文章编号:1001-9014(2022)06-1102-11 DOI:10.11972/j.issn.1001-9014.2022.06.020

结合跨尺度特征融合与瓶颈注意力模块的轻量型红外小目标检测网络

林再平*, 李博扬, 李 淼, 王龙光, 吴天昊, 罗伊杭, 肖 超, 李若敬, 安 玮
(国防科技大学 电子科学学院,湖南 长沙 410073)

摘要:提出一种结合跨尺度特征融合与瓶颈注意力模块的轻量型单帧红外小目标检测网络。该网络在不引入额外神经元的前提下,直接在编码层和解码层之间进行高频多尺度特征交互,从而维持小目标在网络深层的响应幅值,实现小目标浅层空间结构特征与深层高级语义特征之间的交互融合。同时,该网络在编码器瓶颈处级联轻量型混合注意力模块,进一步增强目标特征在网络深层的响应幅值。实验结果表明,该网络能有效抑制复杂背景杂波,并以较低参数量实现红外小目标检测。

关 键 词:红外小目标检测;轻量型算法;跨尺度融合;瓶颈注意力模块
中图分类号:TP753 **文献标识码**:A

Light-weight infrared small target detection combining cross-scale feature fusion with bottleneck attention module

LIN Zai-Ping*, LI Bo-Yang, LI Miao, WANG Long-Guang, WU Tian-Hao, LUO Yi-Hang, XIAO Chao, LI Ruo-Jing, An Wei
(College of electronic science and technology, National University of Defense Technology, Changsha 410073, China)

Abstract: This paper proposed a light-weight single frame infrared small target detection network that combined cross-scale feature fusion and bottleneck attention module. Instead of bringing extra huge neurons, the network directly performs cross-scale feature interaction between the encoding and decoding sub-networks, maintain the response of small target in the deep CNN layers, and thus achieves the full fusion between the spatial structure features from shallow layers and high-level semantic features from deep layers. Based on cross-scale feature fusion module, a light-weight bottleneck attention module is introduced to further enhance the response the target feature in the deep layers of the network. Experimental results demonstrate that the network can effectively suppress the complex background clutter and achieve high performance of infrared small target detection with low amount of parameters.

Key words: infrared small target detection, cross-scale feature fusion module(CFM), bottleneck attention module, light-weight method

引言

红外小目标检测在海洋资源利用、高精度导航、生态环境监测等领域有着广泛应用。区别于常规目标,如图1所示,红外小目标的目标尺寸更小,辐射强度更低且无显著形态特征,目标检出较为困难。

为实现高效目标检测,早期红外小目标检测算法主要采用图像滤波方法[1-3](Filtering-based Method,FM),该方法依据整张红外图像的不连续性来抑制背景噪声,从而实现目标检出。之后,Philip等人[4]对人类视觉系统(Human Vision System,HVS)进行了深入研究,并依据人类视觉系统关注局部区域目标显著性的特点提出了基于局部对比度的目标检测新范式(Local Contrast Based Method, LCM)。接下来,基于低秩(Local Rank Based Method, LRM)的方法[5-7]获得了广泛关注,该类方法将原始图像映

收稿日期:2022-06-13,**修回日期**:2022-08-31 **Received date**:2022-06-13,**Revised date**:2022-08-31
基金项目:国家自然科学基金(61972435, 61401474, 61921001, 62001478)
Foundation items:Supported by National Natural Science Foundation of China (61972435, 61401474, 61921001, 62001478)
作者简介(Biography):林再平(1982-),男,浙江人,副研究员,博士学位,主要研究领域为空间信息处理、天基光学弱小目标检测
E-mail: linzaiping@sina.com
* **通讯作者**(Corresponding author): E-mail: linzaiping@sina.com

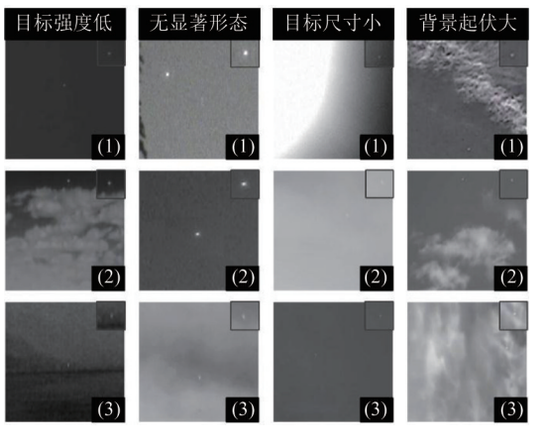


图1　红外小目标检测面临主要挑战

Fig. 1　The main challenges of infrared small target detection.

射在低秩子空间内,通过在该子空间内度量目标和背景间的不连续性来实现目标检出。总的来说,早期红外小目标检测算法的基本设计思路是将原始图像投影在局部显著性区域或多重线性子空间来扩大目标和背景之间的显著性差异,进而根据目标与背景间的不连续性进行目标检出。这些工作对红外小目标检测领域的初步发展起到了重要的推动作用。但由于其本质上过度依赖于专家知识与特征模板,当目标尺寸、目标形状、局部信噪比剧烈变化时,传统方法受固定模型参数的限制,难以适配多变的场景并且容易导致虚警和漏检情况的发生,从而导致算法鲁棒性不足。

近年来,基于数据驱动的深度学习方法在红外小目标检测领域得到了广泛关注。现有的红外小目标检测方法根据算法框架和标签形式差异可以分为基于框标注的目标检测方法以及基于逐像素标注的图像分割方法。Liu 等人[8]研究了多层感知机在红外小目标检测中的应用潜力,他们设计了一个5层深的多层感知机,并通过数据驱动的方式进行了端对端训练,取得了初步的检测效果。McIntosh 等人[9]进一步探索了通用目标检测方法 Faster RCNN[10], Yolo-v3[11]的应用性能,他们对通用目标检测框架生成的特征向量进行优化取得了进一步的性能提升。

近年来,基于图像分割目标检测范式成为了领域的研究焦点。该类方法可以直接输出逐个像素的目标定位及识别结果,便于开展后续的目标轨迹关联和定位识别任务。Dai 等人[12]提出了第一个基于图像分割方式的红外小目标检测网络(Asymmetric Contextual Module,ACM),他们设计了一个对称上下文模块来逐渐来集成浅层和深层的图像特征信息。接下来,Dai 等人[13]进一步改进了 ACM。他们将传统方法中的局部对比度度量思想融入检测算法中,设计了一种特征循环位移模块实现端对端网络训练框架,取得了进一步的性能提升。此外,Wang 等人[14]将红外小目标解耦成两个彼此对立的子任务,即高检测率子任务和低虚警率子任务,并采用条件生成对抗式网络(Conditional Generative Adversarial Network, CGAN)来实现两个子任务之间的纳什均衡。值得注意的是,Li 等人[15]首次提出深度学习范式下红外小目标检测的核心问题,即高层语义信息获取与多次下采样目标丢失之间的对立关系。为缓解两者对立,取得双赢结果,他们设计了一种基于密集嵌套并融合注意力模块的红外小目标检测网络(Dense Nested Attention Network, DNANet)。该方法通过密集地特征交互和反复地视觉注意力增强来维持小目标在深层网络的高幅值响应,从而有效挖掘目标特征中的高层语义信息,实现多层特征交互。该方法在不同尺度、形态和信噪比条件下都取得了显著的效果提升。

