

# 门控卷积和低频特征融合的红外小目标检测

刘奎<sup>1</sup>, 唐慧萍<sup>1+</sup>, 苏本跃<sup>2</sup>

1. 安庆师范大学 计算机与信息学院, 安徽 安庆 246000

2. 铜陵学院 数学与计算机学院, 安徽 铜陵 244000

+ 通信作者 E-mail: 2591377544@qq.com

**摘要:**针对在远距离复杂场景下红外小目标尺寸和形状差异大、检测精度欠佳的问题,提出了一种基于门控卷积和低频特征融合的红外小目标检测模型。为了解决复杂场景下目标尺寸差异的问题,通过在U-Net下采样过程中使用门控快速傅里叶卷积(gated fast Fourier convolution, GFFC)模块提取多尺度的全局和局部特征,为提升模型在不同复杂程度数据集上的训练效果,利用超参数门控调节网络对目标全局和局部特征的权重,以平衡对全局和局部特征的需求。为解决红外小目标形状差异的问题,采用了低频特征融合(high-frequency feature fusion, HFF)模块,进一步提取高频子带特征,增强红外小目标细节纹理信息。在SIRST和IRSTD数据集上的实验结果表明,相比基准UCF方法,提出的方法在两个数据集的评价指标下分别提升了0.83个百分点、0.40个百分点和5.18个百分点、0.23个百分点,证明了该方法的有效性。

**关键词:**门控快速傅里叶卷积;低频特征融合;红外小目标检测

**文献标志码:**A **中图分类号:**TP391.41 **doi:**10.3778/j.issn.1002-8331.2311-0326

## Gated Convolution and High-Frequency Feature Fusion for Infrared Small Target Detection

LIU Kui<sup>1</sup>, TANG Huiping<sup>1+</sup>, SU Benyue<sup>2</sup>

1. School of Computing and Information Technology, Anqing Normal University, Anqing, Anhui 246000, China

2. School of Mathematics and Computer Science, Tongling College, Tongling, Anhui 244000, China

**Abstract:** Aiming at the problem of poor detection accuracy of infrared small targets with large differences in size and shape in remote complex scenes, an infrared small target detection model based on gated convolution and high-frequency feature fusion is proposed. In order to solve the problem of differences in target sizes in complex scenes, multi-scale global and local features are extracted by using the gated fast Fourier convolution (GFFC) module during U-Net downsampling, the hyperparameter gating is utilized to adjust the weights of the target's global and local features to balance the demand for global and local features in order to improve the training effect of the model on datasets with different levels of complexity. In order to solve the problem of shape difference of infrared small targets, the high-frequency feature fusion (HFF) module is used to further extract high-frequency sub-band features to enhance the detailed texture information of infrared small targets. The experimental results on the SIRST and IRSTD datasets show that compared to the benchmark UCF method, the proposed method improves 0.83 percentage points, 0.40 percentage points, and 5.18 percentage points, 0.23 percentage points under the evaluation metrics of the two datasets, respectively, which proves the effectiveness of the method.

**Key words:** gated fast Fourier convolution (GFFC); high-frequency feature fusion (HFF); infrared small target detection

红外小目标检测是计算机视觉领域的一项重要任务,在军事<sup>[1]</sup>、安防<sup>[2]</sup>和工业<sup>[3]</sup>等领域具有广泛的应用价值。由于成像距离的增加和红外辐射能量的衰减,红外小目标在远距离下成像效果较差。反映在图像上,这些目标通常表现出对比度低、尺寸小、视觉效果模糊以及噪声干扰等特点,给复杂场景下远距离红外小目标的检

测精度带来了挑战。

传统的红外小目标检测方法通常包含三种不同的技术途径<sup>[4]</sup>。滤波器方法<sup>[5]</sup>通过对目标场景进行估计来实现目标分离;基于人类视觉系统的方法<sup>[6-9]</sup>针对差异度较大的目标进行局部对比;低秩分解方法<sup>[10]</sup>通过分解目标信号矩阵来实现检测。然而,这些传统方法<sup>[11]</sup>存在一

**基金项目:**安徽省科技重大专项(201903a06020006);安徽省高校自然科学基金重点项目(KJ2017A353)。

**作者简介:**刘奎(1975—),男,博士,教授,研究方向为目标检测和图像去噪;唐慧萍(1999—),女,硕士研究生,研究方向为目标检测。

**收稿日期:**2023-11-22 **修回日期:**2024-03-12 **文章编号:**1002-8331(2025)07-0306-09

些不足之处。对于红外小目标检测而言,这些方法首先要求高度的精心挑选和设计特征;其次需要繁琐的模型超参数优化。更关键的是,它们过于专注于目标的物理特性,而忽视了模型在不同场景下的泛化能力,从而导致较差的检测性能。

近年来,随着深度学习的快速发展<sup>[11-12]</sup>,Dai等人在ACM<sup>[13]</sup>的基础上进一步提出了ALCNet<sup>[14]</sup>,该模型利用局部对比度模块在高层网络中突显目标。Liu等人<sup>[15]</sup>使用非重叠移位窗口自注意力结构来进行效率的提高,但由于其是基于窗口的局部特征提取,感受野仍被限制。Zhang等人<sup>[16]</sup>提出了AGPCNet网络。该网络在最深地使用注意力引导上下文学习全局特征关联,然后通过非对称融合低级和高级的语义。然而,由于缺乏对局部特征学习的关注,导致细粒度的特征信息缺失。前述研究主要着眼于对局部特征和全局特征的单独学习。其中,局部特征的学习侧重于保留细粒度信息,但缺乏全局信息的引导,因此难以清楚地区分在背景和目标中学到的细节信息。同样全局特征的学习可以有效地关注区域之间的关系,若没有局部特征的指导,也会导致缺失细节和空间位置信息。综合考虑,研究特征的学习需要将局部特征和全局特征进行联合学习。Wang等人<sup>[17]</sup>提出了UCFNet,通过在编码器的最后一层使用7次快速傅里叶卷积(fast Fourier convolution,FFC),进行全局特征与局部特征融合,提取全局信息的同时防止小目标丢失。然而,本文研究重点是解决复杂场景下目标尺寸差异大的问题,所以将FFC放在了U-Net编码器中四层下采样的位置,提取多尺度的全局和局部特征,以适应不同尺度的目标。将FFC直接放置在编码器四层下采样的位置,虽然可以实现提取多尺度特征的目的,但也会导致参数量成倍增加的问题。原始的FFC虽然考虑到了对全局和局部分支的处理,但在网络架构的训练中一直保持固定值状态,未能根据网络环境的变化进行优化。为了减轻参数量的成倍增加,提升模型在不同复杂程度数据集上的训练效果。本文利用原始分支结构并将其命名为门控,然后使用超参数门控来优化FFC结构,形成GFFC。GFFC就是利用经验性的门控合理地调节网络对目标全局和局部特征的权重,从而学习到适应性强的全局和局部特征表示,平衡网络对全局和局部特征需求。

