SSCE-YOLO：基于频域–空间协同增强的红外遥感目标检测算法

程德强[1]，岳阳[1]，王衍辰[1]，寇旗旗[2]，徐飞翔[1]，孙凤乾[1]
（1.中国矿业大学信息与控制工程学院，江苏 徐州 221116；2.中国矿业大学计算机科学与技术学院/人工智能学院，江苏 徐州 221116）

摘要：针对红外遥感目标检测中的噪声与复杂背景干扰问题，提出了基于频域–空间协同增强的算法SSCE-YOLO。该方法构建由通道预筛、频域调制、深层门控传递和全局稳定性约束组成的层次化协同增强机制。通过频域–空间协同（FSC）滤波抑制噪声相关频段并保持目标结构；利用频域–空间门控高阶卷积（SGNC）将增强特征注入多阶通道交互以持续传递抑噪先验；引入通道感知全局上下文（CGC），结合均值–方差统计抑制高波动通道，并采用高效通道预筛注意力（ECPA）模块控制计算开销。在VEDAI-IR与HIT-UAV数据集上的实验结果表明，SSCE-YOLO的mAP@0.5分别由59.2提升至63.7、由79.1提升至82.2，优于当前主流算法，且兼顾精度和效率平衡。

关键词：红外遥感；目标检测；YOLO；频域–空间协同
中图分类号：TP391.41
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260204

SSCE-YOLO: spectral-spatial collaborative enhancement for infrared remote-sensing object detection

Cheng Deqiang[1], Yue Yang[1], Wang Yanchen[1], Kou Qiqi[2], Xu Feixiang[1], Sun Fengqian[1]
1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
2. School of Computer Science and Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116, China

Abstract: To address noise interference and complex background disturbance in infrared remote sensing object detection, a spectral-spatial collaborative enhancement algorithm termed SSCE-YOLO was proposed. The method constructed a hierarchical collaborative enhancement mechanism consisting of channel pre-screening, frequency-domain modulation, deep gated propagation, and global stability constraints. The frequency-spatial collaborative (FSC) filtering suppressed noise-related frequency bands while preserving target structures. The SpectralGNConv (SGNC) module injected FSC-enhanced features into multi-order channel interactions to continuously propagate denoising priors. Meanwhile, the channel-aware global context (CGC) employed mean-variance statistics to suppress high-fluctuation channels, while the efficient channel pre-selection attention (ECPA) module controlled computational overhead. Experiments on the VEDAI-IR and HIT-UAV datasets show that SSCE-YOLO improves mAP@0.5 from 59.2 to 63.7 and from 79.1 to 82.2, respectively, outperforming representative state-of-the-art methods while maintaining a favorable balance between detection accuracy and computational efficiency.

Key words: infrared remote-sensing, object detection, YOLO, spectral-spatial collaboration

---

收稿日期：2026–04–15；修回日期：2026–07–14
通信作者：孙凤乾，sunfengqian@cumt.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2023YFC2907600）；徐州高校服务“343”产业发展基金资助项目（No.gx2024004）；江苏省研究生科研与实践创新计划基金资助项目（No.KYCX25_2989）；中国矿业大学研究生创新计划基金资助项目（No.2024WLKXJ080）
Foundation Items: The National Key Research and Development Program Project (No. 2023YFC2907600), Xuzhou Universities Serving the “343” Industrial Development Project (No. gx2024004), Postgraduate Research and Practice Innovation Program of Jiangsu Province (No. KYCX25_2989), The Graduate Innovation Program of China University of Mining and Technology (No.2024WLKXJ080)

0　引言

红外遥感目标检测[1-2]作为对地观测领域的核心分支，因其全天时成像与强环境适应能力，在军事侦察、边海防监控和灾害评估等场景中具有重要应用价值。随着高分辨率[3]红外成像技术和遥感平台的发展，目标检测算法对精度、鲁棒性和效率提出了更高要求。然而，红外遥感成像仍面临两类挑战：一是热噪声等因素会引入大量随机高频干扰，削弱目标与背景对比度，使小目标易被淹没；二是俯视成像下目标尺度变化大、背景复杂，云层、地表和人工设施等干扰会降低检测稳定性与检出率。

针对以上问题，已有研究主要从空间域注意力增强、频域特征建模和全局上下文抑制等方向展开探索。在空间域注意力增强方面，CBAM[4]、Coordinate Attention[5]等方法通过通道、空间或位置信息重标定增强目标相关特征。然而，这类方法主要依据空间位置或通道响应强度进行增强，缺乏对噪声频谱成分与目标结构频谱成分的显式区分，容易在增强目标边缘的同时放大高频噪声伪响应。

在频域特征建模方面，FcaNet[6]、GFNet[7]等方法分别通过离散余弦变换（discrete cosine transform, DCT）和傅里叶域全局滤波建模频率信息，为频域增强提供了有效思路。进一步，在红外小目标检测领域，Deng 等[8]提出了 GSFANet，通过全局空间–频域注意机制挖掘目标与噪声在频域的差异。Xu 等[9]提出了 HDNet，通过混合域建模和多尺度高频信息增强提升红外小目标表征能力。上述研究表明，频域信息对红外弱小目标检测具有重要作用。然而，现有频域方法多侧重于单层频谱调制、全局频域依赖建模或高频细节增强，对噪声相关频段与目标边缘频段的区分仍不明显，且频域抑噪先验难以在深层通道中持续传递，残余高频噪声伪响应仍可能随着网络加深而被放大。

在全局上下文抑制方面，GCNet[10]通过全局上下文聚合增强长程依赖关系，提升复杂场景中的语义表达能力。Dai 等[11]提出了 ALCNet，通过注意力局部对比机制增强目标与邻域背景之间的差异，以提升红外小目标在复杂背景下的可分性。Li 等[12]提出了 DNA-Net，通过密集嵌套特征交互和级联通道–空间注意力保持小目标信息，并增强多层特征表达。然而，红外遥感图像中的复杂背景干扰不仅表现为空间位置混淆，还表现为部分通道在整幅图像范围内剧烈波动。云层、地表纹理、建筑边缘和道路结构等背景区域可能激活与目标相似的通道响应，仅依靠全局均值上下文、位置注意力或高频细节增强，难以有效刻画通道响应的稳定性差异。

综上所述，现有研究虽在空间域注意力增强、频域特征建模和全局上下文抑制方面取得了一定进展，但仍存在三方面局限。其一，常规空间注意力方法缺乏对噪声频谱成分的显式区分，容易造成目标边缘与噪声伪响应同步增强；其二，已有频域增强方法多停留在单层频谱调制，难以将频域抑噪先验持续注入深层通道交互过程；其三，现有全局上下文模块对复杂背景导致的通道响应波动缺少稳定性约束。针对上述问题，本文提出了频域–空间协同增强的红外遥感目标检测（spectral-spatial collaborative enhancement for infrared remote-sensing object detection, SSCE-YOLO）算法，构建通道预筛、频域子带选择、高阶门控传递和全局背景约束的协同增强机制。具体而言，高效通道预筛注意力（efficient channel pre-selection attention, ECPA）模块利用通道均值与方差统计筛除不稳定冗余响应；频域–空间协同滤波（frequency-spatial collaborative filter, FSC）模块通过局部空间分支和可学习复数频谱权重实现噪声相关频段的选择性抑制与目标边缘保持；频域–空间门控高阶卷积（SpectralGNConv, SGNC）模块将 FSC 模块生成的频域–空间增强特征转化为多阶门控核；通道感知全局上下文（channel-aware global context, CGC）模块则利用均值–方差全局上下文约束复杂背景引起的高波动通道。由此，本文方法并非对已有注意力或频域模块的简单组合，而是围绕红外遥感检测中的噪声干扰、复杂背景干扰问题，形成由浅层预筛、同层滤波、深层传递到全局校正的层次化协同增强框架。本文主要贡献如下。

（1）提出了一种面向红外遥感目标检测的频域–空间协同增强框架。该框架针对红外图像中的噪声干扰和复杂背景干扰问题，构建由通道预筛、频域子带选择、高阶门控传递和全局背景约束组成的层次化增强机制。

（2）设计了频域–空间协同滤波与高阶门控传

递机制。FSC 模块通过局部空间分支与可学习复数频谱权重实现目标边缘保持和噪声相关频段抑制；SGNC 模块进一步将 FSC 模块的输出转化为多阶门控核，使抑噪先验在深层通道交互中持续发挥作用，从而缓解残余噪声伪响应在网络深层被放大的问题。

（3）设计了通道稳定性感知的全局背景约束机制。CGC 模块通过均值–方差联合统计刻画通道响应强度与波动程度，从全局范围抑制复杂背景引起的高波动通道；同时结合 ECPA 模块的轻量通道预筛，在控制模型复杂度的前提下提升红外弱小目标检测的鲁棒性与稳定性。

1　SSCE-YOLO 算法模型

1.1　模型整体架构

SSCE-YOLO 算法模型以 YOLO11 为基础框架，由主干网络、特征融合网络和多尺度检测头组成。针对红外遥感图像中的噪声干扰和复杂背景干扰问题，本文在主干网络中引入 SSCE 频域–空间协同增强框架，整体结构如图 1 所示。

SSCE 框架由 ECPA、FSC、SGNC 和 CGC 模块组成，各模块分别作用于通道预筛、同层频域增强、深层门控传递和全局背景约束 4 个环节。其中，ECPA 模块用于抑制不稳定冗余通道，FSC 模块用于实现频域子带选择与空间结构保持，SGNC 模块用于将频域–空间先验注入多阶通道交互，CGC 模块用于约束复杂背景引起的高波动通道。四者在功能上具有明确分工，并在统一框架内形成由浅层预筛、同层增强、深层传递到全局校正的协同增强机制。

