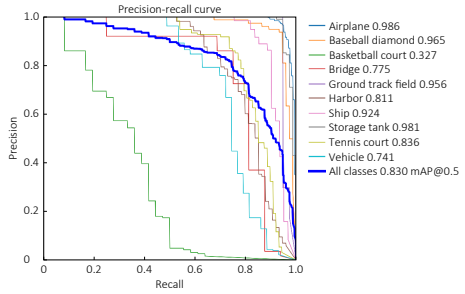
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 3 月, 第 53 卷, 第 3 期

# 融合注意力与特征增强的遥感图像小目标检测




沈学利, 郭　旭*

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105

**摘要：**为了解决遥感图像中小目标特征不明显、背景复杂及目标密集分布导致的检测遗漏问题，提出一种融合注意力与特征增强的遥感图像小目标检测算法。首先，构造多层级特征协同策略，结合通道注意力和空间注意力机制并利用门控机制动态调整信息流动，有效捕捉不同层级特征间的潜在关联，加强模型在复杂场景中的鲁棒性。其次，设计一个多尺度特征增强模块，将三维卷积和多尺度特征编码结合形成一个增强型颈部网络，优化信息提取能力与小目标的特征表达。最后，采用 NWD 结合 SIoU 作为损失函数，在提升模型训练效率的同时进一步加强小目标的定位准确性。在 NWPU VHR-10 数据集、DIOR 数据集和 RSOD 数据集上的实验结果表明，改进算法相较于原始模型，mAP@0.5 数值分别提高了 9.9%、3.1% 和 3.2%，可视化实验结果显示改进算法可以检测出原始模型漏检的小目标，与其他主流检测算法相比具有更优的检测精度。

**关键词：**遥感图像；小目标检测；多尺度；特征增强

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250267 | CSTR: 32245.14.oee.2026.250267 | 中图分类号：TP391.4 | 文献标识码：A

沈学利, 郭旭. 融合注意力与特征增强的遥感图像小目标检测 [J]. 光电工程, 2026, **53**(3): 250267

Shen X L, Guo X. Fusion attention and feature enhancement of small targets detection in remote sensing images[J]. *Opto-Electron Eng*, 2026, **53**(3): 250267

## 1　引　言

随着计算机视觉技术研究的深入，遥感图像目标检测在资源勘察、城市规划管理、灾害应急处理等领域中具有重要价值。然而，遥感图像中的目标一般存在尺度差异显著、场景复杂、细节模糊等特点，同时受成像高度和大气条件的制约，导致目标边缘不够清晰、纹理信息不足。尤其是在低分辨率影像中，目标与背景之间的区分度进一步降低，使得精准识别和检测成为一项非常困难的任务。现阶段遥感图像目标检测算法的研究已取得显著进展，但在小目标检测方面仍面临诸多挑战，交并比 (intersection over union, IoU) 是衡量预测框与真实框重叠度的重要指标，在检测微小物体时 IoU 的敏感性可能过高，目标框的细微变化也会导致 IoU 大幅度波动，使损失函数在训练过程中不稳定，影响小目标的定位精度；样本分配策略也是小目标检测的难点之一，在训练数据中小目标样本数量相对较少，导致样本类别分配不均衡，训练模型时可能会忽视小目标的特征，发生漏检现象；在密集场景中，多个目标可能重叠，同一个目标可能出现多个检测框，非极大值抑制 (non-maximum suppression, NMS) 算法会删除多余的重叠框，由于小目标之间的重叠，NMS 可能会错过小目标，造成漏检。此外，在复杂场景下，小目标常与背景中相似纹理区域混杂，难以有效定位。这使得将小目标与背景进行有效

收稿日期：2025-09-05；修回日期：2025-12-04；录用日期：2025-12-09；网络出版日期：2026-03-19

基金项目：国家自然科学基金面上项目 (62173171)

*通信作者：郭旭，guoxu9907@163.com

版权所有©2026 中国科学院光电技术研究所

区分并对小目标实现精准定位变得尤为困难。因此，提升遥感图像小目标检测的准确性并优化算法的推理效率具有重要的研究意义[1-7]。

近年来，自然图像领域的主流目标检测方法可分为单阶段与双阶段两大类。单阶段方法，如 SSD[8] (single shot multibox detector)、YOLO[9-13] (you only look once)系列，通过单次前向传播直接确定目标类型和边界框位置，具备响应快速、实时性高的优势。双阶段方法，如 Faster R-CNN[14] (faster region-based convolutional neuralnetwork)、Mask R-CNN[15] (mask region-based convolutionalneural network)，在生成候选区域后进一步执行精确的分类与定位，凭借两阶段的级联优化机制，在复杂场景中表现出较高的检测准确率。然而，遥感图像因其独特的成像视角与背景复杂度，直接套用自然图像检测方法在小目标检测任务中效果欠佳。

在近几年的研究中，基于卷积的特征融合已广泛应用于目标检测任务。Liu 等[16] 开创性地探索了不同的基于卷积神经网络的融合架构，并表明中层特征融合可以带来更好的性能。虽然融合特征能更好地捕捉图像中各类物体的细节，但是简单的融合方法会干扰多尺度特征的语义一致性，影响检测效果。特别是在遥感图像中，目标尺寸差异大，单一的尺度或层次特征往往无法有效捕捉小目标的信息。Dong 等[17] 引入可变形卷积机制实现多级特征融合，构建了具有自适应感受野的特征图，以适配遥感目标在形状和方向上的多样性。Yan 等[18] 设计了一种自适应语义网络，同时捕捉全局依赖关系和局部细粒度特征，通过整合低层细节优化高层语义表达以提高检测性能。上述方法引入自适应的特征选择机制，降低了不同特征之间的互相干扰，但未能有效考虑特征之间的相关性和差异性。为进一步提升目标检测性能，一些方法在特征融合的同时引入注意力机制。Xu 等[19] 采用基于注意力的特征优化方法，增强了目标区域显著性与网络特征提取能力。Liu 等[20] 结合注意力机制与 Transformer 架构，建模时空上下文信息，有效弥补了双时态图像的时间域差异。Shamsolmoali 等[21] 提出了一种多实例注意力机制，强化局部特征蒸馏过程，有效区分前景与背景，在保持检测精度的同时实现模型轻量化。Yang 等[22] 基于注意力的软阈值过滤模块剔除高层特征图中的冗余信息，强化目标语义表达，并引入渐进回归损失函数提升定位精度。Ye 等[23] 提出了一种具有自适应注意力融合机制的卷积网络，平衡多尺度目标比例并学习特征图的深层语义信息。Lin 等[24] 利用位置增强和边缘特征描述目标大致位置，并通过注意力与反向注意力机制逐步优化预测结果。虽然上述方法结合特征融合与注意力机制双重优势，使模型有效聚焦于小目标，但是在复杂背景下，不平衡的特征融合可能导致信息的缺失，同时引入多种注意力可能导致信息的偏离。此外，肖振久等[25] 利用归一化 Wasserstein 距离[26](normalized Wasserstein distance, NWD)缓解最小点距离交并比[27] (minimum point distance IoU, MPDIoU)对遮挡场景的敏感性，两者结合作为损失函数有效缓解小目标定位偏差问题，但面对非水平矩阵或旋转偏移的物体仍存在一定局限性。

为解决上述问题，本文在 YOLOv11 算法的基础上，提出了一种融合注意力与特征增强的遥感小目标检测算法。首先，引入一种多层级特征协同策略 (cascade feature synergism, CFS)，通过空间注意力与通道注意力的协同作用，建立两种注意力机制的潜在关联，在实现适度敏感性的同时确保模型稳定地关注重要的特征信息，精准定位目标区域并有效抑制背景噪声干扰，提升对小目标的特征提取效果。其次，设计了一个多尺度特征增强模块 (multiscale feature enhancement, MSFE)，结合三维卷积和多尺度特征编码，通过平衡不同尺度特征图的权重进行特征选择，在保留全局信息的同时关注局部细节，减少了特征融合过程中小物体的信息损失。此外，为了全面考虑目标物体的几何特征，捕获角度及形状信息，增强 IoU 计算的稳定性，本文引入 SIoU[28] (SCYLLA-IoU)与 NWD 联合回归损失函数，通过 SIoU 的形状损失与 NWD 的高斯建模相互协调，更平滑地处理复杂场景中的尺度变化，加快了模型收敛速度。

## 2 融合注意力与特征增强的遥感图像小目标检测

为解决遥感图像小目标特征提取困难和特征信息不足等问题，提出一种融合注意力与特征增强的遥感图像小目标检测算法，其整体架构如图 1 所示。改进检测算法主要包含三个部分：1) 多层级特征协同模块，通过结合空间注意力和通道注意力机制，突出小目标特征，优化特征提取过程，有效捕捉不同层次的特征信息；2) 多尺度特征增强模块，采用维度特征融合策略聚合全局语义信息，并结合多尺度特征编码与双向特征金字塔网络[29](bidirectional feature pyramid network, BiFPN)改进颈部网络结构，提升小目标的检测精度；3) 损失函数部分，通过引入额外的角度信息并结合 NWD 度量，形成新的损失函数，提高了模型的收敛能力。

