑制静态噪声。在 DSAT 数据集上, 本文算法的精度达到 0.6356, F 得分为 0.6475, 性能得以显著提升。该算法通过多尺度特征融合与时序建模, 有效解决了红外小目标检测算法的漏检和虚警问题, 为无人机检测提供了高精度解决方案。

关键词: 目标检测; 信号处理; 实例分割; 空中小目标; 无人机检测

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2025.07.002

An Infrared Small Target Detection Algorithm for Aerial UAVs Based on Multi-Frame Set Prediction

WU Han*, LI Jia-wei

(School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Aiming at the recognition difficulties of infrared small target detection caused by few pixels, weak texture and no color, an infrared small target detection algorithm for aerial unmanned aerial vehicle (UAV) is proposed. Based on multi-frame infrared sequences, the ResNet-50 network is used to extract deep features frame by frame, and the multi-scale deformable attention mechanism and cubic interpolation are combined to enhance the detail representation of small targets; the frame decoder and target decoder are designed to generate the spatiotemporal mask of the target instance through the video-level query vector. The Hungarian algorithm is used to optimize the classification, mask and similarity losses during training; the high-confidence query mask is fused during inference, and the inter-frame difference method is used to suppress static noise. On the DSAT dataset, the accuracy of the algorithm in this paper reaches 0.6356 and the F -score is 0.6475, and the performance is significantly improved. Through multi-scale feature fusion and temporal modeling, the algorithm effectively solves the problems of missed detection and false alarm of infrared small target detection al-

收稿日期: 2025-04-17

作者简介: 吴晗(1995-), 女, 河南濮阳人, 博士, 主要研究方向为视频目标检测、识别与跟踪。

*通讯作者: E-mail: zzuwuhan@zzu.edu.cn

gorithm, and provides a high-precision solution for UAV detection.

Key words: target detection; signal processing; instance segmentation; small aerial target; UAV detection

0 引言

红外小目标检测广泛应用于空中监视、海洋监测、智能导航等方面^[1-2], 同时也是计算机视觉领域的难点问题。本文所提小目标是指离成像设备远而导致成像缺少纹理信息、像素个数少的无人机目标。它们只占若干像素且成像个数差异大, 像素个数变化范围在一个至数百个之间^[3]。小目标检测的难点不仅在于目标成像个数少, 而且成像背景中包含大量噪声和干扰物。如图 1 所示, 图像取自 DSAT 数据集 data5 序列^[4]: 左边图像像素为 256×256 , 红色框表示的目标占 6 个像素; 右边图像是对目标进行 5 倍插值后的成像。可以看出, 目标小, 缺少颜色、形状、纹理信息, 且图像中存在大量外观相似的干扰物。因此, 如何确定目标在图像中的位置, 实现对目标的精准定位是本文着手解决的问题^[4]。

早期经典红外目标检测算法通过设计滤波器来增强特定目标和抑制背景^[5]。随着深度学习的发展, 基于数据驱动的深度卷积神经网络被用于红外目标检测^[3,6-8]。然而, 这些算法主要聚焦于单幅红外图像, 只关注目标外观信息, 因此在具有复杂背景干扰的数据中仍存在漏检或虚警现象。

传统“先跟踪再检测”算法通过分析目标可能的轨迹来确认目标, 即使用目标跟踪算法获得视频中目标的所有潜在轨迹, 并沿着这些轨迹整合能量^[9-14]。此外, 基于低秩稀疏矩阵优化的多帧检测算法通过堆叠帧构建时空张量(包括低秩背景张量、稀疏目标张量和噪声张量), 将红外目标检测任务视为时空张量分解和优化问题^[15-16]。上述方法能够提高信噪比, 进而提升检测性能。然而此类做法通常将目标当作质点处理, 只关注目标的运动信息, 并没有提取目标的外观(灰度)信息。而且这些方法受限于成像平台静止的情况。因此, 本文提出基于集合预测的多帧红外目标检测模型, 旨在同时利用目标的外观信息及其在多帧上的运动信息, 解决现有算法的漏检和虚警问题。使用视觉变换器模型对多帧之间的关系进行建模, 并通过收集具有多种分辨率的红外数据集训练所提出的神经网络。在红外无人机目标检测基准上的实验展示出本文方法的有效性。