1.2　频域–空间协同表征增强模块

在 SSCE 框架中，FSC 模块先在同层内并行建模空间局部结构与频域谱权重，空间分支保持边缘细节，频域分支选择性抑制噪声子带，从而输出更高信噪比的频域–空间增强特征作为后续学习的输入基础。然而，单次滤波仍可能使残余伪响应在深层通道交互中被放大，因此 SGNC 模块进一步将 FSC 模块输出蕴含的抑噪先验显式转化为门控核并注入多阶通道交互过程，使抑噪约束在全网络中沿网络层级持续传递。

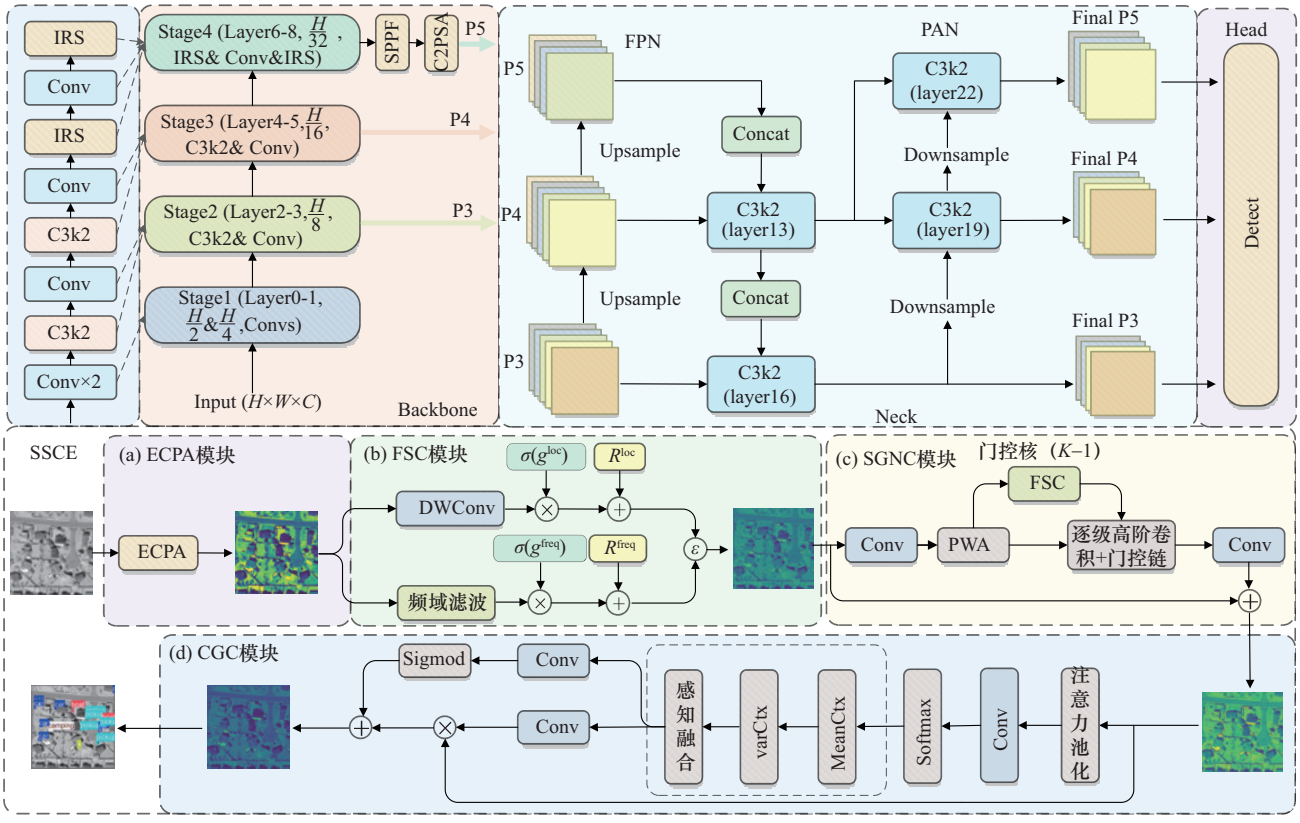


图 1　SSCE-YOLO 网络结构

1.2.1　频域–空间协同滤波模块

为应对红外图像中的噪声干扰问题，本文设计了FSC模块，其在同一层将输入特征沿通道维划分为局部与频域分支。前者保持目标边缘与纹理连续性，后者通过可学习复数频谱权重对噪声相关子带进行选择性抑制，随后通过门控融合与残差输出实现定向抑噪，FSC模块网络结构如图2所示。

设输入特征为$X \in \mathbb{R}^{C \times H \times W \times B}$，其中$B$为批大小，$C$为通道数，$H$,$W$为空间尺寸，将$X$按通道划分为局部分支特征$X^{\text{loc}}$与频域分支特征$X^{\text{freq}}$，对应残差分别记为$R^{\text{loc}}$和$R^{\text{freq}}$，分支门控权重分别记为$g^{\text{loc}}$和$g^{\text{freq}}$，其经Sigmoid激活后得到门控权重$G^{\text{loc}}$和$G^{\text{freq}}$，文中$\odot$表示逐元素乘法。FSC模块首先对$X$进行层归一化以稳定分布，然后将通道维度划分为局部和频域分支。

$$[X^{\text{loc}}, X^{\text{freq}}] = X \tag{1}$$

局部分支采用深度可分离卷积对空间邻域进行建模，得到：

$$\hat{X}^{\text{loc}} = \text{DWConv}_{3\times 3}(X^{\text{loc}}) \tag{2}$$

频域分支则对X^{freq}在空间维度执行二维实数傅里叶变换，得到频谱表示$F = \mathcal{F}(X^{\text{freq}})$。

为在频域实现可学习子带选择，FSC模块为每个通道引入复数形式的频谱权重$W \in \mathbb{C}^{\frac{C}{2} \times H' \times W'}$，其中$H'$和$W'$表示频域特征尺寸，对$W$的实部与虚部分别进行双线性插值后重构复数权重，并参与频域滤波。需要说明的是，该插值作用于可学习频谱权重W，而非输入特征频谱F本身，因此不会改变输入频谱采样坐标，可避免由输入频谱重采样引起的频率混叠问题。随后，经逆傅里叶变换回到空间域得到。

$$\hat{X}^{\text{freq}} = \mathcal{F}^{-1}(\tilde{F}), \hat{X}^{\text{freq}} \in \mathbb{R}^{\frac{C}{2} \times H \times W \times B} \tag{3}$$

实现对噪声相关频段的定向抑制以及对目标相关频段的选择性增强。

为了自适应控制两分支的修正强度，FSC模块在局部和频域分支分别设置可学习门控$g^{\text{loc}}, g^{\text{freq}} \in \mathbb{R}^{\frac{C}{2} \times 1 \times 1 \times 1}$，并通过Sigmoid激活函数得到门控权重。

$$G^{\text{loc}} = \sigma(g^{\text{loc}}), \quad G^{\text{freq}} = \sigma(g^{\text{freq}}) \tag{4}$$

最终两分支输出分别为：

$$Y^{\text{loc}} = R^{\text{loc}} + G^{\text{loc}} \odot \hat{X}^{\text{loc}} \tag{5}$$

$$Y^{\text{freq}} = R^{\text{freq}} + G^{\text{freq}} \odot \hat{X}^{\text{freq}} \tag{6}$$

将两分支特征按通道维度拼接并归一化，得到FSC模块输出特征。

$$Y = \text{LayerNorm}(\text{Concat}(Y^{\text{loc}}, Y^{\text{freq}})) \tag{7}$$

通过局部与频域分支的协同建模，FSC模块能够在保持目标边缘和局部结构连续性的同时，对噪声相关频段进行选择性抑制。该模块输出的频域–空间增强特征为后续SGNC模块的高阶门控交互提供了更稳定的输入基础。

1.2.2　频域–空间门控卷积模块

在红外遥感低信噪比场景中，高频随机噪声容

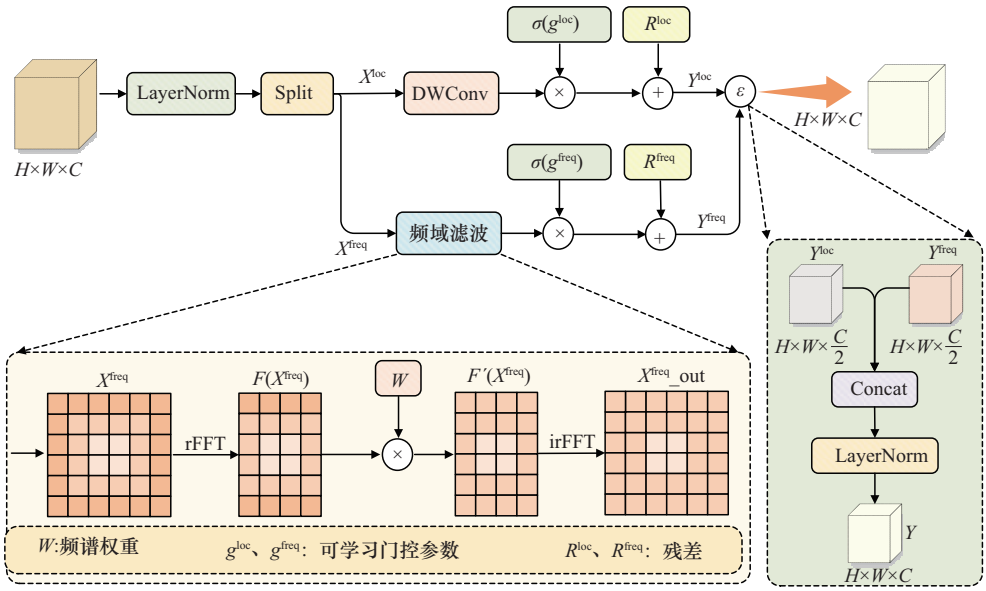


图2　FSC模块网络结构

易激发局部高频伪响应，若直接进入后续卷积交互过程，这类伪响应会在层级叠加中被放大并压制弱目标细节，仅在前端进行一次频域抑噪难以保证深层特征持续受控。为此，本文设计了SGNC模块，在通道维进行多阶分解，并将FSC模块输出的频域–空间增强特征显式切分为多阶门控核，在每一阶通道组合过程中以逐级乘法重标定的方式注入频谱与局部结构先验，从而实现对噪声伪响应的持续压制。

