

逐像素注意力驱动的红外小目标检测网络

王啸林, 方厚章, 李雪婷, 吴辰星, 王黎明

(西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071)

摘 要: 红外小目标检测在军事和民用领域获得了广泛应用,但其存在目标尺度小、细节少、复杂背景干扰等问题,现有经典深度学习检测方法往往适用于通用目标检测,对红外小目标适配性较差。针对上述问题,构建了一种新的基于 U 形注意力块和逐像素注意力块的红外小目标检测网络。设计了 U 形注意力块,在单层级内通过局部 U 形子网络提取多尺度特征,并通过逐像素注意力精细化增强小目标特征,丰富多尺度小尺度目标特征表示,提升网络对小尺度目标判别能力;通过稠密融合方式进一步保留小目标信息,缓解不同层特征融合时的语义鸿沟,降低漏检率;将空间与通道 2 个维度逐像素注意力块应用于融合后的特征图,避免小目标特征被衰减,同时抑制复杂背景干扰。实验结果表明,提出的网络在 2 个红外小目标数据集 NUDT-SIRST 与IRSTD-1k 上交并比、检测概率、虚警率指标均超过最新基准方法。此外,所提网络在检测精度和效率上也达到较好平衡。

关 键 词: 红外小目标检测;U 形注意力块;逐像素注意力

中图分类号: TP391

文献标志码: A

文章编号: 1000-2758(2024)02-0335-09

红外搜索和跟踪系统在军民领域得到广泛应用,如遥感、制导^[1]、无人机^[2]等,红外小目标检测是该系统的关键技术,其在实际应用时面临的主要难点为:①目标尺度小。由于成像距离长或物体自身较小,一般地,红外小目标的定义为红外图像中像素面积小于 9×9 的目标,本文网络适用于此尺度范围。②目标亮度暗,或图像中信杂比较低。③目标细节少。因为红外热成像原理,小目标往往没有纹理信息,边缘细节较模糊;④复杂背景干扰。红外小目标易受到强云、山林、建筑物、海杂波等复杂背景的干扰。以上难点严重制约了红外小目标检测在实际场景中的应用。

到目前为止,已有很多方法被提出用于红外小目标检测。现有方法可以大致分为 2 大类:一类为

手工建模的传统方法^[3-5],另一类为数据驱动的深度学习方法。传统模型驱动的方法容易受到复杂背景与噪声的影响,且对参数敏感,鲁棒性较差,此外基于低秩建模类方法需要多次迭代优化,实时性也难以与部分单阶段深度学习类方法比较。深度学习因其具有自适应特征学习能力^[6],在红外小目标检测中获得广泛应用,但现有的红外小目标检测方法往往在通用目标检测方法上进行改进,如文献[7]是在通用目标检测框架下进行调整,其在 Faster R-CNN 和 YOLO-v3 的基础上设计了目标杂波比指标,以此为目标优化现有的通用目标检测方法,其并没有根据红外小目标特性构建有针对性的网络架构。经典编码-解码式像素级目标检测架构^[8]利用跳转连接联合浅层结构信息和深层语义信息,可在一定程度上捕获和提取目标的特征信息,实现对目标的像素级检测,但现有方法提取红外小目标多尺度细节信息的能力仍然有限,并且普遍缺少增强小目标与抑制复杂背景的有效机制。

在本文的工作中,构建了一种新的 U 形注意力块和逐像素注意力机制的红外小目标检测网络。首先,本文设计了 U 形注意力块,在单层级内通过局部 U 形子网络提取多尺度特征以丰富小尺度目标

收稿日期:2023-03-07

基金项目:国家自然科学基金(41501371)、多谱信息处理技术国家重点实验室开放课题(6142113220303)、中央高校基本科研业务费专项资金与西安电子科技大学研究生创新基金(YJSJ24015)资助

作者简介:王啸林(1998—),博士研究生

通信作者:王黎明(1982—),副教授

e-mail: wanglm@mail.xidian.edu.cn

特征表示;其次,在层级特征进行融合时,使用稠密融合方式,每次将同级跳转连接的特征和相邻层上采样与下采样后的特征进行融合,充分保留小目标信息,避免深层丢失小目标特征,降低漏检率;最后,在跨层特征融合后,构建了融入空间和通道注意力机制的逐像素注意力模块,通过空间与通道维持的方式保留目标特征,并自适应获取逐像素注意力图,对小目标的空间与通道中间特征进行精细化增强,避免小目标特征融合后被衰减,抑制背景干扰,使检测模型对复杂背景具有更强的鲁棒性。实验结果表明,本文的算法在红外小目标检测数据集 NUDT-SIRST^[9]上,交并比(intersection over union, IoU)、检测概率 P_d (probability of detection) 和虚警率 F_a (false-alarm rate) 达到 92.49%, 98.84%, 2.92×10^{-6} , 在IRSTD-1k^[10]上达到 70.25%, 92.18%, 7.44×10^{-5} , 并在推理时间与精度上取得了较好平衡。

