

## 基于 YOLOX-Tiny 的轻量级遥感图像目标检测模型

郎磊<sup>1</sup>, 刘宽<sup>2</sup>, 王东<sup>1\*</sup><sup>1</sup>北京交通大学计算机与信息技术学院, 北京 100044;<sup>2</sup>郑州轻工业大学计算机与通信工程学院, 河南 郑州 450001

**摘要** 为了解决高分辨率遥感图像目标检测存在的背景复杂多样、密集物体分布及目标尺度差异大等问题,同时考虑到应用场景对模型资源的限制,提出了一种基于 YOLOX-Tiny 的轻量级遥感目标检测网络。首先采用多尺度预测方法增强对密集目标的检测能力;其次引入协同注意力模块提高重要特征关注度,抑制背景噪声;再者通过可变形卷积替换关键预测卷积层,强化空间建模能力;最后优化损失函数,改善遥感目标定位精度。在公开的遥感图像目标检测数据集 DIOR 上评估了所提算法的有效性,实验结果表明:相比基准算法(YOLOX-Tiny),所提算法在平均精度(AP)和 AP50 指标上分别提升 4.1 个百分点和 4.42 个百分点;在保持较高精度的前提下,每秒检测帧数(FPS)达到 46,可满足实时检测的需求,优于其他先进算法。

**关键词** 遥感; 目标检测; YOLOX-Tiny; 遥感图像; 协同注意力模块; 可变形卷积; CIOU

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212699

## Lightweight Remote Sensing Object Detector based on YOLOX-Tiny

Lang Lei<sup>1</sup>, Liu Kuan<sup>2</sup>, Wang Dong<sup>1\*</sup><sup>1</sup>School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;<sup>2</sup>Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, Henan, China

**Abstract** To solve problems in the complex geometry scene, dense object distribution, and the large range of object size variations in high-resolution remote sensing object detection and to address the limitations of model resources in application scenarios, a lightweight remote sensing object detector based on YOLOX-Tiny is proposed. A multi-scale prediction method is used to enhance the detection capability of dense objects. Moreover, a coordinate attention module is introduced to improve the attention of important characteristics while suppressing background noise. The key prediction convolution layer is replaced by deformable convolution to strengthen the spatial modeling capability. Finally, the loss function is optimized to increase the localization accuracy of remote sensing objects. The effectiveness of the proposed algorithm is evaluated on the public remote sensing image target detection dataset DIOR. The experimental results show that compared with the benchmark algorithm (YOLOX-Tiny), the proposed algorithm improves the average precision (AP) and AP50 indexes by 4.1 percentage points and 4.42 percentage points respectively; on the premise of maintaining high accuracy, the number of detection frames per second (FPS) reaches 46, which can meet the needs of real-time detection and is superior to other advanced algorithms.

**Key words** remote sensing; object detection; YOLOX-Tiny; remote sensing image; coordinate attention module; deformable convolution; CIOU

## 1 引言

近年来,随着航天遥感技术的快速发展,高分辨率光学遥感图像的获取变得越来越便利,针对光学遥感

图像的分析研究逐渐成为了关注的焦点<sup>[1]</sup>。图像的分类、检测、分割和跟踪等任务成为遥感图像处理方向的热门话题<sup>[2]</sup>,其中目标检测任务在现实生活中呈现出广阔的应用前景,受到研究人员的追捧<sup>[3]</sup>。

收稿日期: 2021-10-11; 修回日期: 2021-11-11; 录用日期: 2021-11-29; 网络首发日期: 2021-12-08

基金项目: 国家重点研发计划项目(2019YFB2204200)、北京市自然科学基金(4202063)、中央高校基本科研项目(2020JBM020)

通信作者: \*wangdong@bjtu.edu.cn

在目标检测任务中,基于深度神经网络的方案展现出优于传统方法的性能<sup>[4]</sup>。目前,典型目标检测模型可分为两类:1)以 SSD<sup>[5]</sup>和 YOLO<sup>[6]</sup>等算法为代表的一阶段检测算法,该类算法直接对输入图像进行回归,预测目标类别和位置坐标,虽然检测速度快但精度低;2)以 Faster R-CNN<sup>[7]</sup>和 Mask R-CNN<sup>[8]</sup>等算法为代表的两阶段检测算法,该类算法首先利用区域建议网络找到可能包含目标的区域,然后使用第 2 个网络对候选区域进行分类和位置的预测,其检测精度高但速度慢。如今,基于深度神经网络的目标检测算法已成功应用于各个领域,其中不乏对遥感图像目标检测任务的尝试。然而,遥感图像与自然图像之间存在较大差异,背景复杂多样、密集的对象分布及目标尺度差异大等特点使得遥感图像的目标检测任务成为一个难题。Zhao 等<sup>[9]</sup>提出了一种基于 SSD 的多尺度遥感目标检测方法,该方法通过通道注意力模块加强对象间的长期语义依赖关系。戴媛等<sup>[10]</sup>提出了一种基于 Faster R-CNN 改进的遥感目标检测方法,且该方法取得了良好的检测结果。吴湘宁等<sup>[11]</sup>提出了一种基于 Mask R-CNN 的小目标检测方法,该方法实现了对遥感图像中船只的稳定检测。

虽然基于深度学习的目标检测方法在遥感图像上取得了较高的检测精度,但是在现实环境中,遥感目标检测任务常用于灾害监测、野外救援、国防安全和城市规划等民用<sup>[12]</sup>和军事领域<sup>[13]</sup>,这对算法的准确性、实时

性和部署性都有着严格的要求。农元君等<sup>[14]</sup>以 YOLOv3-Tiny 作为基础网络,通过改善网络结构来提高识别能力,并在通用遥感图像上实现了实时检测。沈丰毅等<sup>[15]</sup>提出了一种基于改进 YOLOv3 的神经网络舰船检测算法,该算法权衡了识别精度和速度。闫皓炜等<sup>[16]</sup>提出了一种基于 CenterNet 模型的实时目标检测算法,并将其应用于电力基础设施的规划和建设中,实现了对输电杆塔的自动识别。尽管这些算法能够在特定遥感任务中满足实时性需求,但其精确度远低于当前先进算法。

为了平衡算法在遥感目标检测任务上的检测精度和速度,本文提出了一种轻量级遥感目标检测算法。以 YOLOX-Tiny 为基础网络框架,通过修改原始网络结构、优化模型训练损失函数,改善算法对遥感图像中复杂信息的提取能力。在大型遥感目标检测数据集上的结果表明了所提算法的优秀检测性能。

## 2 YOLOX-Tiny 算法介绍

YOLO 系列<sup>[17-18]</sup>有着十分优秀的实时检测性能,其中 YOLOX<sup>[19]</sup>作为当下最新的版本,检测性能达到了新的高度。YOLOX-Tiny 则是其轻量化版本,结构精简、检测速度快。因此,本研究选用 YOLOX-Tiny 作为遥感目标检测任务的基准,所提算法框架主要包括 3 个部分:骨干网络(Backbone)、颈部(Neck)和检测头(Head),如图 1 所示,Conv 表示深度卷积运算,BN