设输入特征为$X \in \mathbb{R}^{C \times H \times W \times B}$，记$P$为基础通道分量，$A$为待进行高阶分解的门控特征块，$K$为高阶分解阶数，$G_i$为按通道划分得到的第$i$阶门控核，$\mathrm{Conv}^{(i)}(\cdot)$为第$i$个逐点卷积映射，$\gamma$为可学习残差缩放系数。SGNC模块首先通过1×1卷积在通道维度进行线性投影，得到：

$$U = \mathrm{Conv}_{1\times 1}(X), U \in \mathbb{R}^{2C \times H \times W \times B} \tag{8}$$

并将其划分为基础分量和高阶门控分量。

$$P \in \mathbb{R}^{d_0 \times H \times W \times B},\quad A \in \mathbb{R}^{(\sum_{k=0}^{K-1} d_k) \times H \times W \times B} \tag{9}$$

其中，P用于保留基础特征，A由若干高阶通道块构成，按预设阶数K进行分解。

随后，将高阶通道块A输入1.2.1节的FSC模块，完成局部深度卷积与可学习频域滤波的联合建模，得到频域–空间协同增强后的特征。

$$\hat{A} = \mathrm{FSC}(A) \tag{10}$$

再按通道维数将$\hat{A}$划分为多阶门控核。

$$\hat{A} = [G_0, G_1, \cdots, G_{K-1}],\quad G_k \in \mathbb{R}^{d_k \times H \times W \times B} \tag{11}$$

在此基础上，SGNC模块采用卷积加门控的逐级结构实现高阶特征组合。首先，利用第0阶门控核对基础分量进行初始重标定。

$$Z_0 = P \odot G_0 \tag{12}$$

然后，依次通过逐点卷积与对应门控核，得到输出。

$$Z_k = \mathrm{Conv}_{1\times 1}^{(k)}(Z_{k-1}) \odot G_k,\quad k = 1,2,\cdots,K-1 \tag{13}$$

这种逐级门控形式可以理解为在通道维上构造高阶非线性叠加，使网络能对小目标、弱对比区域中的细微差异进行更精细的放大，从而降低噪声干扰。

最后，SGNC模块将最高阶输出映射回原始通道维数，并结合归一化与残差连接得到模块输出。

$$\hat{Y} = \mathrm{LayerNorm}(\tilde{Y}),\quad Y = X + \gamma \hat{Y} \tag{14}$$

$$\tilde{Y} = \mathrm{Conv}_{1\times 1}^{\mathrm{out}}(Z_{k-1}) \tag{15}$$

其中，γ为可学习残差缩放系数，初始接近0，以保证网络在训练初期平滑过渡。

通过将FSC模块的输出显式构造为多阶门控核，SGNC模块能够在深层通道交互过程中持续传递频域–空间抑噪先验，从而增强弱小目标判别特征，并降低噪声伪响应在层级传播中的累积放大。SGNC模块网络结构如图3所示。

1.3 通道感知全局上下文模块

红外遥感图像中容易出现大面积的复杂背景区域，这些区域会掩盖目标的纹理与轮廓，导致模型产生误检或漏检，且仅依赖局部卷积容易忽略全局

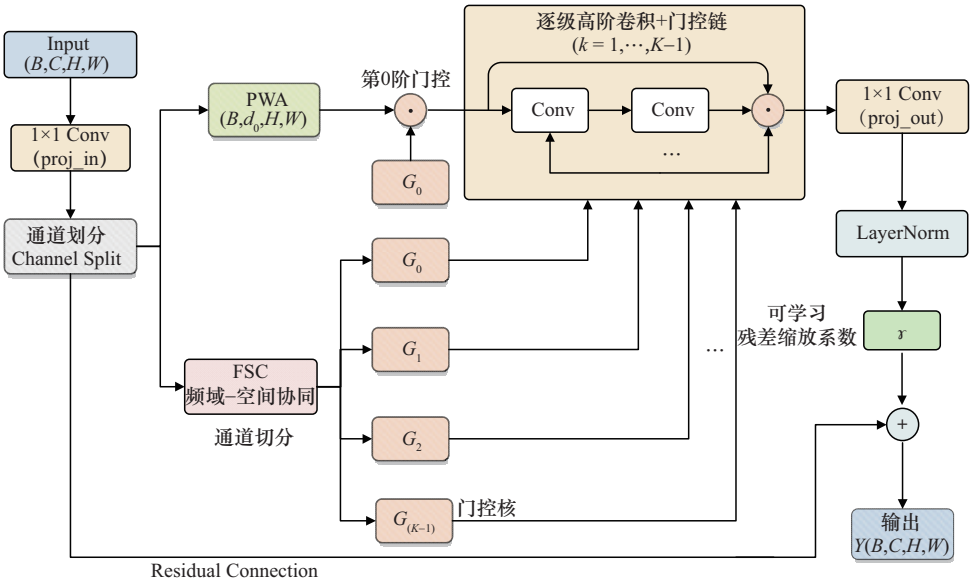


图3 SGNC模块网络结构

语义关联，难以有效区分目标相关通道与受背景干扰的通道。为此，本文设计了CGC模块，该模块通过注意力池化获取通道级全局均值上下文，并结合方差统计刻画通道波动程度，在此基础上生成通道调制系数，以抑制高波动背景通道并增强目标相关通道，从而提升复杂背景下的判别一致性，模块网络结构如图4所示。

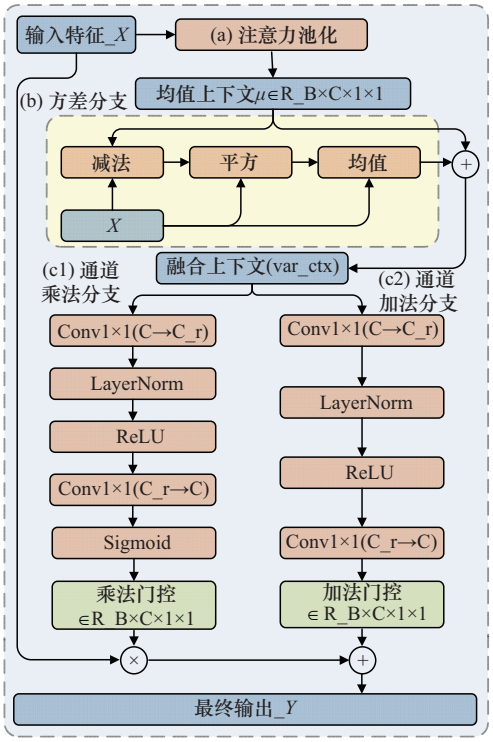


图4　CGC模块网络结构

具体而言，设输入特征为$X \in \mathbb{R}^{C \times H \times W \times B}$，记Ctxmean为注意力池化得到的通道级全局均值上下文，Ctxvar为基于均值参考计算的通道方差上下文，λ为可学习融合系数，Ctx为融合后的最终上下文表示。

CGC模块首先通过注意力池化获得通道级全局均值上下文，利用1×1卷积生成空间权重图，并经过均匀先验与Softmax融合，得到更平滑的注意力分布$\tilde{\alpha}_{b,i,j}$。在此基础上，对每个通道做加权求和，得到：

$$\mathrm{meanCtx}_{b,c} = \sum_{i=1}^{H}\sum_{j=1}^{W} \tilde{\alpha}_{b,i,j} X_{b,c,i,j} \tag{16}$$

从而得到反映整体响应强度的通道均值向量$\mathrm{meanCtx}_{b,c}$。

为了刻画通道在全局范围内的稳定性，CGC模块进一步以$\mathrm{meanCtx}_{b,c}$为参考，计算对应的通道方差为：

$$\mathrm{varCtx}_{b,c} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{b,c,i,j} - \mathrm{meanCtx}_{b,c}\right)^2 \tag{17}$$

其中，方差越大表明该通道在整幅图像上波动更明显，在复杂背景区域中更可能包含不稳定响应。

随后，通过可学习的标量参数γ将通道级全局均值上下文与方差上下文进行融合，得到最终的通道级上下文表示。

$$\mathrm{Ctx}_{b,c} = \mathrm{meanCtx}_{b,c} + \lambda \mathrm{varCtx}_{b,c} \tag{18}$$

