

一种多深度特征连接的红外弱小目标检测方法

王维佳^{1,2} 熊文卓¹ 朱圣杰^{1,2} 宋策¹ 孙翥¹ 宋玉龙¹

1 中国科学院长春光学精密机械与物理研究所 长春 130033

2 中国科学院大学大珩学院 北京 100049

(549272937@qq.com)

摘要 针对红外弱小目标像元数量少、图像背景复杂、检测精度低且耗时较长的问题,文中提出了一种多深度特征连接的红外弱小目标检测模型(MFCNet)。首先,提出了多深度交叉连接主干形式以增加不同层间的特征传递,增强特征提取能力;其次,设计了注意力引导的金字塔结构对深层特征进行目标增强,分离背景与目标;提出非对称融合解码结构加强解码中纹理信息与位置信息保留;最后,引入点回归损失得到中心坐标。所提网络模型在SIRST公开数据集与自建长波红外弱小目标数据集上进行训练并测试,实验结果表明,与现有数据驱动和模型驱动算法相比,所提算法在复杂场景下具有更高的检测精度及更快的速度,模型的平均精度相比次优模型提升了5.41%,检测速度达到100.8FPS。

关键词: 红外弱小目标;深度学习;目标检测;特征连接;注意力机制

中图分类号 TP391

Method of Infrared Small Target Detection Based on Multi-depth Feature Connection

WANG Weijia^{1,2}, XIONG Wenzhuo¹, ZHU Shengjie^{1,2}, SONG Ce¹, SUN He¹ and SONG Yulong¹

1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

2 Daheng College, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Small infrared targets have the characteristics of a small number of pixels and a complex background, which leads to the problems of low detection accuracy and high time-consumption. This paper proposes a multi-depth feature connection network. Firstly, the model proposes a multi-depth cross-connect backbone to increase feature transfer between different layers and enhance feature extraction capabilities. Secondly, an attention-guided pyramid structure is designed to enhance the deep features and separate the background from the target. Thirdly, an asymmetric fusion decoding structure is proposed to enhance the preservation of texture information and position information in decoding. Finally, the model introduces point regression loss to get the center coordinates. The proposed network model is trained and tested on the SIRST dataset and the self-built infrared small target dataset. Experimental results show that compared with existing data-driven and model-driven algorithms, the proposed model has higher detection accuracy and faster speed in complex scenes. Compared with the suboptimal model, the average precision of the model is improved by 5.41%, and the detection speed reaches 100.8FPS.

Keywords Infrared small target detection, Deep learning, Object detection, Feature connection, Attention mechanism

1 引言

红外搜索与跟踪系统(Infrared Search and Track,IRST)是现代战争中重要的系统之一,相比雷达系统,IRST系统以被动接收方式进行检测,具有较强的隐蔽性^[1]。同时,与可见光相比,红外光穿透力强,受大气影响较小,具有全天时、抗干扰等优点。因此,红外检测技术被广泛用于航空预警、导弹跟踪、遥感等多个领域^[2]。

在远距离探测时,红外目标存在像元数量少、目标强度弱、背景结构复杂等特点,使检测变得更加困难,红外弱小

目标检测任务也引起许多学者的关注^[3]。经典的单帧红外弱小目标检测方法大多采用模型驱动的方式,主要包括以Tophat^[4]为代表的通过设计滤波器平滑背景得到感兴趣目标的方法;以IPI^[5]为代表的利用低秩矩阵张量分解的目标提取方法;以LCM^[6]为代表的利用局部对比度特征的目标分离方法。这些方法虽然计算量小且无需学习,但它们依赖于先验模型,对场景变化的辨识能力不足,且具有超参数敏感性,在不同数据集集中测试的结果通常不稳定,在复杂场景下鲁棒性佳。

随着深度学习的发展,人们开始利用数据集训练卷积神经网络学习目标特征,以FasterRCNN^[7],YOLO^[8]系列为

到稿日期:2023-02-06 返修日期:2023-03-29

基金项目:国家自然科学基金(62205332)

This work was supported by the National Natural Science Foundation of China(62205332).

通信作者:熊文卓(xwenzi@tom.com)

代表的算法通过数据驱动的方法逐渐在常规目标检测任务中成为主流。然而,现有算法的设计往往更关注大/中尺度目标的检测性能,网络深度增加反而会使目标特征消失在背景之中,加之红外弱小目标自身特性所带来的难度,现有算法在红外小目标检测上普遍表现不佳,仍存在着误检率高、准确性低的问题。

对于小目标检测任务的算法设计也逐渐走进人们视野^[9-11],S3FD^[12]模型通过设计更小的锚框来匹配有效的感受野,但随着计算量增加,模型结构无法满足实时性;SSRD-Net^[13]模型剔除冗余结构,通过网络剪枝的方法提升了检测效率,但输出特征图上分辨率过小,弱小目标特征占比少,存在漏检多的问题。在特征提取方面,主干通常采用 VGG^[14], ResNet^[15],DenseNet^[16]等网络,这些结构经过一系列的下采样提取了原始图像的全局语义信息,但其对于局部的细节以及全局对比信息的关注度不够。为此,HRNet^[17]保持了深层和浅层的特征图并行,并在不同的层之间反复交换特征,但深层与浅层特征图在语义性与空间性上没有做到很好的均衡,运算效率低;U-Net^[18]采用跳跃连接交换深层和浅层信息,提高了特征利用率,使浅层信息得以保留,但简单融合的方式使得特征整合粗糙,也会带来信息损失;AGPCNet^[19]在处理深层特征时采用注意力机制进行特征增强,但模型中大量超参数需要通过实验调试,不同数据集下泛化能力弱。此外,现有红外弱小目标数据集存在图像数量少、目标场景简单的问题,也限制了卷积模型的发展。

综上所述,为了实现对红外弱小目标的准确检测,本文提出了一种基于神经网络的数据驱动网络模型,针对

红外小目标特点,在有限的卷积层内提升特征的利用率,增强深层网络语义理解能力的同时避免目标覆盖。本文的主要贡献包括:

- 1)提出了多深度交叉连接结构,特征提取时通过不同深度间的特征传递,增强了语义信息的提取能力;
- 2)提出了非对称融合解码方法,特征融合时在保留目标纹理信息的同时增强模型对空间位置信息的保留;
- 3)提出了注意力引导金字塔结构,在多感受野下采用注意力机制增强目标的特征表征,分离背景噪声和目标。