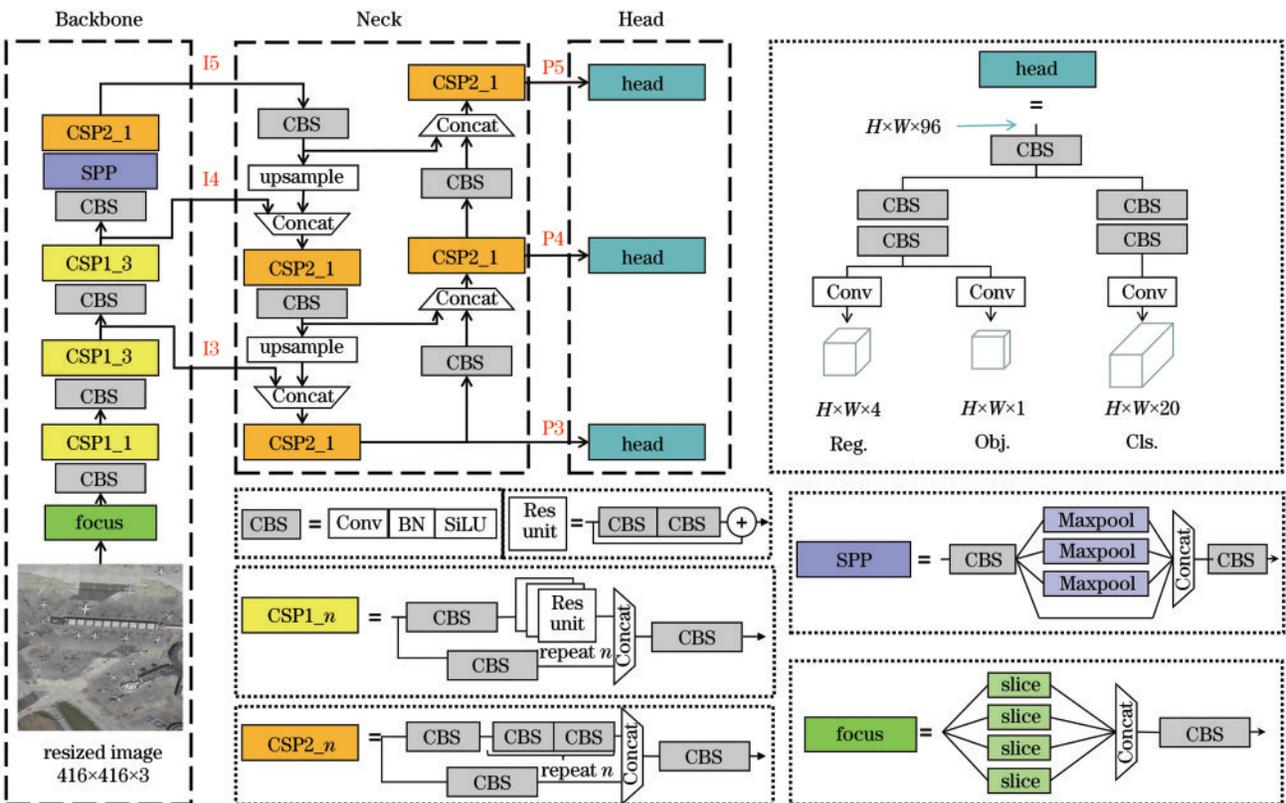


图 1 YOLOX-Tiny 网络结构图  
Fig. 1 YOLOX-Tiny network structure

表示批归一化操作,upsample是采用最近邻算法的上采样操作,Concat是特征图的拼接操作,CBS由Conv、BN和SiLU组成。Backbone采用Modified CSPNet<sup>[20]</sup>和激活函数SiLU来提取输入图像的深度语义信息:其中focus层<sup>[21]</sup>由切片操作和基础卷积运算组成,在保证低参数量和计算量的同时完成下采样操作;空间金字塔池化(SPP)层由池化核大小分别为 $5\times 5$ 、 $9\times 9$ 、 $13\times 13$ 的最大池化操作(Maxpool)和基础卷积组成,用于扩大感受野且不显著增加模型尺寸。Neck位于Backbone和Head之间,采用feature pyramid network(FPN)<sup>[22]</sup>和pyramid attention network(PAN)<sup>[23]</sup>来融合浅层细节和深层语义信息(特征层由浅至深分别为I3、I4、I5),并输出3个尺度的特征图(P3、P4和P5)。Head由1个共享卷积、2个分支的额外卷积和各个任务单独的卷积组成,用于预测目标的定位偏移量(Reg.)、交并比(IoU)感知分类分数(Obj.)和分类分数(Cls.)。

虽然YOLOX-Tiny在自然图像下表现良好,但其在光学遥感图像上仍存在以下问题:

1) YOLOX-Tiny网络深度较浅,对高维特征语义信息提取不完善,极大限制了模型的发挥。此外,模型在单层通道数量上相对较少,不足以保留遥感图像中

丰富的信息量,特别是在经过不断卷积后,重要的细节信息丢失严重,影响目标边界和小物体的判断。

2) YOLOX-Tiny在检测时不能有效地覆盖所有物体,尤其是分布密集且比例差异大的遥感物体。对于输入分辨率为 $416\text{ pixel}\times 416\text{ pixel}$ 的遥感图像,模型用于预测的特征图仅有 $13\times 13$ 、 $26\times 26$ 和 $52\times 52$ 等3个尺度,可以有效识别大小为 $32\text{ pixel}\times 32\text{ pixel}$ 、 $16\text{ pixel}\times 16\text{ pixel}$ 和 $8\text{ pixel}\times 8\text{ pixel}$ 以上的物体,但对于小于 $8\text{ pixel}\times 8\text{ pixel}$ 的物体,检测丢失的可能性较大<sup>[24]</sup>。图2展示了不同尺度下目标位置与预先设定检测框在原图上的映射,实线框体为储罐具体位置,虚线框体为预先设定检测框,其中储罐在缩放后图像中的分辨率为 $5\text{ pixel}\times 5\text{ pixel}$ ,储罐在检测时容易被忽略。

### 3 改进 YOLOX-Tiny 算法

为了解决遥感目标检测任务中存在的背景复杂多样、密集的对象分布及目标尺度差异大等问题,同时考虑到模型轻量化的需求,对YOLOX-Tiny算法进行了如下改进:1)改进网络结构,提升遥感目标的检测能力。首先,采用多尺度预测方法,添加额外预测特征图,降低小目标的漏检概率,提升密集目标检测性能。然后,在

