DOI:10.3979/j.issn.1673-825X.202405060112

基于记忆池的红外序列小目标检测

陈　林[1],高陈强[1],黄　骁[2]
(1.重庆邮电大学 通信与信息工程学院,重庆 400065;2.中国舰船研究设计中心,武汉 430064)

摘　要:为了缓解数据稀缺问题,收集并标注了两个红外跟踪的数据集用于序列小目标检测,命名为 ATR-ISTD、UAV-ISTD;提出了一种融入记忆池的序列小目标检测网络,有效利用前后帧关联信息,通过查询帧与记忆帧之间的记忆匹配读取内存信息,解决红外小目标检测在高杂波背景下的高虚警和低准确率的问题。针对下采样造成的小目标特征丢失问题,设计了前向语义引导融合模块(pre-semantic guided fusion module,PSGF)来整合不同尺度特征;在记忆向量编码器中设计了伪标签引导的特征增强模块(pseudo label guided feature enhancement module,PLG-FE)来强化小目标的局部特征表达能力。实验结果表明,与当前主流单帧目标检测方法相比,提出的方法在降低虚警率方面取得了显著成效,分别在 ATR-ISTD 和 UAV-ISTD 数据集上实现了 16.87%和 10.49%的改善,在目标级 F_1 上提高了 4.89%和 6.54%,在像素级 F_1 上提高了 7.69%和 11.63%。

关键词:红外图像;小目标检测;记忆池;特征增强;特征融合
中图分类号:TP391　　**文献标志码**:A　　**文章编号**:1673-825X(2025)05-0769-12

Infrared sequence small target detection based on memory pool

CHEN Lin[1], GAO Chenqiang[1], HUANG Xiao[2]
(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China;
2. China Ship Development and Design Center, Wuhan 430064, P. R. China)

Abstract: To alleviate the scarcity of data, this paper collects and annotates two infrared tracking datasets for sequential small target detection, named ATR-ISTD and UAV-ISTD. This paper proposes a sequential small target detection network integrating a memory pool, which effectively utilizes the correlation information between frames before and after, reads memory information through memory matching between the query frame and the memory frame, and solves the problems of high false alarm and low accuracy in infrared small target detection under high clutter background. To reduce the loss of small target features caused by downsampling, a forward semantic guided fusion module (PSGF) is designed to integrate features of different scales. In the memory vector encoder, a pseudo label guided feature enhancement module (PLG-FE) is designed to enhance the local feature expression ability of small targets. Experimental results show that, compared with mainstream single-frame detection methods, the proposed method significantly reduces false alarm rates, achieving improvements of 16.87% and 10.49% on the ATR-ISTD and UAV-ISTD datasets, respectively. Target-level F1 scores increased by

收稿日期:2024-05-06　**修订日期**:2025-06-30　**通讯作者**:陈　林　632559679@qq.com
基金项目:国家重点研发项目(2022YFA1004100)
Foundation Item: National Key Research and Development Program (2022YFA1004100)

4.89% and 6.54%, and pixel-level F1 scores improved by 7.69% and 11.63%.

Keywords: infrared image; small target detection; memory pool; feature enhancement; feature fusion

0 引 言

红外小目标检测(infrared small target detection, ISTD)是红外成像分析领域一项关键技术之一,具有广泛的应用,如交通管理[1]、海上监视[2]等。红外成像技术通过捕捉物体自身发出的红外辐射成像,在夜间、雨雾、低光照等复杂条件也能保持出色表现,具有较强的抗干扰能力[3-4]。与一般目标检测不同,红外小目标检测具有独有的挑战性特征:①目标尺寸极小。红外小目标通常在远距离下成像,其在图像中的实际尺寸相对较小,通常目标的像素尺寸小于 30×30。②缺乏纹理信息。在红外图像中,小目标仅占图像的小部分,并且通常以点状形式出现,几乎不含有纹理信息,这使得可供分析和检测的信息非常有限。③背景干扰的问题严重。红外成像的特性使得在多变的环境尤其在城市场景中,背景热源与目标热源难以区分。

在 ISTD 的早期研究中,主要依赖于模型驱动的方法[5-20],这类传统方法往往高度依赖于先验知识,不能较好地适应多变的复杂场景。随着深度学习在自然图像的成功[21-22],研究者们开始探索将深度学习方法应用于红外小目标检测研究。基于数据驱动的深度学习方法显示出了特征学习潜力[23-32],能够更好地适应复杂多变的环境,模型鲁棒性显著提升。基于深度学习的单帧检测方法研究尽管在红外小目标检测领域已取得较为显著的进展,但在应对视频序列场景时无法利用时域信息,因此在复杂场景中的表现有待提高。针对这一挑战,本文提出一种新的红外序列小目标检测方法,利用连续帧之间的关联信息捕捉小目标的运动轨迹,显著提升了小目标在复杂场景中的检测准确性。此外,为了缓解红外序列小目标检测数据集稀缺的问题,本文收集并标注了来自红外目标跟踪领域的两个不同数据集[33-34],形成了两个可用于红外序列小目标检测的数据集 ATR-ISTD 和 UAV-ISTD。

考虑到小目标的尺寸较小,本文没有选择传统的目标检测方法用边界框来定位目标,而是将小目标检测视为一个分割问题[35-37]。本文选用 U-Net 作为网络基础结构,通过编码器提取特征。为了尽可能减少在特征提取过程中的局部信息的丢失,本文设计了前向语义引导融合模块(pre-semantic guided fusion module, PSGF),该模块能够有效整合全局和局部的信息,提高检测准确性。此外,在编码器与解码器之间设计了一个记忆池,实现查询帧与记忆帧之间的特征交互,以更好地定位小目标。考虑到小目标的特征可利用性相对较少,且在高维空间中易被背景噪声干扰,在记忆向量编码器中设计了伪标签引导的特征增强模块(pseudo label guided feature enhancement module, PLG-FE),突出小目标特征区域表示,减小背景噪声的干扰。

综上所述,本文的贡献可以归纳如下:①提出一种新的红外小目标检测方法,通过记忆池的记忆信息匹配来合理利用帧间信息,能够在视频序列中准确地捕捉和追踪红外小目标,在复杂背景和动态场景中展现出良好的性能。②设计前向语义引导融合模块,减少了下采样过程中的局部信息丢失;设计伪标签引导的特征增强模块,增强小目标局部特征。③整理了两个红外跟踪的数据集,专用于红外序列小目标检测。

1 相关工作

1.1 传统红外小目标检测

1.1.1 基于背景抑制的方法

文献[5]利用密度峰搜索和最大灰度区生长的检测方法,来检测不同尺寸的小目标,并消除各种复杂形状杂波造成的干扰。文献[6-7]提出 Top-Hat 方法和 Max-Mean/Max-Median 方法,通过抑制简单且均匀的背景信息从而突出显示小目标。文献[8]从局部图像分割的角度,通过新型的局部描述符实现对杂波的抑制和小目标增强。