得益于数据驱动模式的赋能,红外小目标检测领域得到了蓬勃发展。然而,上述工作过度关注于提高算法的检测性能,缺乏对于实际部署应用环节中模型参数量、运算量等问题的综合考量。本文在前人研究的基础上,提出一种结合跨尺度融合与瓶颈注意力模块的轻量型红外小目标检测网络(Light-weight Infrared small target Detection Network, LIRDNet),该网络采用轻量型特征融合与特征增强模块来构建检测网络,在保证检测性能的同时,极大程度地减少模型参数量,便于后端开展部署应用。

1　网络结构

1.1　整体结构

整体网络结构遵循经典的编码器-解码器范式,即U型结构。如图2所示,该网络以单帧红外小目标图像作为输入,先后对其进行特征提取,特征融合,并对最终输出的逐像素预测结果进行聚类输出。1.2节介绍了跨尺度融合以及级联混合注意力模块的设计动机。首先对所有输入图像进行裁剪、翻转等常规预处理操作。然后,预处理后的图像被送入轻量化特征提取模块进行多个尺度下的特征提取。接下来,特征向量在编码器和解码器之间进行反复跨尺度跳跃交互,高层语义特征和底层语义

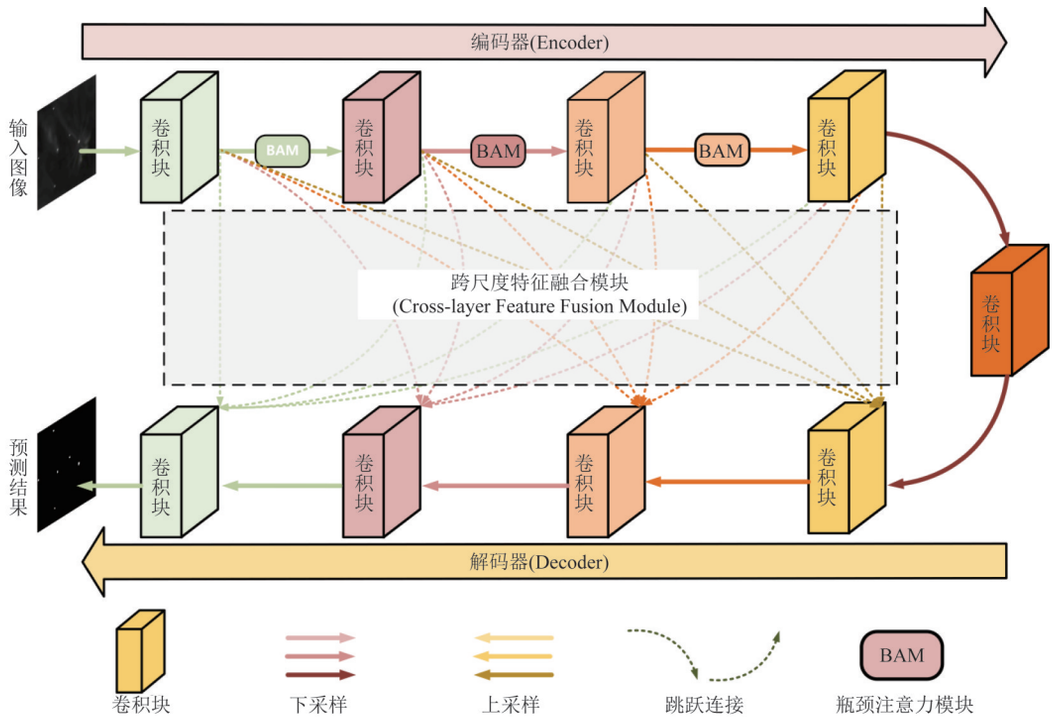


图2 轻量型红外小目标检测网络结构示意图

Fig. 2 An illustration of the proposed light-weighted infrared small target detection network

特征之间可由此进行初步的交互融合。由于红外小目标在全图占比较小,下采样过程中容易丢失甚至削弱该区域响应值,本文在编码器中的每个卷积块后级联一个轻量级的混合注意力模块,该模块由一个空间域注意力分支和一个通道域注意力分支并联组成。其可以有效增加红外小目标特征区域在网络深层的目标响应幅值,从而确保多尺度特征的交互融合。1.3节介绍了八连通聚类模块。经过上述检测网络输出的逐像素分类结果可以经过该模块进行判断,从而确定出各个像素点的所属的目标编号,方便进行评测和后续的目标关联及识别。

1.2 特征提取模块

1.2.1 设计动机

经典的U型分割网络如图3(a)所示,其由编码器、解码器和平行跳跃连接组成。其中,编码器通过多次卷积及其级联的下采样操作可以充分提取到不同尺度和不同语义层次的特征图。解码器用于恢复图像尺寸并实现多层特征图的渐进式融合。平行跳跃连接则起到连通两个子模块的作用,将底层目标轮廓信息和高层语义信息持续不断地传递至解码器。

为了有效地进行图像特征间的上下文信息交互,直观的解决方案是不断增加网络层数。通过这种方式,关联更多上下文信息的高层语义特征可以

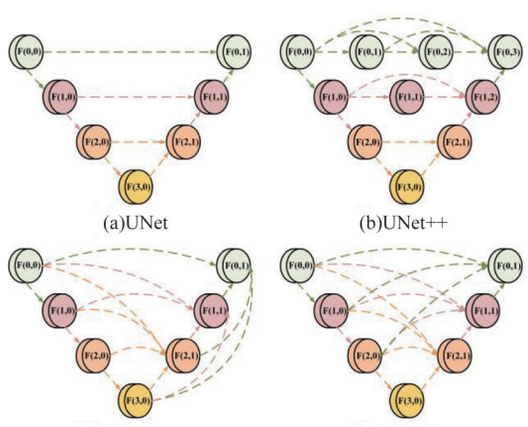


图3 典型U型分割网络结构及本文提出的LIRDNet

Fig. 3 The network of classic U-shape and our proposed LIRDNet

被有效提取出来。然而,红外小目标尺寸较小,通常在几个到几十个像元之间。随着网络层数的不断加深,内含多次下采样操作的池化过程使得神经网络在不断获取高层语义特征的同时,逐渐丢失小尺寸目标在其深层网络中的响应幅值。为解决上述问题,常规解决方案如图3(b)所示,该网络通过在编码器和解码器之间布置多个神经元,使图像特征可以在编码器和解码器之间的主干网络之外,通过跳跃连接结构进行特征间的高频交互,削弱下采

