

中图法分类号: TP18; TP75 文献标识码: A 文章编号: 1006-8961(2025)12-3955-14

论文引用格式: Li Q, Wang Z, Cui S W and He M Y. 2025. Lightweight spaceborne remote sensing object detection algorithm with multi-attention mechanism. Journal of Image and Graphics, 30(12):3955-3968(李强, 王智, 崔书玮, 何明一. 2025. 融合多注意力机制的轻量化天基遥感目标检测算法. 中国图象图形学报, 30(12):3955-3968)[DOI: 10. 11834/jig. 250002]

# 融合多注意力机制的轻量化天基遥感目标检测算法

李强<sup>1</sup>, 王智<sup>2</sup>, 崔书玮<sup>2</sup>, 何明一<sup>1\*</sup>

1. 西北工业大学, 西安 710072; 2. 北京空间机电研究所, 北京 100094

**摘要:** 目的 天基遥感图像覆盖范围大、被探测目标小而密集、成像背景复杂, 大型神经网络在实时性要求高且计算资源受限的天基遥感任务中难以部署。针对以上问题, 提出了一种融合空间与通道多注意力机制的轻量化天基遥感图像目标检测算法。方法 首先, 在YOLOv11n(you only look once version11)的基础上引入CPCA(channel prior convolutional attention)注意力机制, 通过提取通道和空间上的注意力特征来丰富网络的目标特征, 提升网络的特征提取能力。通过采用基于部分卷积(partial convolution, Pconv)的二维卷积层设计, 避免了模型增加注意力模块带来的参数量过大的问题, 使得改进后的模型参数量相较于YOLOv11n原始模型减少了0.48 M(约18.53%)。结果 实验结果表明, 本文算法在DIOR(object detection in optical remote sensing images)数据集上的验证中, 模型参数量为2.11 M, 并且与多种YOLO算法进行的目标检测实验均取得主要性能的提升。本文算法的模型参数量降低到81.47%, 目标检测精度平均提高了1.9%, 召回率平均提高了1.2%。此外, 本文算法的NPU(neural network processing unit)推理耗时仅为14.8 ms, 相对于YOLOv11n原始模型减少了4.8 ms, 推理速度提升了24.49%。结论 本文提出了一种融合多注意力机制和上下文信息的轻量化遥感图像目标检测算法, 在有效降低模型参数量的同时显著提高了检测精度, 为遥感图像目标检测算法的星上部署、天基遥感系统的快速在轨处理、解译及实时准确跟踪目标提供重要的技术支持。

**关键词:** 天基遥感图像; YOLOv11n; 注意力机制; 目标检测; 模型轻量化; 算法部署

## Lightweight spaceborne remote sensing object detection algorithm with multi-attention mechanism

Li Qiang<sup>1</sup>, Wang Zhi<sup>2</sup>, Cui Shuwei<sup>2</sup>, He Mingyi<sup>1\*</sup>

1. Northwestern Polytechnical University, Xi'an 710072, China;

2. Beijing Institute of Space Mechanics & Electricity, Beijing 100094, China

**Abstract: Objective** With the advancement of image processing and artificial intelligence, deep learning-based algorithms have become increasingly important in the tasks of image target detection and recognition. In the aerospace domain, satellite remote sensing object detection consistently confronts challenges, including cluttered imaging backgrounds, numerous minuscule targets, and wide dynamic imaging ranges. In recent years, convolutional neural network-based approaches have witnessed significant progress in satellite remote sensing object detection, particularly in fine-grained target recognition. These advancements play crucial roles across domains such as military reconnaissance, postdisaster reconstruction,

收稿日期: 2025-02-12; 修回日期: 2025-07-30; 预印本日期: 2025-08-07

\* 通信作者: 何明一 myhe@nwpu.edu.cn

基金项目: 国家自然科学基金(62001396, 61671387)

Supported by: National Natural Science Foundation of China (62001396, 61671387)

and resource exploration. Given the challenges of large coverage, small and dense targets, and complex imaging backgrounds in satellite-based remote sensing images, large and complex neural networks have been utilized to represent image features for further target detection. Although large neural networks exhibit certain detection capabilities, they are difficult to deploy in space-based remote sensing tasks because of the high real-time requirements and limited computing resources. To address these issues, this study proposes a lightweight space-based remote sensing image target detection algorithm that integrates multiattention mechanisms in the spatial domain and channels. It deploys remote sensing image data processing and target detection algorithms to a remote sensing edge intelligent computing platform, achieving efficient and accurate target recognition and analysis for remote sensing images. This approach provides a solution for future in-orbit fast target detection algorithm processing and real-time tracking of detection targets.

**Method** Based on a You Only Look Once version 11 model (i. e., YOLOv11n), the proposed algorithm integrates the channel prior convolutional attention (CPCA) mechanism, which combines channel and spatial attention mechanisms. It utilizes the channel attention mechanism to generate a channel attention map. Subsequently, this map is multiplied element-wise with the model's input feature map to produce a channel-weighted feature map. This channel-weighted feature map is then fed into a depthwise convolution module to generate a spatial attention feature map. The CPCA mechanism can dynamically allocate attention weights across channel and spatial dimensions, enriching the network's target features by extracting channel-wise and spatial attention features, thereby enhancing the network's feature extraction capability. By employing a 2D convolutional layer based on partial convolution (Pconv), which convolves only a subset of input channels, it leverages redundant compression in interchannel feature maps. This approach avoids the issue of excessive parameters typically introduced by adding attention modules. Consequently, the improved model reduces the parameter count by 0.48 M (approximately 18.53%) compared with the original YOLOv11n. This approach partially addresses the challenge of deploying network models on embedded devices. For ensuring consistent dimensions between the two branches of Pconv, a max-pooling operation is applied to the nonconvolved channels, downsizing the feature maps to half their original dimensions. Through leveraging pointwise convolution to fully utilize the representational capacity of channel-wise features, this design reduces the computational load while preventing significant degradation in the model's feature extraction capability.

**Result** During validation on the DIOR dataset, the proposed algorithm was compared with various YOLO algorithms for object detection. Experimental results demonstrate that real-time detection transformer (RTDETR) has the largest parameter count at 9.42 M, YOLOv11n has 2.59 M parameters, and YOLOv11n\_CBAM has 2.74 M. By contrast, the proposed model contains only 2.11 M parameters, accounting for 81.47% of those of the original YOLOv11n. Meanwhile, compared with the original YOLOv11n algorithm, the proposed method achieves a mean improvement of 1.9% in accuracy and 1.2% in recall. The neural network processing unit (NPU) inference latency of YOLOv11n is 19.6 ms, whereas the proposed algorithm achieves only 14.8 ms. This result indicates a reduction of 4.8 ms in comparison with the original model, representing a 24.49% speed improvement. Additionally, the NPU-deployed YOLOv11n model attains an accuracy of 0.799 and a recall of 0.642, whereas the proposed algorithm achieves 0.819 accuracy and 0.652 recall. Accordingly, no potential accuracy degradation occurs during model migration and deployment. Compared with merely adding the CPCA module, the proposed algorithm exhibits a slight accuracy decrease of 0.10% but reduces the parameter count by 0.66 M. When contrasted with solely incorporating the Pconv module, it shows a marginal parameter increase of 0.08 M, yet it improves the accuracy by 1.7%.

**Conclusion** Targeting space-based remote sensing minute object detection tasks, this study draws inspiration from the YOLOv11n model to propose a lightweight object detection algorithm that integrates multiattention mechanisms in the spatial domain and channels and contextual information. This approach significantly enhances detection accuracy while effectively reducing model parameters. By refining the attention mechanism in YOLOv11n, we introduce an improved architecture incorporating the CPCA module. This architecture enables comprehensive feature extraction for minute objects across spatial and channel dimensions, effectively mitigating missed detections and false alarms in spaceborne imagery. The conventional 2D convolutional layers in YOLO are replaced with Pconv-based designs, circumventing parameter inflation typically caused by attention modules. This replacement achieves an 18.53% parameter reduction and model lightweighting. Finally, through NPU-optimized deployment, the model's hardware compatibility is enhanced. Compared with the original YOLOv11n, the proposed algorithm reduces inference time by 4.8 ms while maintaining detection accuracy, meeting real-time monitoring requirements. The

solution proves exceptionally resource efficient for space-based engineering deployment with constrained computational resources and memory, providing crucial technical support for onboard implementation in spaceborne remote sensing systems.