此外,为解决红外小目标形状差异大的问题。Dai等人<sup>[14]</sup>设计端到端的局部对比度特征细化层,并且使用自下而上注意力,将低层细节集成到深层特征中,维持小目标的细节特征。Liu等人<sup>[18]</sup>设计多头注意力和特征增强模块,前者学习交互信息,后者学习小而暗目标的细节特征。近年来,在频域中学习有效的特征受到了关注。频域中常见的方法是通过增强图像的高频子带特征来提高目标的清晰度和辨识度,而小波变换在区分高频部分的目标和低频部分的背景方面具有显著效果。易图明等人<sup>[19]</sup>利用小波变换提取图像的低频和高频系

数,针对低频分量进行加权平均融合,对高频分量进行滤波处理,实现图像中目标细节的增强。Yu等人<sup>[20]</sup>通过离散小波变换提取频带特征,在图像恢复中实现图像空间域的补全和缓解频间冲突。尽管这些研究表明离散小波变换可以融合高低频信息,然而直接将离散小波嵌入在网络中并不适合红外小目标检测任务,这由于图片中小目标通常淹没在背景噪声中,简单地融合高频和低频信息可能会造成信息的混叠。

为了解决远距离复杂场景下小目标尺寸和形状差异大的问题,本文定义了门控快速傅里叶卷积(GFFC)和低频特征融合(HFF)两个模块,提出了一个新的GFUWAVE-Net框架。首先,GFFC放在U-Net编码器四层下采样的位置,提取多尺度的全局和局部特征,根据数据集的复杂程度,利用超参数门控确保合理地调节网络对目标全局和局部特征的分配,平衡网络对全局和局部特征需求。其次,对于形状差异大的问题,在获得前馈层输出适应性强的全局和局部特征基础上,HFF模块提取高频子带特征,得到的高频信息与每次上采样的特征图进行融合,增强红外小目标的纹理细节信息。最后,通过将两个模块进行协同使用,提高红外小目标检测精度。

## 1 方法

### 1.1 总体框架

不同尺寸的目标所需要的感受野存在差异,因此通过设计网络提取不同深度特征可以适用不同尺寸的目标。而在远距离复杂场景下红外小目标尺寸和形状差异大。基于以上的思想,本文设计了GFUWAVE-Net模型,如图1所示,首先,在编码器的每一个阶段都进行全局和局部特征学习,获得多尺度的目标特征,接着将目标特征信息通过高频特征融合模块,提取高频子带特征,作为小目标细节补充。最后,将多尺度特征进行上采样融合,有效地增强目标分割效果。具体地,根据目标特点将网络进行重新设计,为了适应目标尺度和提升模型在不同复杂程度数据集上的训练效果,将在1.2节中简要介绍门控快速傅里叶卷积的应用;对于形状差异问题,通过调整小波变换的滤波器,在1.3节中提出高频特征融合模块,使其更加响应高频信号。

### 1.2 门控快速傅里叶卷积

本文考虑到远距离复杂场景下目标尺寸差异问题,基于U-Net架构的红外小目标检测容易出现小目标消失现象。因此,为了适应不同尺度的目标,引入FFC残差结构放在U-Net编码器中每层下采样的位置,以获得多尺度下的目标特征。为了提升模型在不同复杂程度数据集上的训练效果,将FFC优化为GFFC。图2详细描述了该模块主要内容。具体来说,在特征图输入GFFC时,新增加两次卷积操作,结合批归一化和ReLU激活函数,首先进行特征筛选,以消除部分图像背景中的噪声信息。

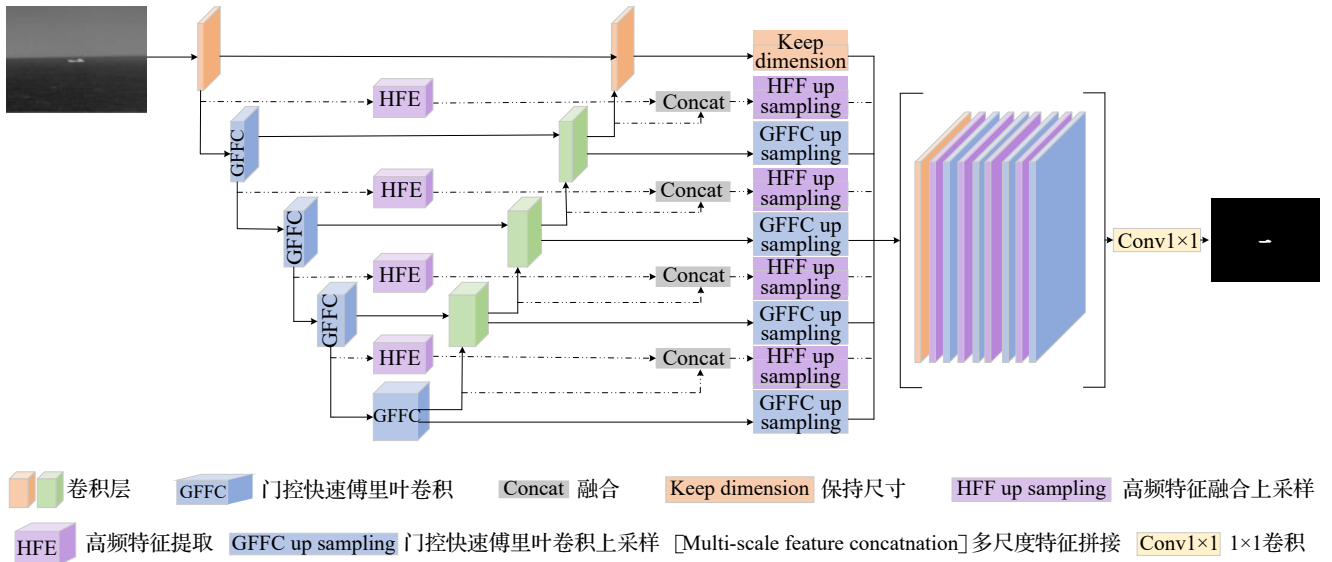


图1 门控卷积和低频特征融合的网络结构图(GFUWAVE-Net)

