

基于混合注意力和动态采样的遥感图像目标检测^①



蔡清, 王净雨, 梁宏涛

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 梁宏涛, E-mail: lht@qust.edu.cn

摘要: 针对遥感图像目标检测中的复杂背景干扰, 目标密集堆积导致遥感图像目标检测模型效果差的问题, 通过对 YOLOv5s 目标检测模型进行改进. 首先使用混合注意力机制对 CBAM (convolutional block attention module) 加以改进并添加至骨干网络中, 使得模型所提取到的特征含有局部信息和全局信息, 以加强模型识别复杂背景下目标的能力; 其次使用超轻量级的动态上采样器 DySample 结构, 减少模型参数并提高检测效率, 最后使用 EIoU 损失函数提高对于待检测目标的定位水平. 在 RSOD 和 DIOR 数据集上进行了实验验证, 结果表明, 改进后的 YOLOv5s 在遥感图像目标检测中的准确度要比原始模型高 7.8%, 同时能够满足遥感图像实时目标检测的需求; 此外与其他目标检测模型相比, 改进模型也能保有一定优势.

关键词: 机器学习; 损失函数; 遥感图像; 目标检测; 注意力机制

引用格式: 蔡清, 王净雨, 梁宏涛. 基于混合注意力和动态采样的遥感图像目标检测. 计算机系统应用, 2025, 34(3): 171-179. <http://www.c-s-a.org.cn/1003-3254/9795.html>

Remote Sensing Image Target Detection Based on Hybrid Attention and Dynamic Sampling

CAI Qing, WANG Jing-Yu, LIANG Hong-Tao

(School of Information Science and Technology, Qingdao University of Science & Technology, Qingdao 266061, China)

Abstract: Ineffective object recognition models occur in remote sensing images through complex background interference and dense target integration. To this end, this study improves the YOLOv5s object output model. First, a mixed attention menu is utilized to improve the convolutional attention model (CBAM) and add it to backbone networks. Accordingly, the extracted features of the model contain local and global information to enhance the model's ability to identify targets in complex backgrounds. Then the study uses the ultra-light sampler DySample to reduce model parameters and improve model performance. Finally, the study employs the EIoU loss function to improve the positioning level of the target to be detected. Experimental verification of RSOD and DIOR data sets shows that the improved YOLOv5s has a 7.8% higher accuracy than the original model in detecting targets in remote sensing images, meeting the real-time detection requirements of targets in remote sensing images. In addition, the improved model retains the advantages it has in comparison to other object recognition models.

Key words: machine learning; loss function; remote sensing image; target detection; attention mechanism

近年来, 遥感技术进入了一个全新的阶段, 随着高质量遥感图像的不断增多, 遥感数据已不再是遥感图像分析研究和应用的瓶颈, 怎样从海量的遥感数据中准确获取有价值的信息已成为亟待解决的问题^[1].

基于深度学习的目标检测算法正在遥感图像处理中表现优异, 在军事和民用领域都有着重要的应用^[2,3]. 在军事领域, 可以及时发现敌方目标, 提高战场感知能力, 为军事战略的制定提供有利信息. 在民用领域,

① 收稿时间: 2024-09-04; 修改时间: 2024-09-24; 采用时间: 2024-10-14; csa 在线出版时间: 2025-01-21
CNKI 网络首发时间: 2025-01-22

可以监测交通状况,分析路况,缓解交通压力.因此,目标检测技术在遥感图像分析中具有重要的意义和价值.

与自然图像相比,遥感图像有自己的特殊性:图像尺寸大,成像视角高,拍摄范围广,内容信息丰富,微小物体分布密集,目标细节丢失多,背景区域复杂程度高等问题^[4],然而,传统的遥感目标检测方法依赖于人为设计的特征表达,对物体的大小和形变比较敏感,鲁棒性差,泛化能力弱;其次,大多数目标检测方法都是针对自然场景设计的,如果应用于微小物体分布密集、与背景差异小的遥感图像,容易出现漏检甚至错检等问题.因此,遥感图像目标检测仍然是一项具有挑战性的任务^[5].

相较于传统检测方法难以应对遥感图像目标检测,深度学习方法则有着良好的适应性、鲁棒性以及准确性.汪鹏等^[6]以 Faster R-CNN 为基础进行调整,通过使用特征自适应网络和上下文特征金字塔来提高模型对于特征的提取能力,检测精度有一定的提升.张绍文等^[7]将二维余弦变化与注意力机制融合,在加权多重感受野金字塔池化结构与跨层特征融合相互作用下,对于遥感图像小目标检测方面效果优异.宋爽爽等^[8]在 YOLOv5 的基础上采用轻量化 GhostConv 替换原有卷积模块,并与端到端坐标注意力结合,有效减少计算参数,尽可能保证检测精度.周沁坤等^[9]通过加强特征同时提出 DCA 模块使网络保留有效的空间信息,对于航空遥感小目标的检测效果有明显改善.孙文赞等^[10]提出一个全局上下文增强网络提高网络的表征能力,引入层内调节方法,同时重设特征融合金字塔网络,实验结果较为优秀.

