**中图法分类号**:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-17

**论文引用格式**:Xiao Qin, Zhang Yuming, Zhang Yingmei, Yang Yong, Wan Weiguo. Dual-branch interactive guidance network for infrared small target detection[J/OL]. Journal of Image and Graphics, XXXX:1-17. DOI: 10.11834/jig.260484. (肖沁, 张与明, 张迎梅, 杨勇, 万伟国. 面向红外小目标检测的双分支交互引导网络[J/OL]. 中国图象图形学报,XXXX:1-17. DOI: 10.11834/jig.260484. )[DOI:10.11834/jig.260484]

# 面向红外小目标检测的双分支交互引导网络

肖沁[1],张与明[2],张迎梅[2],杨勇[3],万伟国[2]

1. 江西财经大学信息管理与数学学院,南昌　330032; 2. 江西财经大学软件与物联网工程学院,南昌　330032; 3. 天津工业大学计算机科学与技术学院,天津　300387

**摘　要**: **目的**　红外小目标检测(infrared small target detection,IRSTD)旨在从复杂背景中检测空间尺度小、对比度低且易被杂波淹没的弱目标。现有基于U-Net的方法虽然具有较强的多尺度表征与跨层融合能力,但大多仅针对主干特征提取或跳跃连接进行局部优化,未能同时兼顾主路径表征增强与跨层融合调控,因此在复杂场景下仍容易出现弱目标响应不足、跨层语义不一致以及背景噪声传播等问题。为此,本文提出一种双分支交互引导网络(dual-branch interactive guidance network, DIGNet)。**方法**　具体而言,本文采用分层编码器—解码器结构,在主干端引入残差U型特征提取块及其空洞扩展变体,以增强局部细节保持和多尺度上下文建模能力。在跨层融合端,提出双分支交互融合模块,对解码特征与跳跃连接特征进行显式交互,以缓解跨层语义差异并抑制噪声传播。进一步地,设计分组跨空间交互注意力模块,通过分组建模与方向敏感重标定增强目标响应并抑制结构化背景干扰。结合多尺度深监督策略,DIGNet能够实现更稳健的目标重建与背景抑制。**结果**　在三个公开基准数据集上的实验结果验证了所提出的DIGNet网络在交并比(intersection over union,IoU)、$P_d$、$F_a$三个指标上均优于现有先进方法,其中,本文方法的$F_a$指标在上述基准数据集上比次优方法分别降低$1.29\times10^{-6}$、$4.97\times10^{-6}$和$4.04\times10^{-6}$。**结论**　本文所提方法在检测精度、虚警抑制和复杂场景鲁棒性方面均优于现有主流方法。

**关键词**: 红外小目标检测;双分支交互引导网络;双分支交互融合模块;分组跨空间交互注意力模块;多尺度监督策略

# Dual-branch interactive guidance network for infrared small target detection

Xiao Qin[1], Zhang Yuming[2], Zhang Yingmei[2], Yang Yong[3], Wan Weiguo[2]

*1. School of Information Management and Mathematics, Jiangxi University of Finance and Economics, Nanchang 330032, China; 2. School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China; 3. School of Computer Science and Technology, Tiangong University, Tianjin 300387, China*

**Abstract**: **Objective** Infrared small target detection (IRSTD) aims to accurately identify and segment dim targets from complex infrared backgrounds and is essential for maritime search and rescue, early-warning surveillance, and intelligent transportation. This task remains challenging because distant targets usually occupy only a few pixels, contain limited texture and contour information, and exhibit weak contrast against their surroundings. Meanwhile, sea-surface highlights,

**收稿日期**:2026-08-04;**修回日期**:2026-10-28

**基金项目**:国家自然科学基金项目(62261025);江西省自然科学基金项目(20262BAC200017);江西省职业早期青年科技人才培养专项项目(20244BCE52075, 20244BCE52088)

**Supported by**: National Natural Science Foundation of China (62261025); Natural Science Foundation of Jiangxi Province, China (20262BAC200017); Early-Career Young Science and Technology Talent Training Project of Jiangxi Province, China (20244BCE52075, 20244BCE52088)

© 中国图象图形学报版权所有

cloud edges, isolated thermal sources, and directional ground textures can produce target-like responses, leading to missed detections and false alarms. U-Net-based models have become a dominant framework for IRSTD because their encoder-decoder structures support multi-scale representation learning and their skip connections preserve high-resolution details. However, two limitations remain insufficiently addressed. First, repeated convolution and downsampling may progressively weaken the already sparse responses of small targets, whereas conventional multi-scale structures do not always preserve fine localization cues while acquiring sufficiently discriminative context. Second, shallow encoder features contain detailed spatial information but also substantial background clutter, whereas deep decoder features provide stronger semantics but lose precise target boundaries. Direct concatenation or addition treats these heterogeneous features indiscriminately, enlarges the semantic gap across layers, and propagates background noise into the reconstruction path. Most existing methods improve either backbone feature extraction or skip-connection fusion but do not jointly regulate target representation, bidirectional cross-layer interaction, and structured clutter suppression. To address these issues, this paper proposes a dual-branch interactive guidance network (DIGNet), which integrates dual-branch multi-scale representation, cross-layer interactive guidance, and direction-aware background suppression within a unified hierarchical encoder-decoder framework. **Method** DIGNet contains four principal components: residual U-shaped blocks (RUB) and their dilated variants (RUBD), a dual-branch interactive fusion (DIF) module, a grouped cross-spatial interactive attention (GCIA) module, and multi-scale deep supervision. RUB is used in shallow and intermediate stages to construct an internal lightweight encoder-decoder path within each block. This design aggregates features at different resolutions while retaining a residual shortcut, thereby preserving local structures and stabilizing the transmission of weak target information. In deeper stages, RUBD replaces selected standard convolutions with dilated convolutions. It enlarges the effective receptive field without introducing additional downsampling, enabling the network to acquire broader contextual evidence while limiting spatial-detail loss. The hierarchical use of RUB and RUBD therefore supports fine-detail preservation in early layers and context-aware discrimination in deeper layers. At each decoding stage, the current skip-connected feature is first recalibrated by GCIA and then fused with the upsampled decoder feature through DIF. Unlike static skip fusion, DIF explicitly establishes bidirectional guidance between the two inputs through complementary channel- and spatial-interaction branches. In the channel-interaction branch, projected decoder and skip features are correlated to generate a cross-branch weighting matrix. Decoder semantics consequently select target-relevant channels from the high-resolution skip feature and suppress channels dominated by clutter. In the spatial-interaction branch, mapped and normalized skip information generates a spatial gate that constrains the decoder response, allowing detailed positional cues to correct coarse semantic reconstruction. The outputs of the two branches are adaptively combined using learnable coefficients and further refined with coordinate attention, which preserves direction-sensitive positional information. This reciprocal design does not merely transfer encoder details to the decoder; instead, it uses decoder semantics to filter skip information and skip details to regularize decoder localization, thereby reducing cross-layer semantic inconsistency and background-noise propagation. GCIA is applied before and after cross-layer fusion to improve the distinction between true targets and structured background responses. It divides the input channels into groups and performs attention modeling within each group, reducing computational redundancy while improving the stability of local statistics. A direction-aware branch aggregates features along the horizontal and vertical axes to characterize directional patterns such as sea streaks, cloud bands, and elongated ground structures. In parallel, a within-group interaction branch combines local convolution and global aggregation to estimate spatial saliency. The two branches jointly produce recalibration weights that enhance compact target responses and suppress structured clutter. Finally, multiple decoder outputs are supervised using a combination of Soft-IoU and binary cross-entropy losses. Deep supervision provides direct gradient signals to intermediate reconstruction stages, while the joint loss promotes an appropriate balance among target-region overlap, pixel-level classification, and background suppression. **Result** To verify the effectiveness of the proposed DIGNet, extensive experiments were conducted on three widely used IRSTD datasets: Nanjing University of Aeronautics and Astronautics single-frame infrared small target (NUAA-SIRST), infrared small target detection 1K (IRSTD-1K), and National University of Defense Technology single-frame infrared small target (NUDT-SIRST). DIGNet was compared with six traditional methods (Top-Hat, Max-Median, WSLCM, TLLCM, IPI, and NOLC) and eight deep learning-based methods (ACM, RDIAN, DNANet, UIU-Net, RPCANet, MSHNet, SCTransNet, and L$^2$SKNet).

© 中国图象图形学报版权所有

Three standard metrics were employed for objective evaluation: intersection over union (IoU), probability of detection ($P_d$), and false alarm rate ($F_a$). The experimental results show that DIGNet consistently achieves the best overall performance across the three datasets. Specifically, compared with the second-best methods, DIGNet improves IoU by 1.75%, 1.77%, and 1.51% on NUAA-SIRST, IRSTD-1K, and NUDT-SIRST, respectively, while reducing $F_a$ by $1.29\times10^{-6}$, $4.97\times10^{-6}$, and $4.04\times10^{-6}$. These results demonstrate the effectiveness and robustness of DIGNet in detecting dim targets and suppressing background interference across diverse infrared scenes. **Conclusion** DIGNet addresses weak-target attenuation and background-noise propagation by coordinating multi-scale main-path extraction, bidirectional cross-layer interaction, and structured clutter suppression. RUB/RUBD preserve fine spatial evidence while expanding contextual perception; DIF converts static skip transfer into reciprocal semantic and spatial guidance; and GCIA suppresses directional pseudo-target patterns through grouped cross-spatial modeling. Extensive quantitative, qualitative, ablation, statistical-stability, and cross-dataset evaluations verify that these components are mutually complementary and that the resulting network achieves accurate target reconstruction, high detection probability, and strong false-alarm suppression across diverse infrared scenes. Future work will focus on reducing computational cost, improving adaptability to more complex environments, and enhancing infrared small target detection under limited-sample conditions.