Fig.1 Network architecture diagram for gated convolution and high-frequency feature fusion(GFUWAVE-Net)

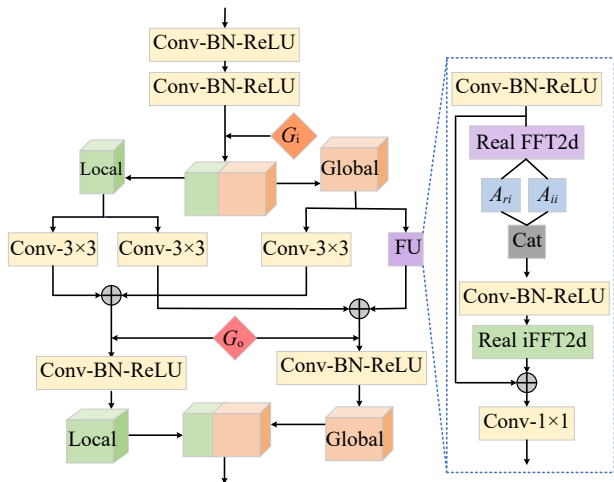


图2 门控快速傅里叶卷积的详细结构

Fig.2 Detailed structure of gated fast Fourier convolution

其次,由于在红外场景下拍摄目标和背景复杂程度的不同,数据集复杂程度也会随之变化<sup>[17]</sup>。在原始FFC分支结构中训练过程参数不会发生变化。为了使网络可以根据数据集复杂程度进行经验性优化,将原始的FFC模块中的分支部分命名为门控。门控模块 $G_i$ 用于控制全局和局部维度的输入,其中局部分支采用两个并行的 $3 \times 3$ 普通卷积来提取局部细节特征。同时,全局分支也进行并行处理,包括 $3 \times 3$ 的普通卷积和傅里叶单元(Fourier unit, FU)。FU本质利用频域信息卷积来学习空间位置之间的相互作用域, $G_i$ 用来调控输出以此来达到输入和输出的平衡。每次模型训练时,通过调节门控结构来控制输入全局和局部信息的分配比例,平衡网络对全局和局部特征需求,从而提升模型的训练效果。最后,将得到的全局信息与局部信息进行加操作,然后进行卷积、批归一化和ReLU操作,两部分结果融合输出,通过这些操作获得全局上下文信息。

GFFC的过程可以描述为:

$$Y_L = Conv_{L \rightarrow L}((X_L) \times (1 - G_i)) + Conv_{G \rightarrow L}((X_G) \times G_i) \quad (1)$$

$$Y_G = Conv_{L \rightarrow G}((X_L) \times (1 - G_i)) + Conv_{G \rightarrow G}((X_G) \times G_i) \quad (2)$$

$$out = Conv(Y_L \times (1 - G_o)) + Conv(Y_G \times G_o) \quad (3)$$

其中,  $X_L$  表示进入Local部分的卷积核,  $X_G$  表示进入Global部分的卷积核,  $G_i$  表示门控值,它通过实验得出的经验值,  $Y_L$  表示通过加操作得到的局部分支,  $Y_G$  表示通过加操作得到的全局分支,  $G_o$  表示门控的输出控制值,  $out$  为GFFC模块的输出结果。

### 1.3 高频特征融合

为了更好地捕捉红外小目标的细节纹理信息,并克服由于形状差异引起的轮廓分割不佳问题。本文进一步提出了HFF模块(high-frequency feature extraction, HFE), HFF结构如图3所示。前馈GFFC输出适应性强的全局和局部信息,为了更好地筛选出红外小目标的特征信息,并考虑到其与背景之间存在明显的颜色差异,可以引入离散小波变换进行高频子带特征提取。通过前者实现空间目标定位,后者提取目标与背景之间的对比度特征。

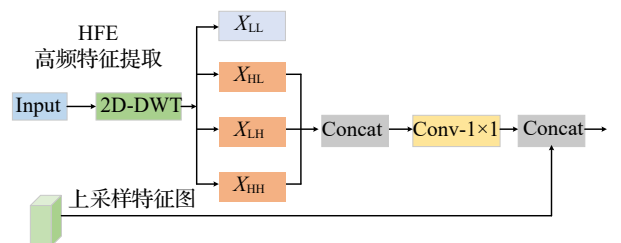


图3 高频特征融合详细结构图

Fig.3 High frequency feature fusion detailed structure

其中,二维离散小波变换具有一个低通滤波器 ( $F_{LL}$ ) 和三个高通滤波器 ( $F_{LH}, F_{HL}, F_{HH}$ ), 首先在使用的方法中引入 Haar DWT,  $F_{LL} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ ,  $F_{LH} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$ ,  $F_{HL} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$ ,  $F_{HH} = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$ 。再与滤波器卷积之后,输入分解成四个分量,为了获取目标和背景之间对比度特征,并突出目标的边缘和细节,本文选择其中的三个高频分量,即  $X_{LH}$ 、 $X_{HL}$  和  $X_{HH}$ 。 $X_{LH}$  可以表示为公式(4),其中  $i$  和  $j$  表示像素的索引,其他两个子带  $X_{HL}$  和  $X_{HH}$  的表达式与  $X_{LH}$  相似。

$$X_{LH} = -X(2i-1, 2j-1) - X(2i-1, 2j) + X(2i, 2j-1) + X(2i, 2j) \quad (4)$$

其次,将三个高频分量进行融合,通过卷积进一步对高频特征进行归纳整合,并将整合后高频信息与每次上采样的特征图进行融合,作为小目标细节特征的补充,使网络更加关注背景和目标的对比度信息,增强红外小目标细节纹理特征。

最后,在原始多尺度融合的基础上并行融合每层提取的同尺度高频特征,小目标的上下文信息通过多尺度聚焦融合被增强,使模型可以达到更好的检测效果。

## 2 实验

### 2.1 数据集和指标

#### 2.1.1 数据集

本文在 SIRST<sup>[13]</sup>、IRSTD<sup>[20]</sup> 和 MSRS\_N<sup>[21]</sup> 三个公开数据集上进行实验。SIRST 数据集包含 427 幅真实红外外场景的代表性图像,这些红外图像分布短波、中长波和 950 nm 波长中,许多目标在图片中表现得非常模糊,并且隐藏在复杂背景中。IRSTD 数据集包含 1 000 幅具有不同目标形状、大小和丰富杂波背景的图像,这些红外图像是从较远成像距离拍摄到的不同位置的小型目标,数据集涵盖背景非常广泛,这使得目标在图像中更加难以检测。

