中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-15

论文引用格式:Qian Menghao, Liu Kui, Zhang Fengbo, Su Benyue. XXXX. Infrared small target detection with multi-branch perception and cross-layer semantic fusion. Journal of Image and Graphics, XX(XX):0001-0015(钱孟豪, 刘奎, 章丰博, 苏本跃. XXXX. 多分支感知与跨层语义融合的红外小目标检测. 中国图象图形学报, XX(XX):0001-0015)[DOI:10.11834/jig.250448]

多分支感知与跨层语义融合的红外小目标检测

钱孟豪[1],刘奎[1],章丰博[1],苏本跃[2]

1. 安庆师范大学计算机与信息学院,安庆　246000; 2. 铜陵学院数学与计算机学院,铜陵　244000

摘　要:**目的**　红外小目标检测在军事和民用等领域具有重要应用价值。然而,由于目标尺度极小且常处于复杂背景之中,如何有效提取边缘等判别性特征仍然是亟待解决的难题。同时,现有基于U-Net的检测网络在跨层特征融合过程中存在明显的语义差异,导致浅层细节信息与深层语义特征难以充分结合,从而进一步限制了检测精度的提升。**方法**　基于U-Net结构,提出一种多分支感知与跨层语义融合的红外小目标检测网络(multi-branch perception and cross-layer semantic fusion network,MPCF-Net)。在编码器阶段,为增强边缘等判别性特征的提取,引入了多分支感知融合注意力(multi-branch perception fusion attention module,MPFM)。该模块通过局部分支、全局分支及串行卷积分支实现多尺度特征提取,并结合局部-全局引导注意力(local-global guided attention,LGGA)与全局通道空间注意力(global channel spatial attention,GCSA),分别强化小目标的响应能力与特征表达能力。随后,为缓解跨层特征间的语义差异并建模上下文依赖关系,采用空间-通道交叉Transformer块(spatial-channel cross transformer block,SCTB)替代传统的跳跃连接,从而提升多层特征融合效果。在解码器阶段,虽然深度可分离卷积能够有效降低参数量和计算复杂度,但由于缺乏跨通道特征交互,削弱了小目标的细节特征。为此,在输出端引入轻量梯度门控模块(lightweight gradient gating,LGG),利用Sobel梯度引导的空间注意力进一步强化小目标的边缘与细节特征。**结果**　在SIRST、IRSTD和NUDT-SIRST三个公开红外小目标数据集上的实验表明,MPCF-Net在交并比(intersection over union,IoU)和归一化交并比(normalized intersection over union,nIoU)指标上分别达到80.12%、66.28%和84.26%,以及78.23%、64.58%和86.48%。同时,该方法在检测概率(probability of detection,Pd)上分别达到99.88%、94.23%和98.21%,虚警率(false alarm,Fa)仅为1.12×10^{-6}、4.39×10^{-6}和14.57×10^{-6},展现了更优的检测性能。**结论**　所提方法通过多分支感知和跨层语义融合,有效增强了红外小目标的边缘等判别特征提取能力及上下文建模能力,从而实现了更高精度的红外小目标检测。

关键词:红外小目标检测;多分支感知;跨层语义融合;注意力机制;Transformer

Infrared small target detection with multi-branch perception and cross-layer semantic fusion

Qian Menghao[1], Liu Kui[1], Zhang Fengbo[1], Su Benyue[2]

1. School of Computer and Information, Anqing Normal University, Anqing 246000, China; 2. School of Mathematics and Computer Science, Tongling University, Tongling 244000, China

收稿日期:2025-09-16;**修回日期**:2025-12-19

*通信作者**:刘奎　liukui@aqnu.edu.cn

基金项目:安徽省科技重大专项(201903a06020006);安徽省教育自然科学研究重点项目(KJ2017A353)

Supported by: Major Special Science and Technology Project of Anhui Province (201903a06020006); Key Project of the Educational Natural Science Research of Anhui Province (KJ2017A353)

Abstract: **Objective** Infrared small target detection plays a critical role in military surveillance, security monitoring, remote sensing, and unmanned aerial vehicle (UAV) operations. Accurate detection of such targets is essential for threat assessment, object tracking, and decision-making in complex operational scenarios. In real-world environments, infrared targets are often extremely small in size, have low contrast against complex backgrounds, and are easily affected by noise. These characteristics make it challenging to extract discriminative features that can effectively separate targets from surrounding clutter. Moreover, differences in semantic content across feature layers impede effective fusion of shallow detailed features with deep semantic features, further limiting the accuracy and robustness of detection. Designing efficient, accurate, and robust methods for infrared small target detection under complex backgrounds remains a critical research problem in the field of computer vision and defense-related applications. **Method** To address these challenges, we propose a novel multi-branch perception and cross-layer semantic fusion network (MPCF-Net) for infrared small target detection. The network is built on an encoder–decoder architecture designed to capture both local and global contextual information while mitigating computational cost. In the encoder, we introduce a multi-branch perception fusion attention module (MPFM) that simultaneously processes local, global, and serial convolutional branches to extract multi-scale features. The MPFM module is combined with local-global guided attention (LGGA), which integrates local spatial features and global semantic information to enhance target response, and global channel spatial attention (GCSA), which models channel dependencies and spatial relationships to improve feature representation quality. To effectively bridge semantic gaps between shallow and deep layers, a spatial-channel cross transformer block (SCTB) replaces conventional skip connections. The SCTB models cross-layer context dependencies, allowing the network to capture hierarchical feature correlations while alleviating semantic differences between layers. This module ensures that shallow detailed features and deep semantic features are adequately fused, improving the overall discriminative power of the network. In the decoder, depthwise separable convolution is employed to reduce computational complexity without sacrificing representational capability. Additionally, a light gradient gate (LGG) guided by Sobel gradients is introduced at the output stage to enhance small target edge details and reinforce feature localization precision. By combining attention mechanisms with multi-branch and cross-layer strategies, MPCF-Net is capable of robustly detecting infrared small targets in challenging and cluttered backgrounds. **Result** The effectiveness of MPCF-Net is evaluated on two public infrared small target datasets, SIRST、IRSTD and NUDT-SIRST, which contain diverse scenes with varying target sizes, shapes, and noise levels. Quantitative results demonstrate that MPCF-Net achieves intersection over union (IoU) scores of 80.12%, 66.28%, and 84.26%, and normalized intersection over union (nIoU) scores of 78.23%, 64.58%, and 86.48%, respectively, across the three datasets. In terms of detection performance, MPCF-Net attains a high probability of detection (Pd) of 99.88%, 94.23%, and 98.21%, while maintaining extremely low false alarm (Fa) rates of 1.12×10^{-6}, 4.39×10^{-6}, and 14.57×10^{-6}, respectively. In addition to quantitative performance, we analyze the effectiveness of individual components. Ablation studies reveal that the MPFM module significantly enhances multi-scale feature extraction, LGGA improves the network's sensitivity to target regions, and GCSA strengthens feature representation by emphasizing spatial-channel relationships. The SCTB module demonstrates clear advantages in reducing semantic conflicts between layers, leading to more coherent and accurate feature fusion. LGG ensures that edge details are preserved and small targets are precisely localized, which is crucial for applications where missed detection could result in operational failure. **Conclusion** MPCF-Net effectively combines multi-branch perception, attention mechanisms, and cross-layer semantic fusion to enhance discriminative feature extraction and hierarchical context modeling for infrared small targets. The network demonstrates high accuracy, robustness, and low false alarm rates in complex scenes. Its design provides a practical solution for real-world surveillance, military monitoring, and UAV-based remote sensing applications. Future work may explore adaptive integration with other sensor modalities and real-time deployment strategies to further enhance operational effectiveness.

Key words: Infrared small target detection; Multi-branch perception; Cross-layer semantic fusion; Attention mechanism; Transformer

0 引 言

红外小目标检测作为目标检测任务的重要分支，在军事侦察（李其昌 等，2016；李俊宏 等，2020）、边境安防（王云杰 等，2023）及工业监控（赵兴科 等，2021）等领域展现出显著的应用价值 。与依赖外部光照的可见光成像相比，红外成像基于目标的热辐射特性，即使在弱光、雾霾或夜间等复杂环境下，也能稳定获取清晰的图像信息，为目标检测任务提供了可靠支持。

尽管红外成像在复杂环境中具备一定优势，但在远距离成像条件下，目标的热辐射能量通常较弱，加之目标尺寸小、对比度低，且易受背景杂波干扰，导致红外图像中的小目标常呈现模糊、弱显性等特征，显著增加了检测的难度（寇人可 等，2024）。早期研究多依赖于传统图像处理技术，主要可分为三类方法：滤波器方法 （Zeng 等，2006）、人类视觉感知模型 （黄磊 等，2023；Gao 等，2013；Dai 等，2017；王志武 等，2022）以及低秩分解方法 （Zhang 等，2019；Zhu 等，2020）。这些方法在特定条件下虽能取得一定效果，但普遍存在对人工特征设计依赖强、泛化能力不足以及检测精度较低等问题，难以满足实际应用需求。

