

脉冲与卷积深度融合的红外弱小目标检测



李乐潇,纪禄平,陈圣嘉,朱思成,郭朝祥

电子科技大学,四川 成都 611731

摘要:随着无人机和其他飞行器技术的快速发展,机载红外弱小目标检测成为重要研究领域。然而,由于目标距离远、体积微小或部分遮挡,传统方法在检测过程中面临巨大挑战。因此,提出更为精确的目标检测算法显得尤为重要。本文提出了一种融合脉冲和卷积技术的新型机载红外弱小目标检测模型,即脉冲增强的融合特征网络(SEFFN)模型。与以往的基于深度学习的红外弱小目标检测算法不同,该模型通过仿生的脉冲神经网络结构增强与弱小目标检测相关的特征表达,从而更精确地提取小目标区域。SEFFN 主要包含4个关键模块:扩展金字塔卷积(DPC)、双路注意力融合(DAF)、多脉冲增强(MSE)和监督注意力(SAM)。这些模块共同作用,提高了对小目标区域的关注,同时在排除噪声干扰的前提下保留小目标的主要特征,显著提升了模型的检测性能。在两个数据集上的试验结果表明,与现有的模型驱动和数据驱动方法相比,SEFFN 的表现更为出色,特别是在 Sirst Aug 数据集上,其 F 和 mIoU 分别达到 85.74% 和 75.04%。这表明 SEFFN 在红外弱小目标检测任务中的有效性和优势。SEFFN 适用于航空领域,能够部署在飞行器的边缘设备上,提高空中平台对远距离目标的检测能力,支持远程监控、反制敌方威胁等任务。

关键词:红外弱小目标;脉冲神经网络;目标检测;深度学习;注意力机制

中图分类号:TP391

文献标识码:A

DOI: 10.19452/j.issn1007-5453.2025.01.002

近年来,机载红外弱小目标检测算法的重要性越发凸显,航空领域的一些后续任务,如监视、侦查、救援和飞行制导^[1]等都需要有更加准确识别和跟踪目标的能力。但是由于红外成像的特点和空中平台远距离采集图像的需求,关注的目标在红外图像中通常表现为尺寸小、亮度弱、缺乏明显的纹理和形状信息^[2],因此非常有必要根据红外弱小目标的特点,有针对性地设计高效的检测方法。

单帧红外弱小目标检测算法是红外目标跟踪检测的重要基础,主要分为模型驱动方法和数据驱动方法。早期的模型驱动算法大都是由先验知识指导设计滤波器或模块,利用滤波器估计背景,从而抑制背景,突出目标^[3]。Deshpande 等^[4]研究了最大均值和最大中值滤波器,两个滤波器分别利用最大均值和最大中值滤波处理原始图像,然后从图像中减去滤波的输出以增强目标区域。受到人类视觉系统启发而被提出的基于局部对比度的方法,以及同样

能抑制背景的低秩稀疏分解方法。Asrabadi 等^[5]根据综合光谱特征构建了冗余的目标字典,研究了以稀疏表示分类器为基础的目标检测算法。但是这些模型依赖领域知识,对模型参数和场景变化敏感,在复杂场景和低对比度条件下表现不佳。

近年来,随着深度学习技术在计算机视觉领域的发展,研究者们将深度学习引入红外弱小目标检测领域,以实现自动提取特征和高效检测,从而得到更好的检测性能。McIntosh 等^[6]改进了经典的两阶段和一阶段的目标检测算法(即 Faster-RCNN^[7]和 YOLOv3^[8]),实现了一个红外弱小目标检测网络,该网络能够优化其输出对目标和杂波响应比。Dai Yimian 等^[9]提出了非对称上下文调制模块(ACM),其中的自上而下的全局上下文反馈路径和基于点级通道注意的自下而上调制路径互相补充,能够实现高层和低层的语义交换和突出小目标。Li Boyang 等^[10]提出了密集嵌套

收稿日期: 2024-04-22; 退修日期: 2024-07-31; 录用日期: 2024-09-27

基金项目: 航空科学基金(2022Z071080006)

引用格式: Li Lexiao, Ji Luping, Chen Shengjia, et al. Deep fusion of spiking and convolutional networks for infrared small target detection[J]. Aeronautical Science & Technology, 2025, 36(01): 11-21. 李乐潇, 纪禄平, 陈圣嘉, 等. 脉冲与卷积深度融合的红外弱小目标检测[J]. 航空科学技术, 2025, 36(01): 11-21.

注意网络(DNA),其中的密集嵌套交互模块和空间注意模块能够实现逐级特征融合和自适应的特征增强,避免小目标在深层网络中丢失。Wang Huan等^[11]提出的条件生成对抗网络受到对抗性学习的启发,将红外目标检测转换为平衡漏检(MD)和虚警(FA)问题,经过对抗学习使MD和FA达到最佳平衡,从而提高检测性能。上述已有的工作在一定程度上解决了红外弱小目标检测的难点,但是由于飞行器采集的图像场景复杂多样,以上方法还欠缺对背景杂波的抑制,容易错误地将噪声识别为目标,以及对目标的分割不够精确,影响最终检测效果。

针对以上问题,本文提出了脉冲增强的融合特征网络模型(SEFFN),用来解决目前机载红外弱小目标检测技术不够精确的问题。在SEFFN中,尝试引入脉冲神经网络结构,构建多脉冲增强模块(MSE)实现对小目标特征的增强。脉冲神经网络作为第三代神经网络,具有低能耗、高效率、易实现等特点,更接近大脑的生物结构,其使用脉冲序列作为载体处理和传递信息,在训练过程中所需参数少,操作简便。此外,SEFFN中还采用了双路注意力融合模块(DAF)用于自适应融合红外小目标在深层和浅层的空间位置和语义信息特征,以及为解决小目标丢失问题设计的扩展金字塔卷积模块(DPC)。为了解决噪声干扰,本文提出的监督注意力模块(SAM)从全局角度增强对小目标区域的关注。经过试验验证,SEFFN具有更好的检测性能。

1 研究背景

1.1 机载红外弱小目标检测的应用背景

图1为机载红外弱小目标检测场景示意图,图中的预警机搭载红外探测设备,能够识别多种空中目标,包含导弹、直升机、无人机、战斗机等。图1左下角为红外探测设备捕获的红外弱小目标图像,红框中为待检测目标。通过部署在边缘设备的检测算法分析图像,准确定位图中红外弱小目标位置,从而支持进一步跟踪、检测、目标打击等任务。