**Key words:** space-based remote sensing images; YOLOv11n; attention mechanism; target detection; model lightweighting; algorithm deployment

## 0 引言

卫星遥感图像目标检测常常面临成像背景复杂、目标小且数目多、成像动态范围大等挑战。近年来,基于卷积神经网络的方法在卫星遥感图像目标检测,尤其在细粒度目标检测方面取得显著进展,在军事侦察、灾后重建及资源勘探等领域发挥了重要作用。基于卷积神经网络的目标检测算法大致可分为基于候选区域的目标检测算法和无区域回归目标检测算法两大类。

1) 基于候选区域的目标检测算法。以 R-CNN (region-based convolutional neural network) (Girshick 等, 2014) 为代表。R-CNN、Fast R-CNN (faster region-based convolutional neural network) (Girshick, 2015) 通过去除对单个候选区域单独传递网络的设计, 仅一次卷积操作生成候选区域, 从而显著提升了检测速度。随后, 为了进一步提高检测效率, Faster R-CNN 引入了区域预测网络 (region proposal network, RPN), 将候选区域选择和目标检测合并成一个统一主干网络 (Ren 等, 2017)。通过一系列优化迭代, R-CNN 系列算法在提高目标检测精度方面取得了显著成效。然而, 该类算法在处理大规模数据集时, 检测速度较慢, 导致其在天基遥感图像处理领域的实用性受限。

2) 无区域回归目标检测算法。其代表性算法由最初的 SSD (single shot multiBox detector) 发展为 YOLO (you only look once) 类算法。SSD (Liu 等, 2016) 通过在不同尺寸的特征图上使用不同大小的预测框, 实现在一次网络传递中检测不同尺寸的目标, 从而提高了目标检测的实时性。然而, SSD 在处理多层特征图时忽略了不同层级特征图之间的相关性, 导致图像特征提取不够丰富。另一种基于回归分析的端对端目标检测算法是 YOLO 系列算法。YOLOv1 (Redmon 等, 2016) 是 YOLO 系列算法的初始版本, 其通过将图像划分为多个网格, 直接对网格中的目标进行识别, 无需生成候选区域, 检测速度较

快, 但是精度相对较低。为解决 YOLOv1 的不足, YOLOv2 (Redmon 和 Farhadi, 2017) 通过将全连接层替换为卷积层, 并引入批归一化, 加速了训练过程并提升了网络的稳定性。Zhang 等人 (2017) 提出的多阶段渐进框架 (multi-phase progressive salient object detection, MP-SOD) 方法利用多尺度特征融合和金字塔空间池检测不同大小的显著对象区域, 大幅提升了检测性能。张曼等人 (2020) 将特征金字塔 (feature pyramid network, FPN) 模型融合进 YOLOv2 中, 提出了多尺度网络结构, 并且裁剪大幅遥感图像进行检测, 解决了目标检测网络无法处理大幅遥感图像的问题, 但是对于遥感图像中的微小目标检测精度仍然较低。为了解决上述问题, Redmon 和 Farhadi (2018) 对于 YOLOv2 进一步改进, 提出了 YOLOv3, 将图像的多尺度特征融合和残差网络结构引入网络, 从而提升目标检测性能。然而, YOLOv3 的特征提取模块仍使用较为简单的 Darknet-53, 未能优化深层特征的提取能力。曲振方和朱福珍 (2022) 将双重特征提取网络引入 YOLOv3 的主干网络中, 采用注意力机制对主干网络和辅助网络之间的特征进行融合, 将改进型的 YOLOv3 引入遥感图像处理领域, 提升了网络的处理效能, 相较于 YOLOv3 平均精度均值 (mean average precision, mAP) 增加了 8.68%。吴杰等人 (2021) 在 YOLOv3-tiny 网络 (Redmon 和 Farhadi, 2018) 的基础上提出 DS-YOLO 网络, 引入了深度分离可卷积操作, 对遥感图像的各个通道分别进行卷积操作, 将各个通道之间的特征信息进行融合, 提升了网络在目标小、动态范围大以及背景环境干扰强的遥感图像的目标检测能力。Bochkovskiy 等人 (2020) 将主干网络改为 CSPDarknet, 提出了 YOLOv4, 通过跨阶段部分连接降低了网络的计算量, 并对多级的图像特征进行融合, 提高了对微小目标的检测能力。与此同时, 应用 Mish 激活函数提高了网络的准确性和泛化能力, 这些改进都使得 YOLOv4 在目标检测速度和目标检测精度达到了平衡。刘继和杨军 (2022) 对 YOLOv4 的主干网络中的卷积层应用密集连接和稀疏连接两种连接方式, 通

过深层特征信息与浅层特征信息的融合,实现了对遥感图像的全局信息与局部信息的提取,缓解了遥感图像中密集小目标误检、漏检的压力(袁翔等,2023)。YOLOv4发布仅一个月后,Jocher等人(2023)发布了YOLOv5,YOLOv5根据模型参数规模不同分为不同的版本,如YOLOv5n、YOLOv5s、YOLOv5m等,YOLOv5s是其中最轻量、检测速度最快的模型。成倩等人(2023)利用改进型的路径聚合网络对YOLOv5s模型进行改进,设计的图像信息特征增强模块使得网络能够提取更加丰富的浅层特征,增强了对遥感图像中微小目标的提取能力。刘晶宇和杨鹏(2023)使用空间金字塔池化特征融合模块(spatial pyramid pooling feature aggregation module, SPPFAM),对YOLOv5s中C5层的特征进行提取,并且将模型的特征注意力集中到目标区域,减少了遥感图像复杂噪声对于图像目标提取的干扰。目前YOLO系列中较为流行的版本为YOLOv8和YOLOv11n。YOLOv8虽然在精度方面能够满足对遥感图像目标检测的要求,但是该模型还是过于庞大,对星上资源占用率较高。而YOLOv11n对模型的深度和宽度都进行了较大的调整,通过提出的C3k2机制对网络结构进行创新,减少了模型的参数量和计算量,其更高的轻量化设计使得YOLOv11n更容易部署到资源有限的环境中,因而受到广泛关注。

尽管YOLO系列算法中的YOLOv11n在一定程度上具有优势,但在天基遥感成像的实际工程应用中,由于特定成像场景下目标小而密集,卫星遥感图像目标检测仍会面临微小目标漏检与错检等问题,且天基遥感成像对目标检测算法的检测速度与精度要求较高,亟需能够在确保检测精度的同时提高检测速度的解决方案。

综合考虑卫星遥感图像中密集微小目标的检测和识别需求,以及现有模型在精度、速度和计算资源上的平衡困难,本文聚焦于YOLOv11n算法的改进,针对卫星遥感图像目标检测任务设计了一种轻量化高精度算法。

1)本文提出了一种融合空间与通道多注意力机制和上下文信息的YOLOv11n遥感图像目标检测算法。在YOLOv11n的骨干网络中加入通道优先卷积的注意力模块来引导主干网络,通过为每个通道分配不同的注意力权重,有效增强了遥感图像中微小

目标特征的代表能力,进而提高了检测精度。通过合理的注意力分配,所提方法能够更准确地选择与微小目标相关的特征通道,优化了特征提取过程。

2)为了更好地适应天基遥感实际应用中的轻量化需求,本文提出了一种基于部分卷积(partial convolution, Pcov)的二维卷积层,以替代YOLOv11n中的传统二维卷积层。通过对主干网络卷积层的轻量化优化,不仅保持了增加的模型检测精度,而且显著减少了模型的参数量,进而提升了运算速度。