随着深度学习技术的迅速发展，端到端的学习框架逐渐成为红外小目标检测的研究主流 。借助深层神经网络强大的特征学习能力，小目标检测的精度与鲁棒性均得到了显著提升。例如，Wang 等人 （2017）基于ILSVRC预训练模型对红外小目标进行特征提取，有效增强了网络的表达能力。Lin 等人 （2018）通过引入过采样策略，提高了模型对弱小目标的感知灵敏度。Dai 等人在ACM （2021）的基础上进一步提出了ALCNet （2021），该模型利用局部对比度模块在高层网络中突显目标。

随后，越来越多的深度学习方法尝试将红外小目标检测转化为语义分割任务，以实现对目标的像素级精确定位。其中，U-Net（Ronneberger 等，2015）成为该类方法的基础框架之一。然而，传统 U-Net 虽然通过编码器与解码器的跳跃连接结合了不同层级的特征，但融合方式仅为简单拼接，缺乏对局部细节与全局上下文的有效融合。因此，如何高效融合局部与全局特征成为关键问题。针对这一问题，Li 等人（2022）提出 DNA-Net，通过密集嵌套的交互式模块，促进局部和全局特征之间的充分交互。Wu 等人（2022）提出了 UIU-Net 框架，在 U-Net 结构中嵌套子 U-Net，从而有效增强了全局与局部信息的融合。上述方法虽在信息融合方面取得了一定进展，但在复杂背景下仍存在细节特征提取不足的问题。为此，Wang 等人（2023）提出了 UCFNet，结合快速傅里叶卷积和中心差分卷积，更好地提取细节特征并高效融合全局与局部信息。随后，Xu 等人（2024）提出了 HCF-Net，通过多尺度特征提取与自适应通道融合，实现了细节特征的高效提取。尽管这些方法在特征融合能力上取得了显著进展，但在构建复杂特征交互结构的过程中，解码器往往难以充分利用编码器所提取的高层语义信息，导致语义特征在传递过程中出现流失。为缓解这一问题，研究开始关注编码器与解码器之间的语义差异建模，以提升语义信息的完整性与表达能力。例如，Wu 等人（2023）提出的 MTU-Net 通过多层级 Transformer 模块（multi-level vit module，MVTM）提取具有长距离依赖的多层特征，并将其融合到解码器中，从而进一步增强语义信息的保留与表达。进一步地，为弥补引导式学习利用不足的问题，Li 等人（2025）提出了 MMLNet，利用多层特征融合策略提升模型的泛化能力。然而，单一结合 Transformer 与 U-Net 的网络结构在全局语义建模方面仍存在局限，难以满足复杂场景下对精细语义的理解需求。此外，编码器中不同层级特征在语义抽象程度与空间细节表达上差异显著，缺乏对其上下文关联的有效建模，导致语义融合不足，从而限制了解码器对浅层细节与深层语义信息的充分利用。同时，在复杂背景干扰下，红外图像中小目标的判别特征仍难以充分提取，从而影响整体检测性能。针对上述问题，本文基于 U-Net 框架提出一种多分支感知与跨层语义融合的红外小目标检测网络。

此框架在网络编码器中引入了多分支感知融合注意力模块用于提升小目标的判别特征提取能力。MPFM 通过并行构建局部、全局和串行卷积分支，实现多尺度特征的高效提取。为增强特征表达效果，一方面在局部与全局分支中引入 LGGA，有效融合细节信息与全局语义，增强对目标区域的响应与识别能力；另一方面，在分支融合阶段引入 GCSA，建模通道与空间维度之间的长程依赖关系，进一步提

升语义融合效率与整体特征表达。为加强不同层级特征之间语义融合，引入了空间-通道交叉Transformer块。该模块融合空间嵌入的单头信道交叉注意力机制（spatial embedded single-head channel-cross attention，SSCA）与互补前馈网络（comple mentary feed-forward network，CFN），其中SSCA用于强化空间与通道信息的交互建模，CFN则实现多尺度语义特征的互补表达，两者协同作用，从而增强跨层特征的融合。

此外，本文在解码器中引入深度可分离卷积以显著降低计算开销，但在减少参数量与计算复杂度的同时，其因通道特征交互受限，往往削弱了对边缘等细节特征的捕捉能力。为此，解码器末端引入轻量梯度门控模块，该模块通过Sobel梯度（Sobel等，1968）引导的空间注意力增强边缘特征，并结合简化通道缩放机制抑制无效响应，从而在几乎不增加计算开销的情况下有效保留目标细节。

1 方 法

1.1 总图框架

本节将详细介绍MPCF-Net的整体结构，其框架如图1所示。MPCF-Net以红外图像为输入，在编码阶段引入MPFM模块和最大池化操作，通过多分支结构提取多尺度特征。随后，这些特征被输入SCTB模块，实现深浅层跨层语义融合，并将融合特征传递至解码器。解码阶段结合深度可分离卷积与转置卷积，在解码器末端引入LGG模块以增强目标细节，并通过多尺度辅助监督有效提升训练性能。

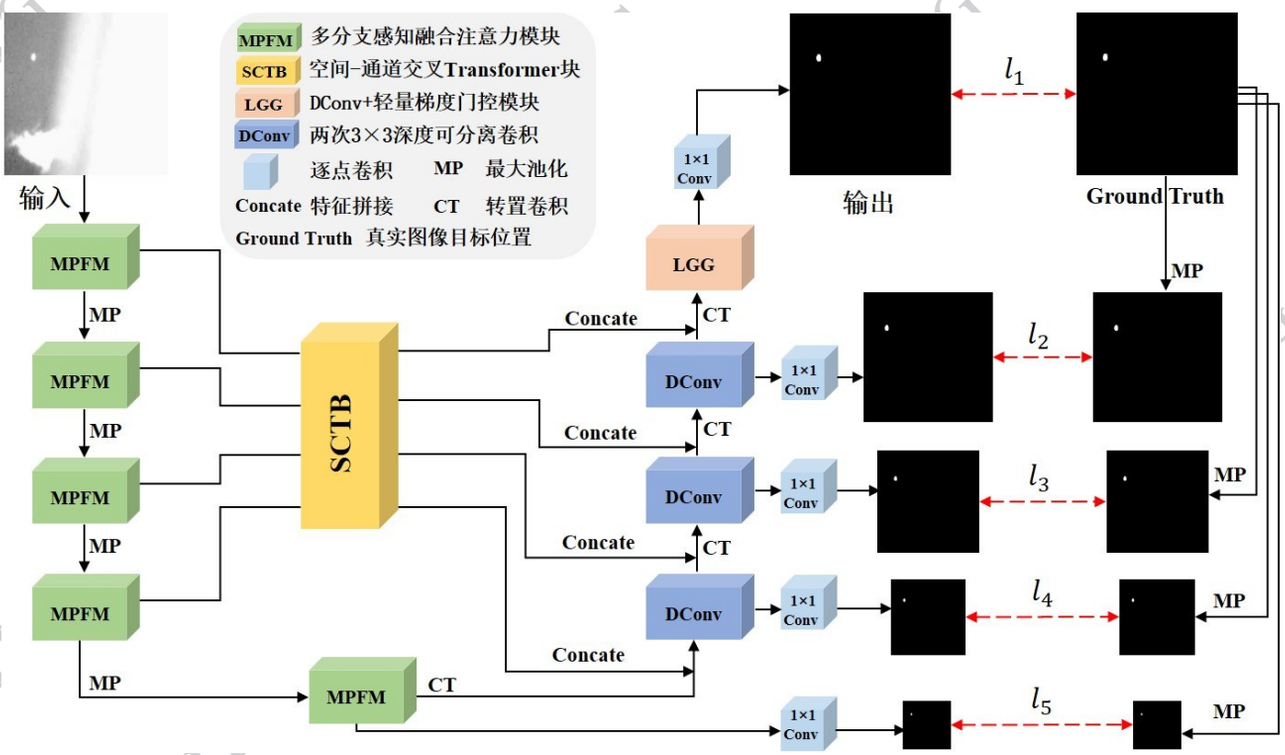


图1 MPCF-Net网络结构

Fig. 1 The network architecture of MPCF-Net

1.2 多分支感知融合注意力模块

红外图像中的小目标通常具有尺寸小、对比度低等特征。在复杂背景（如地面纹理、云层结构和热噪声）中，目标与背景在灰度、纹理和边缘上差异不明显，限制了判别特征提取。

因此，如图2所示，本文引入MPFM，通过并行局部分支、全局分支和串行卷积分支，有效提取多尺度特征。具体而言，先将输入特征$F \in \mathbf{R}^{C\times H'\times W'}$转置成$F' \in \mathbf{R}^{H'\times W'\times C}$后，在通过逐点卷积对其进行维度调整，得到$F'' \in \mathbf{R}^{H'\times W'\times C'}$。然后将$F''$分别输入三个并行分支计算，分别得到局部特征$F_l \in \mathbf{R}^{H'\times W'\times C'}$和全局特征$F_g \in \mathbf{R}^{H'\times W'\times C'}$、串行卷积特征