1 逐像素注意力驱动的红外小目标检测算法设计

1.1 网络整体架构

红外小目标检测面临尺度小、亮度暗、细节少、背景复杂等困难,为了解决此类难题,本文设计了如图 1 所示的基于 U 形注意力块和逐像素注意力块的红外小目标检测网络。其主要构成为:左右两侧基于 U 形注意力块的编码器与解码器,中部的逐像素注意力块的特征稠密融合部件。

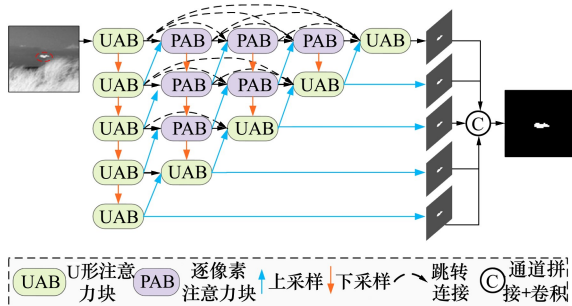


图1 基于U形注意力块与逐像素注意力块的
红外小目标检测网络架构图

输入红外图像后,首先,由 U 形注意力块构建的编码器进行多尺度特征提取,U 形注意力块在单层级内通过 U 形结构提取细粒度多尺度特征,可以进一步捕获红外小目标鉴别性特征,增强网络的识

别能力。其次,通过多级稠密融合方式进行特征融合,采取稠密融合可有效缓解不同层的语义差异,避免直接融合而使小目标特征衰减,提升网络对小目标的特征表达能力。此外,每次特征融合后使用空间与通道双维度的逐像素注意力模块进行有效特征筛选,自适应增强小目标特征,提升小目标检测的精度。最后,稠密融合后的特征再次经过基于 U 形注意力块的解码器,最终将不同层级的多尺度特征沿通道拼接,充分利用浅层丰富的空间信息与深层的强语义信息,进行卷积操作得到像素级检测结果。

1.2 U 形注意力块

经典 U-Net 的编码器经过多次下采样,可以得到目标丰富的全局语义信息,经过长跳转连接,将底层丰富的原始特征传递给解码器,解码器经过多次上采样得到分割结果。但经典 U 形编码-解码结构执行红外小目标检测任务仍存在以下问题:编码器与解码器结构单一,只能通过多次下采样捕获多尺度信息,而对红外小目标而言,每次下采样都会损失细节信息,因此该结构难以实现复杂背景下的小目标精准检测。

受残差 U 型块(RSU)^[11]启发,本文提出了一种新的 U 形注意力块(U-shaped attention block, UAB),在编码器单层级部分通过残差 U 形块获取更细粒度的多尺度信息,并通过逐像素注意力块自适应增强小目标信息。其结构如图 2 所示,其中 H 、 W 、 C_{in} 、 C_{out} 分别代表高度、宽度、输入通道数、输出通

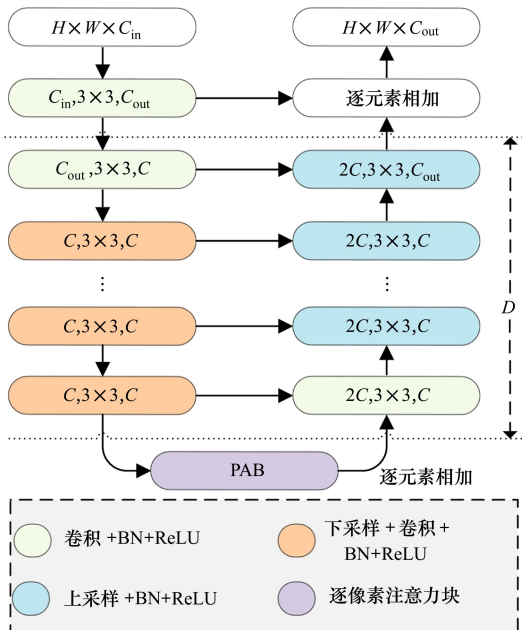


图2 U形注意力块结构示意图

道数, C 表示 UAB 内部的通道数, D 表示深度, 可根据块所在层级不同构造不同深度的 UAB, 在本文中, 最浅层 $D=5$, 逐层递减, 最深层 $D=1$ 。

本文设计的 U 形注意力块, 一方面通过局部 U 形子网络增大感受野, 提升网络的多尺度特征提取能力, 丰富了网络获取的小目标全局判别性知识, 另一方面通过逐像素注意力, 自适应增强小目标特征, 抑制背景干扰。两者结合, 进一步提升网络对小尺度目标的检测能力。