3)面向天基遥感领域的实际应用需求,本文在完成改进型YOLOv11n模型性能优化的基础上,采用昇腾计算芯片进行模型部署,优化了模型与计算芯片的匹配,进一步提升了遥感数据处理的效率。该研究对新型网络的实际工程化应用提供了有价值的参考,也为AI(artificial intelligence)赋能遥感技术提供了新的思路。

## 1 相关工作与研究思路

### 1.1 注意力机制

随着深度学习的发展,神经网络借鉴人类视觉注意力机制的原理,在处理图像特征信息时,能够将更多的计算资源投入到所关注的细节信息上,同时抑制不感兴趣的细节信息,实现有限计算资源的合理分配。得益于其可解释性和高效性等优势,各种注意力机制的算法应运而生,例如空间注意力机制和通道注意力机制等(赖杰等,2024)。

空间注意力机制通过将特征图的空间位置作为注意力加权维度,利用与图像特征图二维空间尺寸相匹配的注意力权重矩阵,根据特征图中不同位置的关联性,对特征信息进行加权融合,并通过打分得到空间掩码,从而实现图像不同区域的权重分配。非局部神经网络(Wang等,2018)通过计算特征图中所有位置特征的加权和,对特征图中的特定位置进行响应,以捕捉长距离依赖关系,进而解决了传统卷积神经网络的网络深度冗余的问题。通道注意力机制通过将网络的注意力集中在特征图的不同通道上,使每个通道作为类特征检测器来捕获特征图不同通道之间的信息。通过学习特征信息,网络能够自动调整输入信息的关注度,提升网络对特征图中关键信息的敏感度。通道注意力的代表性算法如挤压与激励模型(squeeze-and-excitation network, SENet)

(Hu 和 Ramanan, 2017), 该网络通过挤压操作实现特征图全局信息的压缩, 在通道维度上进行特征学习, 自适应地调整各个通道之间的特征响应, 最后通过激励机制对每个通道分配不同的特征权重提高网络的表征能力。结合以上两种注意力的特点, Woo 等人(2018)提出了卷积块注意力模型(convolutional block attention module, CBAM)。该模型首先通过通道注意力机制对特征图中的特征表达进行增强, 然后通过空间注意力机制实现特征图重要区域特征的表达, 实现了模型对关键特征信息的高效捕捉, 提高了网络性能。

### 1.2 网络轻量化设计

神经网络的轻量化设计在尽可能减少对网络精度影响的前提下, 通过降低网络的计算复杂度和模型参数量, 实现网络在计算资源有限的设备上部署。何明一和保铮在《神经网络与信号处理系统: 有限精度设计理论》(何明一和保铮, 1998)一书中, 系统地研究了网络结构、灵敏度、误差界等方面的有限精度设计理论, 为神经网络轻量化、低功耗研究和工程应用提供了基础理论。近年来, 关于深度神经网络的轻量化设计得到广泛关注, 并提出了参数压缩、稀疏表示等方法。Sandler 等人(2018)提出了一种轻量化深度卷积网络结构设计方法, 通过改变卷积结构减少计算量和参数量的网络模型, 在网络的中间展开层使用深度分离可卷积操作对图像特征进行非线性过滤, 将特征信息的空间相关性和通道相关性进行分离, 减少了网络的参数量, 实现了网络的轻量化。

### 1.3 研究思路

受以上启发, 本文在 YOLOv11n 基础上, 研究通过融合注意力机制来提高网络对天基遥感图像中微小目标检出的精确度, 其中还通过二次具有轻量化的部分二维卷积结构的处理(即 Pconv 模块)大大减少了网络参数量, 实现天基遥感图像密集微小目标检测的轻量化高精度检测算法及其实验验证。

## 2 本文算法

### 2.1 网络结构

本文网络以 YOLOv11n 主干网络为基础, 对其内部注意力模块进行了改进。YOLOv11n 网络结构主要包括 4 个部分: 输入层、骨干网络(Backbone)、

颈部架构(Neck)和输出层(Head)。

在输入层部分, YOLOv11n 与先前版本类似, 对处理数据进行缩放、归一化等图像预处理工作。YOLOv11n 中的骨干网络负责对输入端的图像数据进行多尺度的特征提取, 通过堆叠的卷积层和特征提取模块生成特征图。相较于 YOLOv8 模型, YOLOv11n 创新性地采用 C3k2 模块代替 YOLOv8 模型的 C2f 模块, C3k2 采用可变卷积核的设计, 通过调整卷积核的大小改变感受野的范围, 使得模型能够精确地捕捉图像中的上下文信息, 尤其适合天基遥感图像中复杂背景条件下的目标检测任务。C3k2 模块将输入层的特征信息进行两部分操作, 一部分是通过卷积操作直接进行特征参数的传递, 另一部分通过多个 C3k 模块进行特征的提取, 最终将两部分的特征提取结构通过  $1 \times 1$  卷积核进行二者的融合, 这样的结构设计在进行特征提取更加高效的情况下保证了一定的轻量化设计。

颈部 Neck 结构位于 Backbone 和 Head 输出层之间, 能够改善网络的鲁棒性和多样性。YOLOv11n 网络在 SPPF 模块后面增加了 C2PSA 模块, 该模块将 PSA (position-sensitive attention) 块引入到 C2f 模块中, 相当于在 C2 机制中嵌入了多头注意力机制, 通过多头注意力的结构和前馈神经网络增强对图像特征的提取能力。

在遥感图像中, 被检测的目标所占的像素较小并且对图像的检测速率要求较高, 所以 YOLOv11n 中结构相对复杂的 C2PSA 不适用于该下游任务。本文在 YOLOv11n 网络结构基本架构上进行了改进, 如图 1 所示。输入遥感图像的特征信息通过输入层进入改进型 YOLOv11n 的主干网络, 经过两层设计的部分卷积的二维卷积层 Pconv, 再经过 C3k2 结构块后, 在 CPCA (channel prior convolution attention) 进行多尺度特征信息注意力的提取, 主干网络堆叠 3 层此结构。改进型网络的瓶颈层中利用上采样结构和两次轻量化的部分二维卷积模块 Pconv 结构以及两次 C3k2 结构, 在背景复杂的遥感图像中挖掘微小目标的特征信息, 解决了多次的卷积操作和下采样迭代造成的图像信息严重丢失的问题。

### 2.2 混合注意力模块

遥感目标一般尺寸较小, 且图像中被检测目标的背景复杂, 需要从遥感图像的通道与空间两个维度对被检测目标特征信息进行提取。因此, 为了增

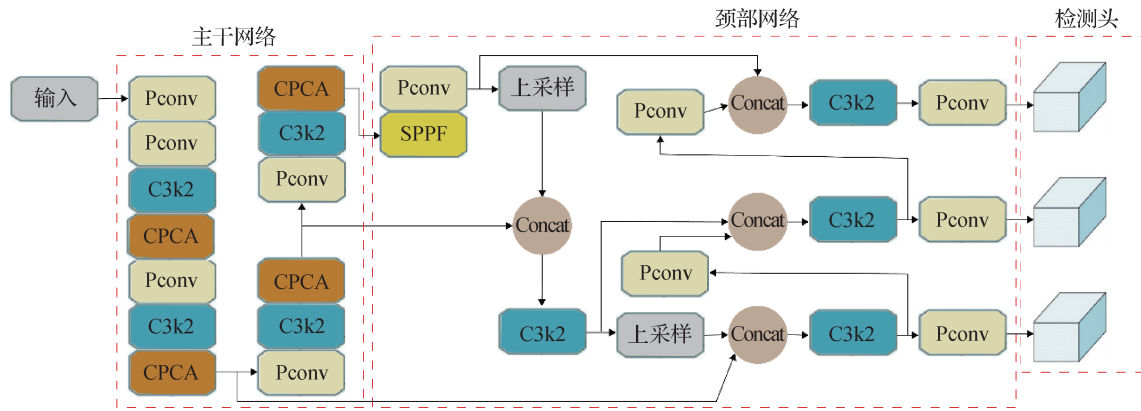


图1 改进型YOLOv11n网络结构

Fig. 1 Architecture of improved YOLOv11n network

强检测模型对遥感目标的检测能力,在YOLOv11n骨干网络 Backbone 中加入通道优先卷积注意力模块 (Huang 等, 2024)。CPCA 通过为每个通道分配不同的注意力权重,增强了微小目标相关特征的表示,从而提高了检测准确性。通道先验机制的引入,使得注意力分配更加合理,网络能够更好地选择与微小目标相关的特征通道,提升了微小目标的检测效果。CPCA 注意力机制结构可分为通道注意力和空间注意力两个模块。