在上下文向量Ctx的基础上，CGC模块通过两条轻量的1×1卷积分支分别生成乘性权重和加性偏置，最终输出特征为：

$$Y = X \odot \sigma\left(f_{\mathrm{mul}}(\mathrm{Ctx})\right) + f_{\mathrm{add}}(\mathrm{Ctx}) \tag{19}$$

其中，$f_{\mathrm{mul}}(\cdot)$和$f_{\mathrm{add}}(\cdot)$分别表示乘性调制分支与加性补偿分支，二者均由1×1卷积与非线性激活构成。

CGC模块通过均值−方差联合统计实现通道稳定性建模，能够在全局范围内抑制由复杂背景引起的高波动通道，并增强目标相关通道响应，从而提升复杂背景下的特征判别性。

1.4　高效通道预筛注意力模块

在红外遥感图像中，由于热噪声及环境辐射的影响，特征图中存在大量与目标无关的冗余响应。为了进一步降低网络负担，本文设计了ECPA模块。

该模块仅依赖全局池化和一维卷积即可完成通道权重的自适应学习，结构简单、开销较低，适合在主干网络中多层重复使用。

设输入特征为$X \in \mathbb{R}^{C \times H \times W \times B}$，首先在空间维度上进行全局统计，得到每个通道的均值与方差。

$$\mu_{b,c} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{b,c,i,j} \tag{20}$$

$$\sigma_{b,c}^2 = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{b,c,i,j} - \mu_{b,c}\right)^2 \tag{21}$$

其中，$\mu_{b,c}$表示第b个样本第c个通道的平均响应，$\sigma_{b,c}^2$刻画该通道在整幅图像上的波动程度。为增强数值稳定性并突出高噪声通道，ECPA模块在对数域中对方差进行建模。

$$v_{b,c} = -\lg\left(\sigma_{b,c}^2 + \varepsilon\right) \tag{22}$$

其中，ε为极小正数。将$\mu_b = \left[\mu_{b,1},\mu_{b,2},\cdots,\mu_{b,C}\right]$,$v_b = \left[v_{b,1},v_{b,2},\cdots,v_{b,C}\right]$视为长度为$C$的一维序列，经由两条

一维卷积分支在通道维上建模局部依赖关系。

$$\tilde{\mu}_b = \mathrm{Conv}_{\mu}(\mu_b),\quad \tilde{v}_b = \mathrm{Conv}_{v}(v_b) \tag{23}$$

其中，Conv_{μ}与Conv_{v}具有相同结构但参数不共享，卷积核大小k随通道数自适应确定。

随后，ECPA模块通过少量可学习系数对均值与方差分支进行感知融合。记$\alpha,\beta \in \mathbb{R}^C$为逐通道融合参数，则融合结果为：

$$Z_b = \alpha \odot \tilde{\mu}_b - \beta \odot \tilde{v}_b \tag{24}$$

其中，$\odot$表示逐通道乘法。α倾向于保留整体响应较高的通道，β用于抑制方差较大、波动剧烈的噪声通道。经Sigmoid激活后得到归一化通道权重。

$$S_b = \sigma(Z_b), S_b \in (0,1)^C \tag{25}$$

再将其映射到特征图上，完成通道重标定。

$$Y_{b,c,i,j} = S_{b,c} \times X_{b,c,i,j} \tag{26}$$

ECPA模块通过通道均值与对数方差建模实现轻量级通道预筛，可在较低计算开销下抑制不稳定冗余响应，并为后续频域–空间协同增强模块提供更高信噪比的输入特征，模块网络结构如图5所示。

2　实验设置与评价指标

2.1　实验环境与配置

实验平台为Windows 11专业版。该平台的CPU型号为Intel(R) Core(TM) i5-14600KF (3.50 GHz)，GPU型号为NVIDIA GeForce RTX 4070 TiSUPER (16 GB)，CUDA版本为12.8，平台环境为Python3.10.18，PyTorch版本为2.7.1。具体实验参数设置如表1所示。

表1　　实验参数设置

参数	值
迭代次数/次	300
学习率	0.001
动量	0.937
批大小	16
优化器	SGD
输入分辨率	640×640

2.2　训练细节说明

模型训练基于Ultralytics框架实现。为保证对比公平性，所有方法均在相同输入分辨率、批大小、优化器和训练轮次下进行训练。训练阶段采用YOLO11官方默认的增强策略，包括Mosaic、随机翻转和尺度缩放，学习率调度采用余弦退火策略，推理阶段统一使用相同的置信度阈值与非极大值抑制（non-maximum suppression, NMS）阈值。除特别说明外，其余参数均保持默认设置。

2.3　实验数据集

本文采用VEDAI-IR、HIT-UAV[13]、IRSTD-1K和SIRST数据集进行实验。VEDAI-IR是在VEDAI[14]（vehicle detection in aerial imagery）基础上重整得到，包含1 268张红外影像和3 700多个已注释的目标，数据集有8个类别的交通工具。按8:1:1的比例划分为训练集、验证集和测试集，并且得到的结果已多次取平均值。HIT-UAV包含2 898张无人机红外图像，涵盖行人、汽车、自行车及其他车辆等类别，目标总数约2.5万个，严格按照官方设

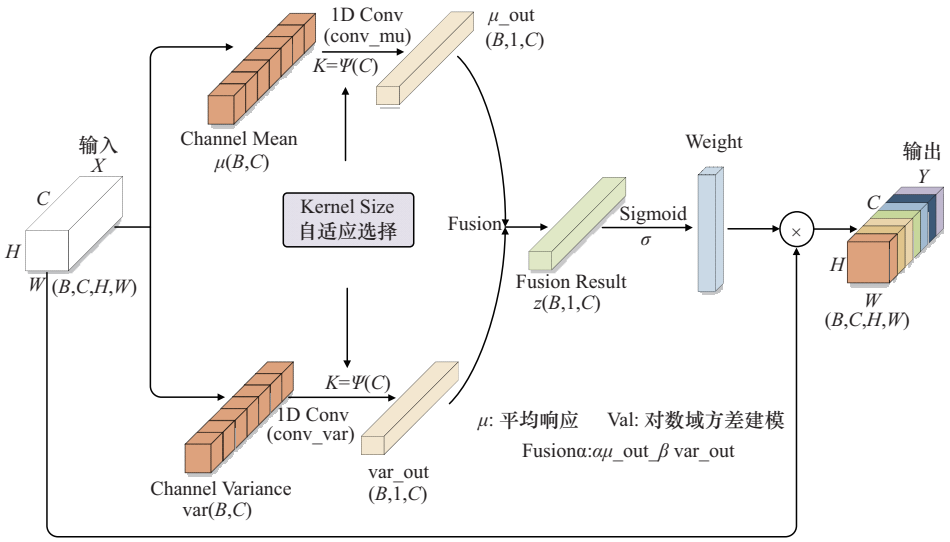


图5　ECPA模块网络结构

置的8:1:1比例划分为训练集、验证集和测试集。

2.4　极端工况测试集构造与评价指标

为进一步验证模型在低信噪比与信息缺失条件下的性能，本文在原始测试集基础上额外构造了强噪声测试子集与严重遮挡测试子集，用于评估模型在极端工况下的性能。需要说明的是，为避免与原始测试集结果混淆，本文将极端工况样本作为独立测试子集单独评估，不与原始测试集混合统计。

在强噪声场景构造中，本文以VEDAI-IR测试集为基础，选取50张图像构造低信噪比测试子集。噪声类型采用高斯噪声，对应的信噪比约为18 dB。在严重遮挡场景构造中，在相同测试集上构造遮挡子集，选取50张图像构造严重遮挡场景测试子集。遮挡方式采用目标区域遮挡，遮挡率设置为80%，遮挡区域相对于目标区域面积进行计算。

在评价方式上，本文对原始测试集、强噪声测试子集及严重遮挡测试子集分别进行独立测试，统一采用mAP@0.5、mAP@0.5:0.95、精确率（Precision）和召回率（Recall）作为评价指标。极端工况实验均在与常规实验一致的推理设置下完成，包括相同输入分辨率、置信度阈值和NMS阈值，以保证对比的公平性。为保证评估结果的稳健性，本文采用K折交叉验证法进行实验。

为进一步评估模型在天气和成像退化条件下的鲁棒性，本文基于VEDAI-IR测试集构造雨天退化、雾天退化和夜间低照度退化3类测试子集。雨天退化通过叠加方向性雨线和轻微模糊模拟降雨条纹干扰；雾天退化基于大气散射模型模拟对比度下降和雾化效应；夜间低照度退化通过Gamma变换、亮度压缩和弱噪声模拟低照度成像质量下降。所有退化样本仅用于测试阶段，不参与模型训练，且由于退化处理不改变目标位置，原始标注框保持不变。

2.5　模型评价指标

为更全面地衡量模型性能，该实验从检测准确性、定位精度以及模型复杂度等多个角度入手，选取了精确率、召回率、平均精度均值、参数量（Params）、计算量（GFLOPs）、推理帧率（FPS）以及单张图像平均推理时延（Latency）等关键指标进行综合评估。其中，精确率、平均精度均值（mean average precision, mAP）定义分别为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{27}$$

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{28}$$

3　算法分析与对比

3.1　模型消融实验

为验证SSCE-YOLO算法中各关键模块的有效性，本文以YOLO11s为基线，在VEDAI-IR数据集上分别对ECPA、SGNC和CGC模块进行消融实验，结果如表2所示。其中，“√”表示启用对应模块，其余训练参数、输入分辨率和测试设置均保持一致。

