2025 年 第 25 卷 第 4 期
2025, 25(4): 01555-08

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2309878
引用格式:费选, 郭梦瑶, 吴思佳, 等. 融合多层特征与上下文信息的 YOLO 改进算法[J]. 科学技术与工程, 2025, 25(4): 1555-1562.
Fei Xuan, Guo Mengyao, Wu Sijia, et al. Improved YOLO algorithm *via* fusing multilayer features and contextual information[J]. Science Technology and Engineering, 2025, 25(4): 1555-1562.

# 融合多层特征与上下文信息的 YOLO 改进算法

费选, 郭梦瑶, 吴思佳, 靳子泷, 马丁
(河南工业大学人工智能与大数据学院, 郑州 450001)

**摘　要**　遥感图像目标检测在军事侦察、智慧农业等领域意义重大,特别是小目标检测一直获得持续关注。然而,遥感图像中的小目标面临特征信息不足、检测难度大等问题,成为困扰遥感检测应用发展的最大障碍。为此,提出 YOLO-HF(you only look once-hybrid feature)算法,该算法在传统 YOLOv7 模型的网络中,引入通道注意力和自注意力的混合注意力机制提取目标深层特征,并将浅层特征和深层特征进行融合,增加局部特征的丰富性;为进一步加强对全局信息的关注,在提取特征后为小尺度目标添加全局注意力机制,实现全局特征表达能力的提升;为避免传统损失函数对小目标位置偏差敏感,导致检测效果不佳,选择使用一种新的度量方式,将其嵌入边界框损失函数的计算中,从而加快损失函数的收敛,实现小目标检测精度的提升。实验结果表明:与传统 YOLOv7 算法相比,所提算法在 RSOD 和 NWPU VHR-10 数据集上均表现出优越性,特别地,在 RSOD 数据集上均值平均精度提升了 2.90%,在 NWPU VHR-10 数据集上均值平均精度实现了 3.61% 的提升。

**关键词**　遥感图像; 目标检测; YOLOv7; 多层特征; 注意力机制
**中图法分类号**　TP391.4;　　**文献标志码**　A

# Improved YOLO Algorithm *via* Fusing Multilayer Features and Contextual Information

FEI Xuan, GUO Meng-yao, WU Si-jia, JIN Zi-long, MA Ding
(School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China)

[**Abstract**]　Remote sensing image target detection is one of great significance in military reconnaissance, intelligent agriculture and other fields, especially small target detection has been gaining continuous attention. However, small targets in remote sensing images face the problems of insufficient feature information and difficult detection, which have become the biggest obstacles plaguing the development of remote sensing applications. To this end, the you only look once-hybrid feature(YOLO-HF) algorithm was proposed, which introduced a hybrid attention mechanism of channel attention and self-attention in the network of the traditional YOLOv7 model to extract the target's deep features, and fused the shallow and deep features to increase the richness of local features; to further strengthen the attention to the global information, a global attention mechanism was added for the small-scale targets after the extraction of the features, to achieve the ability of global feature expression enhancement. In order to avoid that the traditional loss function was sensitive to the positional deviation of small targets, which leaded to poor detection effect, a new metric was selected for use, which was embedded into the computation of the bounding box loss function, so as to accelerated the convergence of the loss function and realized the enhancement of the detection accuracy of small targets. The experimental results show that compared with the traditional YOLOv7 algorithm, the proposed algorithm shows superiority on both RSOD and NWPU VHR-10 datasets, and in particular, the mean average accuracy on RSOD dataset is improved by 2.90%, and the mean average accuracy on NWPU VHR-10 dataset realizes an improvement of 3.61%.