通道注意力模块根据各通道的重要程度动态分配权重,计算方法为首先采用平均池化和最大池化操作,获得每个通道的最大特征图和平均特征图,再将得到的两个特征图分别输入到多层感知器层 (multi-layer perceptron, MLP),通过训练多层感知器的参数,网络可以自适应地决定各通道的权重系数,增强包含重要信息的通道权重,最后,将两个特征图相加并经过 GELU (Gaussian error linear units) 激活函数处理得到具有通道注意力的特征向量,具体计算为

$$CA(\mathbf{F}) = \sigma(MLP(f_{AvgPool}(\mathbf{F})) + MLP(f_{MaxPool}(\mathbf{F}))) \quad (1)$$

式中,  $\mathbf{F}$  为输入特征图, AvgPool 和 MaxPool 分别为平均池化和最大池化操作, MLP 为多层感知器,  $\sigma$  为 GELU 激活函数,  $CA(\mathbf{F})$  为通道注意力特征图。

空间注意力模块使用多尺度结构增强特征之间的空间关系,并且采用深度可分离卷积降低计算复杂性,最终使用  $1 \times 1$  的卷积进行通道混合,具体计算为

$$SA(\mathbf{F}) = Conv_{1 \times 1} \left( \sum_{i=0}^3 Branch_i(DwConv(\mathbf{F})) \right) \quad (2)$$

式中,  $\mathbf{F}$  为输入特征图,  $Conv_{1 \times 1}$  为  $1 \times 1$  卷积,  $Branch_i$  为第  $i$  个分支,  $DwConv$  为深度可分离卷积,  $SA(\mathbf{F})$  为

空间注意力模块的输出结果。

整个计算可以表示为

$$\mathbf{F}_c = CA(\mathbf{F}) \otimes \mathbf{F} \quad (3)$$

$$\mathbf{Y} = SA(\mathbf{F}_c) \otimes \mathbf{F}_c \quad (4)$$

式中,  $\mathbf{F}$  为输入特征图,  $\mathbf{F}_c$  为通道注意力模块输出,  $\mathbf{Y}$  为注意力模块输出特征图,  $\otimes$  为逐元素相乘。

### 2.3 网络轻量化设计

在嵌入式 AI 目标检测领域,由于计算资源匮乏,对神经网络的计算量有着严格的限制,为了实现神经网络的低延迟和高响应速度,本文提出了一种基于部分卷积 (Pconv) 的二维卷积层,如图 2 所示,以替换 YOLO 中的二维卷积层。

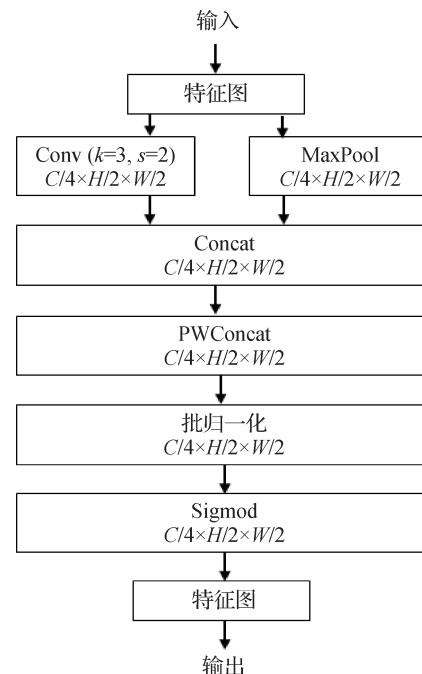


图2 轻量化二维卷积结构——Pconv

Fig. 2 Lightweight 2D convolutional structure (Pconv)

Pconv 仅对部分输入通道进行卷积, 考虑到通道间特征图中的冗余参数以及保持内存的连续性, 选取前段连续  $c_p$  个通道代表整个特征图。那么部分卷积的参数量 (Params) 为

$$Params_{pconv} = k^2 \times c_p^2 + c_p \quad (5)$$

式中,  $k$  为卷积核尺寸,  $c_p$  为选取的通道数。

逐点卷积的参数量为

$$Params_{pwconv} = 1^2 \times c^2 + c \quad (6)$$

式中,  $c$  为输入特征图的通道数。

总参数量为

$$Params = k^2 \times c_p^2 + c_p + c^2 + c \quad (7)$$

式中,  $c$  为输入特征图的通道数,  $c_p$  为选取的通道数。

而普通二维卷积的参数量为

$$Params_{conv} = k^2 \times c^2 + c \quad (8)$$

式中,  $k$  为卷积核尺寸。

且 Pconv 的计算量, 即每秒浮点运算次数 (floating point operations per second, FLOPs) 为

$$FLOPs = h \times w \times k^2 \times c_p^2 \quad (9)$$

式中,  $h$  为特征图高度,  $w$  为特征图宽度。

而普通卷积的计算量为

$$FLOPs = h \times w \times k^2 \times c^2 \quad (10)$$

内存访问量为

$$Memory = h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \quad (11)$$

可以看出, 若选取的通道数  $c_p$  为输入通道数的 1/4, 则计算量为普通卷积的 1/16, 内存访问量为普通卷积的 1/4, 可以降低卷积运算所需算力, 增强算法的实时性。

为了充分利用剩余通道的信息, 在部分卷积后连接一个逐点卷积 (pointwise convolution, PWConv), 通过在通道维度上的  $1 \times 1$  卷积运算, 融合通道间的信息。

Pconv 与 PWConv 连接后, 计算量 (FLOPs) 变为

$$FLOPs = h \times w \times (k^2 \times c_p^2 + c^2) \quad (12)$$

式中,  $c$  为输入特征图的通道数。

从式 (8) 中可以看出, 当  $c_p = \frac{1}{4}c$ ,  $k = 3$  时, 本文的模型参数量是原始模型卷积层的 17.36%, 且该轻量化结构可以降低 84% 计算量, 在一定程度上解决了网络模型在嵌入式设备中部署的难题。由于 YOLOv11n 的卷积层的卷积步长  $s = 2$ , 导致输出特征图的尺寸为输入特征图的一半, 为了保证部分卷积两个分支的尺寸相同, 本文对未做卷积的通道进

行最大池化操作, 将特征图的尺寸缩小为原来的一半。通过逐点卷积发挥各个通道的特征信息的表征能力, 这样的设计在减少模型计算量的同时, 不会显著降低模型的特征提取能力。

### 3 实验结果与分析

#### 3.1 实验设计和实验数据集

本文实验采用的硬件显卡为 Nvidia RTX 6000 Ada, 显存为 48 GB, PyTorch 版本为 1.10, CUDA 版本为 11.8, 编程语言为 Python 3.8。由于实验平台显卡显存为 48 GB, 为提高模型训练稳定性以及泛化能力, 实验对处理的遥感图像, 批量大小设置为 32, Epoch 为 100, 为平衡模型的收敛速度, 优化模型训练的稳定性, 通过实验验证, 本文选择最适合的学习动量和学习率, 学习动量设置为 0.937, 学习率设置为 0.01。