2 基本原理

本章将详细阐述所提模型的结构,并通过对比分析各个结构的原理及其优势。

如前所述,与 VOC,COCO 等经典目标检测任务不同,红外弱小目标的像元大小仅为 3×3 像素左右,且目标缺乏纹理信息,仅依靠远距离采集图像无法识别目标类别,难以利用深层网络提取目标的语义信息。为提升红外弱小目标的检测能力,本文提出了多深度特征连接网络(Multi-depth Feature Connection Network,MFCNet),模型的整体结构如图 1 所示。MFCNet 模型主要由 3 部分构成:特征主干(Backbone)、融合解码(Decoder)、检测头(Head)。首先,将给定的红外图像经过预处理后输入到交叉连接的特征主干中提取目标信息;然后将多深度的特征进一步融合解码;最后利用解耦的检测头输出目标响应热图(Heatmap)、检测框尺度(Size)、中心点偏置(Offset),将其映射到原始图像之中得到最终的检测结果。

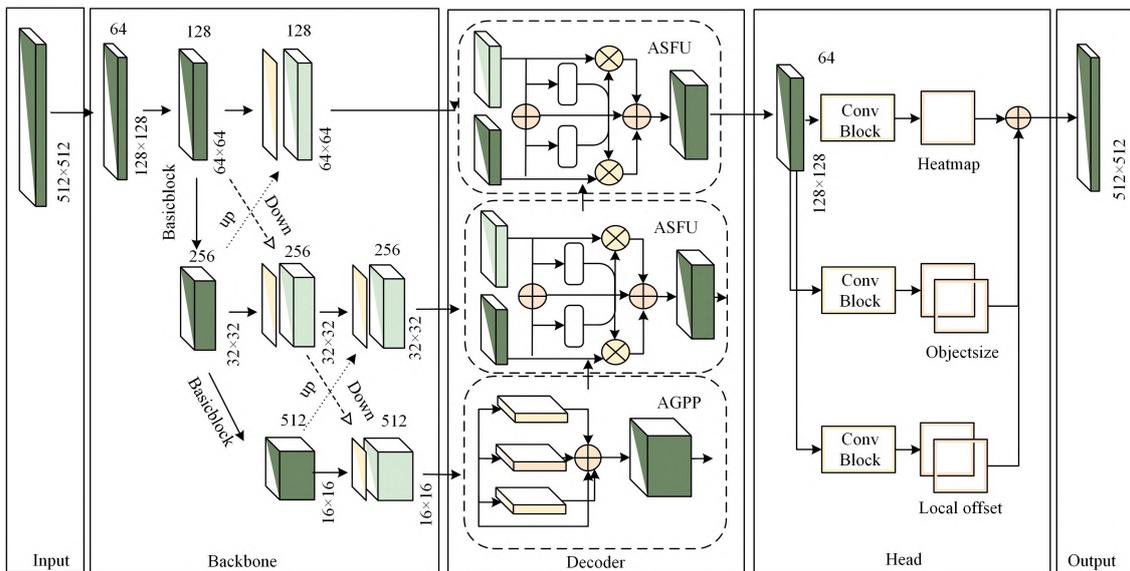


图 1 MFCNet 模型整体结构图

Fig. 1 Overall structure of MFCNet model

2.1 CCS 主干网络

在特征提取部分,一方面,由于红外弱小目标通常只占据数个像素,其背景特征会随着网络深度的增加覆盖目标信息,因此需要控制网络深度,使目标特征能够在深层特征图中得以保留。另一方面,由于红外图像成像过程中存在信噪比低、

干扰严重的问题,要在缺少目标纹理形状信息等条件下进行检测,需要增加卷积层,加强语义理解能力。为了平衡上述问题,本文提出了一种多深度交叉连接结构(Cross Connect Structure,CCS),在网络中引入相邻层的特征,在有限的深度内提升语义信息交换能力。

如图 2 所示,本文将提出的 CCS 网络与主流的主干网络 Hourglass^[20], U-Net, HRNet 进行了对比。Hourglass 的特点是阶段架构,它使用组成的初始模块,重复自下而上、自上而下的处理中间跳过层的特征连接,为网络提供多尺度接受阈,但中间处理部分采用初始模块,具有一定的局限性。U-Net 通过 U 型结构,采用多个下采样卷积块

和上采样反卷积块来恢复原始分辨率,但连接方式过于简单,横向传递时仅利用了同层信息。HRNet 通过浅层特征与深层特征并行卷积重复进行信息交换,随着网络的加深,逐步添加深层特征,从而具备多深度特征表示,但其结构着重于高分辨率下的特征图,模型结构复杂,使得计算量增加。

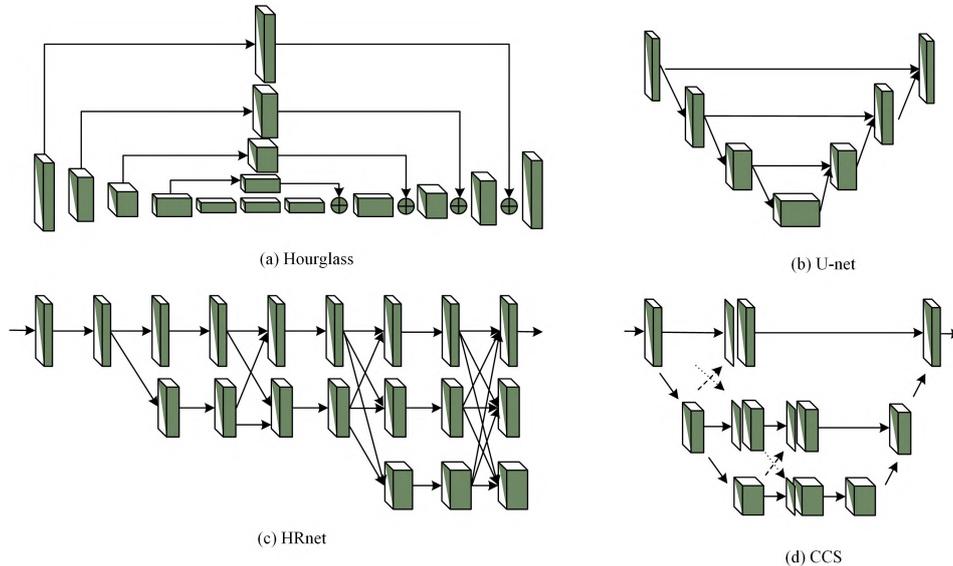


图 2 主干网络结构对比

Fig. 2 Comparison of backbone network structure

图 2(d)展示了本文提出的 CCS 结构,与图 2(b)类似,主体仍采用 U 型网络结构,但优化了 U-Net 中单一跳跃连接的问题。相比 Hourglass 拓展串行结构,CCS 结构增加了多深度特征传递和交叉深度特征传递,继承前一阶段一个特征流以增强不同深度特征间的融合。同时,相较于 HRNet,其优化了不同深度特征分支的计算分配,侧重于语义能力较强的深层特征分支。对比上述主干网络,本文提出的结构在提取浅层特征方面结构简单,同时增强了多深度特征传递,能够有效防止目标特征丢失,更好地提取有效信息。网络的具体结构如表 1 所列。