1 相关工作

随着深度学习技术的突破性进展, 基于深度学习的目标检测方法已成为研究热点。作为目标检测的难点, 近年来红外小目标检测与跟踪技术在学术界和工业界备受瞩目。Wang H

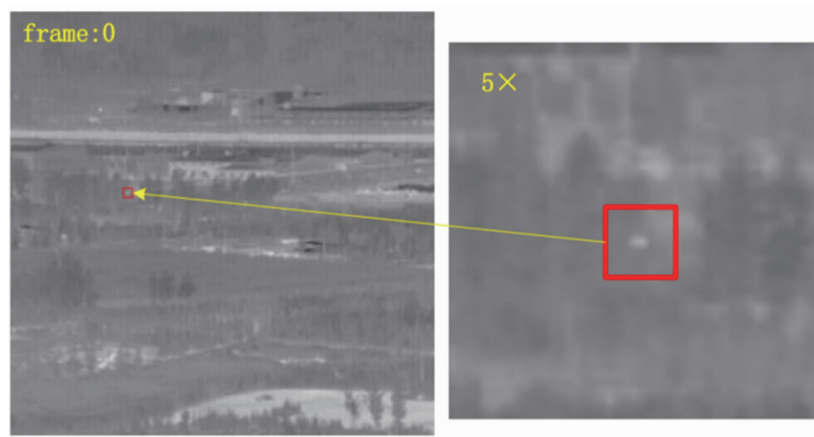


图 1 红外图像及目标示例^[4]

等人为红外目标检测构建了基于双生成器-判别器架构的深度对抗学习框架^[6]。文献^[7]提出了一种用于提升模型特征表达能力的非对称上下文调制模块。Dai Y 等人设计的逐点通道注意力机制通过自底向上的调制路径实现高低层特征交互^[17]。

在特征建模方面,局部相似性金字塔模块旨在捕获目标的多尺度特征^[18],全局注意力网络则通过注意力聚焦机制提升小目标检测精度^[19]。ISNet^[8]融合泰勒有限差分边缘块和双向注意力聚合块,强化特征表征能力。DNA-Net 通过密集嵌套交互模块与通道-空间注意力模块协同实现^[3]。Wang K 等人提出的 IAA-Net 网络采用由粗到精的内部注意力机制^[20]。

最近,随着大模型技术的成功,越来越多的研究者采用视觉 Transformer 技术。比如,Liu F 等人利用自注意力机制实现大范围特征交互^[21],Chen G 等人提出分层视觉 Transformer 架构^[22]。尽管取得诸多进展,但是红外小目标检测仍面临严峻挑战——目标尺寸小、信噪比低、形态动态变化以及复杂背景干扰等因素导致现有算法仍存在漏检和虚警现象。

2 所提方法

在红外小目标检测问题中,对于一帧红外视频来说,大部分区域是背景,而目标只占据少量像素。直接从原始序列中提取时空特征会耗费大量不必要的计算时间和资源,所以本文先对每帧进行空间分析,然后进行视频帧间分析。红外目标在图像中所占的像素非常少,甚至只有一个。对于视频图像目标检测问题,图像或目标的分辨率越高,就能提取出越有效的语义特征;反之,目标分辨率越低,感受野中包含的背景信息越多,目标特征就越难提取。因此,本文对输入的红外图像进行插值以提高分辨率。尽管红外无人机图像的外观各式各样,但由于成像距离较远,目标的外观已经变得模糊,所以将所有无人机目标全部视为同一类别。

2.1 网络结构

输入一系列帧序列后,需要定位红外目标并输出其位置。由于目标分辨率较低,对每个像素的预测正确与否都对算法检测性能起着关键作用。因此,本文将小目标检测任务视为逐像素预测问题。基于集合预测的目标检测算法通过以下方式实现:利用 N 个查询向量分析图像或视频;使用视觉 Transformer 对特征进行解码;通过多层感知器计算一组向量对,其中每对向量包含类别预测嵌入和掩码或边界框嵌入。