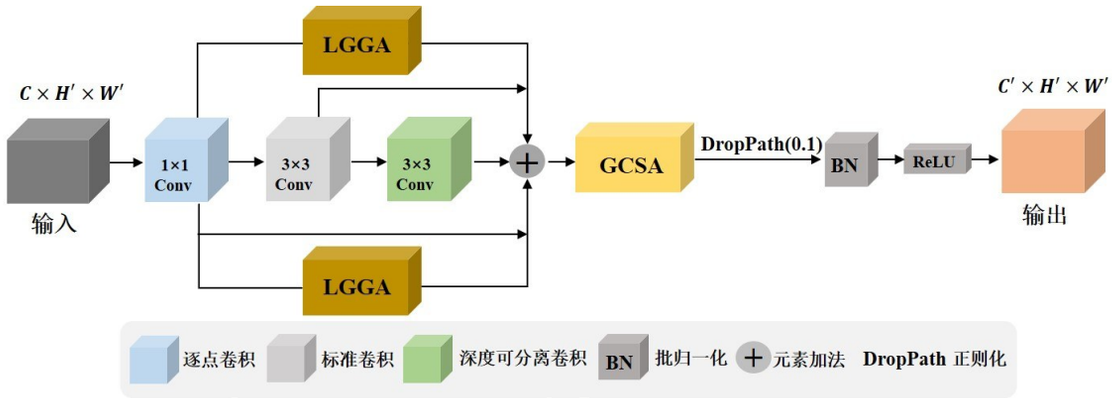


图2　多分支感知融合注意力模块

Fig. 2　Multi-Branch Perception Fusion Attention Module

$F_c \in \mathbf{R}^{H' \times W' \times C'}$。最终，将三个分支的输出结果相加，得到融合特征 $F_f \in \mathbf{R}^{H' \times W' \times C'}$。最后将 F_f 经过 GCSA 注意力等操作得到输出特征 $F''' \in \mathbf{R}^{H' \times W' \times C'}$。

1.2.1　局部-全局引导注意力

在局部分支和全局分支中，引入 LGGA 模块融合局部信息与全局语义，以增强目标响应并抑制背景干扰。针对不同层级特征，LGGA 对输入特征块设置不同尺寸(p)：浅层取较小值(2、4)捕捉细节，深层取较大值(4、8)加强上下文和语义对比。

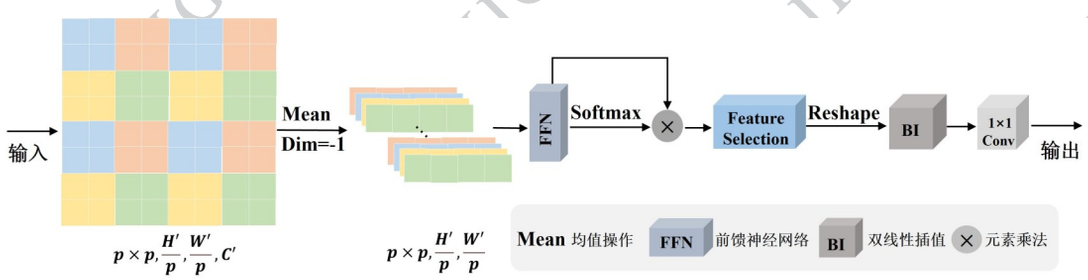


图3　局部-全局引导注意力

Fig. 3　Local-Global Guided Attention

具体流程如图3，首先，将 F' 划分为一组在空间维度上连续的非重叠块(尺寸 p) $F_p \in \mathbf{R}^{\frac{H'}{p} \times \frac{W'}{p} \times C'}$。随后，对每个斑块进行信道平均，得到表示向量 $\boldsymbol{F}_m$，并引入前馈网络(Ashish 等，2017)进行线性变换。接下来，应用激活函数来获取线性计算出的特征在空间维度上的概率分布，并相应地调整其权重得到加权后的结果 $F_d \in \mathbf{R}^{\frac{H'}{p} \times \frac{W'}{p}}$。公式如下：

$$F_m = \frac{1}{C}\sum_{c=1}^{C} F_p, F_d = \delta(\boldsymbol{F}_m) \odot \sigma(FN(\boldsymbol{F}_m)) \tag{1}$$

式中，C 表示通道数，δ 表示前馈网络，由两层多层感知机(MLP)和一层归一化层组成，σ 表示采用 Softmax 激活函数，$\odot$ 表示元素级相乘。

在加权特征处理后，采用特征选择方法(Shi 等，2023)，从维度中筛选与当前任务最相关的特征。具体的，使用可学习的全局引导向量 $\boldsymbol{P}$，对每个局部特征进行余弦相似度计算。随后，通过对相似度进行阈值裁剪生成掩码与加权后的特征元素级相乘，得到输出 F'_d。随后，通过一个可学习的线性变换矩阵进行引导得到 F_T。并通过重塑将其尺寸恢复，最终经过插值与卷积等操作，生成局部和全局特征 F_l 与 F_g 公式如下：

$$Sim = \cos(F_d, \boldsymbol{P}) = \frac{F_d \cdot \boldsymbol{P}}{\|F_d\| \cdot \|\boldsymbol{P}\|} \tag{2}$$

$$Mask = Clamp(sim, 0, 1) \tag{3}$$

$$F'_d = F_d \odot Mask, F_T = F'_d \cdot \boldsymbol{T} \tag{4}$$

$$F_l = Conv_{1\times1}(BI(RE(F_T))) \tag{5}$$

式中 $\boldsymbol{P}$ 为可学习的引导向量，Sim 为余弦相似度函数，其取值范围限定于 [0，1] 之间，clamp 为阈值裁剪，$\boldsymbol{T}$ 为特定 learnable 线性变换矩阵，BI 为双线性插值操作，RE 为重塑操作，$Conv_{1\times1}$ 表示 1×1 卷积操作。

1.2.2　全局通道空间注意力

在并行分支完成特征提取后，引入 GCSA 以增强特征表达。该模块先建模通道依赖以捕捉关键信息，再通过通道洗牌促进跨通道交互，并结合大感受野卷积有效获取空间结构，实现局部与全局语义的统一建模。

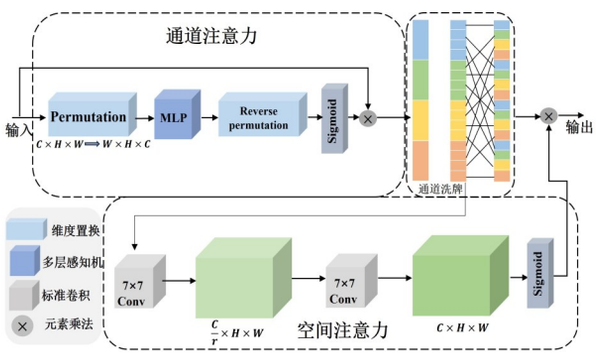


图4　全局通道空间注意力

Fig. 4　Global Channel Spatial Attention

具体而言，在通道注意力中，输入特征首先进行维度置换操作，从 $F_f \in \mathbf{R}^{H' \times W' \times C'}$ 变换到 $F_f' \in \mathbf{R}^{W' \times H' \times C'}$。接着，通过 MLP 对通道间依赖关系进行捕捉。之后，进行逆维度置换以还原特征维度，并通过 Sigmoid 激活函数生成通道特征。最后，将该特征与输入特征逐元素相乘，得到增强后的特征 F_c。随后进行通道洗牌，先将增强后的特征 F_c 划分为 4 组，每组通道大小为原来的四分之一。随后对分组特征进行转置以打乱组内通道排列，最后将其重新整合恢复至原始通道维度。

在空间注意力中，输入特征先经 7×7 卷积压缩至四分之一通道，再经另一个 7×7 卷积恢复至原始通道数。最后经 Sigmoid 激活生成空间特征，并与洗牌后的特征逐元素相乘得到最终输出特征 F_s。上述流程公式如下：

$$F_c = \sigma\left(MLP\left(F_f'\right)\right) \odot F_f', F_c' = \mathrm{CS}\left(F_c\right) \tag{6}$$

$$F_s = \sigma(\mathrm{Conv}_{7\times7}(BN(\upsilon(\mathrm{Conv}_{7\times7}(F_c'))))) \odot F_c' \tag{7}$$

式中，σ 表示 Sigmoid 激活函数，CS 表示通道洗牌操作，$\mathrm{Conv}_{7\times7}$ 表示 7×7 卷积操作，υ 表示 ReLU 激活函数，BN 表示批量归一化。

1.3　空间–通道交叉 Transformer 块

尽管多尺度并行结构在一定程度上提升了特征融合能力，但在编码阶段，浅层特征主要包含局部边缘、纹理及空间细节等低级信息，而深层特征则侧重于全局结构、语义上下文等高级抽象表达。由于不同层级特征在表征内容与语义层次上存在显著差异，若缺乏有效的跨层语义建模与上下文引导机制，将难以实现语义特征的充分融合，进而限制了解码器对语义信息的有效利用。

为此，本文引入 SCTB 模块，在跳跃连接中显式建模不同层级特征的上下文依赖，提升跨尺度语义协同表达能力。SCTB 由 SSCA 和 CFN 组成，分别用于增强空间与通道信息的交互和实现多尺度语义特征互补。