为了验证改进算法的有效性, 本文选择 DIOR (object detection in optical remote sensing images) 数据集进行实验。DIOR 数据集的目标标注采用 LabelMe 工具进行, 标注格式为 VOC 格式。为了避免大目标对微小目标检测精度的干扰, 本文筛选出了包含微小目标标注信息的图像。图像尺寸为  $800 \times 800$  像素, 训练集包含 11 725 幅图像, 测试集包含 11 738 幅图像, 数据集中包含 20 个目标类别, 如飞机、船舶、车辆等常见遥感目标。在将图像输入模型之前, 对其进行了缩放、翻转等一系列预处理操作, 以提高训练效果。

#### 3.2 评价指标

为了全面评估所提方法的性能, 实验选取了精确度 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、参数量、计算量 (FLOPs) 和模型推理速度作为模型的评价指标。具体计算为

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$AP = \int p(r) dr \quad (15)$$

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \quad (16)$$

式中,  $TP$  为模型预测正确的正样本,  $FP$  为模型预测

错误的负样本,  $FN$  为模型预测错误的正样本,  $AP$  为检测模型精确率—召回率曲线(P-R曲线)的面积。

参数量指模型的参数个数,与模型占用的内存成正比;计算量反映了模型所需算力,计算量越小,模型运算速度越快;模型推理速度指在给定输入数据下,模型完成输出结果的耗时,反映了模型的实时性。

### 3.3 对比实验

为了验证本文改进型YOLOv11n算法中注意力机制的有效性,在DIOR数据集上进行了相关对比实验。基于YOLO模型具有实时性高、全局感知能力强的特性,使得YOLO模型在遥感图像目标识别任务中广泛应用。因此本文选择将添加了CPCA注意力机制的模型与YOLOv11n原始模型、添加了常用注意力机制的YOLOv11n模型、YOLO系列算法中

较为经典的YOLOv8n以及目标检测领域较新的算法RTDETR(real-time detection transformer)(Zhao等,2024)、FFCA-YOLO(feature enhancement, fusion and context aware YOLO)(Zhang等,2024)进行对比,在保证各模型数据集、硬件配置、软件平台和评价指标一致的前提下开展实验。对比共分为8组,如表1所示。“YOLOv11n\_CBAM”代表在YOLOv11n模块中添加CBAM注意力模块,“YOLOv11n\_GAM”代表在YOLOv11n模块中添加全局注意力模块(global attention mechanism, GAM)(Liu等,2021);“YOLOv11n\_EMA”代表在YOLOv11n模块中添加高效多尺度注意力(efficient multi-scale attention, EMA)(He,2023);“本文”代表在YOLOv11n模块中添加CPCA注意力模块。

表1 不同算法在DIOR数据集上对比

Table 1 Comparison of different algorithms on the DIOR dataset

模型	精确率	召回率	mAP_0.5	参数量/M	FLOPS/B
YOLOv8n	0.768	0.626	0.680	3.01	8.20
YOLOv11n	0.799	0.642	0.707	2.59	6.50
YOLOv11n_CBAM	0.804	0.643	0.709	2.74	6.60
YOLOv11n_GAM	0.811	0.652	0.720	3.60	7.90
YOLOv11n_EMA	0.804	0.646	0.710	2.62	6.70
RTDETR	0.781	0.648	0.615	9.42	16.30
FFCA-YOLO	0.792	0.651	0.620	<b>1.37</b>	11.20
本文	<b>0.819</b>	<b>0.652</b>	<b>0.722</b>	2.11	<b>5.80</b>

注:加粗字体表示各列最优结果。

从表1可以看出,在DIOR数据集上,YOLOv8n的各项结果都低于YOLOv11n。并且目标检测领域较新的算法RTDETR、FFCA-YOLO也比增加注意力机制的YOLOv11n算法得分较低。相对于传统的YOLOv11n模型,添加不同注意力模块的改进网络均能带来性能提升,例如“YOLOv11n\_CBAM”模型的精确率为0.804,召回率为0.643,mAP\_0.5为0.709,但是针对YOLOv11n,其精确率仅提升了0.50%,召回率仅提升了0.10%,性能提升幅度相对较小。本文方法中,采用CPCA模型,相对于YOLOv11n,精确率提升了2.0%,召回率提升了1.0%,mAP\_0.5提升了1.50%,有效地提升了模型的性能,说明本文设计的注意力机制对于遥感图像中微小目标检测更具有效性。与此同时,通过对比

表1中各个模型的参数量和计算量,本文算法的参数量为2.11M,计算量为5.80B,在保证较高精确率和召回率的前提下,本文算法的轻量化设计更加适用于天基遥感图像处理领域。

为了更好地展示验证各算法的检测效果,将本文算法和先进的目标检测算法的准确率和召回率进行比较得到P-R图。P-R曲线图以召回率(R)为横坐标、以精确率(P)为纵坐标,在P=R处的点为图像的平衡点。平衡点值越大,表示模型的检测结果精度高且漏检少。

在DIOR数据集上,各模型对比分析P-R图,如图3所示,可以看出,本文算法的模型性能要高于基于YOLOv11n的其他网络模型,而基于YOLOv11n的网络模型性能要高于基于YOLOv8的网络性能,且

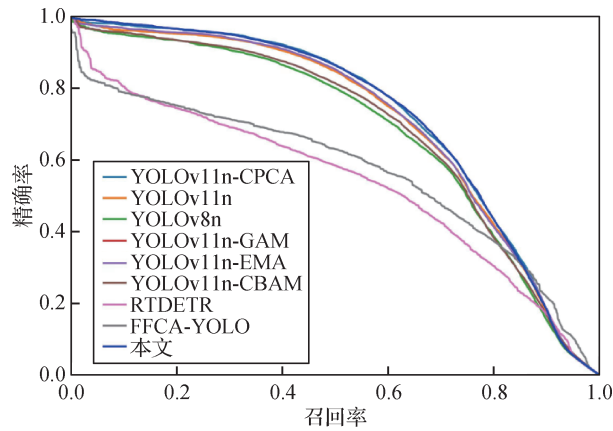


图3 不同算法在DIOR数据集上P-R曲线对比

Fig. 3 Comparison of P-R curves of different algorithms on the DIOR dataset

都优于其他网络模型性能。

### 3.4 改进型模型边缘端嵌入式部署验证实验

面向天基遥感领域的实际应用需求,为实现提出的改进型轻量化YOLOv11n模型能够更好地工程化应用,本文在NPU(neural network processing unit)上部署模型,进行可行性验证实验。为了实现NPU的模型部署,选用了华为Atlas200开发板的运行开发环境。Atlas中配备了一个昇腾310NPU作为计算芯片,与一个8核ARM-CPU作为控制芯片,昇腾系列加速器是华为公司推出的一款面向神经网络加速的DSA硬件。昇腾加速器是一个片上系统,共有两类AI计算引擎:AICore和AICPU,其中AICore提供神经网络算力,AICPU用于承担非矩阵类计算。NPU推理过程硬件配置如表2所示。

NPU推理过程软件开发环境中CANN为Ascend-cann-toolkit\_3.3.0.alpha006,Min-dStudio为MindStudio\_2.0.0\_beta3。

在本文中,NPU的推理过程主要分为5个步骤(如表3所示)。

表3 NPU推理过程步骤

Table 3 NPU inference process steps

推理步骤	推理过程
步骤1	安装模型权重pt文件、PyTorch库、CANN算子包等推理环境
步骤2	权重文件转换和数据处理
步骤3	处理后数据传入加载模型,对其进行推理,生成结果数据
步骤4	将推理的结果数据进行后处理,并与GPU的推理结果进行对比,若精度与GPU的推理结果相同,则证明模型部署成功
步骤5	性能测试,通过测试模型的推理速度以验证模型的性能

