

基于 YOLOv7-tiny 改进的遥感小目标检测算法

王子龙¹ 荣 杰²

(1. 江苏科技大学计算机学院 镇江 212100; 2. 南京邮电大学材料科学与工程学院 南京 210000)

摘 要: 为了解决遥感图像中小目标的误检、漏检难题,提出了一种改进的 YOLOv7-tiny 算法。首先,引入高效多尺度注意力模块(efficient multi-scale attention, EMA),基于此设计了多尺度特征提取模块 ELAN-EMA,这大大增强了骨干网络对于多尺度特征的提取能力;其次,在特征金字塔网络(feature pyramid network, FPN)中引入内容感知特征重组(content-aware reassembly of features, CARAFE)优化最近邻上采样方法,设计了 FPN-CARAFE 结构,扩大了感受野,从而能够获取小目标更多的细节信息和丰富的语义信息;最后,采用归一化距离损失函数(normalized wasserstein distance, NWD)优化 CIoU 损失函数,设计了 NWD-CIoU 损失函数,降低了 CIoU 对小目标位置偏移的敏感性,能够更好地提升小目标的检测效果。在公开的遥感数据集 RSOD 和 NWPU VHR-10 上进行的实验表明,与基准模型相比,在计算量和参数量略增长的情况下,改进的模型在平均精度均值(mAP)mAP@0.5 上分别提升了 3.6% 和 1.8%,有效地提高了遥感图像中小目标的检测精度,综合性能优于其他算法,满足部署在遥感检测系统上的要求。

关键词: 目标检测;小目标;注意力机制;感受野;损失函数

中图分类号: TP399 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved remote sensing tiny object detection algorithm based on YOLOv7-tiny

Wang Zilong¹ Rong Jie²

(1. College of Computer Science, Jiangsu University of Science and Technology, Zhenjiang 212100, China;

2. College of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210000, China)

Abstract: Seeking to resolve the issue of missed and incorrect detection of small targets in remote sensing images, this study proposes an optimized YOLOv7-tiny algorithm. Firstly, a multi-scale attention efficient multi-scale attention (EMA) module is introduced, and based on this, the ELAN-EMA, a multi-scale feature extraction module, is incorporated to to greatly enhance the backbone network's proficiency in extracting features across various scales. Secondly, the feature pyramid network (FPN) is introduced with the content-aware reassembly of features (CARAFE) optimization, which expands the receptive field and enables the acquisition of more detailed information and rich semantic information of small targets. Finally, this study adopts the normalized wasserstein distance (NWD) loss function to optimize the complete intersection over union (CIoU) loss function, and designs the NWD-CIoU loss function, which reduces the sensitivity of CIoU to small target position shifts and can better improve the detection performance of small targets. Experiments conducted on the publicly available remote sensing datasets RSOD and NWPU VHR-10 show that compared with the baseline model, the optimized model achieves a 3.6% and 1.8% increase in mAP@0.5, respectively, with slightly increased computational and parameter requirements, markedly enhancing the accuracy with which small targets are detected in remote sensing images. The comprehensive performance meets the requirements for deployment in remote sensing detection systems.

Keywords: object detection; tiny object; attention mechanism; receptive field; loss function

0 引言

近些年来,随着遥感技术的快速发展^[1],其在日常生活和工业界的应用越来越广泛。然而,遥感图像中的目标通常较小,这使得现有的目标检测算法在识别精度上面临挑战。因此,开发一种针对小目标检测具有高精度的目标检测算法成为了一个迫切的需求^[2]。

当今卷积神经网络(convolutional neural networks, CNN)在各行各业已经被广泛应用,取得了不错的效果,已经成为了主流的研究方向^[3-5]。目前,主要的目标检测算法可以分为单阶段和双阶段两大类。以 R-CNN^[6]、Fast R-CNN^[7]、Faster R-CNN^[8] 等为代表的双阶段目标检测算法,其基本流程是先生成候选框,再对这些候选框进行分类和回归。这种双阶段的设计方式通常可以获得更高的准确率,但在执行速度上较慢。另一方面,单阶段目标检测算法以 YOLO^[9] 和 SSD^[10] 等为代表,其方法是将图像划分为 $S \times S$ 个网格,然后通过每个小网格的中心区域对图像中的目标进行分类和回归。由于这种方法不需要预先生成候选框,因此相比双阶段目标检测算法具有更高的检测效率。

针对遥感图像中小目标检测^[11-14]的问题,主要存在如下4种解决思路。1)采用数据增强策略^[15]。CNN 效果好的一个重要原因就是依赖于大量有标注数据的训练。而标注大量数据集是一件极其消耗时间的工作,Wang 等^[16]提出了一个基于透视变换的自动化样本标注框架,无需手动标注便可以自动生成新的标注数据,不仅有效的扩充了训练数据集数量,而且节省了大量人力资源。2)采用多尺度学习策略^[17]。由于网络的深度增加,图像的深层语义信息越来越丰富,但是目标的特征信息和位置信息逐渐丢失,导致了图像的浅层表征信息越来越少,尤其不利于小目标的检测,虽然浅层网络的表征信息较为丰富,但是深层语义信息较少,不利于目标的分类,因此将深层的语义信息与浅层的表征信息结合变得十分重要。而 FPN 正是将深层的语义信息与浅层的表征信息结合,融合了多尺度的特征图,显著的提高了模型对小目标的检测精度。张朝阳等^[18]通过在 YOLOv5s 的检测头中增加小目标检测层,颈部采用双向特征金字塔,加入多头自注意力机制方式等方式增强网络的多尺度特征提取和融合能力,在一定程度上提高了小目标的检测精度。张寅等^[19]首先在骨干网络中设计了特征增强模块,增强小目标的特征表达能力,其次设计了级联注意力机制,更精确的捉小目标的特征信息。Wang 等^[20]设计了全局感知特征重组模块来提取图像的多尺度信息和全局特征信息,并设计混合特征增强模块从空间和通道维度增强全局特征,最后,设计边界优化模块以细化边界特征,显著的提升了模型的检测效果。3)采用优化损失函数的策略。Liu 等^[21]为了解决小目标和大目标之间损失存在显著差距的问题,设计了反馈驱动损失函数,反馈驱动损失函数使用损失分布信息作为反馈信