如图 5 所示，已知第 i 级特征 $I_i \in \mathbf{R}^{C_i \times h \times w}$（i=1, 2, 3, 4），其中 $h = \frac{H}{16}, w = \frac{W}{16}$。SCTB 的具体过程公式如下：

$$J_S = \mathrm{LN}\left(\left[I_1, I_2, I_3, I_4\right]\right), J_i = \mathrm{LN}\left(I_i\right) \tag{8}$$

$$P_i = \mathrm{SSCA}\left(J_1, J_2, J_3, J_4, J_S\right) + I_i, O_i = \mathrm{CFN}_i\left(P_i\right) \tag{9}$$

式中 LN 表示层归一化，$J_i \in \mathbf{R}^{C_i \times h \times w}$ 和 $J_S \in \mathbf{R}^{C_\Sigma \times h \times w}$ 是 SSCA 的五个输入。P_i 表示 SSCA 和输入第 i 级特征的和，O_i 表示 SCTB 的输出。

1.3.1　空间嵌入的单头信道交叉注意力机制

在图 5(a)中，SSCA 使用四个输入标记 J_i 作为 Query，一个拼接标记 J_S 作为键和值。公式如下：

$$Q_i = W_{dt}^{Q} W_p^{Q} J_i, K = W_d^{K} W_p^{K} J_S, V = W_d^{V} W_p^{V} J_S \tag{10}$$

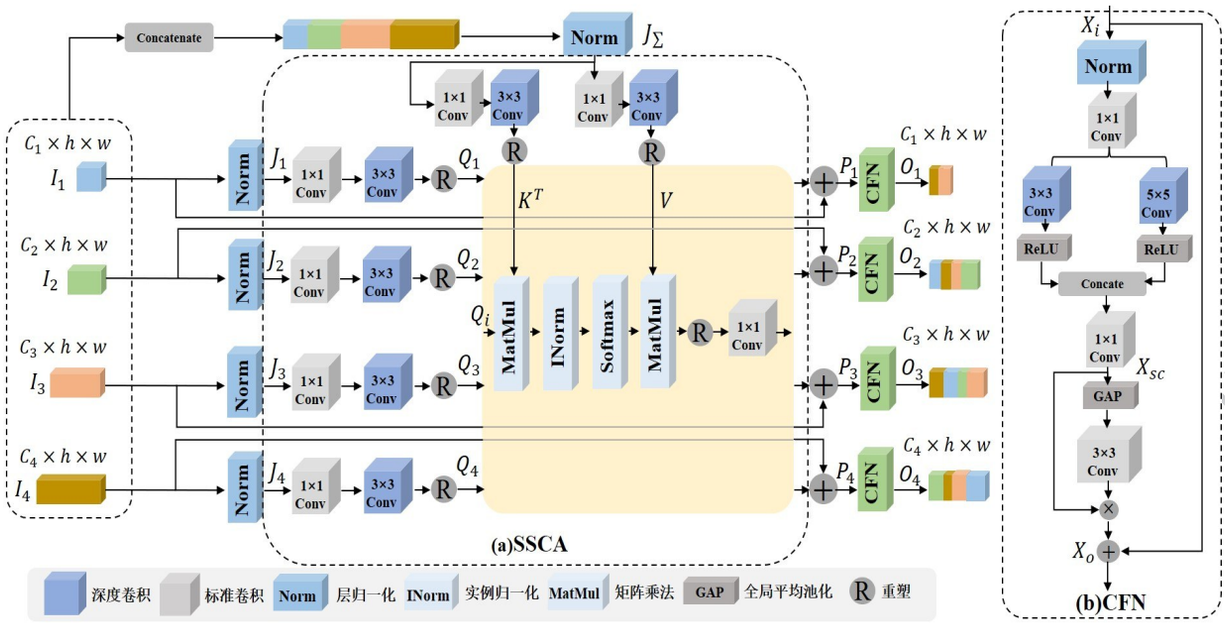
式中 $W_{pi}^{(i)} \in \mathbf{R}^{C_i \times 1 \times 1}$ 和 $W_p^{(i)} \in \mathbf{R}^{C_i \times 1 \times 1}$ 是 1×1 的逐点卷积，$W_d^{(i)} \in \mathbf{R}^{C_i \times 3 \times 3}$ 和 $W_d^{(i)} \in \mathbf{R}^{C_\Sigma \times 3 \times 3}$ 是 3×3 的深度卷积。接下来，将 $Q_i \in \mathbf{R}^{C_i \times h \times w}, K, V \in \mathbf{R}^{C_\Sigma \times h \times w}$ 分别 Reshape 为 $\mathbf{R}^{C_i \times hw}$ 和 $\mathbf{R}^{C_\Sigma \times hw}$。过程如下：

$$CA_i = W_{pi} A_i V, A_i = \mathrm{Softmax}\left\{I \frac{Q_i K^T}{\lambda}\right\} \tag{11}$$

式中 CA_i 是 SSCA 的输出，A_i 代表不同层级的基于协方差的注意力图，I 表示实例规范化操作，而 λ 是一个可选的温度因子，定义为 $\lambda = \sqrt{C_\Sigma}$。

1.3.2　互补前馈网络

在图 5(b)中，给定一个输入张量 $X_i \in \mathbf{R}^{C_i \times h \times w}$，CFN 首先刻画多尺度的局部空间特征与全局通道信息。具体来说，在层归一化处理后，CFN 先利用 1×1 卷积按比例 η 扩展通道维度，然后将特征图划分为 $X_{3\times3}$ 和 $X_{5\times5}$，分别经过 3×3 和 5×5 深度卷积以强化局部空间特征，接着通过通道拼接融合多尺度信息，并将维度还原至初始大小 X_{sc}。最后通过全局平均池化等一系列加权操作得到输出特征 X_o。过程



((a)spatial embedded single-head channel-cross attention;(b)comple mentary feed-forward network)

图5　空间-通道交叉Transformer块

Fig. 3　Spatial-channel cross transformer block

如下：

$$X_{3\times3},X_{5\times5}=\mathrm{Chunk}\left(\mathrm{Conv}_{1\times1}\left(LN\left(X_i\right)\right)\right) \tag{12}$$

$$X_{sc}=\mathrm{Conv}_{1\times1}\left[\upsilon\left(\mathrm{D}_{3\times3}\left(X_{3\times3}\right)\right),\upsilon\left(\mathrm{D}_{5\times5}\left(X_{5\times5}\right)\right)\right] \tag{13}$$

$$X_o=\mathrm{Con\,v}_{3\times3}\left(\mathrm{GAP}\left(X_{sc}\right)\right)\odot X_{sc}+X_i \tag{14}$$

式中，Conv表示普通卷积，υ表示ReLU激活函数，D代表深度卷积，Chunk表示沿着通道维度将特征向量划分为两个相等的部分，GAP表示全局平均池化。

1.4　轻量梯度门控模块

本文在解码器中采用深度可分离卷积以显著降低计算成本，但缺乏通道交互，削弱了细节特征的提取能力。为此，在解码器末端引入LGG模块，通过Sobel梯度显式建模边缘特征，并结合空间门控与轻量通道加权，有效保留了目标细节，同时未显著增加计算开销。

如图6所示，首先对输入特征$X\in\mathbf{R}^{C\times H\times W}$进行高斯平滑以抑制噪声干扰，得到特征图$G$。随后，利用Sobel算子提取梯度响应$\nabla G$，提取的梯度特征经过一个7×7卷积层及Sigmoid激活函数，得到

空间注意力权重A_s。通过该空间注意力权重对特征进行门控，得到调制后的特征X'。接着，采用1×1卷积层和Sigmoid激活函数对通道维度进行选择性加权，生成通道注意力权重A_c。最后，将调制后

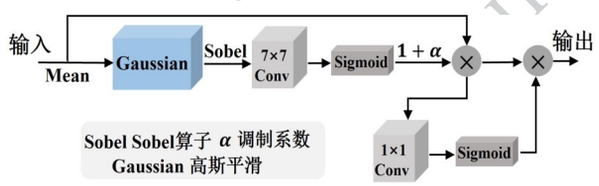


图6　轻量梯度门控模块

Fig. 6　Lightweight Gradient Gating Module

的特征与通道注意力权重逐元素相乘，获得增强后的特征表示Y。公式如下：

$$G=\mathrm{Gaussian}\left(\mathrm{Mean}\left(X\right)\right) \tag{15}$$

$$\nabla G=\sqrt{\left(G*S_x\right)^2+\left(G*S_y\right)^2} \tag{16}$$

$$A_s=\sigma\left(\mathrm{Conv}_{7\times7}\left(\nabla G\right)\right),X'=X\odot\left(1+\alpha M_s\right) \tag{17}$$

$$A_c=\sigma\left(\mathrm{Conv}_{1\times1}\left(X'\right)\right),Y=X'\odot A_c \tag{18}$$