表2 NPU硬件配置

Table 2 NPU hardware configuration

硬件开发环境	
处理器	ArmCortex-A55 8-Core 1.6GHZ
内存	8 GB
NPU	2-CoreAscend310
操作系统	Ubuntu18.04

通过实验,将本文模型与YOLOv11n原始模型的推理时间进行比较,如表4所示。YOLOv11n的NPU推理耗时为19.6ms,而本文算法的NPU推理耗时仅为14.8ms,相对于YOLOv11n原始模型减少了4.8ms,推理速度提升了24.49%。另外,YOLOv11n在NPU部署的模型精确率为0.799,召回率为0.642,本文算法在NPU部署的模型精确率为0.819,召回率为0.652。由此可见,模型的迁移部署并不会对模型带来潜在的精度下降。通过模型部署可行性实验,验证了本文算法提出的基于部分卷积的二维卷积层模型轻量化设计能够有效地减少模型的推理时间,削弱由于引入注意力机制所带来的参数量过大而导致模型推理速度慢的问题。与此同时,验证了本文算法在NPU模型上有很好的鲁棒性,能够符合天基遥感图像处理的工程化应用。

### 3.5 消融实验结果

为验证本文提出的增加CPCA模块以及Pconv模块对模型的影响,在DIOR数据集上进行消融实验。对比实验总共分为4组:分别为YOLOv11n、YOLOv11n+CPCA、YOLOv11n+Pconv以及本文算法模型。4组方法综合对比结果如表5所示,结论如下。

1)在YOLOv11n模型基础上只增加CPCA模块时,改进型注意力机制使得模型对遥感图像微小目

表4 NPU推理时间  
Table 4 NPU inference time

模型	NPU推理时间/ms	精确率	召回率
YOLOv11n	19.6	0.799	0.642
本文	<b>14.8</b>	<b>0.819</b>	<b>0.652</b>

注:加粗字体表示各列最优结果。

标特征提取上更加充分和全面,精确率提升2.0%,召回率提升1.0%,能够更好地表征天基遥感图像中微小目标的特征信息。但是,由于增加了CPCA注意力机制,导致模型的参数量增加了0.18 M,计算量增加了0.6 GFLOPs。

2)在YOLOv11n模型上只增加Pconv模块时,相对于YOLOv11n模型,Pconv模块的轻量化设计在保

证模型对微小目标检测精度的前提下,能够将模型参数量减少0.56 M,将计算量减少1.4 GFLOPs,提升模型的推理速度。

3)本文提出的算法,相比于只加CPCA注意力模块,虽然精确率略小0.10%,但是模型的参数量降低了0.66 M;相比于只加Pconv模块,虽然参数量略大0.08 M,但是模型的精确率提升了1.7%。

4)综合考虑,针对天基遥感图像中微小目标检测任务,本文提出的算法中注意力模块增强了模型对遥感图像关键特征的捕捉能力,通过动态分配权重,使模型关注更有价值的特征信息,显著提升遥感图像中目标特征信息的表征质量,即使增加了Pconv模块轻量化设计,在降低了模型参数量的前提下,也能够实现更高的检测精度。

表5 消融实验结果  
Table 5 Ablation experimental results

模型	精确率	召回率	mAP_0.5	参数量/M	计算量/GFLOPs
YOLOv11n	0.799	0.642	0.707	2.59	6.5
YOLOv11n + CPCA	<b>0.819</b>	0.652	<b>0.722</b>	2.77	7.1
YOLOv11n + Pconv	0.801	0.643	0.708	<b>2.03</b>	<b>5.1</b>
本文	0.818	<b>0.654</b>	<b>0.722</b>	2.11	5.8

注:加粗字体表示各列最优结果。

为了直观展示消融实验中各个模块的有效性,本文使用梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)热力图的方法(Selvaraju等,2017)来可视化YOLOv11n、YOLOv11n+CPCA和本文算法这3类模型的目标识别特征。该方法利用图像特征图中保留原始位置信息的特点,采用梯度加权类激活映射,利用热力图的方式叠加到原始图像上直观展示被关注区域。热力图中颜色的深浅反映数据的强度,颜色越深的区域表明数据点越密集,颜色越浅的区域表明数据点越稀疏,如图4所示,红色越深的区域表示关注程度越高。融合CPCA注意力模块的可视化效果显示该模块可以引导模型更加集中关注遥感图像中对微小目标识别更有意义的区域。如图4(c)中可以看到,模型将篮球场和停车场的车辆标记为颜色较深的区域,表明模型准确地关注到遥感图像中篮球场和停车场的车辆实际位置区域,而图4(a)YOLOv11n算法对篮球场以及停车场的车辆存在漏检的情况,说明本文算法

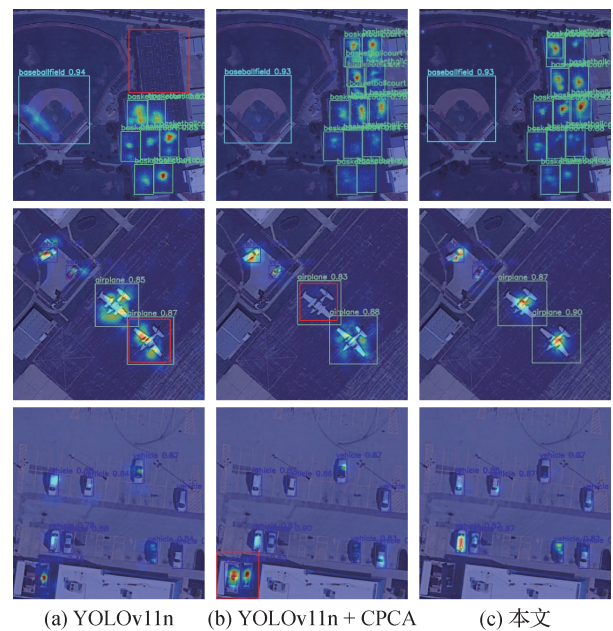


图4 消融实验中不同算法的热力图可视化结果  
Fig. 4 Heatmap visualization results of different algorithms in the ablation experiment ((a) YOLOv11n; (b) YOLOv11n + CPCA; (c) ours)

对于目标的检测精确率更高;图4(b)中只添加注意力机制的网络将停车场左下角的房屋误检为车辆,图4(c)展示了本文算法不存在误检目标,表明本文算法对复杂背景和目标特征的理解能力更强。并且,图4(c)中热力图更精准地捕捉到机场跑道上被探测飞机的边缘信息与形状特征,更符合其特征,增强了对目标的提取能力;图4(b)中只添加注意力机制的网络,模型反而只关注飞机某一集中区域,并未准确捕捉飞机的边缘与形状;图4(a)YOLOv11n算法虽然检测出飞机,但是热力图对其中一架飞机未显示为高亮区域。以上结果都说明了本文模型能够

在遥感图像的复杂背景下有效提取出微小目标的关键特征,具有更高的精准度、稳定性和泛化能力。

图5为消融实验中各算法在DIOR数据集的混淆矩阵。混淆矩阵可以直观地展示模型在不同目标类别上的预测结果,并且能够体现预测结果的概率分布。

图5(c)为本文算法结果。可以看出,在对角线上的值最高,表示本文算法的目标识别性能更好。非对角线上的概率值相较于其他模型较低。综上所述,本文模型能够在学习到各类目标更多信息的同时,也增加各类目标之间的特征区分度。