尽管上述方法在遥感图像目标检测中有效地提升了对小目标的检测性能,但由于未能同时考虑全局-局部信息,导致对多尺度和困难目标的遗漏丢失.由此,本文提出了一种基于 YOLOv5s^[11]的遥感图像目标检测算法.该方法首先将融合了混合注意力的卷积注意力机制模块 CBAM-MLCA^[12]添加到网络骨干,引入通道信息、空间信息、局部信息和全局信息,增强网络提取特征的能力,使网络更关注于感兴趣区域,降低复杂背景的干扰;其次使用 DySample^[13]模块替换原有的上采样模块,减少运算量,提高模型的运算效率以及精确度;最后将损失函数替换为 EIoU^[14],在处理边界框的回归更精确的同时提高检测速度.

1 改进的 YOLOv5s 目标检测模型

YOLO 系列算法作为单阶段目标检测算法,无需对目标进行预选,能够直接由单个神经网络完成目标的分类和定位. YOLO 系列算法利用 CNN 将输入图像按网格划分为小块区域,若检测目标中心点落入某网格,则该网格负责预测此目标的边界框和置信度,再通过非极大值抑制完成筛选. YOLOv5 作为该系列算法之一,网络模型与算法思想受到普遍关注与广泛应用. YOLOv5s 轻量化程度高、检测速度快,适于作为遥感目标检测的基线算法.

本文改进模型在 YOLOv5s 的骨干部分增设了含有混合注意力的卷积注意力模块 CBAM-MLCA,能够提取含有全局信息、局部信息、空间信息以及通道信息的综合特征,以提高网络对于感兴趣区域的注意程度,之后使用 DySample 采样模块以降低由于注意力模块的引进所增加的复杂度,最后使用 EIoU 损失函数替代原有损失函数,改进后的模型结构如图 1 所示.

1.1 注意力模块 CBAM-MLCA

近几年,深度学习领域广泛运用了注意力机制,本文为减少对无关通道的关注、增强重要的特征通道、降低背景噪声和干扰、提高鲁棒性和适应性,通过对 CBAM 注意力模块加以改进,提出了一个同时含有通道信息、空间信息、局部信息以及全局信息的改进注意力机制模块 CBAM-MLCA.

传统的卷积注意力 CBAM 结构能够加强网络对于图像当中关键信息的注意程度,同时抑制背景干扰.接下来简要介绍一下 CBAM 的工作原理:首先是通道注意力模块 (channel attention module, CAM),输入特征 $F \in \mathbb{R}^{C \times H \times W}$ 会先通过 CAM 的最大池化以及平均池化压缩,以此能够获得对应的两个通道信息,之后这两个通道信息特征 $F_{\max}^c, F_{\text{avg}}^c \in \mathbb{R}^{C \times 1 \times 1}$ 将经过一个 3 层全连接的多层感知机,将得到的特征进行加和,然后经 Sigmoid 函数处理得到通道注意力权重 $M_c \in \mathbb{R}^{C \times 1 \times 1}$,计算过程如式 (1) 所示:

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\max}^c))) \end{aligned} \quad (1)$$

其中, σ 为 Sigmoid 激活函数, W_0 和 W_1 为 3 层全连接网络中的权重.

将通道注意力权重 M_c 与输入特征 F 进行内积运算后就能获得与通道注意力进行融合的特征映射 $F_1 \in$

$\mathbb{R}^{1 \times H \times W}$, 并视作空间注意力模块 (special attention module, SAM) 的输入, 经过 SAM 中最大池化和平均池化后得到对应的两个空间信息特征 $F_{\max}^s, F_{\text{avg}}^s \in \mathbb{R}^{1 \times H \times W}$; 将两个特征沿通道拼接, 并通过 7×7 卷积运算后由 Sigmoid 激活获得空间注意力权重 $M_s \in \mathbb{R}^{1 \times H \times W}$, 计算过程如式 (2) 所示:

$$M_s(F_1) = \sigma \left(f^{7 \times 7} \left(\begin{bmatrix} \text{AvgPool}(F); \\ \text{MaxPool}(F) \end{bmatrix} \right) \right) = \sigma \left(f^{7 \times 7} \left([F_{\text{avg}}^s; F_{\max}^s] \right) \right) \quad (2)$$

其中, $f^{7 \times 7}$ 为 7×7 的卷积运算. 将空间注意力权重 M_s 与特征映射 F_1 进行内积运算, 就能获得含有输入特征 $F \in \mathbb{R}^{C \times H \times W}$ 中的空间信息和通道信息的特征图.