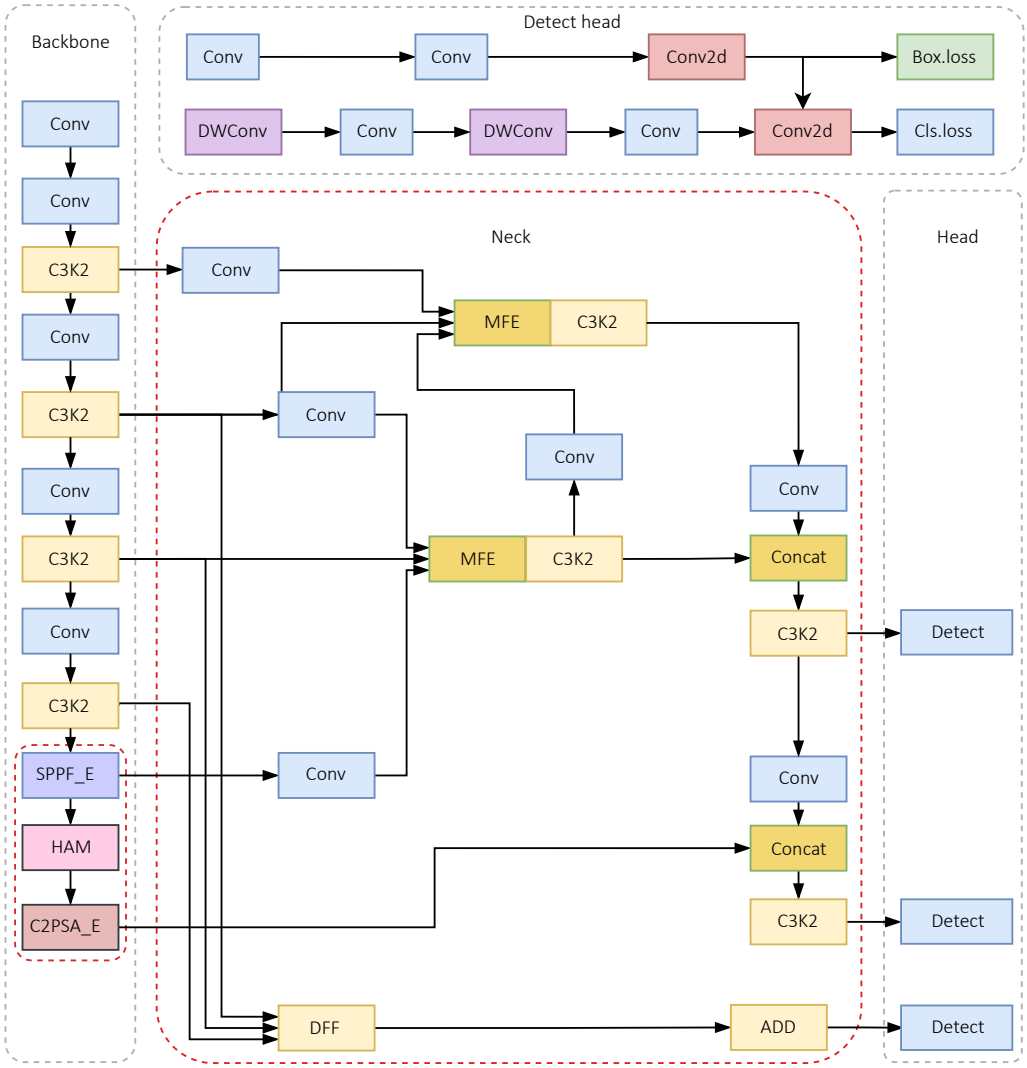



图 1 模型总体结构

Fig. 1 Overall model structure

## 2.1 多层级特征协同模块

对于遥感图像中的小目标检测，背景噪声往往对检测结果产生显著影响，而注意力机制能够显著提升模型对小目标的敏感度。其中空间注意力可以增强目标的局部特征，提升对小目标的定位精度，通道注意力可以优化通道间的信息流，突出重要特征，为充分利用两者的优势以提升在复杂场景中对不同尺度小目标的检测性能，采用一种多层级特征协同机制，结构如图 2 所示。通过融合通道注意力、空间注意力以及多尺度特征处理策略，优化特征提取与聚合过程，显著增强模型在复杂背景下的语义信息捕获能力，有效提升模型的鲁棒性。在传统的多尺度特征融合中，通常采用拼接或加权平均的方式，这往往忽略了不同尺度间的语义差异，导致信息不充分整合。为解决这一问题，引入高效多尺度注意力机制[30]，首先对输入特征进行分组，并行处理多个子特征，其中包含$\mathrm{Conv}_{1\times1}$分支从高度和宽度方向上对通道进行编码和$\mathrm{Conv}_{3\times3}$分支用来捕捉局部跨通道交互特征，可以有效地区分不同尺度的特征，并根据尺度差异调整不同尺度之间的关系。其公式表示为

$$X \in \mathbb{R}^{C\times H\times W}, \tag{1}$$

$$X = \{X_0, X_1, ..., X_G\}, \quad X_g \in \mathbb{R}^{C/G\times H\times W}, \tag{2}$$

$$z^H \in \mathbb{R}^{C/G\times H}, \quad z^W \in \mathbb{R}^{C/G\times W}, \tag{3}$$

$$z \in \mathbb{R}^{C/G\times H\times W}, \tag{4}$$

式中：$X$为输入特征，$C$为通道数，$G$为分组数，$H$和

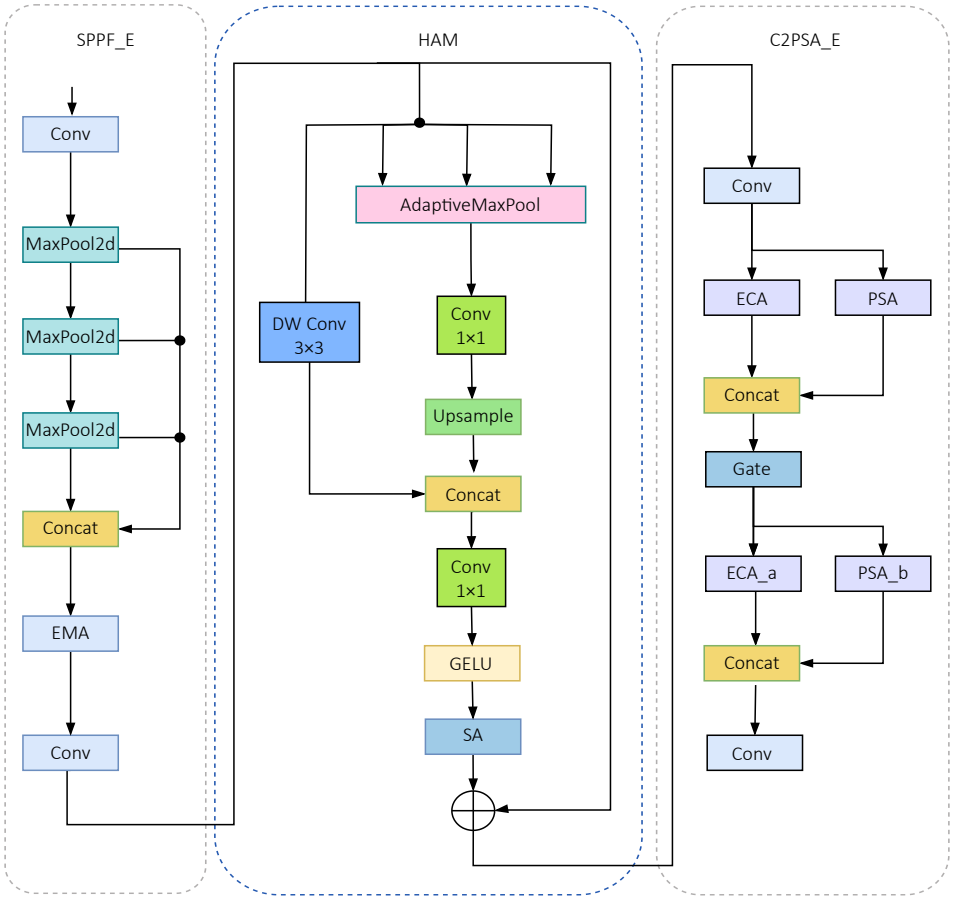



图 2 特征协同模块

Fig. 2 Feature collaboration module

$W$分别表示高度和宽度；$X_g$为分组特征；$z^H$和$z^W$分别为高度编码信息和宽度编码信息；$z$ 为空间展开后的特征图。接着通过二维全局平均池化进一步编码全局空间信息，完成全局特征向量和空间展开特征图的配对建模，有效捕捉长程和短程的依赖关系，其公式表示为

$$Z_c^{(g)} = \frac{1}{H \times W} \sum_j^H \sum_i^W x_c^{(g)}(i,j)\,, \tag{5}$$

$$Z^{(g)} \in \mathbb{R}^{C/G}, \tag{6}$$

式中：$Z_c^{(g)}$表示第 $c$ 个通道的全局平均响应，第 $g$ 组的全局平均响应。通过对不同尺寸卷积核的并行子网络输出的特征图进行有效融合，可以整合来自不同空间位置的特征信息，提高小目标特征表示的丰富性和准确性。

为进一步提升空间信息处理能力，引入一种联合通道注意力和空间注意力的协同策略。传统固定模式特征映射由于空间权重共享会导致背景噪声被同等放大，所以本文采用动态调整空间特征的方式，增强模型对空间信息的处理能力。通过对输入特征进行通道分割与深度可分离卷积处理，使模型在空间上能有效分配注意力资源到关键区域，提升遥感小目标的特征表达能力，该过程表示为

$$[X_0, X_1, X_2, X_3] = Split(X)\,, \tag{7}$$

$$\hat{X}_0 = DW\text{-}Conv_{3\times3}(X_0)\,, \tag{8}$$

$$\hat{X}_i = \uparrow_p (DW\text{-}Conv_{3\times3}(\downarrow_{\frac{p}{2^i}}(X_i))), 1 \leqslant i \leqslant 3\,, \tag{9}$$