1.1.2 基于人类视觉系统

通常,一个像素块中心区域与其周围的像素值亮度存在差异,文献[9]提出一种基于滑动窗口的局部对比方法(local contrast method, LCM),通过测量每个中心像素与邻域像素值的对比度作为提取特征信息。后续不断有学者在 LCM 的基础上提出改进方法,以检测不同尺寸目标,并同时有效增强真实目标和抑制各种类型干扰。文献[10]提出一种改进的增强局部对比测量方法,用来生成更加准确的显著性图。

1.1.3 基于局部块模型分解的方法

文献[11]提出了红外斑块图像(infrared patch

image,IPI)方法,利用红外背景图像的非局部自相关特性,通过恢复低秩矩阵,实现小目标与背景分离。文献[12]利用序列图像的时域信息构建时空块图像,然后采用马尔可夫随机场引导的混合高斯噪声模型进行小目标检测建模。最后,通过贝叶斯方法从复杂背景中分离出小目标分量。尽管如此,由于目标和背景分布差异较大,上述传统方法在实际应用中面临着一定的挑战。

红外序列小目标检测的研究主要集中在以模型驱动的传统方法。文献[16]对红外视频进行时空相关性分析,利用目标周围的边缘和角点信息,从背景的稀疏残差中突出显示目标。文献[17]通过为不同的奇异值赋予不同权重,自适应地准确估计背景信息。文献[18]提出利用滑动窗口机制的方法,逐个提取图像块,以此捕获相邻图像中的非重叠块信息。文献[19]提出一种稀疏正则化扭转张量模型,充分利用红外图像的时空信息,强化模型对小目标的捕捉能力。文献[20]提出将三维向量扩展为四维张量,通过张量列和张量环技术分解为低维张量,将 ISTD 问题建模为稀疏加低秩分解问题,处理空间和时间信息的不平衡。但是,传统的序列检测方法,不能主动进行特征学习,比较依赖于先验知识,在复杂场景中检测虚警率较大。

1.2　深度学习红外小目标检测

深度学习方法在红外小目标检测领域有着广泛应用。文献[23]利用生成对抗网络来平衡漏检和虚警。文献[24]设计了非对称上下文调制模块,同时结合自上而下的全局上下文反馈和自下而上的调制通路,实现了在深层语义和浅层细节之间的信息交换。文献[25]提出一种从全局出发关注局部的网络,融合多尺度特征进行小目标检测。文献[26]提出了密集嵌套注意网络,实现高层信息和底层信息之间的渐进交互,并保持小目标在深层的信息。文献[27]提出新型的红外形状网络,以此关注小目标的精确边缘和形状。文献[28]设计了基于 Transformer 的鲁棒的、通用的红外小目标检测方法。文献[29]设计了 UIU(U-net In U-net)中内嵌 UIU 的结构,实现多层次多尺度的学习 UIU 结构。文献[30]为了解决背景与目标之间的类别不平衡问题,设计了感受野和方向注意力网络。文献[31]设计了一种新的损失函数,通过采用更简洁的模型,能够更有效地关注小目标的尺度和位置敏感性。尽管如此,这些单帧检测方法直接应用于序列检测,由于不能充分利用时序特征,检测效果较差。文献[32]提出一种新的序列检测框架 ST-Trans,它利用时空变换模块来学习前后帧的依赖信息,取得不错的效果。

2　本文方法

2.1　网络整体结构

图 1 为本文设计的红外序列小目标检测网络。

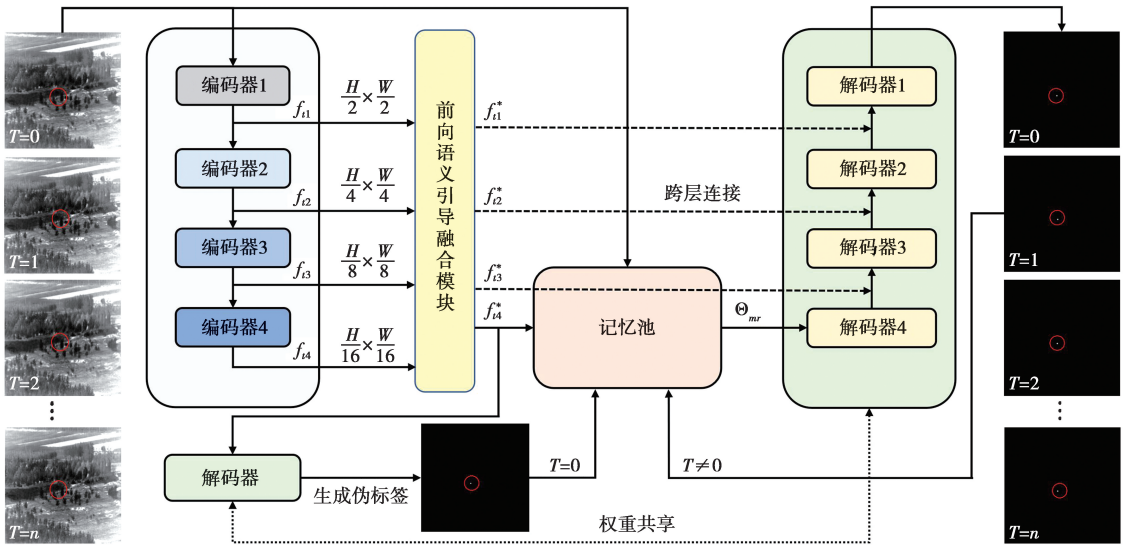

图 1　基于记忆池的红外序列小目标检测网络结构图

Fig.1　Infrared sequence small target detection network structure diagram based on memory pool

该网络以 U-Net 架构为基础,结合编码器和解码器进行构建;使用 Resnet50[38] 从输入的红外视频 V 中提取特征,V 包含一个长度为 t 的视频帧,即 $V=\{f_1,f_2,f_3,\cdots f_t\}$,经过编码后,得到 4 个不同尺度的特征 f_{ti},$i\in\{1,2,3,4\}$,代表不同尺度的索引,这一过程可以表示为

$$f_{t_i}=\text{Resnet50}(V_t)\quad t\in\{1,2,3\} \tag{1}$$

式(1)中,Resnet50 为 50 层残差网络。随后,f_{ti} 被送入 PSGF,该模块综合不同尺度的特征信息,最大程度减少下采样过程中小目标信息丢失,进一步提升深层小目标的特征表达,这一过程可以表示为

$$f_{ti}^{*}=\text{PSGF}(f_{ti})\quad t\in\{1,2,3,4\} \tag{2}$$

式(2)中,f_{ti}^{*} 表示经过 PSGF 融合后的特征映射。紧接着,前三个尺度的特征通过跨层连接与解码器输出的特征图结合,f_{t4}^{*} 记作 $\boldsymbol{Y}_2$ 被送入记忆池中。当时间帧 $T=0$ 时,复用解码器为第 0 帧生成伪标签,伪标签作为引导信息用来引导后续帧的分割。在记忆池中,记忆向量编码器会对输入的信息进行编码,生成记忆向量 $\boldsymbol{Y}_1$。为了更好地突显小目标特征,嵌入在记忆向量编码器中的伪标签引导的特征增强模块会对 $\boldsymbol{Y}_2$ 和记忆向量 $\boldsymbol{Y}_1$ 做特征增强,以提高小目标检测的性能,增强后的记忆向量为 $\boldsymbol{V}_m$。