采用基于集合预测的目标检测框架设计红外目标检测算法,并通过联合分析视频帧中的红外小目标特征来实现红外小目标检测与分割任务。

本文所提红外目标检测算法的整体框架如图 2 所示。输入视频帧序列 $I_1, I_2, \dots, I_T$; 主干网络逐帧提取特征,得到空间分辨率逐层降低、通道数目逐层增多、语义信息逐层增强的多尺度深度特征集合 $F_1, F_2, \dots, F_T$ 。空间分辨率的降低不利于对小目标的识别,因此像素解码器对深度特征逐级插值,获得具有与输入图像相同空间分辨率的特征 $M_{t=1}^T \in R^{C \times T \times H \times W}$ 。其中, H 和 W 表示特征的高和宽(空间尺寸),也是视频帧的高和宽; C 表示特征通道数。随后,帧解码器从一组可学习的帧查询向量 $f_{t=1}^T \in R^{T \times C_f \times N_f}$ 解析特征。接着,目标解码器构建 T 帧标记集合 $f_{t=1}^T$ 的时序关系, N_v 个可训练视频查询向量 $v \in R^{C_v \times N_v}$ 解码目标实例的信息,模型为标准 Transformer 编码器-解码器结构。视频查询向量对应视频目标的时空掩码。网络结构的输出包含两支:类别分支由单层线性分类器组成,直接预测每个视频查询向量的类别概率 $p \in R^{N_v \times (K+1)}$, 这里 K 表示目标类别总数;掩码分支生成融合多帧信息的目标掩码嵌入向量 $w_v \in R^{C_v \times N_v}$, 目标掩码嵌入向量 w_v 和逐像素嵌入特征 $M_{t=1}^T$ 经过矩阵相乘获得最终预测的目标掩码 $m \in R^{N_v \times T \times H \times W}$ 。

像素解码器的目标是获取用于分割的高分辨率特征, 这对于小目标检测尤其重要。本文采用多尺度可变形注意力层解析主干网络 C5、C4 和 C3 阶段的特征。在顶层特征图上, 使用双三次插值和卷积层来生成每个像素的嵌入。

目标编码器汇集所有帧的目标标记 $f_{t=1}^T$, 沿着时间轴的多头自注意力模型对不同时序的标记进行通信。

2.2 红外小目标检测

2.2.1 离线训练

本文构建的红外目标检测网络结构需要采用离线有监督学习。通过匈牙利匹配算法完成预测值与真实值之间的最优配对, 进而构建目标函数。损失函数公式如下:

$$L_{\text{total}} = \lambda_f L_f + \lambda_v L_v + \lambda_{\text{sim}} L_{\text{sim}} \quad (1)$$

式中, 超参数 $\lambda_f = 1$, $\lambda_v = 1$, $\lambda_{\text{sim}} = 0.50$ 。 L_f 函数用于计算单帧预测结果差异, 由分类损失函数和掩码相关损失函数两部分组成: 分类损失函数为交叉熵; 掩码相关损失函数包括二元

交叉熵和 Dice 损失函数^[23]。 L_v 函数用于计算最终预测结果与目标真值之间的差异, 损失函数和 L_f 函数相同。 L_{sim} 函数用于按编号对视频查询和帧查询进行聚类。具体而言, 收集与目标真值标签匹配的帧查询和视频查询。依据矩阵乘法测量所有收集查询的相似性, 在估计的相似度与真实情况之间计算二值交叉熵。其中, 标注为 1 的是编号相同的配对, 标注为 0 的则是编号不同的配对。为了训练红外小目标检测模型, 首先收集整理现有的公开红外数据集, 包括 NUST-SIRST^[6]、NUAA-SIRST^[7]、NUDT-SIRST^[3]、IRST640^[22]以及 IRSTD-1k^[8]数据集。经筛选后, 获得 13471 张训练图像(含 15473 个目标)和 385 张验证图像(含 556 个目标)。图像分辨率范围从 101×96 到 640×512 像素, 目标分辨率范围从 1 到 1065 像素, 目标/全帧像素比在 1.52×10^{-6} 至 0.01 之间。广泛的数据分布旨在提升模型泛化能力(见表 1)。