式中：$Split(\cdot)$表示通道分割操作；$DW\text{-}Conv_{3\times3}(\cdot)$表示具有 3 pixel×3 pixel 的深度可分离卷积；$\uparrow_p(\cdot)$表示将特征向上采样至 $p$ 大小；$\downarrow_{\frac{p}{2^i}}(\cdot)$表示将输入特征向下采样到 $p/2^i$大小。$X_0$用来放大数据的细节提取深层特征，$X_1$、$X_2$、$X_3$只保留最重要的信息生成简化的特征图。接着对多路径特征信息进行拼接，聚合局部与全局关系并利用$1\times1$卷积优化特征紧凑性，然后应用 GELU 非线性

激活函数生成注意力图来标注重点区域，并根据注意力图通过与原始特征逐元素相乘的方式实现空间调制，突出遥感小目标信息，公式表示为

$$\hat{X} = Conv_{1\times1}\left(Concat([\hat{X}_0, \hat{X}_1, \hat{X}_2, \hat{X}_3])\right), \tag{10}$$

$$\overline{X} = \phi(\hat{X}) \odot X, \tag{11}$$

式中：$Concat(\cdot)$表示通道维度的拼接操作；$Conv_{1\times1}(\cdot)$为1×1卷积；$\phi(\cdot)$表示 GELU 函数；⊙是逐元素乘积。为减轻空间调制产生的无关干扰信息，并实现子特征间的信息流动，引入混洗注意力机制[31]。首先对输入特征沿通道维度划分为多个子特征，分别对通道间依赖关系和空间位置重要性进行建模，然后利用混洗操作使模型学会如何在不同通道间进行信息交换，提高对小物体信息的关注度，同时抑制背景噪声，公式表示为

$$X_k = \left[\hat{X}_{k1}, \hat{X}_{k2}\right], \tag{12}$$

$$s = \frac{1}{H\times W}\sum_{j}^{H}\sum_{i}^{W}\hat{X}_{k1}(i,j), \tag{13}$$

$$n = GN(\hat{X}_{k2}), \tag{14}$$

式中：$k$表示分组索引，$X_k$为第$k$组子特征，$\hat{X}_{k1}$为通道注意力分支并对其进行全局平均池化操作；$\hat{X}_{k2}$为空间注意力分支并对其进行组归一化操作，分别捕捉通道和空间信息实现跨通道的特征交互，增强遥感小目标的信息表达。为进一步加强邻居通道交互同时保持通道维度的完整性，采用高效通道注意力机制[32]，根据通道数自适应地调整一维卷积的核大小，捕捉通道间的动态依赖关系，筛选关键通道特征。此外，利用位置空间注意力，关注目标的空间分布特性，弱化不相关区域的特征，并通过门控机制动态地分配权重以优化通道与空间特征的整合。融合特征将会再次利用通道和空间注意力，挖掘更深层次的通道与空间特征，通过多阶段通道-空间注意力交互，显著提升了特征的有效性与目标定位精度。

## 2.2 多尺度特征增强模块

遥感图像通常具有高分辨率和复杂的背景，并且目标往往位于不同的尺度范围内，在进行目标检测时需要同时捕捉到多尺度和多通道的特征，而三维卷积能够同时处理空间和深度信息，增强模型对不同尺度小目标的感知能力以减少漏检现象，提升了对遥感小目标特征提取的鲁棒性。为此本文提出一种多尺度特征增强的颈部网络，通过维度特征融合促进模型对复杂场景的全面理解，同时通过多重特征编码聚焦小目标的局部特征，确保细节信息的精准提取，增强模型对小物体的捕捉能力。

### 2.2.1 维度特征融合

不同层级的特征图承载着不同的语义内容，低层特征图包含丰富的细节与位置信息，高层特征图则包含更全面的目标信息。传统特征金字塔网络通过简单拼接或求和的方法进行特征融合，忽略了各层级特征图间的尺度相关性。为此，提出了一种维度特征融合模块(dimension feature fusion, DFF)，通过三维卷积在层级轴上有效整合深层特征图的高级语义信息与浅层特征图的细节信息，提升模型对不同尺度特征的鲁棒性，其结构如图 3 所示，图中 $D$ 表示特征图堆叠的深度。

通过提取骨干网络输出的特征图如 P3、P4、P5，

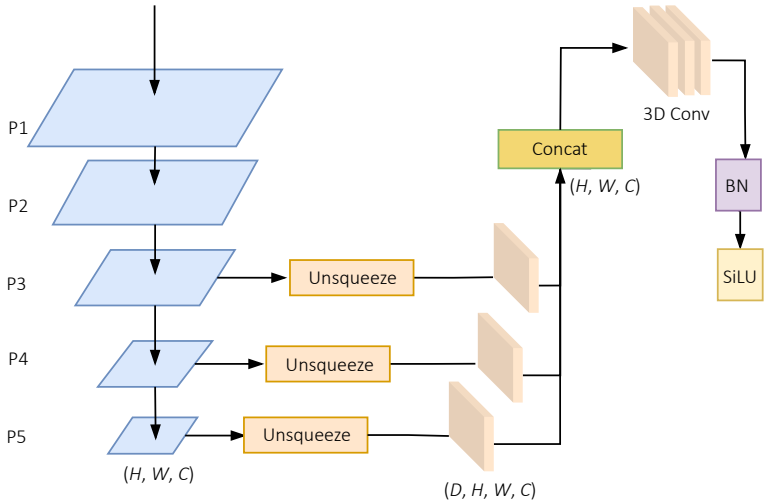



图 3 维度特征融合模块

Fig. 3 Dimensional feature fusion module

生成特征的序列表示，实现不同尺度和细节层次图像内容的提取，有效捕获各层级特征图中不同大小和形状目标的空间尺度信息。P3、P4 和 P5 特征图首先通过标准偏差递增的高斯核进行卷积平滑处理，公式为

$$F_\sigma(i,j)=\sum_u\sum_v f(i-u,i-v)\times G_\sigma(u,v), \tag{15}$$

$$G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-(x^2+y^2)/2\sigma^2}, \tag{16}$$

式中：$f$表示二维特征图；$F_\sigma$表示通过卷积平滑处理后得到的特征图，$(u,v)$表示相对于高斯核中心点的偏移坐标，$\sigma$表示标准差。P3 作为高分辨率特征图，包含小目标检测所需的关键细节信息，因此被选为基准分辨率。由于 P3、P4、P5 的分辨率不同，采用最近邻插值方法将 P4 和 P5 的空间尺寸上采样与 P3 一致。同时，通过 1×1卷积将 P4 和 P5 的通道数统一调整为 256，以确保特征维度一致。为捕获特征图间的尺度相关性，受到对视频帧序列处理中二维和三维卷积操作的启发[33]，将平滑后的特征图水平堆叠，生成三维特征序列。通过增加深度将每个特征层变成一个更立体的结构，然后沿深度维度拼接四维特征图，整理成适合后续处理的三维特征格式。最后，采用三维卷积结合批量归一化和 SiLU 激活函数把不同尺度的特征进行融合。DFF 模块通过三维卷积技术，有效整合了不同层级的特征信息，从多个视角和尺度描述目标，使特征表示更加丰富和全面。这种方法显著提升了特征的判别性，增强了模型对不同类型目标的区分能力。

### 2.2.2 多尺度特征编码

为了识别遥感图像中的小物体，可以通过拉伸图片来观察它们在各种大小下的形状或外观差异。然而，传统特征金字塔网络只是简单把低分辨率特征向上采样后融合至前一层特征，忽略了高分辨率特征层中丰富的细节。BiFPN 可以在不同尺度特征图之间进行双向传递，增强对遥感小目标的敏感度，解决传统特征金字塔在信息融合时的效率和信息损失问题，因而提出了一种多尺度特征编码模块 (multi feature encoding, MFE)，其结构如图 4 所示，采用独立处理的方法，将高、中、低分辨率的元素分离，扩展其中的丰富特征，并通过 BiFPN 优化不同尺度的特征融合过程，大幅强化对微小目标的精细描绘。

在特征整合前，首先选定中等分辨率特征作为核心基准，利用卷积模块把高、中、低三种分辨率特征层调整到相同的通道数目。对于高分辨率层，结合峰值提取与均值平滑的方式进行压缩，有效缩小尺寸并保持对位置偏移的鲁棒性，提升网络对图片的平移、旋转等变化的适应能力。对于低分辨率层，细节容易丢失，利用最近邻插值法进行上采样至基准大小，确保小目标的局部纹理不丢失并减少数据损失。为优化跨尺度信息的整合效果，引入加权特征融合机制，为每个输入特征图学习自适应权重。最后，将标准化后的高、中、低三层沿通道维度进行叠加生成融合特征图，公式表示为

$$F_{\mathrm{MFE}}=Concat(F_{\mathrm{l}},F_{\mathrm{m}},F_{\mathrm{s}}), \tag{17}$$

