

DOI: 10.13382/j.jemi.B2407208

基于改进 YOLOv7-tiny 的高效轻量遥感图像目标检测方法*

邝先验^{1,2} 程福军^{1,2} 吴翠琴^{2,3} 雷 卉^{1,2}

(1. 江西理工大学电气工程与自动化学院 赣州 341000; 2. 江西理工大学多维智能感知与控制江西省重点实验室 赣州 341000; 3. 江西理工大学机电工程学院 赣州 341000)

摘要:针对现有遥感图像目标检测方法在受资源限制的小型设备中检测精度不足问题,提出了一种基于改进 YOLOv7-tiny 算法的高效轻量遥感图像目标检测方法。首先,针对遥感图像小目标分布密集问题,设计了一种低跨度的上下文解耦检测头模块,通过融合深层和浅层特征分别实现目标检测的分类和回归任务,有效解决了遥感图像小目标漏检和误检的问题。同时,针对遥感图像目标多尺度问题,设计了一种并行级联注意力机制,通过并行三支网络与空间注意力模块相结合,增强了网络对多尺度目标特征的提取能力。此外,引入 Focal-EIoU 损失函数,提高模型泛化能力。对模型进行了对比实验、消融实验、部署实验和泛化实验,结果表明,在 DIOR-5s 和 NWPU VHR-10 数据集上的检测精度分别达到了 85.4%、90.6%,相较原模型分别提高了 2.6%、1.7%。且模型大小仅为 19.1 MB,检测速度为 64.1 fps,验证了算法的有效性。

关键词: 遥感图像; 目标检测; YOLO; 低跨度上下文解耦; 并行级联注意力

中图分类号: TP751.1; TN919.5 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Efficient and lightweight target detection method for remote sensing images based on improved YOLOv7-tiny

Kuang Xianyan^{1,2} Cheng Fujun^{1,2} Wu Cuiqin^{2,3} Lei Hui^{1,2}

(1. School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China; 2. Jiangxi Provincial Key Laboratory of Multidimensional Intelligent Perception and Control, Jiangxi University of Science and Technology, Ganzhou 341000, China; 3. School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract: To address the issue of low detection accuracy in remote sensing image target detection methods for small devices with limited resources, an efficient and lightweight method based on the improved YOLOv7-tiny algorithm is proposed. To address the issue of dense distribution of small targets in remote sensing images, a low-span context decoupling detection head module is designed. This module fuses deep and shallow features to perform classification and regression tasks for target detection. It effectively solves the problems of leakage and misdetection of small targets in remote sensing images. Meanwhile, a parallel series attention mechanism is proposed to enhance the network's ability to extract multi-scale target features for remote sensing image targets. This is achieved by combining the parallel three-branch network with the spatial attention module. Additionally, the model's generalization ability is improved through the introduction of the Focal-EIoU loss function. The comparison experiments, ablation experiments, deployment experiments and generalization experiments were conducted on the model. Experimental results indicate that the detection accuracy on DIOR-5s and NWPU VHR-10 datasets improved by 2.6% and 1.7%, respectively, compared to the original model. The model size is only 19.1 MB, and the detection speed is 64.1 fps, verifying the algorithm's effectiveness.

Keywords: remote sensing image; target detection; YOLO; low pitch context decoupling head; parallel series attention mechanism

0 引言

遥感图像目标检测是计算机视觉领域重要的研究方向之一,在国防军事和国民经济领域都有广泛的应用^[1]。传统的遥感图像目标检测方法包括尺度不变特征变换、方向梯度直方图(histogram of oriented gradient, HOG)、形变部件模型(deformable part model, DPM)等。通过手动设计特征来表示目标的形状、纹理和边缘等信息,然后使用分类器或检测器进行目标的识别和定位。依赖人工设计的特征,检测精度有限,实时性差,难以适应多样的地物类型和复杂场景。

随着深度学习的发展,基于深度学习的方法在遥感图像目标检测中取得了显著的进展^[2],能够更好地处理遥感图像中的复杂情况,并提升检测精度和实时性。可根据检测步骤将深度学习目标检测方法分为双阶段检测法与单阶段检测法。双阶段方法如 Faster R-CNN(region-based convolutional neural networks)^[3]、Mask R-CNN^[4]等,先生成候选区域,再对候选区域进行分类和定位。单阶段方法如 YOLO(you only look once)^[5-7]、SSD(single shot multibox detector)^[8]、RetinaNet^[9]等,直接在图像中进行目标检测和定位,一次完成目标位置和类别的预测。

双阶段检测法采用候选区域生成网络,能够生成较为准确的候选框,减少误检和漏检,提高对复杂场景的适应能力。但由于提取的候选区域目标框数量较多,候选区域的生成具有不确定性,所以处理速度较慢,对后续的分类和定位阶段造成一定的干扰。

基于回归的单阶段检测法不再受限于候选框的生成,能够快速地对目标进行检测和定位。这类算法网络结构简单,且同时兼顾检测精度和检测速度,因而在遥感图像目标检测中受到广泛关注与应用。Liu 等^[10]在 YOLO 算法的基础设计了一个连接不同深度层的特征映射网络,提高了检测精度,但仍难解决遥感图像中目标尺度差异大、目标分布密集、背景复杂等问题。陈欣等^[11]对 SSD 算法进行了改进,将分辨率高的浅层特征图与具有丰富语义信息的深层特征图进行融合,对先验框的比例进行调整,提高了遥感图像多尺度目标的检测精度,但对密集分布的小目标检测效果不理想。左露等^[12]在主干网络加入全局注意力机制,通过放大维度交互作用,提高模型对小目标的检测精度,但也随之增加了模型的计算量。Liu 等^[13]基于特征金字塔架构提出了通道滤波模块(CFM)和空间滤波模块(SFM),通过学习通道之间的相关性和引入全局信息来提高网络的检测性能。Luo 等^[14]在原始批量归一化模块中增加定心和缩放校准,增强有效特征,消除低分辨率特征层,在提高检测精度和速

度的同时更易于收敛。Han 等^[15]通过扩张卷积来捕获多尺度信息,并采用不同的扩张速率增大感受野。胡欣等^[16]设计了多分支注意力模块,通过跨维度融合信息提取遥感图像目标中的位置信息和语义信息,提高了预测准确率。

上述研究在遥感图像目标检测上有所突破,但在复杂场景下仍然存在许多挑战,例如遥感图像通常具有的小目标分布密集、目标多尺度等特点,容易导致误检和漏检。同时,上述大多数模型结构较复杂、体积较大,需要较多计算资源,使模型不易部署。本文针对以上问题,在轻量化模型 YOLOv7-tiny 基础上设计了更为高效的改进型遥感图像目标检测模型 CDA-YOLO(YOLOv7-tiny with context decoupled head and attention mechanism)。主要贡献如下:

1) 针对 YOLO 序列网络原始检测头的不足,设计了一种低跨度的上下文解耦目标检测头(low pitch context decoupling head, LPCD-Head),通过逐层融合深层和浅层特征,生成用于分类和定位的解耦特征图,在减小信息融合跨度的同时降低了小目标密集分布的漏检率。