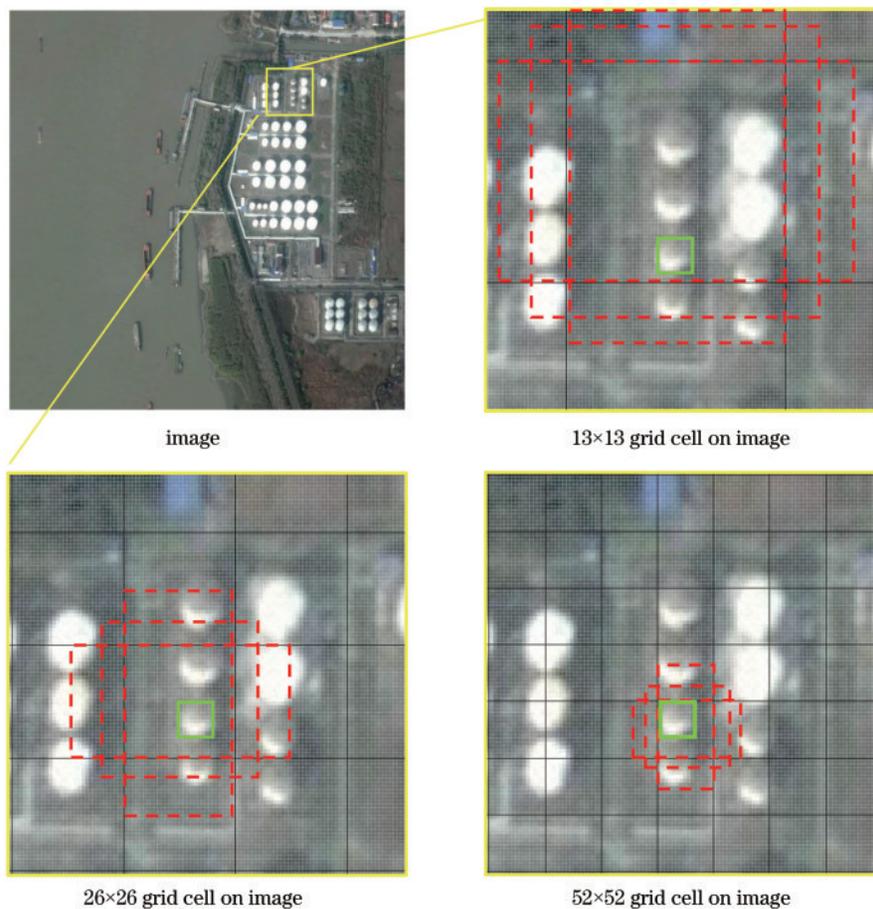


图2 不同尺度下预先设定检测框在原图上的映射

Fig. 2 Mapping of anchor in original image at different scales

几乎不增加模型尺寸情况下,引入协同注意力模块(CA),更加关注目标的有效信息,抑制背景噪声。最后,在Head中使用可变形卷积(DConv)替换深度卷积,强化模型几何变换的建模能力。2)优化网络训练的损失函数,使用CIoU损失函数代替原损失函数中边界框坐标预测损失项,提高遥感目标定位精度和收敛速度。

### 3.1 改进的网络结构

所提网络以YOLOX-Tiny为基础框架,改进后的网络结构如图3所示,亮色部分为主要改进点,分别是多尺度预测、协同注意力模块和可变形卷积。

首先将输入图像缩放至416 pixel×416 pixel,经Backbone后得到包含不同层次信息的特征图I2~I5,其中I2~I5特征图尺寸大小分别为104×104、52×52、26×26和13×13;随后,I5经过CA模块对关键特征进行增强,参与Neck中多个尺度的特征融合,并输出预测特征图P2~P5;最后,预测特征图P2~P5(尺寸分别为104×104、52×52、26×26和13×13)经过1个共享卷积和2个分支额外卷积对定位和分类任务解耦,其中分支额外卷积由可变形卷积和深度卷积单元CBS组成,并通过各任务卷积得到预测目标的位置、类别及IoU感知分类分数。

对于多尺度预测,参考FPN<sup>[22]</sup>的思想,采用直接添加的方式构建额外预测特征图P2,具体实现如下:选取充分融合上下文信息的P3层,通过卷积和上采样操作来扩大特征图尺寸,并与包含大量细节信息的I2层拼接,经过基础构建块CSP2\_1充分混合特征,输出特征图P2。相对于原YOLOX-Tiny的3层预测特征,

额外的特征图能够有效提升物体检测可能性,这种改进对密集目标分布的遥感任务更加有效<sup>[25]</sup>。

协同注意力模块可以发掘现有数据的关联性,突出重要特征,其将位置信息嵌入信道注意力中,这不仅建模了数据通道间的关系,而且捕获了位置信息间的依赖关系<sup>[26]</sup>。此外,该模块即插即用,计算量和参数量较小,几乎可以忽略不计,非常适合在轻量级算法中使用。协同注意力模块结构如图4所示。给定输入特征图 $I \in \mathbb{R}^{C \times H \times W}$ 。首先,通过两个沿着不同空间维度的平均池化层[池化核大小分别为(1, W)和(H, 1)],得到两个聚合了不同维度信息的特征向量,这样不仅捕获了当前空间方向的依赖关系,而且保留了另一个空间方向精确的位置信息;然后,通过维度转换来拼接二者,再通过1×1的卷积,充分利用二者捕获的位置和关联信息,准确地突显感兴趣的区域,得到的特征向量大小为 $\frac{C}{r} \times 1 \times (H + W)$ ,其中参数r为控制通道数缩小的比例,可降低计算量、减少推断时间;紧接着,沿着空间维度再次分割为两个独立的特征向量,并通过1×1卷积和激活函数变换为通道数量相同的特征图;最后,与输入特征图逐元素相乘。

为了学习密集区域遥感目标的形变信息,引入可变形卷积来改善定位和分类分支的空间建模能力。可变形卷积<sup>[27]</sup>在深度卷积基础上使用带偏移的采样代替固定位置采样,并通过可学习参数调整采样位置,达到卷积采样点自适应变化的目的,从而适应尺寸变化范围广阔的遥感目标。对于一个3×3的卷积操作,其卷积

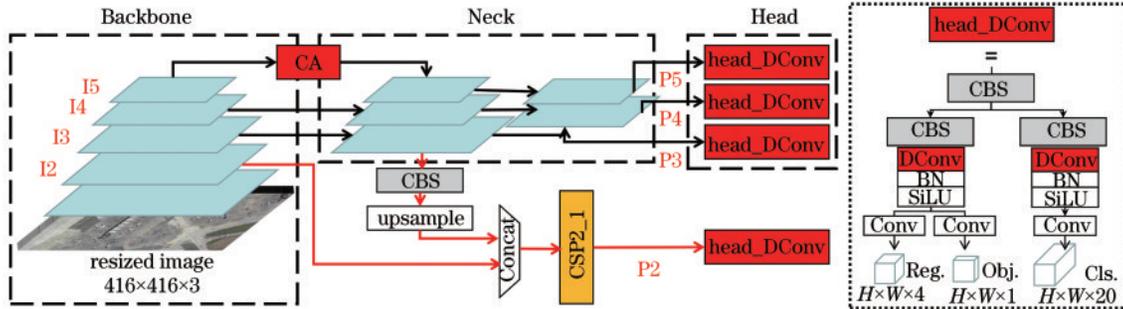


图3 改进YOLOX-Tiny  
Fig. 3 Improved YOLOX-Tiny

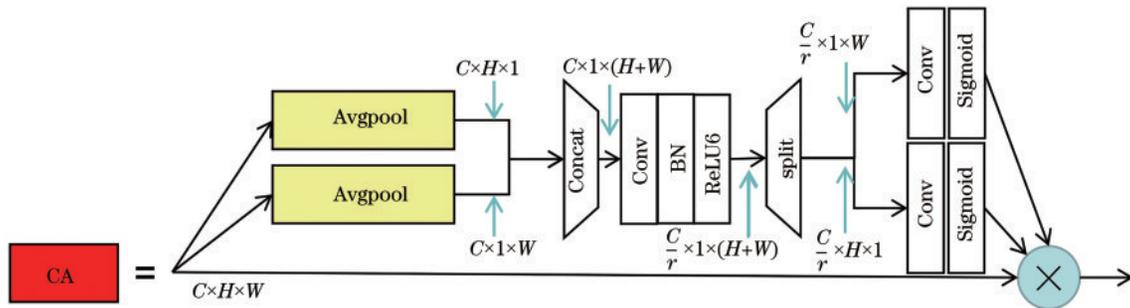


图4 协同注意力模块  
Fig. 4 Coordinate attention module

核采样范围  $R = (-1, -1), (0, -1), \dots, (0, 1), (1, 1)$ , 卷积计算时, 输出特征图中  $P_0$  位置的数值满足:

$$y(p_0) = \sum_{p_i \in R} \omega(p_i) \cdot x(p_0 + p_i + \Delta_{p_i}) \cdot \Delta_{m_i}, \quad (1)$$

式中:  $\omega$  为卷积核权重;  $x$  为输入特征图;  $p_i$  为卷积核采样位置;  $\Delta_{p_i}$  为每个特征位置偏移量;  $\Delta_{m_i}$  为每个采样点

的权重值。可变形卷积结构如图 5 所示, 整体可分为两个部分: 在图上半部分中, 输入特征图  $x$  通过卷积运算学习采样点偏移量  $\Delta_{p_i}$  和权重值  $\Delta_{m_i}$ , 得到特征图 offset map,  $N$  表示深度卷积运算核的尺寸; 在图 5 下半部分, 输入特征层  $x$  经过由 offset map 微调后的深度卷积运算, 完成空间特征自适应采样, 聚集重要的空间特征。

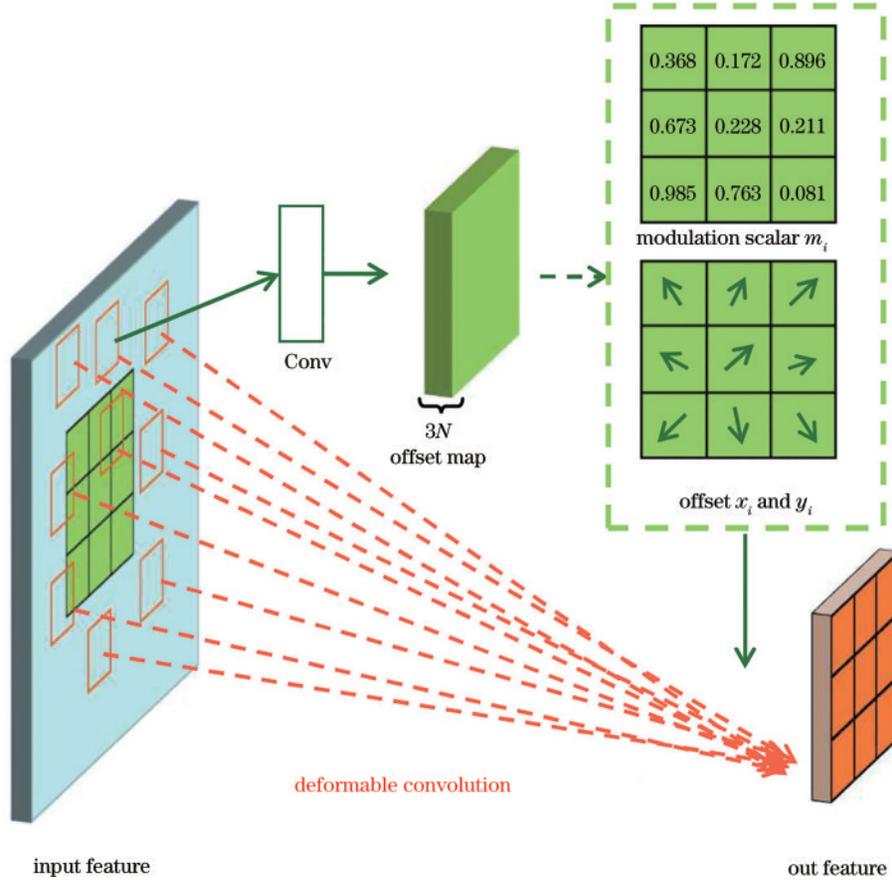


图 5 可变形卷积示意图

Fig. 5 Deformable convolution schematic diagram

### 3.2 优化损失函数

YOLOX-Tiny 算法中, 损失函数主要由 3 部分组成: 边界框回归损失、IoU 感知分类损失和类别损失。在遥感目标检测任务训练期间发现, 目标定位损失收敛速度慢, 尤其是遥感图像中的小目标。由于目标太小, 真实框和预测框之间常处于一种包含关系, 往往是预测框在真实框之中。采用 IoU 损失函数作为边界框回归损失时, 如果出现重叠, 将不能为边界框提供准确的移动方向, 导致在遥感图像定位性能不佳。CIoU 损失函数<sup>[28]</sup>是 IOU 损失函数改良版, 引入了中心位置和长宽比误差计算: 一方面, 解决了与目标框的重叠问题; 另一方面, 对重叠区域、中心点距离和长宽比方面的考量, 使得损失函数收敛速度更快。因此, 本研究采用 CIoU 损失函数作为边界框回归损失:

$$\begin{cases} L_{CIoU} = 1 - R_{IoU} + \frac{\rho(b, b^{gt})}{c^2} + \alpha v \\ \alpha = \frac{v}{(1 - R_{IoU}) + v} \\ v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \end{cases}, \quad (2)$$

式中:  $R_{IoU}$  为预测框和真实框之间的交并比;  $\rho$  表示两点欧氏距离;  $b$  和  $b^{gt}$  分别表示预测框和真实框的中心点;  $c$  代表预测框和真实框最小外接矩形的对角线距离;  $\alpha$  为协调比例参数;  $v$  为用于衡量框的长宽比一致性的参数;  $w^{gt}$ 、 $w$  和  $h^{gt}$ 、 $h$  分别表示预测框和真实框的宽度和高度。

IoU 感知分类损失和类别损失采用交叉熵损失函数。在模型训练后期, 添加  $L_1$  范数<sup>[29]</sup>边界框损失函数, 进一步降低模型定位误差。改进算法的总损失函数为

$$L = \frac{1}{N_{\text{pos}}} \text{CrossEntropy}(P_{\text{Cls.}}^{\text{pos}}, G_{\text{Cls.}}^{\text{pos}}) + \frac{\lambda_1}{N_{\text{pos}}} \text{CrossEntropy}(P_{\text{Obj.}}^{\text{pos+neg}}, G_{\text{Obj.}}^{\text{pos+neg}}) + \frac{\lambda_2}{N_{\text{pos}}} L_{\text{CIoU}}(P_{\text{box}}^{\text{pos}}, G_{\text{box}}^{\text{pos}}) + \frac{\lambda_3}{N_{\text{pos}}} L_1(P_{\text{Reg.}}^{\text{pos}}, G_{\text{Reg.}}^{\text{pos}}), \quad (3)$$