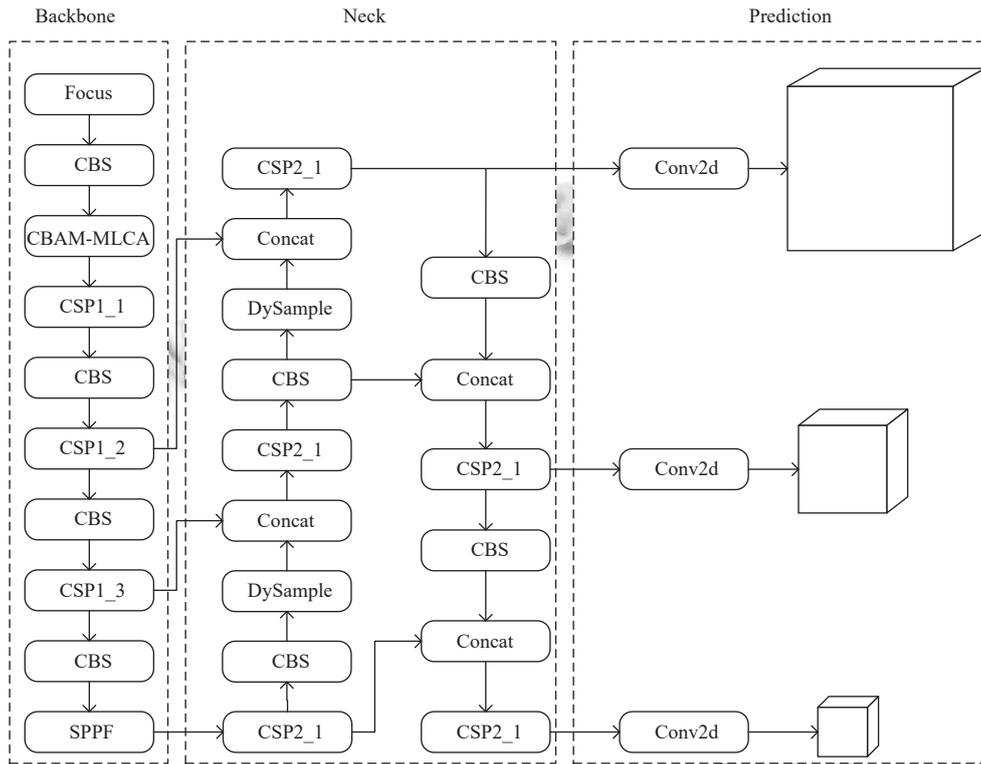


图 1 改进后的模型结构

CBAM 是基于全局信息的注意力机制, 即是基于整个特征提取过程得到的特征信息或整幅图像的特征信息, 其计算过程中需要大量参数. 所以, 近年来, 许多学者开始研究基于局部特征信息的注意力机制, 而局部信息是对特征或图像特定区域的特征提取, 最终得到的信息是若干局部信息的融合. 学者们还通过实验证明此类注意力机制同样可以取得预期效果^[15-17].

文献[18]研究了 CBAM 机制的不足, 认为在 CBAM 处理特征图的过程中, 使用全局平均池化对输入特征进行压缩是冗余操作, 而且会导致部分特征信息损失, 在全局平均池化处理过程中, 特定局部区域的权重分布趋于均匀, 会忽略掉其他有用的特征信息, 从而丢失重要的局部区域特征信息. 尽管最大池化会对特征信息进行一定程度的补充, 但无法弥补重点区域的特征

信息.

受文献[18]启发, 本文在 CBAM 的基础上增加了能够提取到全局信息和局部信息的分支 MLCA 部分, 将 CBAM 改进为 CBAM-MLCA, 模块结构如图 2 所示.

经过改进的 CBAM-MLCA 能够在压缩特征的过程中引入局部信息和全局信息, 避免因全局平均池化导致的重点区域特征缺失, 提高网络鲁棒性, 加强网络对于图像中待检测目标的关注程度, 同时减少特征提取过程中的冗余操作.

MLCA 的具体实现方式如图 3 所示.

首先对于输入特征向量进行全局范围的局部平均池化, 原输入被转化为 $1 \times C \times k_s \times k_s$ 的向量, 其中 k_s 表示在 W 或 H 维度进行局部平均池化的分块数量, 影响全局局部信息的提取, 为了减少计算规模和参数数量,

本文将 ks 设置为 5, 以便通过第 1 个局部池化提取局部空间信息. 在此基础上, 利用两个分支将输入转换为一维向量, 第 1 个分支包含全局信息, 第 2 个分支包含局部信息. 经过一维卷积后, 全局信息可以直接通过反平均池化获得 $1 \times C \times ks \times ks$ 的向量, 局部信息则需要通

过 Reshape 操作获得 $1 \times C \times ks \times ks$ 的向量, 以此保证所获取的局部信息有效性, 将两个分支所提取的特征进行按位加和, 并再次通过反平均池化恢复至输入特征的尺寸, 之后与输入特征进行信息融合, 达到混合注意力的目的.