表 1 中 1-4 为特征提取部分。其中,CBR 由一个 3×3 卷积层、一个归一化层和一个激活函数层组成;BasicBlock^[15]则包含两个 3×3 卷积层、一个归一化层和一个激活函数层。为保证特征的多样性,网络通过 3 次下采样将特征维数扩展到 512 维,通过 $8 \times, 16 \times, 32 \times$ 这 3 个深度特征分支获得了 $64 \times 64, 32 \times 32, 16 \times 16$ 大小的特征图,并作为第一阶段的输出。不同尺度的特征分支分别包含了不同分辨率下的抽象特征。小尺度特征抽象能力强,但会遗失一些细节特征,通过与大尺度的特征分支结合可以进行信息补充。第二阶段为交叉深度特征传输部分,将前一阶段提取的特征图分别向相邻深度特征图传递,下采样采用步幅为 2 的 3×3 卷积核进行卷积,生成深层特征分支。上采样采用双线性插值方式,通过融合不同多尺度特征,减少下采样造成的信息损失。最终网络输出 3 个深度特征层。

表 1 CCS 主干网络结构

Table 1 CCS backbone network architecture

Num	From	Layer	Operator	Output shape	Resolution Branch
0	-	input	-	(1,512,512)	$1 \times$
1	-1	stem	CBR	(64,256,256)	$2 \times$
			Maxpool	(64,128,128)	$4 \times$
2	-1	stage1	BasicBlock $\times 2$	(128,64,64)	$8 \times$
			BasicBlock $\times 2$	(256,32,32)	$16 \times$
3	-1	stage2	BasicBlock $\times 2$	(512,16,16)	$32 \times$
4	-1	stage3	BasicBlock $\times 2$	(256,64,64)	$8 \times$
5	3	Fusestage1	Upsample	(128,64,64)	$8 \times$
6	(-1,2)		Concat	(128,64,64)	$8 \times$
7	2	Fusestage2	Downsample	(256,32,32)	$16 \times$
8	(-1,3)		Concat	(256,32,32)	$16 \times$
9	4	Fusestage2'	Upsample	(512,32,32)	$16 \times$
10	(-1,7)		Concat	(256,32,32)	$16 \times$
11	7	Fusestage3	Downsample	(512,16,16)	$32 \times$
12	(-1,4)		Concat	(512,16,16)	$32 \times$

2.2 注意力引导金字塔模块

考虑到红外弱小目标在深层特征图中不能忽略物体的纹理信息,为了增强物体与背景之间的可区分性,本节设计了一个注意力引导金字塔池化(Attention-Guided Pyramid Pooling, AGPP)模块,如图 3 所示。

对于输入特征,首先通过平均池化和多次最大池化操作获得多感受野下特征,平均池化后的大感受野特征对应背景信息,最大池化后的小感受野特征对应尺度下目标纹理信息。由于单独卷积核表达能力有限,因此本文对池化后的特征采用非局部注意力 Non-local 模块^[21]进行提炼。

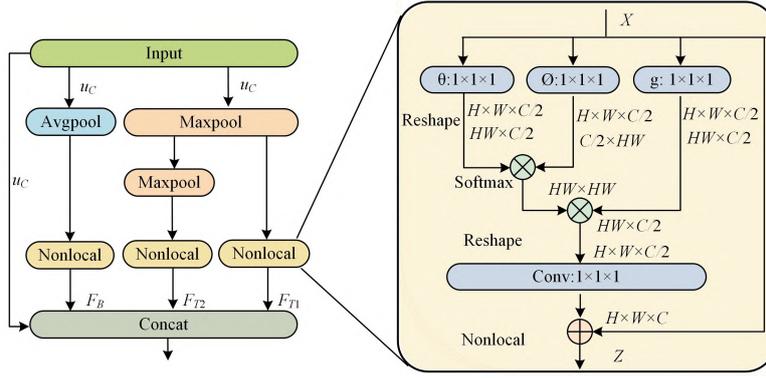


图3 AGPP 结构图

Fig. 3 AGPP structure diagram

Non-local 模块执行全局计算,其计算每个像素与其他像素之间的相关性,打破了卷积层感知范围的局限性,可以捕获像素之间的远距离依赖关系,计算式如下:

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (1)$$

其中, x 为输入特征, i 和 j 代表输出位置, $f(x_i, x_j)$ 函数式用于计算 i 和 j 位置上特征的相似度, $g(x_j)$ 函数计算特征在 j 位置的表示。最终通过响应因子 $C(x)$ 进行标准化处理以后得到输出特征 y 。

根据池化方式的不同, AGPP 具体计算过程可分为语义的整体关联和细节关联,整体关联侧重增强背景特征。计算时首先采用平均池化 (P_{avg}) 遍历特征层,在全局范围内整合背景信息,如式(2)所示:

$$Z_C = P_{avg}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

其中, Z_C , $C=1, 2, \dots, C$ 代表第 C 通道, H 和 W 指输入特征的大小, $u_c(i, j)$ 是第 C 个通道中每个位置的特征。随后将全局池化后特征送入 Non-local 模块分析特征图全局结构信息。

$$F_B = \delta(f_{non}(Z_C)) = \delta(f_{non}(P_{avg}(u_c))) \quad (3)$$

其中, $\delta(\cdot)$ 为校正线性单元(ReLU), f_{non} 为 Non-local 模块。

另一方面,为增强高级语义特征下目标纹理信息,采用了细节关联方式,在相同深度特征图下,采用金字塔池化结构^[22]避免了卷积神经网络重复提取相关特征的问题,输入特征经过两个最大池化层后得到不同尺度的特征图。随后使用 Non-local 模块提取不同感受野下的特征,以便获得更准确的纹理信息。

$$F_{T1} = \delta(f_{non}(P_{max}^{3 \times 3}(u_c))) \quad (4)$$

$$F_{T2} = \delta(f_{non}(P_{max}^{3 \times 3}(P_{max}^{3 \times 3}(u_c)))) \quad (5)$$