2) 针对遥感图像目标多尺度问题,构建了一种并行级联注意力模块(parallel-series attention module, PSAM),采用两条 1×1 卷积分支和一条 3×3 卷积分支构成并行的三支路网络,并与空间注意力机制级联融合,提高了网络对多尺度目标特征的提取能力。

3) 将 Focal-EIoU 损失引入到模型中,并重新调整正负样本权重,增强模型的鲁棒性和泛化能力。

1 网络模型

1.1 网络整体模型结构

改进的 CDA-YOLO 网络模型结构如图 1 所示,首先,预设输入图像尺寸(本文设为 512×512),读取原始图像,保持宽高比,进行图像大小调整、裁剪或填充、归一化等预处理操作,将原始图像缩放为预设大小后输入网络。然后,经过两层 CBL 层初步提取细节特征,再经过 ELAN 和最大池化层(maximum pooling, MP)的进一步提取更深层特征。由于深层特征是相对抽象的语义特征,对于遥感图像而言,容易丢失浅层的多尺度位置信息,所以在 SPPCSPC 池化层后面加入 PSAM 进行多尺度信息聚合,提高网络对多尺度目标的感知能力。特征融合网络采用特征金字塔(feature pyramid network, FPN)结构,通过上采样和下采样操作将不同层次的特征图融合在一起,生成多尺度的特征金字塔,自顶向下部分主要是通过上采样和与更粗粒度特征图的融合,来实现不同层次特征的融合,而自下向上则是通过使用一个卷积层来融合来自不同层次的特征图。由于在目标检测中,分类任务和回

归任务所需的特征信息相反,当两个任务的特征信息来自同一层时,会导致两个任务之间相互影响,因此,设计了低跨度上下文解耦头替换原有的检测头,通过对不同

任务选取所需对应特征层的方法,可以有效降低网络对遥感图像中密集小目标检测的漏检率。

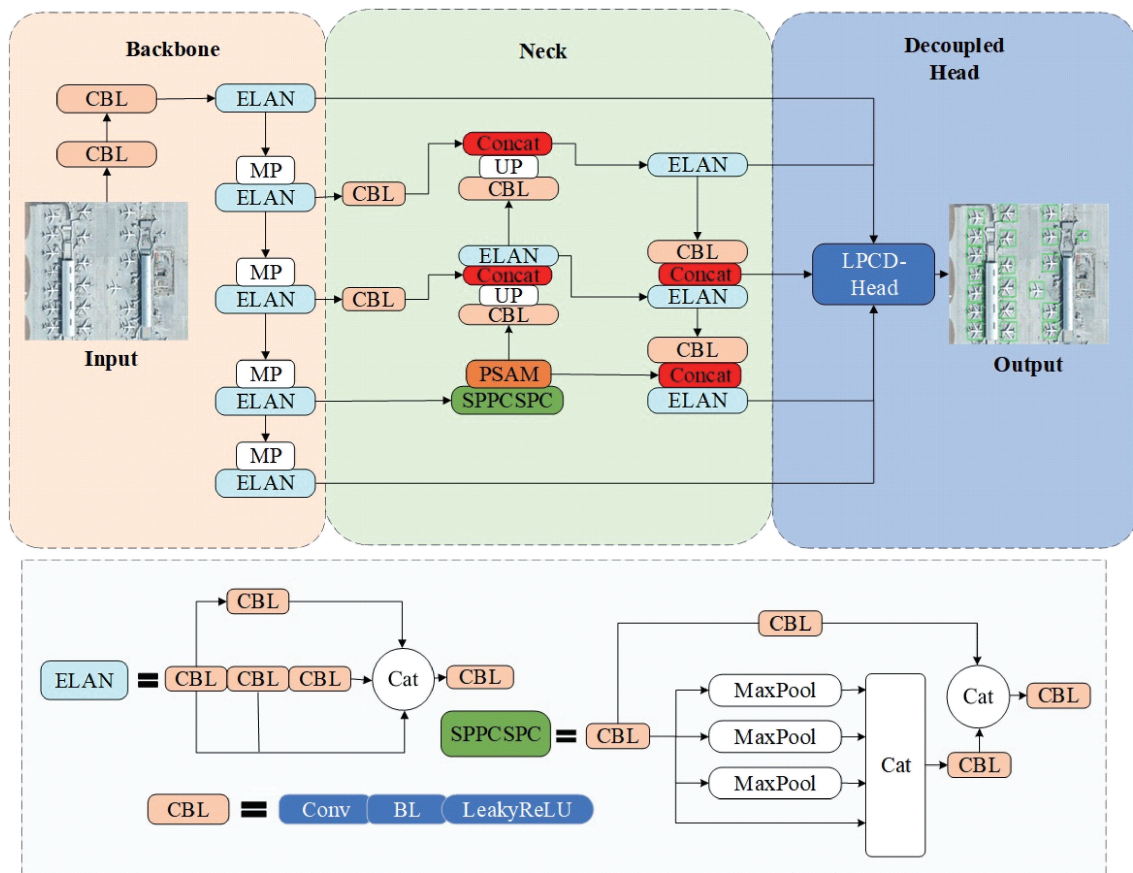


图 1 CDA-YOLO 网络结构

Fig. 1 CDA-YOLO structure

1.2 LPCD-Head 模块

耦合头和解耦合头是目标检测中常见的两种头部设计,用于从检测网络的特征图中提取目标位置和类别信息。其中耦合头通常是将卷积层输出的特征图直接送入几个全连接层或卷积层中,以生成目标位置和类别的输出,但由于其设计思路简单,需要大量的参数和计算资源,容易出现过拟合。Ge 等^[17]提出了解耦合头,通过将目标位置和类别信息分别提取出来,并使用不同的网络分支分别学习,最后再进行融合,可以有效减少参数数量和计算复杂度。但由于目标位置和类别信息来自同一层特征,而分类和回归是目标检测中两个高度相关但互相“矛盾”的任务,对特征信息的需求完全相反,这会导致分类和回归两个任务之间的相互影响,使得处理遥感图像中密集小目标任务时存在较高的漏检率。Zhuang 等^[18]提出一种的 TSCODE 网络,通过一种特定任务上下文解耦头来平衡分类和回归任务相互影响问题,但是该网络在

特征信息融合时跨度大,易发生“语义鸿沟”现象。

综合以上方法,设计了一种低跨度上下文解耦目标检测头 (LPCD-Head),其模块结构如图 2 所示。LPCD-Head 通过接收来自不同层级的特征图,并根据任务需求,逐层进行深层和浅层的特征融合,得到用于分类和回归的解耦特征图。在处理分类任务时,将 P_l 层进行下采样并与更深层的 P_{l+1} 层特征融合构成分类任务模块,在处理回归任务时,将更浅层的 P_{l-1} 层特征下采样与 P_l 层特征融合构成回归任务模块。最终将分类和回归 2 个任务模块通过不同的网络分支分别学习,组成低跨度上下文解耦检测头模块。

分类是一个相对粗颗粒的任务,通常要站在全局的角度去审时度势,需要丰富的语义信息,因此,通过步长为 2 的 3×3 卷积,将 P_l 层特征图缩小 1/2 后,得到特征图 p_m ,然后将 p_m 与含有更丰富语义信息的深层特征层 P_{l+1} 拼接,生成用于分类的特征图 G_l^{cls} ,如图 3(a) 所示。