1.3 逐像素注意力模块

红外小目标具有尺度小、细节少的特点, 易受背景杂波干扰, 在稠密融合操作后, 如果不对目标加以增强, 小目标特征极易被淹没在复杂背景中, 因此, 本文采用注意力机制来自适应增强小目标, 同时抑制背景杂波。但经典注意力机制(如 CBAM^[12])通过全局最大池化与平均池化对空间和通道维度信息进行压缩, 其主要用于增强具有全局分布特性的中大尺度目标, 而红外小目标普遍具有局部的稀疏性, 因此, 经典注意力方法不再适用于红外小目标。如图 3a) 所示, 经典注意力方法难以有效增强复杂背

景下的红外小目标。

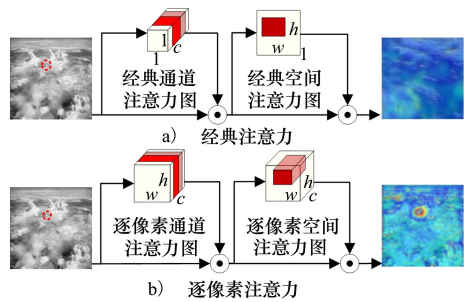


图 3 经典注意力与逐像素注意力效果对比图

针对上述经典注意力机制的缺陷, 本文设计了一个新的无压缩联合空间与通道逐像素注意力模块, 无压缩的注意力机制可有效增强小目标, 其效果如图 3b) 所示。逐像素注意力模块包括空间保持的通道注意力 (spatial maintained channel attention, SMCA) 和通道保持的空间注意力 (channel maintained spatial attention, CMSA), 其网络示意图如图 4 所示, 通过精细化的逐像素注意力图增强小目标特征, 抑制背景干扰。

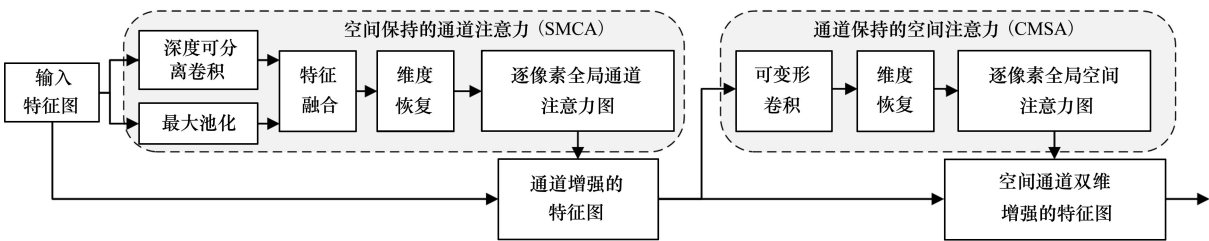


图 4 逐像素注意力模块架构图

首先, 在 SMCA 中, 通过卷积核大小为 n 的深度可分离卷积和最大池化将特征图空间维度压缩为原始大小的 $1/4$ 并逐元素相加, 这样一方面尽可能地保留小目标信息, 另一方面将平均池化替换为深度可分离卷积, 在保证实时性的前提下进一步抑制背景杂波。然后, 用逐点卷积将通道数恢复。使用反卷积将空间大小恢复并进行激活函数操作。最后将得到的通道注意力图与原始输入进行逐元素相乘, 在通道维度上增强小目标。综上所述, SMCA 可表示为: $X_1 = MP(X)$, $X_2 = DWConv(X)$, $Y_c(X) = X \odot \sigma(DeConv(PWConv(X_1 + X_2)))$, 其中 X 代表输入特征图, $MP(\cdot)$ 代表最大池化操作, $DWConv(\cdot)$ 代表深度可分离卷积, $PWConv(\cdot)$ 代

表逐点卷积, $DeConv(\cdot)$ 代表反卷积, σ 代表 sigmoid 函数, $\odot$ 代表逐点相乘, $Y_c(X)$ 代表经过 SMCA 增强的输出特征图。

在 CMSA 中, 首先, 使用逐点卷积调制通道上下文信息并且降低通道数为原始的 $1/r$, 本文中 $r=4$ 。使用核大小为 3×3 的可变形卷积, 可以自适应捕获小目标特征。其次, 通过逐点卷积恢复通道数, 经过激活函数得到空间注意力图。最后, 再与经过通道调制的特征图相乘, 得到空间通道双重维度增强的特征图。综上所述, CMSA 可被表示为: $Y(X) = X \odot \sigma(PWConv(DeformConv(X)))$, 其中 $DeformConv(\cdot)$ 代表可变形卷积操作, $Y(X)$ 代表最终输出的通道空间双重维度逐像素注意力增强的特征图。

本文提出的逐像素注意力模块,一方面可以有效保留红外小目标信息,确保特征传递到深层网络,另一方面通过精细化的特征注意力,细粒度地增强小目标特征,抑制背景干扰,提升小目标检测性能。