**Key words**: Infrared small target detection; dual-branch interactive guidance network; dual-branch interactive fusion module; grouped cross-spatial interactive attention module; multi-scale deep supervision strategy

## 0 引　言

红外小目标检测(infrared small target detection, IRSTD)旨在从复杂红外背景中准确识别空间尺度小、目标对比度低的弱小目标,在海上搜救、预警监视和智能交通等领域具有重要应用价值(Kou等,2024)。由于成像距离较远、传感器分辨率有限,红外小目标通常仅占据少量像素,缺少稳定的纹理和轮廓特征。与此同时,海浪高光、云层边缘和地表热源等背景结构也可能产生与真实目标相似的局部响应,使检测过程同时面临弱目标漏检和背景虚警问题(Zhang等,2026a)。

早期IRSTD研究主要依赖人工构造的目标先验,代表性方法包括背景滤波(Zhu等,2020)、人类视觉系统建模(Han等,2014)以及低秩稀疏分解(Sun等,2021)等。此类方法具有较明确的物理意义,但通常依赖窗口尺度、结构元素和正则化参数,在复杂背景下的适应能力仍然有限(Zhang等,2026b)。近年来,深度学习凭借端到端特征学习能力显著推动了IRSTD的发展。U-Net(Ronneberger等,2015)利用编码器—解码器结构和跨层跳跃连接实现多尺度特征融合,已成为红外小目标检测的重要基础框架。在此基础上,DNANet(dense nested attention network)(Li等,2023)通过密集嵌套交互加强高、低层特征传递;SCTransNet(spatial-channel cross transformer network)(Yuan等,2024)利用空间—通道交叉Transformer建立全尺度长程依赖;UIU-Net(U-Net in U-Net)(Wu等,2023)在整体U形网络中嵌入残差U型块,以获得多层次、多尺度目标表示;ISNet(infrared shape network)(Zhang等,2022)将目标边缘与形状重建引入检测过程;ACM(asymmetric contextual modulation)(Dai等,2021)采用非对称上下文调制协调高层语义和低层细节;HAFNet(hierarchical attention fusion network)(Zhang等,2025a)则同时改进主干特征提取与分层跨层融合。此外,(convolutional block attention module, CBAM)(Woo等,2018)和坐标注意力(Hou等,2021)分别通过通道—空间重标定和方向位置编码提高网络的特征选择能力。

尽管现有方法取得了显著进展,红外小目标在网络中的特征生成和跨层传递仍存在两个值得进一步研究的问题。

首先,弱目标的有效信息容易在主干特征提取过程中逐渐衰减。连续卷积和下采样能够扩大网络感受野并获取背景上下文,但红外小目标只包含有限像素,其响应可能在深层特征中被强背景信息覆盖。现有多尺度结构虽然能够聚合不同范围的上下文,却未必能够同时保持浅层定位细节和深层语义判别信息。特别是在目标对比度较低时,主干网络若缺少稳定的多尺度特征传递路径,便难以形成清晰、连续的目标表示。

© 中国图象图形学报版权所有

其次，编码特征与解码特征之间存在明显的语义差异。浅层跳跃特征具有较高空间分辨率，但同时包含大量背景纹理；深层解码特征具有较强语义信息，却可能丢失精确的位置细节。传统拼接或加和操作通常对两类特征进行无差别融合，不仅难以充分利用二者的互补关系，还可能将浅层噪声持续传递至解码端。即使采用全尺度交互或分层特征聚合，如何在具体融合节点上利用解码语义筛选跳跃信息，并利用跳跃细节反向约束解码响应，仍有进一步细化的空间。

针对上述问题，本文提出双分支交互引导网络（dual-branch interactive guidance network，DIGNet）。在主干特征提取阶段，采用残差U型块（residual U-block，RUB）及带空洞卷积的残差U型块（residual U-block with dilated convolutions，RUBD），通过块内编码—解码路径聚合不同分辨率下的特征，在保留局部细节的同时获得更大范围的上下文信息。RUB主要用于浅层及中层特征提取，以保持小目标的空间结构；RUBD则在深层引入空洞卷积，在避免额外下采样的情况下扩大有效感受野。

在跨层融合阶段，本文设计双分支交互融合模块（dual-branch interactive fusion，DIF）。该模块分别从通道和空间两个维度建立上采样解码特征与同尺度跳跃特征之间的显式联系：通道交互分支利用解码语义筛选跳跃特征中的目标相关通道，空间交互分支利用跳跃特征提供的位置细节约束解码响应。两条分支通过残差门控和可学习权重进行自适应融合，从而减少无关背景信息的跨层传播。

为进一步提高网络对真实目标和背景伪响应的区分能力，本文设计分组跨空间交互注意力模块（grouped cross-spatial interactive attention，GCIA）。GCIA首先对输入特征进行通道分组，再结合水平、垂直方向统计和组内空间交互生成显著性权重。该模块分别作用于融合前的跳跃特征和融合后的重建特征，使网络能够增强目标区域响应，并抑制海浪条纹、云带和复杂地表纹理等结构化背景干扰。结合多尺度深监督策略，DIGNet能够实现更加稳定的目标重建与背景抑制。综上所述，本文的主要贡献如下：

1）提出一种面向红外小目标检测的双分支交互引导网络DIGNet，将主路径中的多尺度目标表征、跨层融合节点中的双向信息约束以及融合后的背景抑制纳入统一框架，以缓解弱目标响应衰减与背景噪声传播问题。

2）设计双分支交互融合模块，在通道和空间两个互补分支中建立解码特征与跳跃特征之间的显式引导关系，并通过残差门控和可学习权重实现自适应融合，从而缩小跨层语义差异并增强目标相关信息。

3）构建分组跨空间交互注意力模块，通过通道分组、方向感知建模和组内空间交互增强弱目标响应，同时抑制具有方向性和结构性的背景伪响应。

4）在NUAA-SIRST、IRSTD-1K和NUDT-SIRST三个公开数据集上开展对比实验与消融实验。实验结果表明，DIGNet在检测精度和虚警抑制方面均取得了具有竞争力的性能，验证了所提方法及各组成模块的有效性。

## 1 相关工作

### 1.1 基于深度学习的红外小目标检测

基于深度学习的方法能够从训练数据中自主学习目标特征，逐渐成为IRSTD的主要研究方向。多尺度表征与跨层特征融合是提升小目标检测性能的重要途径。在通用目标检测领域，FPN（feature pyramid networks）（Lin等，2017）通过自顶向下路径和横向连接融合不同尺度特征，PANet（path aggregation network）（Liu等，2018）进一步引入自底向上的路径增强以缩短低层定位信息的传递距离；NAS-FPN（neural architecture search-feature pyramid network）（Ghiasi等，2019）通过神经架构搜索自动构建跨尺度融合拓扑，BiFPN（bi-directional feature pyramid network）（Tan等，2020）则采用双向跨尺度连接和可学习权重提高多尺度特征融合效率。这些方法推动了跨层信息融合的发展，但其主要关注整体特征金字塔的构建与尺度间信息聚合。

在IRSTD领域，研究者进一步结合红外小目标弱纹理、低对比度及易受背景干扰等特点对特征提取与跨层融合进行针对性设计。ACM（Dai等，2021）通过非对称上下文调制协调高层语义与低层细节；DNANet（Li等，2023）利用密集嵌套交互增强不同层级特征传递；UIU-Net（Wu等，2023）通过嵌套残差U型结构获得多层次、多尺度目标表示；ISNet（Zhang等，2022）引入目标边缘与形状重建以强化

© 中国图象图形学报版权所有

结构约束；SCTransNet（Yuan等，2024）利用空间—通道交叉Transformer建立多层特征之间的全局交互；HAFNet（Zhang等，2025a）通过主干特征增强与分层注意力融合改善多尺度、多层级语义传递。此外，GCI-Net（gaussian curvature inspired network）（Zhang等，2024a）、CFD-Net（computational fluid dynamic network）（Zhang等，2025b）和HDNet（hybrid domain network）（Xu等，2025）分别从几何结构、动态特征传播以及空间—频率联合建模等角度进一步提升了弱目标表征与背景抑制能力。

现有研究已由静态跳跃连接逐步发展至双向调制、密集交互、全尺度建模和分层融合，但多数方法更侧重于扩展跨层信息流或增强多尺度聚合，对具体解码融合节点中深层语义与浅层细节之间的功能性互补关系关注相对有限。与上述方法不同，DIGNet在同尺度融合节点中建立方向明确的双向交互：DIF利用解码语义筛选跳跃特征中的目标相关通道，同时利用浅层细节约束解码端的空间定位；GCIA进一步通过分组方向建模与组内空间交互抑制结构化背景响应。

### 1.2 基于通用视觉模型的红外小目标检测

随着通用视觉模型和新型序列建模架构的发展，IRSTD的技术路线也逐渐扩展至更具泛化能力的模型。IRSAM（advancing segment anything model for infrared small target detection）（Zhang等，2024b）率先将“任意分割模型”（Kirillov等，2023）引入IRSTD任务，并结合基于小波变换的Perona-Malik扩散机制，加强对目标边缘与轮廓信息的刻画。与此同时，状态空间序列模型（state space model，SSM）（Gu等，2021）凭借高效的长距离依赖建模能力，为视觉特征中的全局关系学习提供了新的解决思路。Mamba（Gu等，2023）作为该类模型的代表，引起了广泛关注。MiM-ISTD（Mamba-in-Mamba for efficient infrared small target detection）（Chen等，2024）首次将Mamba架构应用于红外小目标检测，通过状态空间建模增强目标与背景之间的长程关系表达。在此基础上，IRMamba（pixel difference Mamba with layer restoration for infrared small target detection）（Zhang等，2025c）进一步刻画扫描位置与中心邻域之间的强度差异和方向差异，并将这些局部差异信息嵌入状态更新过程，以提升网络对细粒度目标特征的感知与保持能力。

与HAFNet和SCTransNet不同，DIGNet并非主要依赖分层注意力聚合或Transformer全局依赖建模，而是从跨层特征交互与方向敏感空间建模两个方面进行针对性设计。双分支交互融合模块在通道和空间维度协同建模解码特征与跳跃特征之间的互补关系，以缓解语义差异并抑制噪声传播；分组跨空间交互注意力模块则通过分组建模和方向敏感重标定强化小目标结构响应，提高真实目标与复杂背景伪响应之间的可分性。