号,以更平衡的方式训练检测器,有效的提升了数据集中小目标的检测效果,并且没有增加额外的计算量。Deng 等^[22]针对小目标前景背景难以区分的问题,设计了前景背景平衡损失函数,以此来平衡小目标的前景和背景信息,在小目标数据集上检测效果有了显著的提升。4)采用上下文学习策略。目前比较典型的 YOLO 模型没有充分利用图像的上下文信息,张华卫等^[23]在网络中融入了 Attention Gate 模块,可以让网络学习到有用的上下文信息,抑制无用的上下文信息。可以有效的增加网络的上下文信息学习能力。何秋润等^[24]设计了一种加强上下文信息融合的检测算法(CMFFD-YOLO),通过在主干网络中添加上下文网络模块,充分利用局部全局上下文信息,在一定程度上提高了小目标的检测精度。

上述算法虽然在一定程度上可以提高遥感图像中小目标的检测性能,但是同时也存在一些问题,主要原因如下:1)虽然采用特征金字塔网络(feature pyramid network, FPN)融合了网络深层和浅层的特征图,在一定程度上缓解了深层网络特征信息缺失的问题,但是骨干网络多尺度特征提取能力一般;2)传统的上采样方法比如最邻近和双线性插值法,在上采样的过程中,仅局限于借助像素点的空间位置,却未能充分考虑特征图的语义信息,同时有较小的感受野;3)传统的 CIoU 度量方式对小目标的位置偏差非常敏感,微小的位置偏差会造成 CIoU 值的显著下降,造成对小目标的检测精度不高。

针对上述问题,本文对 YOLOv7-tiny 模型进行了改进。1)在骨干网络设计了多尺度特征提取模块 ELAN-EMA,这增强了网络的多尺度特征提取能力,在模型参数量和计算量略微增加的情况下,显著提高了目标检测的精度。2)在 FPN 结构中引入内容感知特征重组(content-aware reassembly of features, CARAFE)上采样算子优化最近邻上采样方法,设计了 FPN-CARAFE 结构,扩大了模型的感受野,使其能更好地关注小目标细节信息,在略微增加计算量和参数量的情况下显著加强了小目标的检测效果,提高了检测精度。3)引入了归一化距离(normalized wasserstein distance, NWD)损失函数,并将其与 CIoU 损失函数结合,设计了 CIoU-NWD 加权损失函数,降低了 CIoU 损失函数对小目标位置偏移的敏感度,在计算量和参数量略微增加的情况下,进一步提高小目标的检测精度。此外分别在 RSOD 和 NWPU VHR-10 遥感数据集上进行了实验,实验结果表明,本文所提出的改进模型对包含大量小目标的遥感图像具有良好的检测效果。

1 改进的 YOLOv7-tiny 算法

YOLOv7-tiny^[25]于2022年7月发布,其在速度和精度上均超越了大多数已有的目标检测模型。YOLOv7-tiny 的网络结构主要由以下4个主要模块构成,输入模块、骨干网络、颈部网络以及检测头。首先,图像经过输入模