式中：$F_{\mathrm{MFE}}$表示 MFE 模块输出的特征图；$F_{\mathrm{l}}$、$F_{\mathrm{m}}$和$F_{\mathrm{s}}$分别表示高、中和低分辨率特征图。$F_{\mathrm{MFE}}$由$F_{\mathrm{l}}$、$F_{\mathrm{m}}$和$F_{\mathrm{s}}$拼接而成的，$F_{\mathrm{MFE}}$的分辨率与$F_{\mathrm{m}}$相同，但通道数是$F_{\mathrm{m}}$的三倍。不同尺度的特征图具有不同的信息结构，大尺度特征图侧重于语义信息，能反映目标的大致类别和位置；中尺度特征图则兼顾了一定的语义信息和部分细节信息；小尺度特征图主要提供了丰富的细节和边缘信息。将它们融合后，可以使特征表达更加丰富和全面，有助于提高模型对目标特征的判别能力，使目标与背景、目标与目标之间的界限更加清晰，减少误判和漏判的情况。

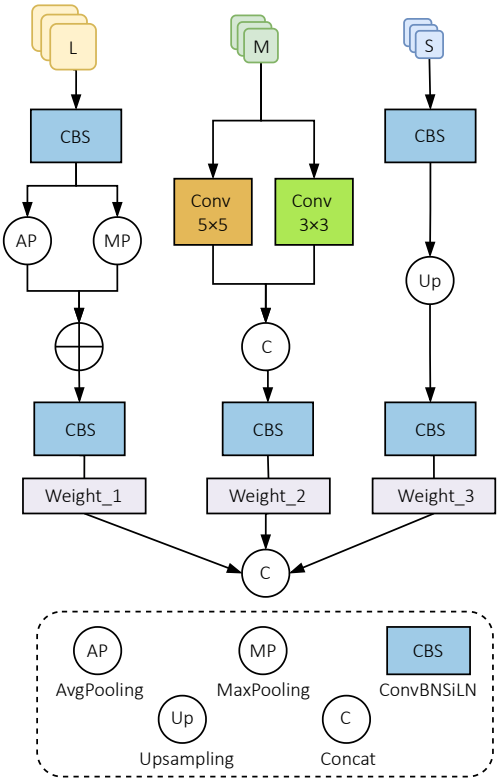



图 4 多重特征编码模块

Fig. 4 Multi-feature encoding module

## 2.3 损失函数优化

YOLOv11 采用的 CIoU 损失函数在常规目标检测中表现良好，但在面对小目标时存在局限性，其对小目标边界框的细微偏移不敏感，且在处理旋转目标和非重叠框的回归问题时效果欠佳。为此，本文结合 SIoU 和 NWD Loss 作为回归损失函数，以提高小目标检测的准确性并增强模型的泛化性能。

### 2.3.1 SIoU loss

传统的 IoU 损失函数通过综合评估预测框与真实框之间的距离、重叠区域以及宽高比等因素，优化了边界框的回归过程。然而，此类方法未充分解决预测框与真实框之间方向性不匹配的问题。SIoU 通过引入角度成本、距离成本、形状成本和 IoU 成本，改进 IoU 损失，提升边界框回归的精度和收敛速度，尤其在小目标、非重叠框和旋转框的检测任务中表现出优越性。

SIoU 损失函数的角度成本通过量化预测框与真实框之间的位置关系来优化几何对齐，公式表示为

$$\Lambda = 1 - 2\sin^2\left(\arcsin x - \frac{\pi}{4}\right), \tag{18}$$

$$x = \frac{c_h}{\sigma} = \sin\alpha, \tag{19}$$

$$\sigma = \sqrt{(c_{xb} - c_{xa})^2 + (c_{yb} - c_{ya})^2}, \tag{20}$$

$$c_h = \max(c_{yb}, c_{ya}) - \min(c_{yb}, c_{ya}), \tag{21}$$

式中：$\Lambda$表示角度成本；$(c_{xa}, c_{ya})$和$(c_{xb}, c_{yb})$分别表示边界框 $A$ 和边界框 $B$ 的中心点坐标，$\alpha$表示预测框与真实框中心连线与水平坐标轴的夹角，$c_h$表示真实框与预测框中心在垂直方向的距离差。考虑到上述定义的角度成本，重新定义距离成本，表示式为

$$\Delta = \sum_{t=x,y}(1 - \mathrm{e}^{-(2-\Lambda)\rho_t}), \tag{22}$$

$$\begin{cases} \rho_x = \left(\dfrac{c_{xb} - c_{xa}}{c_w}\right)^2 \\ \rho_y = \left(\dfrac{c_{yb} - c_{ya}}{c_h}\right)^2 \end{cases}, \tag{23}$$

式中：$\Delta$表示距离成本，用于衡量预测框与真实框在空间位置上的差异；$\rho_x$和$\rho_y$分别表示在 $X$ 轴和 $Y$ 轴方向上，目标边界框中心点之间相对距离的平方。当角度偏差趋近于 0 时，距离成本的贡献会显著降低。相反，当角度偏差越接近 π/4 时，$\Delta$ 就越大，距离成本的贡献也就越大。

SIoU 作为损失函数，其表示式为

$$L_{\mathrm{SIoU}} = 1 - IoU + \frac{\Delta + \Omega}{2}, \tag{24}$$

式中：$\Omega$表示形状成本，它的值对于每一个数据集都是唯一的，用于反映边界框形状的差异。SIoU 损失函数通过考虑角度、距离和形状等多个维度的差异，可以在训练过程中更快地调整预测框的方向，加快模型收敛。

### 2.3.2 NWD loss

在小目标检测任务中，即使是轻微的位置偏差也会导致 *IoU* 值显著下降，使得原本的正样本被错误标记为负样本。这种现象不仅减少了训练样本中的正样本数量，还会对模型的训练效果和检测性能造成负面影响。因此，引入一种基于 Wasserstein 距离的新度量标准，通过将边界框建模为二维高斯分布$N = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$，优化小目标的边界框回归。

$$\begin{cases} \boldsymbol{\mu} = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \\ \boldsymbol{\Sigma} = \begin{bmatrix} \dfrac{w^2}{4} & 0 \\ 0 & \dfrac{h^2}{4} \end{bmatrix}, \end{cases} \tag{25}$$

式中：$(c_x, c_y)$表示水平边界框的中心坐标；$w$ 和 $h$ 分别表示边界框的宽度和高度。通过对边界框$A = [c_{xa}, c_{ya}, w_a, h_a]$和$B = [c_{xb}, c_{yb}, w_b, h_b]$进行计算表达，生成正态分布$N_a$和$N_b$可表示为

$$W_2^2(N_a, N_b) = \left\| \left( \left[ c_{xa}, c_{ya}, \frac{w_a}{2}, \frac{h_a}{2} \right]^{\mathrm{T}}, \left[ c_{xb}, c_{yb}, \frac{w_b}{2}, \frac{h_b}{2} \right]^{\mathrm{T}} \right) \right\|_2^2. \tag{26}$$

然而，$W_2^2(N_a, N_b)$是两个边界框在高斯分布空间中的 Wasserstein 距离，需要通过指数形式的归一化操作重新表示为

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{E}\right), \tag{27}$$

式中：$E$ 表示数据集中目标框的平均尺寸。*NWD* 对不同尺度的目标具有较好的适应性，*NWD* 值越大表示两个框越相似，反之两个框差异越大。准确反映了边界框的实际关系，在小目标检测中更有优势。

通过融合两个损失函数得到的最终损失函数表示为

$$L_{\mathrm{reg}} = \lambda \cdot L_{\mathrm{SIoU}} + \beta \cdot NWD. \tag{28}$$

式中：$\lambda$和$\beta$分别表示两个损失函数的权重，为得到最佳的损失函数，进行参数优化实验，结果如表 1 所示。

可以看出在$\lambda = 0.5$、$\beta = 0.5$时，实验指标达到最佳。SIoU 更多地关注目标框的形状和位置，NWD 通过度量目标框之间的分布差异，有效抑制背景噪声，平衡的设置权重可以加强两者的协同作用以充分发挥各自的优势，显著提升了训练效率和检测精度。

表 1 参数优化实验结果

Table 1 Parameter optimization experiment results

| $\lambda$ | $\beta$ | mAP@0.5/% |
|---|---|---|
| 0.3 | 0.7 | 0.822 |
| 0.5 | 0.5 | **0.830** |
| 0.7 | 0.3 | 0.813 |
| 0.9 | 0.1 | 0.805 |

表 2 参数设置

Table 2 Parameter setting

| Parameter | Value |
|---|---|
| Optimizer | SGD |
| Learning_rate | 0.01 |
| Weight_decay | 0.0005 |
| Momentum | 0.937 |
| Batch_size | 16 |
| Epoch | 200 |

# 3 实验结果与分析

## 3.1 数据集及实验环境

NWPU VHR-10 数据集是一个包含 650 张正例图像和 3651 个目标实例的遥感图像数据集，共有飞机 (PL)、轮船 (SH)、油罐 (ST) 等 10 个类别。相比之下，DIOR 数据集覆盖的对象类别更丰富包含 20 个类别，数据规模更大，能够更全面地验证模型的泛化能力。而 RSOD 数据集只包含四个类别，其中飞机实例数量多达 4993 个，天桥的实例数量仅有 180 个。通过在多个数据集上进行实验评估模型在不同图像数量和目标多样性下的适应性与检测性能。