式中，S_x,S_y表示Sobel算子卷积核，σ表示Sigmoid激活函数，α为调制系数(默认0.5)。

2　实验结果与分析

2.1　数据集和评估指标

2.1.1　数据集

本文在SIRST(Dai 等，2021)、IRSTD(Yu 等，2021)和NUDT-SIRST(Li 等，2023)公开数据集上进

行实验。SIRST数据集共包含427幅真实红外场景图像，覆盖短波、中长波及950 nm波段，许多目标在复杂背景中表现模糊且不易分辨。IRSTD数据集则包含1000幅图像，目标具有不同的形状和尺寸，并伴随丰富的杂波背景，且多来自远距离成像场景。NUDT-SIRST数据集包含1,327张具有挑战性的红外目标图像，涵盖了更广泛的目标类别和尺寸，并具有复杂背景和精确标注。这些数据集的多样性及复杂背景增加了目标检测的难度。实验中，两个数据集均按8∶2的比例划分为训练集和测试集。

2.1.2　评估指标

根据以往研究，我们采用像素级度量（IoU和nIoU）和目标级度量（Pd和Fa）来评估我们的方法。具体而言，IoU和nIoU（Dai等，2021）的定义如下：

$$\mathrm{IoU}=\frac{1}{n}\cdot\frac{\sum_{i=0}^{n}tp_i}{\sum_{i=0}^{n}(fp_i+fn_i-tp_i)} \tag{19}$$

$$\mathrm{nIoU}=\frac{1}{n}\cdot\sum_{i=0}^{n}\frac{tp_i}{fp_i+fn_i-tp_i} \tag{20}$$

式中n表示样本总数，tp、fp和fn分别表示为真阳性、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。

而目标级度量Pd和Fa可定义为：

$$\mathrm{Pd}=\frac{1}{n}\cdot\sum_{i=0}^{n}\frac{N_{pred}^{i}}{N_{all}^{i}} \tag{21}$$

$$\mathrm{Fa}=\frac{1}{n}\cdot\sum_{i=0}^{n}\frac{P_{false}^{i}}{P_{all}^{i}} \tag{22}$$

式中N_{pred}，N_{all}表示正确检测到的对象的数量和总对象的数量，并且P_{false}，P_{all}表示错误检测到的对象的像素和总对象的像素。当预测结果中心与真实值中心的距离小于4时，判定为检测正确。

2.2　实现细节

实验在一台搭载Nvidia RTX A5000 GPU的服务器上，基于Pytorch框架实现。为统一输入规模，所有图像被调整至512×512。训练采用Adam优化器，学习率设为0.0005，批量大小为4，总共迭代300个epochs。

2.3　方法比较

2.3.1　评估指标比较

本文在SIRST、IRSTD和NUDT-SIRST数据集上开展对比实验，以评估所提方法与现有先进方法的性能差异，相关结果见表1。实验结果表明，深度学习方法（Dai等，2021；Liu等，2021；Wu等，2022；Wang等，2023；Xu等，2024）整体优于传统算法（Gao等，2013；Dai等，2017；Zhang等，2019），而本文方法在两类数据集上均展现出更加突出的性能表现。其中，在SIRST、IRSTD和NUDT-SIRST数据集上的IoU分别达到80.12%、66.28%和84.26%，nIoU分别为78.23%、64.58%和86.48%，同时在检测概率（Pd）方面分别达到99.88%、94.23%和98.21%，并保持极低的虚警率（Fa）1.12×10^{-6}、4.39×10^{-6}和14.57×10^{-6}。

进一步的实验结果（表2）显示，在仅小幅增加参数量和计算量的前提下，本文方法依然能够保持较快的推理速度，同时在多项检测指标上取得了更好的性能表现。

2.3.2　可视化结果比较

在SIRST与NUDT-SIRST数据集的可视化结果中（见图7，右上角为放大后的小目标，黄色圆圈标注噪声点），不同场景均展示了本文方法的优势。在第1行的地面场景中，SwinT、ALCNet和HCFNet对小目标的响应较弱，未能有效捕捉细节特征，而MMLNet的目标轮廓则出现了明显的膨胀现象。相比之下，本文方法依托MPFM的多尺度感知与SCTB的跨层上下文建模，在保证检测完整性的同时生成了更为紧致的目标边界。在第2行空中场景中，由于目标数量较多，检测难度显著增加。SwinT虽能完整检测目标数量，但目标丢失了大部分特征；其他方法在保持整体目标结构的同时，边缘轮廓普遍粗糙。相比之下，本文方法能够在保持目标分离的同时保留清晰边界，与GT的一致性更高。在第3行与第4行的复杂空中场景中，背景纹理复杂，容易导致目标与背景混淆。SwinT和AGPCNet边缘过于平滑，缺乏必要的细节。尽管HCFNet、MMLNet在检测精度上表现良好，但其目标轮廓存在一定膨胀，边缘过渡不够锐利。本文方法在SCTB的语义融合基础上引入LGG，并通过Sobel梯度引导进一步增强边缘与纹理特征，从而在保证检测完整性的同时获得更加细致的目标轮廓。然而，由于整体模型设计侧重于多尺度特征提取与跨层语义融合，而缺乏针对背景噪声及传感器噪声的专门抑制机制。如图7黄色圆圈所示，当噪声强度接近或部分覆盖小目标时，模型难以完全区分真实目标与干扰，仍可能残留少量噪声点，对检测结果产生一定影响。

表1　相同评估指标下不同方法的实验结果对比

Table 1　Comparison of experimental results of different methods under the same evaluation metrics

方法	SIRST				IRSTD				NUDT-SIRST			
	像素级度量		目标级度量		像素级度量		目标级度量		像素级度量		目标级度量	
	IoU	nIoU	Pd	Fa	IoU	nIoU	Pd	Fa	IoU	nIoU	Pd	Fa
RIPT(Dai 等,2017)	25.49	33.01	85.32	24.75	8.15	16.12	68.35	26.36	9.28	11.07	43.29	166.30
PSTNN(Zhang 等,2019)	39.44	47.72	83.49	41.07	16.44	25.91	65.32	76.92	14.85	23.57	66.13	44.17
IPI(Gao等,2013)	40.48	50.95	91.74	148.37	14.40	31.29	86.35	450.36	17.76	15.42	74.49	41.23
U-Net(Ronneberger 等,2015)	70.24	69.77	90.52	44.83	57.23	60.21	84.34	22.23	61.37	59.56	91.34	47.21
SwinT(Liu 等,2021)	70.53	69.89	92.19	33.42	59.89	58.78	86.59	17.74	59.34	60.82	94.22	33.24
ACM(Dai 等,2021)	72.45	72.15	93.52	12.39	63.38	60.80	91.58	15.31	63.12	64.40	93.12	55.22
AGPCNet(Zhang 等,2021)	73.69	72,60	98.17	16.99	62.26	60.58	92.83	13.12	80.23	80.77	95.20	19.33
AlCNet(Dai 等,2021)	74.30	73.10	97.35	11.33	62.00	59.60	91.82	14.53	64.74	67.20	94.18	34.61
Res-Vit(Liu 等,2023)	72.82	71.22	98.15	27.15	61.89	60.64	90.91	12.64	68.65	66.28	89.96	66.33
UIUNet(Wu 等,2023)	74.54	73.18	98.17	18.32	64.75	62.32	92.61	18.18	82.61	83.89	97.89	**11.47**
MTU-Net(Wu 等,2023)	79.32	77.13	93.39	11.36	61.12	63.13	90.86	14.07	80.18	81.24	96.57	23.12
HCFNet(Xu 等,2024)	80.09	**78.31**	98.26	15.17	64.69	62.84	91.71	13.28	78.69	77.61	88,12	17.28
MMLNet(Li 等,2025)	79.17	76.16	96.33	12.18	**66.37**	**65.24**	93.08	13.37	**86.19**	86.02	97.11	15.34
Ours	**80.12**	78.23	**99.88**	**1.12**	66.28	64.58	**94.23**	**4.39**	84.26	**86.48**	**98.21**	14.57

注:加粗字体表示各列最优结果

表2　不同方法的计算开销与参数量对比

Table 2　Comparison of computational complexity and model parameters

方法	Params(M)	Computational complexity(GMac)	FPS(f/s)
AGPCNet	**12.43**	172.84	23.89
SwinT	26.14	**27.71**	14.76
UIUNet	50.54	33.64	11.21
HCFNet	15.29	93.16	**54.26**
Ours	21.19	67.37	47.89

注:加粗字体表示各列最优结果

总体而言,尽管存在上述不足,本文方法在小目标的边缘、形状等判别性特征提取方面仍表现出显著优势,从而有效提升了检测精度与鲁棒性。

2.4　消融实验

本文在SIRST与NUDT-SIRST数据集上开展了消融实验,以验证网络结构的有效性。所有实验均以U-Net为基础网络,首先将其解码器替换为由深度可分离卷积构成的新结构(BasNet),随后在此基础上逐步加入或替换本文提出的各个模块,以评估不同模块对模型性能的具体贡献。如表3所示,随着各模块的逐步引入,模型的检测性能持续提升,充分验证了所设计模块在增强目标检测能力方面的有效性。其中,表中符号“√”表示实验采用了相应模块,否则表示使用了BasNet(如图8)模块。

2.4.1　MPFM的有效性

图8 BasNet网络结构

Fig. 8 BasNet network architecture

为验证MPFM模块在提升小目标判别特征提取方面的有效性,本文将原网络中的MPFM模块替换为普通U-Net网络编码块进行对比实验。实验结果如表3所示,在SIRST数据集上的IoU和nIoU分别为75.96%和75.13%,检测精度出现一定程度下降。这是因为普通U-Net编码阶段仅依赖单路径卷