块进行数据增强和其他预处理操作后,被送入骨干网络进行特征提取。接着,在颈部网络进行特征融合处理,得到大、中、小3种尺度的特征。最后,将这些特征进行融合,

再送入检测头进行检测,并输出最终结果。鉴于遥感数据集中小目标检测精度不高的问题,在本文对YOLOv7-tiny模型的网路结构进行了改进,如图1所示。

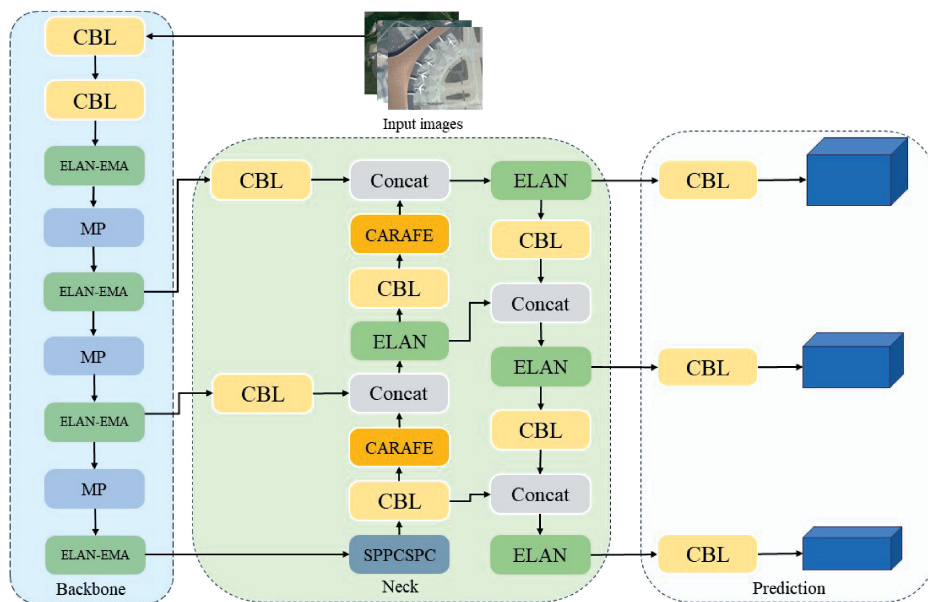


图1 改进后的YOLOv7-tiny网络结构

Fig. 1 The improved YOLOv7-tiny network architecture

1.1 ELAN-EMA 模块

引入注意力机制^[26]能够帮助模型集中关注输入序列中的关键信息,以此来提升模型的检测效果。已有一些典型的注意力机制,如SE^[27](squeeze-and-excitation)、CBAM(convolutional block attention module)、ECA^[28](efficient channel attention)、CA(coordinate attention)等,已被证明对于特征提取具有显著的效果。SE注意力机制通过自动学习每个通道的重要性,从而关注到一些比较重要的通道信息,对空间信息关注较少。而CBAM注意力机制则同时关注空间信息和通道信息,但不能建立长距离的通道依赖关系,而且同时会增加模型的计算量。而CA注意力机制在通道中嵌入位置信息,并且在通道维度捕捉长距离的依赖关系,展现出了较好的性能,但是因为内部采用 1×1 卷积导致了网络较小的感受野,无法有效的从全局空间维度捕捉通道之间的依赖关系和上下文信息。介于以上几种常见注意力机制的缺点,因此,本文引入了一种新的高效的多尺度注意力模块(EMA)^[29]。EMA拥有建立长距离依赖关系和短距离依赖关系的并行分支,并且采用对全局信息编码的方式,以揭示像素级的成对关系,可以将精确的位置信息嵌入到EMA中。EMA注意力机制模块的结构如图2所示。

在EMA模块中,第1步将输入的特征图从通道维度进行分组,对于输入 $C \times H \times W$ 的特征图,分为 G 组,用于学习不同的语义信息,则分组后特征图的大小为 $(C/G) \times W \times H$ 。第2步对每组特征图通过3个并行分

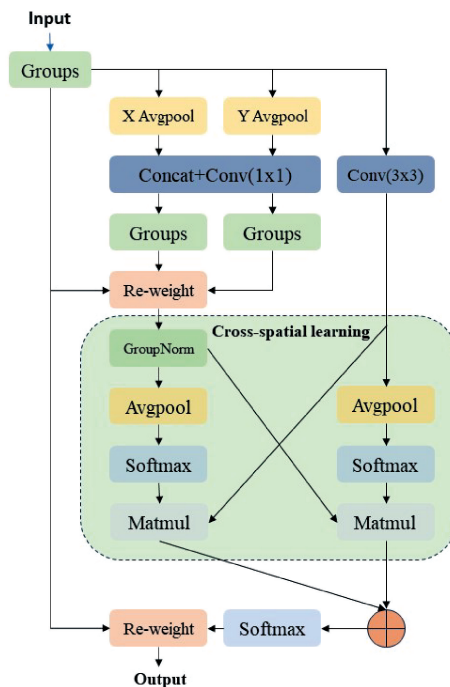


图2 EMA模块结构

Fig. 2 Structure of the EMA module

支提取特征图的注意力权重,在前两个分支通过 1×1 卷积分别对两个空间维度进行全局平均池化,用来对通道进行编码。X Avgpool代表水平全局池化,Y Avgpool代表垂直全局池化。然后对编码后的特征图进行Concat操

作,用 1×1 共享卷积进行特征编码之后生成两组编码向量,利用激活函数对其进行线性拟合,然后对两组特征图进行 Re-weight 操作进行跨通道特征交互。 3×3 分支借助 3×3 卷积捕获局部跨通道交互以扩大特征空间。这样,EMA 不仅对信道间信息进行编码以调整不同信道的重要性,而且将精确的空间结构信息保留到信道中。在 Cross-spatial learning 部分,通过对不同空间维度方向进行跨空间信息聚合,以实现更丰富的特征聚合,用来建立短期依赖和长期依赖的关系。首先对 1×1 分支输出的张量进行归一化操作,然后利用二维全局平均池化对 1×1 支路的输出进行全局空间信息编码和 3×3 的输出进行跨空间的信息交互,采用二维高斯映射的自然非线性函数 Softmax 来拟合以上的线性变换。在通道特征联合激活机制之前,将最小支路的输出直接转化为对应的维度形状,用来进行跨空间特征融合。最后,通过跨空间维度交互模块汇聚两个通道注意力权重值,将每组内的输出特征映射计算为生成的两个空间注意力权重值的集合,然后使用 Sigmoid 函数。这个过程能捕获像素级的成对关系,并突出显示所有像素的全局上下文信息。最后,每个分组通过矩阵进行点积运算,得到 $C \times H \times W$ 的特征图。

因此,本文将 EMA 模块与 ELAN 模块相结合,将其称为 ELAN-EMA 模块。改进后的 ELAN-EMA 模块在骨干网络中不仅可以加强网络的多尺度特征信息提取能力,而且可以获得更多的上下文信息,增强了网络的特征学习和表示能力,可以有效的提升模型对小目标的检测能力,从而优化了模型的性能。ELAN-EMA 模块的结构如图 3 所示。

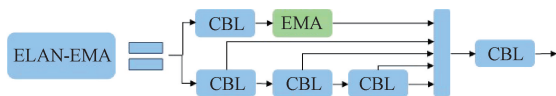


图 3 ELAN-EMA 结构

Fig. 3 Structure diagram of the ELAN-EMA module

1.2 改进的 FPN-CARAFE 结构

特征上采样的设计对于小目标检测任务至关重要,典型的上采样方法有最近邻、线性插值、二次插值等方法。YOLOv7-tiny 模型采用的是最近邻插值法进行上采样操作,然而,仅局限于借助像素点的空间位置来生成上采样,而忽视了特征图语义信息,并且很容易导致特征信息的丢失。为了解决这些问题,本文在 FPN 结构中引入 CARAFE^[30]上采样模块,改进了最近邻上采样模块。CARAFE 能够在一个大的接收域内聚合上下文信息,有更大的感受野,同时,该模块可以在不同的位置使用自适应和优化的重组核,可以更好地适应不同区域的特征,而不是简单地进行统一的上采样操作,从而能提取到更加丰富的小目标的特征信息和语义信息,更好的在 FPN 结构中进行特征融合,相较于最近邻上采样操作,能实现更优的性