样过程带来的不利影响,从而保留红外小目标在网络深层的响应值。

尽管这种基于密集嵌套的解决方案可以实现图像高层语义信息的获取,保持红外小目标在网络深层的高频响应,但过于密集的叠加运算也不可避免地带来了过多的参数量和运算量,不利于红外小目标算法的低成本部署及快速应用。受如图3(c)所示的UNet+++[16]以及轻量化瓶颈注意力模块[17]的启发,本文首先设计了一种基于UNet+++改进的轻量化特征融合模块。如图3(d)所示,该模块在实现上下文信息交互融合的基础上,极大程度地降低了模型参数量。此外,本文引入的特征增强模块仅部署在各个卷积块之间,从而实现必要的红外小目标特征增强,相较于之前的密集注意力增强方案,极大程度地降低了模型运算量和参数量。

1.2.2　跨尺度融合模块

区别于密集嵌套的多尺度融合模式,如图3(d)所示,本文提出的跨尺度融合模块(Cross-Scale Feature Fusion Module, CFM)不依赖于额外引入的卷积单元,可直接实现编码器和解码器之间的特征交互融合。区别于常规的UNet+++,设计的CFM模块可以使得解码器的每一层都可接收到编码器的逐层语义特征。通过这种方式,囊括高层语义信息的深层特征和含有丰富图像轮廓信息的浅层特征可以充分融合,从而实现有效的上下文信息交互。

采用I层编码器结构来进行多层特征提取。为不失一般性,以图3(d)中编码器的第i层和第j个卷积块为例,对该跨尺度特征融合模块(CFM)的功能进行公式化描述。假设$\mathbf{F}_{\mathrm{en}}^{i,0}$代表编码器中第$i$层卷积块提取到的图像特征,$\mathbf{F}_{\mathrm{en}}^{i,0}$可以被表示为:

$$\mathbf{F}_{\mathrm{en}}^{i,0}=\mathbf{P}_{\max}(\varepsilon(\mathbf{F}^{i-1,0})) \quad , \quad (1)$$

其中$\varepsilon(.)$代表该层的级联卷积操作,$\mathrm{P}_{\max}(.)$代表步长为2的最大池化操作。$\mathbf{F}_{\mathrm{de}}^{i,1}$代表解码器中第$i$层卷积块提取到的图像特征,$\mathbf{F}_{\mathrm{de}}^{i,1}$可以被表示为:

$$\mathbf{F}_{\mathrm{de}}^{i,1}=\varepsilon(\varepsilon_{1\times1}\{[\mathrm{P}_{\max(2\times m)}(\mathbf{F}^{i-k,0})]_{k=1}^{K}],\ [\mathrm{U}_{2\times m}(\mathbf{F}^{i+m,0})]_{m=1}^{M},\mathbf{F}^{i,0}\}) \quad , \quad (2)$$

其中[·]代表特征图在通道维度的叠加操作,$\varepsilon_{1\times1}(.)$代表尺寸为1×1的卷积操作。$\mathrm{U}_{2\times m}(.)$代表步长为2×$m$的下采样操作,且$K$的取值范围为$K\in(1,i)$。$\mathrm{P}_{\max(2\times m)}(.)$代表步长为2×$m$的最大池化操作,且$M$的取值范围为$M\in\{1,I-i\}$。

1.2.3　瓶颈注意力模块

为保持红外小目标在网络深层的高频响应,密集嵌套范式在编码器的每一个卷积层后都级联一个混合注意力模块。尽管该方案有利于维持小目标在神经网络深层的响应幅值,但却引入了高额的参数量。区别于上述工作,本文引入轻量化注意力模块(Bottleneck Attention Module, BAM),仅在编码器各层末端布置一个级联的瓶颈注意力模块(BAM),在持续增强网络深层红外小目标响应幅度的同时,最大程度地降低网络参数量。如图4所示,对于给定的输入特征图$\mathbf{F}\in\mathbb{R}^{C\times H\times W}$,BAM可以产生一个与其尺寸相同的3D注意力权重图$M(\mathbf{F})\in\mathbb{R}^{C\times H\times W}$,通过将输入特征和3D注意力权重图相乘可以得到优化后的特征图,具体计算过程如下所示:

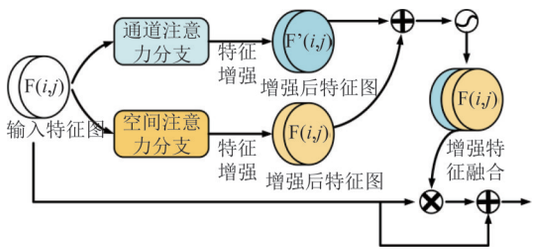


图4　瓶颈注意力模块

Fig. 4　Bottleneck attention module

$$\mathbf{F}'=\mathbf{F}+\mathbf{F}\otimes M(\mathbf{F}) \quad , \quad (3)$$

其中,⊗代表逐像素乘法操作。瓶颈注意力模块(BAM)由空间域注意力单元$M_{\mathrm{s}}(.)$和通道域注意力单元$M_{\mathrm{c}}(.)$并联组成,其对特征图的增强过程如式(3)所示:

$$M(\mathbf{F})=\sigma(M_{\mathrm{c}}(\mathbf{F})+M_{\mathrm{s}}(\mathbf{F})) \quad , \quad (4)$$

其中,$\sigma(.)$代表sigmoid函数。待增强的特征图$\mathbf{F}$经过两个注意力分支处理后,分别产生对应的空间域增强特征$M_{\mathrm{s}}(\mathbf{F})$和通道域增强特征$M_{\mathrm{c}}(\mathbf{F})$。通过非线性激活函数将增强后的两部分特征有效整合,最终生成空间域和通道域共同增强后的特征图。下面将分别介绍上文所提到的通道和空间注意力分支。

(1) 通道注意力分支:该分支主要用于定向增强小目标高频响应通道的权重。该分支首先对输入的特征图$\mathbf{F}\in\mathbb{R}^{C\times H\times W}$进行全局平均池化操作,从而产生一个通道数为$C$,空间尺寸为1×1特征向量$\mathbf{F}\in\mathbb{R}^{C\times1\times1}$。该特征向量将特征图的全局信息软耦合在了其内含的各个通道中,然后,该模块采用含有一个隐藏层的多层感知机(Multi-layer Perception, MLP)来度量各个通道的重要性,进而计算出其对应权重。为了进一步缩减参数量,降低运算复

杂度，本文将隐藏层神经元的尺寸设置为 $\mathbb{R}^{C/r\times 1\times 1}$，其中 r 代表通道数下采样比例，该参数是一个可以调节的超参数。最后，通过批归一化操作（Batch Normalization，BN）将多层感知器处理后特征图 $\mathbf{F}$ 的空域尺寸进行重调，使之和空间注意力分支的特征图尺寸保持一致。综上所述，通道注意力分支可以通过下述公式计算得到：

$$M_c(\mathbf{F}) = BN(MLP(AvgPool(\mathbf{F}))) = BN(\mathbf{W}_1(\mathbf{W}_0 AvgPool(\mathbf{F}) + \mathbf{b}_0) + \mathbf{b}_1), \quad (5)$$