由表2可知，单独引入ECPA、SGNC或CGC模块后，模型性能均较基线有所提升。其中，引入ECPA模块后，mAP@0.5由59.2提升至61.4，mAP@0.5:0.95由38.0提升至38.8，同时Params和GFLOPs分别由9.42×10^6和21.3降至8.04×10^6和20.2。说明ECPA模块能够通过通道预筛抑制冗余响应，在提升特征信噪比的同时降低模型复杂度。引入SGNC模块后，mAP@0.5提升至62.0，Recall提升至59.0%，表明频域–空间门控高阶卷积有助于增强弱小目标和低对比目标的检出能力。引入CGC模块后，mAP@0.5提升至62.7，mAP@0.5:0.95提升至39.4，Precision提升至70.0%，说明通道感知

表2　模型消融实验（VEDAI-IR数据集）

ECPA	SGNC	CGC	mAP@0.5	mAP@0.5:0.95	Precision	Recall	Params	GFLOPs	FPS	Latency/ms
			59.2	38.0	61.4%	56.4%	9.42×10^6	21.3	476.2	2.1
√			61.4	38.8	67.3%	57.2%	**8.04×10^6**	**20.2**	**476.2**	**2.1**
	√		62.0	38.4	64.8%	59.0%	9.10×10^6	21.0	416.7	2.4
		√	62.7	39.4	70.0%	56.6%	8.17×10^6	20.3	434.8	2.3
√	√		62.4	38.8	66.3%	59.0%	9.10×10^6	21.0	434.8	2.3
	√	√	61.8	38.5	67.0%	58.5%	9.23×10^6	21.1	357.1	2.8
√		√	62.3	38.7	67.2%	56.8%	9.23×10^6	20.3	416.7	2.4
√	√	√	**63.7**	**39.6**	**71.7%**	**59.4%**	9.22×10^6	20.8	344.8	2.9

全局上下文能够有效抑制复杂背景引起的误检。

从组合实验结果来看，任意两个模块联合使用均能取得一定增益，但整体性能低于完整模型。当ECPA、SGNC和CGC模块同时启用时，模型取得最优结果，mAP@0.5、mAP@0.5:0.95、Precision和Recall分别达到63.7、39.6、71.7%和59.4%，相比YOLO11s基线分别提升4.5、1.6、10.3和3.0个百分点。同时，完整模型的Params为9.22×10^6，GFLOPs为20.8，仍低于YOLO11s基线，说明本文方法在提升检测精度和召回能力的同时，保持了较好的模型复杂度。

表3给出了典型注意力、频域建模和全局上下文模块的替换对比结果。可以看出，CBAM、Coordinate Attention、GC Block、FcaNet和GFNet均能带来一定性能提升，说明空间注意力、频域建模和全局上下文聚合对红外遥感检测具有积极作用。然而，这些模块大多侧重单一增强路径，在mAP@0.5:0.95或Recall上仍低于本文方法。相比之下，SSCE-YOLO在mAP@0.5、mAP@0.5:0.95和Recall上取得更优结果。结果表明，本文方法性能提升并非源于单一通用注意力或频域模块的简单替换，而是得益于面向红外遥感复杂场景的层次化协同增强机制。

此外，为进一步分析频域门控核与高阶门控结构的贡献，本文设置了两组对照实验（表4）：FSC-Only（仅保留FSC模块的频域–空间门控分支）与SGNC-w/o-FSC（保留SGNC，但将其门控核退化为常规卷积核）。表4表明，仅保留FSC或不含FSC门控的SGNC均可带来一定性能提升，但综合收益不及完整SGNC。其中，FSC-Only在HIT-UAV数据集上精度提升更明显，而SGNC-w/o-FSC在VEDAI-IR数据集上对mAP@0.5:0.95略有改善，但Precision和Recall不及完整方案。结果说明，仅依赖频域子带选择或高阶通道重组均难以同时兼顾噪声抑制与背景鲁棒性，只有将FSC门控核显式注入SGNC，才能形成更完整的协同增强闭环。

为进一步分析SSCE插入位置对检测性能的影响，本文设置了4组对照实验，如表5所示。表5表明，将SSCE模块同时插入Layer 6和Layer 8时，在VEDAI-IR与HIT-UAV数据集上均取得最佳结果，

表3 典型增强模块替换对比实验

方法	模块类型	mAP@0.5	mAP@0.5:0.95	Precision	Recall	Params	GFLOPs
YOLO11s	Baseline	59.2	38.0	61.4%	56.4%	9.42×10^6	21.3
+ CBAM	通道/空间注意力	60.0	38.4	66.8%	52.6%	8.02×10^6	19.4
+ Coordinate Attention	位置注意力	62.3	38.7	66.9%	55.9%	**7.72×10^6**	19.2
+ GC Block	全局上下文	63.2	38.2	**75.3%**	54.2%	7.78×10^6	19.2
+ FcaNet	频域通道注意力	61.7	39.2	66.7%	56.5%	7.73×10^6	19.2
+ GFNet	频域全局滤波	60.7	37.9	64.3%	59.0%	7.86×10^6	**19.1**
+ SSCE	本文协同增强	**63.7**	**39.6**	71.7%	**59.4%**	9.22×10^6	20.8

表4 FSC模块结构贡献验证

方法	VEDAI-IR				HIT-UAV			
	mAP@0.5	mAP@0.5:0.95	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Precision	Recall
FSC-Only	60.5	37.3	65.0%	56.5%	81.5	54.1	87.6%	74.4%
SGNC-w/o-FSC	61.7	38.4	61.9%	61.0%	80.2	55.7	84.2%	72.8%

表5 SSCE模块插入位置验证

SSCE插入位置	VEDAI-IR				HIT-UAV			
	mAP@0.5	mAP@0.5:0.95	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Precision	Recall
Layer 6&Layer 8	**63.7**	**39.6**	**71.7%**	**59.4%**	**82.2**	**55.0**	**86.1%**	**77.2%**
Layer 6	63.2	38.6	63.6%	60.8%	81.1	53.9	82.6%	76.4%
Layer 8	61.8	39.2	67.9%	55.7%	80.8	54.8	80.7%	74.3%
Layer 13	60.8	36.6	63.1%	58.2%	79.1	52.6	74.8%	77.5%

整体优于仅在单一阶段或更深层插入的方案。其原因在于，Layer 6对应中层高分辨率特征，利于提前抑制噪声并保留纹理细节；Layer 8对应更高层语义特征，利于稳定跨尺度聚合后的目标响应。两处协同插入可兼顾小目标细节增强与高层语义约束，从而获得更稳定的综合收益。并且根据YOLO11的网络配置，SSCE插入位置分别对应P4/16与P5/32两级特征提取阶段，通过在中高层语义中实现频域–空间协同增强，提升复杂背景下的目标检测能力。

3.2　参数敏感性实验

表6表明，当FSC频谱权重插值系数取1、SGNC高阶分解阶数取3时模型性能最优。其原因在于，频谱权重插值系数决定了FSC对不同频段的调制粒度。当插值系数较小时，频谱权重分辨率较低，容易造成频域调制过于粗糙，难以区分目标边缘相关频段和噪声相关频段。红外遥感图像中弱小目标通常具有低对比度和边缘响应弱等特点，而热噪声和复杂背景纹理容易产生中高频伪响应，因此更细粒度的频谱权重有助于保留目标相关结构并抑制离散噪声响应。当插值系数为1时，频谱权重与输入频率分布匹配更充分，模型取得较优性能。

对于SGNC模块，高阶分解阶数K决定了通道间高阶交互的表达能力。当K较小时，通道组合关系较弱，难以充分建模弱小目标的跨通道判别信息；当K增大后，多阶门控能够增强非线性表达能力，并使FSC模块提供的频域–空间抑噪先验在深层交互中持续传递。已有高阶门控卷积研究表明，递归门控机制能够实现高阶空间交互并增强视觉特征表达。但当K过大时，多级乘法门控可能增加优化难度并放大不稳定响应。因此，K=3在表达能力和训练稳定性之间取得较好平衡。

3.3　小目标专项评测

为验证所提方法对弱小目标的检测能力，本文开展了按目标尺度划分的小目标专项评测，且本文参考COCO目标尺度划分标准，以目标标注框面积作为判定依据。设目标框宽度和高度分别为w和h，当目标面积满足$A = wh < 32^2$时，将该目标定义为小目标，并据此计算mAP@small和Recall@small。结果如表7所示。可以看出，SSCE-YOLO在小目标检测指标上取得了最佳结果，mAP@small达到37.0、Recall@small达到53.3，显著优于其他方法，说明所提频域–空间协同增强机制能够有效提升弱小目标的分辨能力与检出能力。

表7　小目标专项评测（VEDAI-IR数据集）

方法	mAP@small	Recall@small
YOLOv8s	27.0	47.2
YOLOv10s	26.9	<u>50.9</u>
YOLO11s	28.0	45.3
YOLO12s	<u>30.9</u>	46.6
SSCE-YOLO	**37.0**	**53.3**

3.4　频谱响应可视化与机理分析

为进一步分析FSC模块中可学习复数频谱权重的作用机理，本文对FSC模块输入特征频谱、学习到的复数频谱权重、频域分支响应以及最终输出频谱差异进行可视化，如图6所示。图6中低频位于频谱中心，高频位于外侧区域。需要说明的是，高频区域并不完全等同于噪声，其同时包含目标边缘、细碎纹理和随机噪声等快速变化成分。从图6可以看出，FSC频谱权重呈现非均匀分布，说明模型学习到了具有频段选择性的静态调制模式。由于FSC模块中的复数频谱权重为训练后得到的参数，而非针对每张图像动态生成的权重，因此不同样本中的权重分布整体保持一致。滤波分支响应表明，频域分支主要在部分结构化频率区域产生响应；输出频谱进一步显示，残差门控融合后，FSC模块对输入特征的频谱调整主要集中在低中频结构响应区域和部分中高频离散响应区域。结果说明，FSC并非简单滤除高频信息，而是通过可学习复数频谱权重对噪声相关频段和目标结构频段进行差异化调制。