能。CARAFE 上采样模块结构图和改进后 FPN-CARAFE 的结构分别如图 4 和 5 所示。

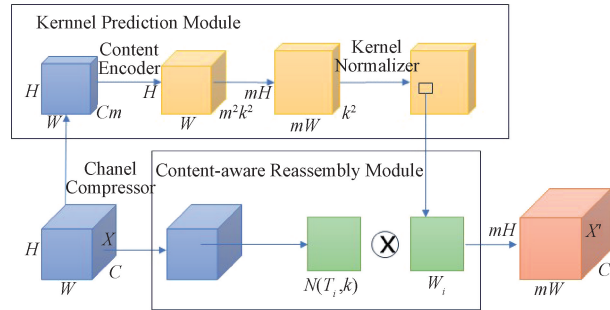


图 4 CARAFE 模块结构

Fig. 4 Structure diagram of the CARAFE module

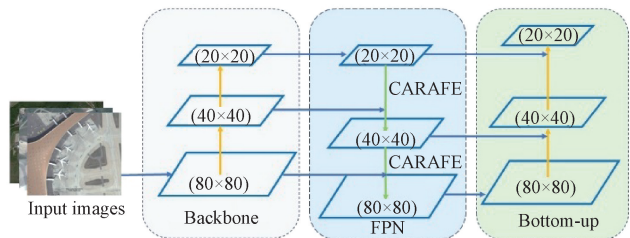


图 5 改进后的 FPN-CARAFE 模块结构

Fig. 5 Structure diagram of the improved FPN-CARAFE module

CARAFE 的上采样过程主要由两个步骤完成。第 1 步根据每个目标的具体位置,动态地生成一个上采样核。首先借助 1×1 卷积对 $C \times H \times W$ 特征图将通道数从 C 压缩到 C_m ,旨在降低后续计算的复杂性和参数量;随后,借助 $K_{encoder} \times K_{encoder}$ 卷积层来生成上采样核,上采样的倍率为设定为 σ ,得到具有 $\sigma^2 K_{up}^2$ 个输出通道的特征图,接着在空间维度展开,对得到形状为 $K_{up} \times K_{up} \times \sigma H \times \sigma W$ 的上采样核进行归一化处理,从而确保卷积核的权重为 1。第 2 步利用第 1 步得到的重组核进行特征重组。特征重组模块对于输出特征图中的每个位置,映射到输入特征图中大小为 $K_{up} \times K_{up}$ 的区域,然后与预测出的上采样核进行点积计算,最终得到一个大小为 $\sigma H \times \sigma W \times K_{up} \times K_{up}$ 的特征图。

实验结果表明,在算法中应用改进后 FPN-CARAFE 模块,参数量和计算量略微增加的情况下,对于小目标检测模型的性能有了显著的提升。

1.3 CIoU-NWD 加权损失函数

传统的 CIoU 损失函数对于小目标位置的微小偏移非常敏感。这种敏感性主要源于边界框位置的离散变化特性,这意味着 CIoU 度量对于具有离散位置偏差的目标不再保持尺度不变。这样的现象最终导致了标签分配的不准确性,从而使得 CIoU 成为了一个针对微小目标的不良度量,这进一步导致基于锚点的目标检测器的性能大幅

下降。为了解决这个问题,本文引入了NWD^[31]损失函数来度量两个边界框之间的相似性。具体来说,首先用2D高斯分布来建模边界框,然后采用了一种新的度量方法,即NWD,通过计算它们对应的高斯分布来衡量它们之间的相似性。由于NWD对不同尺度的目标不敏感,因此更适合于测量微小目标之间的相似性。在损失函数中,引入了NWD,并保留了原有的CIoU度量,通过设置一个平衡因子 α 来调和这两种度量方式。在遥感数据集上,这种新的度量方式表现出了优良的检测性能。

对于微小的物体,本文用二维(2D)高斯分布来建模边界框,具体而言,对于一个以 (c_x, c_y) 为中心,宽度为 w ,高度为 h 的水平边界框 $R = (c_x, c_y, w, h)$,可以将其建模为二维高斯分布 $N(\mu, \Sigma)$ 。在这种情况下,将边界框的相似性转化为高斯分布的距离。建模过程如下:

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (1)$$

采用Wasserstein距离来衡量两个边界框 A 和 B ,其坐标分别为 $A = (cx_a, cy_a, w_a, h_a)$ 和 $B = (cx_b, cy_b, w_b, h_b)$,定义它们的高斯分布距离如下:

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_2^2 \quad (2)$$

$W_2^2(N_a, N_b)$ 作为距离度量,不适合直接用于相似度量(即使用介于0~1的值来表示IoU)。因此,采用其指数形式进行归一化,从而得到了一种新的度量方法,被称NWD,公式如下:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (3)$$

式中: C 是一个常数。

本文采用CIoU-NWD加权损失函数优化CIoU损失函数。引入了一个平衡因子 α ,以便在这两种度量方式之间取得平衡。改进后的损失函数定义如下:

$$L_{obj} = \alpha \times (1 - IoU) + (1 - \alpha) \times (1 - NWD(N_a, N_b)) \quad (4)$$