其中，$\mathbf{W}_0 \in \mathbb{R}^{C/r\times C}$，$\mathbf{b}_0 \in \mathbb{R}^{C/r}$，$\mathbf{W}_1 \in \mathbb{R}^{C\times C/r}$，$\mathbf{b}_0 \in \mathbb{R}^{C}$ 代表多层感知机的隐藏层权重参数。

（2）空间注意力分支：空间注意力分支的作用是定向增强具有视觉显著性的局部区域的响应权重。该分支首先对输入的特征图F进行一个1×1卷积操作，将特征图 $\mathbf{F}\in\mathbb{R}^{C\times H\times W}$ 的尺寸重调为 $\mathbf{F}\in\mathbb{R}^{C/r\times H\times W}$。然后，对处理后的特征图先后施加两次3×3空洞卷积操作。在该分支中，使用空洞卷积替代传统卷积的目的在于空洞卷积可以进一步扩张卷积核的感受野，从更大的上下文邻域中捕捉出具备目标局部显著性的高辨别力区域，进而增加该分支的特征增强效果。最后，通过1×1卷积和批归一化操作（Batch Normalization，BN）将增强后的特征图尺寸与通道注意力分支的输出尺寸一致。综上所述，空间注意力分支可以通过下述公式计算得到：

$$M_s(\mathbf{F}) = BN(\varepsilon^3_{1\times1}(\varepsilon^2_{3\times3}(\varepsilon^1_{3\times3}(\varepsilon^0_{1\times1}(\mathbf{F}))))), \quad (6)$$

其中 $\varepsilon_{1\times1}(.)$ 和 $\varepsilon_{3\times3}(.)$ 分别代表尺寸为1×1和3×3的卷积操作。各层编码器、解码器卷积参数如表1所示。

1.3 八连通聚类模块

基于图像分割范式的红外小目标检测网络可以输出与原图尺寸一致的逐像素定位及分类结果，但缺少将对应坐标点聚合为特定目标的聚类过程。因此，本文引入了通用的八连通聚类算法[18]来解决上述问题。对于网络输出结果 $\mathbf{G}$ 中的任意两个像素点 $g_{(m0,\,n0)}$ 和 $g_{(m1,\,n1)}$，如果将其判定为同一个类别需要满足以下两个条件。

条件1：如果两点在其相邻的八个邻域中存在交集区域，即满足：

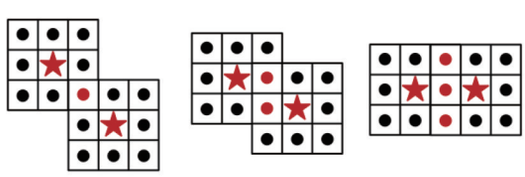
$$\mathbb{N}_{8(m_0,\,n_0)} \cap \mathbb{N}_{8(m_1,\,n_1)} \neq \varnothing \quad , \quad (7)$$

其中，$\mathbb{N}_{8(m_0,\,n_0)}$ 和 $\mathbb{N}_{8(m_1,\,n_1)}$ 分别代表 $(m_0,\,n_0)$ 和 $(m_1,\,n_1)$ 两个像素点的八邻域，则判定两个像素点满足条件1。

表1 LIRDNet模型参数及输入输出尺寸

Table 1 Number of model parameters for LIRDNet and the corresponding size of input and output feature map

编号	滤波器数量	输入尺寸	输出尺寸
预处理	–	1,256,256	3,256,256
$F'^{0,0}_{de}$	8	3,256,256	8,256,256
$BAM^{0,0}$	–	8,256,256	8,256,256
$F'^{1,0}_{de}$	16	8,128,128	16,128,128
$BAM^{1,0}$	–	16,128,128	16,128,128
$F'^{2,0}_{de}$	32	16,64,64	32,64,64
$BAM^{2,0}$	–	32,64,64	32,64,64
$F'^{3,0}_{de}$	64	32,32,32	64,32,32
$F'^{2,1}_{de}$	32	64,64,64	32,64,64
$F'^{1,1}_{de}$	16	32,128,128	16,128,128
$F'^{0,1}_{de}$	8	16,256,256	8,256,256
F'_{final}	1	8,256,256	1,256,256



(a)单域连通 (b)双域连通 (c)三域连通

图5 八连通域聚类示例。两个候选点（★）的八邻域有交集区域（红色●）则判定上述候选点同属一个目标ID

Fig. 5 Samples of eight connected neighborhood clustering module. If the eight neighborhoods of two candidate points have intersection area, they are identified as the same target ID.

条件2：如果两个像素点有相同的取值（0或1），即满足：

$$g_{(m_0,\,n_0)} = g_{(m_1,n_1)}, \forall g_{(m_0,\,n_0)}, g_{(m_1,\,n_1)} \in \mathbf{G} \quad , \quad (8)$$

其中，$g_{(m_0,\,n_0)}$ 和 $g_{(m_1,\,n_1)}$ 代表 $(m_0,\,n_0)$ 和 $(m_1,\,n_1)$ 两个像素点的灰度值，则两个像素点满足条件2。同时满足条件1和条件2的两个像素点可以被判定为同属于一个相同的目标ID。通过八连通聚类模块，可将LIRDNet预测输出的全图所有像素点聚类为多个特定的目标ID，方便进行算法检测率和虚警率评估以及后续目标关联和目标识别任务的开展。

2 实验结果与分析

2.1 评价指标

精准的目标定位与目标轮廓评估是红外小目

标后续任务(如目标关联和目标识别)的基础和前提。因此,本文采用和DNANet[15]相同的评价指标,使用检测率(Probability of Detection, PD)和虚警率(False-Alarm Rate, FA)评估目标定位精度,交并比(Intersection over Union, IoU)评估目标轮廓精度。各个评价指标的定义和公式如下所示:

(1) 交并比:交并比是图像分割领域常用的评价指标,其主要用于评价算法对于目标轮廓的精准描述能力。该指标通过计算预测结果与真值结果两者间的交集区域占两者间的并集区域的比例得到,计算式如下:

$$\mathrm{IoU}=\frac{A_{\mathrm{inter}}}{A_{\mathrm{Union}}}, \quad (9)$$

其中,A_{inter}和A_{Union}分别代表交集区域和并集区域。

(2) 检测率:区别于交并比,目标检测率并不关注单个像素的判断准确率。其作用在于评估全图像素聚合之后各个目标的定位精度,计算式如下:

$$P_{\mathrm{d}}=\frac{T_{\mathrm{correct}}}{T_{\mathrm{all}}}, \quad (10)$$

如果该目标的质心与真值质心之间的距离小于预先设定的距离阈值T_{distance},则认为目标被正确检出。本文设置默认距离阈值T_{distance}等于3。

(3) 虚警率:目标虚警率的作用在于评估错误检出的像素数占全图像素数的比值,计算式如下:

$$F_{\mathrm{a}}=\frac{P_{\mathrm{false}}}{P_{\mathrm{All}}}, \quad (11)$$

如果该目标的质心与真值质心之间的距离大于预先设定的离阈值T_{distance},则认为该目标所含像素点为虚警像素点。本文设置默认距离阈值T_{distance}等于3。