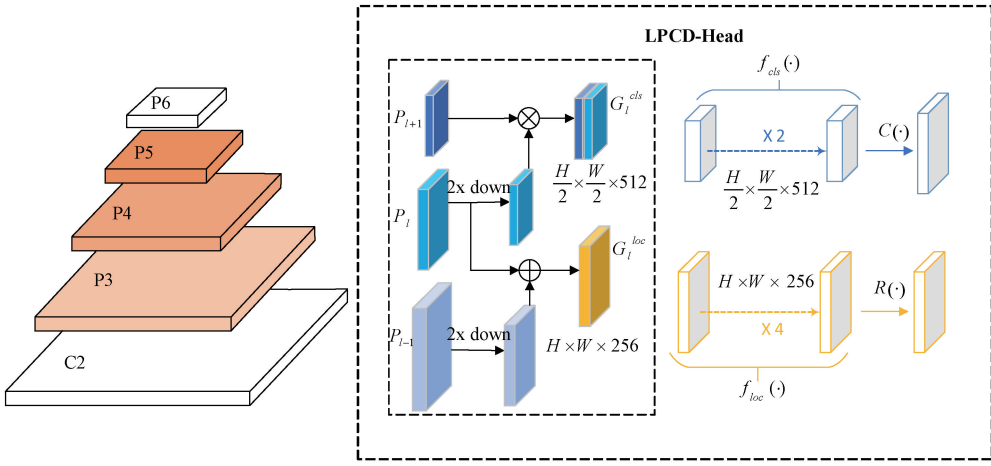


图2 LPCD-Head 模块

Fig. 2 LPCD-Head module

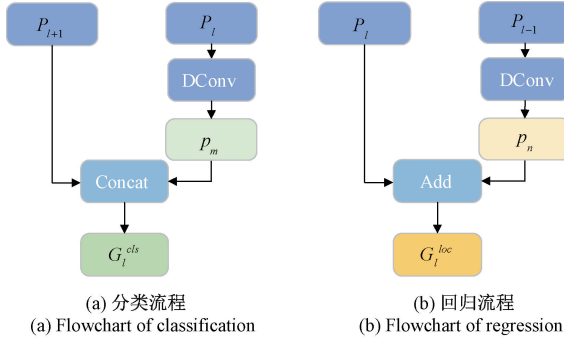


图3 分类与回归特征生成流程

Fig. 3 Flowchart of classification and regression feature generation

图3中 $DConv(\cdot)$ 表示降采样卷积层, $Concat(\cdot)$ 表示连接特征信息。 G_l^{cls} 会被输入到 $F_c(\cdot) = \{f_{cls}(\cdot), C(\cdot)\}$ 中进行分类预测。这里没有采用4个256个通道的卷积层,而是采用了一种浅而宽的 $f_{cls}(\cdot)$ 设计,采用2个512个通道的卷积层来保留更多的语义信息,可以减少额外的计算成本,提升分类的准确率。由于 G_l^{cls} 融合了 P_l 两步降采样的特征,因此将 P_l 中4个最相邻的分类平均得分值作为 G_l^{cls} 中每个位置 (x, y) 的预测值,记为 $\tilde{C} = R^{H_{l+1} \times W_{l+1} \times 4N}$, N 代表类别分类,然后根据式(1)将 $\tilde{C}$ 重新排列为 $\hat{C} = R^{H_l \times W_l \times N}$ 以恢复原分辨率。

$$\hat{C}[2x+i, 2y+j, c] = \tilde{C}[x, y, (2i+j)c], \forall i, j \in \{0, 1\} \quad (1)$$

这样即保留了 P_l 层特征的稀疏性,又融合了 P_{l+1} 层丰富的语义信息,对于遥感图像中分布密集,被严重遮挡的目标对象,具有更高效的类别推断能力。

与分类不同,回归是一个更细粒度的任务,它依赖于

更多的纹理细节和边界信息以更好的回归对象边界。浅层特征层 P_{l-1} 含有更多细节和边缘特征,通过步长为2的 3×3 卷积,将 P_{l-1} 的特征图缩小1/2后,得到特征图 p_n ,然后将 p_n 与 P_l 特征图聚合,生成用于定位的特征图 G_l^{loc} ,如图3(b)所示。

然后将 G_l^{loc} 输入到 $F_r(\cdot) = \{f_{loc}(\cdot), R(\cdot)\}$ 用于预测第 l 层金字塔的边界框,这样可以有效解决遥感图像中小目标定位困难的问题。

LPCD-Head 模块的流程步骤如下:

- 1) 输入特征: $P_l(H \times W)$, $P_{l-1}(2H \times 2W)$, $P_{l+1}(H/2 \times W/2)$;
- 2) P_l 经过卷积 $Conv(c_1, c_1, k=3, s=2)$ 得到 p_m ;
- 3) 将 p_m 和 P_{l+1} 通过 $concat(p_m, P_{l+1})$ 进行特征拼接,得到 G_l^{cls} ;
- 4) 将 P_{l-1} 经过卷积 $Conv(c_1, c_1, k=3, s=2)$ 得到 p_n ;
- 5) 将 p_n 与 P_l 进行 add 特征融合,得到 G_l^{loc} ;
- 6) 将 G_l^{cls} 输入到 $F_c(\cdot) = \{f_{cls}(\cdot), C(\cdot)\}$ 中进行分类预测;
- 7) 将 G_l^{loc} 输入到 $F_r(\cdot) = \{f_{loc}(\cdot), R(\cdot)\}$ 中进行回归预测。

在遥感图像目标检测任务中,通过针对性的对目标的分类和回归任务进行处理,在分类时融合更深层的语义信息,在回归时融合较浅层的细节信息,能有效解决遥感图像检测时小目标分布密集容易发生漏检和误检问题。

1.3 PSAM 模块

在对遥感图像进行目标检测时,特征提取的好坏很大程度上决定了最终的检测效果。针对遥感图像中多尺度目标的特征提取问题,通过构建3支路网络结构设计通道注意力(parallel channel attention, PCA)同时级联空

间注意力^[19] (spatial attention, SA) 的方法, 构建了一种并行级联注意力模块 (PSAM), 模块结构如图 4 所示。PSAM 能对遥感图像多尺度目标提取丰富的通道和空间信息, 突出遥感图像中重要区域, 抑制噪声和背景信息, 能有效提高遥感图像中多尺度目标的检测精度。