2 实验结果与分析

2.1 数据集与实验设置

采用 NUDT-SIRST^[9] 与 IRSTD-1k^[10] 2 个数据集验证本文的方法,NUDT-SIRST 数据集主要包含 5 类真实背景:城市、田野、海洋、强云、高亮场景,涵盖了大部分红外应用场景,单张图片大小为 256×256,其中目标均为真实红外小目标,通过自适应算法有效组合成数据集中的图片。IRSTD-1k 包括 1 000 张真实红外图像,所有数据均为实际拍摄,图像大小为 512×512。其包含不同种类的小目标,如无人机、生物、船只和车辆,背景包含了海洋、河流、田野、山区、城市和云。本文参考文献[9]的数据集划分方式与参数设置,NUDT-SIRST 的训练集为 664 张图片,测试集为 664 张图片,IRSTD-1k 的训练集为 800 张图片,测试集为 200 张图片。为了提升模型的泛化性与鲁棒性,本文对训练数据采用随机翻转、模糊和随机裁剪的方式进行数据增强,经过增强后可将训练数据量提升 8 倍以上。批大小(batch size)设置为 32,迭代期数(epoch)设置为 2 000,初始学习率为 0.05,损失函数选用 Soft-IoU 函数,优化器使用 Adam 方法。

实验硬件环境:服务器系统为 Ubuntu 18.04,显卡为 GeForce RTX 3090,CPU 为 Inter Xeon Silver 4210R @ 2.40GHz。软件环境为 python 3.6,CUDA 11.1,Pytorch 1.8.0。

2.2 实验评估指标

红外小目标因为像素数过少,单个错误像素值将导致指标显著变化,因此本文使用适合红外小目标检测的指标进行评估,包括交并比(intersection over union, IoU)、检测概率(probability of detection, P_d)、虚警率(false-alarm rate, F_a)。其中 IoU 和 P_d 越接近于 1 代表方法越好, F_a 越接近于 0 代表方法越好。此外,使用受试者特征曲线(receiver operation characteristics, ROC)来动态评估模型。指标具体细节为:

1) IoU:像素级的评价指标,可以评估算法的形状描述能力。IoU 可通过 $A_{\text{inter}}/A_{\text{union}}$ 计算得到,其中 A_{inter} 与 A_{union} 分别代表预测与真实标签的像素交集面积与像素并积面积。

2) P_d :目标级的评价指标,表示预测正确的目标数 T_{correct} 与所有目标数 T_{all} 的比值,即: $P_d = T_{\text{correct}}/T_{\text{all}}$ 。当目标质心与真实质心的距离小于 3 个像素时,将此目标视为预测正确。

3) F_a :目标级的评价指标,表示预测错误的像素 P_{false} 和所有图像像素 P_{all} 的比值,即: $F_a = P_{\text{false}}/P_{\text{all}}$ 。同样地,当目标质心与真实质心的距离大于 3 个像素时,将此像素视为错误预测。

4) ROC 曲线:用来描述在不同虚警率下,检测概率的变化趋势,可以动态地评估模型性能,曲线下面积越大,代表此方法越好。

2.3 实验结果与分析

2.3.1 定量分析

为了证明本文设计方法的先进性,选取多种先进方法进行对比,包括基于低秩分解类方法,如 RIPT^[3], PSTNN^[4];深度学习类方法,如 MDvsFA-cGAN^[13], ACM^[14], ALCNet^[15], DNANet (ResNet18)^[9],实验结果如表 1 所示。

表 1 不同方法的量化实验结果对比

算法	单帧推理 时间/s	NUDT-SIRST			IRSTD-1k		
		IoU/ 10^{-2}	$P_d/10^{-2}$	$F_a/10^{-6}$	IoU/ 10^{-2}	$P_d/10^{-2}$	$F_a/10^{-5}$
RIPT	0.641	29.44	91.85	344.30	7.45	17.27	67.64
PSTNN	0.762	14.85	66.13	44.17	6.96	27.03	1 029.78
MDvsFA	0.034	75.14	90.47	25.34	0.59	42.55	28 648.11
ACM	0.027	67.08	95.97	10.18	31.21	15.56	24.44
ALCNet	0.028	81.40	96.51	9.26	26.13	85.81	53.82
DNANet	0.029	87.09	98.73	4.22	65.58	89.46	10.63
本文	0.039	92.49	98.84	2.92	70.25	92.18	7.44

从表 1 结果可知,首先传统模型驱动类的算法在运行效率与检测效果上都难以超过深度学习类的

算法,手工建模类方法已无法建模复杂的场景,导致检测概率较低,虚警率较高。其次本文构造的算法与其他深度学习类算法相比,在 NUDT-SIRST 和 IRSTD-1k 数据集上,3 个指标都表现最佳,IoU 达到 92.49% 与 70.25%, P_d 为 98.84% 与 92.18%, F_a 为 2.92×10^{-6} 与 7.44×10^{-5} 。本文算法也在保证运行效率的同时提升检测效果,单帧推理时间为 0.039 s,约为 25.6 frame/s。此外,图 5a)~5b)为提出的算法与 4 种对比方法在 2 个数据集上的 ROC 曲线。对比方法中 RIPT, PSTNN 为低秩类方法,ALCNet 与