总体来说,本文网络通过记忆池进行当前帧与记忆帧之间的关联信息匹配,得到当前帧在记忆池中的强关联信息匹配结果;随后,将记忆池中匹配得到的结果与解码器输出的特征图进行拼接;逐步上采样进行解码,最终实现对小目标的准确分割。

2.2 前向语义引导融合模块

随着卷积深度的增加,在不断的下采样过程中,现有网络小目标特征会被淹没。受到文献[39]的启发,本文设计了一个 PSGF 来实现全局和局部信息的融合,在捕获全局信息的同时,避免了小目标局部信息的丢失。PSGF 如图 2 所示。其融合操作可以表示为

$$\begin{cases} f_{t1}^{*}=f_{t1}+\sigma(f_{t1})U(f_{t2})+\sigma(f_{t1})U(f_{t2})+\\ \qquad\sigma(f_{t1})U(f_{t2})\\ f_{t2}^{*}=D(f_{t1})+f_{t2}+\sigma(f_{t2})U(f_{t3})+\sigma(f_{t2})U(f_{t4})\\ f_{t3}^{*}=D(f_{t1})+D(f_{t2})+f_{t3}+\sigma(f_{t3})U(f_{t4})\\ f_{t4}^{*}=D(f_{t1})+D(f_{t2})+D(f_{t3})+f_{t4} \end{cases} \tag{3}$$

式(3)中:$D(\cdot)$ 表示下采样;$U(\cdot)$ 表示上采样;σ 表示使用 Sigmoid 函数得到对应层特征权重。图 2 展示了 PSGF 的融合过程,对于 4 个不同的尺度 f_{t1},f_{t2},f_{t3},f_{t4} 均会融合其他尺度的信息。如对于 f_{t1}^{*},首先会利用 σ 生成 f_{t1} 的权重值,接着 f_{t2}、f_{t3}、f_{t4} 上采样之后分别与生成的权重值相乘,之后与 f_{t1} 相加,其他三个尺度的操作同理。最终 f_{ti}^{*} 结合来自不同尺度特征的信息,通过将深层信息和浅层信息进行自适应融合,实现信息的互补与增强,提高模型对目标的检测能力。

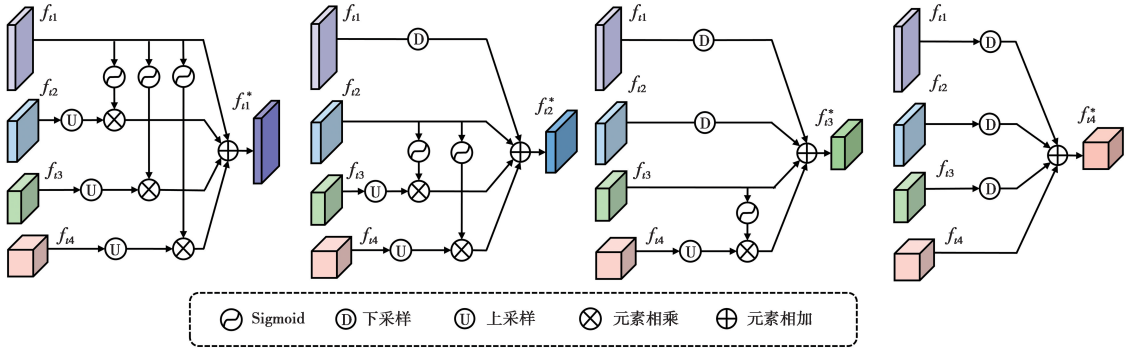

图 2 前向语义引导融合模块

Fig.2 Pre-semantic guided gusion module

2.3 记忆向量编码器

为了记忆当前帧信息,利用该信息帮助后续帧进行分割,本文设计了记忆向量编码器,如图 3 所示。首先将当前帧的图像信息和当前帧的伪标签通道叠加,然后用编码器进行特征编码,生成记忆向量 $\boldsymbol{Y}_1$。为了在局部增强目标特征的映射区域,突出小目标的显示,减小背景噪声对小目标区域的影响,本文在记忆向量编码器中设计了一个基于伪标签引导的特征增强模块,它的输入是 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$。为了充分挖掘 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 特征的通道关系,首先对输入做一个中间特征变换,操作为

$$\varphi(Y_i)=\sigma(\vartheta_1(\mathrm{Cat}(\mathrm{Maxpool}(Y_i),\mathrm{Avgpool}(Y_i)))) \tag{4}$$

式(4)中:ϑ_1 为 1×1 卷积,Maxpool 和 Avgpool 分别代表最大池化和平均池化。为了在空间上增强目标信息,我们将变换后的结果与原始特征图进行逐点相乘,对原始特征图进行加权,加权的权重为中间变换的结果,$\boldsymbol{Y}_i$ 增强的操作表示为

$$\boldsymbol{Y}'_i=\boldsymbol{Y}_i\otimes\varphi(\boldsymbol{Y}_i) \tag{5}$$

式(5)中,⊗表示逐点相乘。在通道上叠加增强后的特征 $\boldsymbol{Y}'_i$ 后做 3×3 卷积进一步整理特征,ϑ_3 代表 3×3 卷积。这一过程可以表示为

$$\boldsymbol{Y}_3=\vartheta_3(\mathrm{Cat}(\boldsymbol{Y}_i,\boldsymbol{Y}'_i)) \tag{6}$$

为了学习特征之间的非线性关系,将 $\boldsymbol{Y}_3$ 输入到由多个全连接组成的多层感知网络中,然后将得到的权重值与原始特征图 $\boldsymbol{Y}_3$ 逐点相乘得到增强后的记忆向量 $\boldsymbol{V}_{\mathrm{m}}$,操作可表示为

$$\boldsymbol{V}_m=\boldsymbol{Y}_3\otimes\sigma(\mathrm{MLP}(\mathrm{Maxpool}(\boldsymbol{Y}_3))+\mathrm{MLP}(\mathrm{Avgpool}(\boldsymbol{Y}_3))) \tag{7}$$