并行通道注意力模块 PCA: 为了增强遥感图像多尺度目标特征提取能力, 同时降低网络深度, 设计了一种并行 3 支路网络结构的通道注意力模块。该模块以两条

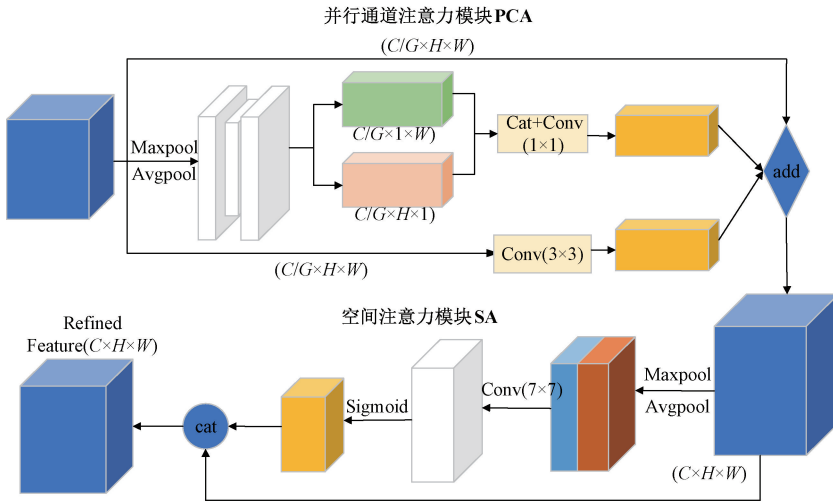


图 4 PSAM 注意力机制模块

Fig. 4 PSAM module

首先, 通过对两个编码特征沿图像高度方向进行连接, 共享相同的 1×1 卷积, 且不降低 1×1 分支的维度。然后将 1×1 卷积的输出分解为两个向量, 并使用两个非线性 Sigmoid 函数来拟合线性卷积后的二维二项分布。为了在 1×1 分支的两条平行路径之间实现不同的跨通道交互特征, 使用简单的乘法将每个组内的两个通道注意力图聚合在一起。同时, 3×3 分支通过 3×3 卷积捕获局部跨通道交互以扩大特征空间, 最后将各支路提取的特征信息通过 Concat 整合。

在特征分组时, PCA 模块将任意输入特征映射 $X \in R^{C \times H \times W}$ 沿着通道维度分成 G 个子特征, 以学习不同的语义信息, 每个子特征表示 $X_i \in R^{C/G \times H \times W}$ 其中一组通道的特征, 将其组合成 $X = [X_0, X_i, \dots, X_{G-1}]$ 。取 $G \ll C$, 使用学习到的注意力权重描述符来增强每个子特征中感兴趣区域的特征表示。

空间注意力模块 SA: 首先将 PCA 模块的输出特征图作为输入进行平均池化和最大池化操作, 分别生成平均值特征图和最大值特征图, 然后通过一个 7×7 的卷积将通道维度降为 1, 最后经过 Sigmoid 激活函数生成一个权重特征图, 并将该特征图与输入特征图进行加权, 得到最终的特征图。在最终的特征图中, 与遥感图像目标相

1×1 卷积分支和一条 3×3 卷积分支共同构建, 采用 CA 注意力机制^[20] 中 1×1 卷积的共享组件作为 1×1 分支, 该分支采用两个 1D 全局平均池化操作, 分别沿两个空间方向对通道进行编码, 以降低网络深度, 减轻了模型的计算量, 再增加一个 3×3 卷积与其并行作为 3×3 分支, 以聚合遥感图像多尺度目标信息, 增强网络对遥感图像多尺度目标特征的提取能力。

关的被赋予了更高的权重, 其余背景部分则降低了权重。该模块突出了遥感图像中的重要区域, 抑制了噪声和背景信息, 有效解决了遥感图像背景复杂的问题。

PSAM 模块的流程步骤如下:

- 1) 输入特征: $X (C \times H \times W)$;
- 2) X 分别沿高度 h 和宽度 w 两个空间方向进行平均池化 AdaptiveAvgPool2d(), 得到 x_h, x_w ;
- 3) x_h 与 x_w 经过 torch.cat() 进行特征拼接, 得到 X_1 ;
- 4) X 经过卷积 $\text{Conv}(c_1, c_1, k=3, s=1, p=1)$ 得到 X_2 ;
- 5) 将 $y = X + X_1 + X_2$ 分别经过最大池化 maxpool(y) 和平均池化 meanpool(y), 得到 y_1 和 y_2 ;
- 6) 将 y_1 与 y_2 经过 torch.cat() 进行特征拼接, 得到 y_3 ;
- 7) 将 y_3 输入到卷积 $\text{Conv}(c, c, s=7)$ 中, 并经过 sigmoid 激活函数得到 y_4 ;
- 8) 将 y 与 y_4 经过 torch.cat() 进行特征拼接, 得到输出特征 output。

1.4 Focal-EIoU

目前目标检测任务中比较新颖的损失函数如

CIoU^[21]、DIoU^[22],在常规检测任务中表现优异,但是由于通常只考虑检测框和真实框的问题,对于复杂场景下遥感图像的检测任务往往难以适应,因此,考虑到遥感图像存在正负样本不平衡的问题,引入 Focal-EIoU 损失函数^[23],调整正负样本权重,减小低质量样本对梯度的负面影响。

Focal-EIoU 损失函数是结合了 EIoU Loss 和 Focal Loss 两个损失函数的优点。此外,CIoU 作为基础损失定义如式(2)所示,其中 b, b^{gt} 表示预测框和实际框的中心点位置, ρ 表示两个中心点之间的欧氏距离,以最小的闭包区域同时包含预测框和真实框,则 c 表示这个闭包区域的对角线距离, v 是关于 w 和 h 的梯度定义如式(3)所示,示意如图 5 所示。该方式是通过矩形框的距离和相对比例作为惩罚项,导致当 w 和 h 满足 $\{(w = kw^{gt}, h = kh^{gt}) \mid k \in R^+\}$ 时,CIoU 的相对比例的惩罚项将不起作用。

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (2)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

$$L_{EIou} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (4)$$

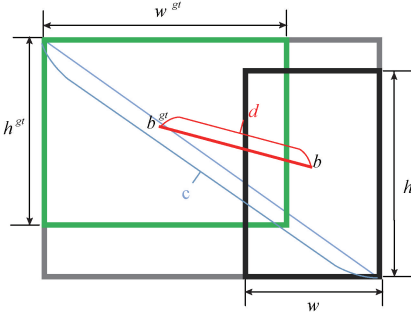


图 5 CIoU 损失函数示意图

Fig. 5 Schematic of the loss function

EIoU 直接用宽和高的预测结果作为惩罚项,由 IoU 损失 L_{IoU} 、距离损失 L_{dis} 和边长损失 L_{asp} 3 部分组成,如式(4)所示,其中 C_w 和 C_h 分别为闭包矩形的宽和高。相对于 CIoU, EIoU 具有更快的收敛速度,但还存在低质量样本会造成损失值剧烈震荡问题。

为解决样本类别不平衡问题,引入 FocalL1 Loss,其梯度表达式如(5)所示, α 用于将梯度值归一化到 $[0, 1]$, β 用于抑制负样本的梯度值,且 β 越大,正样本回归误差越小。

$$g(x) = \frac{\partial_f}{\partial x} = \begin{cases} -\alpha x \ln(\beta x), & 0 < x \leq 1; 1/e \leq \beta \leq 1 \\ -\alpha \ln(\beta), & x > 1; 1/e \leq \beta \leq 1 \end{cases} \quad (5)$$

通过对式(5)求积分得到 FocalL1 Loss 的表达式如式(6)所示,其中 C 为一个常数,取 $x=1$ 代入可得 $C = (2\alpha \ln \beta + \alpha)/4$ 。

$$L_f(x) = \begin{cases} -\frac{\alpha x^2 (2 \ln(\beta x) - 1)}{4}, & 0 < x \leq 1; 1/e \leq \beta \leq 1 \\ -\alpha \ln(\beta) x + C, & x > 1; 1/e \leq \beta \leq 1 \end{cases} \quad (6)$$