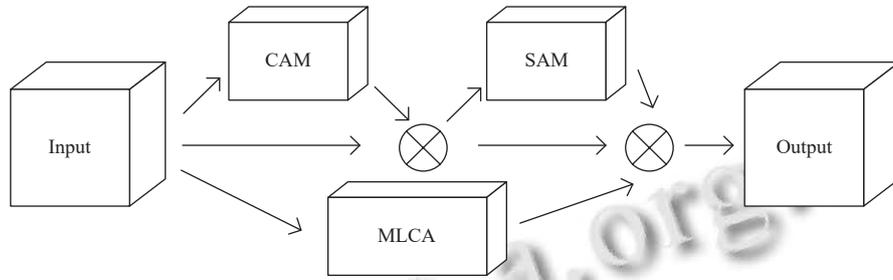


图2 改进后的 CBAM-MLCA 模块

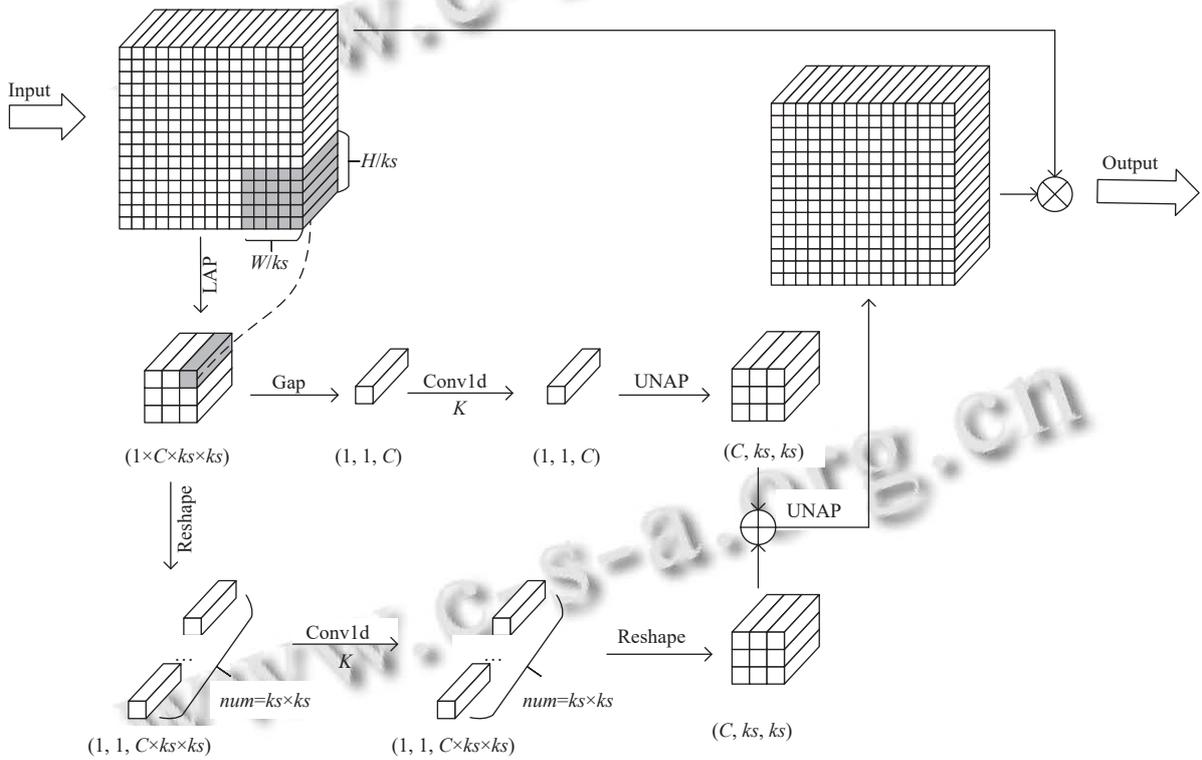


图3 MLCA 实现原理

分支中一维卷积用 Conv1d 表示, 卷积核大小 k 与通道维度 C 成正比, 表明在捕获局部跨通道交互信息时仅考虑每个通道与其 k 个相邻通道之间的关系. k 的选择由式 (3)^[19] 确定.

$$k = \Phi(C) = \left\lfloor \frac{\log_2(C)}{r} + \frac{b}{r} \right\rfloor_{\text{odd}} \quad (3)$$

其中, C 是通道数, k 是卷积核的大小, r 和 b 都是超参数,

默认值为 2, odd 表示 k 只能是奇数, 如果 k 是偶数, 则加 1.

1.2 DySample 上采样结构

上采样操作在 CNN 框架中至关重要, 尤其是在特征融合过程中, 用于强化骨干网络所提取的特征. 就 YOLOv5s 而言, 其特征融合网络藉由最近邻插值方式实现上采样, 但此方式局限于子像素的局部信息, 未能

充分利用像素间的全局关联,从而丢失了图像中像素点的渐变关系,无法满足密集预测任务对丰富语义信息的需求.

对于主干网络中 CBAM-MLCA 模块的加入,模型检测精度固然提升,随之而来的是参数量和运算时间的增加,为了能够更好地捕捉图像中的细节和语义信息,减少上采样过程中可能导致的信息损失,降低模型参数量和复杂程度,本文在采样方式上选用具有更大的视野、拥有内容感知处理、轻量级和快速计算属性的 DySample 模块^[13],使得改进模型在处理遥感图像目标检测任务中表现更为优秀.

DySample 由点采样的角度设计而来,是一个快速有效且通用的上采样结构,主要由采样点生成器和网格采样两部分组成.首先输入特征经过采样点生成器

获得采样集,之后通过网格采样函数对输入特征进行重新采样.对于采样点生成器来说,采样集是生成的偏移量和原始网格位置之和,其中偏移量由线性层生成;由此得到的上采样特征可以加强对于局部区域的关注,获得更为丰富的语义信息. DySample 过程可用公式表示为:

$$X' = \text{grid_sample}(X, S) \quad (4)$$

其中, X' 为采样后的特征, X 为原始采样特征, S 为采样集, grid_sample 过程为 PyTorch 中的网格采样方法; 采样集 S 为偏移量 O 与原始网格 G 的和:

$$S = O + G \quad (5)$$

$$O = \text{linear}(X) \quad (6)$$

DySample 上采样算子结构如图 4 所示.