其中, P_{max} 表示最大池化层, 3×3 为卷积核尺寸。

最后将加强后的背景与纹理特征以及原始特征结合,获得具有深层语义特征的特征加权和,如式(6)所示:

$$F_{OUT} = \delta(\beta(C_{1 \times 1}(f_{cat}(F_B, F_{T1}, F_{T2}, u_c)))) \quad (6)$$

其中, f_{cat} 为 Concat 函数, $C_{1 \times 1}$ 为卷积核大小为 1×1 的卷积, $\beta(\cdot)$ 为批归一化(BN)。

2.3 非对称上下文融合

为了使网络具有从底层动态选择相关特征的能力,保留

编码阶段中更多的有效特征,本节设计了一个非对称交叉融合模块(Asymmetric Fusion, ASFU)来整合上下文信息,将低级语义(X_L)和高级语义(X_H)作为输入,分别对其包含的不同信息进行单独处理,最后将两部分融合在一起进行信息的相互补充。图4详细描述了其体系结构。

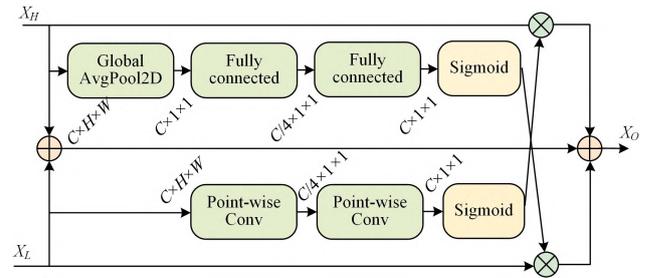


图4 ASFU 结构图

Fig. 4 ASFU structure diagram

深层特征层包含编码的高级语义信息。在向上传递中,高级语义信息不足以恢复详细的空间信息,因此需要从浅层特征映射,对低层次细节和空间信息进行整合,细化粗糙的高级语义特征,实现准确的空间定位。同时,浅层特征虽然对全局的感受范围大,但对特征的挖掘不够深入,在特征融合时,深层特征可以为它提供更准确的语义信息。据此,在对浅层特征图和深层特征图进行融合时,分别针对低级和高级语义信息设计了从上至下和从下至上两种不同策略。

低级语义 X_L 包含了大量的目标位置信息,本文使用逐点卷积(Point-wise Convolution)^[23]将每个空间位置的通道特征上下文单独进行加权组合,以自下而上的方式传递空间信息,生成的权值可以表示为:

$$L(X_L) = \sigma(\beta(PWConv_2(\delta(\beta(PWConv_1(X_L)))))) \quad (7)$$

其中, $PWConv$ 表示逐点卷积,大小分别为 $C \times 1 \times 1$ 和 $C/4 \times 1 \times 1$; δ, β, σ 分别表示 ReLU 激活函数^[24]、批处理归一化(BN)和 Sigmoid 激活函数。

另一方面,对于高级语义 X_H ,首先利用全局池化遍历特征,然后通过两个卷积改变网络通道数,最后使用激活函数加入非线性因素,增强高级语义特征,最小化网络参数。 W_1 和 W_2 卷积核的大小为 1×1 , P_{avg} 为全局平均池化。

$$H(X_H) = \sigma(\beta(W_2 \delta(\beta(W_1 (P_{avg}(X_H)))))) \quad (8)$$

最后将两部分结果与原始输入叠加,在输入特征上增强细节信息与语义信息,最终得到输出。

$$Q(X) = (L(X_L) \otimes X_H) \oplus (H(X_H) \otimes X_L) \oplus (X_H \oplus X_L) \quad (9)$$

2.4 损失函数

红外弱小目标检测任务要求模型满足实时性的要求,所以本文在检测头部分参考无锚框的 Centernet^[25]方法。相比预设锚框方法中锚框数量多而产生冗余问题,无锚框检测方法摆脱了预设锚框带来的计算量,进一步提高了算法速度。同时,针对检测任务中目标小的特点,用目标对象的中心点来表示目标;以热图峰值作为目标中心,再结合坐标偏移以及尺度预测生成最终的检测框。

训练时,首先利用标注文件,采用高斯函数的方式生成标准中心点热图,然后结合偏移值与长宽尺度设计损失函数回归模型参数。检测框的损失函数部分由热图损失、中心点偏移损失和长宽损失组成。热图损失函数的计算公式如式(10)所示:

$$L_r = \frac{-1}{N} \sum_{x_{yc}} \begin{cases} (1 - \hat{Y}_{x_{yc}})^\alpha \log(\hat{Y}_{x_{yc}}), & \text{if } Y_{x_{yc}} = 1 \\ (1 - Y_{x_{yc}})^\beta (\hat{Y}_{x_{yc}})^\alpha \log(1 - \hat{Y}_{x_{yc}}), & \text{otherwise} \end{cases} \quad (10)$$

采用 Focal loss 方法并对其进行改写,增加 $(1 - Y_{x_{yc}})^{\beta c}$ 是为了抑制热图中高斯中心点附近负样本点的损失 ($0 < Y_{x_{yc}} < 1$)。式(10)中 N 为图像中关键点数量,用于标准化; x_{yc} 表示 heatmap 上的坐标点; $\hat{Y}_{x_{yc}}$ 为预测值; $Y_{x_{yc}}$ 为标注真实值; α 和 β 为超参数,用来均衡正负样本及难易样本,其中 $\alpha = 2, \beta = 4$ 。

偏移和长宽损失采用 L2 loss 方法计算。利用偏移损失对特征图中每个像素点的位置偏差进行修正,计算公式如式(11)所示:

$$L_{off} = \frac{1}{N} \sum_p (\hat{O}_p - (\frac{p}{R} - \tilde{p}))^2 \quad (11)$$

其中, \hat{O}_p 表示预测的偏移值, p 为目标中心点坐标, R 为缩放尺度, \tilde{p} 为缩放后中心点的取整坐标。

长宽损失用于计算目标框尺寸的回归损失,计算公式如式(12)所示:

$$L_{size} = \frac{1}{N} \sum_{k=1}^N (\hat{S}_{pk} - S_k)^2 \quad (12)$$

其中, S_k 为真实尺寸, \hat{S}_{pk} 为预测尺寸。

综上,得到总损失函数 L_{det} 如式(13)所示:

$$L_{det} = L_k + \lambda_{size} L_{size} + \lambda_{off} L_{off} \quad (13)$$

其中, λ_{size} 和 λ_{off} 为修正偏移损失和长宽损失的平衡系数,实验证明 $\lambda_{size} = 0.1, \lambda_{off} = 1$ 时能取得最优效果。