本实验环境为 Linux，GPU 为 NVIDIA Tesla P100-PCIE，运行内存为 32 GB，编程语言为 Python 3.10，深度学习框架为 PyTorch 2.5.1，输入图像分辨率设置为 512 pixel，参数设置如表 2 所示。

其中 optimizer 为优化器，learning_rate 为初始学习率，weight_decay 为权重衰减，momentum 为动量设置，batch_size 为批量大小，epoch 为模型的训练周期。为了确保实验的准确性，在模型训练期间没有使用预训练的权重。

## 3.2 评价指标

为衡量模型性能，实验使用精确率 (Precision)、召回率 (Recall)、平均精度 (mAP) 和帧率 (FPS) 作为模型的评价指标，公式表示为

$$Precision = \frac{TP}{TP+FP}, \tag{29}$$

$$Recall = \frac{TP}{TP+FN}, \tag{30}$$

$$AP = \int_0^1 Precision(Recall)\mathrm{d}(Recall), \tag{31}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i, \tag{32}$$

式中：$TP$ 表示正确检测的目标数量；$FP$ 表示误检的目标数量；$FN$ 表示漏检的目标数量。

## 3.3 消融实验

以 YOLOv11 为基础模型进行改进，在 NWPU VHR-10 数据集上进行消融实验，实验环境以及其他参数设置严格一致。实验结果如表 3 所示，A 表示应用多层级特

表 3 在 NWPU VHR-10 数据集中的消融实验结果

Table 3 Ablation study results on the NWPU VHR-10 dataset

| Number | A | B | C | D | Precision/% | Recall/% | Params/$10^6$ | FPS | mAP@0.5/% | mAP@0.5:0.95/% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | × | × | × | × | 82.4 | 65.8 | 2.58 | 344 | 73.1 | 41.2 |
| 2 | √ | × | × | × | 83.2 | 68.7 | 2.76 | 320 | 75.4 | 42.4 |
| 3 | × | √ | × | × | 81.2 | 75.0 | 3.01 | 268 | 80.4 | 47.9 |
| 4 | × | × | √ | × | 82.6 | 71.7 | **2.57** | 370 | 75.8 | 40.6 |
| 5 | × | × | × | √ | 79.8 | 70.5 | 2.62 | 303 | 76.4 | 42.9 |
| 6 | × | × | √ | √ | 81.2 | 72.3 | 2.60 | **384** | 77.0 | 42.8 |
| 7 | √ | √ | × | × | 84.8 | 71.8 | 3.41 | 254 | 79.4 | 47.3 |
| 8 | √ | × | √ | √ | 82.3 | 70.3 | 2.69 | 344 | 77.1 | 42.5 |
| 9 | × | √ | √ | √ | **86.7** | 74.4 | 3.29 | 267 | 82.5 | **49.4** |
| 10 | √ | √ | √ | √ | 83.3 | **77.2** | 3.51 | 272 | **83.0** | 48.8 |

征协同 CFS，B 表示采用多尺度特征增强 MSFE，C 和 D 分别表示替换为 SIoU 和 NWD 损失函数，同时表格中“√”表示使用该改进策略，“×”表示未使用。

基线实验使用未改进过的 YOLOV11 模型，在此数据集下的实验结果为 Precision 为 82.4%，Recall 为 65.8%，mAP@0.5 为 73.1%，mAP@0.5:0.95 为 41.2%。

第二组实验仅加入 CFS 加强提取特征的质量，同时保证计算效率，召回率提高 2.9%，mAP@0.5 提高 2.3%，mAP@0.5:0.95 提高 1.2%。

第三组实验仅加入 MSFE，有效融合了多尺度信息，减少了小目标细节特征的丢失，召回率提升了 9.2%.，mAP@0.5 提升了 7.3%，mAP@0.5:0.95 提升了 6.7%。

第四组实验采用 SIoU 作为损失函数，加快了训练速度并优化了边界框回归，提升了定位精度，召回率提高了 5.9%。

第五组实验采用 NWD 作为相似度度量，对形状一致性进行优化，mAP@0.5 提升 3.3%。

第六组实验结合 SIoU 和 NWD 作为损失函数，综合衡量边界框之间的相似性和差异性，对小目标进行精准定位，有助于提升检测精度，召回率提高 6.5%，mAP@0.5 提高 3.9%。

第七组实验同时加入 CFS 和 MSFE，充分利用高层语义信息和多尺度特征信息。召回率提升 6.0%，mAP@0.5 提升 6.3%。

第八组实验引入 CFS、并结合 NWD 和 SIoU 作为损失函数，增强了特征协同和回归精度，mAP@0.5 提高 4.0%，召回率也有明显提高。

第九组实验加入 MSFE、并结合 NWD 和 SIoU 作为损失函数，准确率提升至 86.7%，mAP@0.5:0.95 提升至 49.4%。

第十组实验为所提算法，通过引入 CFS、MSFF，并结合 NWD 度量和 SIoU 作为损失函数，将算法整体精度 mAP@0.5 提升至 83.0%，召回率为 77.2%。由实验结果可知，所提算法能够有效地提取遥感图像中的细节信息，提升小目标检测准确性。

### 3.4 对比实验

为了评估改进后的算法，在 NWPU VHR-10 数据集上进行对比实验，训练结果如图 5 和图 6 所示。可以看出，改进算法 mAP@0.5 值从 73.1% 提高到 83.0%，提升幅度达 9.9%。具体而言，篮球场 (BC) AP 值从 5.9% 提升到 32.7%，提升了 26.8%；桥梁 (BR) AP 值从 52.0% 提升到 77.5%，提升了 25.5%；港口 (HA) AP 值从 62.4% 提升到 81.1%，提升了 18.7%；小型车辆 (VE) AP 值从 60.2% 提升到 74.1%，提升了 13.9%；同时，其他类别的 AP 值均实现了不同程度的提升，充分验证了所提算法的有效性。

此外，本文算法分别与 Thundernet[34]、RICNN[35]、SSD[8]、Faster R-CNN[14]、YOLOv10、SGFTHR[36]、RT-DETR[37] 和 DFS-DETR[38] 在 NWPU VHR-10 数据集上进行对比实验，结果如表 4 所示，可以看出改进算法在检测性能上相较于其他算法具有明显优势 (本文表格中的加粗数据，表示最优值)。在复杂背景下，Thundernet 和 Faster R-CNN 由于缺乏充分的背景抑制机制，容易发

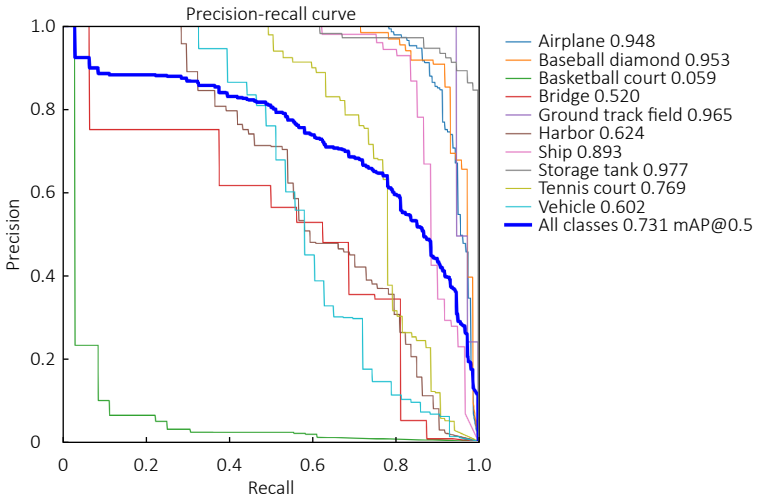



图 5 YOLOv11 算法训练结果

Fig. 5 Training results of the YOLOv11 algorithm

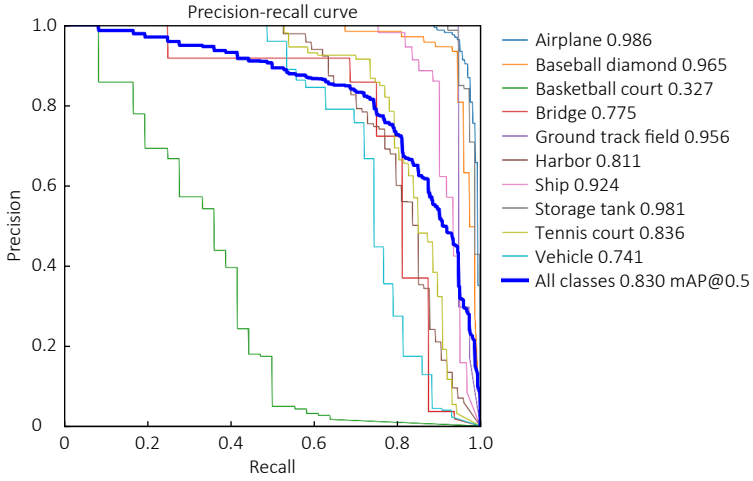



图 6 所提算法训练结果

Fig. 6 Training results of the proposed algorithm

表 4 不同算法在 NWPU VHR-10 数据集上的实验结果 (单位：%)