2.2　实验设置

本文采用公开的NUAA-SIRST数据集进行算法性能评估,并对上述数据集设置两种训练测试条件。训练测试条件1采用和该数据集原文一致的设置条件,即256张图像用于训练,86张图像用于测试。训练测试条件2减少训练样本量至213张图像,并将测试集样本量扩充至214张。网络训练之前,首先将所有输入图像归一化处理。然后,对上述归一化后的图像进行预处理。预处理过程包括翻转、模糊、裁剪等常规操作。接下来,将输入图像的图像分辨率被重调至256×256后送入网络进行训练。

为客观评估本文提出的算法有效性,本文采用了VGG[19]、ResNet[20]系列等领域公认的骨干网络进行广泛的对比实验。值得注意的是,骨干网络仅部署在网络的编码器部分,且编码器的下采样次数为3。在训练过程中,本文采用Soft-IoU损失函数指导网络训练,使用Adagrad[21]和CosineAnnealingLR分别作为训练优化器和学习率调度器。此外,本文将初始学习率,批尺寸和迭代轮数分别设置为0.05,32和1 000。所有模型均采用Pytorch深度学习框架[22]进行训练和评估。计算设备采用AMD Ryzen 9 3950X CPU和一块Nvidia GeForce 3090 GPU。

2.3　对比实验及分析

为评估本文所提出算法在检测精度和检测效率等方面的综合性能,本章节引入了多种基于模型驱动的传统算法进行对比,如Top-Hat[1]、Max-Median[2]、TLLCM[24]、IPI[5]和RIPT[6]。同时与领域内认可

表2　不同方法在交并比、检测率、虚警率等客观评价指标上的表现

Table 2　Performance of different methods on IoU, Pd, and Fa

方法	NUAA-SIRST (条件1)			NUAA-SIRST (条件2)		
	IoU/(%)	Pd/(%)	Fa (10^{-6})	IoU/(%)	Pd/(%)	Fa (10^{-6})
Top-Hat	17.66	82.56	34.95	7.143	79.84	1012
Max-Median	3.90	52.29	49.32	4.172	69.20	55.33
TLLCM	0.96	77.98	5829	1.029	79.09	5899
IPI	22.77	86.23	10.65	25.67	85.55	11.47
RIPT	11.24	77.98	17.03	11.05	79.08	22.61
MDvsFA-CGAN	63.26	90.75	49.33	60.30	89.35	56.35
ACM	71.78	96.33	3.570	70.33	93.91	**3.728**
ALCNet	74.39	97.16	22.77	73.33	96.57	30.47
DNANet-Light	74.46	98.19	15.79	74.72	96.95	18.18
LIRDNet-ResNet10	72.52	97.24	22.17	73.47	**97.71**	26.23
LIRDNet-ResNet18	76.47	98.02	16.85	74.89	97.33	16.09
LIRDNet-ResNet34	**77.81**	**99.21**	**1.240**	**75.18**	97.33	7.060

表3 不同深度学习方法参数量、浮点运算量和检测精度性能比较

Table 3 Performance of different deep learning-based methods on the number of model parameters, FLOPs, IoU, Pd, and Fa

方法	#Params	FLOPs	mIoU/Pd/Fa
ACM	0.52 M	1.75 G	71.78/96.33/3.570
ALCNet	0.50 M	1.48 G	74.39/97.16/22.77
MDvsFA-cGAN	3.76 M	868.75 G	63.26/90.75/49.33
DNANet-Light	0.48 M	1.88 G	74.46/98.19/15.79
LIRDNet-Res18	0.25 M	1.43 G	76.47/98.02/16.85

的多个基于数据驱动的深度学习如MDvsFA-cGAN[14]、ACM[12]、ALCNet[13]、DNANet[15]等对比算法进行了广泛的对比。定量和定性结果分析显示本文提出的LIRDNet在较低参数量和运算量的前提下，取得了更全面的检测效果。具体分析过程见2.3.1和2.3.2节。

2.3.1 定量结果分析

对于所有传统算法，本文采用基于单图自适应阈值的方法进行全图虚警抑制，自适应阈值($T_{adaptive}$)可以根据如下式计算得到：

$$T_{adaptive} = Max[Max(G) \times 0.7, 0.5 \times \sigma(G) + avg(G)], \quad (12)$$

其中Max(G)代表预测图中最亮点的响应值。$T_{adaptive}$代表自适应阈值，$\sigma(G)$和avg(G)代表预测结果全图的标准差和均值。对于基于深度学习的对比算法，本文采用各个算法原文采用的固定阈值。具体来说，对ACM，ALCNet，和MDvsFA-cGAN，DNANet分别采用0，0，0.5，和0的固定阈值进行虚警抑制。定量化对比实验结果如表2所示，本文所提出的LIRDNet相较于传统对比算法有较大程度的性能提升。其原因在于NUAA-SIRST数据集内包含大量有挑战性的场景，例如信噪比多变、背景起伏较大、目标尺寸多样等。LIRDNet可以充分发挥数据驱动的模型优势，从多样化的训练数据中充分学习到具有高度判别性的语义特征，从而实现高鲁棒性的目标检测结果。传统算法如IPI和RIPT等由于过于依赖专家知识，需要人为手动地筛选超参数来使