[**Keywords**]　remote sensing images; target detection; YOLOv7; multilayer features; attention mechanism

目标检测作为计算机视觉领域的重要研究方向,被众多研究者所关注。与传统自然图像领域的目标检测不同,遥感图像的目标检测对环境监测、动物保护、交通管理、国防军事等领域具有重要的理论意义和实用价值[1]。基于不同传感器获取的遥感图像,往往蕴含丰富的数据信息,且类型多样、背景复杂,特别是数据集样本中含有大量特征不明显的小目标。此外,传统目标检测通过人工方式提取图像特征[2],导致图像特征信息提取不足,无法有效甄别小目标,阻碍了遥感图像小目标检测的进

**收稿日期**: 2023-12-14;　　**修订日期**: 2024-11-19
**基金项目**: 国家自然科学基金青年科学基金(62006072);河南省重点研发与推广专项(科技攻关)项目(222102210108);粮食处理与控制教育部重点实验室开放课题(KFJJ2022013);河南工业大学创新基金支持计划专项资助(2022ZKCJ11);河南工业大学青年骨干教师培育计划
**第一作者**: 费选(1986—),男,汉族,河南郑州人,博士,副教授。研究方向:高光谱遥感影像分析。E-mail:feixuan@haut.edu.cn。

一步发展和应用。

近年来,随着软硬件环境和计算资源的不断进步,有很多学者开始研究如何将深度学习方法,尤其是卷积神经网络,与各个领域进行结合,并取得了一定成果,这引起了目标检测领域研究者的广泛关注[3]。Girshick 等[4]在目标检测领域应用卷积神经网络,并借助该网络提取图像区域特征,从而实现目标检测,即 R-CNN(regions with CNN features)。与传统滑动窗通过滑动来逐个判断所有可能包含目标的区域截然不同,Girshick 等[4]提出预先提取最可能的目标候选区域,然后利用卷积神经网络对这些候选区域进行特征提取,以便判断和识别目标。这种创新性想法影响深远,为目标检测的研究开辟了新思路。紧随其后的 Fast R-CNN 算法[5]在 R-CNN 的基础上取得了一定的发展,该算法引入 RoI(region of interest pooling)池化层,目的是将各种尺寸的候选区域映射为统一大小的特征图。Ren 等[6]提出 Faster R-CNN 算法,快速生成候选区域的过程由区域生成网络(region proposal network,RPN)负责实现,同时利用 RPN 与共享卷积特征图的卷积操作,获取生成候选区域的边界框和置信度得分,从而结合 RPN 与 Fast R-CNN 形成了端到端的目标检测系统,能够在一定程度上提高目标检测的精度。

当前,以 R-CNN 系列算法为代表的两阶段方法检测速度无法满足实时性的需求,因而不需要生成候选框的单阶段方法逐渐成为主流。Redmon 等[7]提出的 YOLO(you only look once)算法成为单阶段方法的重要代表,引起相关研究者的极大兴趣。与以往的目标检测工作不同,YOLO 算法采用全新的方法,将目标检测过程视为空间分离的边界框和相关类概率的回归问题,进一步利用分割网格进行目标位置和类别的预测。在此基础上,YOLOv2[8]使用 Darknet-19 作为特征提取网络,并考虑到尺度多样性,针对不同尺度利用锚框预测不同形状和大小的目标,提高了目标检测精度。YOLOv3 利用 Darknet-53 作为特征提取的核心网络,并且在此之上嵌入特征金字塔网络(feature pyramid network,FPN)结构,以实现对多尺度目标的检测,这种更细粒度的锚框可以提升对检测目标的定位能力。YOLOv4[9]选择 CSPDarknet53 作为主干网络,其中 Neck 结构主要采用 SPP(spatial pyramid pooling)模块、FPN 和路径聚合网络(path aggregation network,PAN)。YOLOv5 则通过 Mosaic 数据增强处理数据,用 Focus 结构和 CSP(cross-stage-partial-connections)结构进一步提升 YOLO 算法的目标检测精度。李启明等[10]针对 X 射线图像危险品检测存在的问题对 YOLOv5 的网络进行改进,使用剪枝减小模型,并通过坐标注意力机制使网络聚焦检测目标,进一步使用数据增强实现检测性能优化。YOLOv7 在 YOLOv5 基础上,引入 ELAN 模块代替 CSP(cross-stage-partial-connections)模块,对池化操作进行修改使得目标检测能力得到进一步提高。