MSRS\_N 是基于公开 MSRS 数据集场景下构建的训练和测试数据集,其中包含了夜间同一场景下的 729 幅红外和可见光图像对,这些图像的空间分辨率大小为 480×640。

#### 2.1.2 评价指标

本文主要使用交并比 (IoU) 和归一化的交并比 (nIoU) 作为小目标分割的评价指标。IoU 指标被定义为预测的值和真实值之间的交集区域与并集区域的比值,用于判断检测结果和真实目标之间的重叠程度。nIoU 指标是 IoU 指标的标准化版本,这两个指标的值越大表示效果越好。

$$IoU = \frac{1}{n} \frac{\sum_{i=0}^n tp_i}{\sum_{i=0}^n (fp_i + fn_i - tp_i)} \quad (5)$$

$$nIoU = \frac{1}{n} \sum_{i=0}^n \frac{tp_i}{fp_i + fn_i - tp_i} \quad (6)$$

其中,  $n$  表示样本总数,  $tp$ 、 $fp$  和  $fn$  分别表示真阳性、假阳性和假阴性像素的数目,  $tp_i$ 、 $fp_i$  和  $fn_i$  中的  $i$  表示第  $i$  个样本。

### 2.2 实验细节

本文实验使用了一台 RTX A5000 服务器,采用 Pytorch 框架进行实验。优化器采用 AdamW,学习率采用固定值 0.000 5,完成了 300 次迭代,批量大小采用的值为 6。损失函数使用了 SoftIoULoss 和二进制交叉熵。

### 2.3 与其他先进方法的对比

本文通过在 SIRST 和 IRSTD 数据集上进行对比实验,评估了提出的 GFUWAVE-Net 方法与最近方法在性能方面的差异,具体结果如表 1 所示。研究发现,相较于传统方法<sup>[7-8, 20]</sup>,深度学习方法<sup>[14-18]</sup>通常表现更为优秀。与其他深度学习方法相比,本文提出的方法在两个数据集上都取得了显著的提升。相比基准 UCF 方法,在 SIRST 数据集上 IoU 和 nIoU 指标分别实现 0.83 个百分点和 0.40 个百分点精度的提升。在 IRSTD 数据集上,本文方法在 IoU 和 nIoU 指标上分别实现了 5.18 个百分点和 0.23 个百分点精度的提升。同时,对于网络参数量和计算复杂度,本文在 SIRST 数据集上与其他方法进行了对比,在 SIRST 数据集上表 2 的结果表明,本文方法在适当牺牲参数量或计算复杂度的前提下可以实现更好的目标检测性能。

表 1 在两个数据集上不同模型评价指标的对比

Table 1 Comparison of evaluation indicators of different models on two datasets 单位:%

Method	SIRST		IRSTD	
	IoU	nIoU	IoU	nIoU
IPI <sup>[7]</sup>	40.48	50.95	14.40	31.29
RIPT <sup>[8]</sup>	25.49	33.01	8.15	16.12
PSTNN <sup>[11]</sup>	39.44	47.72	16.44	25.91
ALCNet <sup>[14]</sup>	74.30	73.10	62.00	59.60
Res-Vit <sup>[18]</sup>	72.82	71.22	61.89	60.64
SwinT <sup>[15]</sup>	64.47	66.56	59.89	58.78
AGPCNet <sup>[15-16]</sup>	73.69	72.60	62.26	60.58
UCF <sup>[17]</sup>	79.74	77.95	63.49	62.73
Ours	<b>80.57</b>	<b>78.35</b>	<b>68.67</b>	<b>62.96</b>

表 2 其他方法参数量和计算复杂度的对比

Table 2 Comparison of number of parameters and computational complexity of other methods

Method	Params/10 <sup>6</sup>	Computational complexity/GMac	SIRST	
			IoU/%	nIoU/%
AGPCNet <sup>[16]</sup>	12.43	172.84	73.69	72.60
SwinT <sup>[15]</sup>	26.14	27.71	64.47	66.56
UCF <sup>[17]</sup>	26.64	60.11	79.74	77.95
Ours	24.82	94.74	80.57	78.35

## 2.4 不同检测方法可视化结果比较

本文在SIRST数据集中选用6个和其他检测方法<sup>[14, 17-18]</sup>部分相同的典型场景图像进行对比测试。在选用的图像中,场景6和场景2相对较小,它们呈现了从空中场景到陆地场景的转变。场景5和场景3的目标相对次小。在场景5中,存在与目标相似的背景,这容易导致误检测。在背景方面,场景3比场景5更复杂。场景1和4都属于天空背景中较大的目标,且场景1中的目标数量较多。在选择典型场景中物体的形状各有差异,例如圆形、方形、飞机形状以及点状。其中,准确检测到的目标使用红色框表示,而误检测的目标则用黄色框表示。

结果如图4显示,传统的红外小目标检测方法IPI<sup>[7]</sup>和PSTNN<sup>[11]</sup>方法,在不同场景下的检测效果都较差。SwinT<sup>[15]</sup>方法在检测目标的细节纹理方面表现不佳,并且容易受到背景物体的干扰,导致误检测问题。这是因为SwinT方法采用局部特征提取的移位加窗自注意力操作,感受野受到限制,所以在处理包含多个小目标的图像时,不能很好地对目标形状进行分割。ALCNet<sup>[14]</sup>方法使用局部对比度细化层,仅将低级语义与深层语义进行整合,未考虑对象与全局信息的直接联系,因此对所有场景中小目标轮廓细节表现欠佳。AGPCNet<sup>[16]</sup>学习全局特征关联,可以检测到目标位置信息,但由于缺乏局部特征的学习,导致对一些较复杂的目标细粒度特征信息缺失,例如场景3、4和5。

然而,当背景和背景之间非常相似时,这些方法仍然难以有效区分目标和背景的差异,对小目标的轮廓细

节检测仍然存在一定的挑战。UCF方法<sup>[17]</sup>在网络编码器末端使用了7次FFC残差块,只考虑到小目标高级语义的全局与局部信息,丢弃了低级语义中的全局与局部信息,不利于目标进行更好的检测。本文提出的模型能够有效地提取多尺度的全局和局部特征,从而适应不同场景下的目标尺度。此外,还对训练模型进行了优化,并根据数据集的复杂程度,调节超参数门控,以平衡网络对全局和局部特征的需求。通过提取特征图中的高频信息,突出小目标的对比度特征,增强小目标的细节纹理检测能力。