1.2 脉冲神经网络技术背景

脉冲神经元是脉冲神经网络中处理信息的基本单元,其中一个典型的模型就是Gerstner等^[12]在1995年提出的泄漏积分点火(LIF)神经元模型,极大地简化了动作电位的产生过程,保留了神经元膜电位的泄漏、积累、阈值激活三个关键特征,LIF神经元会对输入的电流积分,当膜电位大于阈值时发放脉冲,同时神经元膜电位被重置为静息电位。但是LIF通过二值的脉冲信号传递信息,其性能相比卷积



图1 机载红外弱小目标检测场景示意图

Fig.1 Aircraft-mounted infrared small target detection scene diagram

网络下降明显,Wu Zhenzhi等^[13]提出的泄漏积分模拟点火(LIAF)神经元在LIF的基础上保留了其动力学基础,又将神经元中传输的信号改为模拟值,使得LIAF能够和传统人工神经网络联合使用,用这种神经元构建的网络使用替代梯度^[14]反向传播的方式进行训练。本文在LIAF神经元的基础上增加了负脉冲的发放,并进一步引入邻域神经元对中心神经元的关注,将其融合到整体的网络结构中,实现对目标相关特征的增强。

2 算法流程

2.1 模型整体结构

为了增强红外弱小目标图像的有效特征,提高对红外弱小目标的检测能力,本文提出了SEFFN模型。模型的整体结构如图2(a)所示,给定一个输入图像 I ,通过SEFFN的端到端处理,最终得到和输入图像大小相同的小目标分割图。首先将经过预处理的红外图像输入全卷积网络^[15]进行特征提取,此处采用ResNet-18得到三个不同大小的特征层,分别为特征层1、特征层2和特征层3。来自不同层次的特征依次进入相应模块处理,特征层1进入SAM得到全局的注意力掩码,随后施加到被其他模块处理后的更深层特征。特征层2被MSE处理,增强网络对红外小目标的特征表示。特征层3通过DPC补偿小目标的底层语义信息。经过上述模块处理后的特征依次通过DAF对深层和浅层语义进行融合对齐,以获取更准确的目标定位,最后得到小目标分割结果。

2.2 DPC模块

较深层次的特征相比浅层特征经过更多次数的卷积处理,一些更精细的信息被过滤,包含更多的语义信息。为增加对小目标相关特征的感知,本文设计了DPC,如图2(b)所示。其中采用了三种不同膨胀系数的空洞卷积来增加网络

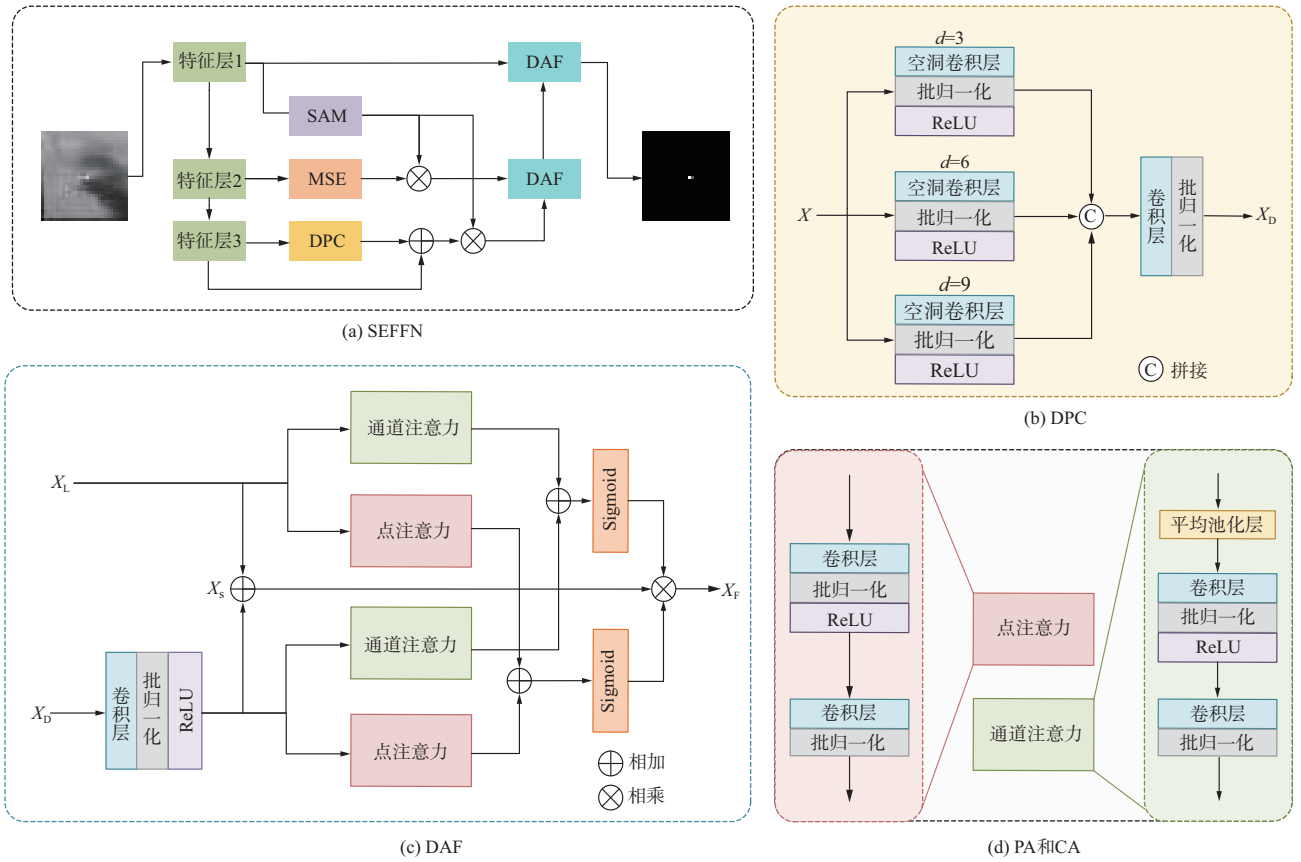


图2 SEFFN整体框架和扩展金字塔卷积模块、双路注意力融合模块

Fig.2 Overview structure of SEFFN and DPC&DAF

的感受野,同时过滤目标周围的噪声,补偿在下采样过程中减少的特征,有助于后续对小目标的提取。空洞卷积的处理过程表示为

$$X_j = \delta(\text{BN}(\text{Conv}_{3 \times 3}(X, d=3, 6, 9))), j=1, 2, 3 \quad (1)$$