在验证与测试时,模型计算输入图像得到一个中心点热图 $P \in [0, 1]^{\frac{W}{R} \times \frac{H}{R} \times C}$ 。通过 3×3 最大池化操作,提取热图中大于或等于其 8 个相邻的所有峰值,只保留前 100 个峰值。对于 (\hat{x}_i, \hat{y}_i) 位置上的峰值,使用偏移预测 $\hat{O}_{x_i, y_i} = (\Delta \hat{x}_i, \Delta \hat{y}_i)$ 和大小预测 $\hat{S}_{x_i, y_i} = (\hat{w}_i, \hat{h}_i)$ 来生成对象的边界框,最后

映射到原始图像中得到最终检测结果。边界框的生成方式如下:

$$\begin{aligned} & (\hat{x}_i + \Delta \hat{x}_i - \hat{w}_i / 2, \hat{y}_i + \Delta \hat{y}_i - \hat{h}_i / 2, \hat{x}_i + \Delta \hat{x}_i + \hat{w}_i / 2, \\ & \hat{y}_i + \Delta \hat{y}_i + \hat{h}_i / 2) \end{aligned} \quad (14)$$

3 实验与分析

本章首先介绍本文提出的数据集和公共数据集,其次介绍实验设置及评估指标,用于比较现有检测方法。然后将本文算法分别在两个数据集上与现有方法进行定量比较。定性结果直观地说明了本文方法的优越性。最后,通过一系列的消融实验,验证了本文方法各模块、操作方式对检测结果的影响。

3.1 数据集

由于红外弱小目标图像采集困难,应用范围较为特殊。现有开源图像少,而数据少会导致网络训练不稳定、模型容易过拟合或无法收敛等问题,因此,部分算法通过点扩散函数(Point Spread Function, PSF)^[26]随机产生模拟目标获得合成图像,而生成目标的内部纹理与真实目标存在一定差距,算法效果验证不准确。

本文选择 SIRST 数据集以及自建数据集 LWIRST 对提出的 MFCNet 模型进行验证。SIRST 数据集共 427 张图像,通过不同成像系统拍摄后,从数百个不同场景的红外序列中选取而来,包含短波、中波、长波等多波段图像。然而,由于 SIRST 数据集图像数量少,成像系统缺乏一致性,为了更好地验证实际工程中模型的效果,本文制作了长波红外弱小目标数据集(Long Wave Infrared Small Target, LWIRST)数据集。该数据集由实际光电系统采集 255×320 像素图像组成,采用长波成像。数据集中每个目标都是通过观察它的移动序列来确认的,以确保它是一个真实目标,而不是像素级的脉冲噪声。目标背景以天空为主,其中包括云背景、无云背景、建筑天空背景和地空背景。LWIRST 数据集共 3000 张图像,采用 Labelimg 标注软件对目标进行标注。

3.2 实验细节

模型在 Intel Core i5-11400F CPU、16GB 内存和 GeForce GTX 3070Ti GPU(8GB 内存)的计算机上进行测试和评估,使用开源的 Pytorch 框架实现。训练过程中,通过边缘填充的方式将图像大小调整为 512×512 像素。训练迭代次数设置为 3000,利用 Adam 优化器更新参数,采用多项式衰减学习率调度策略,初始学习率为 0.0005,动量衰减和权重衰减分别设置为 0.9 和 0.005,预热迭代次数为 100 轮。

本实验采用检测精度(Precision)、召回率(Recall)、综合评价指标(Fmeasure)和平均精度(Average Precision, AP)等经典目标检测评价指标来全面评估模型的性能。FLOPs 用于评估计算量,FPS 用于评估模型检测速度。其中,精度 Precision 为所有检测到的目标为正样本的概率,召回率 Recall 为正样本被正确检测的概率,计算式可表示如下:

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{cases} \quad (15)$$

其中, TP, FP, TN 和 FN 分别为真阳性、假阳性、真阴性、假阴性的数量占比。为了综合衡量模型的性能, 平衡精度和召回率关系, 采用 $Fmeasure$ 评估指标对它们进行加权调和平均, 计算方式被定义为:

$$Fmeasure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (16)$$

AP 是在 $Precision-Recall$ 曲线基础上计算与坐标轴的包围面积的结果, 计算式如式(17)所示:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_{i+1}) \quad (17)$$

其中, r_1, r_2, \dots, r_i 是按升序排列的 $Precision$ 插值段中每个插值处对应的 $Recall$ 值, P_{inter} 为对应点 $Precision$ 值。

3.3 模型性能对比

为验证上述模型的效果, 本文将提出的 MFCNet 模型与现有常见模型进行对比, 包括以 YOLO 为代表的卷积神经网络模型以及以 Tophat 为代表的经典算法模型。分别从数据驱动和模型驱动两方面进行对比, 验证本文算法性能。

3.3.1 数据驱动算法对比

对于数据驱动的方法, 本文选择深度学习目标检测中的多阶段检测算法(FasterRCNN)、单阶段检测算法中基于锚框的检测算法(YOLOv3)、无锚检测算法(CenterNet)、针对小目标检测算法(SSRD-Net)等进行对比。实验对比结果如表 2 所列。

表 2 数据驱动模型的对比分析实验结果

Table 2 Comparative analysis of experimental results of data-driven models

Models	Backbone	LWIRST				SIRST				GFLOPs	FPS
		Precision/%	Recall/%	Fmeasure	AP/%	Precision/%	Recall/%	Fmeasure	AP/%		
FasterRCNN	ResNet50+FPN	57.14	9.41	0.16	9.81	20.00	12.76	0.16	6.53	370.2	7.2
SSD ^[27]	VGG16	42.34	27.06	0.33	42.34	54.55	20.34	0.30	34.53	62.8	45.5
YOLOv3	Darknet53	65.96	50.82	0.57	47.06	55.88	44.19	0.49	41.40	193.9	51.3
YOLOv5m	CSPDarkNet53	73.53	62.50	0.68	60.23	69.78	46.13	0.56	45.51	49.0	67.1
CenterNet	ResNet18	57.00	74.92	0.65	53.09	56.15	59.48	0.58	47.05	35.7	151.7
SSRD-Net	CSPDarkNet53	73.49	70.76	0.72	58.02	63.61	57.61	0.60	49.74	45.6	90.2
CenterFace ^[28]	MobileNetV2	77.26	78.29	0.78	61.95	65.66	60.00	0.62	50.23	39.6	139.2
MFCNet	ResNet18	80.95	85.00	0.83	67.36	68.90	64.97	0.66	54.46	42.1	100.8