## 2.5 门控值的选取

在网络架构确定的情况下,本文对门控值的选取进行对比分析,实验数据来自SIRST和IRSTD数据集。如图5(a)所示为SIRST数据集中选取的图片,如图5(b)所示为IRSTD数据集中选取的图片,不难看出IRSTD数据集比SIRST数据集更为复杂<sup>[17]</sup>。实验采用了从0到1之间的门控值,并以0.1为间隔进行实验。实验研究先对较大范围内的门控值进行粗略的搜索,以获得初始的实验结果和方向。在选择IoU最佳实验结果时,会进一步考虑在两个相邻较好的实验值之间以0.05为单位完善实验结果。通过实验,选择最高IoU所对应的门控值,即为经验性的最佳门控值。这种分阶段的实验设计可以在保证实验效率的同时,充分考虑对最佳实验结果的寻找。

在SIRST数据集上进行实验,因为小目标在背景中的淹没程度较低,根据表3实验结果表明,在0.1到0.9范围内的门控值实验结果中,门控值取0.70以上时,IoU

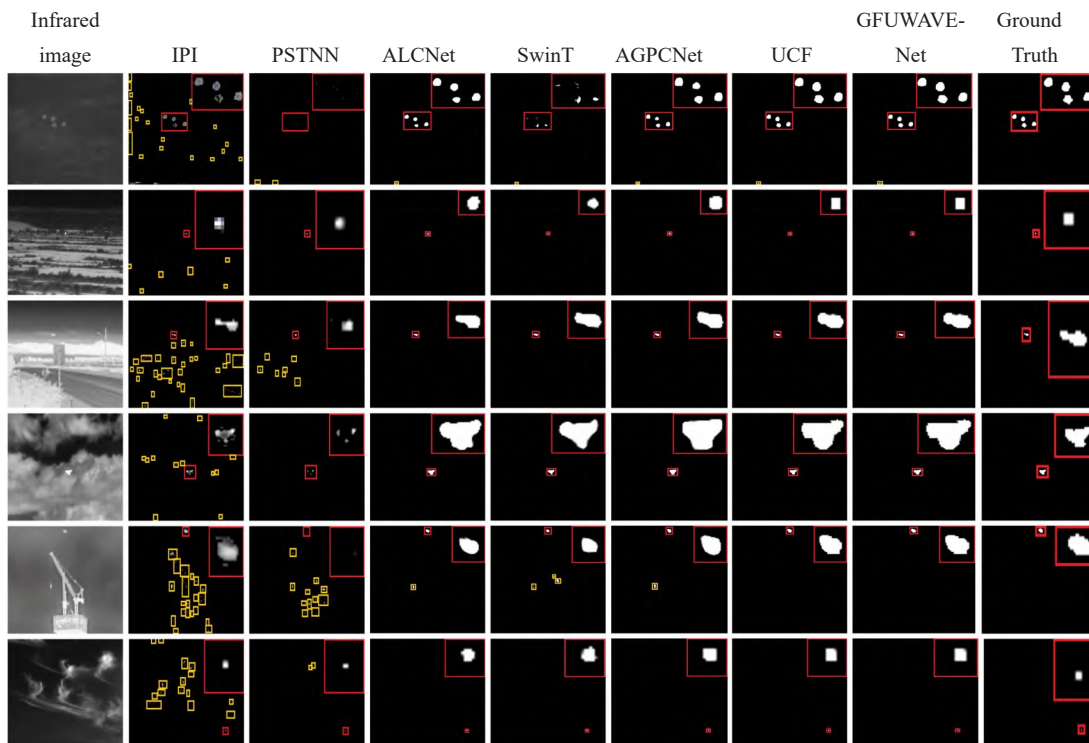


图4 在SIRST数据集上不同方法的检测结果可视化图

Fig.4 Visualization of detection results of different methods on SIRST dataset

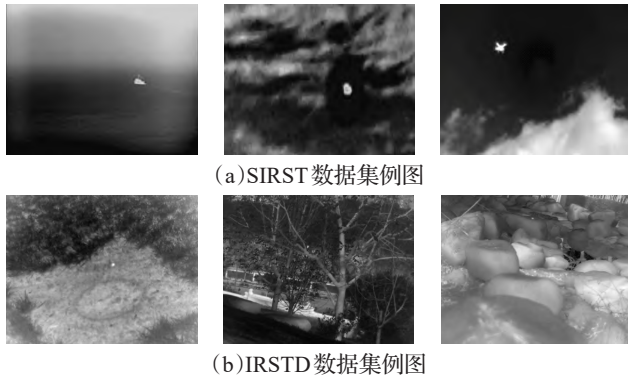


图5 两个数据集上的举例图片

Fig.5 Example pictures on two datasets

指标的波动效果较好,并且 nIoU 所代表的稳定性也有所提升。总体效果表明,检测网络更加关注图像中的全局特征。当门控值取 0.75 时,达到了最优的 IoU 值,同时模型所使用的参数量为  $24.82 \times 10^6$ 。随着 IRSTD 数据集中背景复杂性的增加,特征图会包含更多丰富的细节信息。然而,当 GFFC 增加全局分支的维度时,虽然网络可以捕捉到更多的全局特征,但是过多地忽略局部特征可能会导致无法区分噪声和背景细节。通过观察表 4 中的实验结果分布,门控值取 0.3 至 0.4 之间时, IoU 指标的波动效果较好。实验结果表明,检测网络更加关注图像中的局部特征。当门控值取 0.35 时,达到了最优的 IoU 值,同时模型所使用的参数量为  $28.58 \times 10^6$ 。

表3 SIRST 数据集上不同 Gate 值的 IoU 和 nIoU 结果  
Table 3 IoU and nIoU results for different Gate values on SIRST dataset

Gate	Params/ $10^6$	IoU/%	nIoU/%
0.90	22.72	79.03	77.81
0.80	24.17	79.93	78.03
<b>0.75</b>	<b>24.82</b>	<b>80.57</b>	<b>78.35</b>
0.70	25.45	79.51	77.96
0.60	26.55	78.68	77.22
0.50	27.49	78.55	76.47
0.40	28.27	79.34	77.57
0.30	28.86	78.97	77.49
0.20	29.29	78.93	76.84
0.10	29.54	76.29	75.29