式中, $\text{Conv}_{3 \times 3}(\cdot)$ 是卷积核大小为 3×3 的空洞卷积, 膨胀系数 d 为 3、6、9, 它定义了卷积核中相邻元素之间的间隔数量。BN($\cdot$) 和 δ 分别表示批归一化和 ReLU 激活函数。

特征层 X 输入上述空洞卷积得到 X_j , 三个空洞卷积的结果通过 $\text{Concat}[\cdot]$ 拼接聚合, 后面接卷积核大小为 1×1 的卷积 $\text{Conv}_{1 \times 1}(\cdot)$ 整合通道数

$$X_D = \delta(\text{BN}(\text{Conv}_{1 \times 1}(\text{Concat}[X_j])), j=1, 2, 3 \quad (2)$$

2.3 DAF 模块

不同层次的特征表示对小目标的语义信息和空间位置信息各有侧重, 为了使网络能够充分利用上下文信息, 避免目标细节丢失, 融合语义不一致和尺度不一致的特征, 本文设计了 DAF。

如图 2(c) 所示, X_L 和 X_D 分别表示来自浅层和深层的特征信息。首先对 X_D 进行卷积和上采样, 调整到和浅层特征

层 X_L 相同的大小和维度, 将两者相加初步融合得到 X_S , 作为后续施加注意力的对象。DAF 中包含点注意力 (PA) 模块和通道注意力 (CA) 模块, 如图 2(d) 所示, 两者都会作用于浅层特征 X_L 和深层特征 X_D , 也就是说会学习 4 个不同的注意力。

X_L 和 X_D 分别输入各自的通道注意力模块中计算两个通道注意力, 两者相加后用 sigmoid 进行归一化, 得到对不同特征通道的注意力权重 W_{CA}

$$W_{CA} = \sigma[\text{Conv}_2(\text{Conv}_1(\text{AvgPool}(X_D))) + \text{Conv}_4(\text{Conv}_3(\text{AvgPool}(X_L)))] \quad (3)$$

式中, σ 代表 sigmoid, $\text{Conv}_1(\cdot)$ 、 $\text{Conv}_2(\cdot)$ 、 $\text{Conv}_3(\cdot)$ 、 $\text{Conv}_4(\cdot)$ 均为 1×1 卷积, $\text{Avgpool}(\cdot)$ 是平均池化。为保证公式简略, 通道注意力和点注意力公式中省略批归一化和 ReLU 操作。

然后, X_L 和 X_D 分别输入各自的点注意力模块中计算对应注意力权重, 学习来自不同层次的小目标的细节, 再将两个点注意力相加, 同样使用 sigmoid 进行归一化得到对像素

位置的注意力权重 W_{PA}

$$W_{PA} = \sigma[\text{Conv}_2(\text{Conv}_1(X_D)) + (\text{Conv}_4(\text{Conv}_3(X_L)))] \quad (4)$$

然后对 X_S 依次施加通道注意力和点注意力,得到最终的注意力融合图 X_F ,式中 $\otimes$ 表示对应元素相乘

$$X_F = [X_L + \delta(\text{BN}((\text{Conv}_{1 \times 1}(X_D))))] \otimes W_{CA} \otimes W_{PA} \quad (5)$$

DAF 以这样的方式融合来自深层和浅层的特征图,解决了目标细节丢失、层间信息冗余以及特征融合不充分等问题。

值得一提的是,在对 X_D 进行上采样的操作时,本文选择了同时应用最近邻插值上采样和转置卷积来恢复深层特征层的分辨率,具体操作为

$$X_{Up}(X_D) = \text{Up}(\delta(\text{BN}(\text{Conv}_{3 \times 3}(X_D)))) + \delta[\text{BN}(\text{ConvT}(X_D))] \quad (6)$$

式中, $\text{Up}(\cdot)$ 表示最近邻插值上采样, $\text{Conv}_{3 \times 3}$ 为 3×3 卷积, $\text{ConvT}(\cdot)$ 是转置卷积,卷积核大小为 2。

最近邻插值上采样不会改变特征层的信息,而转置卷积可以根据学习到的卷积核来恢复细节,两者互补能降低块状效应,增加网络对重点区域的关注。

2.4 MSE 模块

对该模块的创新灵感来自脉冲神经元模型,本文将脉冲神经元的膜电位更新机制整合为 MSE,如图 3(a)所示,用来增强小目标特征,防止弱小目标在网络中丢失。

MSE 由三个脉冲模块及其后的非局部模块组成,如图 3(b)所示。特征图压缩通道后,经过多个模块处理,再重新恢复特征通道。在本文提出的脉冲模块中, t 为时间步,输入信号 X_t 通过卷积 $\text{Conv}(\cdot)$ 整合得到因输入信号引起的膜电位变化 I_t

$$I_t = \text{BN}(\text{Conv}(X_t)) \quad (7)$$

然后上一时间步的膜电位 V_{t-1} 和 I_t 累积得到当前时间步的膜电位 U_t ,即

$$U_t = V_{t-1} + I_t \quad (8)$$

所得当前时间步的膜电位 U_t ,随后和人为设定的激活阈值 V_{th} 相比较,大于阈值上界则发放正脉冲,小于阈值下界则发放负脉冲,否则不发放脉冲,脉冲信号 F_t 的计算如下

$$F_t(U_t, V_{th}) = \begin{cases} 1, & \text{if } U_t \geq V_{th} \\ 0, & \text{if } |U_t| < V_{th} \\ -1, & \text{if } U_t \leq -V_{th} \end{cases} \quad (9)$$

发放脉冲之后,膜电位被重置为 R_t ,发放正脉冲的神经元膜电位落回静息电位 V_{rest} ,发放负脉冲的神经元膜电位被抑制,而未发放的则被累积, γ 为累积系数。

$$R_t = F_t \cdot V_{rest} + \gamma \cdot U_t \quad (10)$$

$$\gamma = \begin{cases} 0, & \text{if } F_t = 1 \\ 1, & \text{if } F_t = 0 \\ 0.5, & \text{if } F_t = -1 \end{cases} \quad (11)$$

当前时间步神经元的膜电位会在衰减后保留为 V_t ,接下来进入下一时间步脉冲信号的计算,衰减为模拟生物神经元中的泄漏过程。 V_t 的计算如下

$$V_t = \alpha \cdot R_t + \beta \quad (12)$$

式中, α 和 β 分别为乘法衰减和加法衰减。

当最后一个时间步的膜电位累积结束,该神经元膜电位经过 SELU 激活函数处理,即式(13)中的 $\text{Act}(\cdot)$,随后接受来自邻域神经元膜电位的影响,得到信号输出。

$$Y_t = \text{Act}(U_t) \otimes \text{Atten}(X_t) \quad (13)$$

$\text{Atten}(\cdot)$ 以非局部块^[10]的形式实现,如图 3(c)所示,表示感受野中其他神经元膜电位对中心神经元输出的影响,这种机制更接近生物神经元实际的信息处理机制,能够捕获神经元之间的远距离依赖。非局部块对输入特征用 1×1 卷积进行线性映射,然后通过嵌入高斯函数估计嵌入空间的相似度

$$e_{ij} = \frac{\exp(K_i \cdot Q_j)}{\sum_{i=1}^N \exp(K_i \cdot Q_j)} \quad (14)$$