通过这个过程,网络可以综合利用原始和增强后的特征信息,并通过 MLP 网络的学习调整,实现对特征的进一步优化和调整,实现对小目标的准确检测。

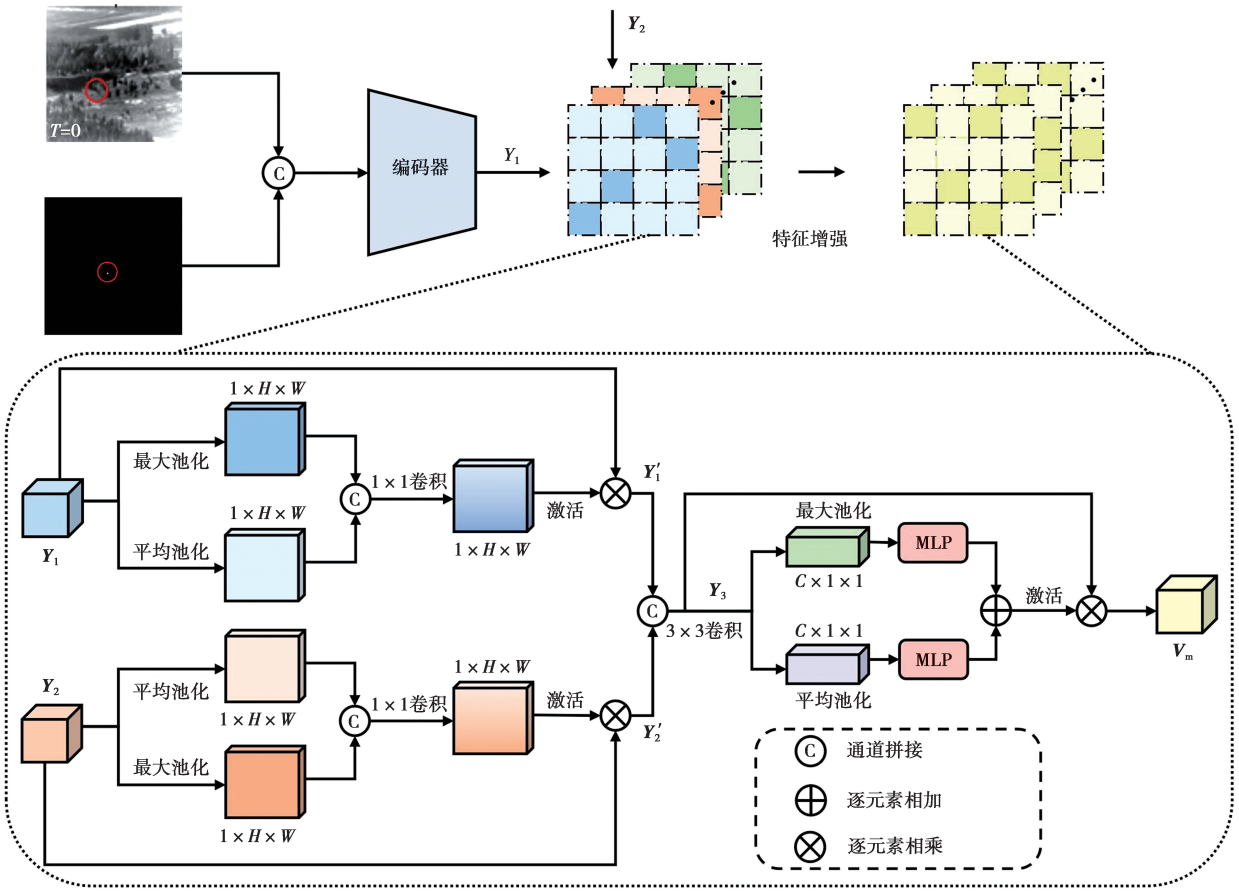

图 3　记忆向量编码器和伪标签引导的特征增强模块

Fig.3　Memory vector encoder andpseudo-label guided feature enhancement module

2.4　记忆匹配策略

在记忆池中,本文基于注意力机制读取记忆的结果,这是一种高效的记忆读取策略。通过注意力机制,模型会根据查询的内容自适应地从记忆池中提取最相关的信息。首先,查询向量的特征映射 $\boldsymbol{K}_{\mathrm{q}}$ 与记忆池中记忆向量特征映射 $\boldsymbol{K}_{\mathrm{m}}$ 进行矩阵乘法操作得到记忆矩阵 $\boldsymbol{W}$,其中特征映射通过卷积实现。对该记忆矩阵 $\boldsymbol{W}$ 应用 Softmax 函数,得到记忆权重矩阵 $\boldsymbol{W}_O$,记忆权重矩阵记录着每个记忆向量对于当前查询的重要性,得到记忆权重的这一操作过程可以表示为

$$W_o=\mathrm{Softmax}(K_q\otimes K_m) \tag{8}$$

得到 $\boldsymbol{W}_O$ 后,将 $\boldsymbol{W}_O$ 与 $\boldsymbol{V}_{\mathrm{m}}$ 相乘得到记忆匹配结果 $\boldsymbol{M}_{\mathrm{r}}$,$\boldsymbol{W}_O$ 中的每个权重代表了对应记忆向量在当前

查询上下文中的相关性。$\boldsymbol{V}_m$ 对于当前查询帧来说是记忆池存储的实际信息,它们相乘的结果反映了当前查询与记忆池中信息的匹配程度。将 $\boldsymbol{M}_r$ 与查询值向量 $\boldsymbol{V}_q$ 进行整合得到在记忆池中最终查询结果 $\boldsymbol{\Theta}_{mr}$,具体操作为

$$\boldsymbol{\Theta}_{mr} = \mathrm{Cat}(\boldsymbol{M}_r, \boldsymbol{V}_q) \tag{9}$$

通过记忆整合,$\boldsymbol{\Theta}_{mr}$融合了当前帧信息和记忆帧的历史信息。通过记忆匹配策略,模型不仅可以利用当前帧信息,还可以有效利用历史帧的信息来增强当前帧的处理,充分利用了小目标运动信息,在复杂视频场景中,能够有效检测小目标。

2.5 损失函数

本文将小目标检测问题转化为逐像素的二分类任务,选择二值交叉熵损失函数作为本文的损失函数。二值交叉熵损失是评估二分类任务预测精度的常用方法,通过比较网络输出的二值化图像与真实标签之间的差异来计算损失,定义为

$$\mathbf{Loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log p(y_i) + (1 - y_i)\log(1 - p(y_i)) \tag{10}$$

式(10)中:y_i 表示一个二值标签,$y_i \in \{0,1\}$;$p(y_i)$ 是预测给定标签的概率 $p(y_i) \in \{0,1\}$;N 表示图像中的总像素数。

3 实验分析

3.1 实验数据集

鉴于红外小目标检测的特殊性,目前适用于此类检测任务的公开数据集数量相对较少。目前,用于单帧红外小目标检测的主要数据集包括 NUAA-SIRST、MFIRST、IRSTD-1k。其中,NUAA-SIRST 数据集提供了一定数量的图像,但总量有限;MFIRST 数据集主要包含合成图像;IRSTD-1k 数据集则提供了 1 000 多张真实场景中不同目标的图像,这些目标在大小和形状上各不相同,并覆盖了多种不同的场景,这些数据集的引入极大地推动了单帧红外小目标检测技术的发展。当前存在多帧数据集主要是 IRSTD[32],不过其采用的数据标注形式是数据框标注。基于此,本文从红外图像弱小飞机目标跟踪数据集和红外无人机跟踪数据集中筛选出了运动红外小目标图像,并对其进行了像素级目标的标注,最终构建了 ATR-ISTD 和 UAV-ISTD 两个可用于红外序列小目标检测的数据集,如表 1 所示。