式中:CrossEntropy、 $L_{\text{CIoU}}$ 和 $L_1$ 分别表示交叉熵损失函数、CIoU损失函数和 $L_1$ 范数; $N_{\text{pos}}$ 表示正样本数量; $P$ 和 $G$ 分别代表预测样本和真实样本;pos和neg则代表训练正样本和负样本;Cls.、Obj.、box和Reg.表示分类数值、IoU感知分数值、定位框和定位偏移量。

## 4 实验与结果分析

### 4.1 实验数据和平台

实验选择由西北工业大学发布的一个大规模、公

开的遥感目标检测基准数据集 DIOR<sup>[30]</sup>来评估和测试算法性能。DIOR数据集总共包含 23463 张图像, 192472 个实例样本, 图像尺寸均为 800 pixel×800 pixel, 有 20 个类别, 分别为飞机、机场、棒球场、篮球场、桥梁、烟囱、水坝、高速公路服务区、高速公路收费站、高尔夫球场、田径场、港口、立交桥、船只、体育场、储罐、网球场、火车站、车辆和风车, 并用 C1~C20 表示。该数据集具有丰富尺度变化的目标、较高的类间相似度和类内多项性及成像差异较大的图像, 可以代表遥感目标检测任务多数应用场景。图 6 展示了该数据集中遥感图像和目标, 其中框出的位置为遥感目标, 图 7 则展示该数据集中目标大小的分布, Small 表示像素低于 32×32 的目标, Middle 表示像素大于 32×32 且小于 96×96 的目标, Large 则表示像素大于 96×96 的目标。



图 6 DIOR数据集图像和目标  
Fig. 6 Images and objects in DIOR dataset

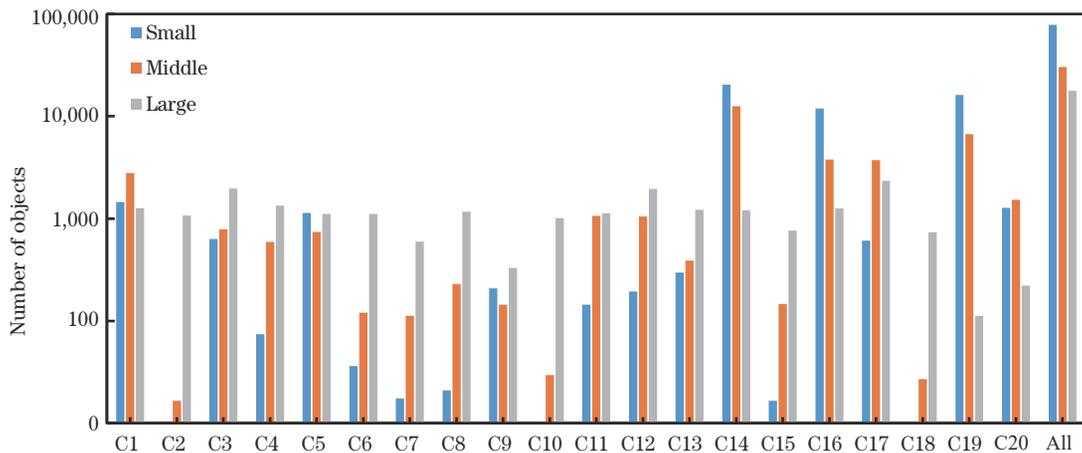


图 7 DIOR数据集各类目标大小分布  
Fig. 7 Size distribution of objects in DIOR dataset by category

模型训练和性能评价实验在 GPU 服务器上完成,服务器硬件配置如下:Inter Core i9-7960X CPU (2.80 GHz), 内存为 64 GB, 显卡型号为 NVIDIA GeForce RTX2080 GPU (11 GB); 软件环境为 Ubuntu 16.04, Python 3.7, Pytorch 1.7.0, MMDetection 工具箱<sup>[31]</sup>, 其中工具箱版本为 mmdet 1.3.9 和 mmdet 2.15.1。为了公平比较,实验设置统一训练参数,如表 1 所示,优化器采用标准 SGD,其动量、衰减系数和 Nesterov 功能分别设置为 0.9、0.0005 和 True,学习率调度器采用余弦退火学习率策略,最小学习率为所设置当前学习率的 5%,并在最后 15 轮采用最小学习率。单次训练样本数量(batch)设置为 16,迭代次数(epoch,训练完所有样本为一轮)为 100。训练集预处理采用 Mosaic 数据增强策略,随机选择 4 张图片,在所有图片缩放后进行拼接,得到 1 张包含多个目标的新图像;再通过随机几何变换对新图像进行随机平移、旋转、缩放等操作,输出尺寸大小为 640 pixel×640 pixel 的增强图像;随后采用光度失真从亮度、对比度、饱和度和色相方面随机调整图像,提高模型泛化能力;最后对图像进行随机缩放,输出图像大小在 320 pixel×320 pixel~640 pixel×640 pixel

表 1 训练参数

Table 1 Training parameters

Name	Value
Optimizer	SGD
Momentum	0.9
Weight decay	$5 \times 10^{-4}$
Nesterov	True
Learning rate scheduler	Type is CosineAnnealing, Learning rate is 0.0025, Min_lr_ratio is 0.05
Batch	16
Epoch	100
Mosaic	Img_scale is (640, 640)
Random affine	Scaling_ratio_range is (0.5, 1.5) Brightness_delta is 32,
Photometric distortion	Contrast_range is (0.5, 1.5), Saturation_range is (0.5, 1.5), Hue_delta is 18

表 2 算法改进逐级叠加性能比较

Table 2 Comparison of progressively improved algorithms

Method	Parameters / $10^6$	AP / %	AP50 / %
YOLOX-Tiny	5.04	46.0	71.66
+ multi-scale prediction method	5.40	47.4(+1.4)	75.10(+3.44)
+ coordinate attention module	5.41	47.7(+0.3)	75.30(+0.20)
+ deformable convolution	5.6	49.8(+2.1)	75.60(+0.30)
+ loss function(proposed optimized model)	5.6	50.1(+0.3)	76.08(+0.48)

之间。测试数据集时,图像缩放至 416 pixel×416 pixel 进行预测。

#### 4.2 评价指标

目前,针对小目标定义尚未达成统一标准<sup>[32]</sup>,考虑到网络结构的特点、人类对场景能有效识别需要的彩色图像最小尺寸和多数学术论文的应用,本研究采用 MS COCO 目标检测评价指标作为算法性能的评估标准<sup>[33]</sup>。平均精度(AP)表示在预测样本与真实样本各种阈值下交并比的平均准确度,阈值在 0.5~0.95,间隔为 0.05,交并比大于阈值为检测正确,反之则错误;AP<sub>50</sub>则表示阈值为 0.5 的准确度;AP<sub>S</sub>表示目标为 Small 的平均准确度;AP<sub>M</sub>表示目标为 Middle 的平均准确度;AP<sub>L</sub>表示目标为 Large 的平均准确度。根据 DIOR 数据集目标尺寸分布,AP<sub>S</sub>、AP<sub>M</sub>和 AP<sub>L</sub>有助于对遥感目标检测任务算法的评估。