2 实验结果与分析

2.1 实验环境以及参数设置

实验的运行环境设定如下:操作系统为Ubuntu 20.04,显卡型号为NVIDIA GeForce RTX 3090,服务器运行内存大小为6 GB,CUDA 11.7版本,Pytorch 2.0版本,Python 3.9版本。

本文实验一共进行了300个训练周期(Epochs),每个批次(Batch)的大小设定为16,优化器选择为SGD优化器,动量参数设定为0.937,初始学习率设定为0.01。

2.2 实验数据集

实验使用RSOD和NWPU VHR-10遥感数据集。其

中,RSOD数据集由武汉大学标注,涵盖了飞机、油罐、游乐场和立交桥4类目标。该数据集总共收录了976张图片,包含6950个目标实例,这些实例包括了446张图片中的4993架飞机、189张图片中的191个游乐场、176张图片中的180个立交桥,以及165张图片中的1586个油罐。另一方面,NWPU VHR-10数据集由西北工业大学标注,由800张遥感图像,以及3651个目标实例构成,包含了飞机、网球场、船舶、车辆、储罐、桥梁、棒球场、地面跑道、篮球场、港口等10类目标。

2.3 评价指标

沿用先前研究的评价方式,精确率(precision, P),召回率(recall, R),所有类别平均精度(average precision, AP)和平均精度均值(mean average precision, mAP)。参数量(Param),反映了实验模型占用内存大小。平均每帧检测时间(Time),反映了实验模型检测速度。模型复杂度计算量(FLOPs),用于衡量实验模型的计算复杂度。

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

式中: TP 意味着预计是正样本真实也是正样本; FP 意味着预计是正样本实际是负样本; TN 意味着预计是负样本实际是正样本; FN 意味着预计是负样本实际是负样本。

$$AP = \int_0^1 p(R) dR \quad (7)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (8)$$

2.4 消融实验

为了验证CIoU-NWD损失函数对小目标检测改进效果的有效性,本文选择在RSOD数据集上设定不同的平衡因子 α 进行了一系列的对比实验,通过对比实验的结果,以此找到模型最优性能下平衡因子的值,实验结果如表1所示。

表1 在RSOD数据集不同平衡因子 α 的实验结果
Table 1 Different balance factors in the RSOD dataset α experimental results

方法	α	mAP@0.5/%	浮点数 /GFLOPs	Time/ms
基线	0	96.5	13.0	3.11
改进1	0.2	97.3	13.0	3.30
改进2	0.4	96.8	13.0	3.19
改进3	0.6	96.5	13.0	3.60
改进4	0.8	96.6	13.0	3.76
改进5	1.0	94.5	13.0	3.80

从表1可以看出,在CIoU-NWD加权损失函数中,通过设定不同的平衡因子 α (α 代表在加权损失函数中CIoU度量方式的权重),模型的计算量没有变化,检测速度有略

微的变化,但是检测精度都优于基线模型,这验证了改进方法的有效性。当 α 值设置为0.2,即NWD度量方式的权重为0.8,CIoU度量方式的权重为0.2时,在平均检测时间略微增加的情况下,模型的检测精度达到了最佳,比基线模型提高了2.8%,而平均每帧检测时间仅增加了0.19 ms。可以很好的提升模型的检测性能。

为了验证在模型骨干网络中引入EMA注意力机制能有效提升模型性能,验证EMA注意力机制的有效性,在SE,ECA,CBAM,CA等经典注意力机制进行了对比实验,实验结果如表2所示。

表2 嵌入不同注意力机制的ELAN模块对比实验				
Table 2 Comparative experiment of ELAN modules embedded with different attention mechanisms				
注意力机制	mAP@0.5/%	Param/($\times 10^6$)	浮点数/GFLOPs	Time/ms
YOLOv7-tiny	94.5	6.02	13.0	3.11
+SE	96.0	6.03	13.11	4.29
+CBAM	96.2	6.03	13.16	17.24
+ECA	95.6	6.02	13.13	4.00
+CA	93.7	6.04	13.14	19.60
+EMA	96.3	6.04	13.31	3.80

从对比实验结果来看,只有CA注意力机制降低了检测精度,而其他几种注意力机制均在一定程度上提升了检测精度,在这其中,采用ECA注意力机制对模型检测效果提升最小,而EMA注意力机制的提升效果最显著。在模型检测速度方面,加入了CBAM注意力之后平均每帧检测时间大幅度增加,加入EMA注意力则略微增加,模型的参数量和计算量也是略微的增加,这进一步验证了EMA注意力机制的有效性,能够有效地提升目标检测的性能。

为了验证改进的CIoU-NWD损失函数的性能,本研究在CIoU,DIoU,EIoU,GIoU以及CIoU-NWD等损失函数中进行了一系列的实验对比,以此来表明CIoU-NWD性能最优,实验对比结果如表3所示。

从实验对比可以观察到,DIoU,EIoU,GIoU以及CIoU-NWD等损失函数都在不同程度上提高了模型的检测精度。其中,CIoU-NWD损失函数对模型精度的提升效

表3 不同损失函数对比实验			
Table 3 Comparative experiments on different loss functions			
损失函数	mAP@0.5/%	浮点数/GFLOPs	Time/ms
CIoU	94.5	13.0	3.11
DIoU	95.7	13.0	3.70
EIoU	95.6	13.0	3.60
GIoU	96.1	13.0	3.10
CIoU-NWD	97.3	13.0	3.30