Table 4 Experimental results of different algorithms on the NWPU VHR-10 dataset (unit: %)

| Algorithm | PL | BD | BC | BR | GT | HA | SH | ST | TC | VE | mAP@0.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Thundernet | 86.1 | 89.6 | 68.3 | 86.3 | 86.8 | 60.4 | 85.0 | 47.3 | 51.3 | 63.8 | 72.5 |
| RICNN | 88.4 | 88.1 | 58.5 | 61.5 | 86.7 | 68.6 | 77.3 | 85.2 | 40.8 | 71.1 | 72.6 |
| SSD | 83.7 | 89.4 | **86.7** | **84.7** | 88.6 | 82.1 | 88.5 | 50.9 | 77.7 | 64.5 | 79.7 |
| Faster R-CNN | 78.1 | 95.2 | 43.3 | 49.0 | 84.9 | 36.6 | 66.3 | 84.9 | 76.6 | 79.7 | 69.6 |
| YOLOv10 | 82.8 | 81.4 | 16.4 | 54.6 | 85.0 | 51.7 | 76.5 | 92.7 | 54.7 | 45.7 | 64.1 |
| RT-DETR | 95.4 | 75.6 | 78.9 | 54.9 | **95.8** | **86.7** | 91.3 | 80.6 | **98.3** | 84.5 | **85.3** |
| SGFTHR | 78.2 | 97.4 | 46.1 | 50.9 | 87.4 | 37.9 | 67.1 | 86.4 | 85.8 | **93.3** | 73.1 |
| DFS-DETR | 95.1 | **98.2** | 78.0 | 52.0 | 95.3 | 84.8 | 70.3 | 78.3 | 79.3 | 71.7 | 81.7 |
| Ours | **98.6** | 96.5 | 32.7 | 77.5 | 95.6 | 81.1 | **92.4** | **98.1** | 83.6 | 74.1 | 83.0 |

生误检和漏检。尤其是在城市或密集的建筑环境中，小目标容易与背景混淆，但由于轻量化的设计使其在实时检测任务中表现良好。在检测小目标时，RICNN 对目标的边缘细节关注不足，且在特征融合过程中未对小目标信息赋予足够权重，因此难以有效定位微小目标，但其优化了候选区域机制在面对遮挡、光照变化时有更好的鲁棒性。SSD 的感受野较大无法有效保留细节信息，并且小目标特征容易在融合过程中丢失，造成小物体定位困难导致识别准确度下降，但在资源受限条件下仍能保持应有的效果。YOLOv10 在遥感小目标边界框预测中提取的特征信息不够丰富，导致边界框回归的准确性不足，存在真实框和预测框偏差较大的情况，但双标签分配策略使其在高效检测方面有很大优势。SGFTHR 是基于 transform 的遥感目标检测算法，通过结构张量提取模块弥补深层语义特征中空间细节的丢失，采用逐像素预测方法，更适合密集小目标场景，但对于结构弱、纹理少的目标适应性不足。DFS-DETR 和 RT-DETR 降低了传统 DETR 模型的计算成本，并针对航拍图像中的小物体检测进行优化，在棒球场 (BD)、田径场 (GT) 和港口 (HA) 类别的检测中达到最优效果，但在面对高度杂乱图像时缺少鲁棒性。根据结果可以看出，改进算法虽然在篮球场 (BC) 和车辆 (VE) 等角度特征不明显的非旋转目标检测效果较差，但整体平均精度相比其他对比算法仍有很大优势，特别是面对飞机 (PL)、轮船 (SH) 等有多个旋转方向的目标时，由于对小目标特征信息的增强以及角度信息的引入，改善了漏检情况，充分验证了改进算法在遥感小目标检测任务中的适用性。

为进一步验证本文方法在小目标检测方面的有效性，在 NWPU VHR-10 数据集上按目标尺度划分检测指标，实验结果如表 5 所示，其中 mAP-L 为大尺度目标，

mAP-M 为中等尺度目标，mAP-S 为小尺度目标。可以看出，通用检测算法例如 YOLOv10、Faster R-CNN 在小目标检测中仍受限于特征下采样导致的细节损失，而面向遥感场景的改进算法例如 RT-DETR、SGFTHR 虽然在一定程度上缓解了该问题，但特征融合过程仍存在信息冗余与多尺度权重不均衡的问题。相比之下，所提算法通过引入多层级特征协同策略实现了空间与通道注意力的动态平衡，并结合多尺度特征增强模块在不同尺度特征间进行自适应权重分配，有效提升了在小尺度目标检测任务的准确率。

为测试改进算法的泛化性能，基于 DIOR 数据集和 RSOD 数据集进行对比实验，结果如表 6 和表 7 所示。可以看出，在 DIOR 数据集上，改进方法的 mAP 指标达到了 75.6% 的优秀水平，与同为遥感图像检测算法 ARS-DETR[39] 相比也提升了 2.5%，表明本文算法在遥感图像小目标检测方面的有效性，且展现了其在多类别检测任务中的良好泛化能力；在 RSOD 数据集中 mAP 指标也达到了 92.6%，说明在面对样本数量较少的数据集也能发挥出稳定的性能。实验结果表明，相较于其他方法本文算法具有明显优势。

表 5　在 NWPU VHR-10 数据集上不同目标尺度的实验结果 (单位：%)

Table 5　Experimental results for different target scales on the NWPU VHR-10 dataset (unit: %)

| Method | mAP@0.5 | mAP-L | mAP-M | mAP-S |
|---|---|---|---|---|
| Thundernet | 72.5 | 68.6 | 80.7 | 65.3 |
| RICNN | 72.6 | 53.6 | 82.9 | 77.8 |
| SSD | 79.7 | **83.0** | 85.9 | 67.9 |
| Faster R-CNN | 69.6 | 56.3 | 73.7 | 76.9 |
| YOLOv10 | 64.1 | 41.9 | 75.2 | 71.6 |
| RT-DETR | **85.3** | 77.4 | 88.4 | 85.4 |
| SGFTHR | 73.1 | 60.9 | 75.3 | 82.2 |
| DFS-DETR | 81.7 | 69.7 | **93.3** | 73.4 |
| Ours | 83.0 | 64.6 | 93.0 | **88.2** |

表 6　不同算法在 DIOR 数据集上的实验结果 (C1~C20) (单位：%)

Table 6　Experimental results of different algorithms on the DIOR dataset (C1–C20)(unit: %)

| Algorithm | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Thundernet | 90.1 | 68.8 | 90.8 | 67.4 | 38.3 | 86.8 | 59.5 | 68.6 | 61.0 | 67.3 | |
| RICNN | 39.1 | 61.0 | 60.1 | 66.3 | 25.3 | 63.3 | 41.1 | 51.7 | 36.6 | 55.9 | |
| SSD | 59.5 | 72.7 | 72.4 | 75.7 | 29.7 | 65.8 | 56.6 | 63.5 | 53.1 | 65.3 | |
| Faster R-CNN | 53.6 | 49.3 | 78.8 | 66.2 | 28.0 | 70.9 | 62.3 | 69.0 | 55.2 | 68.0 | |
| YOLOv8 | 91.5 | 76.3 | 93.7 | 79.0 | 41.9 | 89.4 | 61.7 | 73.9 | 60.8 | 70.8 | |
| YOLOv10 | **91.7** | 74.7 | 93.8 | 79.6 | 38.4 | 88.3 | 60.9 | 73.7 | 60.6 | 71.9 | |
| ARS-DETR | 88.0 | 74.1 | 74.4 | **81.6** | 41.1 | 75.6 | **81.9** | 73.1 | **64.8** | **76.1** | |
| Ours | 91.4 | **79.6** | **94.1** | 80.4 | **42.2** | **90.5** | 60.5 | **75.1** | 64.2 | 74.5 | |
| Algorithm | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | mAP@0.5 |
| Thundernet | 57.7 | 51.8 | 56.0 | 18.3 | 89.5 | 72.0 | 80.4 | 56.5 | 34.8 | 54.2 | 63.5 |
| RICNN | 58.9 | 43.5 | 39.0 | 9.1 | 61.1 | 19.1 | 63.5 | 46.1 | 11.4 | 31.5 | 44.2 |
| SSD | 68.6 | 49.4 | 48.1 | 59.2 | 61.0 | 46.6 | 76.3 | 55.1 | 27.4 | 65.7 | 58.6 |
| Faster R-CNN | 56.9 | 50.2 | 50.1 | 27.7 | 73.0 | 39.8 | 75.2 | 38.6 | 23.6 | 45.4 | 54.1 |
| YOLOv8 | 65.3 | 67.8 | 62.0 | 88.1 | 92.7 | 80.3 | 91.2 | 50.8 | 63.0 | 72.7 | 73.6 |
| YOLOv10 | 63.2 | 66.6 | 58.2 | 87.7 | 92.8 | 79.2 | 90.5 | 44.2 | 62.3 | 71.0 | 72.5 |
| ARS-DETR | **78.6** | 66.3 | 65.4 | 84.5 | 90.1 | 72.2 | 81.1 | **61.5** | 60.6 | 70.3 | 73.1 |
| Ours | 72.3 | **70.4** | **64.0** | **90.0** | **94.7** | **81.8** | **91.8** | 50.9 | **66.6** | **77.1** | **75.6** |

表 7 不同算法在 RSOD 数据集上的实验结果 (单位：%)

Table 7 Experimental results of different algorithms on the RSOD dataset (unit: %)

| Algorithm | Aircraft | Oiltank | Overpass | Playground | mAP@0.5 |
|---|---|---|---|---|---|
| FRCN | 85.8 | 86.7 | 88.1 | 90.3 | 87.7 |
| SSD | 69.1 | 71.2 | 70.2 | 81.3 | 72.9 |
| Faster R-CNN | 71.3 | 90.7 | **90.9** | **99.7** | 88.1 |
| YOLOv8 | 92.4 | **99.2** | 73.3 | 99.4 | 91.1 |
| YOLOv10 | 90.8 | 99.1 | 63.5 | 96.6 | 87.5 |
| YOLOv12 | 90.1 | 98.4 | 54.9 | 99.4 | 85.7 |
| Ours | **94.1** | 99.1 | 77.7 | 99.5 | **92.6** |