表6　参数敏感性实验（VEDAI-IR数据集）

性能指标	FSC频谱权重插值系数				SGNC高阶分解阶数			
	0.25	0.5	0.75	1.0	1	2	3	4
mAP@0.5	62.4	63.3	63.3	**63.7**	62.6	62.8	**63.7**	63.3
mAP@0.5:0.95	39.4	39.1	39.0	**39.6**	38.0	38.6	**39.6**	39.5
Precision	68.2%	67.1%	61.8%	**71.7%**	61.9%	67.2%	**71.7%**	64.8%
Recall	58.3%	56.3%	59.8%	**59.4%**	59.4%	57.7%	**59.4%**	58.1%

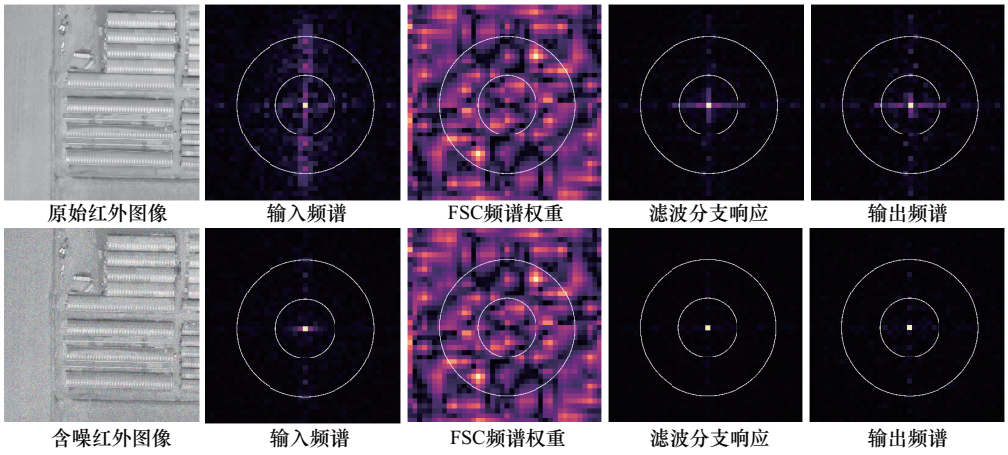


图6　FSC频谱权重与特征频谱响应可视化

3.5　频谱相位与通道统计机理分析

为分析FSC模块中复数频谱权重的幅相调制作用，并验证ECPA与CGC模块中全局均值–方差统计对弱小目标响应的影响，本文从频谱相位建模和通道特征分布两方面进行补充分析。一方面，结合表8分析不同频谱权重参数化方式在mAP、Precision和Recall上的性能差异，以验证实虚联合建模相较于仅幅度调制和幅相解耦调制的综合检测稳定性。另一方面，通过表9中的去方差统计实验，分析全局方差统计对弱小目标召回能力的影响。

对于FSC频域分支，本文中插值对象为可学习复数频谱权重W，而非输入特征频谱F本身。对W的实部与虚部分别进行双线性插值，等价于在复平面内对邻域复权重进行线性组合，因此不会改变输入频谱的采样坐标，同时可避免由输入频谱重采样引起的频率混叠问题。若将插值后的复权重表示为$Me^{j\theta}$，其作用可理解为对输入频谱幅度的加权调制和对相位的轻量补偿。为减小相位扰动带来的潜在影响，FSC模块通过门控融合和残差连接控制频域分支的修正幅度，从而降低振铃伪影风险。

表8给出了不同频谱权重参数化方式的对比实验结果。其中，Amp-only采用实值幅度权重，仅对输入频谱幅度进行调制，不显式改变相位结构；Amp-Phase分别建模幅度权重和相位补偿项，用于分析幅度调制与相位补偿的解耦作用；Real-Imag则采用复数权重，并通过实部与虚部联合插值实现频谱调制。由表8可知，Amp-only在mAP@0.5上达到64.8，说明仅对频谱幅度进行调制能够提升目标检出能力；Amp-Phase在mAP@0.5:0.95上达到39.7，表明显式相位补偿对定位精度有一定帮助。相比之下，Real-Imag在mAP指标上并非最高，但其Precision和Recall分别达到71.7%和59.4%，均优于另外两种方式，说明实虚联合建模在误检抑制和目标召回之间取得了更稳定的平衡。结合图6中的频谱响应可视化可以看出，FSC模块输出频谱主要集中于低中频结构响应区域和部分中高频离散响应区域，未出现大范围异常高频扩散，表明Real-Imag在引入幅相联合调制能力的同时，仍能够通过残差门控和局部分支保持较稳定的空间响应。

表8　不同频谱权重参数化方式对比实验

方法	mAP@0.5	mAP@0.5:0.95	Precision	Recall
Amp-only	64.8	39.5	69.8%	58.6%
Amp-Phase	64.6	39.7	66.7%	58.1%
Real-Imag	63.7	39.6	71.7%	59.4%

表9给出了方差统计对弱小目标召回率的影响。其中，Rsmall@0.5表示弱小目标样本在交并比（intersection over union, IoU）阈值为0.5时的召回率，且与前文Recall@small的统计范围和实验设置不同，因此不作直接比较。由表9可知，当仅保留均值统计时，模型Precision、Recall和Rsmall@0.5分别为70.4%、57.6%和75.7%；引入方差统计后，上述指标分别提升至71.7%、59.4%和85.2%。其中，Rsmall@0.5提升最为明显，说明方差统计有助于提升弱小目标召回稳定性。从特征分布角度看，弱小目标响应通常具有局部稀疏性，其激活区域在整幅特征图中占比较小；而复杂背景干扰更容易在道路纹理、建筑边缘、云层和地表结构等区域产生较大范围的持续波动。均值-方差联合统计能够同时刻画通道响应强度与全局波动程度，并通过可学习参数进行软调制，而非直

接依据方差大小硬性抑制通道。因此，该统计方式有助于增强模型对局部目标响应和广域背景扰动的区分能力，从而提升弱小目标召回稳定性。

表 9　方差统计对弱小目标召回率的影响

方法	Mean	Var	Precision	Recall	Rsmall@0.5
SSCE-w/o-Var	√	×	70.4%	57.6%	75.7%
SSCE-YOLO	√	√	**71.7%**	**59.4%**	**85.2%**

3.6　极端工况与天气退化实验结果分析

由表 10 可知，SSCE-YOLO 在原始噪声和原始遮挡场景下的 mAP@0.5 分别为 79.7 和 83.5，均高于 YOLO11s；在强噪声场景下，其 mAP@0.5、mAP@0.5:0.95 和 Recall 分别达到 24.3、11.6 和 33.7%；在严重遮挡场景下，其 mAP@0.5、mAP@0.5:0.95、Precision 和 Recall 分别达到 31.1、18.5、47.1% 和 35.9%，均优于 YOLO11s，说明本文方法在强噪声和严重遮挡场景下仍具有更好的检测鲁棒性。

在雨天、雾天和夜间低照度三类退化场景下，SSCE-YOLO 的 mAP@0.5 和 mAP@0.5:0.95 均高于 YOLO11s，说明本文方法在退化场景下保持了更好的总体检测精度。雨天退化场景下 SSCE-YOLO 的 Recall 略低于 YOLO11s，说明方向性雨线可能仍会对部分弱小目标召回造成影响，但其 Precision 和 mAP 指标更高，表明模型在抑制误检和保持整体检测性能方面具有一定优势。在雾天退化和夜间低照度退化场景下，SSCE-YOLO 在 mAP 和 Recall 上均取得更明显提升，说明频域–空间协同增强机制能够缓解雾化低对比和低照度弱响应造成的性能下降。

3.7　模型对比实验

为全面验证 SSCE-YOLO 的检测性能，本文在 HIT-UAV 与 VEDAI-IR 两个数据集上与典型目标检测方法进行对比。对比方法包括两阶段检测器 Faster R-CNN、一阶段检测器 RetinaNet、Transformer 类检测器 RT-DETR，以及 YOLO 系列轻量模型和近年来红外小目标检测方法。对于具备公开实现且可复现的方法，本文尽可能在相同数据划分和训练设置下进行复现实验。对于缺少完整复现条件的方法，则引用原文报告结果，并以“—”表示未公开指标。评价指标包括 mAP@0.5、mAP@0.5:0.95、Precision、Recall、Params 和 GFLOPs。

在 HIT-UAV 数据集上（表 11），SSCE-YOLO 取得 82.2 的 mAP@0.5 和 55.0 的 mAP@0.5:0.95，均优于对比方法。与 YOLO11s 相比，SSCE-YOLO 的 mAP@0.5 提升 3.1 个百分点，Recall 由 75.1% 提升至 77.2%，同时 Params 和 GFLOPs 基本保持不变。

与 RT-DETR-R50 相比，SSCE-YOLO 在显著降低参数量和计算量的情况下仍取得更高的检测精度。与 GSFANet、HDNet 等红外小目标检测方法相比，SSCE-YOLO 在 mAP@0.5 和 Recall 上也表现更优，说明所提频域–空间协同增强机制对无人机红外场景中的小目标和复杂背景具有较好的适应性。