在许多方面,研究者都选用 YOLO 作为基础进行详尽的研究,周孟然等[11]通过 FReLU 所形成的新卷积块来对空间的解析能力进行提升,引入位置注意来解决钢材缺陷的检测问题,取得成效。郭华玲等[12]利用 RepVGG 和 YOLOv5 的结合对交通标志小目标进行检测得到了不错的效果。蒋启超等[13]将 Transform 和 YOLO 相结合用于驾驶员的疲劳检测,其算法的检测精度和轻量上都具有一定的优势。

随着深度学习方法在遥感图像处理领域的引入和快速发展,遥感图像目标检测的精度得到较大提升。Li 等[14]在 YOLOv4 的基础上,将主干网络替换为 MobileNext 网络以减少参数量,并添加了 RFB(receptive field block)和 ECA(efficient channel attention)结构,通过实验验证了模型在遥感数据集上的检测有效性。张朝阳等[15]针对遥感图像的多尺度、形态多样等问题,引入双向特征金字塔网络,并融合 Swin Transformer 的多头注意力机制,重构网络结构,对 YOLO 算法进行优化。此外,针对遥感图像中普遍存在的小目标遮挡导致漏检及误检问题,如何充分利用被检测目标所具有的独特先验知识对提高检测效果而言很重要。Li 等[16]考虑遥感图像中大量背景先验知识可能提供有效信息,首次在遥感目标检测中探索大卷积核机制,提出了 LSKNet(large selective kernel network),优势在于相对其他检测器较轻量,检测精度在多个数据集上得到大幅提高,极大降低了误检率。针对遥感图像小目标检测性能不理想的问题,Rabbi 等[17]将 GAN 引入遥感图像目标检测领域,结合其他模块能够进一步提高遥感小目标检测能力。Zhang 等[18]通过融合多模态遥感图像中的互补信息来改善小目标检测能力,删除相应模块保留高分辨特征,利用像素级多模态融合提取信息,并通过超分辨辅助分支学习高分辨特征,在低分辨率输入的大背景下区分小物体,从而更好检测小目标。

基于此,针对遥感图像小目标特征信息过少的检测难点,YOLO-HF(you only look once-hybrid feature)以 YOLOv7 算法[19]为基础,通过改进和优化,提升小目标检测能力。将混合注意力转换器(hybrid attention transformer,HAT)模型[20]中对输入图像的浅

层和深层特征提取模块添加到 YOLOv7 的主干网络所提取的特征信息之后,对所得到的特征信息进行处理,提取更多小目标的特征信息;引入全局注意力机制(global attention mechanism,GAM)[21],增强模型对上下文信息的理解,提升模型检测性能;将归一化的 Wassertein 距离(normalization Wassertein distance,NWD)[22]嵌入边界框损失函数中,并调整相应参数,打破传统基于 IoU 度量对小目标位置偏差敏感的局限,提升检测器的检测性能,从而准确评估小目标间的相似度,进一步提高检测精度。