通过整合 EIoU 和 FocalL1 loss,最终得到 Focal-EIoU 损失函数如式(7)所示,其中 γ 用于调节正负样本的权重。本文将正负样本权重 γ 设置为 0.5,提高模型的鲁棒性和泛化能力。

$$L_{Focal-EIoU} = IoU^\gamma L_{EIou} \quad (7)$$

2 实验结果与分析

2.1 数据集及评价指标

为了检验模型性能,采用 DIOR-5s 遥感图像数据集进行验证实验,采用 NWPU VHR-10^[24] 公共数据集进行泛化性实验。DIOR-5s 是在 DIOR^[25] 公共数据集上挑选出 5 类复杂场景下的小目标图像共 1 129 张,包含网球场、船舶、油桶、车辆、飞机等 5 个类别,图像分辨率大小为 800×800 ,共标注 11 326 个检测对象,并对目标尺度分布情况进行统计,其中包含 8 772 个像素数量 α 为 $0 < \alpha \leq 32^2$ 的小目标样本,879 个 $32^2 < \alpha \leq 96^2$ 的中目标样本,以及 1 675 个 $\alpha > 96^2$ 的大目标样本。图像中背景复杂多样,且存在阴影、畸变、遮挡等干扰。实验将数据集按照 8 : 2 分为训练集和验证集。

NWPU VHR-10 数据集包括高分辨率遥感图像的 10 种构造特征,该数据集的重点主要服务于遥感图像目标检测任务。图像中的目标种类包括油罐、棒球场、网球场、篮球场、田径场、港湾、桥梁和水坝、港口、桥梁和车辆。数据集包括 85 个空间锐化的空间分辨率为 0.08 m 的彩色红外(CIR)图像、空间分辨率为 0.08 m 的 85 幅空间锐化彩色红外(CIR)图像、715 幅空间分辨率为 0.5 ~ 2.0 m 的图像,以及总共 800 幅高分辨率光学图像,其中 650 幅为目标图像,150 幅为光学图像。

实验主要采用精度率 P (Precision)、召回率 R (Recall)、平均精确率 mAP (mean average precision),检测速度 (frame per second, FPS) 作为评价指标来检测模型性能,公式如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\%$$

(9)

$$Ap = \int_0^1 P(r) dr$$

(10)

$$mAP = \frac{\sum_{i=1}^c Api}{C}$$

(11)

式中, TP (true positive) 表示真的正样本, FP (false positive) 表示假的正样本, FN (false negative) 表示假的负样本, AP 表示单个类别的平均精度, mAP 表示对所有类别的平均精度再取平均值。

2.2 实验环境及参数设置

实验环境为: Windows11 操作系统, CPU: R7-7735H, 主频 3.2 GHz, RAM: 16 GB, 显卡: GeForce RTX4050, DL 框架: Pytorch1. 10. 2。

整个训练过程借鉴了迁移学习的思想, 首先采用预训练模型进行冻结训练 50 轮, 然后再解冻训练 250 轮, 这种训练方式可以加快模型收敛, 减少训练迭代次数, 模型训练超参数设置: 冻结学习率为 0. 01、解冻学习率 0. 001、动量为 0. 937、置信度阈值 0. 5。为保证实验结果的可比性, 本文的所有实验均采用相同的实验环境, 相同的参数设置下进行。

2.3 模型对比实验

为了验证本文 CDA-YOLO 模型在遥感图像复杂场景目标检测中的优越性, 采用 DIOR-5s 遥感图像数据集进行测评, 与当前主流的基于单阶段深度学习目标检测算法进行对比实验, 实验结果如表 1 所示。SSD 采用 vgg16 网络为主干网络, 浅层特征的提取表征能力不够, 平均精度和检测速度较低; YOLOv3 引入残差网络, 采用 Darkent53 作为主干网络, mAP 有所提高, 但由于网络深度加深, 模型太大, 检测速度较低; YOLOv5 采用更小的模型结构, 对检测头进行了解耦操作, 提高了模型检测速度, 检测速度达到 76. 9 fps; YOLOv7 引入卷积重参化, 采用引导标签分配策略, 提高了检测精度, mAP 达到了 85. 3%, 但模型较大, 检测速度也下降至 26. 2 fps; YOLOv7-tiny 在 YOLOv7 的基础上减少了两层特征提取卷积, 缩小了模型结构, 模型精度有所下降, 但模型大小仅为 11. 7 MB; 与上述算法模型相比, 本文 CDA-YOLO 的 P 、 R 、 mAP 分别为 90. 1%、79. 5%、85. 4%, 都达到最高值, 模型大小和检测速度也都在可接受范围之内, 验证了改进算法在遥感图像复杂场景目标检测中具有较好的优越性。

使用不同算法训练的权重, 在验证集上进行逐张的可视化验证实验, 结果如表 2 所示, 其中误检率: $FPR = FP / (FP + TN) \times 100\%$, 漏检率: $FNR = FN / (FN + TP) \times 100\%$, 正检率: $TPR = TP / (TP + FN) \times 100\%$ 。

表 1 模型对比实验结果
Table 1 Results of model comparison experiments

模型	P /%	R /%	mAP @ 0. 5/%	Size /MB	检测速度/ fps
SSD	72. 1	55. 8	60. 3	92. 6	39. 4
YOLOv3	81. 7	71. 3	75. 2	142. 6	21. 7
YOLOv5	88. 2	88. 2	79. 6	40. 3	76. 9
YOLOv7	89. 9	78. 9	85. 3	71. 3	26. 2
YOLOv7-tiny	86. 8	78. 1	82. 8	11. 7	71. 4
CDA-YOLO	90. 1	79. 5	85. 4	19. 9	64. 1

表 2 模型正检、误检和漏检对比实验

Table 2 Comparative experiments of model positive, false and missed detections

模型	TPR /%	FPR /%	FNR /%
SSD	65. 8	24. 3	35. 1
YOLOv3	69. 2	30. 1	22. 5
YOLOv5	74. 3	18. 5	25. 3
YOLOv7	77. 1	20. 1	21. 1
YOLOv7-tiny	73. 3	18. 8	26. 2
CDA-YOLO	85. 9	21. 4	12. 9

由表 2 可知, 本文的 CDA-YOLO 在正检率 (TPR) 达到 85. 9%, 且漏检率 (FNR) 仅为 12. 9%。YOLOv5 的误检率 (FPR) 为 18. 5%, 但其正检和漏检的表现均不如 CDA-YOLO。因此, 本文的 CDA-YOLO 在遥感图像中的小目标检测任务上具有更优越的表现。

图 6 为不同算法在同一张图片上的检测效果, 其中 TP 表示正检, FP 表示误检, FN 表示漏检。可以看出本文 CDA-YOLO 模型具有更好的检测效果, 特别是在小目标密集和多尺度目标的检测上表现更好。