## 2 本文方法

### 2.1 总体网络结构

针对红外小目标尺度小、对比度低且易受复杂背景干扰的问题，本文提出双分支交互引导网络DIGNet，如图1所示。DIGNet采用分层编码器—解码器架构，主要由RUB/RUBD、DIF、GCIA以及多尺度深监督组成。其中，RUB/RUBD用于编解码主路径中的多尺度特征提取与重建，分层跳跃连接（hierarchical skip connection，HSC）用于传递编码阶段的多尺度细节信息，解码特征注入式跳跃连接（decoded feature injection skip connection，DSC）用于向当前层注入更深层的解码语义信息。在解码阶段，当前层编码特征先经GCIA重标定，再与上一级上采样后的解码特征在DIF中进行通道—空间双分支交互融合，融合结果随后经GCIA增强和RUB/RUBD精炼，最终用于目标重建。

设输入红外图像为$\boldsymbol{X}\in\mathbb{R}^{H\times W\times C}$，编码阶段共有6个层级，其特征提取过程表示为

$$\boldsymbol{E}_1=F_1(\boldsymbol{X}),\boldsymbol{E}_i=F_i\left(P\left(\boldsymbol{E}_{i-1}\right)\right),i=2,3\cdots,6\#(1)$$

式中，$F_i(\cdot)$表示第$i$个RUB/RUBD块，$P(\cdot)$表示$2\times2$最大池化操作。解码阶段从瓶颈特征$\boldsymbol{E}_6$开始，

逐层恢复空间分辨率，其过程表示为：

$$\boldsymbol{D}_i=R_i\left(GCIA\left(DIF\left(Up\left(\boldsymbol{D}_{i+1}\right),\boldsymbol{E}_i\right)\right)\right),i=5,4,,,1\#(2)$$

式中，$Up(\cdot)$表示双线性上采样，$R_i(\cdot)$表示解码阶段的特征精炼过程。为提高训练稳定性，网络在多个解码层输出侧做预测，并将其融合得到最终结果：

$$\boldsymbol{O}_i=Up\left(Conv_{1\times1}\left(\boldsymbol{D}_i\right)\right),$$
$$\hat{\boldsymbol{Y}}=Conv_{1\times1}\left(\left[\boldsymbol{O}_1,\boldsymbol{O}_2,\cdots,\boldsymbol{O}_5\right]\right)\#(3)$$

式中，[·]表示通道拼接，$\hat{\boldsymbol{Y}}$为最终预测图。

© 中国图象图形学报版权所有

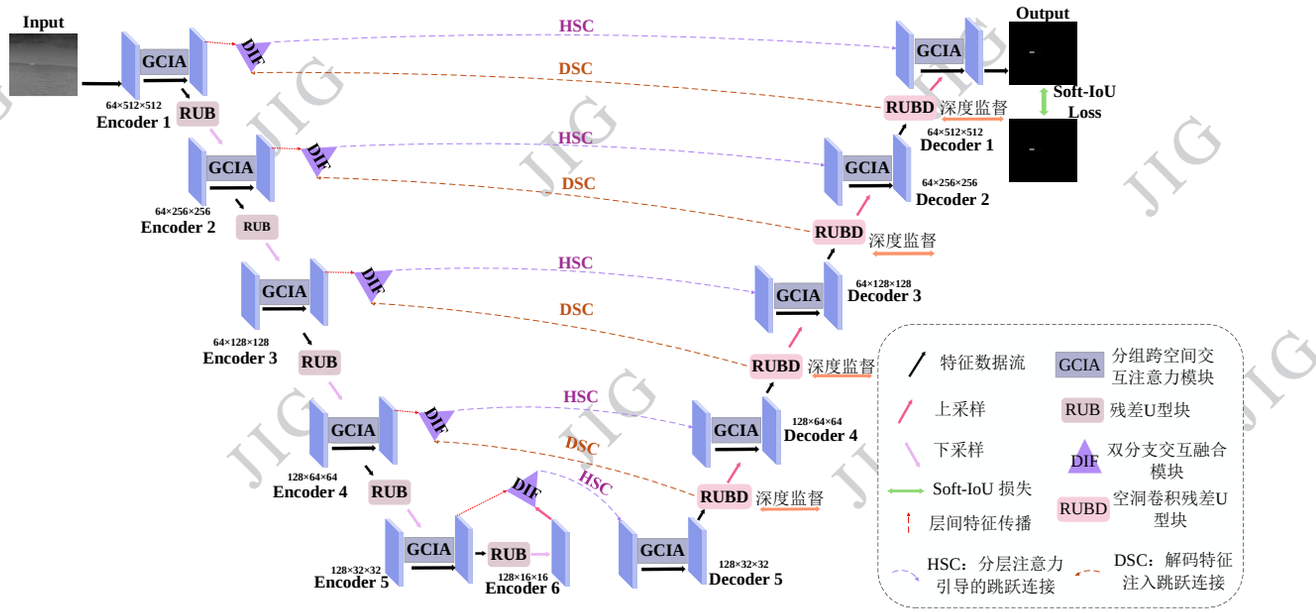



图1　DIGNet网络整体架构

Fig. 1　Overall Framework of the DIGNet Network

## 2.2　残差U型特征提取块

为了兼顾小目标的局部细节、边界结构和上下文判别信息，DIGNet采用残差U型块作为基本特征提取单元。对于输入特征$\boldsymbol{X}$，RUB在块内构建轻量级编码-解码路径以提取多尺度信息，同时保留残差捷径以稳定特征传递，其形式为：

$$RUB(\boldsymbol{X}) = U(\boldsymbol{X}) + \phi(\boldsymbol{X}) \#(4)$$

式中，$\phi(\cdot)$表示块内的浅层映射，$U(\cdot)$表示内部U形编码-解码聚合路径。该结构能够在单个模块内部同时获得局部细节和多尺度上下文信息，有利于弱小目标的稳定建模。在深层特征阶段，为避免频繁下采样带来的空间细节损失，本文采用RUBD替代标准RUB。RUBD在整体结构上与RUB保持一致，但在块内部分卷积位置引入空洞卷积，以在尽量保持分辨率的同时扩大感受野，其形式可表示为：

$$RUBD(\boldsymbol{X}) = U_d(\boldsymbol{X}) + \phi(\boldsymbol{X}) \#(5)$$

式中，$U_d(\cdot)$表示包含空洞卷积的块内编码-解码聚合过程。通过RUB与RUBD的分层配合，网络能够在浅层保留精细空间结构，在深层获得更强上下文感受野，从而提升复杂背景中的目标判别能力。

## 2.3　双分支交互融合模块

标准U-Net通常通过直接拼接实现解码特征与跳跃特征融合，但这种固定融合方式容易将浅层背景噪声直接引入解码端，同时也难以充分利用高层语义特征与低层细节特征之间的互补关系。为此，本文提出DIF模块，如图2所示，对上采样解码特征$\boldsymbol{U}_i$与同尺度跳跃特征$\boldsymbol{S}_i$进行双分支交互建模，分别从通道和空间两个角度增强跨层融合效果。

在通道交互分支中，首先对解码特征和跳跃特征进行线性映射，并利用二者之间的相关性生成权重矩阵，用于引导跳跃特征的重标定。其过程可表示为：

$$A_c = Softmax\left(\frac{((\boldsymbol{W}_q\boldsymbol{U}_i)^\top|\boldsymbol{W}_k\boldsymbol{S}_i)}{\tau}\right) \#(6)$$

$$\boldsymbol{g}_c = \sigma(MLP(A_c(\boldsymbol{W}_v\boldsymbol{S}_i)^\top)) \#(7)$$

$$\boldsymbol{Z}_c = \boldsymbol{S}_i \odot (1 + \boldsymbol{g}_c) \#(8)$$

式中，$A_c$表示由解码特征$\boldsymbol{U}_i$与跳跃特征$\boldsymbol{S}_i$建立的跨分支相关性矩阵。$\boldsymbol{W}_q$，$\boldsymbol{W}_k$和$\boldsymbol{W}_v$为可学习映射参数，$\tau$为温度系数，用于调节相关性分布的平滑程度。$MLP(\cdot)$表示轻量通道映射函数，$\sigma(\cdot)$表示Sigmoid激活，$\odot$表示逐元素乘法。该分支通过解码特征对跳跃特征进行引导，有助于突出目标相关通道并抑制噪声响应。

在空间交互分支中，首先对跳跃和解码特征进行映射与归一化处理，生成空间门控图，并用于增强解码特征，其过程可写为：

$$\boldsymbol{g}_s = \sigma(LN((\boldsymbol{W}_q\boldsymbol{S}_i)^\top(\boldsymbol{W}_k\boldsymbol{U}_i))) \#(9)$$

$$\boldsymbol{Z}_s = \boldsymbol{U}_i \odot (1 + \boldsymbol{g}_s) \#(10)$$

式中，$LN(\cdot)$表示层归一化操作。在获得两条分支输

© 中国图象图形学报版权所有

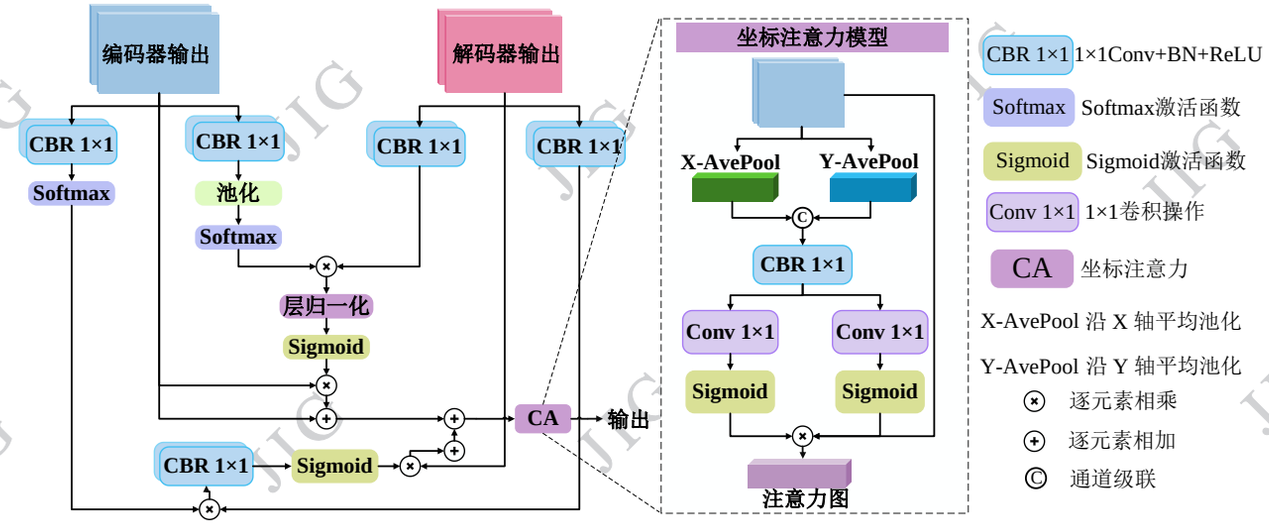



图2　双分支交互融合模块结构

Fig. 2　Dual-branch interactive fusion module structure

出后，采用可学习系数对其进行自适应融合，并结合坐标注意力进一步强化方向位置敏感建模，融合结果 $\boldsymbol{F}_i$ 表示为：

$$\boldsymbol{F}_i = Conv\left(CA\left(\alpha_i \boldsymbol{Z}_c + \beta_i \boldsymbol{Z}_s\right)\right) \#(11)$$