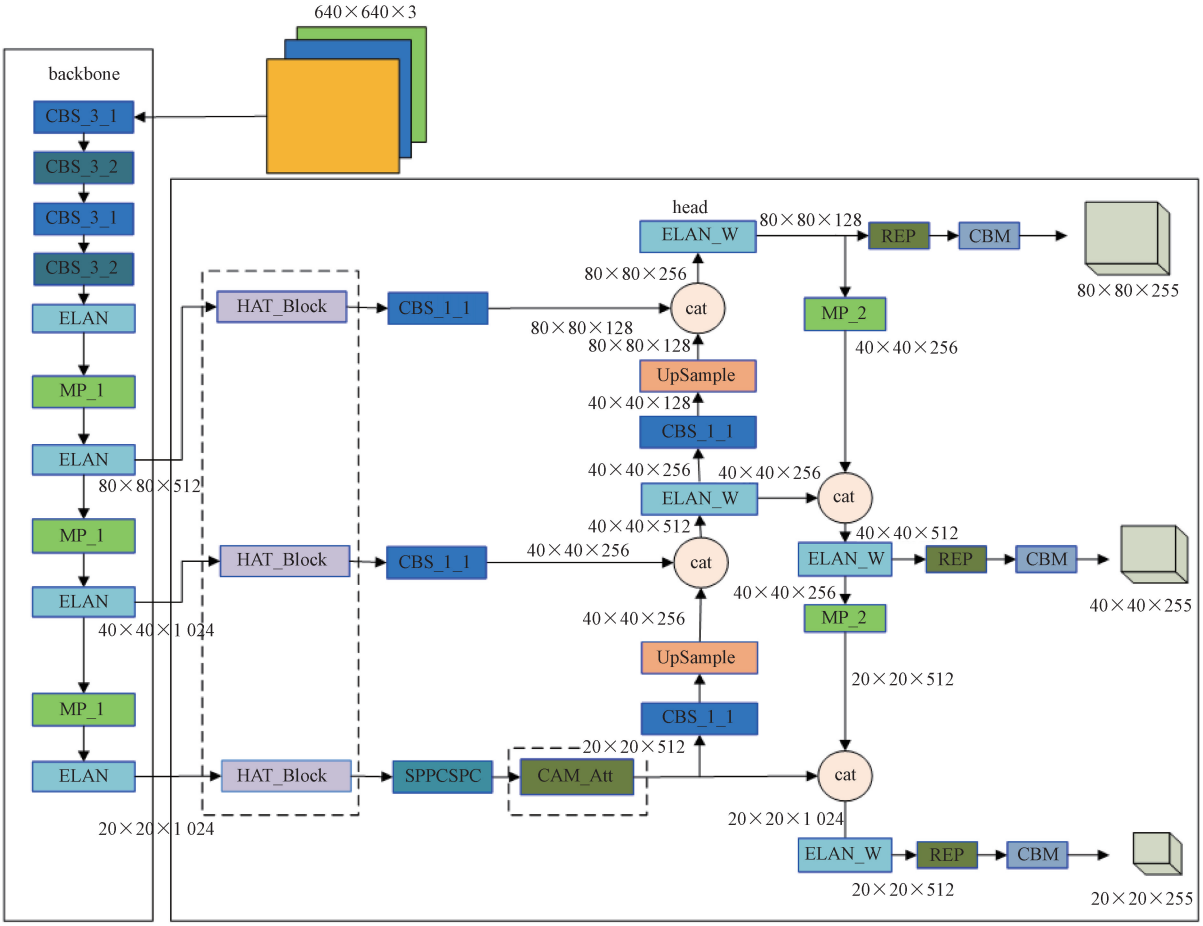
# 1　YOLO-HF

## 1.1　网络模型

YOLOv7 的原始网络中,主干网络是由卷积层堆叠得到,提取的不同层次特征信息不够充分。受图像超分辨可对小目标放大并增强信息的启发,在 YOLOv7 主干网络中增加 HAT_Block 对浅层特征和深层特征进行提取,以增强小目标特征提取能力,便于后续步骤中有较为充分的特征信息进一步对目标进行识别和定位。此外,通过在 SPPCSPC 模块之后引入全局注意力机制 GAM_Att,增强对小目标上下文信息的关注,进一步提升小目标的检测性能。研究发现,原始 YOLOv7 使用基于 IoU(intersection of union)的度量,对小物体位置偏差敏感,导致基于锚框的检测模型性能降低。因此,将 NWD 度量嵌入边界框与预测框的回归损失计算中,将加快损失函数的收敛速度,提高检测精度。修改后的整体网络模型图如图 1 所示。