DNANet 为最新的基于深度学习的红外小目标检测方法。ROC 曲线下面积越大表明该方法性能越好,从图 5 可看出,深度学习类方法性能优于建模类方法,本文的方法在 2 个数据集上都可在较低的虚警率下,保持最高的检测概率,这表明本文的方法具有较高的鲁棒性。这得益于本文为红外小目标特定设置的网络模型,U 形块充分提取目标细粒度特征,稠密融合方式充分保留了小目标特征,逐像素注意力可以在丰富的特征中增强红外小目标特征,抑制背景干扰。

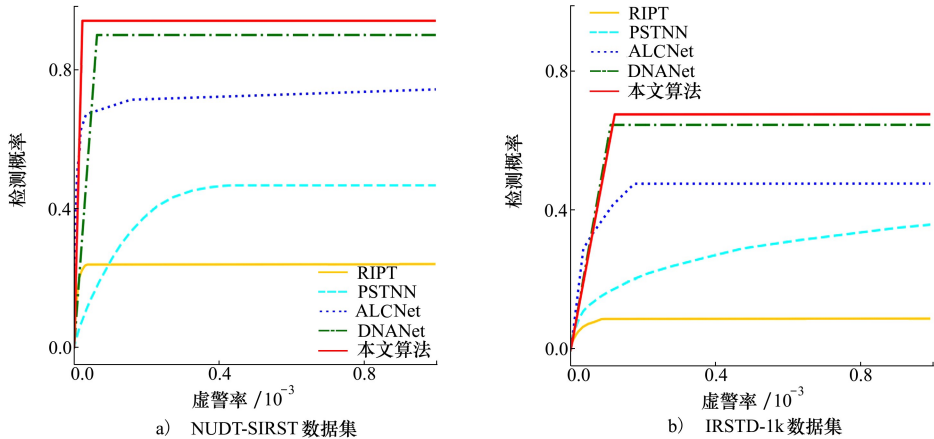


图 5 5 种方法的检测概率相对于虚警率变化的 ROC 曲线

2.3.2 检测结果可视化分析

图 6 与图 7 分别是不同方法在 NUDT-SIRST 数据集与 IRSTD-1k 数据集上的主观检测结果,图 6a)