从表 2 可以看出, 在 SIRST 数据集上, 本文算法在 $Recall, Fmeasure, AP$ 指标上取得了最好的效果, 在 $Precision$ 指标上, 与最优算法 YOLOv5m 仅相差 0.88%。在 LWIRST 数据集上, 本文算法均取得了最好的效果。在 SIRST 数据集上, 本文算法 MFCNet 在 AP 指标上比次优算法(CenterFace)高出 4.23%, 在 LWIRST 数据集上也有 5.41% 的提升。

与双阶段检测算法 FasterRCNN 相比, 由于 FasterRCNN 中锚框生成与检测分离, 运算量大, 检测速度极慢, 因此不适用于红外目标检测任务。

与基于锚框检测模型 SSD, YOLOv3, YOLOv5m, SSRD-Net 相比, 本文算法的检测速度与检测精度均有提高。一方面, 在主干结构设计上, SSD 模型只是在不同深度特征尺度上预测目标, 没有融合不同深度特征图, 召回率和精度均不理想; YOLOv3 和 YOLOv5m 使用 FPN 网络和 PANet 网络进行特征融合, 但在融合阶段没有针对目标和背景的具体特征进行辨别, 造成特征遗漏和冗余, 召回率低; SSRD-Net 对固定感受野进行了优化, 剔除了其他分支结构, 使得检测速度加快, 但红外弱小目标的像素数量与模型感受野不匹配, 导致召回率较低。另一方面, 在同一位置生成一定数量锚框会增大计算量, 且模型最终精度依赖于预设参数的选取, 具有不确定性。与同属基于无锚框检测模型的 CenterNet 和 CenterFace 相比, MFCNet 检测精度提升, FPS 降低。虽然它们在检测速度上优于本文模型, 但在特征处理时并未区分深层特征中的目标与噪声, 进行目标增强, 识别精度低于本文模型。

综上, 本文提出的 MFCNet 模型参数量较少, 在保证模型精度的条件下实现了网络结构轻量化, 更适用于红外弱小目标检测任务。

为了直观评价各模型间的检测差异, 本文从两个数据集中各选择了 3 个典型的红外弱小目标场景, 比较了 8 种算法的检测结果, 结果如图 5 所示, 方框为检测的目标结果。

从整体效果来看, 在图 5(a) 等目标亮度大、背景干扰少的检测状态下, 算法均能准确检测出目标, 但本文算法位置更为准确。针对图 5(d) 目标出现在云层中的情况, YOLOv5m 等网络模型会倾向于提高在较大目标上的性能, 以最小化损失函数, 并减少对较小目标的关注, 但这样反而会导致目标召回率下降。而对于图 5(e) 中目标出现在云边缘等情况, CenterNet 等网络结构由于没有对目标特征进行针对性增强, 会对碎云造成误检。本文算法保留了浅层的网络结构特点, 增加了多深度交叉连接加深学习, 在复杂红外场景中针对困难位置下红外弱小目标也有着很好的效果, 目标检测更加准确。

由原始图片中目标放大图可以看出, SIRST 数据集目标与背景的对比度较大、信噪比高, 在检测中易于识别, 算法评估时区分性较小; 同时由于数据集数量少, 且不同图片中目标形态相差较大, 利用 SIRST 数据集进行检测时会造成检测精度较低的情况。LWIRST 数据集背景包含云边、云上、无云等多种情况, 信噪比低, 识别难度更大; 同时, 数据集红外图像数量多, 利于充分训练, 对于算法的评估更加准确。

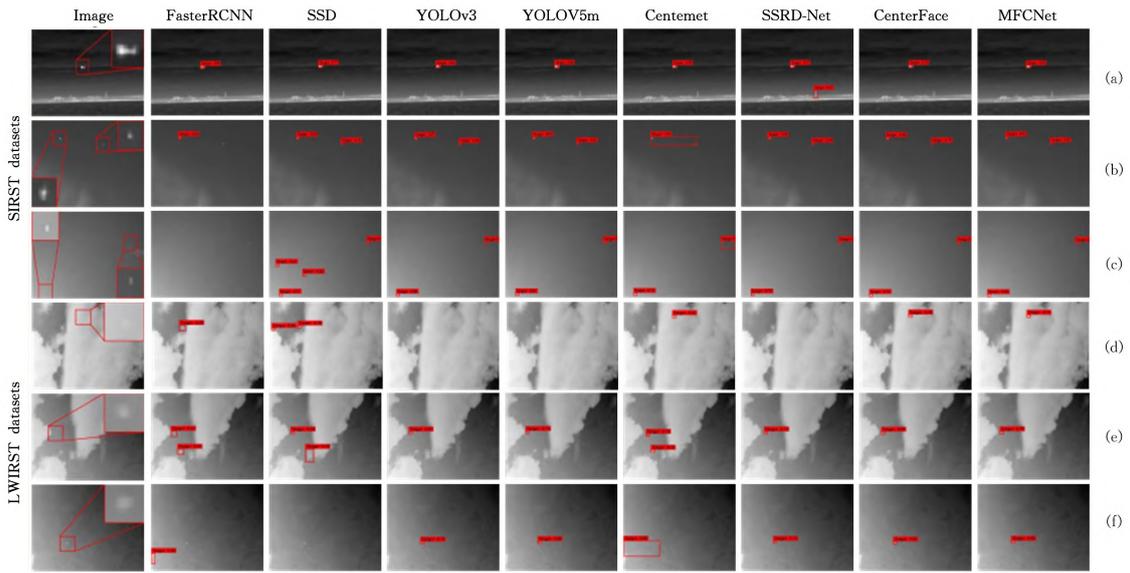


图 5 各算法的检测效果图

Fig. 5 Detection effect of each algorithm

3.3.2 模型驱动算法对比

对于模型驱动算法,本文选取 3 类经典红外弱小目标算法作为对比算法在 LWIRST 数据集上进行测试,具体实验结果如表 3 所列。

表 3 模型驱动算法对比实验结果

Table 3 Comparative experimental results of model-driven algorithms

Models	Precision/%	Recall/%	Fmeasure
Tophat	5.26	13.65	0.07
HBMLCM ^[29]	26.49	34.29	0.30
ADMA ^[30]	16.75	26.67	0.20
LIG ^[31]	14.35	40.37	0.21
IPI	23.57	34.28	0.28
MPCM ^[32]	21.14	37.19	0.27
MFCNet	80.95	85.00	0.83