对所有类别在不同场景各挑选一张图片进行逐张检测, 检测结果如表 3 所示。改进后的网络对网球场、船舶、车辆和飞机的检测精度 AP 分别提升了 3. 3%、7. 1%、6. 8% 和 1. 9%, 对油桶 AP 值仅下降了 0. 3%。此外, 改进后的 CDA-YOLO 有效降低了遥感图像目标检测的漏检率和误检率。图 7 展示了网络模型改进前和改进后在不同场景下的检测效果。

表 3 各类型目标检测结果

Table 3 Detection results for each type of target

模型	类别	TP	FP	FN	AP /%
YOLOv7-tiny	网球场	12	1	2	62. 5
	船舶	93	12	18	80. 3
	油桶	81	0	1	92. 5
	车辆	29	1	1	80. 6
	飞机	31	0	0	92. 6
CDA-YOLO	网球场	13	0	1	65. 8
	船舶	103	2	3	87. 4
	油桶	83	1	0	92. 8
	车辆	30	0	0	87. 4
	飞机	31	0	0	94. 5

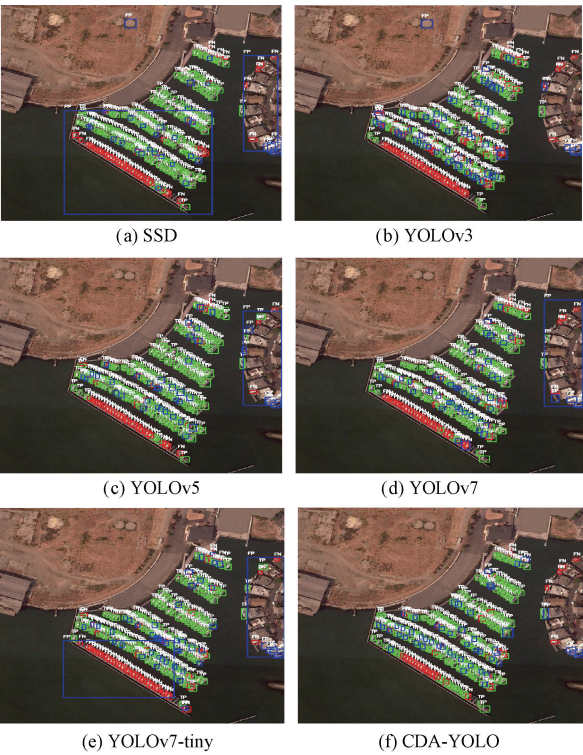


图 6 不同模型检测结果

Fig. 6 Detection results of different models

2.4 消融实验

为了定量分析改进算法中低跨度上下文解耦模块,并行级联注意力模块,损失函数改进对遥感图像小目标检测的影响,设计消融实验逐一进行评估,其中“√”表示添加对应模块,实验结果如表 4 所示。

表 4 消融实验结果

YOLOv7- tiny		LPCD- Head	Focal- PSAM EIoU Loss	P/%	R/%	mAP@ 0.5/%	检测速度/ fps
√				86.6	78.1	82.8	71.4
√	√			86.8	82.1	84.2	66.7
√		√		87.1	80.6	83.9	69.5
√			√	88.2	78.9	83.7	69.9
√	√	√		84.8	81.1	84.7	61.1
√	√		√	88.9	78.4	84.4	67.6
√	√	√	√	90.1	79.5	85.4	64.1

由表 4 可知,仅使用 LPCD-Head 的 YOLOv7-tiny 的 mAP 从 82.8% 提高到 84.2%,整体提高了 1.4%。召回率提高了 4%,而精度基本保持不变。新检测头的特征融合机制导致检测速度从 71.4 降至 66.7 fps。仅使用 PSAM 融合时, mAP 增加了 1.1%,但速度降低了 1.9 fps。只改进损失函数时, mAP 从 82.8% 增加到 83.7%,而速



图 7 YOLOv7-tiny 与 CDA-YOLO 方法在多尺度区域(上)、复杂背景区域(中)和小目标密集区域(下)检测结果

Fig. 7 Detection results of YOLOv7-tiny and CDA-YOLO under multi-scale regions(top), complex background regions(middle) and small target-intensive regions(bottom)

度却降低了 1.5 fps。在改进检测头的基础上再融合 PSAM,模型平均精度由 82.8% 提升至 84.7%,整体提升 1.9%,检测速度下降至 61.1 fps;同时改进检测头和损失函数后, mAP 从 82.8% 增加到 84.4%,速度则下降到 67.6 fps。所有改进后, mAP 从 82.8% 增加到 85.4%,速度则为 64.1 fps。尽管检测速度略有下降,但在保证实时性足够的基础上,检测精度提升了 2.6%。

1) LPCD-Head 消融实验

为了验证 LPCD-Head 的有效性,以 YOLOv7-tiny + PSAM+Loss 为基础模型,对不同特征层进行消融实验,实验结果如表 5 所示。

表 5 LPCD-Head 消融实验结果

$G_{loc,l}$			$G_{cls,l}$			P/%	R/%	mAP/%
P_{l-1}	P_l	P_{l+1}	P_{l-1}	P_l	P_{l+1}			
	√			√		88.2	76.9	82.3
	√			√	√	86.8	79.1	83.4
√	√			√		89.3	78.9	84.8
√	√			√	√	90.1	79.5	85.4
√	√	√	√	√	√	84.9	80.8	83.9

当分类任务融入相邻深层特征时,平均精度 mAP 提升了 1.1%,证明了语义信息对分类任务的重要性;当回

归任务融入相邻浅层特征时, mAP 提升了 2.5%, 证明了细节信息对回归任务的重要性, 同时也体现出回归任务对遥感图像检测的影响因素更大; 将以上两者相结合, mAP 由 82.3% 提升至 85.4%, 整体提升了 3.1%; 最后, 对分类和回归同时融合 3 层特征, 发现效果明显下降, 再次证明了不同任务对特征层的需求是不同的, 验证了本文设计的 LPCD-Head 的有效性。

2) PSAM 实验

为了进一步检验 PSA 模块的优势, 以改进检测头和损失函数为基础模型 (DL-YOLOv7-tiny), 分别引入经典的 CBAM 注意力机制模块和 CA 注意力机制模块进行对比实验, 实验结果如表 6 所示。

表 6 融入不同注意力机制对比实验结果

Table 6 Experimental results comparing the different attention mechanisms

DL-YOLOv7-tiny	CBAM	CA	PSAM	P/%	R/%	mAP/%	参数量
✓				88.9	78.4	84.4	10 792 609
✓	✓			87.9	79.5	84.8	10 800 900
✓		✓		87.3	80.6	84.7	10 795 233
✓			✓	90.1	79.5	85.4	10 795 332

CBAM 通过自适应地学习通道和空间的重要性权重, 增强有用的特征并抑制无关特征, 但对于遥感图像中多尺度目标检测效果不理想, mAP 仅提高了 0.4%, 且参数量增加较多; CA 模块将精准的位置信息嵌入到通道中, 并捕获了长程依赖关系, 回归率 R 提高了 2.2%, 但缺乏在空间维度上的信息, mAP 仅提高 0.3%; PSA 模块在 CA 模块的基础上增加一层 3×3 卷积支路构成并行三支路网络, 并引入空间注意力模块, 在增强网络对多尺度目标特征的提取能力的前提下, 还实现了自适应地学习通道和空间的重要性权重, mAP 由 84.4% 提升至 85.4%, 整体提升了 1%; 验证了 PSA 模块在遥感图像检测中的优势。