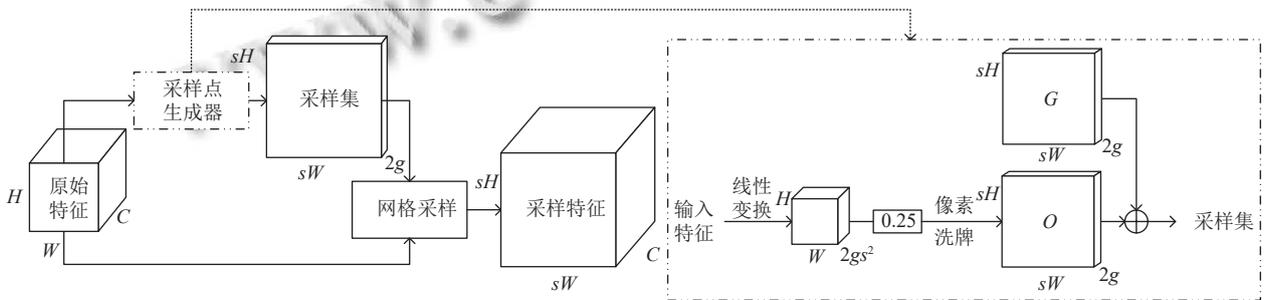


图 4 DySample 上采样结构

1.3 损失函数改进

在深度学习模型中,损失函数发挥着极其关键的作用. YOLOv5s 的损失函数可分为分类损失,置信度损失以及边界框损失这 3 部分. 其中前两部分损失函数需要借助 BCE Loss 来实现的,其公式如下:

$$BCE\ Loss = \begin{cases} -\frac{1}{N} \times \sum_{n=1}^N \log(1 - x_n), & y = 0 \\ -\frac{1}{N} \times \sum_{n=1}^N \log x_n, & y = 1 \end{cases} \quad (7)$$

其中, x_n 表示由 Sigmoid 函数获得的预测置信度或目标概率, N 为样本总量, y 表示样本的真实类别.

目标检测的损失函数一般使用 IoU 损失函数,包括 IoU Loss、GIoU Loss、CIoU Loss 等.

1) IoU Loss

IoU 是指相交率,能够表示预测的检测框和标签框的检测效果. IoU Loss 的计算公式如下:

$$IoU\ Loss = 1 - IoU = 1 - \frac{(A \cap B)}{(A \cup B)} \quad (8)$$

其中, A 、 B 分别表示检测框和标签框.

2) GIoU Loss

GIoU Loss 引入了检测框和标签框的最小外接矩形,这样 GIoU 能够不仅关注相交区域,非相交区域也会有所关注. 设 B^{gt} 为标签框, B 为预测的检测框, C 为检测框 A 和标签框 B 的最小外接矩形,则 GIoU Loss 的计算公式如下:

$$GIoU\ Loss = 1 - GIoU = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|} \quad (9)$$

GIoU 能够让训练在检测框和标签框在不相交时也能进行下去,但是在检测框和标签框重叠时,无法区分检测框和标签框之间的位置关系.

3) CIoU Loss

CIoU Loss 函数为 YOLOv5s 在训练过程所采用的损失函数,除了检测框和标签框的覆盖面积外,还考虑到了中心点距离和长宽比,有着更强的鲁棒性且能够更准确地衡量框与框之间的相似程度. CIoU Loss 的计算公式如下:

$$CIoU Loss = 1 - CIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (10)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (11)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (12)$$

在式(10)–式(12)中, w , h 分别代表检测框的宽和高, w^{gt} , h^{gt} 则各自对应为标签框的宽和高. b 为检测框的中心点坐标, b^{gt} 对应为标签框的中心点坐标, c^2 代表两框最小外接矩形的对角线距离, α 是权重系数, v 用于描述纵横比.

4) *EIoU Loss*

为了进一步提高模型对于遥感图像的检测效果, 本文将使用效率更高的 *EIoU Loss*. *EIoU Loss* 的计算公式如下:

$$EIoU Loss = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c^2} + \frac{\rho^2(h, h^{gt})}{c^2} \quad (13)$$

EIoU 在 *CIoU* 的基础上将纵横比拆开, 分为具体的宽高损失, 更为准确的表示 3 个几何因素之间的差异, 在保留 *CIoU Loss* 效果的同时, 使得预测框和标签框的宽高差异最小化, 从而在更快收敛的同时有着更精准的定位效果.

2 实验环境与评价指标

2.1 实验环境与参数设置

本文在 Windows 11 系统、Python 版本 3.9 情况下进行实验训练, 具体环境信息如表 1 所示.

表 1 实验参数配置

名称	配置
CPU	Intel Core i7-13650HX
GPU	NVIDIA RTX 4070 (8 GB)
运行内存	32 GB
深度学习环境	PyTorch 2.0.1
CUDA版本	CUDA 11.7

主要训练参数具体包括: batch-size 为 4, 迭代次数 epoch 为 100, 训练图像输入尺寸为 640×640, 动量初始值设置为 0.937, 初始学习率设置为 0.01, 模型训练优化器选择为 SGD.

2.2 实验数据集

本文选取 RSOD^[20]数据集和 DIOR^[21]数据集两个

数据集为实验检测基准.

RSOD 数据集由飞机、油箱、操场和立交桥 4 个类别共 936 张有标签标注图片以及 40 张背景图片组成, 数据集共有 6 950 个标注实例, 本文实验按照 7:2:1 随机划分, 分别用作训练验证以及测试过程.

DIOR 是一个用于目标检测和图像分割的数据集, 其中含有 23 463 张高分辨率遥感图像, 共计 20 个不同的目标种类和 192 472 标注实例, 本文选取其中 5 862 张用作训练, 5 863 张用作验证, 11 738 张用作测试集.