综合以上分析,在网络架构确定的情况下,根据数据集的复杂程度来优化门控值的选取,从而提升模型训练效果。在 SIRST 和 IRSTD 数据集上,经验性地得到了最佳门控值 0.75 和 0.35,对应的参数量分别为  $24.82 \times 10^6$  和  $28.58 \times 10^6$ 。值得注意的是,相较于 IRSTD 数据集而言,在 SIRST 数据集上选择最佳门控值时,减少了必要参数的使用,同时平衡全局和局部特征的需求。

## 2.6 消融实验

本文在 SIRST 和 IRSTD 数据集上进行了消融实验,以验证 GFFC 和 HFF 模块的有效性。在 MSRS\_N 数据

表4 IRSTD 数据集上不同 Gate 值的 IoU 和 nIoU 结果  
Table 4 IoU and nIoU results for different Gate values on IRSTD dataset

Gate	Params/ $10^6$	IoU/%	nIoU/%
0.90	22.72	63.67	62.86
0.80	24.17	64.86	62.47
0.70	25.45	62.46	62.55
0.60	26.55	63.38	61.42
0.50	27.49	63.89	62.19
0.40	28.27	66.37	60.56
<b>0.35</b>	<b>28.58</b>	<b>68.67</b>	<b>62.96</b>
0.30	28.86	65.78	59.95
0.20	29.29	62.99	59.56
0.10	29.54	64.12	62.19

集上,研究模型在复杂的红外和可见光场景下的适用性。在表 5 和表 6 中,不同的数据集下显示了消融实验的结果,这些实验表明每个模块都能提升基准性能,并且它们之间具有互补的效果。其中,“√”表示实验中采用了该方法,“×”表示实验中不采用该方法。

表5 SIRST 数据集上不同模块的消融实验对比  
Table 5 Comparison of ablation experiments with different modules on SIRST dataset 单位:%

Baseline	GFFC	HFF	IoU	nIoU
√	×	×	72.22	73.90
√	√	×	78.92	77.84
√	×	√	73.65	75.11
√	√	√	80.57	78.35

表6 IRSTD 数据集上不同模块的消融实验对比  
Table 6 Comparison of ablation experiments with different modules on IRSTD dataset 单位:%

Baseline	GFFC	HFF	IoU	nIoU
√	×	×	61.62	60.30
√	√	×	64.27	62.72
√	×	√	62.55	62.39
√	√	√	68.67	62.96

### 2.6.1 GFFC 的有效性

GFFC 利用超参数门控提升模型在不同复杂程度数据集上的训练效果,通过调节全局和局部特征的权重,学习到适应性强的全局和局部特征表示,以获取多尺度的全局和局部特征。因此,在处理不同数据集时,GFFC 需要平衡门控值和网络特征需求之间的关系,确保合理地调节全局和局部特征的权重。本文在变体中除去了 HFF 模块,只采用跳过连接进行编码和解码的连接。SIRST 数据集上,增量 GFFC 模块的 IoU 值从 72.22% 增加到 78.92%,nIoU 值则是从 73.90% 增加到 77.84%,IRSTD 数据集上,增量 GFFC 模块的 IoU 值从 61.62% 增加到 64.27%,nIoU 值则是从 60.30% 增加到 62.72%,体现出了 GFFC 模块的有效性。

本文通过可视化实验分析,旨在评估GFFC模块在不同场景下利用多尺度的全局和局部特征去适应目标的能力。在基础网络中只加入GFFC模块验证多尺度特征的适应能力,选取上采样到同等尺度的特征图命名为Layer1、Layer2、Layer3和Layer4。在SIRST数据集中,选择了三幅不同尺度的目标图片进行可视化分析,具体结果如图6所示。这些图片中,场景1和场景2都展示了海面上船只的检测图像。与场景1相比,场景2成像距离更远,并且包含了尺度差异的两种船只。根据图6中的特征图可视化结果Layer1、Layer2、Layer3和Layer4观察,有助于理解网络在不同深度层次上的响应模式和感受野大小。对于场景2中不同尺度的船只,检测模型能够学习到更多抽象和高级的特征表示,随着网络深度的增加,G4中只能感知到两个船只的高级语义和位置信息。将图6场景1中可视化和基础U-Net网络可视化结果进行对比,也能体现出GFFC模块对于感知目标全局和局部特征的适应性。

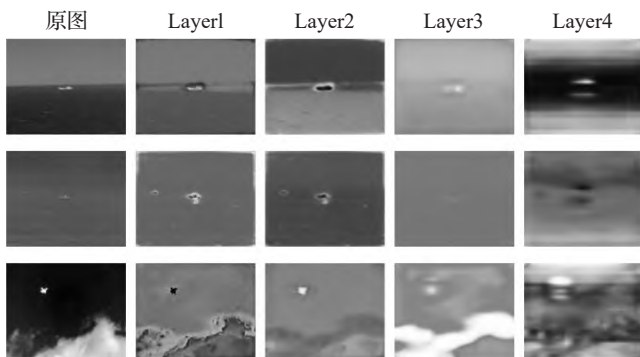


图6 GFFC模块可视化图

Fig.6 GFFC module visualization

### 2.6.2 HFF的有效性

每次池化后通过HFF模块,网络利用适应性更强的全局和局部信息提取高频子带特征,高频子带特征再与上采样结果融合,增强红外小目标细节纹理信息,小目标的上下文信息通过多尺度聚焦融合被增强。本文在变体中去除GFFC模块,采用两次卷积替代。SIRST数据集上,增量HFF模块的IoU值从72.22%增加到73.65%,nIoU值则是从73.90%增加到75.11%,IRSTD数据集上,增量

模块的IoU值从61.62%增加到62.55%,nIoU值则是从60.30%增加到62.39%,体现出了HFF模块的有效性。

本文通过可视化实验分析,旨在评估HFF模块在增强目标纹理细节的能力。将基础的U-Net网络<sup>[22]</sup>与仅加入HFF模块的U-Net网络进行了消融实验,并通过可视化进行结果对比。为了更清晰地观察HFF模块对目标纹理细节的学习能力,将HFF模块上采样到同尺度后的输出结果命名为V1、V2、V3和V4。然后,将这些输出部分与图7所示的基础U-Net网络中的S1、S2、S3和S4部分进行了细节纹理的分析,如图8所示。从图8(a)的观察结果可以明显看出,U-Net架构通过四次池化操作对背景噪声进行了过滤。然而,在过滤同时也导致了小目标的细节纹理信息丢失。尽管可以通过多尺度融合来补充缺失的信息,但仍然对目标的细节关注不够。根据图8(b)可以观察到,在U-Net架构中加入HFF模块后,特征图中提取到的高频特征更多地关注目标和背景之间的对比度信息。这种对比度信息的提取进一步增强了对小目标周围纹理和边缘信息的感知能力。V4层到V1层是对目标细节的具体化,池化去除背景噪声的同时提取高频信息,V3和V4可以清晰地观察到HFF模块对于边缘纹理的感知,V1和V2在感知边缘的同时定位。