式中，$\alpha_i$ 和 $\beta_i$ 为可学习标量，$CA(\cdot)$ 表示坐标注意力模型。

通过上述双分支交互，解码特征中的高层语义信息能够对跳跃特征进行通道筛选，减少无关背景信息的跨层传递；与此同时，跳跃特征中保留的细粒度空间信息能够反向约束解码响应，补偿深层特征中的定位信息损失。两种交互分别侧重语义判别与空间定位，并在融合过程中形成互补，从而使 DIF 在缓解跨层语义差异的同时增强弱小目标表征并抑制背景噪声。

## 2.4　分组跨空间交互注意力模块

为进一步抑制结构化杂波并突出目标区域，本文设计 GCIA 模块，对跳跃特征和融合特征进行轻量重标定。如图 3 所示，不同于仅从通道或空间单一维度进行建模的注意力方式，GCIA 首先将输入特征按通道划分为 $G$ 组，并在组内分别建模，从而在降低计算开销的同时增强局部统计稳定性。设输入

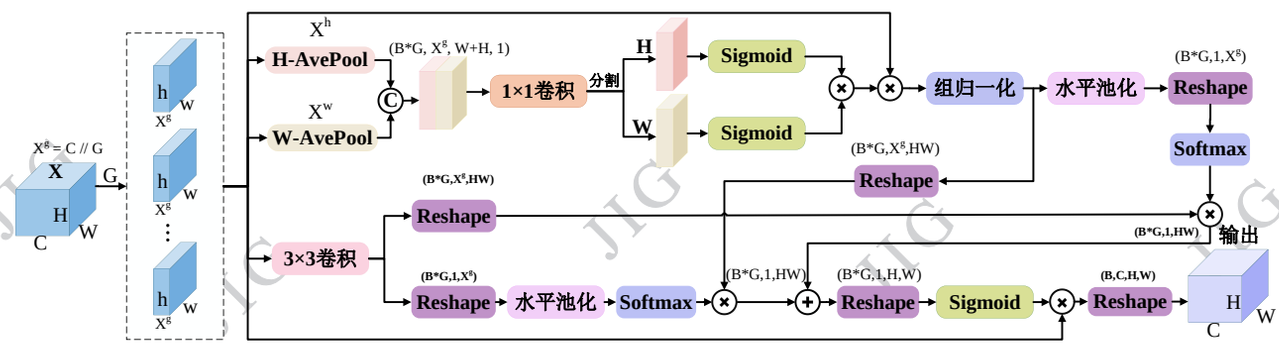



图3　分组跨空间交互注意力模块结构

Fig. 3　Grouped cross-spatial interactive attention module structure

特征为 $\boldsymbol{X}$，分组后的第 $g$ 组特征记为 $\boldsymbol{X}^g$。

在此基础上，GCIA 并行构建方向感知分支和组内交互分支。对于方向感知分支，首先沿高度与宽度两个方向对 $\boldsymbol{X}^g$ 进行池化，并通过共享映射生成方向感知权重：

$$\boldsymbol{M}_{\text{dir}} = \sigma\left(\psi\left[P_{\text{h}}\left(\boldsymbol{X}^g\right), P_{\text{w}}\left(\boldsymbol{X}^g\right)\right]\right) \#(12)$$

式中，$P_{\text{h}}(\cdot)$ 和 $P_{\text{w}}(\cdot)$ 分别表示沿高度和宽度方向的平均池化，$[\cdot]$ 表示拼接操作，$\psi(\cdot)$ 表示卷积映射。该分支能够捕获条纹、云带等背景的方向性统计特征，

© 中国图象图形学报版权所有

从而为后续抑噪提供基础。

同时，为增强组内显著区域建模，GCIA进一步利用局部卷积与全局聚合构建交互权重$\boldsymbol{M}_{\text{int}}$，并与方向权重共同作用于输入特征。其输出可写为

$$\hat{\boldsymbol{X}} = Re^{-1}\left(\boldsymbol{X}^{g}\odot\sigma\left(\boldsymbol{M}_{\text{dir}} + \boldsymbol{M}_{\text{int}}\right)\right)\#(13)$$

式中，$Re^{-1}(\cdot)$表示将分组特征恢复为原始特征维度。组内交互关系生成的显著性权重。通过方向感知分支与组内交互分支的联合建模，GCIA能够在保持较低计算复杂度的同时有效抑制结构化背景响应，并增强红外小目标的显著性表示。

### 2.5 损失函数

为同时兼顾区域重叠质量与像素级分类精度，本文采用Soft-IoU（soft intersection over union loss，Soft-IoU loss）损失（Rahman等，2016）与二元交叉熵（binary cross-entropy，BCE）损失（$\mathcal{L}_{\text{B}}$）（Guo等，2025）联合监督，并对多尺度侧输出实施深监督。设预测概率图为$\boldsymbol{P}$，真值掩码为$\boldsymbol{Y}$，则Soft-IoU损失定义为

$$\mathcal{L}_{\text{S}} = 1 - \frac{\sum \boldsymbol{P}\odot\boldsymbol{Y} + \varepsilon}{\sum \boldsymbol{P} + \sum \boldsymbol{Y} - \sum \boldsymbol{P}\odot\boldsymbol{Y} + \varepsilon}\#(14)$$

式中，$\varepsilon$为平滑项。进一步地，记第$k$个尺度输出对应的损失权重为$\lambda_k$，则总损失为

$$L = \sum_{k=0}^{K}\lambda_k\left(\mathcal{L}_{\text{B}}^{(k)} + \gamma\mathcal{L}_{\text{S}}^{(k)}\right)\#(15)$$

式中，$k=0$表示最终输出，$\gamma$为平衡系数。该联合监督方式能够在训练早期为浅层特征提供更直接的梯度信号，同时促进最终输出在目标完整性与背景抑制之间取得更好平衡。

## 3 实验结果与分析

### 3.1 数据集与实验设置

为验证DIGNet在不同场景下的有效性与泛化能力，本文在三个公开红外小目标检测数据集上进行实验：NUAA-SIRST（Nanjing University of Aeronautics and Astronautics—single-frame infrared small target）（Dai等，2021）、IRSTD-1K（infrared small target detection）（Zhang等，2022）和NUDT-SIRST（National University of Defense Technology）（Li等，2023）。三个数据集在目标尺度、对比度及背景复杂度方面具有明显差异，能够较全面地评估模型在弱小目标增强与复杂背景抑制方面的性能。其中，IRSTD-1K按4∶1划分训练集与测试集，NUAA-SIRST和NUDT-SIRST均按1∶1划分训练集与测试集。所有输入图像统一调整为512 × 512，并进行标准化处理。训练阶段采用随机翻转、随机缩放、随机裁剪和随机高斯模糊等增强策略，以提高模型对复杂背景和成像扰动的鲁棒性；测试阶段不使用随机增强，仅对图像进行尺寸调整后评估。优化器为Adan，初始学习率0.001，权重衰减为$1\times10^{-4}$；采用带warm-up的余弦退火策略，前10个epoch预热，随后按iteration级衰减。批次大小为4，训练800个epoch并使用混合精度。推理时以0.5为阈值二值化显著图。本文模型在256 × 256输入下含4.052 M参数，32位浮点数（32-bit floating point，FP32）和自动混合精度（automatic mixed precision，AMP）推理速度分别为36.0 FPS和37.6 FPS。

### 3.2 评估指标

为了全面评估检测性能，本文在实验中采用了三个广泛使用的指标（Zhang等，2026a）：交并比（intersection over union，IoU）、检测率（probability of detection，$P_d$）和虚警率（false alarm rate，$F_a$）。

### 3.3 定量评估

表1给出了DIGNet与现有主流方法在三个基准数据集上的定量比较结果，包括6种传统方法（Top-Hat（top-hat transformation）（Bai等，2010）、Max-Median（max-median filter）（Deshpande等，1999）、WSLCM（weighted strengthened local contrast measure）（Han等，2021）、TLLCM（tri-layer local contrast method）（Han等，2020）、IPI（infrared patch-image）（Gao等，2013）、NOLC（non-convex optimization with lp-norm constraint）（Zhang等，2019））与8种深度学习方法（ACM（Dai等，2021）、RDIAN（receptive-field and direction induced attention network）（Sun等，2023）、DNANet（Li。等，2023）、UIU-Net（Wu等，2023）、RPCANet（robust principal component analysis）（Wu等，2024）、MSHNet（multi-scale head）（Liu等，2024）、SCTransNet（Yuan等，2024）、L$^2$SKNet（learnable local saliency kernel network）（Wu等，2025））。总体来看，DIGNet在IoU、$P_d$和$F_a$等指标上均取得了最优表现，说明该方法能够在提升目标检测能力

的同时有效抑制背景误响应。与次优方法相比，DIGNet在NUAA-SIRST、IRSTD-1K和NUDT-SIRST三个数据集上的IoU分别提升1.75、1.77和