最后将映射结果添加到原始特征映射 X 中,得到非局部块的结果 X' ,标量 α 用于自适应学习两者的比值

$$X'_j = \alpha \sum_{i=1}^N (e_{ji} V_i) + X_j \quad (15)$$

被输入上述生物可解释性的脉冲神经元进行计算的 X_t ,其中和小目标相关的特征会在膜电位累积过程中被不断增强。由于输入的图像是单帧的,直接输入特征层 2 并不含有时间维度,因此 MSE 中的脉冲模块会通过复制的方式对输入特征图增加时间维度,用于脉冲模块中的膜电位计算和累积。在计算结束后,为了恢复特征图的维度,对时间维度作取平均值的处理。

在 MSE 中总共堆叠了三个脉冲模块,多次累积增强小目标特征,并且融入了对全局信息的理解,以更加准确地定位。

2.5 SAM 模块

由于小目标所占的像素极少,和背景之间的存在极端类别不平衡,背景中的很多像素对检测小目标无用或者有干扰。为此,本文设置了 SAM 用来对更深层特征进行过滤,抑制背景,在一定程度上增加对小目标区域的关注。

该模块采用了空间组增强模块 SGenet^[17]中的注意力结构,如图 3(d)所示,它通过为每个语义组 group 中的每个

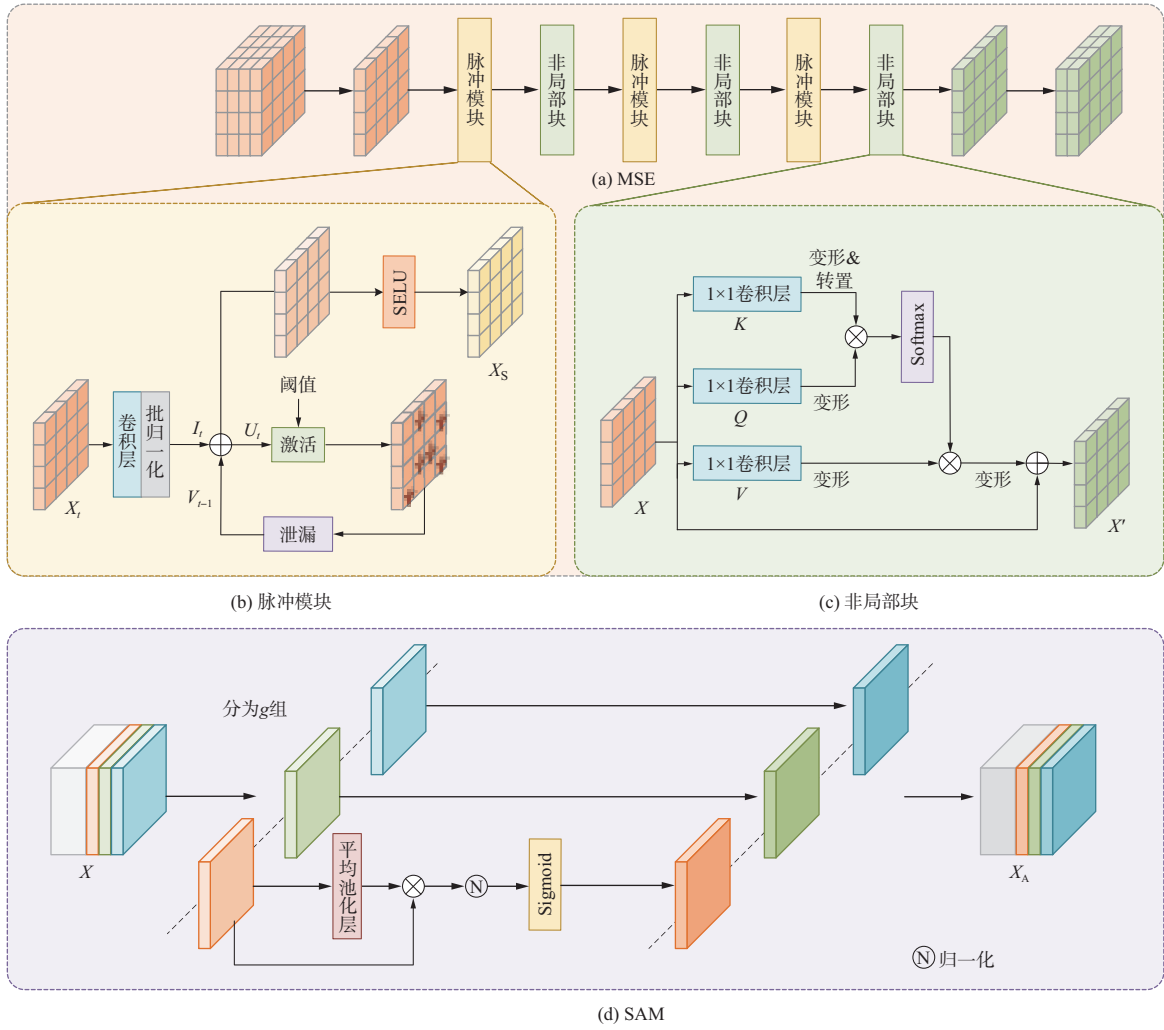


图3 多脉冲增强模块和监督注意力模块

Fig.3 MSE and SAM

空间位置生成注意因子来调整每个子特征的重要性,从而单独的组能够自主地增强其学习的特征表示,并抑制噪声。

在SAM中,通道数为 C ,大小为 $H \times W$ 的输入特征图按通道方向分为 G 个组,此处 G 是人为设定的组数, P_i 是局部统计特征

$$p = \{P_1, \dots, P_n\}, P_i \in \mathbb{R}^{\frac{C}{G}}, n = H \times W \quad (16)$$