表 1 ATR-ISTD 和 UAV-ISTD 数据集介绍

Tab.1 Introduction to ATR-ISTD and UAV-ISTD datasets

指标	ATR-ISTD	UAV-ISTD
图像尺寸	256×256	640×512
图像类型	真实	真实
标注类型	掩码	掩码
图像帧数量	11 243	11 900
背景类型	田野、雨林	湖泊、城市 天空、云层

3.2 评价指标

本文用文献[11]中的评价指标 F_a 来衡量预测中的虚警率,以 N_f 表示在整个视频中检测到的虚假目标总数,N_I 代表视频中图像帧的总数。F_a 值是评估小目标检测性能的重要指标,它直接关联到检测系统的可靠性和实用性。F_a 定义为

$$F_a = \frac{N_f}{N_I} \tag{11}$$

同时为了全面评价预测的精度,还采用了准确率 P(Precision),召回率 R(Recall),目标级 F_1^t 和像素级 F_1^p 来综合评价。准确率衡量被模型预测为正例的样本中真正为正例的样本个数,准确率越高表明被误判的数量越少,定义为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{12}$$

R 反应真正为正例的样本有多少被模型正确识别,定义为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{13}$$

通过计算准确率和召回率,我们能够使用它们的调和平均数,目标级 F_1^t 和像素级 F_1^p 来综合评估模型的性能。F_1^t 和 F_1^p 定义为

$$F_1^t = \frac{2 \times P_t \times R_t}{P_t + R_t} \tag{14}$$

$$F_1^p = \frac{2 \times P_p \times R_p}{P_p + R_p} \tag{15}$$

式(12)—(13)中,N_{TP} 为真正例;N_{FP} 为假正例;N_{FN} 为假反例。

3.3 网络训练细节

本文基于 Pytorch 框架,在 NVIDIA GeForce RTX 3090 显卡上迭代了 30 000 次;以 Adam 优化器来训练模型,参数为 $\beta_1 = 0.9$,$\beta_2 = 0.999$;设置初始学

习率为 10^{-5},Batch size 为 8,每个 Batch 数据为视频序列中随机有序的 3 帧,3 帧之间的采样间隔按照 3→6→9→12→9→3 进行变化。训练开始前,对训练样本进行随机翻转增强处理,然后统一将图像尺寸调整为 256×256。为了证明本文方法的优势,本文与几种最先进的方法进行比较,如表 2 所示。表 2 中,最优结果加粗显示。所有方法均使用作者公开提供的源代码,且在默认参数下进行。

表 2　不同方法在 ATR-ISTD 和 UAV-ISTD 数据集上的结果

Tab.2　Results of different methods on ATR-ISTD and UAV-ISTD datasets

方法	ATR-ISTD 数据集							UAV-ISTD 数据集						
	目标级别			像素级别			F_a	目标级别			像素级别			F_a
	P	R	F_1^t	P	R	F_1^p		P	R	F_1^t	P	R	F_1^p	
IPI[11]	45.68	86.83	59.86	53.63	26.30	35.29	103.24	8.72	17.20	11.57	21.11	0.65	1.27	183.76
NRAM[13]	42.89	63.17	51.09	51.57	12.32	19.89	84.13	10.09	21.83	13.80	20.79	0.74	1.43	198.21
NOLC[14]	39.00	47.38	42.78	48.44	10.11	16.73	74.11	7.43	12.70	9.38	19.69	0.56	1.10	158.15
PSTNN[15]	41.52	89.37	56.70	46.70	25.51	33.00	125.89	5.28	9.89	6.88	14.02	0.37	0.71	181.15
ECA[19]	57.17	90.90	70.20	68.78	25.70	37.41	68.10	12.37	12.45	12.41	20.88	0.26	0.52	89.93
4D[20]	58.36	**91.99**	71.41	70.91	29.79	41.96	65.64	39.8	48.57	43.75	60.54	1.27	2.48	74.93
MDvsFA[23]	50.93	40.79	45.30	42.64	38.50	40.46	39.30	57.15	42.64	48.84	71.72	51.42	59.9	31.97
LPNET[25]	67.35	61.76	64.43	53.77	57.92	55.77	29.97	52.72	53.54	53.13	66.54	45.94	54.36	48.97
ISNET[27]	70.92	70.26	70.59	62.17	47.87	54.09	28.81	74.14	59.37	65.94	73.64	54.74	62.80	21.06
ISTUC[28]	76.71	88.98	82.39	57.66	59.45	58.54	27.02	73.54	65.51	69.29	58.16	71.29	64.06	24.00
UIUNET[29]	64.04	66.41	65.20	57.25	55.80	56.52	37.29	59.14	60.28	59.71	63.18	54.76	58.67	41.73
RDIAN[30]	81.97	87.9	84.83	68.04	56.99	62.02	19.34	81.31	63.76	71.47	67.45	61.34	64.25	14.67
MSHNET[31]	82.40	88.37	85.28	60.28	56.77	58.47	18.88	70.58	63.40	66.80	59.72	64.85	62.18	26.45
本文方法	**97.13**	**83.36**	**89.72**	**76.66**	**62.00**	**68.56**	**2.47**	**94.23**	**68.24**	**79.16**	**75.54**	**76.23**	**75.88**	**4.18**

3.4　与已有方法对比结果

3.4.1　定量结果对比

由表 2 可见,在 ATR-ISTD 和 UAV-ISTD 数据集上,本文的方法在 F_1^t、F_1^p、F_a 三个指标上均取得了最好的性能,这是由于本文充分利用了帧间关联信息。具体来看,传统的单帧方法在准确性上不及传统序列方法。前者 F_1^t 最高为 59.86 和 13.80,F_1^p 最高为 35.29 和 1.43,而后者 F_1^t 最高为 71.41 和 43.75,F_1^p 最高为 41.96 和 2.48。在 F_a 评价上后者最优为 65.64和 74.93,也明显优于传统单帧方法。这是由于序列方法可以整合连续帧之间的信息。