积,缺乏多尺度感知及局部和全局特征交互能力。

相比之下,MPFM通过并行局部、全局及串行卷积分支实现多尺度特征的充分提取,并结合LGGA与GCSA机制,有效增强了特征表达能力和目标区域响应。可视化结果如图10所示,当MPFM被替换为普通U-Net编码块时,检测结果中的小目标轮廓模糊,边缘呈现锯齿状,部分目标出现断裂或形变,且目标区域缺乏高频细节,表现较为粗糙。在云层和地面等复杂背景中,该替换模型对噪声抑制不足。相比之下,本文模型能够更突出地表征小目标的边缘及判别特征,且在复杂背景下有效抑制干扰,表明

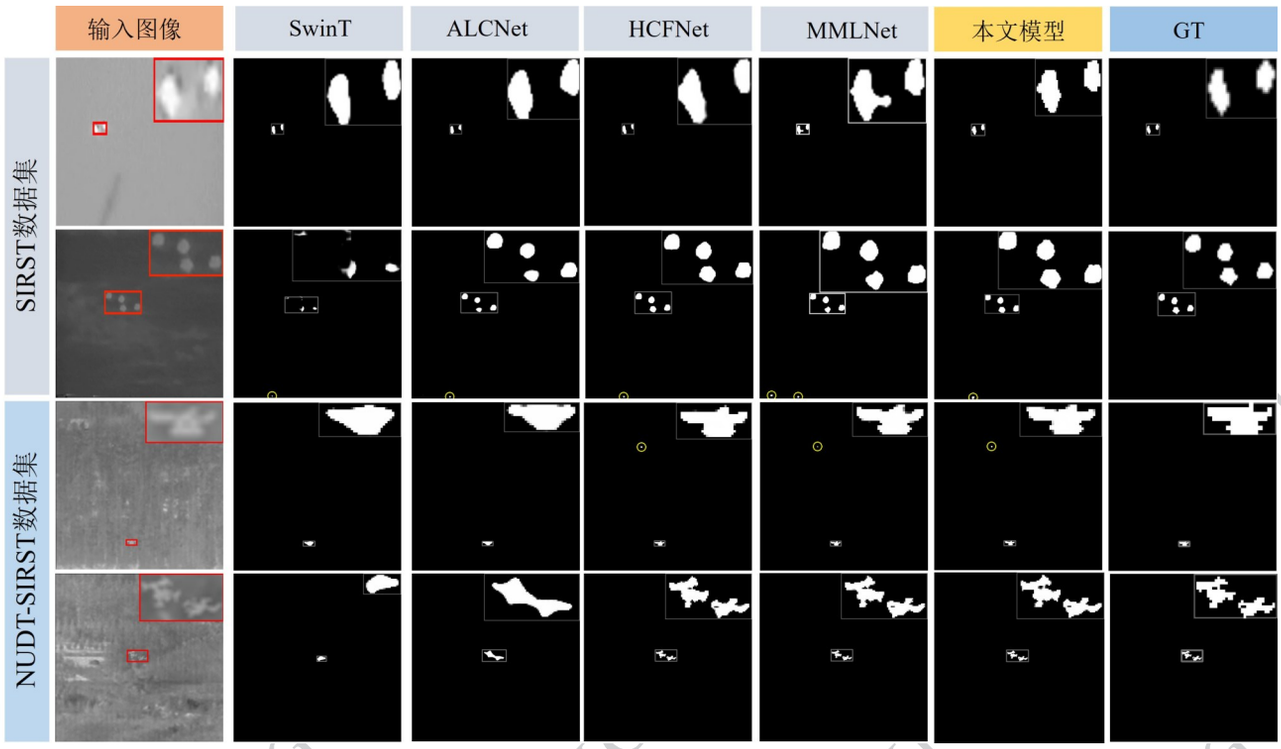


图7　不同方法可视化图比较

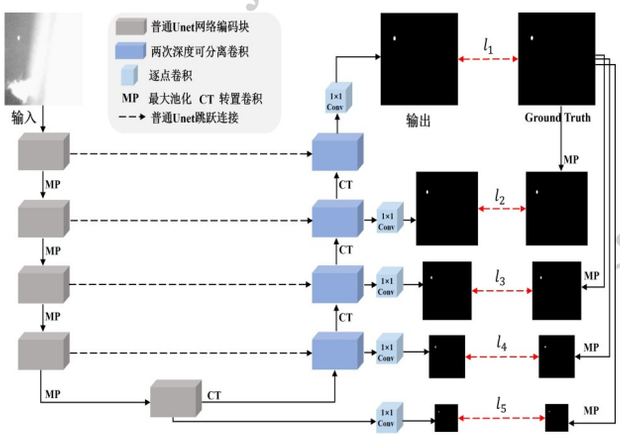


Fig. 7　Comparison of visualizations using different methods

MPFM在下采样阶段能够充分提取小目标的判别性信息并提升检测精度。

此外，为评估MPFM模块中两类注意力机制的实际贡献以及其可能产生的信息冗余对检测性能的影响，我们对模块内部各注意力分支进行了消融实验。实验结果如表4所示，随着不同注意力机制逐步加入，模型检测性能稳步提升；而不同注意力排列方式带来的性能差异较小，说明两类注意力在特征建模中相互补充，共同提升了特征表达能力。

2.4.2　LGGA模块p值设置的合理性分析

为验证不同层级LGGA模块中输入特征图划分

表3　不同模块对检测性能的影响

Table 3　Effect of different modules on detection performance

BasNet	MPFM	SCTB	LGG	IoU %	nIoU %
√				72.34	72.87
√	√			76.33	74.72
√	√		√	77.69	75.97
√		√		75.12	74.84
√		√	√	75.96	75.13
√	√	√		78.92	77.87
√	**√**	**√**	**√**	**80.12**	**78.23**

注：加粗字体为本文模型设计，实验结果来自SIRST数据集

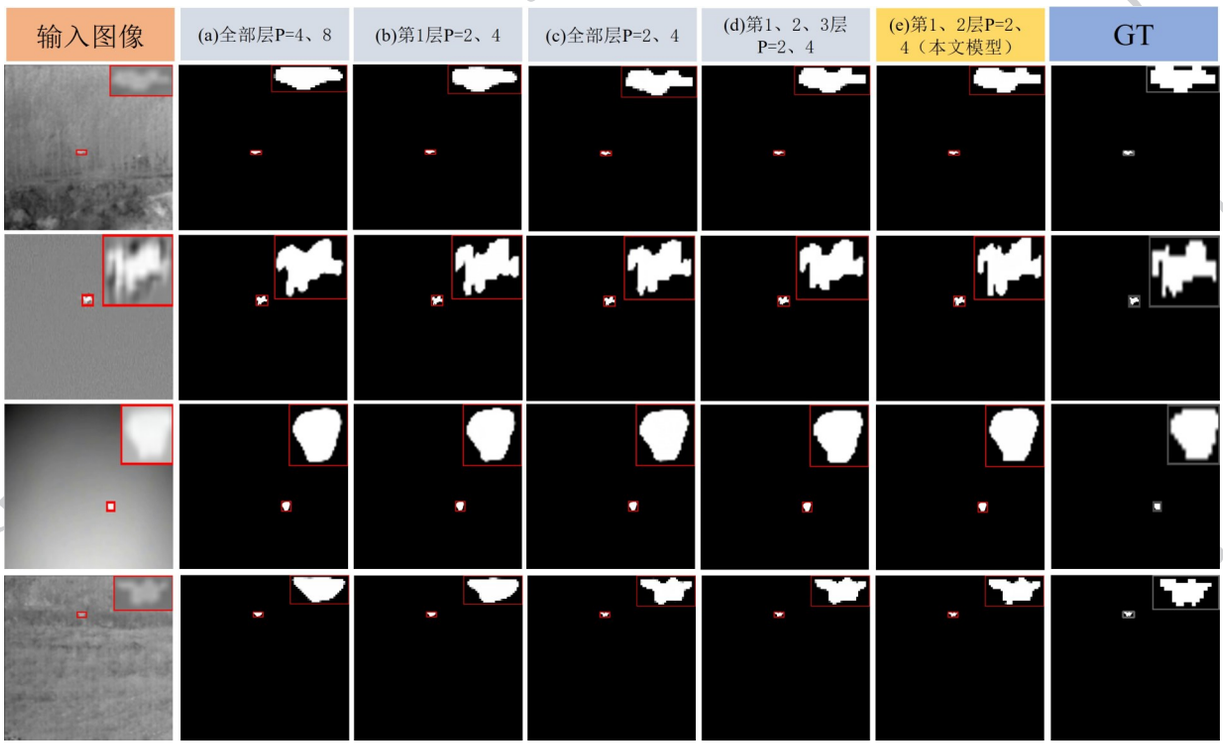
表4　MPFM中不同注意力分支对检测性能的影响

Table 4　Effect of different attention branches in MPFM on detection performance

MPFM	LGGA	GCSA	IoU %	nIoU %
①			76.94	75.97
①	②		78.73	77.72
①		②	77.29	76.91
①	③	②	79.76	77.87
①	②	③	**80.12**	**78.23**