3) 损失函数实验

为了进一步展现 Focal-EIoU 在遥感图像上的检测效果, 以 YOLOv7-tiny+LPCD-Head+PSAM 为基础模型, 引入不同的损失函数做对比实验, 实验结果如表 7 所示, 原模型的损失函数为 CIoU, 由表 7 可知, DIoU 通过惩罚项直接最小化中心点之间的距离, 加快了收敛速度, 但未考虑 box 回归的纵横比, 存在边长被错误放大问题, mAP 仅提升了 0.2%; EIoU 直接将边长作为惩罚项, 进一步的加快了收敛速度, 但由于遥感图像中样本难易程度极不平衡, mAP 反而下降了 0.8%; Focal-EIoU 在 EIoU 的基础上引入 Focal Loss, 解决了遥感图像中正负样本不平衡问题, 在加快了收敛速度的同时, 平均精度 mAP 提升了 0.7%, 验证了 Focal-EIoU 在遥感图像上检测效果。

表 7 引入不同损失函数对比实验结果

Table 7 Experimental results comparing different loss functions

Loss	P/%	R/%	mAP/%
CIoU	84.8	81.1	84.7
DIoU	87.6	82.2	84.9
EIoU	88.2	79.7	83.9
Focal-EIoU	90.1	79.5	85.4

2.5 模型部署对比分析

模型体积和检测速度是衡量模型轻量化的重要指标, 较小的模型体积和较快数据响应速度可以更容易在资源有限的设备上部署和应用模型。因此, 为验证本文模型的轻量高效性, 选取表 1 中的数据将 CDA-YOLO 与其他主流模型的体积和响应速度进行对比。实验结果分别如图 8 和 9 所示, 其中圆形图标越大, 代表模型体积越大。由图 8 可知, 相比于大模型 YOLOv3, YOLOv5、SSD、CDA-YOLO 无论在体积还是检测精度上都更具优势。CDA-YOLO 相比于 YOLOv7-tiny 在体积相近的情况下, mAP 由 82.8% 上升至 85.4%, 整体提高了 2.6%。此外, YOLOv7 检测精度与 CDA-YOLO 持平, 但其模型体积太大, 难以在资源受限的设备上部署。由图 9 可知, CDA-YOLO 的 mAP 和 FPS 分别达到 85.4% 和 64.1 fps, 数据均优于 YOLOv3 和 SSD。YOLOv7 的检测精度取得较好表现, 但其响应速度太低, 实时性较弱。YOLOv7-tiny 和 YOLOv5 的 FPS 值最高, 但其 mAP 值分别仅为 82.8% 和 79.6%。综合考虑模型的轻量化和检测精度, CDA-YOLO 是一种高效轻量的遥感图像目标检测模型。

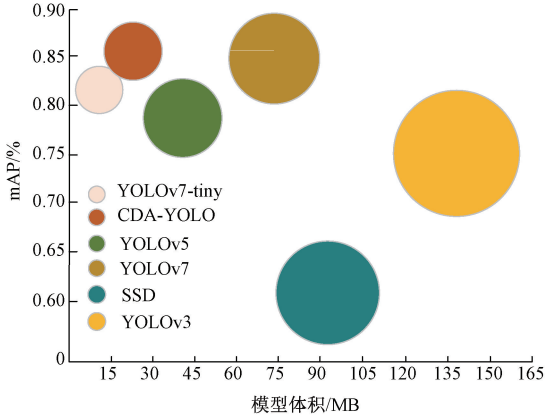


图 8 各模型体积和检测精度对比

Fig. 8 Comparisons of volume and detection accuracy of various models

2.6 模型泛化能力实验

为了进一步验证本文模型的先进性和泛化能力, 选择在 NWPU VHR-10 公共数据集上进行对比实验。所有实验结果如表 8 所示, 由表中实验数据可知, YOLOv5 在

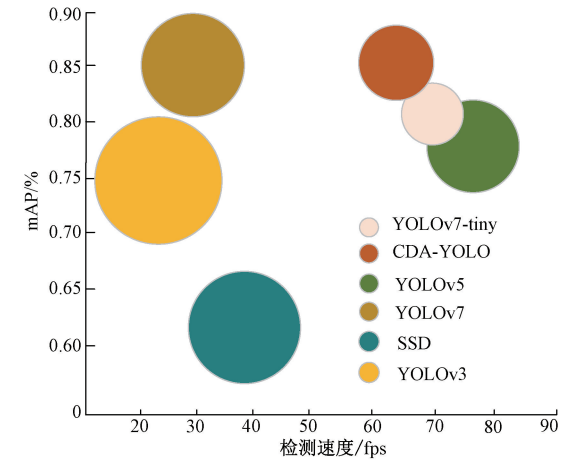


图 9 各模型实时性和检测精度对比

Fig. 9 Comparisons of real-time and detection accuracy of various models

检测速度上最快,检测速度达到 76.9 fps,但模型不够轻量,检测精度只有 88.6%。YOLOv7-tiny 模型大小仅为 11.7 MB,但其检测精度也只有 88.9%。相比之下,本文提出的 CDA-YOLO 模型大小仅为 19.9 MB,其检测速度和检测精度分别达到 64.1 fps 和 90.6%,在保证足够轻量且快速的基础上其精确度达到最高。图 10 为改进前后算法在 NWPU VHR-10 数据集中复杂场景图像的检测效果。由图可知,改进后的 CDA-YOLO 算法有效的降低了目标在复杂场景下的误检和漏检问题,同时也验证了 CDA-YOLO 具有较好的泛化能力。

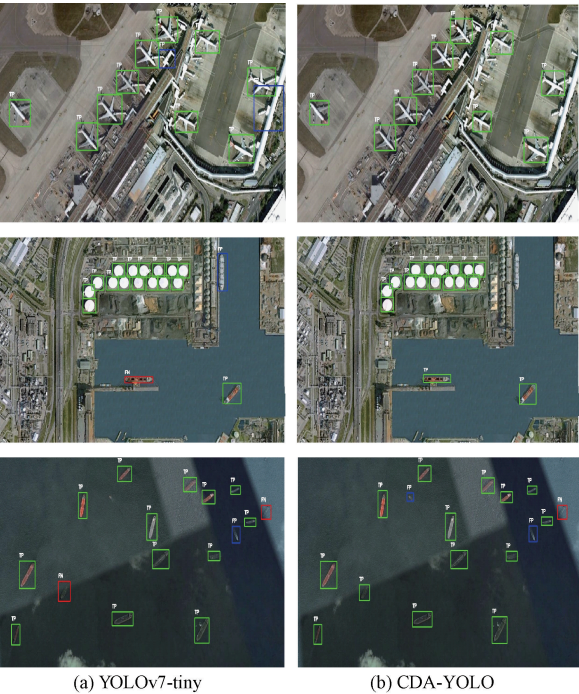


图 10 NWPU VHR-10 遥感数据集对比

Fig. 10 Comparison diagram of NWPU VHR-10 remote sensing dataset

表 8 NWPU VHR-10 遥感数据集对比结果
Table 8 Comparison results of various models on the NWPU VHR-10 dataset