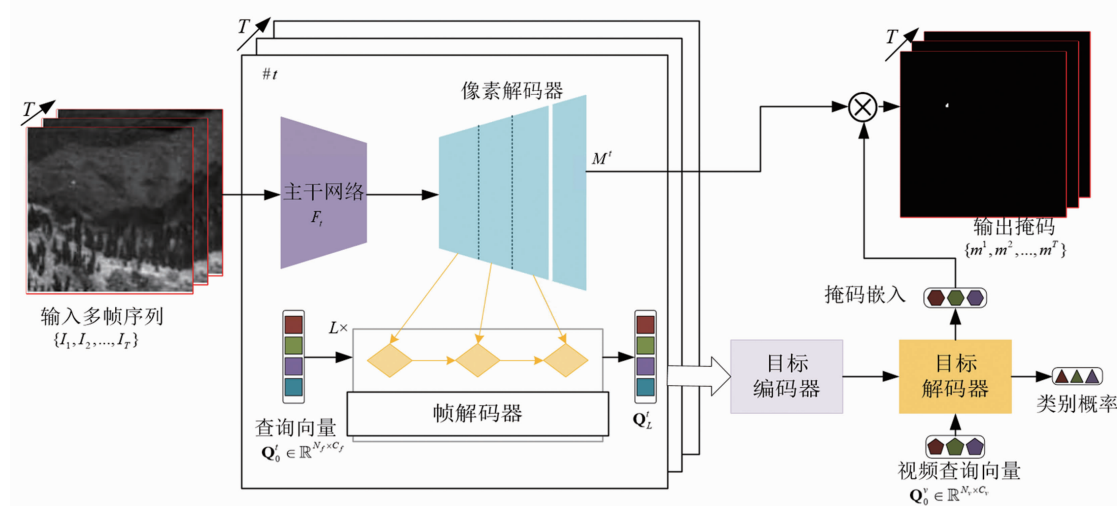


图 2 本文所提红外目标检测算法的整体框架

表 1 用于离线训练的红外数据集

数据集名称	出处	训练集数目/条	验证集数目/条
NUST-SIRST	文献[6]	9978	100
NUAA-SIRST	文献[7]	341	85
NUDT-SIRST	文献[3]	1327	0
IRST640	文献[22]	1024	0
IRSTD-1k	文献[8]	801	200

本文采用在图像分类任务(ImageNet 数据集)上预训练的残差网络 ResNet-50 作为主干网络。设置超参数帧查询向量的数目 $N_f = 100$, 视频查询向量的数目 $N_v = 100$ 。训练时选择 AdamW 优化器。训练过程分为以下四步:

(1) 迁移学习。基于 COCO 数据集预训练的模型, 使用收集的红外图像进行微调。

(2) 单帧训练。输入帧数设为 $T=1$, 在自定义红外数据集上训练 $36k$ 次迭代, 学习率在 $32k$ 和 $35k$ 时衰减, 初始学习率为 0.0001 。

(3) 伪视频生成。通过随机仿射变换将单幅图像裁剪 4 次(尺寸范围为 $[384, 600]$), 生成长度 $T=4$ 的伪视频。

(4) 伪视频训练。设置 $140k$ 次迭代, 学习率在 $75k$ 时衰减, 初始学习率为 0.00005 , 每帧短边调整为 360 像素。

2.2.2 在线推断

在线推断过程中, 整个视频序列被划分为长度为 T 的不重叠视频片段。本文逐个处理每个视频片段, 选择具有高置信度的查询向量对应的预测的掩码, 然后对这些掩码逐元素求和, 输出最终的分割结果。将分类概率和掩码表示为 $z = (p_i, m_i)_{i=1}^{N_{\text{target}}}$, 最终目标掩码为

$$m = \sum_{i=1}^{N_{\text{target}}} (p_i > th) \cdot m_i \quad (2)$$

式中, N_{target} 表示被分类为目标查询的数量, $th = 0.5$ 。通过二值图像连通区域分析获得目标中心坐标。本文采用帧间差分辅助红外目标检测。具体而言, 计算序列中两个相邻帧之间的差分:

$$\Delta(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \quad (3)$$

$$I_{FD}^t(x, y) = \frac{\Delta(x, y)}{\max_{x, y}(\Delta)} \times 255 \quad (4)$$

式中, $I_t(x, y)$ 表示帧 I_t 在坐标像素 (x, y) 处的灰度值。对于第一帧, 计算 $\Delta(x, y) = |I_1(x, y) - I_2(x, y)|$ 。通过该方式获得的帧差图像可消除大量背景并突显目标。本文将帧差序列作为输入发送到训练好的模型。模型进行前向推断后输出每帧检测结果。