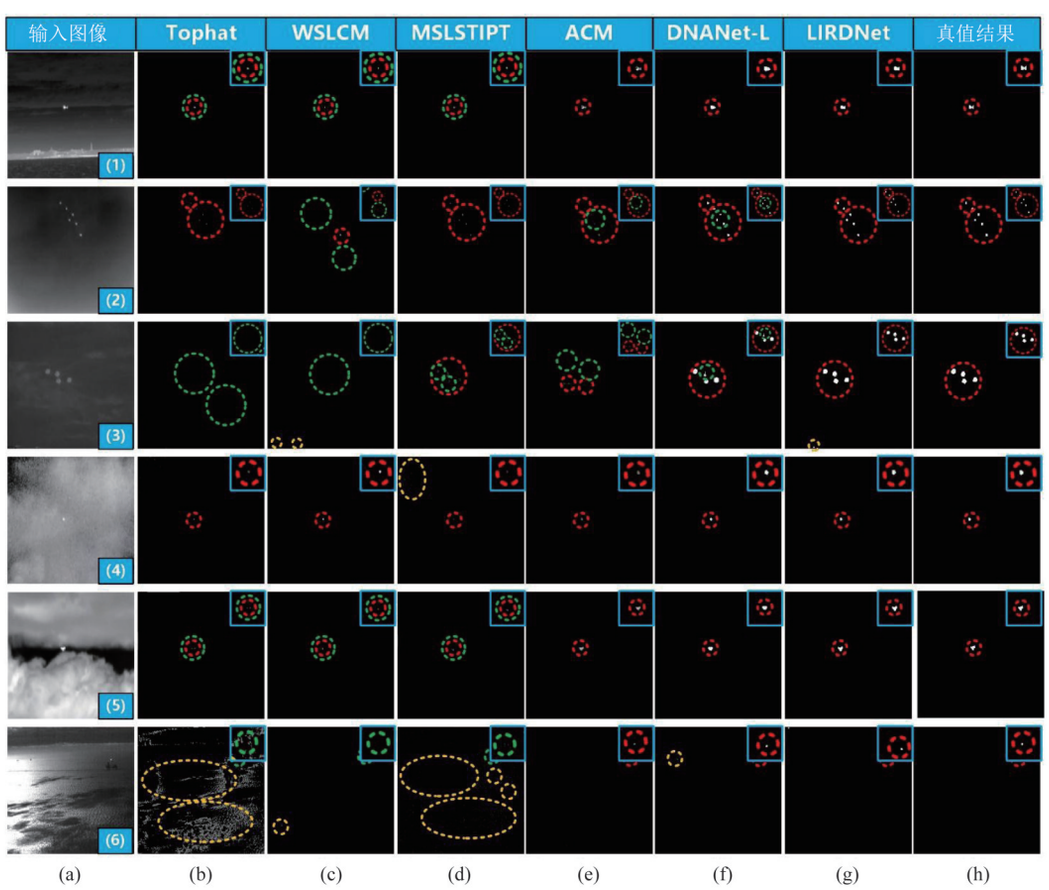


图6 本文算法和6种对比算法在NUAA-SIRST上的定性实验结果对比图

Fig. 6 Examples of (a) original images and corresponding qualitative comparison results on (b) Tophat, (c) IPI, (d) RIPT, (e) ACM, (f) DNANet, (g) LIRDNet, (h) ground truth masks.

算法适应多样化应用场景,当多种挑战并存,场景变化剧烈时,上述算法的检测性能被极大约束。

表 4　不同深度学习方法在不同计算单元下的推理时间、浮点运算量性能比较

Table 4　Inference time and FLOPs performance of different deep learning-based methods on different computational units (Smart Phone-Chip, PC-GPU)

方法	计算单元	推理时间 /s	浮点运算量 /G
DNANet-Light	天玑 800U	0. 322	1. 88 G
DNANet-Light	麒麟 980	0. 211	1. 88 G
DNANet-Light	Nvidia 1070	0. 102	1. 88 G
DNANet-Light	Nvidia 3090	0. 005	1. 88 G
LIRDNet	天玑 800U	0. 198	1. 43 G
LIRDNet	麒麟 980	0. 097	1. 43 G
LIRDNet	Nvidia 1070	0. 076	1. 43 G
LIRDNet	Nvidia 3090	0. 002	1. 43 G

相较于本领域的深度学习对比算法,如 ACM、ALCNet,本文所提出的 LIRDet 在更低的参数量以及更小的运算量前提下取得了更高精度的检测效果。其原因在于本文所提出的 CFM 和 BAM 模块可以很好地保持并增强红外小目标在网络深层的特征响应。此外,如表 3 所示,相较于 DNANet 等基于密集嵌套和持续增强的网络结构,本文提出的算法可以在检测精度降低不多的前提下,极大程度地降低网络的参数量和运算量。其原因在于本文所提出的 CFM 模块直接在编码器和解码器之间进行多尺度交互,该过程并不会引入额外的神经元。同时相较于 DNANet 中反复特征增强的网络结构,本文采用的 BAM 模块仅在各个卷积块之后进行必要且低频次的特征增强,从而使得深层目标特征被增强的同时,不引入额外参数量。

为评估不同红外小目标检测算法在实际部署应用环节的应用性能。本文将多种红外小目标检测算法在多种端侧计算单元(如华为麒麟 980、天玑 800U、英伟达 1070、英伟达 3090)上的进行了实际应用部署测试。实验结果如表 4 所示,无论是在智能芯片端侧还是移动 PC 端侧,本文提出的 LIRDNet 的实际所需推理时间远小于与其检测性能相近的 DNANet-Light 所需的推理时间,该结果论证了本文

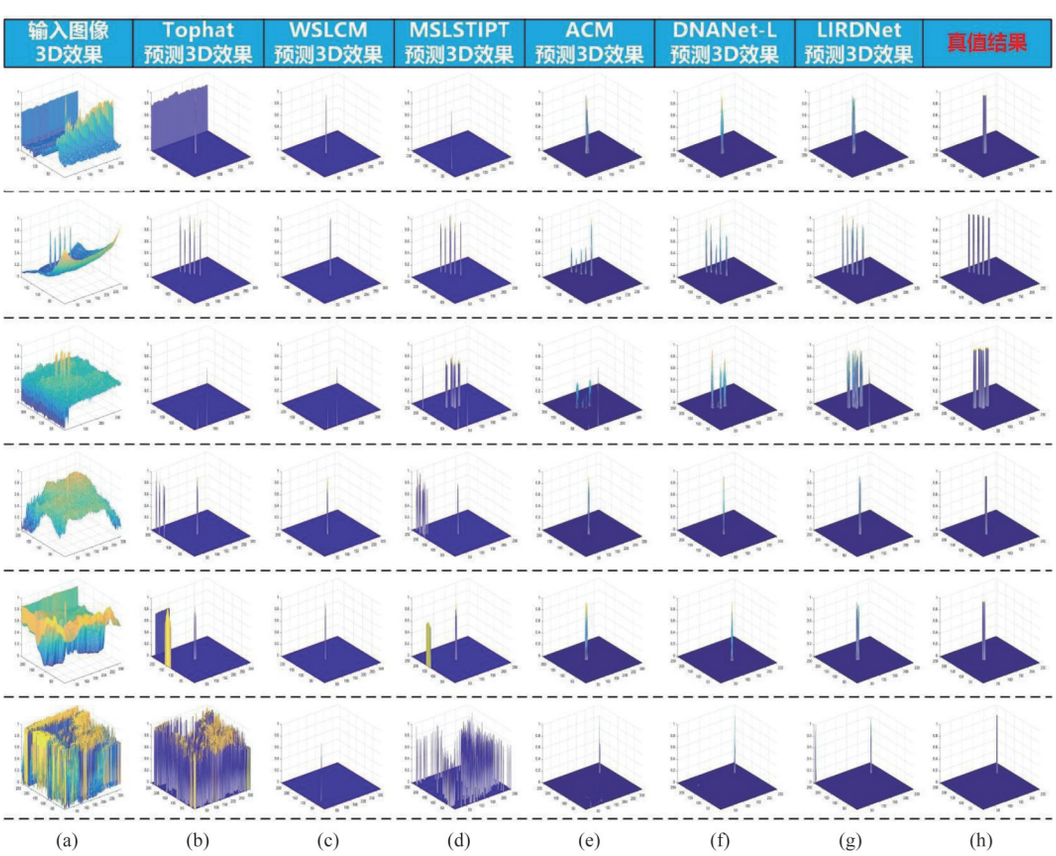


图 7　本文算法和 6 种对比算法在 NUAA-SIRST 上的 3D 可视化结果对比图

Fig. 7　Examples of (a) original images and corresponding 3D visualization results on (b) Tophat, (c) IPI, (d) RIPT, (e) ACM, (f) DNANet, (g) LIRDNet, (h) ground truth masks.