模型	mAP@ 0.5/%	模型体积/MB	检测速度/fps
SSD	68.3	92.6	39.4
YOLOv3	83.2	142.6	21.7
YOLOv5	88.6	40.3	76.9
YOLOv7	89.3	71.3	26.2
YOLOv7-tiny	88.9	11.7	71.4
CDA-YOLO	90.6	19.9	64.1

3 结 论

本文针对目前目标检测算法在受资源限制的小型设备中检测精度不足,且在光学遥感复杂场景条件下目标易误检和漏检的问题,提出了 CDA-YOLO 算法,并介绍了网络的整体结构和工作原理;在特征提取阶段,通过构建三支路并行网络和空间注意力网络结合,设计了一种并行级联注意力模块,有效解决了遥感图像中目标多尺度问题;在检测阶段,通过对分类任务融合深层特征,回归任务融合浅层特征的方法,设计了一种低跨度上下文解耦的检测头,降低了遥感图像中密集小目标检测的漏检率;最后引入 Focal ElIoU 损失函数,提高了模型的泛化能力。实验结果表明:在 DIOR-5s 数据集下的检测平均精度由 82.8% 提升至 85.4%,检测速度为 64.1 fps,模型大小仅为 19.9 MB,在保证足够轻量的前提下,一定程度上降低了模型的漏检率,并在 NWPU VHR-10 数据集上检测精度达到 90.6%,验证了模型的泛化能力。在未来的工作中,将基于 CDA-YOLO 算法在物理平台上完成算法的植入,对算法的落地部署进行研究。

参考文献

[1] 陈燕,杨艳,杨春兰,等. 基于阶段聚焦损失和并行增广策略的遥感图像场景分类[J]. 电子测量与仪器学报, 2023, 37(1): 116-122.
CHEN Y, YANG Y, YANG CH L, et al. Remote sensing image scene classification via stage-based focal loss and parallel data augmentation strategy[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 116-122.
[2] 闫钧华, 张琨, 施天俊, 等. 融合多层级特征的遥感图像地面弱小目标检测[J]. 仪器仪表学报, 2022, 4 (3): 221-229.
YAN J H, ZHANG K, SHI T J, et al. Multi-level feature fusion based dim small ground target detection in remote sensing images[J]. Chinese Journal of Scientific Instrument, 2022, 4 (3): 221-229.

- [3] LORAN T, DA SILVA A B C, JOSHI S K, et al. Ship detection based on Faster R-CNN Using range-compressed airborne radar data [J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 3500205.
- [4] ULLO S L, MOHAN A, SEBASTIANELLI A, et al. A new mask R-CNN-based method for improved landslide detection [J], IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14:3799-3810.
- [5] QIU M, HUANG L, TANG B H, et al. ASFF-YOLOv5: Multielement detection method for road traffic in UAV images based on multiscale feature fusion[J]. Remote Sensing, 2022, 14(14): 3498.
- [6] 钱承山, 沈有为, 孙宁, 等. 改进 YOLOv5s 的遥感图像检测研究[J]. 国外电子测量技术, 2022, 41(11): 57-66.
QIAN CH SH, SHEN Y W, SUN N, et al. Improved YOLOv5s remote sensing image detection research [J]. Foreign Electronic Measurement Technology, 2022, 41(11): 57-66.
- [7] LIN J, ZHAO Y, WANG S, et al. YOLO-DA: An efficient YOLO-based detector for remote sensing object detection [J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 6008705.
- [8] NIE Y, BIAN C, LI L, et al. LFC-SSD: Multiscale aircraft detection based on local feature correlation[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6510505.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42 (2): 318-327.
- [10] LIU W, MA L, WANG J, et al. Detection of multiclass objects in optical remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(5): 791-795.
- [11] 陈欣, 万敏杰, 马超, 等. 采用多尺度特征融合 SSD 的遥感图像小目标检测[J]. 光学 精密工程, 2021, 29(11): 2672-2682.
CHEN X, WAN M J, MA CH, et al. Recognition of small targets in remote sensing image using multi-scale feature fusion-based shot multi-box detector[J]. Optics and Precision Engineering, 2021, 29(11): 2672-2682.
- [12] 左露, 牛晓伟, 朱春惠. 基于改进 YOLOX 的航拍遥感图像检测模型[J]. 电子测量技术, 2023, 46(16): 179-186.
ZUO L, NIU X W, ZHU CH H. Aerial remote sensing image detection model based on improved YOLOX[J]. Electronic Measurement Technology, 2023, 46 (16): 179-186.
- [13] LIU M, LIN K, HUO W J, et al. Feature enhancement modules applied to a feature pyramid network for object detection[J]. Pattern Analysis & Applications, 2023, 26(2): 1-13.
- [14] LUO S, YU J, XI Y J, et al. Aircraft target detection in remote sensing images based on improved YOLOv5[J]. IEEE Access, 2022, 10: 5184-5192.
- [15] HAN W, KUERBAN A, YANG Y, et al. Multi-vision network for accurate and real-time small object detection in optical remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6001205.
- [16] 胡欣, 马丽军. 基于 YOLOv5 的多分支注意力 SAR 图像舰船检测 [J]. 电子测量与仪器学报, 2022, 36(8): 141-149.
HU X, MA L J. Multi-branch attention SAR image ship detection based on YOLOv5 [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (8): 141-149.
- [17] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J]. ArXiv preprint arXiv: 2107, 08430, 2021.
- [18] ZHUANG J Y, QIN Z, YU H, et al. Task-specific context decoupling for object detection [J]. ArXiv preprint arXiv: 2303, 01047, 2023.
- [19] ZHU X Z, CHENG D Z, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 6688-6697.
- [20] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [21] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
- [22] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [23] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IoU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [24] ZHU X, CHENG D, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks [C]. Proceedings of the IEEE/CVF International Conference

on Computer Vision, 2019; 6688-6697.

[25] RORIZ R, CAMPOS A, PINTO S, et al. Dior: A hardware-assisted weather denoising solution for Lidar point clouds[J]. IEEE Sensors Journal, 2021, 22(2): 1621-1628.

作者简介



邝先验, 2014 年于华南理工大学获得博士学位, 现为江西理工大学教授, 主要研究方向为智能感知与控制、深度学习与图像处理。

E-mail: xianyankuang@163.com

Kuang Xianyan received his Ph. D. degree from South China University of Technology in 2014. Now he is a professor in Jiangxi University of Science and Technology. His main research interests include intelligent sensing and control, deep learning and image processing.



程福军, 江西理工大学硕士研究生, 主要研究方向为深度学习与图像处理。

E-mail: 2500432750@qq.com

Cheng Fujun is a M. Sc. candidate at Jiangxi University of Science and Technology.

His main research interests include deep

learning and image processing.



吴翠琴 (通信作者), 2006 年于江西理工大学获得硕士学位, 现为江西理工大学讲师, 主要研究方向为智能感知与机器人、深度学习与图像处理。

E-mail: cuiqinwu@163.com

Wu Cuiqin received her M. Sc. degree from Jiangxi University of Science and Technology in 2006. Her main research interests include intelligent perception and robotics, deep learning and image processing.