2.3 评价指标

1) 精准率 (Precision, *P*)

精确率定义为分类预测为正的样本中实际为正的样本比率. 如公式所示:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

其中, *TP* (true positive) 为真正, 即实际为正被分类预测为正的样本数量. *FP* (false positive) 为假正, 即实际为负却被分类预测为正的样本数量.

2) 召回率 (Recall, *R*)

召回率定义为实际为正的样本中被分类预测为正的样本比率. 如公式表示:

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

其中, *FN* (false negative) 为假负, 即实际为正却被分类预测为负的样本数量.

3) 平均精度均值 (mean average precision, *mAP*)

平均精度均值 *mAP* 表示各类目标的平均精度的算术平均值. 如公式所示:

$$mAP = \frac{\sum_0^N AP_n}{N} \quad (16)$$

其中, AP_n 表示第 n 类目标的平均分类精确率, 其数值等于 *Precision-Recall* 函数与坐标轴所覆盖的面积.

对于多目标分类, 各类别目标的分类精度 AP_n 用其平均值 *mAP* 表示. *IoU* 参数阈值为 0.5 时的平均精度均值表示为 *mAP@0.5*; *IoU* 阈值分别取 0.5、0.55、0.6、…、0.95 时对应的 *mAP* 值, $mAP = \sum_0^N AP_n / N$ 取平均值得到 *mAP@0.5 : 95*. 此外, 为了验证改进前后模型的大小、计算量和推理时间, 综合判别模型性能, 本文引入参数量 (params) 衡量模型的大小; 浮点运算数 (FLOPs) 衡量模型的复杂度; 推理速度 (inference time) 来评估

检测效率。

3 实验结果分析

3.1 消融实验

为了验证 CBAM-MLCA 层、DySample 上采样层以及损失函数 $EIoU$ 对于改进模型的性能影响,我们在 RSOD 数据集上进行了模型组合训练与测试,结果如表 2 所示。当 YOLOv5s 增加 CBAM-MLCA 层后,模型参数量 $params$ 提升 1.74M,模型复杂程度也有所

提高, $mAP@0.5$ 提高了 0.8%,测试集中各目标类别的检测召回率 R 提高了 1.8%,推理时间增加了 0.7 ms。说明模型增加 CBAM-MLCA 层能够提高模型对于待检测目标的关注程度,但会增加模型规模;当模型改进 DySample 层后,模型参数量和复杂程度均有所下降, $mAP@0.5$ 提高了 1.7%,同时推理时间减少了 0.1 ms,对于模型的检测精准度提升较为明显;最后是 $EIoU$ 损失的增加,不仅不会降低模型的检测精度,同时也减少模型的参数量和复杂程度。

表 2 改进模型在 RSOD 数据集上的消融实验

CBAM-MLCA	DySample	$EIoU$	Params (M)	GFLOPs	P (%)	R (%)	$mAP@0.5$ (%)	Inference time (ms)
—	—	—	7.20	16.5	94.6	92.0	95.3	6.8
√	—	—	8.94	17.7	92.9	93.8	97.1	7.5
—	√	—	7.04	15.9	94.5	96.1	96.0	6.7
—	—	√	7.01	15.4	95.7	94.4	96.3	6.9
√	√	—	8.03	17.3	91.7	96.3	97.2	7.8
√	—	√	8.26	16.8	93.5	96.7	97.0	7.5
—	√	√	6.50	15.6	95.6	94.7	96.5	6.8
√	√	√	7.47	16.9	95.5	96.4	98.1	7.1

当同时改进 CBAM-MLCA、DySample、 $EIoU$ 后,能够使得模型精准率 P 提高 0.9%,召回率 R 提高 4.4%,提升较为明显; $mAP@0.5$ 提高了 2.8%,参数量和复杂程度有小幅度上升,推理时间仅增加 0.3 ms。说明本模型对于遥感图像目标检测方面的精度有着明显的性能提升,同时模型的参数量,复杂程度以及推理时间会有轻微的增加。

消融实验结果表明,注意力模块的加入、采样方式的替换、损失函数的改进三者相结合,对于模型 mAP , AP 的提升有一定的帮助。虽然注意力模块会小幅度的增加网络参数量和检测速度,但是对于检测精准度有巨大提升;新型轻量化的采样结构替换和损失函数的改进能够在保证检测精准度的情况下有效降低参数量;最终改进模型的参数量和检测速度较原模型比未有提升,但并不会对模型的轻量化和实时性产生负面效果,以此换取检测精准度的提高是合理的。

3.2 与其他模型测试对比实验

为了量化分析本文改进模型的检测性能,我们分别在 RSOD 数据集和 DIOR 数据集上,与 YOLOv8-S, YOLOv7^[22], M2Det^[23], ARST-YOLOv7^[9], YOLOX-S^[24], CWFY-YOLOX^[7], Faster R-CNN, IF-RCNN^[6], FCOS^[25], YOLOv5s, GC-YOLOv5^[8], 共 11 种检测模型进行对比分析,其中包括测试结果如表 3 及表 4 所示。

在 RSOD 数据集上的测试结果表明,本文的改进模型 mAP 值较最新检测模型 YOLOv8-S 要高出 0.74%,虽然 AP_S 值略有不足,但是有着更少的参数量,相较于其他检测模型与相对应改进模型,本文改进模型能够在保证检测精度的情况下拥有较低的参数量。