表5 CFM模块的消融实验结果

Table 5 Ablation study on our proposed CFM module

Method	#Params	FLOPs	mIoU/Pd/Fa
LIRDNet-Res18 w/o CFM	0.232M	1.184G	73.01/96.58/24.13
LIRDNet-Res18 w/o CFM L1/L2	0.234M	1.204G	73.39/97.16/34.59
LIRDNet-Res18 w/o CFM L1	0.243M	1.362G	74.23/97.16/24.37
LIRDNet-Res18	0.248M	1.435 G	76.47/98.02/16.85

提出的轻量化设计算法在实际部署应用环节的应用价值。本实验所涉及到的部署过程将于不久后以APP的形式公开在https://github.com/YeRen123455/Infrared-Small-Target-Detection.

表6 BAM模块的消融实验结果

Table 6 Ablation study on our introduced BAM module

Method	#Params	FLOPs	mIoU/Pd/Fa
LIRDNet-Res18 w/o BAM	0.245M	1.415 G	74.52/96.58/21.29
LIRDNet-Res18 w/o BAM SA	0.247M	1.422 G	75.37/97.16/16.14
LIRDNet-Res18 w/o BAM CA	0.247M	1.434 G	75.43/97.24/25.54
LIRDNet-Res18	0.248M	1.435 G	76.47/98.02/16.85

2.3.2 定性结果分析

定性结果如图6所示，红黄绿虚线框分别代表目标、虚警和漏检区域。相较于传统算法，本文提出的LIRDNet在更高检测率和更低虚警率的前提下，具备更强的目标定位和目标轮廓分割能力。从图6中可以看到，LIRNet对于图6-(2)(3)所示的斑目标和扩展目标的目标轮廓分割更为精准，且成功定位到了所有弱小目标。定性实验结果表明，相较于深度学习算法，LIRDNet对复杂场景的适应能力更强，且具备对多种目标类型，如无形态点目标及有形态斑目标和扩展目标的适应能力。

此外，为评估本文提出的算法在不同信杂比(signal-clutter-ratio, SCR)条件下的检测性能波动情况。本文将NUAA-SIRST数据集中的测试集按照信杂比区间分为了信杂比小于3，信杂比介于3-6之间和信杂比大于6三个区间，并在上述三个区间上针对不同阈值条件下的目标检测率和虚警率变化情况绘制了ROC曲线。如图8所示，随着信杂比

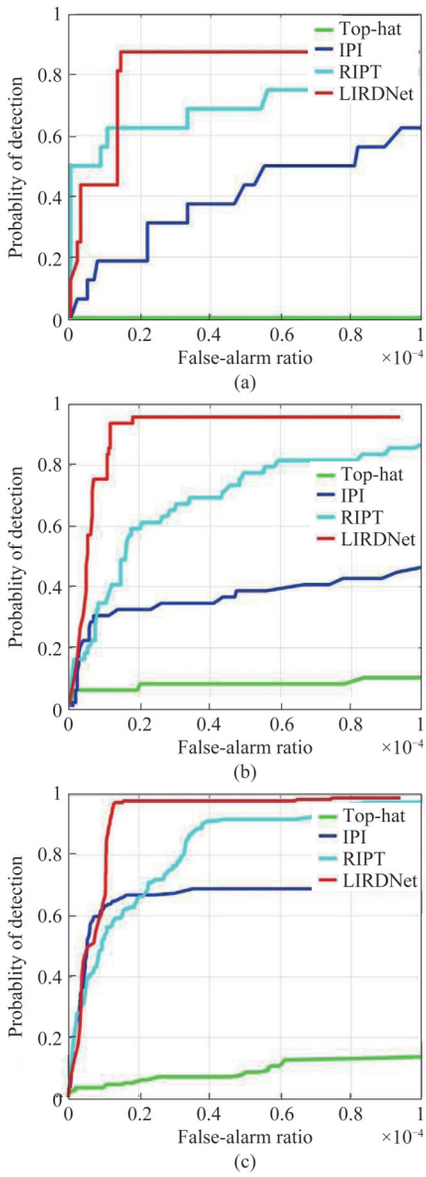


图8 本文算法在不同信杂比条件下(a)信杂比小于3，(b)信杂比介于3-6，(c)信杂比大于6的ROC曲线

Fig. 8 The ROC curve of our proposed LIRDNet under different signal-clutter-ratio (SCR) values (a) SCR<3, (b) 3<SCR<6, (c) 6<SCR.

不断提升，本文提出的LIRDNet的检测性能总体呈微弱的上升趋势，这主要得益于深度学习算法良好的泛化性能。受益于不同信杂比区间的充足训练数据，LIRDNet总体受信杂比变化的波动较小。此外，从图8中可以看到LIRDNet相较于Tophat等传统检测算法在不同信杂比区间都可以取得更好的综合检测能力。

2.4 消融实验

2.4.1 跨尺度融合模块

跨尺度融合模块通过直接连接编码器和解码

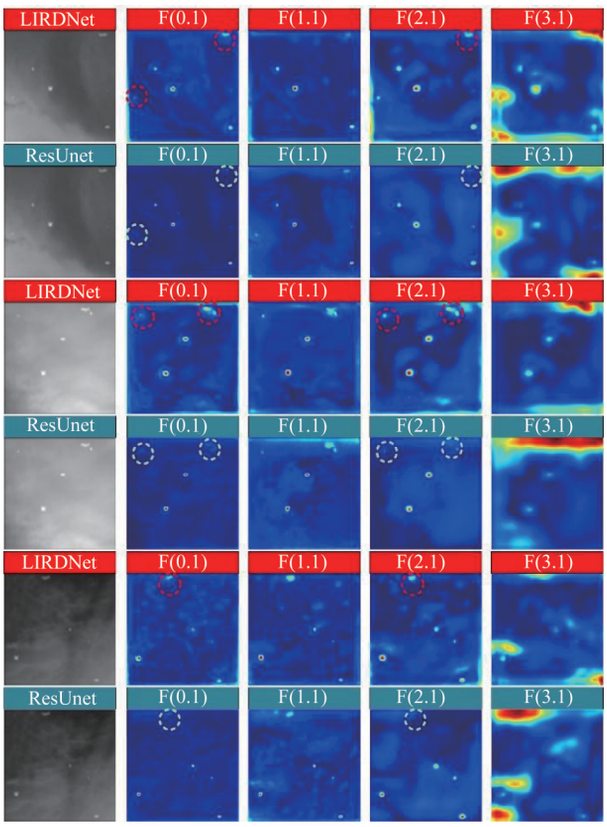


图9　本文所提出的LIRDNet和骨干网络ResUnet在不同层神经元上的检测结果特征响应图

Fig. 9　Visualization map of our proposed LIRDNet and backbone network ResUnet on different convolutional layers