采用的对比算法包括基于背景抑制算法、基于局部对比度算法和基于稀疏分解算法等 6 种模型。由表 3 可知,相比

背景抑制算法(Tophat),MFCNet 模型可以更好地处理背景干扰;相比局部对比度方法(HBMLCM),本文算法可以增强目标,从含有误检的复杂背景中识别出目标;相比稀疏分解方法(IPI),本文模型可以从少见的结构中压制误检。

为了更直观地展示检测结果,图 6 给出了不同模型驱动算法下目标结果的显著性图。

在图 6(a)和图 6(b)所示目标亮度较大的情况下,天空背景经典模型驱动算法基本可以检测出目标,但结果也受到了少量杂波影响;在图 6(d)和图 6(e)等存在云层干扰的情况下,模型驱动算法在少数碎云处响应较高,无法避免虚警出现;图 6(c)和(f)中,背景出现建筑物时,大多数经典算法检测峰值出现在楼房角点处,导致检测结果与真实标签差异较大,算法易受背景影响,鲁棒性不强。综合来看,本文算法能够避免强边缘信息,过滤尖锐噪声,更适用于红外弱小目标检测。

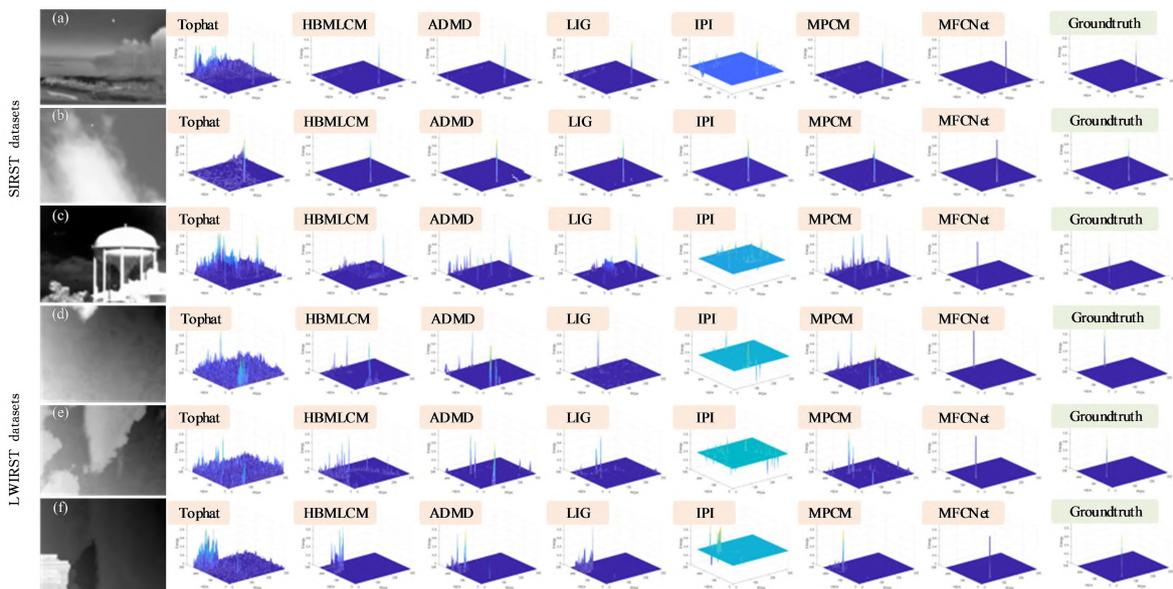


图 6 模型驱动算法显著性图对比结果

Fig. 6 Comparison results of model-driven algorithm significance diagram

3.4 消融实验

为了验证 MFCNet 模型中各结构的有效性,本文通过在 ResNet18 上逐个添加 CCS, ASFU 和 AGPP 的方式进行消融

实验,从检测性能和效率两方面对网络进行评估。共进行 4 组对比实验(num1-num4),实验结果如表 4 所列,4 组实验中增加的结构分别在表格中进行了标注。

表 4 MFCNet 中不同模块对检测性能的影响

Table 4 Effect of different modules in MFCNet on detection performance

Num	CCS	ASFU	AGPP	Precision/%	Recall/%	F1	AP/%	GFLOPs	FPS
1	—	—	—	57.00	74.92	0.65	53.09	35.7	151.7
2	✓	—	—	63.25	77.24	0.70	58.53	38.4	145.1
3	✓	✓	—	71.85	82.57	0.77	62.43	40.6	133.5
4	✓	✓	✓	80.95	85.00	0.83	67.36	42.1	100.8

从表 4 中可以看出引入不同改进部分对实验结果的影响。对比 num1 和 num2 可以看出,在特征提取的改进部分,增加的交叉连接提高了检测的精度和召回率,Fmeasure(F1)指标提升了 5%。这表明,交叉连接的网络主干能够提高网络的性能,以少量的参数换来精度的提升且对网络的检测速度影响不大。对比 num2 和 num3,在融合方式上,本文采用的 ASFU 方式相比上采样简单连接,将不同深度下的信息相互补充,增强了目标位置的回归能力,召回率提升了 5.33%。

对比 num3 和 num4,本文模型引入深层特征处理 AGPP 部分,将目标与背景噪声区分开,模型精度提升了 9.1%,证实了注意力机制在整体和局部感知的能力。对于特征融合和特征增强操作,为了展示本文结构的优越性,本节将 ASFU 与 AGPP 结构与 AGPCNet 模型中的 AFM 和 CPM 模块相比较。在整体架构不变的情况下,修改特征融合和特征增强两部分,在 LWIRST 数据集上的实验结果如表 5 所列。4 组对比实验(num1-num4)中采用的结构分别在表格中进行了标注。