3 实验

实验中使用 2 块 NVIDIA GeForce RTX-3090 GPU 以及 PyTorch 框架。为验证方法效果, 采用 DSAT 红外序列数据集^[4]进行评估。该数据集包含天空和地面背景下的 22 段红外序列, 共 16177 帧, 涵盖 16944 个固定翼无人机目标。帧分辨率为 256×256 像素, 目标分辨率为 $1 \sim 400$ 像素, 采用点标注法(目标由中心坐标表示)。

为评估算法, 对于检测结果 d , 根据相应目标标注 gt 评估检测性能。若 $d \in N_3(gt)$, 表示精准定位出目标, 则正确且精确的检测结果个数 N_{accurate} 累加一次。这里, $N_3(\cdot)$ 表示像素 3×3 邻域。若 $d \notin N_3(gt) \cap d \in N_9(gt)$, 表示检测结果不在目标的 3×3 邻域内, 而在目标的 9×9 邻域内, 此时能够定位到目标, 但是还不够精确, 则正确但不精确的检测结果个数 N_{inexact} 累加一次。这里, $N_9(\cdot)$ 表示像素 9×9 邻域。如果在检测结果 d 的邻域内没有对应的目标标注 gt , 即检测结果 d 是虚警, 那么检测到的假目标个数(虚警) N_{false} 累加一次。对于目标标注 gt , 如果其邻域内没有检测结果 d , 那么表明该目标被遗漏, 未检测到的目标个数(漏检) N_{missed} 累加一次。基于 N_{accurate} 、 N_{inexact} 、 N_{missed} 以及 N_{false} , 获得最终评价指标——召回率 R 、精度 P 以及 F 得分:

$$R = N_{\text{right}} / N_{\text{total}} \times 100\% \quad (5)$$

$$P = \frac{N_{\text{right}}}{N_{\text{right}} + N_{\text{false}}} \times 100\% \quad (6)$$

$$F = 2 \times \frac{P \cdot R}{P + R} \times 100\% \quad (7)$$

式中, $N_{\text{right}} = N_{\text{accurate}} + N_{\text{inexact}}$ 。

3.1 红外小目标检测

表 2 列出了各种算法在 DSAT 数据集上的检测结果。对于 MDvsFA-CGAN 算法^[6], 从 <https://github.com/ZhexuanZhou/MDvsFA> 获取代码并重新训练。对于 IAANet 算法^[21], 使用 <https://github.com/kwwcv/iaanet> 的代码和模型权重。DNA-Net 算法^[3]的代码和模型权重来自

<https://github.com/YeRen123455/InfraredSmall-Target-Detection>。AGPCNet 算法^[24]的代码和预训练模型取自 <https://github.com/TianfangZhang/AGPCNet>。ACM 算法^[7]的代码和预训练模型取自 <https://github.com/TianfangZhang/acm-pytorch>。基于静态图像的红外小目标检测算法逐帧处理 DSAT 数据集。IAANet 算法^[20]通过输出矩形框定位目标, 其他算法及本文算法输出二值图像, 通过连通区域分析^[25]获取目标位置。

如表 2 所示, 本文算法的召回率为 0.6599, 高于 AGPCNet 算法(召回率为 0.3709)和 IMNN-LWEC 算法(召回率为 0.4515), 与 ACM 算法(召回率为 0.6712)、

DNA-Net 算法(召回率为 0.6275)、MFSTPT 算法(召回率为 0.6595)相当; 虽然低于 NOLC 算法(召回率为 0.9515)、MDvsFA-CGAN 算法(召回率为 0.8694)、IAANet 算法(召回率为 0.8109)和 SRSTT 算法(召回率为 0.9606), 但本文算法虚警极少, 精度得分为 0.6356, 远高于 NOLC 算法(精度为 0.0163)、MDvsFA-CGAN 算法(精度为 0.0603)、IAANet 算法(精度为 0.0928)和 SRSTT 算法(精度为 0.0255)。因此, 本文算法取得最高综合得分, F 得分为 0.6475, 远超其他方法。