和图 7a)是原始输入图像,图 6b)~6f)和图 7b)~7f)是不同方法的检测结果,图 6g)和图 7g)是真实标签。图中红色虚线框代表红外小目标的位置,橙

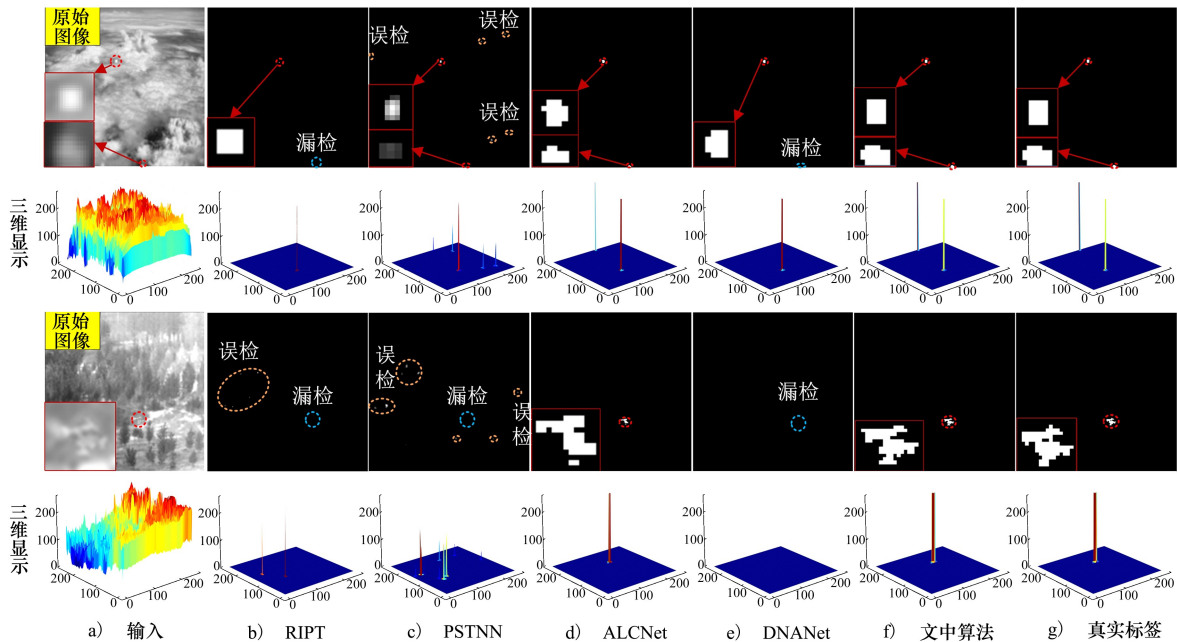


图 6 不同方法在 NUDT-SIRST 数据集上的检测结果

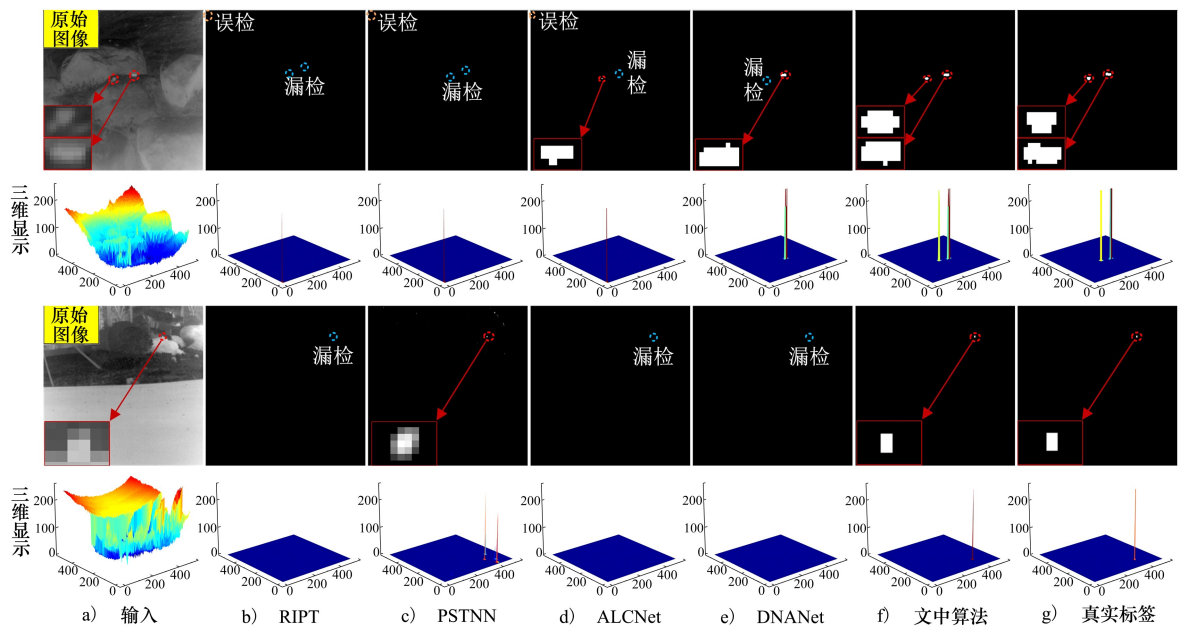


图 7 不同方法在 IRSTD-1k 数据集上的检测结果

色虚线框代表误检目标,蓝色虚线代表漏检目标,左下角为目标局部放大图。

从结果中可以看到,低秩建模的 RIPT 和 PSTNN 方法,在面对复杂背景时难以达到深度学习类方法的水平,漏检与误检情况较多,此外,建模类方法难以检测较暗的目标(如图 7b)和 7c)中的暗目标),且对背景中的亮区产生误检(如图 6b)和 6c)),本文的方法相比其他深度学习类方法漏检情况少,而 DNANet 在图 6 和图 7 均出现了漏检。此外本文方法描述目标形状的能力也较强,如图 6 第 3 行的图片,ALCNet 检测出的目标形状与真实标签仍有一定差距,本文的方法检测出的形状与真实标签更为接近。

此外从三维图中可以更直观地看出,本文的方法与真实标签最为接近,这进一步证明了本文的方法对红外小目标检测的有效性。

2.3.3 消融实验与对比实验

为了验证本文新构建模块的有效性与必要性,进行了消融实验,结果如表 2 所示。表 2 中的基线方法指 DNA-Net^[12],其参数量为 4.697 M,浮点计算量为 14.282 G。表 2 中的“+U 形注意力块”方法代表使用 U 形注意力块代替基线方法中对应模块,相比于基线方法参数量由 4.697 下降至 3.156,下降了约 32%,有利于提升训练的稳定性,降低模型复杂度,减少储存需求和计算开销,并且可以提升模型面