DIOR 数据集含有 20 个类别,将飞机、烟囱、港口、立交桥、存储罐、船、车辆和风车这 8 个类别的检测结果显示。其中改进模型对于飞机、烟囱、港口、船、存储罐、车辆这类目标的准确率远高于其他模型,且本文改进模型 mAP 值相较于 YOLOv8-S 和 ARST-YOLOv7 分别高出 0.69% 和 1.87%。虽然 AP_S 较 YOLOv8-S、CWFY-YOLOX 稍有不足,但是综合考虑模型检测精准度,参数量方面本文改进模型仍有优势,更适于遥感图像目标检测。

3.3 模型检测效果对比

为了比较原模型 YOLOv5s 和改进模型在目标检测方面的表现,我们在同一组图像上展示了两种模型实验结果的对比示意图,如图 5 所示,其中图 5(a) 为 YOLOv5s 原模型的检测结果,图 5(b) 为改进模型的检测效果。可以明显观察到,相较于 YOLOv5s 原模型在检测过程中容易出现的漏检、误检情况;本文改进模型在这方面有明显改善,且精准度有所提高,对于在遥感图像中进行目标检测有很大的帮助。

表3 不同检测方法在RSOD数据集的性能

模型	AP (%)				mAP (%)	AP _S (%)	Params (M)
	aircraft	oiltank	overpass	playground			
YOLOv8-S	96.02	98.38	86.92	100	97.36	45.6	11.2
YOLOv7	97.36	98.42	84.62	100	95.10	41.7	36.90
ARST-YOLOv7	96.82	97.91	86.45	99.68	97.30	42.9	78.80
YOLOX-S	96.37	98.19	83.78	100	94.58	40.2	8.93
CWAF-YOLOX	97.01	98.42	90.39	100	97.75	45.3	9.94
YOLOv5s	95.74	98.18	80.32	99.28	95.30	37.3	7.20
GC-YOLOv5	95.37	97.56	82.61	100	92.30	41.0	11.37
Faster R-CNN	81.78	97.69	93.28	100	93.19	11.6	136.71
IF-RCNN	84.49	95.78	92.61	99	89.64	37.7	145.48
M2Det	88.11	97.87	94.01	100	95.00	14.7	86.50
FCOS	93.95	98.81	85.06	100	94.45	31.4	51.0
Ours	98.60	98.49	95.91	99.40	98.10	45.1	7.47

表4 不同检测方法在DIOR数据集的性能 (%)

模型	P								mAP	AP _S
	airplane	chimney	harbor	overpass	ship	storagetank	vehicle	windmill		
YOLOv8-S	91.24	85.92	70.15	69.01	93.61	80.49	77.27	84.74	76.49	13.7
YOLOv7	90.57	78.31	65.64	61.84	91.04	79.98	56.03	82.03	72.76	11.7
ARST-YOLOv7	91.56	82.81	71.34	65.21	91.85	82.12	73.46	84.81	75.31	12.6
YOLOX-S	83.49	76.83	63.19	58.59	89.82	72.93	52.74	78.79	70.91	11.2
CWAF-YOLOX	84.92	81.67	70.56	59.73	89.23	78.71	77.21	81.23	73.98	13.6
YOLOv5s	86.55	77.91	59.46	57.98	89.30	75.24	52.23	75.10	69.23	11.9
GC-YOLOv5	90.23	70.39	56.89	57.82	87.49	76.48	53.89	75.67	69.52	10.5
Faster R-CNN	58.29	76.27	54.80	53.52	27.13	33.91	17.04	59.65	57.94	1.3
IF-RCNN	68.47	78.45	69.42	67.21	84.67	65.39	55.37	70.49	70.25	8.2
M2Det	72.25	77.49	55.06	51.59	52.34	44.43	28.17	65.66	59.94	4.1
FCOS	87.91	80.44	62.68	58.94	87.09	73.02	49.01	80.35	71.57	11.8
Ours	94.52	92.44	71.05	68.47	94.24	86.05	78.81	83.91	77.18	12.7