3.2 可视化结果

图 3 所示为可视化结果的对比情况。在原始图像中, 黄色框表示目标标注。各种算法的

表 2 在 DSAT 数据集上的检测效果对比

算法	召回率	精度	F 得分
NOLC ^[26]	0.9515	0.0163	0.032
MDvsFA-CGAN ^[6]	0.8694	0.0603	0.1127
ACM ^[7]	0.6712	0.177	0.2801
DNA-Net ^[3]	0.6275	0.4641	0.5336
IAANet ^[20]	0.8109	0.0928	0.1666
AGPCNet ^[24]	0.3709	0.0584	0.1008
IMNN-LWEC ^[15]	0.4515	0.0774	0.1322
MFSTPT ^[5]	0.6595	0.1661	0.2653
SRSTT ^[16]	0.9606	0.0255	0.0496
本文算法	0.6599	0.6356	0.6475

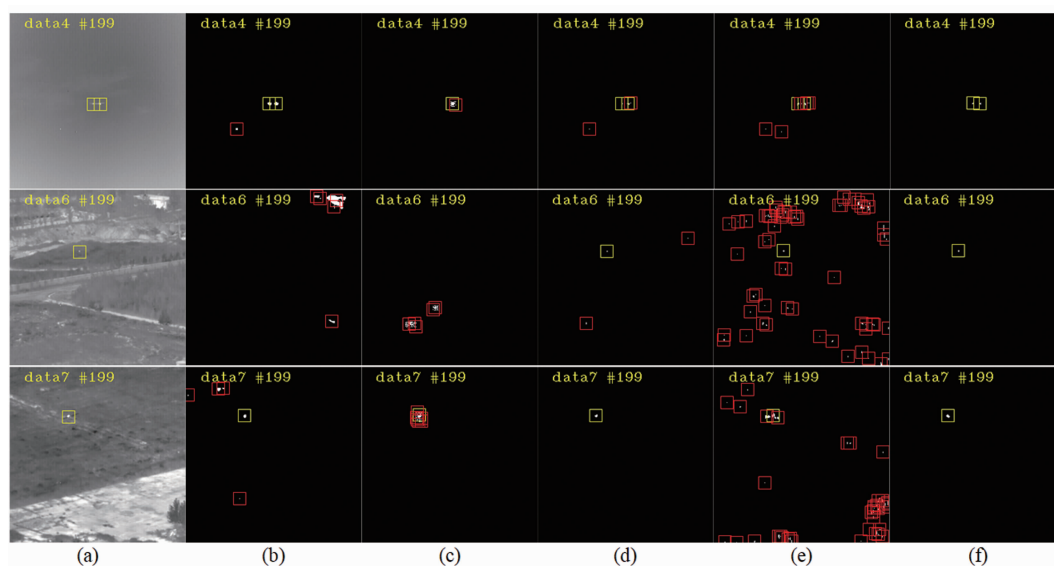


图 3 可视化结果:(a)原始图像;(b) AGPCNet;(c) IMNN-LWEC;(d) MFSTPT;(e) SRSTT;(f)本文算法

可视化结果中,黄色框表示正确检测,红色框表示误报。可以看出,SRSTT 算法虽能检测目标,但产生大量误报,误将背景噪声判为目标。AGPCNet、IMNN-LWEC 和 MFSTPT 等算法既产生误报,又漏检真实目标。相比之下,本文算法能够准确检测目标且误报极少。

4 结束语

本文针对红外小目标检测任务因目标少像素、弱纹理、无颜色而导致的识别难题,提出了一种面向空中无人机的红外小目标检测算法。该算法在 DSAT 数据集上取得了 0.6475 的 F 得分,相较于现有方法有显著的性能提升。实验结果表明,通过多尺度特征融合与时序建模能够有效提升红外小目标检测的精度。本研究不仅在特征提取阶段引入了多尺度可变形注意力机制以增强小目标的细节表征,而且还在时空建模方面进行了创新,通过视频级查询向量生成目标实例的时空掩码,从而更精准地捕捉目标的时空信息。

然而本研究仍存在一些尚未解决的问题。比如,在处理更复杂的背景干扰时,如何有效剔除虚警仍是待解决的问题。未来的研究可以考虑引入更多的先验知识,如目标的运动模型和环境的物理特性,以进一步提升算法的适应性和鲁棒性。此外,在嵌入式系统中,如何在保持高精度的同时降低算法的计算复杂度以实现更高效的实时检测,也是值得深入研究的方向。