果最为显著,在模型检测速度方面,其他几种损失函数略微增加平均每帧检测时间,且没有增加模型的计算量,这进一步证实了CIoU-NWD损失函数的有效性。

消融实验结果如表4所示,可以看出,采用CARAFE上采样算子优化最近邻上采样后,检测精度相比基线模型提高了1.8%,计算量增加了0.1 GFLOPs,参数量增加了 0.03×10^6 ,检测时间增加了0.16 ms,虽然计算量、参数量、检测时间略微增加,但是检测精度有明显的提升,这表明CARAFE上采样算子可以有效地扩大模型的感受野,从而增强模型对小目标特征的提取能力。进一步采用CIoU-NWD损失函数替换传统的CIoU损失函数,检测精度再次提高了1.3%,参数量、计算量没有增加,平均每帧检测时间增加了0.1 ms,由此可见,在检测速度略微变慢的情况下,模型的检测精度则有着较大的提升,由此可以看出,CIoU-NWD损失函数能有效地解决CIoU损失函数对小目标偏移敏感的问题,提高小目标的检测精度。最后,在骨干网络中添加EMA注意力机制,利用ELAN-EMA模块取代ELAN模块,检测精度又提升了0.5%,参数量增加了 0.03×10^6 ,计算量增加了0.3 GFLOPs,平均每帧检测时间增加了1.43 ms,同样的在计算量、参数量、平均每帧检测时间略微增加的情况下,模型的检测精度有着明显的提升,这进一步证实了EMA注意力机制能增强模型的多尺度特征提取能力。最终,本文所提出的改进方法与基线模型相比,参数量增加了 0.06×10^6 ,计算量增加了0.4 GFLOPs,平均每帧检测时间增加了1.79 ms,在这些指标略微增加的情况下,检测精度提高了3.6%,综合性能较优,充分验证了改进方法的有效性。

表4 在RSOD数据集消融实验结果							
Table 4 Results of ablation experiments on the RSOD dataset							
方法	CARAFE	CIoU-NWD	ELAN-EMA	mAP@0.5/%	Param/($\times 10^6$)	浮点数/GFLOPs	Time/ms
YOLOv7-tiny	×	×	×	94.5	6.02	13.0	3.11
改进1	✓	×	×	96.3(+1.8)	6.05	13.1	3.37
改进2	✓	✓	×	97.6(+3.1)	6.05	13.1	3.47
本文	✓	✓	✓	98.1(+3.6)	6.08	13.4	4.90

2.5 主流模型对比试验

本文针对改进的模型,与其他主流模型在 RSOD 数据集和 NWPU VHR-10 数据集上分别进行了比较实验,结果如表 5、6 所示。从表 5 实验结果来看,改进的 YOLOv7-tiny 模型参数量为 6.08×10^6 ,相比其他主流模型,比如 YOLOv3、YOLOv4、YOLOv5、YOLOX,参数量较小,比基线模型仅增加了 0.06×10^6 ,平均每帧检测时间为 4.9 ms,虽然相比基线模型略微增加,但是远小于其他模型检测时间,而且展现出了较高的检测精度。

表 5 在 RSOD 数据集实验结果

Table 5 Experimental results on the RSOD dataset			
方法	mAP@0.5/%	Param/($\times 10^6$)	Time/ms
YOLOv3 ^[32]	81.36	61.60	21.8
YOLOv4 ^[32]	90.23	64.20	17.6
YOLOv5 ^[32]	90.97	7.50	11.9
YOLOX ^[32]	92.10	8.94	14.9
YOLOv7-tiny	94.50	6.02	3.1
本文	98.10	6.08	4.9

为了进一步证明实验的有效性,又在 NWPU VHR-10 遥感数据集中进行了实验,如表 6 实验结果可以发现,改进后的 YOLOv7-tiny 参数量为 6.07×10^6 ,虽然比基线

模型略微增加了 0.04×10^6 ,但是相比其他模型大大减小,平均每帧检测时间为 7.5 ms,与基线模型相比略微增加,但是与其他模型相比远远小于其他模型检测时间,而且本文改进的模型检测效果明显优于其他模型,这充分证明了本文提出的改进策略能够有效地提高遥感数据集中小目标的检测精度。

表 6 在 NWPU VHR-10 数据集实验结果

Table 6 Experimental results on the NWPU VHR-10 dataset

方法	mAP@0.5/%	Param/($\times 10^6$)	Time/ms
Faster R-CNN ^[33]	81.39	60.68	139.3
YOLOv3 ^[33]	82.29	61.95	24.8
YOLOv4 ^[33]	86.31	62.73	23.3
YOLOv5 ^[33]	91.08	7.07	18.4
YOLOv7-tiny	92.10	6.03	5.9
本文	93.90	6.07	7.5

2.6 可视化结果对比

为了验证实验的有效性,在 NWPU VHR-10 数据集上进行了可视化实验,如图 6 所示。从对比效果可以看出,本文所提出的改进模型的检测可以有效的解决原始模型存在漏检、误检的问题,检测效果明显优于原始模型。