表 5 不同模块对检测性能的影响

Table 5 Effect of different modules on detection performance

Num	AFM	ASFU	CPM	AGPP	Precision/%	Recall/%	Fmeasure	AP/%
1	✓	—	✓	—	78.73	81.25	0.80	64.88
2	—	✓	✓	—	79.43	83.68	0.81	65.82
3	✓	—	—	✓	79.89	83.16	0.81	66.19
4	—	✓	—	✓	80.95	85.00	0.83	67.36

num1 和 num2 相比,在特征融合方式上,AFM 将处理后的结果直接叠加,而本文提出的 ASFU 将深层的位置信息和浅层的纹理信息分别加权到原始特征上,可以得到更高的召回率。相较于 CPM,本文提出的 AGPP 采用了快速空间金字塔池化结构,在全局范围内进行注意力关联,平均池化的引入也增加了对背景特征的提取。通过 num2 和 num4 的对比可知本文算法提高了检测精度,证明了设计的合理性。

综上所述,MFCNet 模型中各部分设计是有效的,在提升网络精度与召回率的同时,检测效率仍保持在较高水平。

结束语 本文从提高空中红外小目标检测的精度及速度角度出发,提出了一种多深度特征连接的红外弱小目标检测模型 MFCNet。该模型鲁棒性强,在云层等复杂背景下仍具有较好的性能。MFCNet 网络采用交叉连接进行特征提取,利用注意力机制增强特征提取能力,通过非对称融合进行特征解码,并通过热图回归得到最终目标位置。消融实验结果表明,本文方法与基准模型相比,Fmeasure 提升了 0.18,精度提升了 23.95%,召回率提升了 10.08%。相较于该任务已有的卷积神经网络模型,本文提出的 MFCNet 网络计算参数量少,实时性强,满足地面设备分析的实时性要求。由于现有机载航空预警设备的图形计算能力有限,还需对模型进行进一步压缩。未来,我们将通过进一步优化网络内部特征连接的方式来提升模型的鲁棒性,探索模型压缩等网络轻量设计,并将其部署到嵌入式设备中。

参考文献

[1] DAI X,ZHAO X,LI L P,et al. Infrared dim small target detec-

tion algorithm based on improved YOLOv5 in complex background [J]. Infrared Technology,2022,44(5):504-512.

[2] ZHANG Q,ZHU H T,CHENG H. Light weight infrared dim small target detection algorithm [J]. Progress in Laser and Optoelectronics,2022,59(16):282-288.

[3] CHEN M S,SUN W X,LI M Y,et al. Infrared small target detection under various complex backgrounds[J]. Journal of Jilin University(Engineering and Technology Edition),2020,50(6):2288-2294.

[4] GAO C Q,MENG D Y,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.

[5] GAO C,MENG D,YANG Y,et al. Infrared patch-image model for small target detection in a single image[C]//IEEE Transactions on Image Processing. 2013:4996-5009.

[6] LI X F,WANG S Q,WENG X,et al. Remote sensing of floating macroalgae blooms in the East China Sea based on UNet deep learning model[J]. Acta Optica Sinica,2021,41(2):18-26.

[7] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. MIT Press,2015:91-99.

[8] REDMON J,FARHADI A. Yolov3:An incremental improvement[J]. arXiv:1804.02767,2018.

[9] PAN X H,SHAO Q,LU J G. Small object detection algorithm based on CBD-YOLOv3. Journal of Chinese Computer Systems. 2022,43(10):2143-2149.

- [10] ZHANG H D,ZHANG R Q,TONG L. A vehicle target detection method based on improved YOLOv5s [J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37(7):80-89.
- [11] DAI D E,ZHU R F,CHEN C Z, et al. Aviation small target detection algorithm based on improved Yolov5l [J]. Computer Engineering and Design, 2023, 44(9):2610-2618.
- [12] ZHANG S,ZHU X,LEI Z, et al. S3FD:Single shot scale-invariant face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017:192-201.
- [13] ZHU S J,LIU J,TIAN Y, et al. Rapid Vehicle Detection in Aerial Images under the Complex Background of Dense Urban Areas[J]. Remote Sensing, 2022, 14(9):2088.
- [14] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.
- [15] HE K,ZHANG X,REN S, et al. Identity mappings in deep residual networks[C]//European Conference on Computer Vision. 2016:630-645.
- [16] HUANG G,LIU S,VANDER M L, et al. CondenseNet:An efficient DenseNet using learned group convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:2752-2761.
- [17] 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.
- [18] RONNEBERGER O,FISCHER P,BROX T. U-net:Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. 2015:234-241.
- [19] ZHANG T,CAO S,PU T, et al. Agpnet:Attention-guided pyramid context networks for infrared small target detection[J]. arXiv:2111.03580, 2021.
- [20] NEWELL A,YANG K,DENG J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision. 2016:483-499.
- [21] WANG X,GIRSHICK R,GUPTA A, et al. Non-local neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2018:7794-7803.
- [22] CHEN L C,PAPANDREOU G,KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4):834-848.
- [23] DAI Y,WU Y,ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//IEEE/CVF Winter Conference on Applications of Computer Vision. 2021:950-959.
- [24] CHEN C L,LI H,WEI Y T, et al. A local contrast method for small infrared target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1):574-581.
- [25] DUAN K,BAI S,XIE L, et al. Centernet:Keypoint triplets for object detection[C]//IEEE/CVF International Conference on Computer Vision. 2019:6569-6578.
- [26] LIN L K,WANG S Y,TANG Z X. Point target detection method of infrared oversampling scanned image based on deep convolution neural network [J]. Journal of Infrared and Millimeter Wave, 2018, 37(2):219-226.
- [27] LIU W,ANGUELOV D,ERHAN D, et al. SSD:Single shot multibox detector[C]//European Conference on Computer Vision. Cham:Springer, 2016:21-37.
- [28] XU Y,YAN W,YANG G, et al. CenterFace:joint face detection and alignment using face as point[J]. Scientific Programming, 2020, 2020:1-8.
- [29] SHI Y,WEI Y,YAO H, et al. High-boost-based multiscale local contrast measure for infrared small target detection[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 15(1):33-37.
- [30] 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.
- [31] ZHANG H,ZHANG L,YUAN D, et al. Infrared small target detection based on local intensity and gradient properties[J]. Infrared Physics & Technology, 2018, 89:88-96.
- [32] WEI Y, YOU X,LI H. Multiscale patch-based contrast measure for small infrared target detection [J]. Pattern Recognition, 2016, 58:216-226.



WANG Weijia, born in 1998, master. Her main research interests include computer vision and aerial image target detection.



XIONG Wenzhuo, born in 1967, master, researcher. His main research interests include aerial photoelectric imaging and photoelectric sensor technology.

(责任编辑:何杨)