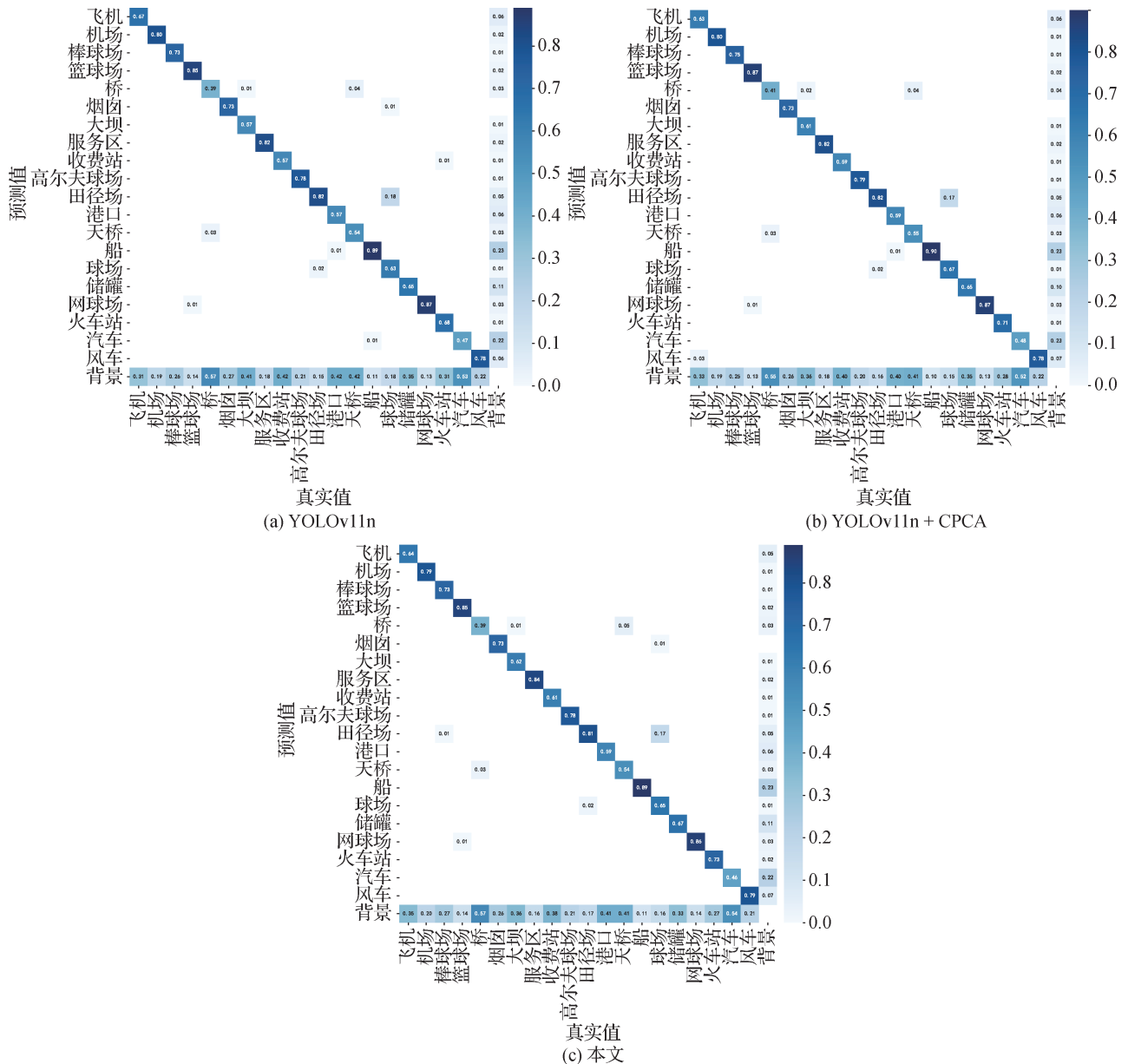


图5 消融实验中不同算法的混淆矩阵

Fig. 5 Confusion matrices of different algorithms in the ablation experiment ((a) YOLOv11n; (b) YOLOv11n + CPCA; (c) ours)

### 3.6 识别可视化结果

为了直观展示本文改进算法相较于经典目标检测算法(如YOLOv8、YOLOv11n)的优势,将3类算法进行了比较。图6为YOLOv8、YOLOv11n和本文算法的部分实验结果。通过图6可以看到,针对遥感图像中无论是港口处的微小船只目标或者是停车场的车辆,还是图像背景较为复杂情况下的操场、停机坪上的飞机,本文模型对其检测效果更好,符合预期效果。

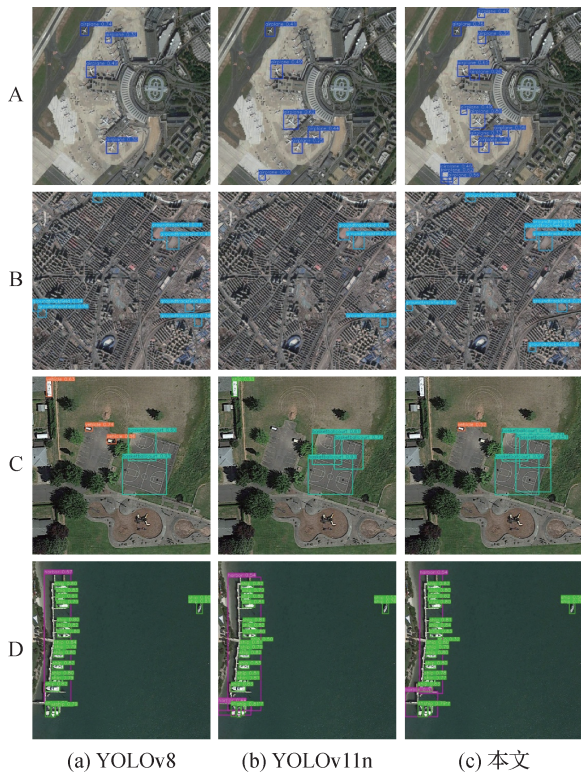


图6 不同算法的目标检测效果图

Fig. 6 Effect images of object detection for different algorithms  
(a) YOLOv8; (b) YOLOv11n; (c) ours

通过4个不同场景,即图6中A、B、C和D的实验对比可以看出,图6A是飞机密集停靠的机场停机坪,本文算法对于飞机的检测可以有效减小密集目标的漏检率。图6B是边缘细节较为模糊的操场,本文算法能够对目标进行识别并且能够准确地进行检测。针对图6C中检测目标大小不一致的情况,传统的YOLOv8和YOLOv11n对大小尺寸目标存在漏检情况,而本文算法通过改进空间与通道相融合的注意力机制模块,能够实现目标的完整检测。

表6给出了3种对比方法的平均精度均值mAP<sub>0.5</sub>。可见,本文算法相较于YOLOv8和YOLOv11n

算法,对不同目标的mAP<sub>0.5</sub>都是最高的。针对图6D中港口处的船只,mAP<sub>0.5</sub>达0.928,表明本文算法的性能优于其他两种。

表6 不同算法的mAP<sub>0.5</sub>

Table 6 The mAP<sub>0.5</sub> of different algorithms

目标类型	YOLOv8	YOLOv11n	本文算法
机动车	0.545	0.550	<b>0.558</b>
船	0.926	0.926	<b>0.928</b>
飞机	0.744	0.741	<b>0.817</b>
操场	0.752	0.781	<b>0.797</b>
港口	0.541	0.560	<b>0.599</b>

注:加粗字体表示各行最优结果。

## 4 结论

针对天基遥感图像微小目标检测任务,本文借鉴YOLOv11n模型,提出了一种融合多注意力机制的轻量化的天基遥感图像微小目标检测算法。首先,通过对YOLOv11n模型中的注意力机制进行改进,提出一种基于CPCA注意力模块的改进型模型架构,能够在空间和通道方向上实现对天基遥感图像微小目标的特征进行提取,避免或显著降低了对微小目标的漏检和错检率;其次,采用基于部分卷积的二维卷积层的设计,替换YOLO中的二维卷积层,避免了模型增加注意力模块带来的参数量过大的问题,参量降低了18.53%,实现了模型的轻量化;最后,通过对模型在NPU上的部署,优化了模型与计算平台的适配性,相比于原始的YOLOv11n模型,在保证目标检出精度的前提下,推理时间缩短了4.8 ms,能够满足实时监测的需求,模型计算资源和内存受限的天基工程化部署非常友好。