注：加粗字体表示本文模型设计，数字符号表示模块的加入顺序，未标注数字符号表示相应模块未加入

块的尺寸(p)设定的合理性，本文设计了p值消融实验。本文网络第1、2层(从上往下)的p值选取为2和4，其他层则采用4和8。表5的实验结果验证了该p值设置的有效性。具体而言，网络浅层(第1、2层)特征具有较高空间分辨率，接近原始输入图像，包含丰富的边缘、纹理及局部细节信息，但语义抽象程度较低。因此，选用较小的p值



((a) All layers with P = 4, 8; (b) Layer 1 with P = 2, 4; (c) All layers with P = 2, 4; (d) Layers 1, 2, and 3 with P = 2, 4; (e) Layers 1 and 2 with P = 2, 4 (Proposed Model))

图9　不同块的尺寸(P)的可视化图

Fig. 9　Visualization of different patch size

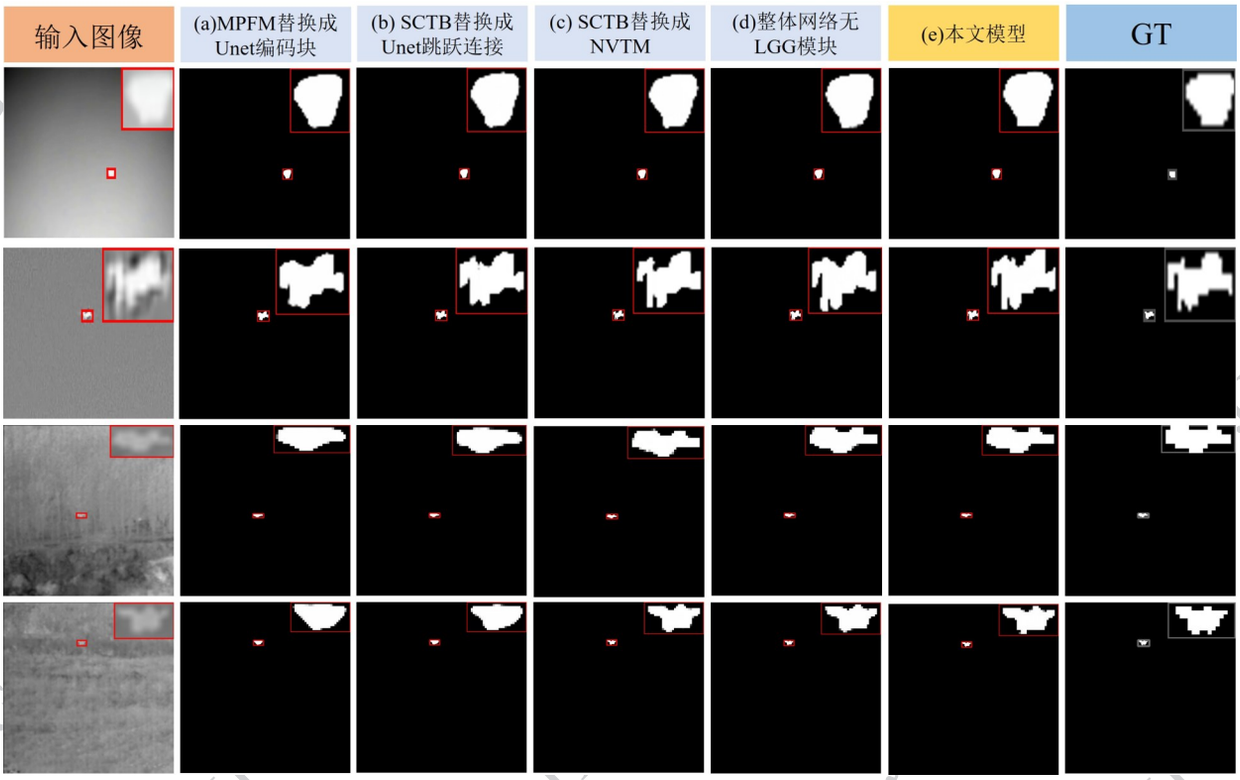
能局部建模细节特征，有助于捕捉目标的边缘特征。

相比之下，深层(第3、4、5层)特征空间分辨率降低，感受野更大，语义信息更加抽象和全局化。

对深层采用较大的p值(如4或8)有利于捕获更大范围上下文，增强全局语义并改善目标与背景区分。如图9所示的天空、陆地等场景检测结果表明，全部层p采用4、8时检测结果虽然保持了目标的整体轮廓，但边缘部分出现模糊和膨胀现象，细节刻画不足。第1层p采用2、4时相比全部层采用4、8，目标边缘较为清晰，但部分细节仍存在欠缺，复杂背景下仍有少量误检；全部层p为2、4尽管小目标边缘细节得到增强，但由于深层未采用较大的p，模型对全局语义的建模能力不足，容易在噪声背景下产生漏检；第1-3层p为2、4时，虽细节有所改善，但全局语义利用仍不足，检测出的小目标不够完整。本文对块的尺寸取值较为合理，在细节刻画方面表现最佳，能够在天空等复杂场景下精准检测出小目标。

2.4.3　SCTB的有效性

为验证连接融合策略对特征融合与检测性能的影响，本文在所提模型基础上，将SCTB分别替换为普通U-Net跳跃连接和MTU-Net的MVTM模块(Wu等，2023)进行对比实验。表6结果显示，替换SCTB后检测精度明显下降。图10化结果表明，采用普通跳跃连接虽能保留整体目标轮廓，但边缘依旧模糊，且在噪声背景中目标显著性较弱。这是因为普通U-Net跳跃连接仅通过直接拼接将编码器与解码器特征融合，虽可保留部分浅层细节，但缺乏深浅层特征间长程依赖建模，融合效率较低。引



((a) MPFM replaced by U-Net encoding block; (b) SCTB replaced by U-Net skip connection; (c) SCTB replaced by NVTM; (d) Entire network without LGG module; (e) Proposed model)

图 10　不同模块替换可视化图

Fig. 10　Visualization of different module replacements

表 5　不同 p 值设置对检测性能的影响

Table 5　Impact of different P-Value settings on detection performance

网路各层对应的 p 值选取(层数从上往下递增)	IoU%	nIoU%
全部层 p=2、4	77.97	75.14
全部层 p=4、8	74.62	74.94
第 1 层 p=2、4，其余 p=4、8	77.70	74.83
第 1、2 层 p=2、4，其余 p=4、8(本文取值)	**80.12**	**78.23**
第 1、2、3 层 p=2、4，其余 p=4、8	78.59	76.97

注：加粗字体表示本文模型设计，结果来自 SIRST 数据集

入 MVTM 模块后检测精度有所提升，目标边界趋于平滑，但真实边缘的细微起伏仍存在缺失。这是由于编码器不同层级特征在语义抽象和空间细节上差异显著，而 MVTM 缺乏对其上下文关联的精细建模，导致语义融合不足，影响细节特征提取。相比之下，本文模型引入 SCTB 后检测精度进一步提升，可视化结果亦验证了 SCTB 在缓解层间语义差异和增强细节建模方面的有效性。

此外，为评估在同一编码阶段同时引入 MPFM 和 SCTB 的通道－空间注意力机制是否会造成功能冗余并影响检测性能，我们设计了针对性的消融实验。表 7 的结果表明：当分别去除 MPFM 或 SCTB 中的通道－空间注意力模块时，模型检测精度均有所下降；而同时保留两者时，性能达到最佳。说明两类注意力机制在特征建模中发挥互补作用，尽管额外的注意力模块会带来一定的计算开销，但相较于其带来的精度提升，这种多重注意力设计依然合理且有效。

表 6　不同连接方式对检测性能的影响

Table 6　Impact of different connection methods on detection performance

连接方式	IoU%	nIoU%
Unet 跳跃连接	77.69	75.97
MVTM	78.49	74.91
SCTB	**80.12**	**78.23**

注：加粗字体为本文模型设计，实验结果来自 SIRST 数据集

表7　不同模块叠加注意力对检测性能的影响

Table 7　Effect of stacking attention mechanisms in different modules on detection performance

注意力设计	IoU %	nIoU %
MPFM中去除GCSA模块，其他不变	78.73	77.72
SCTB中去除通道-空间注意力，其他不变	78.21	77.03
MPFM与SCTB中均去除通道-空间注意力	77.29	76.91
保持所有注意力机制不变（本文模型设计）	**80.12**	**78.23**