对每个组分别进行平均池化得 g

$$g = \text{Avgpool}(p) \quad (17)$$

然后对于每个位置 $i \in \{1, \dots, n\}$,计算全局和局部的相似度,此处 θ 为 g 和 P_i 之间的夹角

$$W_i = \|g\| \times \|P_i\| \times \cos \theta_i \quad (18)$$

为了避免不同样本之间的分布差异,需要归一化到同样的范围来计算最终的注意力,并引入常数 ε 保证数值稳定性。具体操作如下

$$N_i = \frac{W_i - \mu_w}{\sigma_w + \varepsilon} \quad (19)$$

$$\mu_w = \frac{1}{n} \sum_j W_j \quad (20)$$

$$\sigma_w^2 = \frac{1}{n} \sum_j (W_j - \mu_w)^2 \quad (21)$$

最后通过sigmoid计算最终的空间注意力权重 W_A

$$W_A = \sigma(N_i) \quad (22)$$

SAM中计算的注意力权重会分别施加到被MSE和DPC处理的特征层,对背景杂波、噪声等进行过滤,再输入DAF中融合。

SAM采用监督训练的方式,以节省训练所需的时间和迭代次数。但小目标区域和背景之间存在极端类别不平衡,以及在注意力特征层中,关注区域的边界并非阶跃变化,因此对真值标签进行膨胀放大和高斯模糊处理后,生成

以小目标为中心向外扩散的真值模糊标签,作为监督项。

2.6 损失函数

模型训练的损失函数由两部分组成。

一是对于预测结果,希望它更多地关注目标区域,而不是平等地对待所有像素。因此采用在语义分割任务广泛使用的 SoftIoU 作为小目标分割的损失

$$L_{\text{softIoU}} = 1 - \frac{1}{n} \sum_{k=1}^n \frac{P_k \cap G_k}{P_k \cup G_k} \quad (23)$$

式中, n 为批量大小, G 为真值标签, P 为预测结果。

二是对于 SAM, 需要区分像素是目标还是背景, 采用 BCEloss 作为损失函数, 此处用于监督 SAM 的是经过处理后的真值模糊标签 G'

$$L_{\text{BCE}} = -\frac{1}{n} \sum_{k=1}^n (G'_k \log P_k + (1 - G'_k) \log (1 - P_k)) \quad (24)$$

最后, 用于训练的损失函数为两者相加, λ 为平衡两者的超参数, 根据经验设置为 5

$$\text{Loss} = L_{\text{softIoU}} + \lambda L_{\text{BCE}} \quad (25)$$

3 试验与结果分析

3.1 数据集

为了充分评估所提方法的有效性, 本文选择了 SIRST Aug^[18] 和 MDFA^[11] 数据集进行试验, 它们均包含天空、地面等多种场景, 符合机载红外弱小目标检测算法的应用场景。SIRST Aug 数据集的训练集包含 8525 张图像, 测试集包含 545 张图像, 图像分辨率为 256×256, 是对经典的红外弱小目标数据集 SIRST^[9] 的扩充, 包含短波、中波、长波等多波段真实图像。MDFA 数据集的训练集包含 10000 张图像, 测试集包含 100 张图像, 其中既包含真实的红外图像, 又包含将小目标与真实背景合成的图像。

3.2 试验细节

训练过程中, 输入图像大小设置为 256×256, 训练迭代次数为 5, 初始学习率为 0.05, 学习率的衰减策略为多项式衰减。批量大小为 4, 利用 SGD 随机梯度下降优化器更新参数, 其中动量和权重衰减分别设置为 0.9 和 0.0004。

本文采用了 4 个重要指标来全面验证模型的性能, 分别是精度、召回率、 F 度量、平均交并比。

精度 P 表示正样本和预测的正样本的比例

$$P = \frac{TP}{TP + FP} \quad (26)$$

召回率 R 表示正样本被正确预测的概率

$$R = \frac{TP}{TP + FN} \quad (27)$$

F 度量是精确率和召回率的加权调和平均, 能综合考虑两者以更全面地衡量性能

$$F = \frac{2 \times P \times R}{P + R} \quad (28)$$

平均交并比 (mIoU) 是衡量语义分割模型性能的经典指标之一, 它定义为预测与真值标签的交集和并集之比

$$\text{mIoU} = \frac{TP}{FP + FN + TP} \quad (29)$$

3.3 模型性能对比

本文选取了 9 种经典的红外弱小目标检测方法进行对比, 分别是模型驱动的 TopHat^[19]、PSTNN^[20]、WSLCM^[21]、RLCM^[22]、ADMD^[23] 和数据驱动的 ACM^[9]、ALC^[24]、ISTDU^[25]、DNA^[10], 对比试验结果见表 1。

表 1 不同算法在 Sirst Aug 和 MDFA 数据集上的对比试验结果

Table 1 Comparison between experimental results with different algorithms on Sirst Aug and MDFA datasets

模型	Sirst Aug				MDFA			
	P/%	R/%	F/%	mIoU/%	P/%	R/%	F/%	mIoU/%
TopHat ^[19]	71.16	18.24	29.08	17.02	4.64	19.89	7.53	3.91
PSTNN ^[20]	94.34	13.95	24.30	13.83	73.32	25.04	37.34	22.95
WSLCM ^[21]	84.35	5.41	10.16	5.35	84.19	15.03	25.51	14.62
RLCM ^[22]	69.59	19.83	30.86	18.25	22.21	42.64	29.22	17.11
ADMD ^[23]	93.39	11.75	20.87	11.65	55.01	12.84	20.81	11.62
ACM ^[9]	78.24	79.50	78.86	65.10	59.53	63.39	61.40	44.30
ALC ^[24]	79.62	78.52	79.06	65.38	60.50	60.92	60.71	43.59
ISTDU ^[25]	83.20	82.53	82.86	70.74	65.61	59.67	62.50	45.45
DNA ^[10]	82.90	84.98	83.92	72.30	61.09	64.18	62.60	45.56
SEFFN	84.64	86.87	85.74	75.04	63.45	65.22	64.32	47.41

在给出的试验结果中, 所有数据均以百分比形式呈现。SEFFN 在两个数据集上召回率、 F 度量和 mIoU 均达到最佳, 表明该算法在红外弱小目标检测任务方面有显著提升。在 Sirst Aug 数据集上, SEFFN 比第二位的 DNA 在 F 度量和 mIoU 分别高出 1.82% 和 2.74%, 在 MDFA 数据集上也有类似的趋势。

而精度指标方面, 数据驱动方法不占优势, 因为红外弱小目标检测任务中, 背景和目标之间的类不平衡十分显著, 数据驱动方法在训练过程中可能更倾向于提高召回率和 F 度量, 在减少偏差的同时会在一定程度上牺牲精度, 从而得以学习更加复杂的数据和提高泛化性。