在 NWPU VHR-10 和 DIOR 数据集上对可视化结果进行比较，如图 7 和图 8 所示。可以观察到 YOLOv10 算法在识别小目标时存在遗漏，YOLOv11 算法的检测精确率较低，YOLOv12 对密集小目标的位置信息提取能力不足。本文通过门控机制协同通道注意力和空间注意力，利用不同层级的通道信息和空间位置信息加强对小目标的关注度，减轻背景噪声的干扰来增强检测效果。利用改进后的颈部网络，通过对大、中、小等不同尺度特征图进行融合，进一步提取小目标的局部细节和全局信息。将 SIoU 和 NWD 融合作为损失函数，显式建模预测框与真实框的方向差异，有效解决小目标定位模糊的问题。综上表明，本研究提出的检测算法显著提升了小目标识别性能，检测出 YOLOv11 模型漏检的目标，有效解决了在复杂背景中的定位偏差和识别精度不高等问题。

## 4 结 论

为解决遥感图像中复杂场景下小目标特征提取困难，特征融合后细节信息容易丢失等问题，本文提出一种融合注意力与特征增强的遥感图像小目标检测算法，通过多层级特征协同模块和特征增强模块，提升了模型的检测准确性并缓解了漏检问题。多层级特征协同模块通过门控机制加强通道注意力和空间注意力的性能，有效提升模型提取特征的质量，提高对复杂背景噪声的抗干扰能力。多尺度特征增强模块采用三维卷积和特征编码策略，优化整合不同尺度特征图中的信息，提升了模型对全局信息和局部细节的理解能力，进一步强化了小目标的特征表达。此外，通过结合 SIoU 和 NWD 作为损失函数，提高了模型的训练效率和检测精度。实验结果显示，该算法在 NWPU VHR-10、DIOR 和 RSOD 数据集上的各项指标均有较大提升。尽管本文方法在检测精度与收敛效率方面表现出较强的优势，但仍存在一定的局限性。首先，由于引入了多层注意力机制与多尺度卷积结构，模型在计算复杂度与显存占用方面较高，不利于轻量化部署。其次，在极端密集场景或低分辨率遥感影像中，小目标间的特征重叠与边界模糊仍会导致检测精度下降。此外，本文方法主要基于单模态可见光影像，

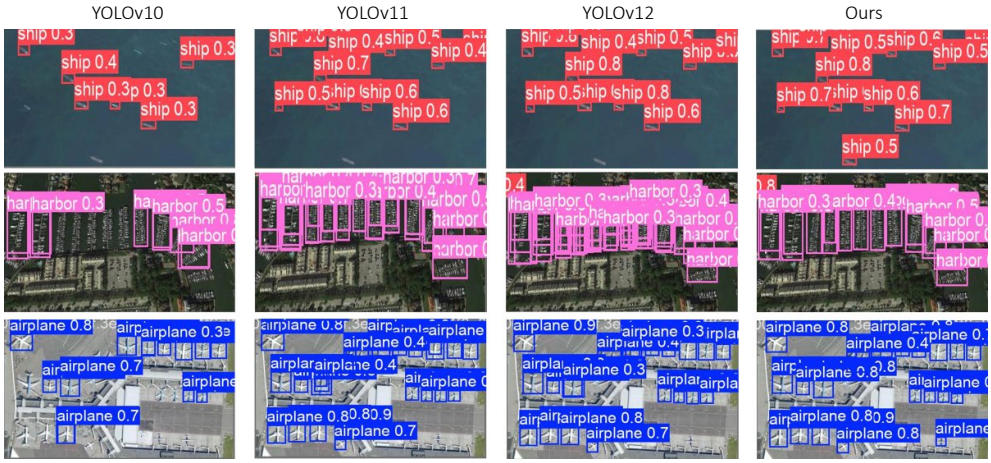



图 7 基于 NWPU VHR-10 数据集的检测效果对比

Fig. 7 Comparison of detection performance based on the NWPU VHR-10 dataset

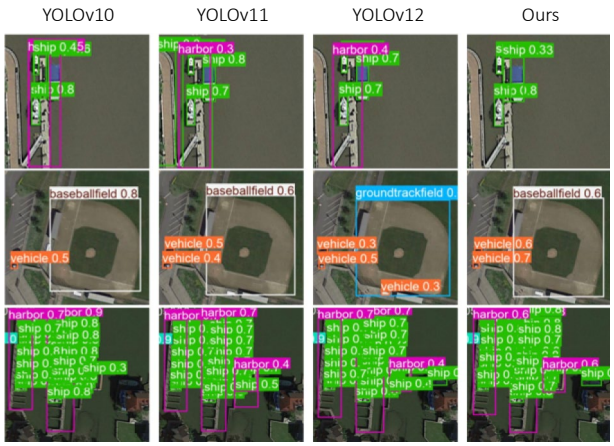



图 8 基于 DIOR 数据集的检测效果对比

Fig. 8 Comparison of detection performance based on the DIOR dataset

模型在多源异构遥感数据例如红外图像、SAR 图像上的适应性仍需进一步验证。在未来研究中，可以探索基于特征蒸馏与稀疏注意力机制的轻量化网络结构，以在保证检测精度的同时提升推理效率，并结合跨模态特征融合与 Transformer 动态建模，提升模型在复杂多源场景下的泛化能力。此外，未来的工作还将考虑与红外-可见光图像融合检测、海量遥感场景检索等实际任务相结合，以进一步提升其实用性与学术价值。

## 作者贡献

沈学利：指导论文选题、论文修改；郭旭：研究资料的收集、分析，论文写作。

## 利益冲突

所有作者声明无利益冲突

# 参考文献


1. Chen L, Zhang J L, Peng H et al. Few-shot image classification via multi-scale attention and domain adaptation[J]. *Opto-Electron Eng*, 2023, **50** (4): 220232.
陈龙, 张建林, 彭昊, 等. 多尺度注意力与领域自适应的小样本图像识别[J]. 光电工程, 2023, **50** (4): 220232.
2. Yang Y D, Ge H B, Xin S A et al. Lightweight small-object detection for remote sensing images integrating super-resolution and feature enhancement[J]. *Comput Eng*, 2024, **50** (11): 284−296.
杨雨迪, 葛海波, 辛世澳, 等. 融合超分辨率和特征增强的轻量化遥感图像小目标检测[J]. 计算机工程, 2024, **50** (11): 284−296.
3. Wang F F, Chen X A, Ren W H et al. Low-illumination object-detection algorithm based on image adaptive enhancement[J]. *Comput Eng*, 2024, **50** (10): 352−361.
王非凡, 陈希爱, 任卫红, 等. 基于图像自适应增强的低照度目标检测算法[J]. 计算机工程, 2024, **50** (10): 352−361.
4. Wang Y, Wang H H, Liu S D et al. Small target detection in sonar images with multilevel feature screening and task dynamic alignment[J]. *Opto-Electron Eng*, 2024, **51** (10): 240196.
王燕, 王宏辉, 刘树东, 等. 多级特征筛选和任务动态对齐的声呐图像小目标检测[J]. 光电工程, 2024, **51** (10): 240196.
5. Shan H L, Wang S Y, Tong J Y et al. Multi-scale optical remote sensing image target detection based on enhanced small target features[J]. *Acta Opt Sin*, 2024, **44** (6): 0628006.
单慧琳, 王硕洋, 童俊毅, 等. 增强小目标特征的多尺度光学遥感图像目标检测[J]. 光学学报, 2024, **44** (6): 0628006.
6. Wang X L, Liang Z Y, Liu T. Feature attention pyramid-based remote sensing image object detection method[J]. *Natl Remote Sens Bull*, 2023, **27** (2): 492−501.
汪西莉, 梁正印, 刘涛. 基于特征注意力金字塔的遥感图像目标检测方法[J]. 遥感学报, 2023, **27** (2): 492−501.
7. Zhang D Y, Zhao Z H, Xie Y G et al. Research on aircraft target detection in remote sensing images based on improved YOLOv8[J]. *Autom Appl*, 2024, **65** (2): 193−195,198.
张德银, 赵志恒, 谢逸戈, 等. 基于改进 YOLOv8 的遥感图像飞机目标检测研究[J]. 自动化应用, 2024, **65** (2): 193−195,198.
8. Lyu Z W, Jin H F, Zhen T et al. Small object recognition algorithm of grain pests based on SSD feature fusion[J]. *IEEE Access*, 2021, **9**: 43202−43213.
9. Redmon J, Divvala S, Girshick R et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
10. Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6517–6525. https://doi.org/10.1109/CVPR.2017.690.
11. Redmon J, Farhadi A. YOLOv3: an incremental improvement[Z]. arXiv: 1804.02767, 2018. https://doi.org/10.48550/arXiv.1804.02767.
12. Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. https://doi.org/10.48550/arXiv.2004.10934.
13. Ge Z, Liu S T, Wang F et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. https://doi.org/10.48550/arXiv.2107.08430.
14. Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
15. He K M, Gkioxari G, Dollár P et al. Mask R-CNN[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 2980–2988. https://doi.org/10.1109/ICCV.2017.322.
16. Liu J J, Zhang S T, Wang S et al. Multispectral deep neural networks for pedestrian detection[C]//*Proceedings of the 27th British Machine Vision Conference*, 2016: 73.1–73.13. https://doi.org/10.5244/C.30.73.
17. Dong X H, Qin Y, Gao Y H et al. Attention-based multi-level feature fusion for object detection in remote sensing images[J]. *Remote Sens*, 2022, **14** (15): 3735.
18. Yan R X, Yan L Q, Geng G H et al. ASNet: adaptive semantic network based on transformer–CNN for salient object detection in optical remote sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5608716.
19. Xu Y L, Bi H B, Yu H F et al. Attention-based contrastive learning for few-shot remote sensing image classification[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5620317.
20. Liu W, Lin Y Y, Liu W J et al. An attention-based multiscale transformer network for remote sensing image change detection[J].