但是本文模型对于过于密集的微小目标也会存在漏检的现象,如港口密集的船只检测的场景。本文下一步工作将在以下两方面开展:1)进一步优化模型结构,针对模型采用更优的损失函数,如用Focal Loss解决类别不平衡,或用完全交并比(complete intersection over union, CIoU),以更适合密集微小目标的回归。另外,可以添加针对微小目标的辅助损失,如在浅层特征上增加监督,强化模型对微小目标的学习,从而提升极小目标的特征提取能力。

2)对卷积层以及注意力模块在模型的不同位置配置进行尝试,如将卷积层在特征提取阶段、特征融合阶段配置比例进行合理分配,将注意力模块配置在卷积层前、卷积层后,或嵌入卷积层进行优化设计,探究模型参数量、计算复杂度以及模型准确性之间的关系。

深度学习的进展无疑对图像目标识别等带来了技术上的突破,大数据与深度学习大模型在提升分类识别性能的同时存在算力需求和功耗巨大的严重问题。

现有轻量化的深度神经网络方法主要是面向资源受限条件的大模型参数压缩、稀疏表示等,能够有所减少参数,便于资源受限条件的部署。真正轻量化、低功耗的神经网络需要在范式、模型、算法和工程等方面实现综合考虑。例如,神经网络的光学学习、绿色神经网络(He, 2023),有可能攻克或者显著缓解这一问题。

## 参考文献(References)

- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2025-02-01]. <https://arxiv.org/pdf/2004.10934.pdf>
- Cheng Q, Li J and Du J. 2023. Ship target detection algorithm of optical remote sensing image based on YOLOv5. *Systems Engineering and Electronics*, 45(5): 1270-1276 (成倩, 李佳, 杜娟. 2023. 基于YOLOv5的光学遥感图像舰船目标检测算法. *系统工程与电子技术*, 45(5): 1270-1276) [DOI: 10.12305/j.issn.1001-506X.2023.05.02]
- Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]
- He M. 2023. Skip-connection neural networks-advance and prospect//Proceedings of 2023 IEEE International Conference on Cyber-Physical Social Intelligence (ICCS123). Xi'an, China: IEEE
- He M Y and Bao Z. 1998. *Neural Networks and Signal Processing Systems: A Finite Precision Design Theory*. Xi'an: Northwestern Polytechnical University Press (何明一, 保铮. 1998. *神经网络与信号处理系统: 有限精度设计理论*. 西安: 西北工业大学出版社)
- Hu P Y and Ramanan D. 2017. Finding tiny faces//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 1522-1530 [DOI: 10.1109/CVPR.2017.166]
- Huang H J, Chen Z G, Zou Y, Lu M, Chen C Y, Song Y Z, Zhang H Q and Yan F. 2024. Channel prior convolutional attention for medical image segmentation. *Computers in Biology and Medicine*, 178: #108784 [DOI: 10.1016/j.compbiomed.2024.108784]
- Jocher G, Chaurasia A and Qiu J. 2023. Ultralytics YOLO (Version 8.0) [EB/OL]. [2025-02-01]. <https://github.com/ultralytics/ultralytics>
- Lai J, Peng R H, Sun D X and Huang J. 2024. Detection of camouflage targets based on attention mechanism and multi-detection layer structure. *Journal of Image and Graphics*, 29(1): 134-146 (赖杰, 彭锐晖, 孙殿星, 黄杰. 2024. 融合注意力机制与多检测层结构的伪装目标检测. *中国图象图形学报*, 29(1): 134-146) [DOI: 10.11834/jig.221189]
- Liu J and Yang J. 2022. A lightweight dense connection network for object detection of remote sensing images. *Bulletin of Surveying and Mapping*, (10): 37-43, 55 (刘继, 杨军. 2022. 遥感影像目标检测的轻量化密集连接网络. *测绘通报*, (10): 37-43, 55) [DOI: 10.13474/j.cnki.11-2246.2022.0291]
- Liu J Y and Yang P. 2023. Remote sensing image target detection based on YOLO-V5 improvement. *Computer Era*, (7): 50-55 (刘晶宇, 杨鹏. 2023. 基于YOLO-V5改进的遥感图像目标检测. *计算机时代*, (7): 50-55) [DOI: 10.16644/j.cnki.cn33-1094.tp.2023.07.012]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer vision. Amsterdam, The Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0\_2]
- Liu Y C, Shao Z R and Hoffmann N. 2021. Global attention mechanism: retain information to enhance channel-spatial interactions [EB/OL]. [2025-02-01]. <https://arxiv.org/pdf/2112.05561.pdf>
- Qu Z F and Zhu F Z. 2022. Remote sensing image target detection based on improved YOLO v3. *Journal of Natural Science of Heilongjiang University*, 39(2): 231-237 (曲振方, 朱福珍. 2022. 基于改进的YOLO v3的遥感图像目标检测. *黑龙江大学自然科学学报*, 39(2): 231-237) [DOI: 10.13482/j.issn1001-7011.2022.03.004]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2025-02-01]. <https://arxiv.org/pdf/1804.02767.pdf>
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards

- real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Woo S Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2\_1]
- Wu J, Duan J, He L Q, Li Y C and Zhu W T. 2021. Research on aircraft detection algorithm of DS-YOLO network in remote sensing images. *Computer Engineering and Applications*, 57(1): 181-187 (吴杰, 段锦, 赫立群, 李英超, 朱文涛. 2021. DS-YOLO网络在遥感图像中的飞机检测算法研究. *计算机工程与应用*, 57(1): 181-187) [DOI: 10.3778/j.issn.1002-8331.1909-0409]
- Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, Yao X W, Huang Z L, Sun X and Han J W. 2023. Progress in small object detection for remote sensing images. *Journal of Image and Graphics*, 28(6): 1662-1684 (袁翔, 程焱, 李戈, 戴威, 尹文昕, 冯瑛超, 姚西文, 黄钟冷, 孙显, 韩军伟. 2023. 遥感影像小目标检测研究进展. *中国图象图形学报*, 28(6): 1662-1684) [DOI: 10.11834/jig.221202]
- Zhang J, Dai Y C, Porikli F and He M Y. 2017. Multi-scale salient object detection with pyramid spatial pooling//*Proceedings of 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. Kuala Lumpur, Malaysia: IEEE: 1286-1291 [DOI: 10.1109/APSIPA.2017.8282222]
- Zhang M, Li J, Ding R L, Cheng H T and Shen J. 2020. Remote sensing image object detection technology based on improved YOLO-V2 algorithm. *Computer Science*, 47(S1): 176-180 (张曼, 李杰, 丁茉莉, 成昊天, 沈霁. 2020. 基于改进 YOLO-V2 算法的遥感图像目标检测技术研究. *计算机科学*, 47(S1): 176-180) [DOI: 10.11896/jsjx.191100206]
- Zhang Y, Ye M, Zhu G Y, Liu Y, Guo P Y and Yan J H. 2024. FFCA-YOLO for small object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5611215 [DOI: 10.1109/TGRS.2024.3363057]
- Zhao Y A, Lyu W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, Liu Y and Chen J. 2024. DETRs beat YOLOs on real-time object detection//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, USA: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]

## 作者简介

李强,男,研究员,博士研究生,主要研究方向为遥感电子信息。E-mail: lqcast@sina.com

何明一,通信作者,男,教授,主要研究方向为神经网络人工智能、机器视觉与图像处理、高光谱遥感、数字化技术。

E-mail: myhe@nwpu.edu.cn

王智,男,工程师,主要研究方向为遥感图像处理。

E-mail: wz3323281646@163.com

崔书玮,男,硕士研究生,主要研究方向为遥感信息智能处理。E-mail: 2606161960@qq.com