注：加粗字体为本文模型设计，实验结果来自SIRST数据集

2.4.4　深度可分离卷积与LGG模块的有效性

为验证解码器中深度可分离卷积与LGG模块设计的有效性，本文进行了消融实验。具体而言，在保持模型其他模块不变的情况下，开展了三组对比实验：解码器仅采用普通卷积；解码器采用深度可分离卷积但未引入LGG模块；解码器采用深度可分离卷积并在末端引入LGG模块。实验结果如表8所示，本文设计在IoU和nIoU上均优于其他两种方案，同时保持较低计算成本，验证了该设计在兼顾计算效率与细节保留方面的优势。其原因在于，采用普通卷积虽能一定程度保留细节，但计算开销较大。而仅使用深度可分离卷积可降低参数量和计算量，但由于缺乏跨通道交互，高频纹理和细微边缘特征的建模能力下降，导致检测性能有所下降。所以本文通过LGG模块利用Sobel梯度显式建模边缘信息，并结合空间门控调制与轻量通道加权机制，有效弥补了深度可分离卷积在细节保留上的不足。可视化结果（图10）显示，本文模型检测到的小目标边缘更清晰，背景干扰得到有效抑制，进一步验证了模块设计的有效性。

表8　不同解码器设计对检测性能的影响

Table 8　The impact of different decoder designs on detection performance

解码器设计	IoU	nIoU	Params(M)	Computational complexity(GMac)
两次普通卷积	78.69	76.97	23.25	92.35
两次深度可分离卷积	78.92	77.87	21.19	67.09
两次深度可分离卷积+末端LGG	**80.12**	**78.23**	**21.19**	**67.37**

注：加粗字体为本文模型设计，实验结果来自SIRST数据集

3　结　论

本文针对复杂背景下红外小目标判别特征提取不足和跨层语义融合不充分的问题，提出了一种基于多分支感知与跨层语义建模的检测方法（MPCF-Net）。该方法在编码器中引入MPFM，结合LGGA和全局通道空间注意力GCSA增强目标响应与关键区域判别能力，并采用SCTB替代跳跃连接以优化特征融合；在解码器中使用深度可分离卷积降低计算成本，并集成LGG以有效保留小目标细节。然而，该方法由于引入MPFM与SCTB等多分支和Transformer结构，计算开销有所增加；同时，LGGA中局部感受野尺寸的动态设定仍依赖人工调节，缺乏完全自适应机制；此外，网络在应对复杂背景噪声方面仍存在一定的局限性。未来的研究将聚焦于进一步优化结构设计，以在提升检测精度的同时降低计算复杂度，并增强对复杂噪声环境的鲁棒性。

参考文献（References）

Dai Y M and Wu Y Q. 2017. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(8): 3752-3767 [DOI: 10.1109/JSTARS.2017.2700023]

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. Los Alamitos: IEEE, pp. 950-959

Dai Y M, Wu Y Q, Zhou F and Barnard K. 2021. Attentional local contrast networks for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 59(11): 9813-9824 [DOI: 10.1109/TGRS.2020.3044958]

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]

Huang L, Yang Y, Yang C Y, Yang W and Li Y H. 2023. FS-YOLOv5: light weight infrared rode target detection method. Computer Engineering and Applications, 59(9): 215-224 (黄磊, 杨媛, 杨成煜, 杨威, 李耀华. 2023. FS-YOLOv5: 轻量化红外目标检测方法. 计算机工程与应用, 59(9): 215-224) [DOI: 10.3778/j.issn.1002-8331.2210-0487]

Kou R K, Wang C P, Luo Y, Zhang Y, Xu Z L, Peng Z M, Wu C Y, Fu Q. 2024. Multiscale small-target detection techniques in single-frame infrared images. 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, 武晨燕, 付强. 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]

Lin L K, Wang S Y and Tang Z X. 2018. Using deep learning to detect small targets in infrared oversampling images. Journal of Systems Engineering and Electronics, 29(5): 947-952 [DOI: 10.21629/JSEE.2018.05.07]

Li Q C, Li B W and Wang H C. 2016. Development trend of uncooled infrared imaging technology and its military application. Dual Use Technologies & Products, (21): 54-57 (李其昌, 李兵伟, 王宏臣. 2016. 非制冷红外成像技术发展动态及其军事应用. 军民两用技术与产品, (21): 54-57)

Li B, Xiao C, Wang L, Wang Y, Lin Z, Li M, An W and Guo Y. 2023. Dense nested attention network for infrared small target detection. IEEE Transactions on Image Processing, 32: 1745-1758 [DOI: 10.1109/TIP.2022.3199107]

Li Q, Zhang W, Lu W and Wang Q. 2025. Multibranch mutual-guiding learning for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-10 [DOI: 10.1109/TGRS.2025.3526754]

Li J H, Zhang P, Wang X W, Huang S Z. 2020. Infrared small-target detection algorithms: a survey. Journal of Image and Graphics, 25(9): 1739-1753 (李俊宏, 张萍, 王晓玮, 黄世泽. 2020. 红外弱小目标检测算法综述. 中国图象图形学报, 25(9): 1739-1753) [DOI:10.11834/jig.190574]

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S and Guo B. 2021. Swin Transformer: Hierarchical vision transformer using shifted windows//Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE, pp. 10012-10022

Liu F C, Gao C Q, Chen F, Meng D Y, Zuo W M and Gao X B. 2023. Infrared small and dim target detection with transformer under complex backgrounds. IEEE Transactions on Image Processing, 32: 5921-5932 [DOI: 10.1109/TIP.2023.3326396]

Ronneberger O, Fischer P and Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, pp. 234-241

Sobel I and Feldman G. 1968. A 3x3 isotropic gradient operator for image processing. A talk at the Stanford Artificial Project, 1968: 271-272

Wang Y J, Wang Y L, Xia R Q and Liu Y. 2023. High-precision azimuth extraction of targets in wide-field infrared warning systems. Laser Technology, 47(2): 200-204 [DOI: 10.7510/jgjs.issn.1001-3806.2023.02.007] (王云杰, 王艳林, 夏润秋, 刘洋. 2023. 大视场红外告警系统中目标高精度方位提取. 激光技术, 47(2): 200-204)

Wang Z W, Zhang Z M, Xu K, Zhang F M and Liu B. 2022. Research on linear pose measurement method based on optimal polarization angle. Infrared and Laser Engineering, 51(3): 329-338 (王志武, 张子淼, 许凯, 张福民, 刘斌. 2022. 基于最佳偏振角的线性位姿测量方法研究. 红外与激光工程, 51(3): 20210241-1)

Wang W T, Qin H L, Cheng W X, Wang C M, Leng H B and Zhou H X. 2017. Small target detection in infrared image using convolutional neural networks//AOPC 2017: Optical Sensing and Imaging Technology and Applications. Bellingham: SPIE, 10462: 1335-1340 [DOI: 10.1117/12.2285689]

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]

Wang C, Wang H and Pan P. 2023. Local contrast and global contextual information make infrared small object salient again [EB/OL]. [2023-01-22]. https://doi.org/10.48550/arXiv.2301.120937

Wu T, Li B, Luo Y, Wang Y, Xiao C, Liu T, Yang J, An W and Guo Y. 2023. MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-15 [DOI: 10.1109/TGRS.2023.3235002]

Xu S B, Zheng S C, Xu W H, Xu R T, Wang C W, Zhang J G, Teng X Q, Li A and Guo L. 2024. HCF-Net: Hierarchical context fusion network for infrared small object detection//2024 IEEE International Conference on Multimedia and Expo (ICME). Los Alamitos: IEEE, pp. 1-6

Yu Y C, Zhan F N, Lu S J, Pan J X, Ma F Y, Xie X S and Miao C Y. 2021. WaveFill: A wavelet-based generation network for image inpainting//Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE, pp. 14114-14123

Zhang T F, Cao S Y, Pu T and Peng Z M. 2021. AGPCNet: Attention-guided pyramid context networks for infrared small targetdetection [EB/OL]. [2021-11-03].

https://doi.org/10.48550/arXiv.2111.03580

Zeng M, Li J X and Peng Z X. 2006. The design of top-hat morphological filter and application to infrared target detection. Infrared Physics & Technology, 48(1): 67-76 [DOI: 10.1016/j.infrared.2005.04.006]

Zhu H, Liu S M, Deng L Z, Li Y S and Fu X. 2019. Infrared small target detection via low-rank tensor completion with top-hat regularization. IEEE Transactions on Geoscience and Remote Sensing, 58(2): 1004-1016 [DOI: 10.1109/TGRS.2019.2942384]

Zhang L D and Peng Z M. 2019. Infrared small target detection based on partial sum of the tensor nuclear norm. Remote Sensing, 11(4): 382 [DOI: 10.3390/rs11040382]

Zhao X K, Li M L, Zhang G, Li N and Li J S. 2021. Object detection method based on saliency map fusion for UAV-borne thermal images. Acta Automatica Sinica, 47(9): 2120-2131 (赵兴科, 李明磊, 张弓, 黎宁, 李家松. 2021. 基于显著图融合的无人机载热红外图像目标检测方法. 自动化学报, 47(9): 2120-2131) [DOI:10.16383/j.aas.c200021]

作者简介

钱孟豪，男，硕士研究生，研究方向为目标检测。E-mail: 770028821@qq.com

刘奎，男，通讯作者，教授，硕士研究生导师，研究方向为目标检测与图像去噪。E-mail: liukui@aqnu.edu.cn

章丰博，男，硕士研究生，研究方向为目标检测。E-mail: 15212975680@163.com

苏本跃，男，教授，硕士研究生导师，研究方向为人工智能与模式识别。E-mail: subenyue@sohu.com