器来实现反复跨尺度跳跃交互,从而保持小目标特征在网络深层的响应幅度,确保高层语义特征提取。为验证跨尺度融合模块(CFM)的有效性,本章节开展了三组消融实验。CFM及其各个变种模块的应用性能如表5所示。LIRDNet相较于不添加CFM模块的网络结构,在NUAA-SIRST数据集上的IoU、Pd和Fa三个指标上分别取得了3.46%, 1.44%和7.28×10^{-6}的性能增益。其原因在于目标尺寸较小,多次池化操作将削弱甚至丢失目标在深层网络的响应值,使得高层语义特征无法正常提取,从而导致性能下降。

接下来,本章节逐层移除LIRDNet中的CFM模块,实验结果如表5所示。随着CFM模块被逐渐移除,网络的各项检测性能也在逐渐下降。相较于完整的CFM结构,移除一层和移除两层CFM结构会在IoU、Pd和Fa三个指标上分别造成2.24%/3.08%, 0.86%/0.86%和7.52×10^{-6}/17.74×10^{-6}的性能下降。该实验结果表明,各层间充分地特征交互有利于多尺度特征的信息交互融合,从而产生高精度的检测结果。

2.4.2　瓶颈注意力模块

瓶颈注意力模块模块通过在编码器各层卷积块之间级联瓶颈注意力单元来进一步增强小目标在网络深层的响应幅值。为验证瓶颈注意力模块(BAM)的有效性,本章节开展了三组消融实验。BAM及其各个变种模块的应用性能如表6所示。

LIRDNet相较于不添加BAM模块的网络结构,在NUAA-SIRST数据集上的IoU、Pd和Fa三个指标上分别取得了1.95%,1.44%和4.44×10^{-6}的性能增益。其原因在于BAM中内含了通道和空间域注意力分支,两个分支可以同时增强具有高度判别性的目标特征图通道和局部空间邻域,进而提高相应区域的响应幅值,保证特征间地顺利传递与交互,从而产生高精度的检测结果。

接下来,本章节逐步移除LIRDNet中的BAM模块的通道域和空间域注意力分支,实验结果如表6所示。随着BAM模块中各个注意力分支被逐渐移除,网络的各项检测性能也在逐渐下降。相较于完整的BAM结构,移除通道域和空间域注意力分支会在IoU、Pd和Fa三个指标上分别造成1.04%/1.10%, 0.78%/0.86%以及8.69×10^{-6}/0.71×10^{-6}的性能下降。该实验结果表明,瓶颈注意力模块可以在仅仅引入0.003M参数量和0.020GFLOPs运算量的前提下,带来显著的检测性能增益,从而论证了必要的特征增强会对红外小目标检测带来正向结果这一观点。

最后,如图9所示,得益于本文所提出的CSF和BAM两个模块带来的多层特征融合和特征增强能力,LIRDNet在各个卷积块后输出的特征响应图在目标区域的响应幅值明显高于仅包含骨干网络的ResUNet。该响应幅值增益随着网络层数的逐渐增加而不断累积,从而使得网络最终得以赋予目标区域更高的权重以支持最后的高精度目标检出。

3　结论

本文创新性地提出了一种基于跨尺度特征融合与瓶颈注意力模块(BAM)相结合的轻量型红外小目标检测网络。具体结论如下:

(1) 跨尺度特征融合模块具备较强的特征提取能力,该模块可以在不引入额外神经元的前提下提取多尺度目标特征,从而保证了小目标在网络深层的特征响应。实验结果表明,CFM模块在仅引入0.016M参数量和0.251GFLOPs的前提下,显著地提升了网络的检测性能。

（2）瓶颈注意力模块进一步增强了红外小目标在网络深层的响应幅值，从而保证了高层语义和高判别性特征的进一步提取，底层目标轮廓特征和深层高层语义特征可以充分交互融合。实验结果表明，BAM模块在仅仅引入0.003M参数量和0.020 GFLOPs运算量的前提下进一步改善了网络的检测性能。

综上所述，本文提出的LIRDNet模型，通过设计和引入轻量型特征提取和特征增强模块，较好地缓解了红外小目标检测中高检测精度和高检测效率之间的对立关系。定性和定量实验结果表明，本文提出的LIRDNet在引入微量参数的前提下，在公开单帧红外小目标检测数据集NUAA-SIRST上的IoU、Pd和Fa三个指标上分别取得了76.47%、98.02%和16.85×10^{-6}的检测性能，相较于最近公开的算法取得了最佳的检测结果。相关代码将于不久后公开在https://github.com/YeRen123455/Infrared-Small-TargetDetection.

References

[1] Rivest J F, Fortin R. Detection of dim targets in digital infrared imagery by morphological image processing[J]. *Optical Engineering*, 1996, **35**(7):1886-1893.

[2] 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. International Society for Optics and Photonics, 1999, **3809**: 74-83.

[3] Wang P, Tian J W, Gao C Q. Infrared small target detection using directional highpass filters based on LS-SVM [J]. *Electronics letters*, 2009, **45**(3):156-158.

[4] 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.

[5] 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.

[6] 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.

[7] Sun Y, Yang J, An W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **59**(5):3737-3752.

[8] Liu M, Du H, Zhao Y, *et al*. Image small target detection based on deep learning with SNR controlled sample generation[J]. *Current Trends in Computer Science and Mechanical Automation*, 2017, **1**:211-220.

[9] 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*, 2020, **57**(1):485-496.

[10] Ren S, He K, Girshick R, *et al*. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. *Advances in neural information processing systems*, arXiv , 2015:28-40.

[11] Redmon J, Farhadi A. "Yolov3: An incremental improvement," *arXiv preprint* arXiv:1804.02767, 2018.

[12] 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.

[13] 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.

[14] 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.

[15] Li B, Xiao C, Wang L, *et al*. Dense nested attention network for infrared small target detection[J]. *IEEE Transactions on Image Processing*, 2022, Early Access.

[16] Huang H, Lin L, Tong R, *et al*. Unet 3+: A full-scale connected unet for medical image segmentation [C]// ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020:1055-1059..

[17] Park J, Woo S, Lee J Y, *et al*. Bam: Bottleneck attention module[J]. *arXiv preprint* arXiv:1807.06514, 2018..

[18] Wu K, Otoo E, Shoshani A. Optimizing connected component labeling algorithms[C]//Medical Imaging 2005: Image Processing. SPIE, 2005, **5747**:1965-1976.

[19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint* arXiv:1409.1556, 2014.

[20] 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. 2016: 770-778.

[21] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. *Journal of machine learning research*, 2011, **12**(7): 102-110.

[22] 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. International Society for Optics and Photonics, 1999, **3809**: 74-83.

[23] Paszke A, Gross S, Massa F, *et al*. Pytorch: An imperative style, high-performance deep learning library[J]. *Advances in neural information processing systems*, 2019, **32**: 190-211.

[24] Han J, Moradi S, Faramarzi I, *et al*. A local contrast method for infrared small-target detection utilizing a tri-layer window [J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, **17**(10):1822-1826.