此外,算法评价不仅要考虑在准确度方面的提升,还需要考虑模型的部署性。采用模型参数量(parameters)、计算量(FLOPs)和每秒检测帧数(FPS)来评估模型的部署性。参数量和计算量用于评估神经网络模型的权重占用和计算复杂度,FPS 则用于评估模型在设备上的处理速度。

#### 4.3 结果和分析

为了验证本研究各阶段改进的成效,对算法改进逐级叠加,实验结果如表 2 所示。所提算法在参数量仅增加  $0.56 \times 10^6$  的情况下,AP 和 AP<sub>50</sub> 指标分别提升了 4.1 个百分点和 4.42 个百分点。在多尺度预测和可变形卷积两处改进精度提升大,当然代价是模型参数数量的激增;协同注意力模块和损失函数则在几乎不增加模型负担的情况下,改善网络检测能力。多尺度预测模块使 AP<sub>50</sub> 提高了 3.44 个百分点,而 AP 仅提高 1.4 个百分点;可变形卷积使 AP 提高 2.1 个百分点,而 AP<sub>50</sub> 仅提升 0.3 个百分点。多尺度预测更关注预测正确的数量,对单个预测框的质量关注度低,从而在阈值超过某一固定数值提升较大;可变形卷积则相反,更加关注每个预测框质量。

图 8 和图 9 分别展示了模型在训练过程中 AP 和损失函数的变化曲线;算法在起始阶段精度提升明显,损失函数波动较大;在 40 个 epoch 后,精度和损失函数趋于平缓,进入微调阶段;在第 85 个 epoch,精度又有了突破,损失函数也断崖式下跌。一方面由于关闭了

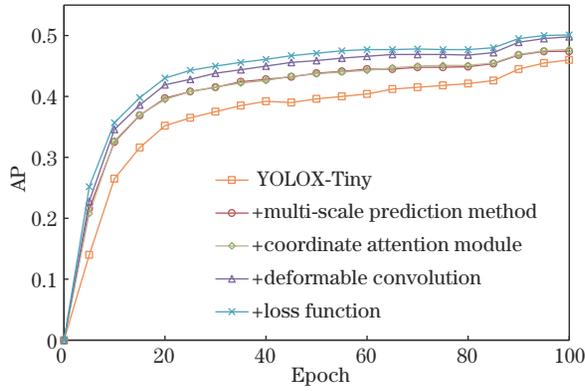


图8 训练过程中 AP 变化曲线  
Fig. 8 AP curve during training

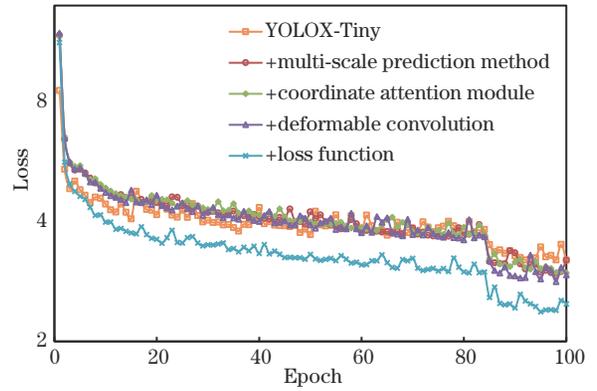


图9 训练过程中损失函数变化曲线  
Fig. 9 Loss curve during training

Mosaic 数据增强引入真实的样本,阻止了过度数据增强所带来的物体标注不准确问题;另一方面,固定最小学习率的方法,改善了权重更新缓慢甚至停滞等问题。此外,引入 L1 范数边界框损失函数,降低了目标定位误差值,加强了算法对定位偏差的敏感性。随着算法的不断改进,模型的收敛速度和识别精度存在明显提升。改进网络结构使精度提升显著,而损失函数数值几乎无太大差异;改进损失函数后,算法在损失函数数值上下下降显著,精度方面也有增益。

从图 7 可知,DIOR 数据集中小目标数量多于中目标和大目标,可从目标尺寸角度分析算法性能,表 3 中展现了不同尺寸下精度。算法在遥感图像检测任务的尺度变化较大的问题上得到改善:其中 AP 指标下大中小目标分别提高 3.3 个百分点、2.9 个百分点和 3.9 个百分点;AP50 指标下大中小目标分别提高 7.4 个百分点、4.6 个百分点和 1.4 个百分点。目标精度提升显著,这极大程度上源于小目标漏检率和误检率的逐渐下降。

表 3 改进算法在不同尺度下检测结果

Table 3 Improved algorithm detection results under different scales

unit: %

Method	AP	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>	AP50	AP50 <sub>S</sub>	AP50 <sub>M</sub>	AP50 <sub>L</sub>
YOLOX-Tiny	46.0	9.5	35.7	66.3	71.66	24.1	58.5	90.4
Proposed optimized model	50.1	12.8	38.6	70.2	76.08	31.5	63.1	91.8

表 4 展示了不同算法 AP50 的结果。C5(桥)和 C7(水坝)作为数据集中相似度较高的类别,所提算法相较于 YOLOX-Tiny 分别提高 4.6 个百分点和 6.2 个百分点;C1(飞机)和 C9(高速公路收费站)作为目标分布最均衡且多样性较高的类别,所提算法相较于 YOLOX-Tiny 分别提高 6.4 个百分点和 10 个百分点。此外,其他类别在精度上均有不同程度提升。改进算法在类间相似度和类内多样性较高的遥感目标检测任务上取得了优秀的表现。

络结构改进前后,对 P2~P5 多尺度预测层特征进行可视化,如图 10 所示。由于多方面的共同作用,所提改进算法的卷积响应可以更好地覆盖不同尺度物体,无论是响应强度还是响应边界都比原始 YOLOX-Tiny 要出色。观察图 10 细节,YOLOX-Tiny 受背景因素影响较大,热力图中物体周围存在热源散漫不集中的问题。所提优化算法的热力图响应效果极佳,这归功于协同注意力模块提取了通道的重要信息,滤除了部分背景噪声,并且可变形卷积根据物体形状发生偏移,采样点可以更精准覆盖物体。

为了便于理解卷积神经网络提取的特征,尤其网

表 4 改进算法在不同类别下 AP50 结果

Table 4 AP50 results under different categories

unit: %

Method	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
YOLOX-Tiny	68	84.6	79.9	86.6	42.6	78.3	71.3	86.5	66.1	80.8	78.8	61.5	59.8	79.7	69.8	61.5	87.5	67.4	40.1	82.3
Proposed optimized model	74.4	89.1	83.8	88.3	47.2	78.6	77.5	88.8	76.1	82.4	81.2	64.5	62.2	87.7	74.3	70.9	88.8	70.3	49.1	86.4