© 中国图象图形学报版权所有

表1　NUAA-SIRST、IRSTD-1K和NUDT-SIRST数据集上不同方法客观指标

Table 1　**The objective metrics for different methods on the NUAA-SIRST, IRSTD-1K, and NUDT-SIRST datasets**。

| Methods | NUAA-SIRST | | | IRSTD-1K | | | NUDT-SIRST | | |
|---|---|---|---|---|---|---|---|---|---|
| | IoU↑ | $P_d$ ↑ | $F_a$ ↓ | IoU↑ | $P_d$ ↑ | $F_a$ ↓ | IoU↑ | $P_d$ ↑ | $F_a$ ↓ |
| Top-Hat | 7.14 | 79.84 | 1012 | 10.06 | 75.11 | 1432 | 20.72 | 78.41 | 166.7 |
| Max-Median | 4.17 | 69.20 | 55.53 | 7.00 | 65.21 | 59.73 | 4.20 | 58.41 | 36.89 |
| WSLCM | 1.16 | 77.95 | 5446 | 3.45 | 72.44 | 6619 | 2.28 | 56.82 | 1309 |
| TLLCM | 1.03 | 79.09 | 5899 | 3.31 | 77.39 | 6738 | 2.18 | 62.01 | 1608 |
| IPI | 25.67 | 84.63 | 16.67 | 27.92 | 81.37 | 16.18 | 17.76 | 74.49 | 41.23 |
| NOLC | 17.70 | 89.69 | 16.41 | 20.46 | 82.82 | 49.64 | 18.61 | 85.61 | 18.19 |
| ACM | 65.28 | 90.49 | 47.13 | 58.84 | 91.92 | 27.59 | 60.63 | 91.75 | 20.06 |
| RDIAN | 70.66 | 93.92 | 43.08 | 63.37 | 88.55 | 30.18 | 77.77 | 95.66 | 26.36 |
| DNANet | 76.34 | <u>95.82</u> | 15.23 | 66.50 | 91.58 | 18.31 | 91.96 | <u>98.94</u> | 9.28 |
| UIU-Net | <u>77.17</u> | 95.44 | <u>14.34</u> | 65.34 | 90.91 | 15.01 | 90.43 | <u>98.94</u> | 7.72 |
| RPCANet | 53.93 | 95.44 | 149.2 | 62.09 | 89.35 | 48.61 | 87.79 | 95.98 | 32.03 |
| MSHNet | 73.88 | <u>95.82</u> | 19.25 | 67.17 | 92.52 | <u>12.60</u> | 76.95 | 95.77 | 20.52 |
| SCTransNet | 75.46 | <u>95.82</u> | 18.80 | 66.54 | <u>93.27</u> | 13.68 | 93.01 | 98.73 | <u>4.80</u> |
| $L^2$SKNet | 72.90 | 93.92 | 40.34 | <u>67.38</u> | 91.58 | 14.78 | <u>93.52</u> | 97.57 | 5.29 |
| **DIGNet (Ours)** | **78.92** | **96.95** | **13.05** | **69.15** | **94.28** | **7.63** | **95.03** | **99.15** | **0.76** |

注：加粗字体表示最优结果，下划线表示次优结果。↑代表值越大越好，↓代表值越小越好

1.51个百分点，$F_a$分别降低$1.29\times10^{-6}$、$4.97\times10^{-6}$和$4.04\times10^{-6}$。这一结果表明，所设计的双分支交互融合与分组跨空间注意机制能够更充分地利用跨层互补信息，并在复杂背景下保持更稳定的检测性能。

图4展示了不同红外小目标检测方法在三个数据集上的ROC（receiver operating characteristic，ROC）曲线，进一步证明了所提出的DIGNet方法的优越性。可以看出，在大多数阈值设置下，DIGNet的ROC曲线始终高于其他对比方法，表明其在检测率和虚警率之间取得了更好的平衡。尤其是在低虚警区域，DIGNet保持了更高的准确率，凸显了其更强的区分真实目标和杂波干扰的能力。这一优势验证了所提出的GCIA和DIF模块能够有效地增强弱目标并抑制背景干扰，使模型能够在具有挑战性的红外场景下实现更可靠的检测。

### 3.4　噪声鲁棒性性能比较

为进一步评估DIGNet在噪声干扰条件下的鲁棒性，本文在NUAA-SIRST数据集上分别加入三种强度的高斯噪声，即$\sigma$=10、20和30，对应信噪比（signal-to-noise ratio，SNR）分别为5.35、3.69和2.76，并采用IoU和nIoU进行评价。

由表2可知，随着噪声强度增加，各方法性能均出现不同程度下降，但DIGNet在三种噪声条件下始终取得最优结果，其IoU分别达到78.97%、75.74%和73.61%，nIoU分别达到77.61%、73.66%和71.05%。尤其在$\sigma$=30的强噪声条件下，DIGNet的IoU和nIoU仍分别较次优方法提升4.84和2.48个百分点。从$\sigma$=10增大至30时，DIGNet的IoU和nIoU分别下降5.36和6.56个百分点，仍保持较稳定的检测性能，表明所提方法在较强噪声干扰下具有良好的鲁棒性。

定的检测性能，表明所提方法在较强噪声干扰下具有良好的鲁棒性。

### 3.5　定性评估

图5展示了不同红外小目标检测方法在NUAA-SIRST（图5(a)、(b)）、IRSTD-1K（图5(c)、(d)）和NUDT-SIRST（图5(e)、(f)）数据集上的定性对比。其中，红色、蓝色和黄色方框分别代表正确检测到的

© 中国图象图形学报版权所有

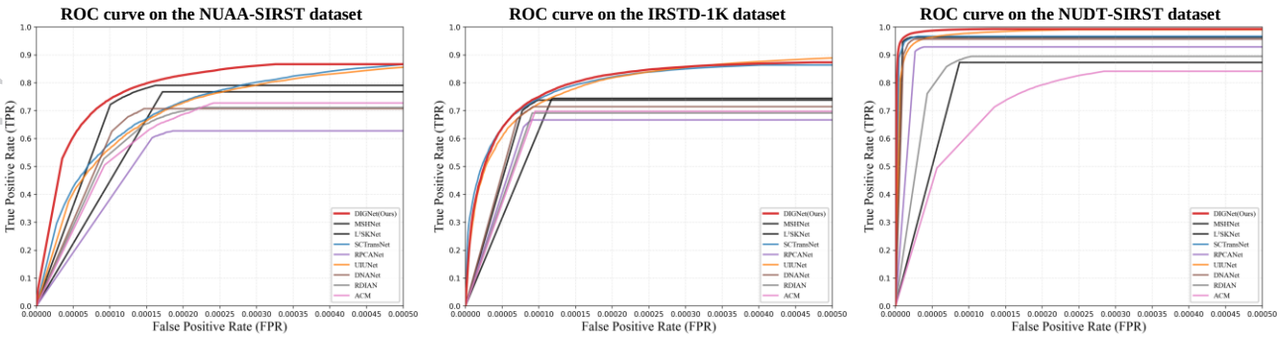



图4　三个公开数据集上不同方法的ROC曲线

Fig. 4　The ROC curve for different methods on the three datasets

**表2　不同高斯白噪声强度下各个方法的性能比较**

**Table 2　Performance comparison of various methods on NUAA-SIRST dataset under different Gaussian white noise levels**

| 方法 | $\sigma$=10 SNR=5.35 | | $\sigma$=20 SNR=3.69 | | $\sigma$=30 SNR=2.76 | |
|---|---|---|---|---|---|---|
| | IoU | nIoU | IoU | nIoU | IoU | nIoU |
| ACM | 67.36 | 68.01 | 68.55 | 64.59 | 62.18 | 62.11 |
| RDIAN | 70.91 | 75.71 | 69.27 | 70.42 | 65.99 | 66.41 |
| DNANet | 77.01 | 74.27 | 70.56 | 68.77 | 68.59 | 62.89 |
| RPCANet | 65.44 | 67.83 | 50.58 | 55.11 | 44.46 | 48.81 |
| MSHNet | 73.54 | 71.21 | 70.91 | 68.40 | 68.77 | 64.12 |
| SCTransNet | 73.61 | 74.81 | 70.81 | 71.31 | 68.51 | 68.57 |
| L²SKNet | 72.29 | 72.73 | 71.06 | 71.12 | 64.14 | 64.19 |
| **DIGNet** | **78.97** | **77.61** | **75.74** | **73.66** | **73.61** | **71.05** |

注：加粗字体表示最优结果，下划线表示次优结果。$\sigma$表示噪声强度，SNR表示信噪比。

目标、漏检的目标和误检。总体来看，传统方法在复杂背景下容易产生较多伪响应，而基于深度学习的方法能够更有效地区分真实目标与背景干扰，误检率明显降低。相比现有方法，DIGNet在目标召回和形状重建方面表现更优。例如，在图5(a)中，尽管MSHNet、SCTransNet和L²SKNet等方法仍存在误检，而DIGNet能够准确检测目标，且检测结果在形态上与真实标签更为接近，说明其在目标保持和分割精度方面具有更好的效果。

此外，图6中的3D响应表面可视化结果进一步验证了不同方法之间的性能差异。在复杂背景条件下，部分对比方法虽然能够在目标位置形成响应峰值，但仍存在背景伪峰残留、峰体分散或响应偏移等问题。相比之下，DIGNet在真实目标位置形成了更尖锐、集中的单峰响应，同时对非目标区域表现出更强的抑制能力，其整体响应分布与真实图像更为接近。这表明DIGNet能够在增强弱小目标显著性的同时，更有效地抑制复杂背景干扰，从而具备更好的检测鲁棒性和定位精度。

### 3.6　消融实验

#### 3.6.1　DIGNet中核心组件的有效性

为验证DIF和GCIA两个关键模块的有效性，本文在NUAA-SIRST、IRSTD-1K和NUDT-SIRST三个数据集上开展了消融实验，分别构建基础模型(UNet)、仅引入DIF、仅引入GCIA以及完整DIGNet四组实验。由表3可知，单独引入DIF或GCIA后，模型在三个数据集上的IoU均得到提升，说明两个模块均能有效改善小目标特征表征。其中，DIF更侧重于加强跨层特征交互与目标定位，GCIA则在抑制背景干扰和降低虚警方面表现出明显优势，但单独使用时在检测率与虚警率之间存在一定权衡。