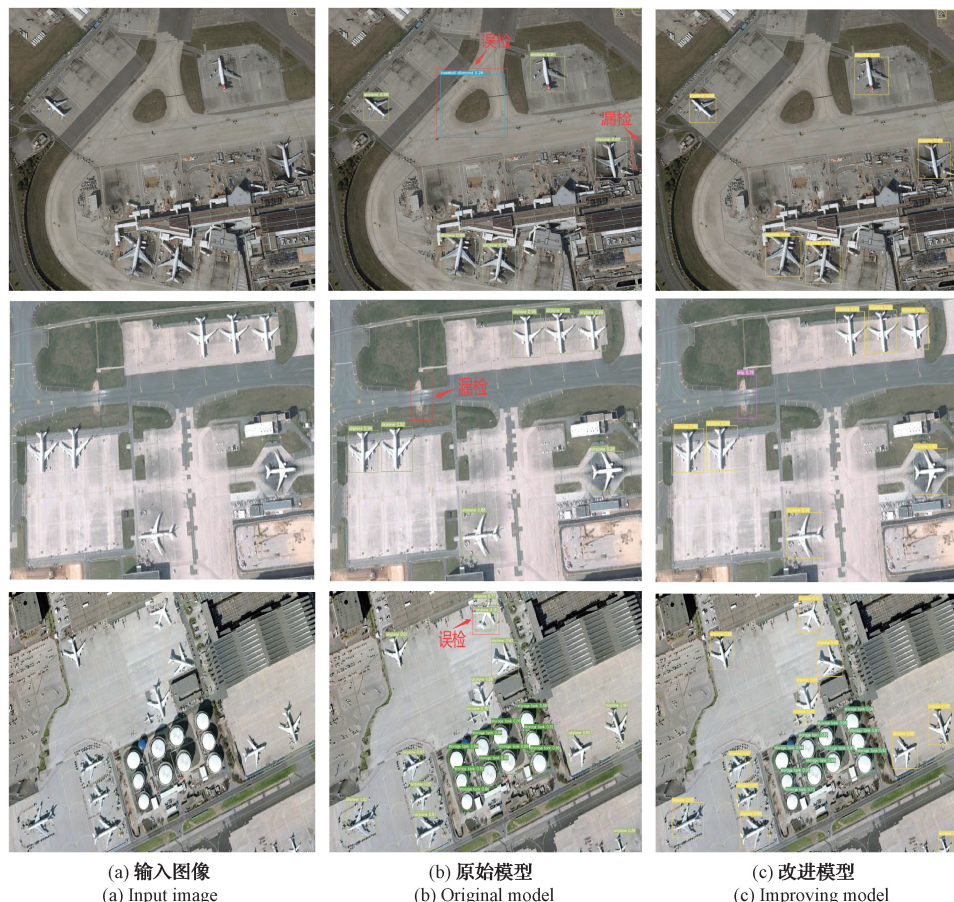


图 6 NWPU VHR-10 数据集模型改进前后检测效果对比

Fig. 6 Comparison of detection performance before and after improvement of NWPU VHR-10 dataset model

3 结论

为了解决遥感数据集中小目标检测的挑战,本文提出了一种改进的 YOLOv7-tiny 算法。首先在骨干网络中引入了 EMA 注意力机制,设计了 ELAN-EMA 模块,这有效地提升了网络的多尺度特征提取能力。其次为了应对深层网络中小目标特征难以提取的问题,采用了 CARAFE 上采样算子优化模型 FPN 中的上采样方法,设计了 FPN-CARAFE 结构,从而有效地扩大了模型对小目标检测的感受野,强化了模型对小目标特征的提取能力。最后针对传统的 CIoU 损失函数对小目标位置偏移敏感的问题,采用了 CIoU-NWD 加权损失函数,这有效地提升了小目标的检测效果。通过对比实验看出,相比于基线模型,改进后的模型在 RSOD 数据集相比基线模型在参数量和平均每帧检测时间略微增加的情况下,mAP@0.5 有着明显的提升,提高了 3.6%,同样在 NWPU VHR-10 数据集上,参数量和平均每帧检测时间略微增加的情况下,mAP@0.5 提高了 1.9%,充分证明了本文所提出的改进模型能够提升遥感数据集中小目标的检测精度,满足部署在遥感检测系统上的要求。

参考文献

- [1] 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. Research on improving remote sensing image detection with YOLOv5s [J]. Foreign Electronic Measurement Technology, 2022, 41(11):57-66.
- [2] 王悦伟,焦良葆,高阳.基于改进 YOLOv5s 的遥感目标检测[J].计算机测量与控制,2023,31(8):70-76.
WANG Y W, JIAO L B, GAO Y. Remote sensing object detection based on improved YOLOv5s [J]. Computer Measurement and Control, 2023, 31(8):70-76.
- [3] 吴建成,郭荣佐,成嘉伟,等.注意力特征融合的快速遥感图像目标检测算法[J].计算机工程与应用,2024,60(1):207-216.
WU J CH, GUO R Z, CHENG J W, et al. Fast remote sensing image object detection algorithm based on attention feature fusion [J]. Computer Engineering and Applications, 2024, 60(1):207-216.
- [4] 刘晶宇,杨鹏.基于 YOLO-V5 改进的遥感图像目标检测[J].计算机时代,2023(7):50-55.
LIU J Y, YANG P. Remote sensing image object detection based on improved YOLO-V5 [J]. Computer Era, 2023(7):50-55.
- [5] 王凯,王伟,蒋志伟.基于改进 YOLO 模型的遥感小目标检测[J].电光与控制,2023,30(6):60-64.
WANG K, WANG W, JIANG ZH W. Detection of small remote sensing targets based on improved YOLO model [J]. Optoelectronics and Control, 2023, 30(6):60-64.
- [6] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]. IEEE International Conference on Computer Vision, 2021: 3500-3509.
- [7] LI J N, LIANG X D, SHEN S M, et al. Scale-aware Fast R-CNN for pedestrian detection [J]. IEEE Transactions on Multimedia, 2018,20(4):985-996.
- [8] 张杰.基于改进 Faster-RCNN 的小目标检测[J].现代计算机,2023,29(14):14-18.
ZHANG J. Small object detection based on improved Faster-RCNN [J]. Modern Computer, 2023, 29(14):14-18.
- [9] CHAO Z, XIN S, XI Y, et al. RDD-YOLO: A modified YOLO for detection of steel surface defects[J]. Journal of the International Measurement Confederation, 2023, 214:112776-112776.
- [10] SEHWAG V, CHIANG M, MITTAL P. SSD: A unified framework for self-supervised outlier detection[C]. International Conference on Learning Representations, 2021.
- [11] 王恒涛,张上,陈想,等.轻量化无人机航拍目标检测算法[J].电子测量技术,2022,45(19):167-174.
WANG H T, ZHANG SH, CHEN X, et al. Lightweight target detection algorithm for UAV aerial photography [J]. Electronic Measurement Technology, 2022, 45(19):167-174.
- [12] 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,2022,43(3):221-229.
YAN J H, ZHANG K, SHI T J, et al. Detection of weak ground targets in remote sensing images by fusion of multilevel features [J]. Chinese Journal of Scientific Instrument, 2022, 43(3):221-229.
- [13] 陈范凯,李士心.改进 YOLOv5 的无人机目标检测算法[J].计算机工程与应用,2023,59(18):218-225.
CHEN F K, LI SH X. Improved UAV object detection algorithm based on YOLOv5 [J]. Computer Engineering and Applications, 2023, 59(18):218-225.
- [14] 单慧琳,吕宗奎,付相为,等.基于 DFECANet 的遥感图像飞机目标检测方法[J].电子测量与仪器学报,2024,38(2):19-29.
SHAN H L, LYU Z K, FU X W, et al. Aircraft object detection method in remote sensing images based on DFECANet [J]. Journal of Electronic