对比传统序列方法和深度学习单帧方法,在 ATR-ISTD 数据集上,传统序列方法 F_1^t 得分超过了 MDvsFA、LPNET、UIUNET 等深度学习单帧方法。然而,在复杂环境下传统方法往往无法有效抑制虚警。这主要是因为传统方法无法主动学习和适应目标的复杂特征,在背景噪声和光照变化等因素的干扰下尤其如此。整体上传统方法的 F_a 都高于深度学习方法。ECA 和 4D 在 ATR-ISTD 数据集上,均表现出准确率较低、召回率较高的特点。这可能缘于传统序列方法不能很好地应对非目标特征的干扰,而错误地将这些非目标区域检测为目标。ISTUC 和 RDIAN 通过融合多尺度信息,结合注意力机制,能够从更广泛的视角捕捉小目标的特征,但是这些单帧方法从本质上讲没有利用前后帧的关联信息。相较之下,本文方法在准确率和召回率之间实现了较好的平衡,所以 F_1^t,F_1^p 均表现不错的效果。通过有效利用前后帧的信息,本文方法能够更准确地区分小目标和背景噪声。

此外,传统方法在不同数据集上的表现存在明显差异,且所有传统方法在 UAV-ISTD 数据集上的表现极为不佳,这进一步说明传统方法往往依赖于特定的先验特征,缺乏足够的鲁棒性和适应性,在多变的应用场景中受限。此外,传统方法在不同数据集上的表现存在明显差异,且所有传统方法在 UAV-ISTD 数据集上的表现极为不佳,这进一步说明传统方法往往依赖于特定的先验特征,缺乏足够的鲁棒性和适应性,在多变的应用场景中受限。

3.4.2 定性结果对比

不同 ISTD 方法在数据集中部分场景的定性比较结果如图 4 所示。图 4 中,图 4e 场景来自于 ATR-ISTD 数据集,其他所有场景均来自于 UAV-ISTD 数据集。红色为正确检测到的目标,绿色为漏检的目标,橙色为错误的检测,右上角矩形区域为放大的目标。由图 4 可见,本文的方法能够在多种复杂背景下准确地检测到目标,并且在大多数场景中极少出现虚警现象。

图 4a 场景中,由于背景干扰较少,红外小目标的特征相对更为明显,大多数现有方法都能实现较为准确的检测。图 4b 场景中,水流和波纹在红外图像中形成了复杂的纹理模式,此时传统方法 4D、ECA 和 IPI 均未能成功检测到目标;UIU 虽能识别小目标,但水面的反射造成周围区域高亮,导致虚警的产生;RDIAN 方法不仅未能检测到小目标,还出现了虚警。图 4c 场景中所有方法都能检测到小目标,但 IPI、4D、ECA 和 UIU 在不同程度上产生了虚警。总体而言,基于深度学习的方法因其能够学习小目标的特征,相对于传统的单帧和序列方法,虚警的情况有所改善。图 4d 场景中,复杂的光照条件对小目标检测造成了干扰,除本文的方法外,其他所有方法都出现了较为严重的虚警问题,错误地将其他光源识别为小目标。在图 4e 和图 4f 这样的野外背景下,由于地形的不均匀和植被的遮挡,传统方法 IPI、4D、ECA 均表现较差,产生了大量的虚假目标,UIU 和 RDIAN 漏检也比较严重,UIU 仅仅微弱地检测到目标。

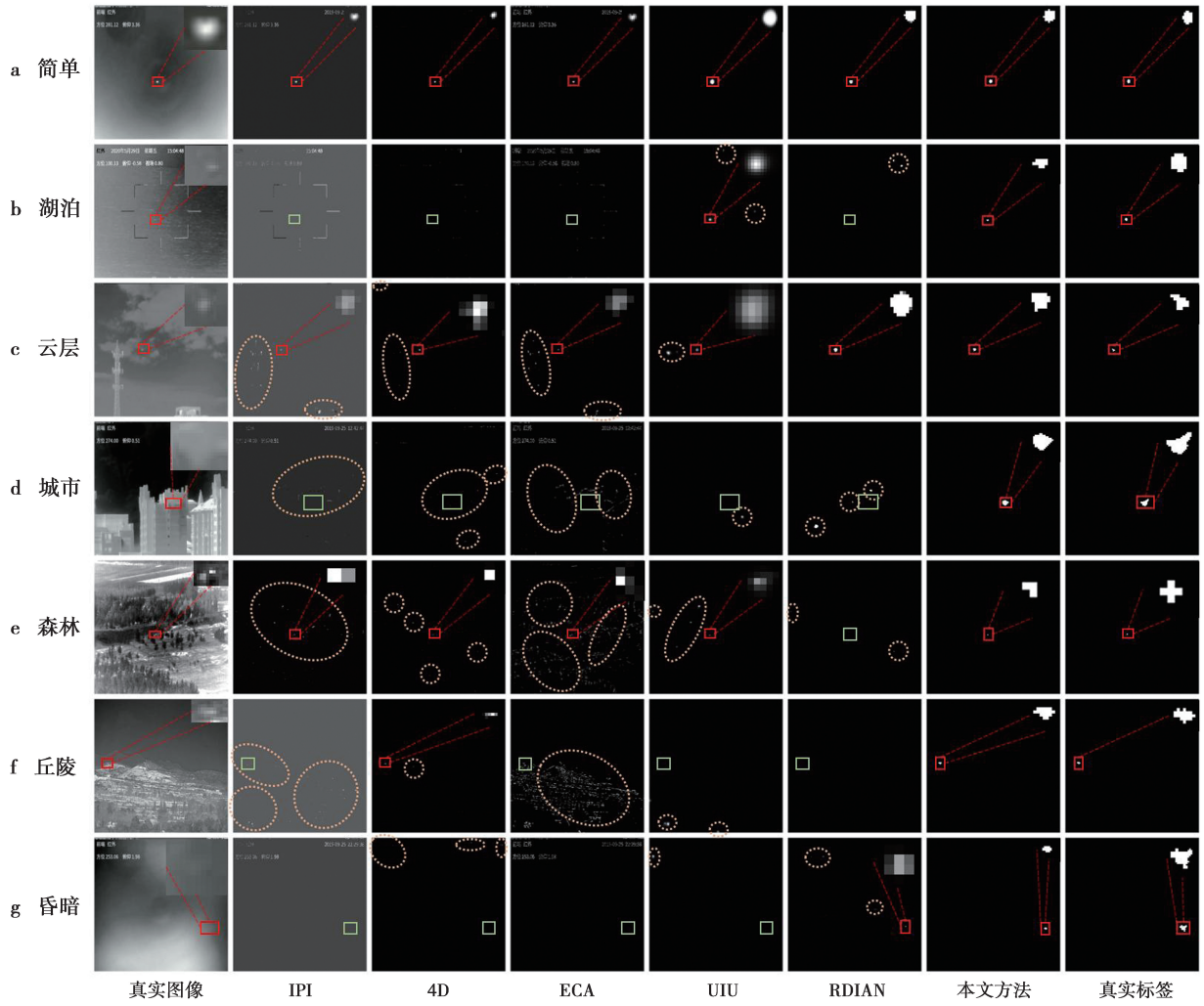

图 4 不同方法在 ATR-ISTD 和 UAV-ISTD 数据集上的定性结果对比

Fig.4 Comparison of qualitative results of different methods on ATR-ISTD and UAV-ISTD data sets