在 VEDAI-IR 数据集上（表 12），SSCE-YOLO 取得 63.7 的 mAP@0.5 和 39.6 的 mAP@0.5:0.95，同样达到最优结果。相比 YOLO11s，SSCE-YOLO 的 mAP@0.5 提升 4.5 个百分点，在 Params 和 GFLOPs 低于基线的情况下取得更高检测精度，说明本文方法能够在控制模型复杂度的同时提升红外遥感目标检测性能。与 SuperYOLO、GSFANet 和 HDNet 等方法相比，SSCE-YOLO 在总体精度和召回能力上具有一定优势，表明其在复杂背景和低对比度目标场景下具有更稳定的检测能力。

此外，为验证所提方法在红外小目标检测任务中的泛化能力，本文在 SIRST 和 IRSTD-1K 数据集上开展补充实验，结果如表 13 所示。需要说明的

表 10　极端工况和天气退化条件下的实验结果对比

场景	YOLO11s				SSCE-YOLO			
	mAP@0.5	mAP@0.5:0.95	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Precision	Recall
原始噪声	75.5	57.8	67.9%	70.2%	79.7	61.0	68.2%	74.4%
强噪声	20.4	8.1	78.9%	27.9%	24.3	11.6	89.4%	33.7%
原始遮挡	79.1	50.2	73.6%	73.5%	83.5	53.7	71.0%	83.2%
严重遮挡	25.4	16.4	29.3%	30.9%	31.1	18.5	47.1%	35.9%
雨天退化	43.5	25.6	51.7%	42.3%	43.9	26.6	55.7%	37.8%
雾天退化	47.8	30.7	58.5%	42.9%	50.6	31.3	59.4%	47.8%
夜间低照度退化	43.2	23.9	55.8%	40.3%	48.4	27.4	52.0%	51.0%

表 11 不同方法在 HIT-UAV 数据集上的实验对比结果

方法	期刊&年份	mAP@0.5	mAP@0.5:0.95	Precision	Recall	Params	GFLOPs
Faster R-CNN[15]	NIPS2015	70.2	40.2	79.7%	67.7%	41.3×10^6	168
RetinaNet[16]	ICCV2017	77.4	46.2	77.3%	73.9%	21.5×10^6	151.5
Deformable DETR[17]	ICLR2021	77.1	45.9	75.0%	71.7%	40.1×10^6	154
TOOD[18]	ICCV2021	76.6	—	75.2%	70.5%	32.0×10^6	158
YOLOv8s	2023	76.0	50.1	84.7%	68.8%	11.1×10^6	28.4
RT-DETR[19]	CVPR2024	78.8	49.6	81.7%	76.5%	42.8×10^6	57.0
RT-DETR-R50[19]	CVPR2024	80.3	51.0	89.9%	76.1%	42.0×10^6	130
EFLNET[20]	TGRS2024	64.1	—	69.9%	61.2%	—	—
YOLOv10s	NIPS2024	77.8	51.7	85.7%	67.9%	8.07×10^6	21.4
YOLO11s	2024	79.1	54.3	79.6%	75.1%	9.42×10^6	21.3
YOLO12s	2025	79.5	53.1	82.1%	75.4%	9.23×10^6	21.2
PWL-RTDETR[21]	Sci.Rep.2025	81.3	—	**93.0%**	74.9%	10.4×10^6	21.7
CRPE-Net[22]	PR2025	76.2	—	85.1%	70.1%	—	—
GSFANet	TGRS2025	81.7	—	84.3%	75.9%	6.13×10^6	**20.5**
HDNet	TGRS2025	79.1	—	85.9%	73.5%	$\mathbf{3.66\times10^6}$	22.7
SSCE-YOLO	本文方法	**82.2**	**55.0**	86.1%	**77.2%**	9.22×10^6	20.8

表 12 不同方法在 VEDAI-IR 数据集上的实验对比结果

方法	期刊&年份	mAP@0.5	mAP@0.5:0.95	Precision	Recall	Params	GFLOPs
Faster R-CNN	NIPS2015	50.6	—	48.0%	51.0%	41.3×10^6	169.2
RetinaNet	ICCV2017	54.2	33.2	47.0%	55.4%	21.5×10^6	155.1
Deformable DETR	ICLR2021	55.6	35.2	48.4%	56.5%	40.1×10^6	162
YOLOrs[23]	J-STARS2021	54.7	32.4	41.6%	52.4%	—	—
SuperYOLO[24]	TGRS2023	62.6	38.2	65.4%	54.6%	$\mathbf{4.83\times10^6}$	**16.6**
YOLOv8s	2023	59.3	37.1	62.6%	58.1%	11.1×10^6	28.5
RT-DETR	CVPR2024	60.2	37.1	65.4%	56.0%	42.8×10^6	58.4
YOLOv10s	NIPS2024	59.9	37.1	62.0%	57.1%	8.07×10^6	21.4
YOLO11s	2024	59.2	38.0	61.4%	56.4%	9.42×10^6	21.3
DFDet	TGRS2024	53.6	—	—	—	—	—
YOLO12s	2025	60.1	36.9	65.9%	55.7%	9.23×10^6	21.2
GSFANet	TGRS2025	63.1	—	66.4%	**61.7%**	6.12×10^6	20.5
HDNet	TGRS2025	59.4	—	50.1%	59.1%	$\mathbf{3.66\times10^6}$	22.7
SSCE-YOLO	本文方法	**63.7**	**39.6**	**71.7%**	59.4%	9.22×10^6	20.8

表 13 不同方法在 SIRST、IRSTD-1K 数据集上的实验对比结果

方法	SIRST				IRSTD-1K			
	mAP@0.5	mAP@0.5:0.95	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Precision	Recall
YOLOv8s	93.5	66.9	93.1%	85.5%	82.0	40.8	80.9%	78.0%
YOLOv10s	93.0	72.1	93.8%	81.8%	80.7	41.4	83.6%	70.6%
YOLO11s	93.9	69.1	94.3%	87.0%	84.0	42.8	83.2%	77.6%
YOLO12s	93.5	62.7	92.4%	87.4%	84.5	42.2	83.9%	77.6%
SSCE-YOLO	**95.7**	**76.2**	**96.0%**	**87.5%**	**85.4**	**43.7**	**84.3%**	**78.8%**

是，SIRST 和 IRSTD-1K 主要面向单类红外弱小目标检测，与 VEDAI-IR 和 HIT-UAV 的多类别遥感目标检测任务存在差异。因此，该部分采用统一 YOLO 基线进行对比，重点考察 SSCE-YOLO 在相同训练与测试设置下的迁移能力。实验结果表明，SSCE-YOLO 在 SIRST 和 IRSTD-1K 数据集上均取得较优结果，说明所提频域–空间协同增强机制不仅适用于红外遥感多类别检测任务，也能够迁移到红外弱小目标检测场景中。

3.8　检测结果可视化分析

为进一步从定性角度分析 SSCE-YOLO 在复杂场景下的检测表现，本文从 VEDAI-IR 测试集中选取复杂背景、噪声干扰以及极端退化场景下的典型样例进行可视化对比，结果如图 7~图 9 所示，其中粗圆圈表示漏检目标，细圆圈表示误检目标。

如图 7 所示，在复杂背景场景中，地面纹理、建筑边缘和道路结构容易产生与目标相似的局部响应，导致部分 YOLO 系列模型出现误检或漏检。相比之下，SSCE-YOLO 能够更稳定地定位目标区域，并减少背景纹理引起的误检问题，说明通道预筛与全局上下文约束有助于增强目标与背景之间的区分能力。

如图 8 所示，在噪声干扰场景中，红外图像中的随机噪声会削弱目标边缘和局部结构，使弱小目标更容易被忽略。部分基线模型在该场景下出现目标漏检或将噪声伪响应误判为目标的情况。SSCE-YOLO 通过频域–空间协同滤波和深层门控传递，在抑制噪声伪响应的同时保留目标结构信息，因此