- Measurement and Instrumentation, 2024, 38(2): 19-29.
- [15] 李博文,张宏帅,赵华东,等. 基于小样本数据驱动模型的硅片线切割质量预测[J]. 机床与液压, 2024, 52(1): 66-73.
LI B W, ZHANG H SH, ZHAO H D, et al. Quality prediction of silicon wafer cutting based on small sample data-driven model [J]. Machine Tool & Hydraulics, 2024, 52(1): 66-73.
- [16] WANG K, FANG B, QIAN J Y, et al. Perspective transformation data augmentation for object detection[J]. IEEE Access, 2020, 8: 4935-4943.
- [17] 曾溢良,张浩,吕志武. 基于多尺度语义的目标检测方法[J]. 计算机工程与设计, 2024, 45(1): 252-260.
ZENG Y L, ZHANG H, LYU ZH W. Object detection method based on multi-scale semantics [J]. Computer Engineering and Design, 2024, 45(1): 252-260.
- [18] 张朝阳,张上,王恒涛,等. 多尺度下遥感小目标多头注意力检测[J]. 计算机工程与应用, 2023, 59(8): 227-238.
ZHANG CH Y, ZHANG SH, WANG H T, et al. Multi-head attention detection of small remote sensing targets at multiple scales [J]. Computer Engineering and Applications, 2023, 59(8): 227-238.
- [19] 张寅,朱桂熠,施天俊,等. 基于特征融合与注意力的遥感图像小目标检测[J]. 光学学报, 2022, 42(24): 140-150.
ZHANG Y, ZHU G Y, SHI T J, et al. Small target detection in remote sensing images based on feature fusion and attention [J]. Acta Optica Sinica, 2022, 42(24): 140-150.
- [20] WANG Z, GUO J X, ZHANG C L, et al. Multiscale feature enhancement network for salient object detection in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-19.
- [21] LIU G, HAN J, RONG W Z. Feedback-driven loss function for small object detection[J]. Image and Vision Computing, 2021, 111: 104197.
- [22] DENG C, WANG M, LIU L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.
- [23] 张华卫,张文飞,蒋占军,等. 引入上下文信息和 Attention Gate 的 GUS-YOLO 遥感目标检测算法[J]. 计算机科学与探索, 2024, 18(2): 453-464.
ZHANG H W, ZHANG W F, JIANG ZH J, et al. GUS-YOLO remote sensing object detection algorithm introducing context information and attention gate [J]. Journal of Computer Science and Exploration, 2024, 18(2): 453-464.
- [24] 何秋润,胡节,彭博,等. 基于上下文信息的多尺度特征融合织物疵点检测算法[J/OL]. 计算机应用, 1-9 [2024-06-12]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240422.1655.010.html>.
- HE Q R, HU J, PENG B, et al. Fabric defect detection algorithm based on multi-scale feature fusion with context information [J/OL]. Computer Applications, 1-9 [2024-06-12]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240422.1655.010.html>.
- [25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 7464-7475.
- [26] 周华平,张杰. 融合上下文信息和注意力的遥感小目标检测[J]. 吉林师范大学学报(自然科学版), 2024, 45(1): 117-125.
ZHOU H P, ZHANG J. Fusion of context information and attention for small remote sensing target detection [J]. Journal of Jilin Normal University (Natural Science Edition), 2024, 45(1): 117-125.
- [27] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 2011-2023.
- [28] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Computer Vision and Pattern Recognition, 2020: 11531-11539.
- [29] OUYANG D L, HE S, ZHANG J, et al. Efficient multi-scale attention module with cross-spatial learning [J]. 2023 International Conference on Acoustics, Speech, and Signal Processing, 2023.
- [30] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-aware reassembly of features [C]. IEEE International Conference on Computer Vision, 2019: 3007-3016.
- [31] WANG J W, XU C, YANG W, et al. A normalized Gaussian wasserstein distance for tiny object detection[J]. ArXiv preprint, arXiv:2110.13389, 2021.
- [32] YANG L, YUAN G W, ZHOU H, et al. RS-YOLOX: A High-precision detector for object detection in satellite remote sensing images [J]. Applied Sciences-Basel, 2022, 12(17): 8707.
- [33] TIAN Z W, HUANG J, YANG Y, et al. KCFS-YOLOv5: A high-precision detection method for

object detection in aerial remote sensing images[J].
Applied Sciences-Basel, 2023, 13(1): 649-649.

作者简介

王子龙(通信作者), 硕士研究生, 主要研究方向为计

算机视觉、目标检测。

E-mail:wzilong007@163.com

荣杰, 硕士研究生, 主要研究方向为计算机视觉、目标检测。

E-mail:rj1617121671@163.com