为了直观表现改进算法在遥感目标检测任务的有效性,挑选了几张具有代表性的遥感图像进行检测,检测结果可视化如图 11 所示。所提算法的检测结果更

加准确,在遥感目标漏检、定位不准确、分类错误等问题上得到改善。在第 1 张和第 2 张图中,对烟囱和火车站的识别和定位结果更为准确。第 3 张和第 4 张图像

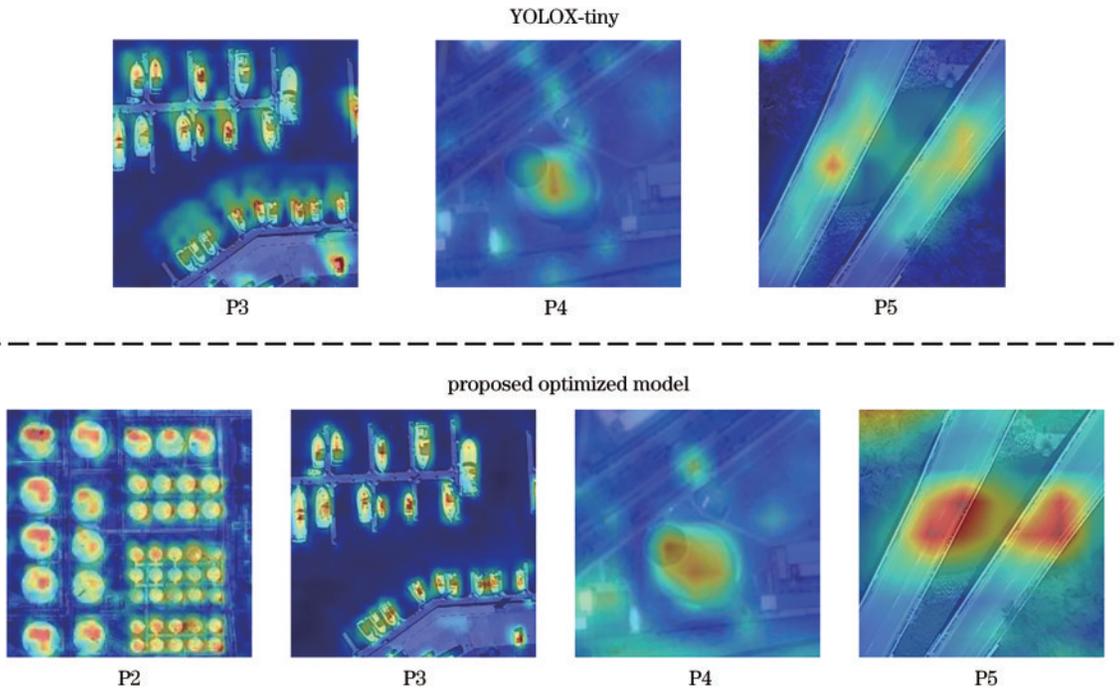


图 10 热力图可视化结果  
 Fig. 10 Thermal map visualization results

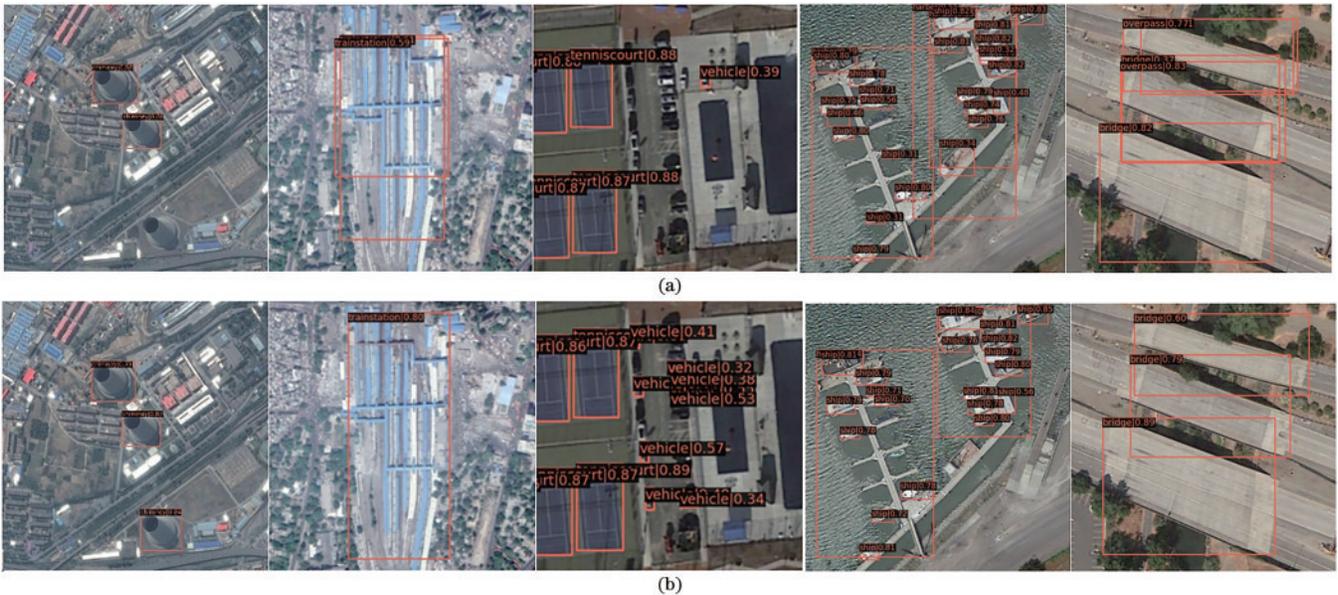


图 11 所提优化算法与基准模型在 DIOR 数据集上检测结果对比。(a) YOLOX-Tiny; (b) 所提算法  
 Fig. 11 Comparison of detection results between YOLOX-Tiny and optimized model. (a) YOLOX-Tiny; (b) proposed algorithm

中目标小而密,漏检、误检问题严重,尤其是车辆和船只,由于物体较小,可获取特征少,所提改进算法可以缓解此问题,准确检测出部分漏检目标。此外,对比第 5 张检测结果可知,基准模型存在将桥梁误检为立交桥的问题,两类别之间目标相似度高,而所提改进算法能够准确识别。

最后,在 DIOR 数据集上,从算法的骨干网络、参数量、计算量、AP50 和 FPS 等 5 个维度对比了当前先进算法,结果如表 5 所示。CF2PN<sup>[34]</sup>、ASSD<sup>[35]</sup> 在

AP50 取得了不错的表现,由于骨干网络采用 VGG16<sup>[36]</sup> 模型,模型整体参数量和计算量相对较大,在性能强劲的显卡上不能达到实时性能,即 FPS 数值小于 25。LO-Det<sup>[37]</sup> 采用轻量级卷积神经网络 MobileNetv2<sup>[38]</sup>,极大地降低了模型的体积和计算负担,但检测精度却不尽如人意。所提算法不仅在识别准确度上表现良好,而且具有体积小、检测快的特点,在各方面性能优于对比算法。

表 5 在 DIOR 数据集上的结果对比  
Table 5 Comparison of results in DIOR dataset

Method	Year	Backbone	Parameters / $10^6$	FLOPs / $10^9$	AP50 /%	FPS	Device
CF2PN <sup>[34]</sup>	2021	VGG16	91.6	>31	67.25	19.7	RTX 2080
ASSD <sup>[35]</sup>	2021	VGG16	>40	>31	71.8	21	RTX TITAN
LO-Det <sup>[36]</sup>	2020	MobileNetv2	6.93	6.424	65.85	60.03	RTX 3090
Proposed optimized model	2021	Modified CSPNet	5.6	7.695	76.08	46.0	RTX 2080Ti