传统方法中除了 TopHat 和 RCLM 在 MDFA 数据集上精度表现不佳, 其余方法在两个数据集上保持较高精度和较低召回率, 但整体性能较差, 也就是说它们能检测到目标的位置, 但只能检测到真实目标中的几个像素。这是因为

传统方法引入的先验知识在抑制背景的同时也会抑制目标像素,检测结果受到先验知识的约束。而数据驱动的方法则不会受此影响,它们的检测结果取决于所学习的特征,因此能检测到大部分的目标像素,在平衡精度和召回率之间达到平衡。

通过比较图4中不同方法分别在两个数据集上的ROC曲线,能明显看出数据驱动方法比传统方法更具优势,因为传统方法的性能很大程度上依赖先验知识,无法适应复杂背景的变化。而SEFFN的性能达到最佳,能准确有效地抑制背景,并保证红外弱小目标的定位和精确分割。

除了对模型性能的试验,本文还计算了模型的浮点运算次数(GFLOPs)、参数量(Params)以及每秒推理的图像帧数FPS作为参考。

从表2可知,SEFFN的模型参数量和浮点计算次数都超过了先前的方法,因此导致FPS有所下降。参数量的增

表2 不同算法性能对比

Table 2 Performance comparison between different algorithms

模型	GFLOPs	Params/M	FPS	mIoU/%
ACM	0.41	0.40	219.14	65.10
ALC	0.38	0.43	235.24	65.38
ISTDU	7.96	2.75	94.61	70.74
DNA	14.28	4.70	63.84	72.30
SEFFN	62.02	20.62	32.66	75.04

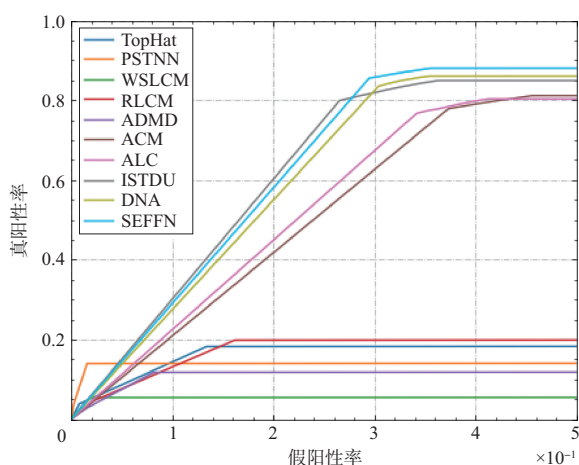
加主要来源于DAF,而FPS的下降是由MSE导致的。MSE中的脉冲模块是一个时空模块,对输入的二维特征图增加了时间维度,对三维特征图的计算导致算法的效率有所下降,因而低于其他数据驱动算法。但是从整体的模型效果而言,更大的参数量提升了模型的表达能力,能够捕捉到更丰富的特征信息,尤其是在多变的背景下对缺乏本征信息的红外小目标的提取方面,SEFFN有明显的检测精度提升。

不同算法的可视化检测结果如图5所示,图中从左至右的每一列,分别为输入图像、TopHat、PSTNN、ADMD、ALC、ISTDU、DNA、本文算法SEFFN的检测结果以及数据集的红外弱小目标分割标签,而每一行则对应着不同的检测场景,其中蓝色框、黄色框、红色框分别框出正确检测、漏检、错误检测的区域。

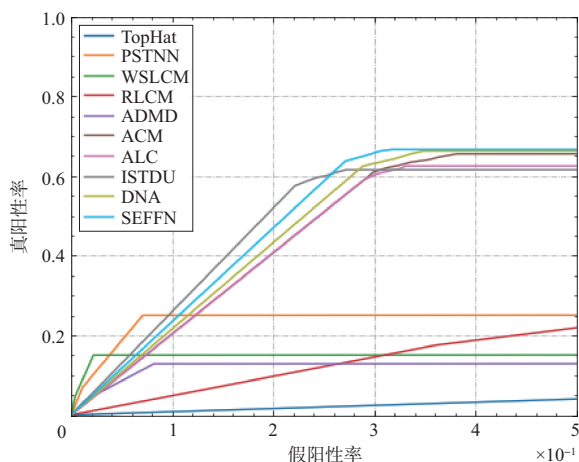
可视化检测结果中,传统方法的误检和漏检情况很严重。从上至下的4个场景中,第一个塔吊场景下,其他对比方法都有不同程度的漏检和误检,错误地将塔吊结构的高亮区域识别为目标,并且TopHat和PSTNN都没有正确定位到目标区域。在第二个海洋场景中,三种传统方法都有不同程度的误检,而数据驱动方法都能正确定位到目标区域,只是对于目标的精确分割程度不同。在第三个场景中存在多个灰暗小目标,TopHat、PSTNN和ALC均存在漏检的情况,而且大多数方法对目标轮廓的分割也不够精确。在第4个场景中,传统方法无法检测出完整的目标区域。这4个场景中,传统方法主要关注目标的定位,它们都无法完整地检出精确的目标形状,其表现很大程度上取决于场景的复杂性;数据驱动方法相比前者有较大的性能提升,但是对目标的检测不够精确。本文提出的SEFFN在不同场景中表现出更强的鲁棒性,能很好地适应各种复杂的背景、目标形状和尺寸的挑战,取得更好的检测性能。

3.4 消融试验

除了和经典方法的对比试验以外,本文还进行了消融试验,试验在Sirst Aug和MDFA数据集上进行了评估,旨在分析模型中不同组成部分的作用和贡献,试验结果见表3。