综上所述,在简单环境中大多数方法表现良好,但在复杂背景下,尤其是有动态纹理、复杂光照和遮挡等因素存在时,传统方法和一些深度单帧的学习方法的性能均受到挑战,而本文方法在处理复杂背

景下小目标检测任务时具有优越性和鲁棒性。

3.5　消融实验

3.5.1　前向语义引导融合模块的影响

深层语义学习捕捉了直观感知不到的信息。在高维抽象空间中,背景噪声可能会掩盖小目标的特征,导致其被遗忘。为了解决这一问题,本文利用浅层信息进行引导,结果如表 3 所示。由表 3 可见,运用本文设计的引导融合模块进行特征学习可以显著提升小目标检测的性能。这验证了 PSGF 模块抑制小目标信息在下采样过程中的丢失以及在不同尺度感受野之间利用信息进行检测方面的有效性。本文设计了 3 种其他融合的方式,说明了 PSGF 模块的合理性和有效性。直接融合方式,没有为向上采样的语义信息赋予指导权重,上采样的过程中无法避免深度衰减问题,没有充分利用浅层信息来帮助深层信息的恢复;单向融合方式仅利用了低于当前尺度的信息,缺少了更深层语义信息;在未使用 PSGF 模块时,各网络层之间的信息交流不足,F_1^t 和 F_1^p 的指标均大幅度下降,检测性能较差。

3.5.2　伪标签引导的特征增强模块的影响

为了增强小目标的局部特征表示,本文设计了伪签引导的特征增强模块,消融实验结果如表 4 所示。由表 4 可见,当 PLG-FE 模块中的池化采用串行设计时,在抑制虚警方面,在 UAV-ISTD 数据集上的 F_a 降为 31.12,明显低于最优结果,检测结果不理想。因此,在本文的增强模块中选择并行方式同时实现最大池化和平均池化,以在捕获关键特征的同时减少信息损失。$\boldsymbol{Y}_1$ 通过较浅的残差网络编码获得,它富含局部和基本特征信息。仅对 $\boldsymbol{Y}_1$ 进行特征增强可能会丢失深层次的语义信息,影响模型在复杂背景中的识别能力。$\boldsymbol{Y}_2$ 代表了从更深层网络编码得到的当前帧特征,包含丰富的深层语义信息。如果仅增强 $\boldsymbol{Y}_2$,模型可能会忽视细节信息,从而无法精确定位小目标。在 UAV-ISTD 数据集中同时增强 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$,各指标表现均表现最佳。然而,ATR-ISTD 数据集中仅增强 $\boldsymbol{Y}_1$ 时 $\boldsymbol{F}_a$ 表现更好。这是因为 ATR-ISTD 主要涉及野外场景,且这些场景的语义复杂多样,$\boldsymbol{Y}_2$ 中的一些深层语义信息可能包含与目标无关的特征,这些特征在语义上容易与目标混淆。当同时增强 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 时,$\boldsymbol{Y}_2$ 中这些易混淆的语义信息可能会干扰目标的准确判断,从而导致 F_a 检测效果不如仅增强 $\boldsymbol{Y}_1$。尽管如此,$\boldsymbol{Y}_1$ 的局部基本特征和 $\boldsymbol{Y}_2$ 的深层语义特征仍有一定的互补性。当选择同时增强 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 时,尽管 $\boldsymbol{F}_a$ 略有下降,但 $\boldsymbol{F}_1^t$ 和 $\boldsymbol{F}_1^p$ 均显著提升。该增强模块通过结合 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 使得模型能够同时利用详细的局部信息和丰富的高级语义信息,更全面地挖掘图像特征,从而提高小目标检测的准确性。

表 3　PSGF 模块在 ATR-ISTD 和 UAV-ISTD 数据集上的结果

Tab.3　Results of PSGF module on ATR-ISTD and UAV-ISTD datasets

方法	ATR-ISTD 数据集			UAV-ISTD 数据集		
	F_1^t	F_1^p	F_a	F_1^t	F_1^p	F_a
直接融合	87.64	64.36	15.56	73.99	72.68	3.30
单向融合	87.00	66.21	10.79	74.77	72.20	4.60
不使用 PSGF	76.04	59.57	3.24	76.41	70.79	7.15
使用 PSGF	**89.72**	**68.56**	**2.47**	**79.16**	**75.88**	**4.18**

表 4　PLG-FE 模块在 ATR-ISTD 和 UAV-ISTD 数据集上的结果

Tab.4　Results of PLG-FE module on ATR-ISTD and UAV-ISTD datasets

方法	ATR-ISTD 数据集			UAV-ISTD 数据集		
	F_1^t	F_1^p	F_a	F_1^t	F_1^p	F_a
串行设计	85.16	62.26	3.16	72.12	67.04	31.12
仅增强 Y1	84.38	68.42	**2.39**	75.14	72.84	6.82
仅增强 Y2	88.54	65.79	4.93	72.62	72.16	7.48
不使用 PLG-FE	87.90	63.27	3.85	71.29	66.93	31.79
使用 PLG-FE	**89.72**	**68.56**	2.47	**79.16**	**75.88**	**4.18**

3.5.3 跨帧信息交互策略的影响

记忆池中存储着不同帧之间的信息交互,不同帧与帧之间的信息融合得到不同的检测性能,为了找到合适的帧间信息融合策略,本文设计了跨帧信息交互策略来验证不同帧数量检测性能对比。训练时采样间隔的变化满足

$$f(n,t)=\begin{cases} n\times t & t\leqslant 3 \\ n\times(5-t) & t>3 \end{cases} \tag{16}$$

式(16)中,n 是策略编号,$n\in(1,5)$;t 是训练迭代次数的阶段编号,$t\in(1,4)$;测试时的采样间隔变化满足

$$g(n)=n \tag{17}$$

这意味着对于策略 n,每隔 n 帧存入一帧记忆帧。本文中,通过记忆池控制帧间信息的交互,采用不同的记忆管理策略来优化小目标的检测性能。跨帧信息交互策略的影响如表 5 所示。由表 5 可见,交互策略 3 的跨帧信息交互在性能上最为出色。这表明,在构建记忆池时,并非简单地增加记忆帧就能带来更好的效果,过多的记忆帧反而会消耗更多计算资源,增加整体的计算负担。此外,记忆间隔的设置也需在合理的范围之内。小目标在连续运动中相邻几帧的信息往往高度相似,如果记忆间隔设置得太小,在进行注意力匹配时,模型就可能会过度集中在临近帧的信息上,导致早期的记忆信息被忽略,因此,策略 1 和策略 2 相比于策略 3 在准确率和误报方面的表现较为逊色。反之,若记忆间隔过大,则帧与帧之间的关联性会减弱,模型可能无法有效捕捉到小目标的运动轨迹,影响记忆特征的匹配效果,当采用策略 4 或 5 时,由于记忆帧逐渐变多 F_1^{t}、F_1^{p}、F_{a} 都呈现下降趋势。总之,表 5 结果强调了为小目标检测精心设计记忆池管理策略的重要性和合理性。