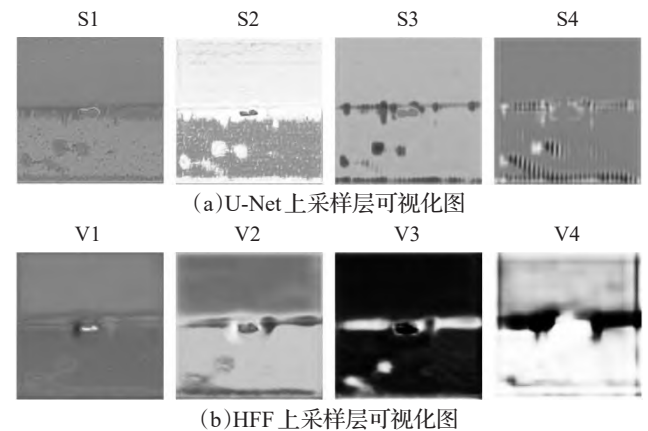


图8 HFF有效性可视化对比图

Fig.8 HFF validity visualization comparison chart

根据图9所示,Mask是检测图片的真实标签,S\_out是U-Net架构的检测结果可视化,V\_out是加入HFF模

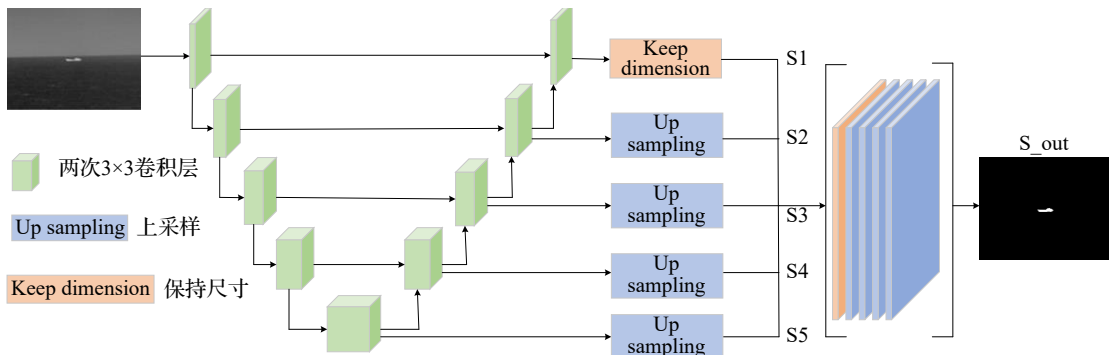


图7 基础U-Net网络结构

Fig.7 Basic U-Net network structure

块后检测结果可视化,以上图片都是从512×512尺度的图片上截图放大。通过对比这三张图片,可以观察到V\_out图像中的小目标感知纹理和边缘信息较好。这说明HFF模块在突出描绘目标形状方面具有显著作用。

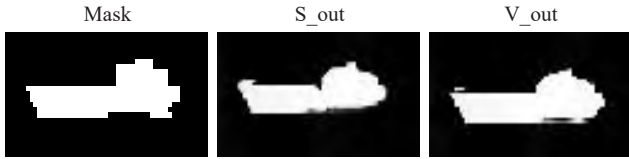


图9 标签和检测结果可视化对比

Fig.9 Visual comparison of labels and test results

### 2.6.3 不同模块作用效果对比

红外小目标的尺度大小在几个像素到几十像素不等,在SIRST数据集中选用不同尺寸和形状的小目标进行各模块有效性的验证,如表7所示。选择的目标中最小是9×7像素,最大的目标是49×16像素,如图10所示,在基础架构中无模块<sup>[22]</sup>和只使用HFF模块的情况下,场景2和场景6无法很好地分割出背景中过小的目标,并且容易出现误检测。而在使用了GFFC模块的情况下,通过利用多尺度特征能够有效适应目标尺度的变化,区分目标和背景的区别,从而更准确地定位目标。例如场景6中,GFFC对误检问题的解决。

表7 远距离红外不同尺度和形状的目标场景

Table 7 Remote infrared target scenes of different scales and shapes

场景	目标像素大小	形状
场景1	49×16	船只
场景2	15×9,41×15	两艘尺度不同船
场景3	27×18	车辆
场景4	23×19	鸟/飞机
场景5	38×32	飞机
场景6	9×7	未知飞行物体

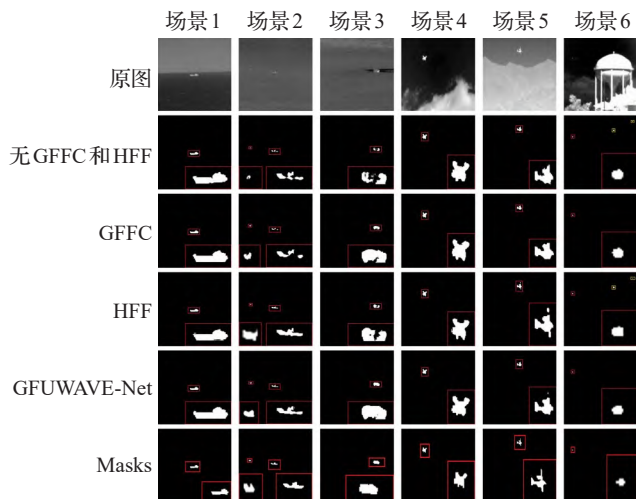


图10 不同模块的检测结果可视化

Fig.10 Visualization of detection results for different modules

观察场景1到6的比较结果可以发现,HFF模块在感知不同形状目标的边缘和纹理细节方面表现出了明显的优势。结合以上分析,本文采用GFFC模块确定目标位置,并通过HFF模块对目标的细节特征进行补充。两个模块相互补充,从而实现了较好的分割效果。