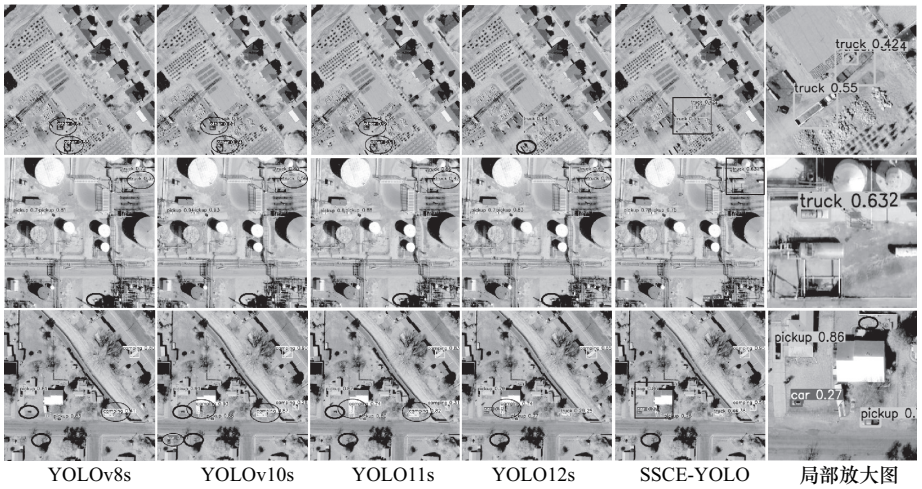


图 7　复杂背景场景下的检测结果对比

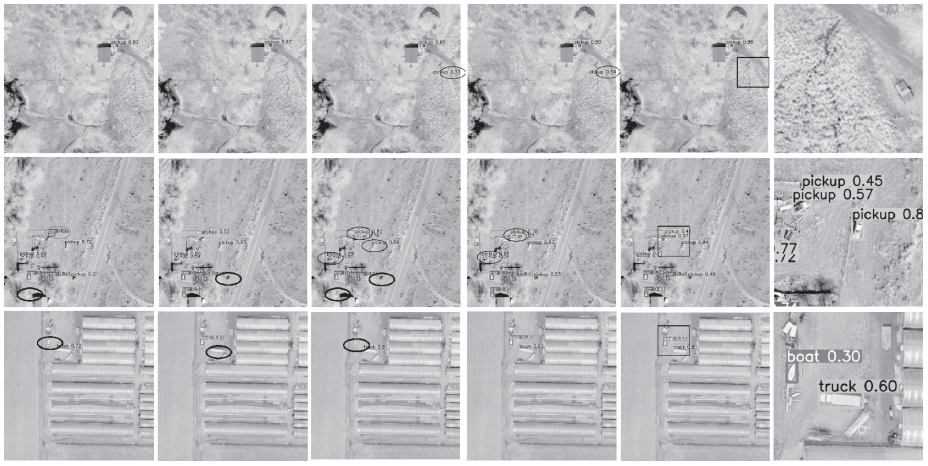


图 8　噪声干扰场景下的检测结果对比

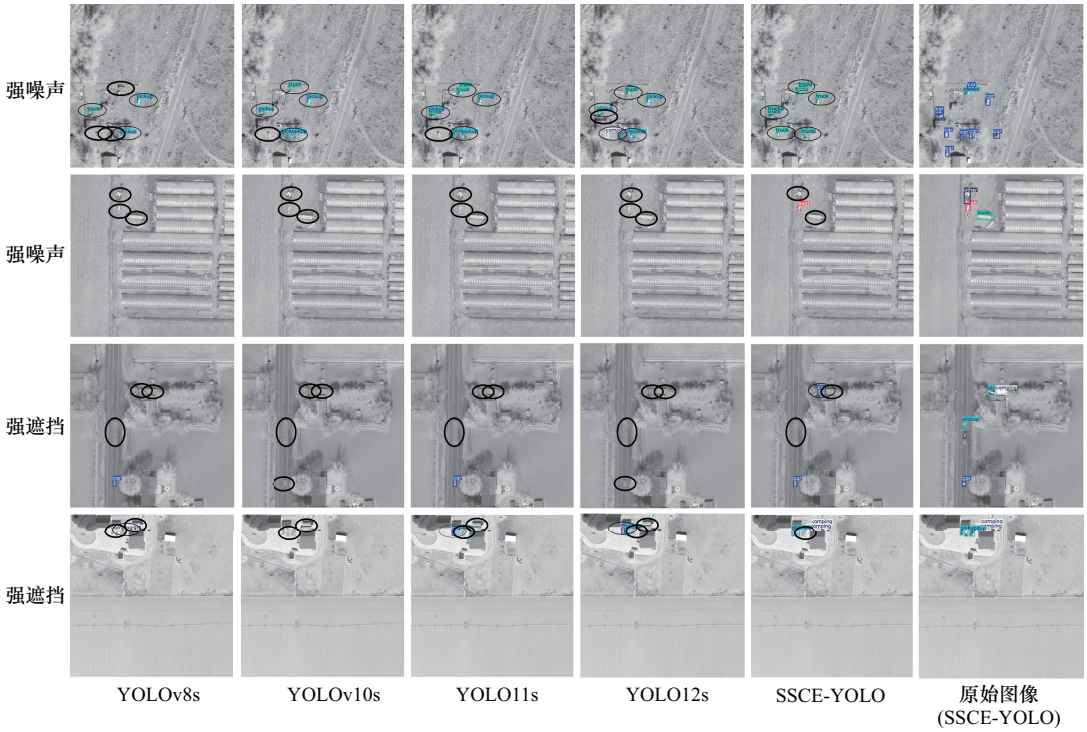


图9　强噪声、强遮挡场景下的检测结果对比

在所选样例中表现出更少的漏检和误检。

如图 9 所示，在强噪声、强遮挡等极端场景下，基线模型的检测结果更容易受到噪声伪影和目标信息缺失的影响。相比之下，SSCE-YOLO 仍能够保持较稳定的目标响应，说明本文方法在复杂退化场景下具有一定鲁棒性。

综合来看，SSCE-YOLO 在复杂背景、噪声干扰和极端退化场景中能够减少部分误检和漏检现象，进一步验证了频域–空间协同增强机制对红外遥感复杂场景目标检测的有效性。

4　结束语

本文面向红外遥感目标检测中的噪声干扰、弱小目标响应不足和复杂背景误检问题，提出了 SSCE-YOLO 算法。该方法通过频域–空间协同增强机制，将频域调制、深层门控传递和全局通道稳定性约束统一到 YOLO 检测框架中，在不显著增加模型复杂度的前提下提升了红外目标检测性能。实验结果表明，SSCE-YOLO 在 VEDAI-IR 和 HIT-UAV 数据集上的 mAP@0.5 分别达到 63.7 和 82.2，较 YOLO11s 基线分别提升 4.5 和 3.1 个百分点；在 SIRST 和 IRSTD-1K 补充实验、小目标专项评测以及强噪声、强遮挡、雨雾和低照度退化场景下也表现出较好的泛化性和鲁棒性。后续工作将进一步研究真实跨天气、多传感器和时序红外遥感场景下的稳定检测问题。

参考文献：

[1] Jiang B, Li J X, Lu Y, et al. Eficient image denoising using deep learning: a brief survey[J]. Information Fusion, 2025, 118: 103013.

[2] Wen L, Cheng Y, Fang Y, et al. A comprehensive survey of oriented object detection in remote sensing images[J]. Expert Systems with Applications, 2023, 224: 119960.

[3] 寇旗旗, 刘规, 江鹤, 等. 基于多域信息增强的轻量级图像超分辨率网络[J]. 通信学报, 2025, 46(4): 144-159.
Kou Q Q, Liu G, Jiang H, et al. Lightweight image super-resolution network based on muti-domain information enhancement[J]. Journal on Communications, 2025, 46(4): 144-159.

[4] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.

[5] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 13708-13717.

[6] Qin Z Q, Zhang P Y, Wu F, et al. FcaNet: frequency channel attention networks[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 763-772.

[7] Rao Y M, Zhao W L, Zhu Z, et al. Global filter networks for image classification[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: ACM Press, 2021:

980-993.

[8] Deng C Y, Zhao Z Y, Xu X, et al. GSFANet: global spatial-frequency attention network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-17.

[9] Xu M Z, Yu C L, Li Z X, et al. HDNet: a hybrid domain network with multiscale high-frequency information enhancement for infrared small-target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5004115.

[10] Cao Y, Xu J R, Lin S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE Press, 2019: 1971-1980.

[11] 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.

[12] Li B Y, Xiao C, Wang L G, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2023, 32: 1745-1758.

[13] Suo J S, Wang T Y, Zhang X Z, et al. HIT-UAV: a high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection[J]. Scientific Data, 2023, 10: 227.

[14] Razakarivony S, Jurie F. Vehicle detection in aerial imagery: a small target detection benchmark[J]. Journal of Visual Communication and Image Representation, 2016, 34: 187-203.

[15] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[16] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2999-3007.

[17] Zhu X Z, Su W J, Lu L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[C]//International Conference on Learning Representations (ICLR). 2021:1-16.

[18] Feng C J, Zhong Y J, Gao Y, et al. TOOD: task-aligned one-stage object detection[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 3490-3499.

[19] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 16965-16974.

[20] Yang B, Zhang X Y, Zhang J, et al. EFLNet: enhancing feature learning network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5906511.

[21] Guo C H, Luo P D, Ma Z X, et al. An efficient reparameterized small object detection transformer for thermal infrared images[J]. Scientific Reports, 2025, 15: 44599.

[22] Jiang Z W, Chen S G, Tian C N, et al. CRPE-Net: infrared small target detection transformer with cross-layer relative-position embedding[J]. Pattern Recognition, 2026, 172: 112640.

[23] Sharma M, Dhanaraj M, Karnam S, et al. YOLOrs: object detection in multimodal remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 1497-1508.

[24] Zhang J Q, Lei J, Xie W Y, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5605415.

[作者简介]



程德强（1979– ），男，博士，中国矿业大学教授、博士生导师，主要研究方向为机器视觉与模式识别、图像智能检测与信息处理等。



岳阳（2000– ），男，中国矿业大学硕士生，主要研究方向为三光融合目标检测。



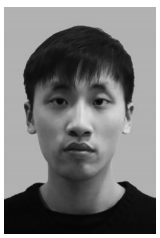
王衍辰（2000– ），男，中国矿业大学博士生，主要研究方向为三光融合目标检测。



寇旗旗（1988– ），男，博士，中国矿业大学副教授、硕士生导师，主要研究方向为图像增强与复原、智能检测与模式识别等。



徐飞翔（1992– ），男，博士，中国矿业大学副教授、博士生导师，主要研究方向为无人系统智能感知与控制、智能机器人等。



孙凤乾（1997– ），男，中国矿业大学博士生，主要研究方向为小目标检测与开放词汇检测。