参考文献

- [1] Chen C L P, Li H, Wei Y T, et al. A Local Contrast Method for Small Infrared Target Detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(1): 574–581.
- [2] 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 Processing*, 2013, **22**(12): 4996–5009.
- [3] Li B, Xiao C, Wang L, et al. Dense Nested Attention Network for Infrared Small Target Detection [J]. *IEEE Transactions on Image Processing*, 2023, **32**: 1745–1758.
- [4] Hui B, Song Z, Fan H, et al. A Dataset for Infrared Detection and Tracking of Dim-Small Aircraft Targets Under Ground/Air Background [J]. *China Scientific Data*, 2020, **5**(3): 74.
- [5] Hu Y X, Ma Y P, Pan Z X, et al. Infrared Dim and Small Target Detection from Complex Scenes via Multi-Frame Spatial-Temporal Patch-Tensor Model [J]. *Remote Sensing*, 2022, **14**(9): 2234.
- [6] Wang H, Zhou L, Wang L. Miss Detection vs. False Alarm: Adversarial Learning for Small Object Segmentation in Infrared Images [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [7] Dai Y, Wu Y, Zhou F, et al. Asymmetric Contextual Modulation for Infrared Small Target Detection [C]. Waikoloa: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021.
- [8] Zhang M, Zhang R, Yang Y, et al. ISNet: Shape Matters for Infrared Small Target Detection [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [9] Mohanty N C. Computer Tracking of Moving Point Targets in Space [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1981, **PAMI-3**(5): 606–611.
- [10] Barniv Y. Dynamic Programming Solution for Detecting Dim Moving Targets [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1985, **AES-21**(1): 144–156.
- [11] Blostein S D, Huang T S. Detecting Small, Moving Objects in Image Sequences Using Sequential Hypothesis Testing [J]. *IEEE Transactions on Signal Processing*, 1991, **39**(7): 1611–1629.
- [12] Xiong Y, Peng J X, Ding M Y, et al. An Extended Track-before-detect Algorithm for Infrared Target Detection [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1997, **33**(3): 1087–1092.

- [13] Grossi E, Lops M, Venturino L. Track-before-detect for Multiframe Detection with Censored Observations [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2014, **50**(3): 2032–2046.
- [14] Wang J, Yi W, Kirubarajan T, et al. An Efficient Recursive Multiframe Track-Before-Detect Algorithm [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2018, **54**(1): 190–204.
- [15] Luo Y, Li X, Chen S, et al. IMNN-LWEC: A Novel Infrared Small Target Detection Based on Spatial-Temporal Tensor Model [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1–22.
- [16] Li J, Zhang P, Zhang L, et al. Sparse Regularization-Based Spatial-Temporal Twist Tensor Model for Infrared Small Target Detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1–17.
- [17] 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.
- [18] Huang L, Dai S, Huang T, et al. Infrared Small Target Segmentation with Multiscale Feature Representation [J]. *Infrared Physics & Technology*, 2021, **116**: 103755.
- [19] Chen F, Gao C, Liu F, et al. Local Patch Network with Global Attention for Infrared Small Target Detection [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2022, **58**(5): 3979–3991.
- [20] Wang K, Du S, Liu C, et al. Interior Attention-Aware Network for Infrared Small Target Detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1–13.
- [21] Liu F, Gao C, Chen F, et al. Infrared Small and Dim Target Detection with Transformer Under Complex Backgrounds [J]. *IEEE Transactions on Image Processing*, 2023, **32**: 5921–5932.
- [22] Chen G, Wang W, Tan S.IRSTFormer: A Hierarchical Vision Transformer for Infrared Small Target Detection [J]. *Remote Sensing*, 2022, **14**(14): 3258.
- [23] Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [C]. Stanford: Fourth International Conference on 3D Vision (3DV), 2016.
- [24] Zhang T, Li L, Cao S, et al. Attention-Guided Pyramid Context Networks for Detecting Infrared Small Target Under Complex Background [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2023, **59**(4): 4250–4261.
- [25] Burger W, Burge M J. Principles of Digital Image Processing: Core Algorithms [M]. London: Springer, 2009.
- [26] Zhang T, Wu H, Liu Y, et al. Infrared Small Target Detection Based on Non-Convex Optimization with L_p -Norm Constraint [J]. *Remote Sensing*, 2019, **11**(5): 559.