### 2.6.4 模型在红外和可见光场景的应用

本文利用包含红外和可见光图像的MSRS\_N数据集<sup>[21]</sup>进行对比实验,以研究模型在不同场景下的适用性。该数据集中包含了同一场景下的红外和可见光图像对,如图11所示。本文将MSRS\_N数据集分为两个部分,即红外场景数据集和可见光场景数据集。为了确保公平性,在同一场景中保持所有实验条件相同,将两个数据集中的门控值都设定为0.5,控制GFFC中全局和局部分支各占一半。然而,根据表8在MSRS\_N红外和可见光数据集上的实验结果可见,本文提出的模型在MSRS\_N红外数据集上IoU和nIoU值分别为68.18%和62.96%。在MSRS\_N可见光数据集上IoU和nIoU值分别为60.19%和51.37%。综合实验结果可知,该模型比较适用于红外图像数据集。



图11 红外和可见光图像对

Fig.11 Infrared and visible image pairs

表8 红外和可见光数据集上对比评估

Table 8 Comparative evaluation on infrared and visible datasets

数据集	单位:%	
	IoU	nIoU
MSRS_N红外数据集	68.18	62.96
MSRS_N可见光数据集	60.19	51.37

## 3 总结与展望

本文提出了一个基于门控卷积和低频特征融合的红外小目标检测框架,针对在远距离复杂场景下红外小目标尺寸和形状差异的问题给出了解决方法。具体来说,为了获得多尺度的全局和局部特征,并提升模型在不同复杂程度数据集上的训练效果,本文将GFFC模块嵌入到U-Net编码器的四层下采样位置。通过这种方式,获得多尺度的全局和局部特征,门控机制被引入,以平衡网络对全局和局部特征的需求。为了更好地感知目标形状特征,采用HFF模块提取高频子带特征,作为小目标细节信息的补充,同时增强红外小目标细节纹理信息。SIRST和IRSTD数据集的与其他先进方法实验对比表明GFUWAVE-Net的优越性和鲁棒性。然而,在编码器中,GFFC中的门控值是超参数,它根据数据集的



复杂程度变化而进行优化,需要人工手动进行调节,对于尺度的灵活性没有很好地把控,后续研究旨在设计可学习的门控结构,自适应地调整网络的特征结构分布,使整体模型更加健壮。因此,未来的工作将在本文基础上继续优化网络,以提高目标检测的准确性和稳定性。

### 参考文献:

- [1] 赵兴科,李明磊,张弓,等.基于显著图融合的无人机载热红外图像目标检测方法[J].自动化学报,2021,47(9):2120-2131. ZHAO X K, LI M L, ZHANG G, et al. Object detection method based on saliency map fusion for UAV-borne thermal images[J]. Acta Automatica Sinica, 2021, 47(9): 2120-2131.
- [2] 李其昌,李兵伟,王宏臣.非制冷红外成像技术发展动态及其军事应用[J].军民两用技术与产品,2016(21):54-57. LI Q C, LI B W, WANG H C. Development trend of uncooled infrared imaging technology and its military application[J]. Dual Use Technologies & Products, 2016(21): 54-57.
- [3] 王云杰,王艳林,夏润秋,等.大视场红外告警系统中目标高精度方位提取[J].激光技术,2023,47(2):200-204. WANG Y J, WANG Y L, XIA R Q, et al. High precision azimuth extraction of target in large field of view infrared warning system[J]. Laser Technology, 2023, 47(2): 200-204.
- [4] 杨亚东,黄胜一,谭毅华.基于低秩和重加权稀疏表示的红外弱小目标检测算法[J].应用科学学报,2023,41(5):753-765. YANG Y D, HUANG S Y, TAN Y H. Infrared dim and small target detection algorithm based on low-rank and reweighted sparse representation[J]. Journal of Applied Sciences, 2023, 41(5): 753-765.
- [5] 张婷,张兴忠,王慧民,等.基于神经网络的变电站场景三维目标检测[J].计算机工程与应用,2023,59(9):329-336. ZHANG T, ZHANG X Z, WANG H M, et al. 3D object detection in substation scene based on graph neural network[J]. Computer Engineering and Applications, 2023, 59(9): 329-336.
- [6] 黄磊,杨媛,杨成煜,等.FS-YOLOv5:轻量化红外目标检测方法[J].计算机工程与应用,2023,59(9):215-224. HUANG L, YANG Y, YANG C Y, et al. FS-YOLOv5: light-weight infrared rode target detection method[J]. Computer Engineering and Applications, 2023, 59(9): 215-224.
- [7] 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.
- [8] DAI Y M, WU Y Q. 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.
- [9] 王志武,张子森,许凯,等.基于最佳偏振角的线性姿态测量方法研究[J].红外与激光工程,2022,51(3):329-338. WANG Z W, ZHANG Z M, XU K, et al. Research on linear pose measurement method based on optimal polarization angle[J]. Infrared and Laser Engineering, 2022, 51(3): 329-338.
- [10] ZHU H, LIU S M, DENG L Z, et al. Infrared small target detection via low-rank tensor completion with top-hat regularization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(2): 1004-1016.
- [11] ZHANG L D, PENG Z M. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. Remote Sensing, 2019, 11(4): 382.
- [12] 陆慧敏,杨朔.基于深度神经网络的自动驾驶场景三维目标检测算法[J].北京工业大学学报,2022,48(6):589-597. LU H M, YANG S. Three-dimensional object detection algorithm based on deep neural networks for automatic driving [J]. Journal of Beijing University of Technology, 2022, 48(6): 589-597.
- [13] DAI Y M, WU Y Q, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2021: 950-959.
- [14] DAI Y M, WU Y Q, 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.
- [15] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 10012-10022.
- [16] ZHANG T F, CAO S Y, PU T, et al. AGPCNet: attention-guided pyramid context networks for infrared small target detection[J]. arXiv:2111.03580, 2021.
- [17] WANG C Y, WANG H, PAN P W. Local contrast and global contextual information make infrared small object salient again[J]. arXiv:2301.12093, 2023.
- [18] LIU F C, GAO C Q, CHEN F, et al. Infrared small and dim target detection with transformer under complex backgrounds[J]. IEEE Transactions on Image Processing, 2023, 32: 5921-5932.
- [19] 易图明,王先全,袁威,等.基于导向滤波和小波变换的红外可见光图像融合改进算法研究[J].现代信息科技,2023,7(6):41-45. YI T M, WANG X Q, YUAN W, et al. Research on improved infrared visible light image fusion algorithm based on guided filtering and wavelet transform[J]. Modern Information Technology, 2023, 7(6): 41-45.
- [20] YU Y C, ZHAN F N, LU S J, et al. WaveFill: a wavelet-based generation network for image inpainting[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 14094-14103.
- [21] MA J Y, TANG L F, FAN F, et al. SwinFusion: cross-domain long-range learning for general image fusion via swin transformer[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(7): 1200-1217.
- [22] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.