当DIF和GCIA同时引入时，完整DIGNet获得了最优的综合性能。与基础模型相比，其在NUAA-SIRST、IRSTD-1K和NUDT-SIRST上的IoU分别提升2.74、2.00和5.91个百分点，$F_a$分别由20.62、10.78和4.73降低至13.05、7.63和0.76($\times10^{-6}$)。值得注意的是，在IRSTD-1K上，完整模型的$F_a$为7.63$\times10^{-6}$，略高于仅使用GCIA时的7.61$\times10^{-6}$，但与此同时，IoU由68.42%提升至69.15%，$P_d$由93.28%提升至94.28%。这一现象反映了目标增强与背景抑制之间的轻微权衡：GCIA更侧重于抑制结构化背景响应，而DIF通过跨层双向交互增强弱目标及其细节信息，在提升目标检测敏感性的同时，也可能保留少量与目标响应相近的背景信息。因此，联合使用DIF和GCIA虽然未在IRSTD-1K上

© 中国图象图形学报版权所有

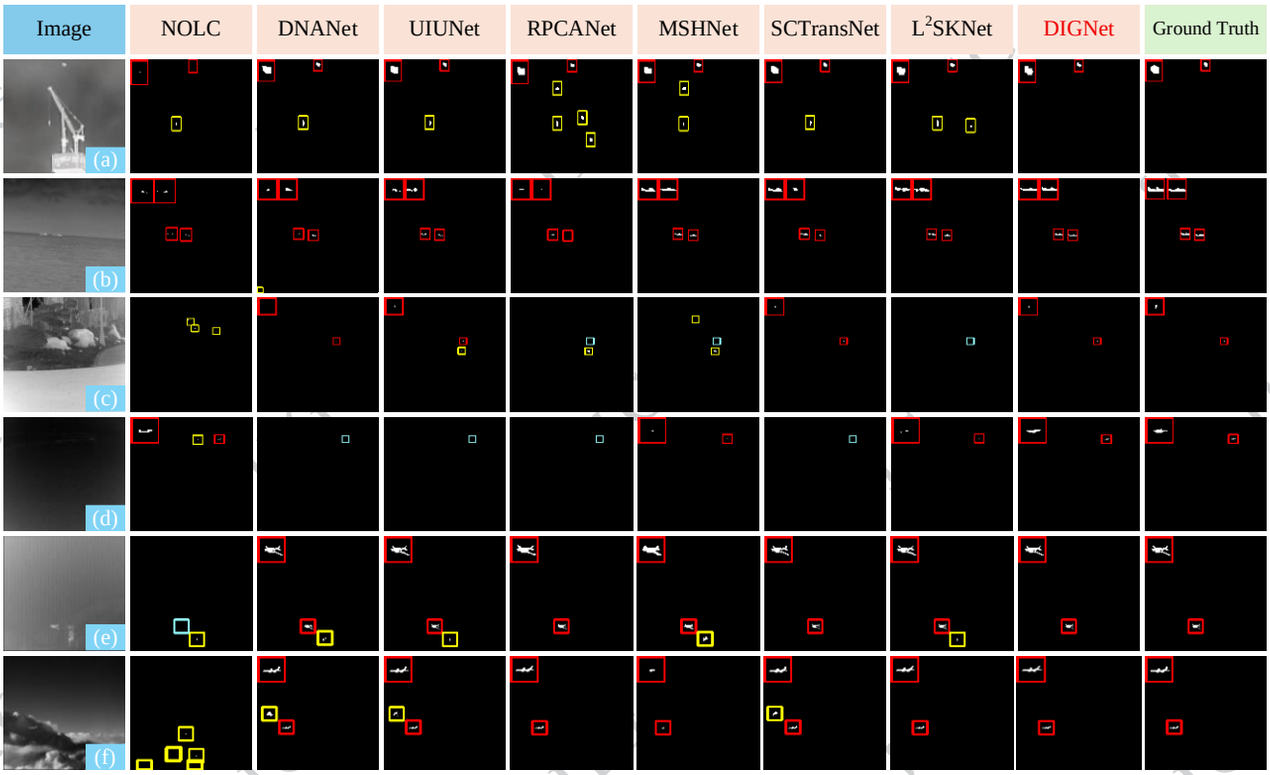



图5　在三个数据集上不同检测方法的可视化结果

Fig. 5　Visualization results of different detection methods on the three datasets




图6　不同检测方法的3D可视化结果

Fig. 6　Visualization results of different detection methods

取得绝对最低的$F_a$，但在保持近似虚警水平的同时获得了更高的区域重叠精度和检测概率。

3.6.2　U-Net主干模块特征提取的有效性

为验证RUB及其空洞扩展变体RUBD的有效性，本文在NUAA-SIRST数据集上开展主干模块整体替换实验。在保持网络整体架构、特征通道数、DIF、GCIA及多尺度深监督策略不变的条件下，采用Conv Block和DConv Block分别替换RUB和RUBD，构建对比模型。

由表4可知，采用RUB/RUBD主干模块组合后，模型的IoU由77.95%提升至78.92%，提高0.97个百分点；$P_d$由95.04%提升至96.95%，提高1.91个百分点；$F_a$由21.48×10$^{-6}$降低至13.05×10$^{-6}$，降低8.43×10$^{-6}$。结果表明，标准卷积块主要侧重局部特征提取，对不同尺度上下文信息的建模能力相对有限；单纯采用空洞卷积虽然能够扩大感受野，但难以

© 中国图象图形学报版权所有

表3　定量消融实验评估DIGNet中核心组件DIF和GCIA的影响

Table 3　Quantitative ablation experiments evaluate the impact of the core components DIF and GCIA in DIGNet

| 模型 | | NUAA-SIRST | | | IRSTD-1K | | | NUDT-SIRST | | |
|---|---|---|---|---|---|---|---|---|---|---|
| DIF | GCIA | IoU↑ | $P_d$ ↑ | $F_a$ ↓ | IoU↑ | $P_d$ ↑ | $F_a$ ↓ | IoU↑ | $P_d$ ↑ | $F_a$ ↓ |
| | | 76.18 | 94.28 | 20.62 | 67.15 | 93.94 | 10.78 | 89.12 | 98.08 | 4.73 |
| √ | | 77.41 | 96.94 | 16.31 | 67.62 | 91.92 | 9.00 | 94.30 | 98.40 | 2.88 |
| | √ | 77.83 | 96.18 | 22.07 | 68.42 | 93.28 | **7.61** | 92.40 | 98.61 | 2.31 |
| √ | √ | **78.92** | **96.95** | **13.05** | **69.15** | **94.28** | 7.63 | **95.03** | **99.15** | **0.76** |

注：加粗字体表示最优结果，下划线表示次优结果。√表示含有该模块，×表示不含有该模块。↑代表值越大越好，↓代表值越小越好。

表4　不同主干特征提取模块的性能对比

Table 4　Performance comparison of different backbone feature extraction modules

| 模型 | 基础模块 | 空洞扩展模块 | IoU↑ | $P_d$ ↑ | $F_a$ ↓ |
|---|---|---|---|---|---|
| 对比模型 | Conv Block | DConv Block | 77.95 | 95.04 | 21.48 |
| 本文模型 | RUB | RUBD | **78.92** | **96.95** | **13.05** |

注：加粗字体表示最优结果。↑代表值越大越好，↓代表值越小越好。

表5　GCIA定量消融实验在GCIA中分组数的影响

Table 5　Quantitative ablation experiments on the number of groups in GCIA

| 分组数G | IoU ↑ | $P_d$ ↑ | $F_a$ ↓ |
|---|---|---|---|
| 16 | 79.78 | 95.57 | 16.79 |
| 32 | 78.92 | **96.95** | **13.05** |
| 64 | **80.66** | 95.80 | 17.31 |

注：加粗字体表示最优结果，下划线表示次优结果。↑代表值越大越好，↓代表值越小越好。

充分兼顾局部细节保持与多尺度特征聚合。相比之下，RUB/RUBD主干模块组合通过块内U形编码—解码路径、残差连接及空洞卷积协同建模局部细节与多尺度上下文信息，并在保持空间分辨率的同时扩大有效感受野，从而增强网络对弱小目标的表征能力，有效抑制复杂背景中的干扰响应。

3.6.3　GCIA模块中分组数的影响

为分析GCIA中分组数G对检测性能的影响，本文在NUAA-SIRST数据集上将G分别设置为16、32和64，其余实验设置保持不变，结果如表5所示。不同分组数在区域重叠精度、检测概率和虚警抑制方面表现出不同侧重。当G=64时，模型取得最高的IoU，但$P_d$较低，且$F_a$最高；当G=16时，IoU表现较好，但检测率和虚警抑制能力仍有限。相比之下，G=32虽未取得最高IoU，但获得了最高的$P_d$和最低的$F_a$，在目标检测与虚警抑制之间表现出更优的综合平衡。因此，本文将GCIA模块的分组数G设置为32。

3.6.4　DIF双向交互机制的有效性

为进一步验证DIF中双向交互设计的有效性，本文在NUAA-SIRST数据集上分别控制两个交互方向。其中，U→S表示利用解码特征U引导跳跃特征S进行通道筛选，S→U表示利用跳跃特征S约束解码特征U的空间响应。关闭某一分支时采用恒等映射，以避免直接丢失对应输入特征。四组实验保持公共融合结构及训练设置一致，仅改变两个交互分支的启用状态。

由表6可知，相较于无交互的G00，单独启用U→S或S→U均能提高IoU并降低$F_a$，说明两个方向

表6　DIF双向交互机制的有效性

Table 6　Effectiveness of the DIF bidirectional interaction mechanism

| 实验设置 | U→S | S→U | IoU↑ | $P_d$ ↑ | $F_a$ ↓ |
|---|---|---|---|---|---|
| G00 | | | 76.20 | **96.95** | 17.40 |
| G01 | √ | | 77.21 | 95.42 | 14.17 |
| G10 | | √ | 76.94 | 96.57 | 14.37 |
| G11 | √ | √ | **78.92** | **96.95** | **13.05** |