对不同场景时的泛化能力,因为 U 形注意力模块中具有较细粒度的多尺度提取能力,虽然其浮点计算量上升了 1.217 G,提升了约 8.5%,但 3 个指标全部有所优化,尤其是 IoU 由 87.09% 提升至 90.12%,U 形注意力块在性能与效率之间取得了较好平衡。表 2 中的“+逐像素注意力块”方法与基线方法对比,虽然逐像素注意力模块增加了模型复杂度与推理时间,但其却可以有效增加算法的检测精度与 IoU。本文方法为在基线方法上增添了 2 个模块的最终算法,在 3 个指标上全面超越了基线方法,这进一步证明了本文构造的 2 个模块可以提高网络对小目标特征的提取能力。从表 2 可看出,在单张图片的推理时间方面,本文的模型对比基线方法微小增加 0.01 s,达到了 25.6 frame/s,这主要是逐像素注意力增加的参数量导致。综合地,本文的方法在精度和效率之间取得了较好平衡。

本文还将逐像素注意力的模型与基线方法的特征图进行可视化对比显示,如图 8 所示。其中第一行是原始图片,第二行代表基线方法的可视化图,第三行代表本文的逐像素注意力的可视化图,通过对比可看出,逐像素注意力能使网络更加关注小目标区域,增强对小目标的响应值,降低背景区域的干扰。

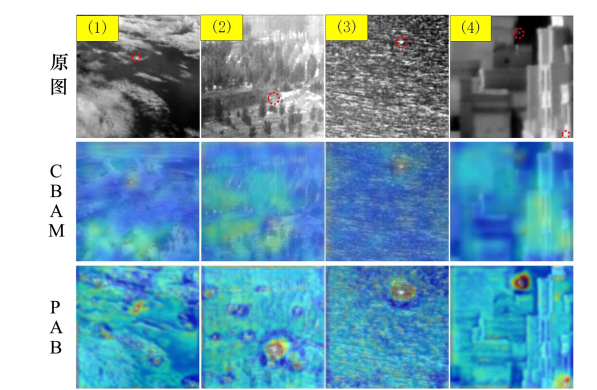


图 8 CBAM 注意力与逐像素注意力 PAB 可视化效果图

为了验证本文方法的鲁棒性,构建了如表 3 所示的交叉实验,其中 I-N 代表在IRSTD-1k 数据集上训练,NUDT-SIRST 进行测试的客观指标,N-I 代表在 NUDT-SIRST 数据集上训练,IRSTD-1k 进行测试的客观指标。可以看出本文的方法在IoU 这一综合指标上仍然最高,MDvsFA 方法 P_d 值虽然较高,但其误检过多,所以其 F_a 指标最差,ACM 方法漏检较多,其 F_a 虽然最好但 P_d 较差。而 ALCNet 与 DNANet 也难以超越本文提出的方法。综上所述,本文的方法具有较高的鲁棒性。

表 2 模型消融实验结果对比

编号	方法名称	$IoU/10^{-2}$	$P_d/10^{-2}$	$F_a/10^{-6}$	参数量/M	浮点计算量/G	单帧推理时间/s
1	基线方法	87.09	98.73	4.223	4.697	14.282	0.029
2	+U 形注意力块	90.12	98.83	4.205	3.156	15.499	0.033
3	+逐像素注意力块	91.09	99.15	3.930	4.797	14.499	0.041
4	本文算法	92.49	98.84	2.918	3.165	15.631	0.039

表 3 交叉实验结果对比

算法	I-N			N-I		
	$IoU/10^{-2}$	$P_d/10^{-2}$	$F_a/10^{-5}$	$IoU(\times 10^{-2})$	$P_d/10^{-2}$	$F_a/10^{-5}$
MDvsFA	18.47	85.61	2641.33	1.51	80.00	985.15
ACM	28.77	26.67	2.58	18.50	20.73	1.65
ALCNet	24.72	18.48	4.66	16.03	58.26	2.90
DNA-Net	40.19	71.53	4.66	36.02	76.97	7.13
本文算法	42.20	70.48	3.27	39.16	79.13	5.77

此外,为了验证本文方法的实用性,使用在NUDT-SIRST 上训练好的权重,在无人机数据集^[16]上挑选 2 个小尺度目标序列(序列名称见表 4),共计 3 000 张图片进行测试,评价指标为精度 P 和召回率 R 。因本文方法为像素级检测,而测试无人机数据集只有实例级标注信息,所以本文方法得到检测结果后,将其最大外接矩形作为实例级检测结果并计算评价指标,其检测指标如表 4 所示,检测结果示意图如图 9 所示,绿色框为标记框,红色为检测结果,从中可看出本文方法在反无人机领域具有一定效果。

表 4 真实无人机数据集测试结果

序列名称	$P/10^{-2}$	$R/10^{-2}$
3700000000002_095542_1	86.5	71.0
3700000000002_104622_06	92.2	61.3



图 9 真实无人机数据集检测效果图

3 结 论

本文针对红外小目标尺度小、细节少、复杂背景干扰的难题,提出了新的由U形注意力块和逐像素注意力块构成的红外小目标检测网络。首先,通过U形注意力块,提取目标细粒度特征信息,提升网络对背景与目标的区分能力;其次,利用稠密融合方式,充分保留小目标信息,确保目标信息传递到网络