*ISPRS J Photogramm Remote Sens*, 2023, **202**: 599−609.
21. Shamsolmoali P, Chanussot J, Zhou H Y et al. Efficient object detection in optical remote sensing imagery via attention-based feature distillation[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5624412.
22. Yang Y, Zang B J, Song C Y et al. Small object detection in remote sensing images based on redundant feature removal and progressive regression[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5629314.
23. Ye Y X, Ren X Y, Zhu B et al. An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images[J]. *Remote Sens*, 2022, **14** (3): 516.
24. Lin Y H, Sun H, Liu N Z et al. Attention guided network for salient object detection in optical remote sensing images[C]//*Proceedings of the 31st International Conference on Artificial Neural Networks*, 2022: 25–36. https://doi.org/10.1007/978-3-031-15919-0_3.
25. Xiao Z J, Zhang J H, Lin B H. Feature coordination and fine-grained perception of small targets in remote sensing images[J]. *Opto-Electron Eng*, 2024, **51** (6): 240066.
肖振久, 张杰浩, 林渤翰. 特征协同与细粒度感知的遥感图像小目标检测[J]. 光电工程, 2024, **51** (6): 240066.
26. Wang J W, Xu C, Yang W et al. A normalized Gaussian Wasserstein distance for tiny object detection[Z]. arXiv: 2110.13389, 2021. https://doi.org/10.48550/arXiv.2110.13389.
27. Ma S L, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression[Z]. arXiv: 2307.07662, 2023. https://doi.org/10.48550/arXiv.2307.07662.
28. Gevorgyan Z. SIoU loss: more powerful learning for bounding box regression[Z]. arXiv: 2205.12740, 2022. https://doi.org/10.48550/arXiv.2205.12740.
29. Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.
30. Ouyang D L, He S, Zhang G Z et al. Efficient multi-scale attention module with cross-spatial learning[C]//*ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023: 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096516.
31. Zhang Q L, Yang Y B. SA-Net: shuffle attention for deep convolutional neural networks[C]//*ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021: 2235–2239. https://doi.org/10.1109/ICASSP39728.2021.9414568.
32. Wang Q L, Wu B G, Zhu P F et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.
33. Park H J, Kang J W, Kim B G. ssFPN: scale sequence ($S^2$) feature-based feature pyramid network for object detection[J]. *Sensors*, 2023, **23** (9): 4432.
34. Qin Z, Li Z M, Zhang Z N et al. ThunderNet: towards real-time generic object detection on mobile devices[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 6717–6726. https://doi.org/10.1109/ICCV.2019.00682.
35. Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2016, **54** (12): 7405−7415.
36. Li J J, Zhang H Q, Song R et al. Structure-guided feature transform hybrid residual network for remote sensing object detection[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 5610713.
37. Zhao Y A, Lv W Y, Xu S L et al. DETRs beat YOLOs on real-time object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605.
38. Cao X Y, Wang H W, Wang X et al. DFS-DETR: detailed-feature-sensitive detector for small object detection in aerial images using transformer[J]. *Electronics*, 2024, **13** (17): 3404.
39. Zeng Y, Chen Y S, Yang X et al. ARS-DETR: aspect ratio-sensitive detection transformer for aerial oriented object detection[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5610315.

## 作者简介



沈学利 (1969-)，男，博士，教授，CCF 高级会员，主要从事推荐系统、网络及信息安全、图像与视觉信息计算方面的研究等。

E-mail：Shenxueli@lntu.edu



【通信作者】郭旭 (1999-)，男，硕士研究生，主要从事机器学习和图像与视觉信息计算方面的研究。

E-mail：guoxu9907@163.com

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文

# Fusion attention and feature enhancement of small targets detection in remote sensing images

Shen Xueli, Guo Xu*

School of Software, Liaoning University of Engineering and Technology, Huludao, Liaoning 125105, China

## Abstract

**Objective** Remote sensing image target detection plays a crucial role in a wide range of practical applications, including resource exploration, urban planning and management, environmental monitoring, and disaster emergency response. With the rapid development of high-resolution sensors and large-scale data acquisition technologies, the demand for accurate and efficient object detection methods has increased significantly. However, targets in remote sensing images are typically characterized by small size, large scale variations, complex and cluttered backgrounds, blurred boundaries, and insufficient texture information, which significantly increases the difficulty of effective feature extraction and reliable detection. In addition, factors such as imaging altitude, atmospheric interference, and illumination variations further degrade image quality, resulting in low contrast between targets and the surrounding environment. Despite substantial progress in recent years, small target detection in remote sensing imagery still faces several critical challenges. From a localization perspective, commonly used metrics such as Intersection over Union are overly sensitive to slight variations in bounding boxes when applied to small targets, leading to unstable regression optimization and reduced localization accuracy. From a data perspective, the limited number of small target samples and the imbalance in category distribution often cause models to underrepresent small object features, resulting in missed detections. Moreover, in complex environments, small targets are often mixed with background regions that share similar textures or visual patterns, making it more difficult to accurately separate foreground from background and achieve precise localization.Therefore, improving the accuracy, robustness, and generalization capability of small target detection in remote sensing images, while enhancing training stability and localization precision under complex conditions, remains a critical and meaningful research problem.

**Methods** Addressing the above challenges, a remote sensing small target detection algorithm integrating attention mechanisms and feature enhancement is proposed. First, a multi-level feature collaboration strategy is constructed by combining channel attention and spatial attention with a gating mechanism. This design enables adaptive weighting of features across different levels, strengthens cross-scale information interaction, and effectively suppresses background interference. By dynamically emphasizing salient regions and reducing redundant responses, the model is able to capture subtle and discriminative features of small targets more reliably. Second, a multi-scale feature enhancement module is designed by integrating three-dimensional convolution with multi-scale feature encoding to form an enhanced neck structure. This module facilitates the fusion of semantic and spatial information across different scales, allowing the network to preserve global contextual information while retaining local details. The use of three-dimensional convolution further enhances the modeling of inter-scale dependencies, improving the representation of small targets under complex backgrounds and varying resolutions. Finally, SIoU and NWD are jointly employed as regression loss functions to optimize bounding box prediction. This combination improves geometric alignment between predicted and ground-truth boxes, reduces sensitivity to scale differences, and stabilizes the optimization process. As a result, the model achieves faster convergence and more accurate localization, especially for small and densely distributed targets.

**Results and Discussions** Extensive experiments are conducted on three widely used remote sensing datasets, including NWPU VHR-10, DIOR, and RSOD, to evaluate the effectiveness and generalization capability of the proposed method. The results show that, compared with the baseline model, the improved algorithm achieves consistent performance gains across all datasets, with mAP@0.5 increasing by 9.9%, 3.1%, and 3.2%, respectively. These improvements demonstrate that the proposed method effectively enhances feature representation and detection performance for small targets. Visualization results further indicate that the model can successfully identify targets that are easily missed by conventional approaches, particularly in scenarios with dense distributions and complex backgrounds. In addition, the proposed method shows improved robustness in distinguishing targets from visually similar background regions. Ablation studies confirm the effectiveness of each component, showing that the multi-level feature collaboration strategy, the multi-scale feature enhancement module, and the optimized loss function contribute synergistically to performance improvement. The results also suggest that the proposed framework maintains stable detection performance under different dataset distributions, indicating good generalization ability.

Foundation item: National Natural Science Foundation of China (62173171)

*Correspondence: Guo X, E-mail: guoxu9907@163.com

**Conclusions** In this study, an attention- and feature-enhancement-based small target detection algorithm for remote sensing images is proposed. By introducing a multi-level feature collaboration strategy and a multi-scale feature enhancement module, the method significantly improves cross-level feature interaction and strengthens the representation of small targets. Meanwhile, the joint optimization of SIoU and NWD loss functions further enhances localization accuracy and training efficiency. The proposed approach effectively addresses key challenges in remote sensing small target detection, including complex backgrounds, scale variations, and insufficient feature information. Experimental results on multiple benchmark datasets confirm the superiority, robustness, and generalization capability of the proposed method over the baseline model. These findings demonstrate its strong potential for practical applications in real-world remote sensing tasks, particularly in scenarios that require accurate detection of small and densely distributed targets. Furthermore, the proposed framework provides a reliable technical reference for future research and can serve as a foundation for further improvements in small target detection methods.

**Keywords:** remote sensing images; small target detection; multiscale; feature enhancement

Shen X L, Guo X. Fusion attention and feature enhancement of small targets detection in remote sensing images[J]. *Opto-Electron Eng*, 2026, **53**(3): 250267; DOI: 10.12086/oee.2026.250267