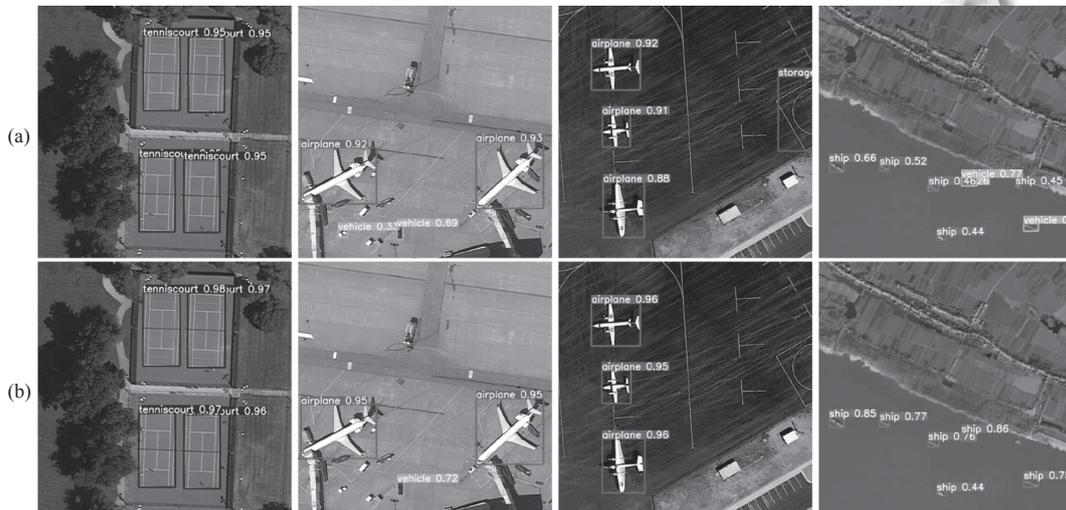


图5 改进模型与YOLOv5s的检测效果对比示意图

4 结论与展望

针对遥感图像内复杂背景干扰、目标密集堆积的问题,本文对YOLOv5s模型加以改进.在骨干部分增加CBAM-MLCA注意力模块,提高模型对于重点特征

的提取能力,采用DySample采样方式以及更换损失函数减少模型复杂度,并在RSOD和DIOR两个遥感数据集上进行实验,相比原模型精准度分别提高2.8%、7.95%,且与其他改进模型相比,仍能保持精准度和参

数量的优势. 证明本模型在遥感图像目标检测应用中具备较强的性能优势. 虽然注意力模块对于模型的精准度提升较为明显, 但是参数量和检测速度的增加也不可忽视, 下一步将以模型的轻量化为目标进行研究.

参考文献

- 1 Wang K, Wang Z, Li Z, *et al.* Oriented object detection in optical remote sensing images using deep learning: A survey. arXiv:2302.10473, 2024.
- 2 Sadgrove EJ, Falzon G, Miron D, *et al.* Real-time object detection in agricultural/remote environments using the multiple-expert colour feature extreme learning machine (MEC-ELM). *Computers in Industry*, 2018, 98: 183–191.
- 3 Kussul N, Lavreniuk M, Skakun S, *et al.* Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(5): 778–782. [doi: 10.1109/LGRS.2017.2681128]
- 4 薛俊达, 朱家佳, 张静, 等. 基于 FFC-SSD 模型的光学遥感图像目标检测. *光学学报*, 2022, 42(12): 1210002.
- 5 Cheng G, Yuan X, Yao XW, *et al.* Towards large-scale small object detection: Survey and benchmarks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(11): 13467–13488.
- 6 汪鹏, 郑文凤, 史进, 等. 基于 MFANet 和上下文特征融合的遥感影像目标检测. *应用科学学报*, 2022, 40(1): 131–144. [doi: 10.3969/j.issn.0255-8297.2022.01.012]
- 7 张绍文, 史卫亚, 张世强, 等. 基于加权感受野和跨层融合的遥感小目标检测. *电子测量技术*, 2023, 46(18): 129–138.
- 8 宋爽爽, 肖开斐, 刘昭华, 等. 一种基于 YOLOv5 的高分辨率遥感影像目标检测方法. *自然资源遥感*, 2024, 36(2): 50–59. [doi: 10.6046/zrzyyg.2023052]
- 9 周沁坤, 周华平, 孙克雷, 等. ARST-YOLOv7: 用于航空遥感图像的小目标检测网络. *计算机工程与应用*, 2024. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240905.1859.05.html>, (2024-09-06)[2024-09-20].
- 10 孙文赞, 车嘉航, 金忠. 基于全局上下文注意力特征融合金字塔网络的遥感目标检测. *计算机系统应用*, 2024, 33(9): 114–122. [doi: 10.15888/j.cnki.csa.009631]
- 11 Zou ZX, Chen KY, Shi ZW, *et al.* Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023, 111(3): 257–276.
- 12 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.
- 13 Liu WZ, Lu H, Fu HT, *et al.* Learning to upsample by learning to sample. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 6004–6014.
- 14 Zhang YF, Ren WQ, Zhang Z, *et al.* Focal and efficient IoU loss for accurate bounding box regression. *Neurocomputing*, 2022, 506: 146–157. [doi: 10.1016/j.neucom.2022.07.042]
- 15 Fu J, Liu J, Tian HJ, *et al.* Dual attention network for scene segmentation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 3141–3149.
- 16 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 17 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 13708–13717.
- 18 Wan DH, Lu RS, Shen SY, *et al.* Mixed local channel attention for object detection. *Engineering Applications of Artificial Intelligence*, 2023, 123: 106442. [doi: 10.1016/j.engappai.2023.106442]
- 19 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 11531–11539.
- 20 Li K, Wan G, Cheng G, *et al.* Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296–307. [doi: 10.1016/j.isprsjprs.2019.11.023]
- 21 Long Y, Gong YP, Xiao ZF, *et al.* Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2486–2498. [doi: 10.1109/TGRS.2016.2645610]
- 22 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 7464–7475.
- 23 Zhao QJ, Sheng T, Wang YT, *et al.* M2Det: A single-shot object detector based on multi-level feature pyramid network. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Honolulu: AAAI, 2019. 9259–9266.
- 24 Ge Z, Liu ST, Wang F, *et al.* YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
- 25 Tian Z, Shen CH, Chen H, *et al.* FCOS: Fully convolutional one-stage object detection. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 9626–9635. [doi: 10.1109/ICCV.2019.00972]

(校对责编: 张重毅)