深层,降低漏检率。最后,为了避免融合后目标特征衰减并且抑制复杂背景,本文构建了逐像素注意力模块,通过空间与通道维持的逐像素注意力块,避免压缩特征丢失过多的小目标信息,精细化逐像素增强小目标特征,降低背景与噪声的干扰。实验表明本文的方法优于现有经典基准检测方法。逐像素注意力虽可以有效提升网络性能,但其推理速度仍有待提高,这将是本文下一步的研究目标。

参考文献:

- [1] 张凯, 刘昊, 杨曦, 等. 基于关键点检测网络的空中红外目标要害部位识别算法[J]. 西北工业大学学报, 2020, 38(6): 1154-1162
ZHANG Kai, LIU Hao, YANG Xi, et al. Identification algorithm based on key-point detection network for vital parts of infrared aerial target[J]. Journal of Northwestern Polytechnical University, 2020, 38(6): 1154-1162 (in Chinese)
- [2] ZHA Y, WU M, QIU Z, et al. Online semantic subspace learning with siamese network for UAV tracking[J]. Remote Sensing, 2020, 12(2): 325
- [3] 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 in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3752-3767
- [4] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. Remote Sensing, 2019, 11(4): 382
- [5] 胡磊力, 张君昌, 张亮中. 基于人类视觉系统的实时红外目标检测方法[J]. 西北工业大学学报, 2017, 35(5): 910-914
HU Leili, ZHANG Junchang, ZHANG Liangzhong. Real-time infrared target detection method based on human vision system[J]. Journal of Northwestern Polytechnical University, 2017, 35(5): 910-914 (in Chinese)
- [6] 龚任杰, 郑智辉, 丛龙剑, 等. 小样本条件下异源图像迁移学习的红外目标检测与识别[J]. 西北工业大学学报, 2021, 39(增刊1): 84-88
GONG Renjie, ZHENG Zhihui, CONG Longjian, et al. Infrared target detection and recognition based on transfer learning with small samples[J]. Journal of Northwestern Polytechnical University, 2021, 39(suppl 1): 84-88 (in Chinese)
- [7] 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 Trans on Aerospace and Electronic Systems, 2020, 57(1): 485-496
- [8] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention, 2015
- [9] LI B, XIAO C, WANG L, et al. Dense nested attention network for infrared small target detection[J]. IEEE Trans on Image Processing, 2022, 32: 1745-1758
- [10] 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, 2022: 877-886
- [11] QIN X, ZHANG Z, HUANG C, et al. U2-Net: going deeper with nested U-structure for salient object detection[J]. Pattern recognition, 2020, 106: 107404
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, 2018: 3-19
- [13] 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, 2019: 8509-8518
- [14] 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, 2021: 950-959

[15] DAI Y, WU Y, ZHOU F, et al. Attentional local contrast networks for infrared small target detection[J]. IEEE Trans on Geoscience and Remote Sensing, 2021, 59(11): 9813-9824

[16] JIANG N, KUIRAN W, XIAOKE P, et al. Anti-UAV: a large-scale benchmark for vision-based uav tracking[J]. IEEE Trans on Multimedia, 2023, 25: 486-500

Pixel-wise attention driven infrared small target detection network

WANG Xiaolin, FANG Houzhang, LI Xueting, WU Chenxing, WANG Liming

(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

Abstract: Infrared small target detection has been widely used in the military and civil fields. The detection difficulties lie in small target size, few details and complex background interference. Existing classic deep learning detection methods are commonly suitable for generic target detection, but have poor adaptability to infrared small targets. To handle these problems, this article build a new infrared small target detection network based on U-shaped attention block and pixel-wise attention block. Firstly, a residual U-shaped attention block to extract multi-scale features through local U-shaped subnetworks in a single layer level and extract multiscale features to enrich the representation of small-scale target features is designed, so as to enhance the network's ability to discriminate small-scale targets. Then, small target information is further preserved through dense fusion method to alleviate the semantic gap feature fusion between different layers and reduce the miss detection rate. Finally, the pixel-wise attention in space and channel dimensions is applied to the fused feature map to enhance small targets and suppress complex background interference. The experimental results show that the present network outperforms the latest benchmark method in the intersection over union, detection probability and false alarm rate of two infrared small target data sets NUDT-SIRST and IRSTD-1k. Moreover, the present network also achieves a good balance between the detection accuracy and the efficiency.

Keywords: small infrared target detection; U-shaped attention; pixel-wise attention

引用格式:王啸林, 方厚章, 李雪婷, 等. 逐像素注意力驱动的红外小目标检测网络[J]. 西北工业大学学报, 2024, 42(2): 335-343

WANG Xiaolin, FANG Houzhang, LI Xueting, et al. Pixel-wise attention driven infrared small target detection network [J]. Journal of Northwestern Polytechnical University, 2024, 42(2): 335-343 (in Chinese)