## 1.2　添加多层特征提取模块

受图像超分辨重建的 HAT 模型启发,在提取浅层特征的基础上,通过引入通道注意力和自注意力的混合注意力机制来提取目标深层特征,为了提升




CBS 为卷积归一化和 SiLU 激活函数模块;ELAN 为高效层聚合网络模块;MP 为最大池化操作;CBM 为卷积归一化和 sigmoid 激活函数模块;REP 为 3×3 卷积和 1×1 卷积和残差连接组合的一个卷积层;CBS_$x$_$x$ 中,C 为 2D 卷积,B 为批归一化层(batch normalization,BN),S 为所使用的激活函数 silu;为了同 sigmoid 区分,用 M 表示 sigmoid 激活函数;$x$_$x$ 中,第 1 个 $x$ 为卷积核尺寸和步长;虚线框所标出部分为修改部分;backbone 为主干特征提取网络;HAT_Block 为本文所引入的混合注意力模块;head 为检测头;cat 为凭借操作;UpSample 为上采样操作;SPPCSPC 为 YOLOv7 结构中特殊的卷积层;GAM_ATT 为全局注意力块

图 1　YOLO-HF 整体结构图

Fig. 1　Improvement of YOlO-HF overall structure

小目标局部特征的多样性,将浅层特征和深层特征进行结合。具体来说,浅层特征提取依然采用卷积层,深层特征提取则使用混合注意力组(residual hybrid attention group,RHAG)结构。紧接着,将浅层特征和深层特征融合,融合是通过使用残差连接方式,最后得到融合了多层特征的结果。

所使用到的多层特征提取模块 HAT_Block 结构如图 2 所示,主要由 RHAG 模块、卷积模块和残差连接构成。其中,RHAG 模块由混合注意力模块 HAB(hybrid attention block)、重叠交叉注意力模块 OCAB(overlapping cross-attention block)和卷积模块组成。由于 HAB 模块计算通道注意力权重时涉及全局信息,所以能激活更多的像素,从而增强网络的表示能力,而 OCAB 模块则通过构建跨窗口的连接进一步提高了表示能力。

### 1.3 添加上下文信息引导模块

为了增强神经网络对全局上下文的信息感知和获取能力,采用 GAM 注意力机制作为上下文信息引导模块是一种比较合适的选择。其主要思想是通过全局上下文的引入指导特征的加权和融合,将每个特征的重要性与全局上下文关联,可以捕捉全局结构,上下文关系和长距离依赖,从而优化网络模型。具体实现过程见式(1)。