表 5 跨帧信息交互策略在 ATR-ISTD 和 UAV-ISTD 数据集上的结果

Tab.5 Results of cross-frame information interaction strategy on ATR-ISTD and UAV-ISTD datasets

方法	ATR-ISTD 数据集			UAV-ISTD 数据集		
	F_1^{t}	F_1^{p}	F_{a}	F_1^{t}	F_1^{p}	F_{a}
交互策略 1	84.54	63.90	7.01	73.39	72.55	6.00
交互策略 2	85.87	67.69	10.48	74.33	72.20	8.27
交互策略 3	**89.72**	**68.56**	2.47	**79.16**	**75.88**	**4.18**
交互策略 4	87.74	67.80	**2.31**	77.03	71.93	11.79
交互策略 5	84.26	64.76	7.94	75.02	73.35	5.64

4 结束语

本文提出了一种新方法来处理复杂背景下的红外序列小目标检测任务,使用记忆池来存储并匹配各帧之间的关系,通过帧间信息交互很好捕捉了小目标的运动信息。本文还设计了前向语义融合模块和伪标签引导的特征增强模块,前者重点考虑全局和局部信息之间的差异,减少下采样过程中信息的丢失;后者专注于在局部增强小目标的特征表示,减小背景噪声对小目标区域的影响。此外,本文还收集了两个可用于红外序列小目标检测的数据集,供未来研究使用。在这两个数据集上的大量实验表明,与现有单帧检测方法相比,本文方法更为有效。

参考文献:

[1] YING X, LIU L, WANG Y, et al. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 15528-15538.

[2] ZHAO M, L W, LI L, et al. Single-frame infrared small-target detection: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(2): 87-119.

[3] ZHANG T, PENG Z, WU H, et al. Infrared small target detection via self-regularized weighted sparse model[J]. Neurocomputing, 2021, 420: 124-148.

[4] ZHU H, ZHANG J, XU G, et al. Balanced ring top-hat transformation for infrared small-target detection with guided filter kernel[J]. IEEE Transactions on Aerospace and Electronic Systems, 2020, 56(5): 3892-3903.

[5] HUANG S, PENG Z, WANG Z, et al. Infrared small target detection by density peaks searching and maximum-gray region growing[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(12): 1919-1923.

[6] ZENG M, LI J, PENG Z. The design of top-hat morpho-

logical filter and application to infrared target detection [J]. Infrared Physics & Technology, 2006, 48(1): 67-76.

[7] DESHPANDE S D, ER M H, VENKATESWARLU R, et al. Max-mean and max-median filters for detection of small targets[C]//Signal and Data Processing of Small Targets 1999. Denver, CO, USA: SPIE,1999: 74-83.

[8] QIN Y, BRUZZONE L, GAO C, et al. Infrared small target detection based on facet kernel and random walker [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 7104-7118.

[9] CHEN C L P, LI H, WEI Y, et al. A local contrast method for small infrared target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 574-581.

[10] HAN J, MA Y, ZHOU B, et al. A robust infrared small target detection algorithm based on human visual system [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(12): 2168-2172.

[11] 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.

[12] GAO C, WANG L, XIAO Y, et al. Infrared small-dim target detection based on Markov random field guided noise modeling[J]. Pattern Recognition, 2018(76): 463-475.

[13] ZHANG L, PENG L, ZHANG T, et al. Infrared small target detection via non-convex rank approximation minimization joint l 2, 1 norm[J]. Remote Sensing, 2018, 10(11): 1821.

[14] ZHANG T, WU H, LIU Y, et al. Infrared small target detection based on non-convex optimization with Lp-norm constraint[J]. Remote Sensing, 2019, 11(5): 559.

[15] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. Remote Sensing, 2019, 11(4): 382.

[16] ZHANG P, ZHANG L, WANG X, et al. Edge and corner awareness-based spatial - temporal tensor model for infrared small-target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(12): 10708-10724.

[17] LIU T, YANG J, LI B, et al. Nonconvex tensor low-rank approximation for infrared small target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021(60): 1-18.

[18] WANG G, TAO B, KONG X, 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): 1-17.

[19] 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.

[20] WU F, YU H, LIU A, et al. Infrared small target detection using spatio-temporal 4D tensor train and ring unfolding[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023(61): 1-22.

[21] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015: 234-241.

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc, 2017: 5998-6008.

[23] WANG H, ZHOU L, WANG L. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, KOR:IEEE, 2019: 8509-8518.

[24] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Wailuku, HI, USA: IEEE, 2021: 950-959.

[25] 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.

[26] LI B, XIAO C, WANG L, et al. Dense nested attention network for infrared small target detection [J]. IEEE Transactions on Image Processing, 2022(32): 1745-1758.

[27] ZHANG M, ZHANG R, YANG Y, et al. ISNet: Shape matters for infrared small target detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 877-886.

[28] 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.

[29] WU X, HONG D, CHANUSSOT J. UIU-Net: U-Net in

U-Net for infrared small object detection[J]. IEEE Transactions on Image Processing, 2022(32): 364-376.

[30] 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.

[31] LIU Q, LIU R, ZHENG B, et al. Infrared small target detection with scale and location sensitivity[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2024: 17490-17499.

[32] TONG X, ZUO Z, SU S, et al. ST-Trans: Spatial-temporal transformer for infrared small target detection in sequential images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024(62): 1-19.

[33] 回丙伟,宋志勇,王琦,等.空中弱小目标检测跟踪测试基准[J].航空兵器,2019,26(6):56-59.
HUI B W, SONG Z Y, WANG Q, et al. A benchmark for dim or small aircraft targets detection and tracking[J]. Aero Weaponry,2019, 26(6): 56-59.

[34] HUANG B, LI J, CHEN J, et al. Anti-UAV410: A thermal infrared benchmark and customized scheme for tracking drones in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(5): 2852-2865.

[35] OH S W, LEE J Y, XU N, et al. Video object segmentation using space-time memory networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE,2019: 9226-9235.

[36] CHENG H K, TAI Y W, Tang C K. Rethinking space-time networks with improved memory coverage for efficient video object segmentation[J]. Advances in Neural Information Processing Systems, 2021(34): 11781-11794.

[37] LI M, HU L, XIONG Z, et al. Recurrent dynamic embedding for video object segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA : IEEE, 2022: 1332-1341.

[38] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.

[39] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3349-3364.

作者简介:

陈 林,硕士,主要研究方向为红外小目标检测。E-mail: 632559679@qq.com。

高陈强,教授,博士,主要研究方向为图像处理、视频分析、机器学习。E-mail:gaochq6@mail.sysu.edu.cn。

黄 骁,高级工程师,博士,主要研究方向为智能无人平台感知规划决策。E-mail:huangxiao_88@outlook.com。

(编辑:田海江)