## 5 结 论

针对遥感目标检测存在的复杂几何背景、密集的物体分布及目标尺度差异大等问题,同时为了平衡遥感检测任务中模型的尺寸和准确度,提出了一种基于改进 YOLOX-Tiny 的轻量级遥感目标检测算法。采用多尺度预测方法,引入协同注意力模块和可变形卷积,优化训练损失函数,在不显著增加模型体积的基础上,强化模型提取特征的能力,提升识别精度。在 DIOR 遥感数据集测试中,与基准网络 YOLOX-Tiny 相比,AP50 从 71.66% 提升至 76.08%,相较于对比算法,所提算法在精度和速度上具有明显的优势。研究结果为今后遥感目标检测任务的实际部署打下坚实的基础。

## 参 考 文 献

- [1] 李树涛,李聪好,康旭东.多源遥感图像融合发展现状与未来展望[J].遥感学报,2021,25(1):148-166.  
Li S T, Li C Y, Kang X D. Development status and future prospects of multi-source remote sensing image fusion[J]. National Remote Sensing Bulletin, 2021, 25(1): 148-166.
- [2] 周成虎,孙九林,苏奋振,等.地理信息科学与技术应用[J].地理学报,2020,75(12):2593-2609.  
Zhou C H, Sun J L, Su F Z, et al. Geographic information science development and technological application[J]. Acta Geographica Sinica, 2020, 75(12): 2593-2609.
- [3] 聂光涛,黄华.光学遥感图像目标检测算法综述[J].自动化学报,2021,47(8):1749-1768.  
Nie G T, Huang H. A survey of object detection in optical remote sensing images[J]. Acta Automatica Sinica, 2021, 47(8): 1749-1768.
- [4] Zhao Z Q, Zheng P, Xu S T, et al. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [7] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [9] Zhao X L, Zhang J, Tian J M, et al. Multiscale object detection in high-resolution remote sensing images via rotation invariant deep features driven by channel attention[J]. International Journal of Remote Sensing, 2021, 42(15): 5764-5783.
- [10] 戴媛,易本顺,肖进胜,等.基于改进旋转区域生成网络的遥感图像目标检测[J].光学学报,2020,40(1):0111020.  
Dai Y, Yi B S, Xiao J S, et al. Object detection of remote sensing image based on improved rotation region proposal network[J]. Acta Optica Sinica, 2020, 40(1): 0111020.
- [11] 吴湘宁,贺鹏,邓中港,等.一种基于注意力机制的小目标检测深度学习模型[J].计算机工程与科学,2021,43(1):95-104.  
Wu X N, He P, Deng Z G, et al. A deep learning model of small object detection based on attention mechanism [J]. Computer Engineering & Science, 2021, 43(1): 95-104.
- [12] 汪晨,张辉辉,乐继旺,等.基于深度学习和遥感影像的松材线虫病疫松树目标检测[J].南京师大学报(自然科学版),2021,44(3):84-89.  
Wang C, Zhang H H, Yue J W, et al. Object detection to the pine trees affected by pine wilt disease in remote sensing images using deep learning[J]. Journal of Nanjing Normal University (Natural Science Edition), 2021, 44(3): 84-89.
- [13] 李竺强,朱瑞飞,马经宇,等.联合连续学习的残差网络遥感影像机场目标检测方法[J].光学学报,2020,40(16):1628005.  
Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. Acta Optica Sinica, 2020, 40(16): 1628005.
- [14] 农元君,王俊杰.基于嵌入式的遥感目标实时检测方法[J].光学学报,2021,41(10):1028001.  
Nong Y J, Wang J J. Real-time object detection in remote sensing images based on embedded system[J]. Acta Optica Sinica, 2021, 41(10): 1028001.
- [15] 沈丰毅,王琴,蒋剑飞.基于改进卷积神经网络舰船的

- 实时目标检测算法[J]. 计算机应用研究, 2020, 37(S1): 316-319.
- Shen F Y, Wang Q, Jiang J F. Real-time automatic ship detection based on modified convolutional neural networks[J]. Application Research of Computers, 2020, 37(S1): 316-319.
- [16] 闫皓炜, 张洁, 燕正亮, 等. 改进 CenterNet 模型在遥感影像输电杆塔中的应用[J]. 遥感信息, 2021, 36(4): 84-91.
- Yan H W, Zhang J, Yan Z L, et al. Tower recognition of satellite imagery with improved CenterNet model[J]. Remote Sensing Information, 2021, 36(4): 84-91.
- [17] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [18] Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-YOLOv4: scaling cross stage partial network[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13024-13033.
- [19] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06) [2021-09-10]. <https://arxiv.org/abs/2107.08430>.
- [20] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.
- [21] 张宏群, 班勇苗, 郭玲玲, 等. 基于 YOLOv5 的遥感图像舰船的检测方法[J]. 电子测量技术, 2021, 44(8): 87-92.
- Zhang H Q, Ban Y M, Guo L L, et al. Detection method of remote sensing image ship based on YOLOv5 [J]. Electronic Measurement Technology, 2021, 44(8): 87-92.
- [22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [23] Li H C, Xiong P F, An J, et al. Pyramid attention network for semantic segmentation[EB/OL]. (2018-11-25) [2021-09-10]. <https://arxiv.org/abs/1805.10180>.
- [24] 余培东, 王鑫, 江刚武, 等. 一种改进 YOLOv4 的遥感影像典型目标检测算法[J]. 测绘科学技术学报, 2021, 38(3): 280-286.
- Yu P D, Wang X, Jiang G W, et al. A typical target detection algorithm in remote sensing images based on improved YOLOv4[J]. Journal of Geomatics Science and Technology, 2021, 38(3): 280-286.
- [25] 李红光, 于若男, 丁文锐. 基于深度学习的小目标检测研究进展[J]. 航空学报, 2021, 42(7): 024691.
- Li H G, Yu R N, Ding W R. Research development of small object tracking based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 024691.
- [26] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [27] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9300-9308.
- [28] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(07): 12993-13000.
- [29] 周巍. L1 范数最小化算法及应用[D]. 广州: 华南理工大学, 2013.
- Zhou W. L1-norm minimization algorithms and applications [D]. Guangzhou: South China University of Technology, 2013.
- [30] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [31] Chen K, Wang J Q, Pang J M, et al. MMDetection: open MMLab detection toolbox and benchmark[EB/OL]. (2019-06-17) [2021-09-10] <https://arxiv.org/abs/1906.07155v1>.
- [32] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3): 391-417.
- Gao X B, Mo M J C, Wang H T, et al. Recent advances in small object detection[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 391-417.
- [33] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.
- [34] Huang W, Li G Y, Chen Q Q, et al. CF<sub>2</sub>PN: a cross-scale feature fusion pyramid network based remote sensing target detection[J]. Remote Sensing, 2021, 13(5): 847.
- [35] Xu T, Sun X, Diao W H, et al. ASSD: feature aligned single-shot detection for multiscale objects in aerial imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5607117.
- [36] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2021-09-10]. <https://arxiv.org/abs/1409.1556>.
- [37] Huang Z C, Li W, Xia X G, et al. LO-det: lightweight oriented object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5603515.
- [38] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.