(a) Sirst Aug数据集上不同算法的ROC曲线



(b) MDFA数据集上不同算法的ROC曲线

图4 不同算法的ROC曲线

Fig.4 ROC curves of different algorithms

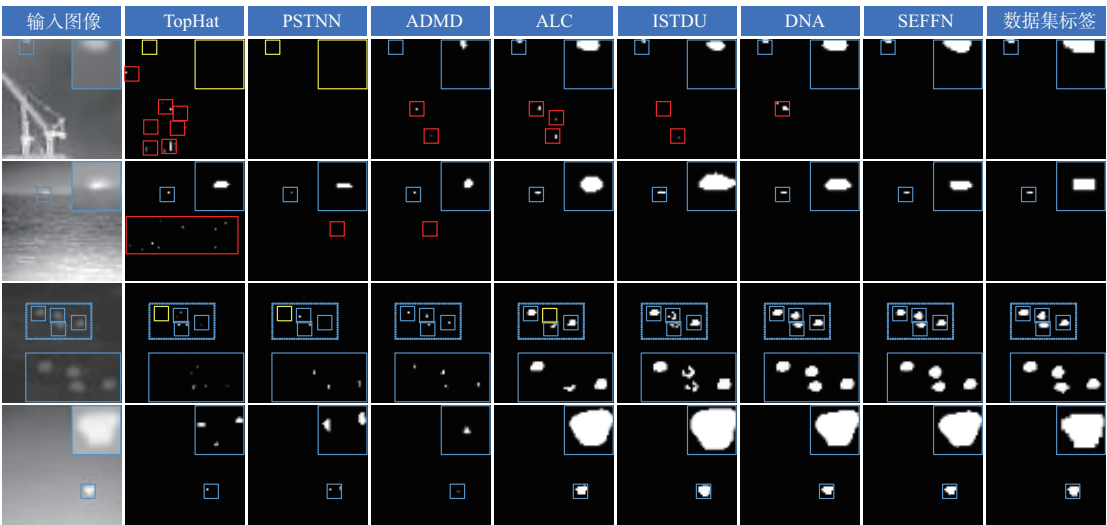


图5 不同算法的可视化检测结果

Fig.5 Visualization results of different algorithms

本文所做消融试验主要探究以下4个问题。

(1) DPC的影响

DPC用于过滤深层信息中的噪声,并且增加网络感受野。通过对比试验1和2、3和6以及8和9可以得知,在添加了DPC模块的网络中,网络模型的性能在两个数据集上均有不同程度的提高,这说明了扩展卷积在网络中发挥了关键作用,不同膨胀率能够增加网络的感受野,有助于捕获更丰富的小目标特征细节。不过DPC也在一定程度上增加了模型的参数量。

(2) MSE的影响

MSE利用了生物可解释性的信息处理方式,用于实现对小目标相关特征的增强。试验1和3、2和6、4和7、5和8中,分别是在几种变体中未添加MSE和添加MSE的网络,比较

这4组试验结果不难发现,加入MSE能够明显使两个数据集上的评价指标得到提升。MSE带来的参数量增加相对较少,但是对FPS影响显著。MSE中的脉冲模块是事件驱动的,当膜电位高于阈值的时候,才会将模拟值发送到后继层,这种设计减少了不必要的数据传输,提高了信息处理的精确度和资源利用率。然而正是因为脉冲模块的这一特点,引入了时间维度的计算,使得模型推理时间增加,导致FPS下降。MSE对小目标相关特征的增强有一定效果,从而显著提升网络检测的效果,但也导致了算法效率的降低。

(3) DAF的影响

试验中未使用DAF的变体,而是采用简单相加来对深层和浅层的特征融合。DAF用于上下文信息的交互,融合深层和浅层的不一致的特征。通过对比试验1和4、3和7

表3 Sirst Aug和MDFA数据集上的消融试验结果

Table 3 Ablation test results on Sirst Aug and MDFA datasets

试验 序号	模块				SirstAug		MDFA				
	DPC	MSE	DAF	SAM	F/%	mIoU/%	F/%	mIoU/%	GFLOPs	Params/M	FPS
1	×	×	×	×	81.29	68.48	60.42	43.29	37.90	11.38	378.44
2	√	×	×	×	82.64	70.42	60.93	43.81	42.20	15.58	307.25
3	×	√	×	×	82.19	69.77	62.50	45.46	38.98	11.42	60.40
4	×	×	√	×	81.99	69.47	62.98	45.96	56.78	16.63	195.34
5	×	×	×	√	83.18	71.20	61.13	44.02	38.14	11.38	274.88
6	√	√	×	×	84.03	72.45	63.07	46.05	43.28	15.62	38.98
7	×	√	√	×	84.17	72.67	63.17	46.16	57.86	16.68	48.55
8	×	√	×	√	83.75	72.04	62.70	45.66	39.22	11.42	53.24
9	√	√	×	√	84.58	73.27	63.43	46.45	43.52	15.62	36.31
10	√	√	√	√	85.74	75.04	64.32	47.41	62.02	20.62	32.62

还有9和10,可以观察到添加了DAF的变体在两个数据集上的表现都优于未添加DAF的变体,在更复杂的网络中DAF的作用更加明显,这表明DAF在增强上下文特征融合方面是有效的,通过关注深层和浅层的空间特征细节和语义特征,实现了特征的高效利用。DAF中的通道注意力、点注意力和多尺度特征融合的操作带来了参数量的增加,这是在提升算法模型精度和鲁棒性目标上所必须付出的代价。

(4) SAM的影响

SAM是施加到更深层的注意力模块,用于过滤深层特征和抑制背景。由表3可知,对比试验1和5、3和8、6和9的结果显示,没有SAM的模型变体性能有所降低。并且,SAM是分组进行注意力计算的,其带来的参数量增加相比其他模块更少,但是由于需要作用到深层特征层,将注意力权重与对应元素相乘,导致效率有一定程度的降低,FPS有所减少。试验1和5的对比中,没有SAM的变体1比只添加SAM的变体5在Sirst Aug(MDFA)上的 F 度量和mIoU分别下降1.89%(0.71%)和2.72%(0.73%),这表明SAM确实能够增进网络对目标所在区域的关注,从而使小目标检测更为准确。

在试验10中,所有的模块都被添加,这也是本文SEFFN最终的结构,其表现出在所有评价指标上最好的性能,相对单独使用各个模块更加优秀。虽然引入的结构导致了模型的参数量增加,以及相对应的算法效率降低,但这是模型优化过程中必须面对的挑战。总之,以上所有模块的引入让SEFFN在整体效果上取得了明显的提升,在Sirst Aug上 F 度量和mIoU分别达到85.74%和75.04%,在MDFA上 F 度量和mIoU分别达到64.32%和47.41%,这也说明这些模块在集成时能够相互补充,优化网络的整体性能,在网络模型中扮演了不可或缺的关键角色。

4 结束语

本文提出了一种新的机载红外弱小目标检测模型——SEFFN。该模型包含融合脉冲神经网络结构的MSE,能有效增强小目标相关的特征,并且使用DPC增加网络感受野,过滤噪声,还有DAF同时应用通道和点注意力来整合不同层次的上下文信息。SAM从全局角度抑制背景噪声,聚焦小目标区域。SEFFN模型能够部署在空中平台的边缘设备上,在执行监控、预警等项目时,实现远距离有效监测小目标,为后续反应和决策提供重要支持。在公共数据集上的大量对比试验表明,本文提出的SEFFN取得了优于现