注：加粗字体表示最优结果，下划线表示次优结果。√表示含有该模块，×表示不含有该模块。→代表特征引导方向，↑代表值越大越好，↓代表值越小越好。

© 中国图象图形学报版权所有

均具有独立作用。其中，G01的IoU由76.20%提升至77.21%，$F_a$由17.40降至14.17；G10的IoU提升至76.94%，$F_a$降至14.37。当两个方向同时启用时，G11取得最佳综合性能，IoU达到78.92%，$P_d$为96.95%，$F_a$进一步降至13.05。相比两个单向设置，双向交互不仅进一步提高IoU和虚警抑制能力，同时避免了单向交互带来的$P_d$下降。结果表明，U→S的语义通道筛选与S→U的空间定位约束具有互补作用，验证了DIF采用双向交互机制的有效性与必要性。

3.6.5　跨数据集泛化能力分析

为评估DIGNet在不同数据分布下的泛化能力，本文分别以NUAA-SIRST、NUDT-SIRST和IRSTD-1K作为源数据集训练模型，并直接在其他数据集上测试，整个过程不采用微调或域适应。表7给出了不同跨数据集设置下的IoU、$P_d$及其相较于同域测试结果的下降幅度。

**表7　DIGNet跨数据集性能及下降幅度（IoU/$P_d$）**

**Table 7　Cross-dataset performance and performance drops of DIGNet (IoU/$P_d$)**

| Train/Test | NUAA-SIRST | NUDT-SIRST | IRSTD-1K |
|---|---|---|---|
| NUAA-SIRST | **78.92/96.95**<br>(0.00/0.00) | 55.77/83.17<br>(23.15/13.78) | 55.39/88.89<br>(23.53/8.06) |
| NUDT-SIRST | 57.85/88.97<br>(37.18/10.18) | **95.03/99.15**<br>(0.00/0.00) | 38.63/73.40<br>(56.40/25.75) |
| IRSTD-1K | 68.62/92.40<br>(0.53/1.88) | 41.67/69.10<br>(27.48/25.18) | **69.15/94.28**<br>(0.00/0.00) |

注：加粗字体表示最优结果。

由表7可知，模型在同域测试时均取得最佳性能，而跨数据集测试后指标出现不同程度下降，说明三个数据集在目标尺度、背景结构和成像分布方面存在明显差异。其中，以NUDT-SIRST为训练集并迁移至IRSTD-1K时性能下降最为显著，IoU和$P_d$分别下降60.40%和25.75%；相比之下，以IRSTD-1K为训练集并在NUAA-SIRST上测试时，两项指标仅下降0.53%和1.88%，表现出较好的迁移能力。此外，多数跨数据集设置下$P_d$的下降幅度小于IoU，表明域偏移对目标区域精确分割的影响通常大于对目标发现能力的影响。

进一步分析表明，不同源域—目标域组合的跨域性能差异，可能与目标尺度、目标—背景对比度以及背景结构复杂度等方面的分布偏移有关。例如，以NUDT-SIRST为训练集并迁移至IRSTD-1K时，IoU和$P_d$分别下降56.40和25.75个百分点；相比之下，以IRSTD-1K为训练集并迁移至NUAA-SIRST时，两项指标仅下降0.53和1.88个百分点。不同迁移设置下性能下降幅度的明显差异表明，源域与目标域之间的目标及背景分布变化会对模型泛化能力产生重要影响。后续将考虑采用多数据集联合训练、域泛化及跨域特征对齐等策略，降低模型对特定数据分布的依赖，进一步提升跨场景适应能力。

3.6.6　模型稳定性的统计评估

为评估DIGNet在不同随机初始化条件下的稳定性与鲁棒性，本文选取ACM和DNANet作为对比方法，在NUAA-SIRST数据集上采用不同随机种子进行8次独立实验，其余训练与评价设置保持一致。表8报告了各方法在IoU、$P_d$和$F_a$上的均值（Mean）、样本标准差（Std）、95%置信区间（95%CI）、中位数（Median）、四分位距（Interquartile Range，IQR）以及Shapiro-Wilk正态性检验结果。

由表8可知，DIGNet取得了最优的平均检测性能，其IoU和$P_d$分别为79.01%±0.86%和96.28%±0.53%，平均$F_a$为$(14.82\pm2.02)\times10^{-6}$。相比DNANet，DIGNet的IoU和$P_d$分别提高4.81%和2.08%，$F_a$降低$7.71\times10^{-6}$。同时，DIGNet在三个指标上的标准差均低于ACM和DNANet，且95%置信区间与四分位距较小，表明其在不同随机种子下的性能波动更低。

此外，DIGNet三个指标的Shapiro－Wilk检验结果均满足$p>0.05$，说明其多次运行结果未显著偏离正态分布，采用均值和标准差进行描述具有合理性。综合来看，DIGNet不仅具有更高的检测精度和更低的虚警率，而且在多次独立训练中表现出更好的稳定性，说明其性能优势并非由个别随机初始化造成。

### 3.7　模型复杂度分析

为综合评估检测性能与模型复杂度，本文将DIGNet与DNANet、UIU-Net、SCTransNet和L$^2$SKNet进行比较，结果如图7所示。DIGNet在三个数据集上的平均IoU达到81.03%，较其余方法提升2.69～3.41个百分点，取得最优检测性能。在模型复杂度方面，DIGNet的参数量为4.052M，浮点运算次数（floating-point operations，FLOPs）为44.054G。其参数量仅高于L$^2$SKNet，相较于DNANet、UIU-Net和

© 中国图象图形学报版权所有

表8　不同方法在8次独立运行下的性能与稳定性比较

Table 8　Performance and stability comparison of different methods over eight independent runs

| Methods | Metrics | IoU | $P_d$ | $F_a$ |
|---|---|---|---|---|
| ACM | Mean ± Std | 66.91 ± 1.36 | 92.11 ± 1.50 | 47.48 ± 17.46 |
| | 95% CI | [65.77, 68.04] | [90.85, 93.36] | [32.88, 62.08] |
| | Median [IQR] | 67.01 [2.12] | 92.21 [2.75] | 45.31 [24.51] |
| | p-value (Shapiro-Wilk) | 0.3647 (✓) | 0.2309 (✓) | 0.7839 (✓) |
| DNANet | Mean ± Std | 74.20 ± 2.08 | 94.20 ± 1.73 | 22.53 ± 9.44 |
| | 95% CI | [72.47, 75.93] | [92.75, 95.65] | [14.64, 30.42] |
| | Median [IQR] | 74.73 [1.13] | 94.29 [2.84] | 18.35 [13.55] |
| | p-value (Shapiro-Wilk) | 0.0324 (×) | 0.4196 (✓) | 0.1166 (✓) |
| DIGNet | Mean ± Std | **79.01　0.86** | **96.28　0.53** | **14.82　2.02** |
| | 95% CI | **[78.29, 79.72]** | **[95.83, 96.72]** | **[13.13, 16.51]** |
| | Median [IQR] | **78.95 [0.46]** | **96.18 [0.58]** | **14.54 [2.31]** |
| | p-value (Shapiro-Wilk) | 0.3429 (✓) | 0.5180 (✓) | 0.0704 (✓) |

注：加粗字体表示最优结果。√表示未拒绝正态性假设，×表示拒绝正态性假设。

SCTransNet 分别减少了 13.79%、91.98% 和 63.79%，体现出较为紧凑的参数规模。虽然 DIGNet 的 FLOPs 高于 DNANet、SCTransNet 和 L²SKNet，但仍较 UIU-Net 降低了 19.06%。这说明 DIGNet 通过增加一定的计算开销，有效增强了特征表征能力，从而获得了更高的检测精度。综上所述，DIGNet 在保持较小参数规模的同时取得了最高的平均 IoU，在检测精度与模型规模之间实现了较好的平衡。

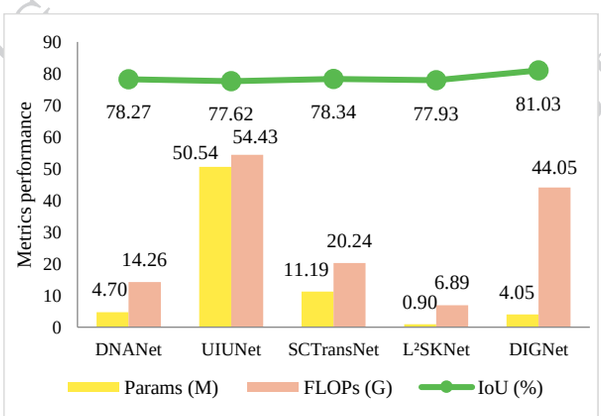



图 7　五种算法在三个指标上的性能评估

Fig. 7　Performance evaluation of five algorithms across three metrics

## 4　结束语

1）针对红外小目标尺度小、对比度低、特征易衰减以及复杂背景干扰易随跳跃连接传播等问题，本文提出了双分支交互引导网络 DIGNet。该网络采用 RUB/RUBD 提取局部细节与多尺度上下文信息，利用双分支交互融合模块 DIF 在通道和空间维度实现解码特征与跳跃特征的双向引导，并通过分组跨空间交互注意力模块 GCIA 增强目标响应、抑制结构化背景干扰，结合多尺度深监督完成目标重建。

2）在 NUAA-SIRST、IRSTD-1K 和 NUDT-SIRST 三个公开数据集上的实验结果表明，DIGNet 在检测精度、检测率和虚警抑制方面均取得了较好的综合性能。与传统红外小目标检测方法及多种主流深度学习方法相比，DIGNet 能够更准确地保持弱小目标的形态与位置，并有效减少复杂背景中的误检。

3）受数据条件和研究范围限制，本文主要基于三个公开数据集开展实验验证，对不同传感器、不同成像距离及极端复杂环境的覆盖仍不够充分。现有实验表明，DIGNet 具有较好的检测性能和一定的跨场景适应能力，但数据分布差异仍可能影响模型的检测效果。

4）本文尚未开展大规模多源异构数据联合训练

© 中国图象图形学报版权所有

以及实际红外探测设备上的部署验证。后续将进一步丰富不同场景和成像条件下的训练样本，研究更有效的跨场景特征学习方法，以提高模型对未知环境的适应能力。

## 参考文献(References)

Bai X Z and Zhou F G. 2010. Analysis of new top-hat transformation and the application for infrared dim small target detection. Pattern Recognition, 43(6): 2145-2156 [DOI: 10.1016/j. patcog. 2009. 12. 023]