有方法的卓越性能,消融试验也验证了各个模块的有效性。虽然脉冲结构和注意力融合结构等模块的引入带来了不可避免的效率损失,但这是实现这些重要改进所必须付出的代价。

在未来的研究中,将在保持或者提升模型精度的前提下,进一步减少模型的参数量和加快推理速度。此外,希望本文工作能够为脉冲神经网络结构在空中平台跟踪检测运动红外弱小目标任务中的应用提供新的思路,推动航空领域红外探测技术的进步。

AST

参考文献

- [1] 赵晓冬,张洵颖,车军,等.精确制导武器末制导目标识别优化算法研究[J].航空科学技术,2022,33(1):126-134.
Zhao Xiaodong, Zhang Xunying, Che Jun, et al. Research on optimization algorithm of terminal guidance target recognition for precision-guided weapons[J]. Aeronautical Science & Technology, 2022, 33(1): 126-134. (in Chinese)
- [2] 董凤禹,魏振忠.面向空中小目标检测任务的YOLOv7改进模型[J].航空科学技术,2023,34(12):111-117.
Dong Fengyu, Wei Zhenzhong. An improved YOLOv7 model for small aerial object detection [J]. Aeronautical Science & Technology, 2023, 34(12): 111-117. (in Chinese)
- [3] 李俊宏,张萍,王晓玮,等.红外弱小目标检测算法综述[J].中国图象图形学报,2020,25(9):1739-1753.
Li Junhong, Zhang Ping, Wang Xiaowei, et al. Infrared small target detection algorithms: a survey[J]. Journal of Image and Graphics, 2020, 25(9): 1739-1753. (in Chinese)
- [4] Deshpande S D, Er M H, Venkateswarlu R, et al. Max-mean and max-median filters for detection of small targets[C]// Proceedings of SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, 1999: 74-83.
- [5] Asrabadi N. Hyperspectral target detection: an overview of current and future challenges[J]. IEEE Signal Processing Magazine, 2014, 31(1): 34-44.
- [6] McIntosh B, Venkataramanan S, Mahalanobis A. Infrared target detection in cluttered environments by maximization of a target to clutter ratio (TCR) metric using a convolutional neural network[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(1): 485-496.
- [7] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN:

- towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6):1137-1149.
- [8] Farhadi A, Redmon J. Yolov3: an incremental improvement [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1804-2767.
- [9] Dai Yimian, Wu Yiquan, Zhou Fei, et al. Asymmetric contextual modulation for infrared small target detection[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021: 950-959.
- [10] Li Boyang, Xiao Chao, Wang Longguang, et al. Dense nested attention network for infrared small target detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 1745-1758.
- [11] Wang Huan, Zhou Luping, Wang Lei. 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*, 2019: 8509-8518.
- [12] Abbott L F. Lapique's introduction of the integrate-and-fire model neuron(1907)[J]. *Brain Research Bulletin*, 1999, 50(5-6): 303-304.
- [13] Wu Zhenzhi, Zhang Hehui, Lin Yihan, et al. LIAF-Net: leaky integrate and analog fire network for lightweight and efficient spatiotemporal information processing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(11): 6249-6262.
- [14] Wu Yujie, Deng Lei, Li Guoqi, et al. Spatio-temporal backpropagation for training high performance spiking neural networks [J]. *Frontiers in Neuroscience*, 2018, 12: 331.
- [15] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [16] Wang Xiaolong, Girshick R, Gupta A, et al. Non-local neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7794-7803.
- [17] Li Yuxuan, Li Xiang, Yang Jian. Spatial group-wise enhance: enhancing semantic feature learning in CNN[C]//*Proceedings of the Asian Conference on Computer Vision*, 2022: 687-702.
- [18] Zhang Tianfang, Cao Siying, Pu Tian, 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.
- [19] Tom V T, Peli T, Leung M, et al. Morphology-based algorithm for point target detection in infrared backgrounds[C]//*Proceedings of the Signal and Data Processing of Small Targets*, 1993, 1954: 25-32.
- [20] Zhang Landan, Peng Zhenming. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. *Remote Sensing*, 2019, 11(4): 382.
- [21] Han Jinhui, Moradi S, Faramarzi I, et al. Infrared small target detection based on the weighted strengthened local contrast measure[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 18(9): 1670-1674.
- [22] Han Jinhui, Liang Kun, Zhou Bo, et al. Infrared small target detection utilizing the multiscale relative local contrast measure [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15: 612-616.
- [23] Moradi S, Moallem P, Sabahi M F. Fast and robust small infrared target detection using absolute directional mean difference algorithm[J]. *Signal Processing*, 2020, 177: 107727.
- [24] Dai Yimian, Wu Yiquan, Zhou Fei, et al. Attentional local contrast networks for infrared small target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59 (11): 9813-9824.
- [25] Hou Qingyu, Zhang Liuwei, Tan Fanjiao, et al. ISTDU-Net: infrared small-target detection U-Net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.

Deep Fusion of Spiking and Convolutional Networks for Infrared Small Target Detection

Li Lexiao, Ji Luping, Chen Shengjia, Zhu Sicheng, Guo Chaoxiang

University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: With the rapid development of UAV and other aerial vehicle technology, aircraft-mounted infrared small target detection has become a crucial research field. However, traditional algorithms face significant challenges in detection due to factors such as long target distances, small target sizes or partial occlusion. Therefore, developing more precise target detection algorithms is of utmost importance. In response to these challenges, this paper proposed a novel aircraft-mounted infrared small target detection model which integrated spiking and convolutional technologies, namely spike-enhanced fusion feature network (SEFFN) model. Unlike previous infrared small target detection algorithms based on deep learning, this model enhanced feature representation related to small targets through a biomimetic spiking neural network structure, enabling more accurate extraction of small target regions. Specifically, SEFFN comprises four key modules: dilated pyramid convolution (DPC), dual attention fusion (DAF), multi-spike enhance (MSE), and supervised attention module (SAM). These modules work collectively to improve the focus on the small target regions while retaining most of the features of small target without noise interference. Experimental results on two datasets demonstrate that SEFFN outperforms existing model-driven and data-driven algorithms, especially with its F-measure and mIoU on Sirst Aug dataset achieving 85.74% and 75.04%, respectively. This breakthrough validates the effectiveness and superiority of SEFFN in infrared small target detection task. SEFFN is suitable for the aviation field and can be deployed on the edge devices of aircraft to improve the air platform's ability to detect long-distance targets, enabling it to carry out remote monitoring missions, counter enemy threats, and perform other related tasks.

Key Words: infrared small target; spiking neural network; object detection; deep learning; attention mechanism

Received: 2024-04-22; **Revised:** 2024-07-31; **Accepted:** 2024-09-27

Foundation item: Aeronautical Science Foundation of China (2022Z071080006)