Chen T X, Ye Z, Tan Z T, Gong T, Wu Y, Chu Q, et al. 2024. MiM-ISTD: Mamba-in-mamba for efficient infrared small-target detection. IEEE Transactions on Geoscience and Remote Sensing, 62: 1-13 [DOI: 10.1109/TGRS.2024.3485721]

Dai Y M, Wu Y Q, Zhou F, and Barnard K. 2021. Asymmetric contextual modulation for infrared small target detection//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: 949-958 [DOI: 10.1109/WACV48630. 2021. 00099]

Deshpande S D, Er M H, Venkateswarlu R and Chan P K. 1999. Max-mean and max-median filters for detection of small targets// Proceedings of the Signal and Data Processing of Small Targets. Bellingham, WA, USA: SPIE: 74-83 [DOI: 10.1117/12.364049]

Gu A, Goel K and Ré C. 2021. Efficiently modeling long sequences with structured state spaces [EB/OL]. [2025-07-21]. https://arxiv.org/pdf/2111.00396.pdf

Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2025-07-21]. https://arxiv. org/pdf/2312. 00752.pdf

Gao C Q, Meng D Y, Yang Y, Wang Y T, Zhou X F and Hauptmann A G. 2013. Infrared patch-image model for small target detection in a single image. IEEE Transactions on Image Processing, 22(12): 4996-5009. [DOI: 10.1109/TIP.2013.2281420]

Guo W, Cai F L, and Quan W Z. 2025. What is in the frequency: wavelet-guided semantic understanding for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-14. [DOI: 10.1109/TGRS.2025.3623373]

Ghiasi G, Lin T Y, and Le Q V. 2019. NAS-FPN: Learning scalable feature pyramid architecture for object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: 7036-7045 [DOI: 10. 1109/CVPR. 2019. 00720]

Han J, Ma Y, Zhou B, Fan F, Liang K, and Fang Y. 2014. A robust infrared small target detection algorithm based on human visual system. IEEE Geoscience and Remote Sensing Letters, 11(12): 2168-2172. [DOI: 10.1109/LGRS.2014.2323236]

Han J, Moradi S, Faramarzi I, Liu C, Zhang H, and Zhao Q. 2020. A local contrast method for infrared small-target detection utilizing a tri-layer window. IEEE Geoscience and Remote Sensing Letters, 17(10): 1822-1826 [DOI: 10.1109/LGRS. 2019. 2954578]

Han J H, Moradi S, Faramarzi I, Zhang H H, Zhao Q, Zhang X, et al. 2021. Infrared small target detection based on the weighted strengthened local contrast measure. IEEE Geoscience and Remote Sensing Letters, 18(9): 1670-1674 [DOI: 10. 1109/LGRS. 2020. 3004978]

Hou Q B, Zhou D Q, and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: 13713-13722. [DOI: 10.1109/CVPR46437.2021.01350]

Kou R K, Wang C P, Luo Y, Zhang Y, Xu Z L, Peng Z M, et al. 2024. Multiscale small-target detection techniques in single-frame infrared images: a review. Journal of Image and Graphics, 29(9): 2625-2649 (寇人可, 王春平, 罗迎, 张勇, 徐泽龙, 彭真明等. 2024. 单帧红外图像多尺度小目标检测技术综述. 中国图象图形学报, 29(9): 2625-2649) [DOI: 10.11834/jig.230788]

Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: 4015-4026 [DOI: 10. 1109/ICCV51070. 2023. 00371]

Lin T Y, Dollár P, Girshick R, He K, Hariharan B, and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: 2117-2125 [DOI: 10.1109/CVPR. 2017.106]

Li B Y, Xiao C, Wang L G, Wang Y Q, Lin Z P, Li M, et al. 2023. Dense nested attention network for infrared small target detection. IEEE Transactions on Image Processing, 32: 1745-1758 [DOI: 10. 1109/TIP. 2022. 3199107]

Liu S, Qi L, Qin H, Shi J, and Jia J. 2018. Path aggregation network for instance segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]

Liu Q, Liu R, Zheng B, Wang H, and Fu Y. 2024. Infrared small target detection with scale and location sensitivity//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: 17490-17499 [DOI: 10. 1109/CVPR52733. 2024. 01656]

Ronneberger O, Fischer P, and Brox T. 2015. U-Net: Convolutional networks for biomedical image segmentation // Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer: 234-241 [DOI: 10. 1007/978-3-319-24574-4_28]

Rahman M A and Wang Y. 2016. Optimizing intersection-over-union in deep neural networks for image segmentation//Proceedings of the 12th International Symposium on Visual Computing. Las Vegas, NV, USA: 234-244 [DOI: 10. 1007/978-3-319-50835-1_22]

© 中国图象图形学报版权所有

Sun Y, Yang J, and An W. 2021. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model. IEEE Transactions on Geoscience and Remote Sensing, 59(5):3737-3752. [DOI: 10.1109/TGRS.2020.3022069]

Sun H, Bai J X, Yang F, and Bai X. 2023. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-13 [DOI: 10.1109/TGRS.2023.3235150]

Tan M, Pang R, and Le Q V. 2020. EfficientDet: Scalable and efficient object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]

Wu X, Hong D, and Chanussot J. 2023. UIU-Net: U-Net in U-Net for infrared small object detection. IEEE Transactions on Image Processing, 32:364-376 [DOI: 10.1109/TIP.2022.3228497]

Woo S, Park J, Lee J, and Kweon I. 2018. CBAM: Convolutional block attention module//Proceedings of the European Conference on Computer Vision. Munich, Germany: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Wu F Y, Zhang T F, Li L, Huang Y, and Peng Z M. 2024. RPCANet: Deep unfolding RPCA based infrared small target detection//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: 4797-4806 [DOI: 10.1109/WACV57701.2024.00474]

Wu F, Liu A, Zhang T, Zhang L, Luo J, and Peng Z. 2025. Saliency at the helm: Steering infrared small target detection with learnable kernels. IEEE Transactions on Geoscience and Remote Sensing, 63:1-14 [DOI: 10.1109/TGRS.2024.3521947]

Xu M, Yu C, Li Z, Tang H, Hu Y, and Nie L. 2025. HDNet: A hybrid domain network with multiscale high-frequency information enhancement for infrared small-target detection. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-15 [DOI: 10.1109/TGRS.2025.3574962]

Yuan S, Qin H L, Yan X, Akhtar N, and Mian A. 2024. SCTransNet: spatial-channel cross transformer network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 62:1-15 [DOI: 10.1109/TGRS.2024.3383649]

Zhang Y M, Bao W T, Xiao Q, Yang Y, Wan W G, Luo Y T, et al. 2026a. Selective attention-based for infrared small target detection. Journal of Image and Graphics, 31(3):797-810. (张迎梅, 鲍王涛, 肖沁, 杨勇, 万伟国, 罗亦韬, 等. 2026. 基于选择性注意力的红外小目标检测. 中国图象图形学报, 31(3):797-810) [DOI: 10.11834/jig.250313]

Zhu H, Zhang J, and Xu G. 2020. Balanced ring top-hat transformation for infrared small-target detection with guided filter kernel. IEEE Transactions on Aerospace and Electronic Systems, 56(5):3892-3903. [DOI: 10.1109/TAES.2020.2982347]

Zhang Y M, Bao W T, Yang Y, Wan W G, Xiao Q and Zou X T. 2026b. MPCNet: Multi-scale perception and cross-attention feature fusion network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 64: 1-15. [DOI: 10.1109/TGRS.2026.3653023]

Zhang M J, Zhang R, Yang Y X, Bai H, Zhang J, and Guo J. 2022. ISNet: Shape matters for infrared small target detection// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: 877-886 [DOI: 10.1109/CVPR52688.2022.00095]

Zhang Y M, Bao W T, Yang Y, Wan W G, Xiao Q and Zou X T. 2025a. HAFNet: Hierarchical attention fusion network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 63:1-16. [DOI: 10.1109/TGRS.2025.3607732]

Zhang M J, Yue K, Li B Y, Guo J, Li Y S, and Gao X B. 2024a. Single-frame infrared small target detection via gaussian curvature inspired network. IEEE Transactions on Geoscience and Remote Sensing, 62:1-13 [DOI: 10.1109/TGRS.2024.3423492]

Zhang M J, Yue K, Guo J, Zhang Q M, Zhang J, and Gao X B. 2025b. Computational fluid dynamic network for infrared small target detection. IEEE Transactions on Neural Networks and Learning Systems, 36 (8): 14777-14789 [DOI: 10.1109/TNNLS.2025.3548984]

Zhang M J, Wang Y C, Guo J, Li Y S, Gao X B, and Zhang J. 2024b. IRSAM: Advancing segment anything model for infrared small target detection//Proceedings of the European Conference on Computer Vision. Milan, Italy: 233-249 [DOI: 10.1007/978-3-031-72855-6_14]

Zhang M J, Li X L, Gao F, and Guo J. 2025c. IRMamba: Pixel difference mamba with layer restoration for infrared small target detection//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, PA, USA, 39(9): 10003-10011 [DOI: 10.1609/aaai.v39i9.33085]

Zhang T, Wu H, Liu Y, Peng L, Yang C, and Peng Z. 2019. Infrared small target detection based on non-convex optimization with Lp-Norm constraint. Remote Sensing, 11(5): 559 [DOI: 10.3390/rs11050559]

## 作者简介

肖沁,女,讲师,硕士生导师,主要研究方向为计算机视觉、数据驱动决策和机器学习。 E-mail: xiaoqin@jxufe.edu.cn

张迎梅,通信作者,女,讲师,硕士生导师,主要研究方向为红外小目标检测、图像融合与计算机视觉。 E-mail: zhangyingmei@jxufe.edu.cn

张与明,男,本科生,主要研究方向为红外小目标检测与图像处理。E-mail: zym0921xx@163.com

杨勇,男,教授,博士生导师,主要研究方向为遥感图像处理、计算机视觉与机器学习。E-mail: yangyong@tiangong.edu.cn

万伟国,男,副教授,硕士生导师,主要研究方向为图像处理、

© 中国图象图形学报版权所有

人脸识别与计算机视觉。E-mail：wanweiguo@jxufe.edu.cn

© 中国图象图形学报版权所有