$$F_3 = M_s[M_c(F_1) \otimes F_1] \otimes F_2 \tag{1}$$

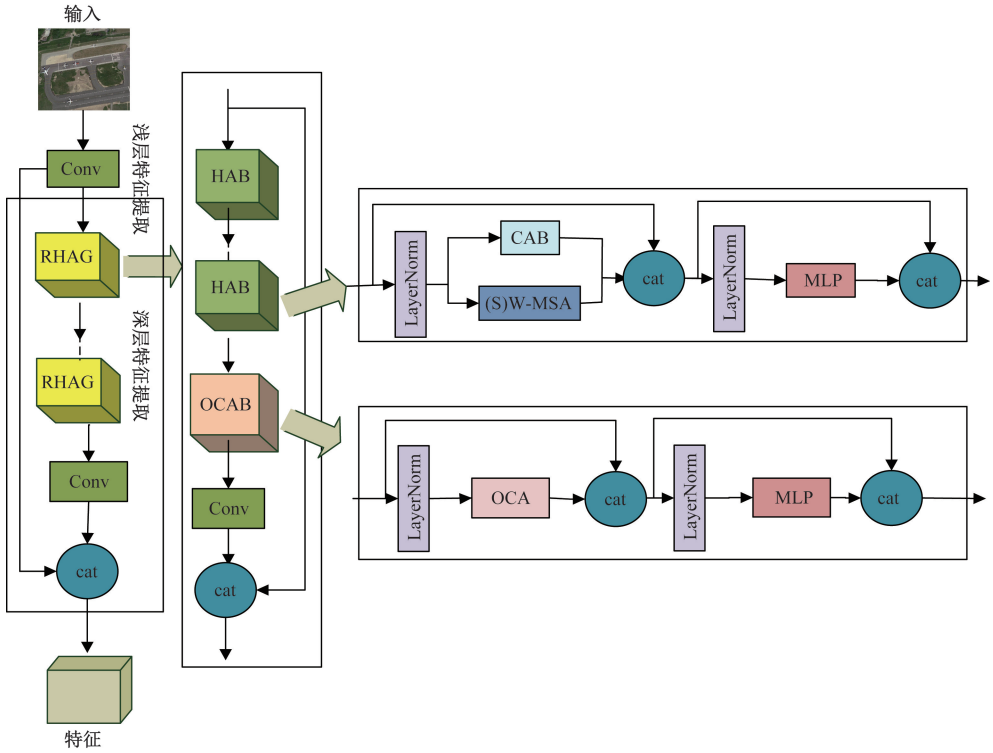
式(1)中:$F_1$ 为输入特征;$M_c(*)$表明 $*$ 经过通道注意力后得到的输出;$F_2$为 $M_c(F_1)$和 $F_1$ 相互作用后的中间结果,也可作为输入,进入后续空间注意力模块中;$M_s(*)$为经过空间注意力后得到的输出;$F_3$为得到的输出特征。

所使用的上下文信息引导模块 GAM_Att 结构如图 3 所示,它是一种全局调度指挥控制机制,借助减少相关信息缩减和放大全局交互操作表示,达到在保留通道和空间信息的基础上,增强跨维度交互,有利于目标检测时对全局及远距离信息的把握,进而提高深度神经网络的性能,提升检测效果。

### 1.4 损失函数改进

YOLOv7 的损失函数由目标置信度损失、类别置信度损失、预测框和真实框的回归损失 3 个指标组成。在预测框与真实框的回归损失计算中,采用的 IoU 度量对小目标位置偏差敏感,导致基于锚框的检测模型性能降低。而 NWD 对不同尺度的物体不敏感,更适合测量微小物体之间的相似性,因此使用 NWD 度量替换原来的 IoU 度量,以获得较好的小目标检测效果。

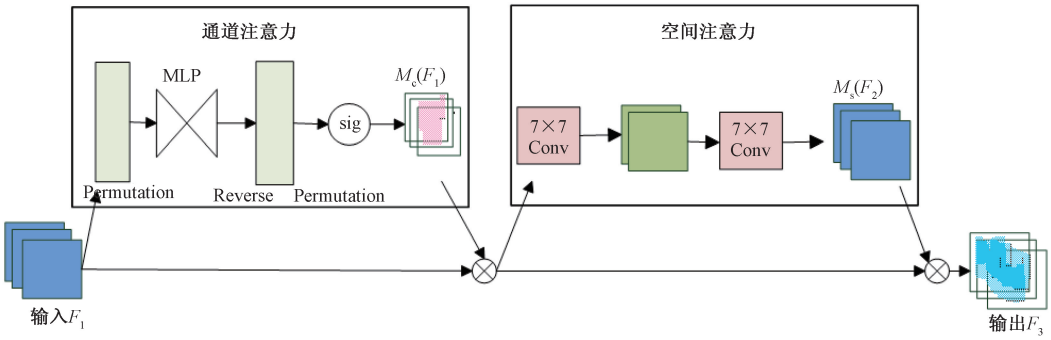
NWD 度量方式的提出是为了减轻 IoU 对小物体位置偏差的敏感性,从而提升模型对小目标检测




Conv 为卷积;RHAG 为残差混合注意力组;cat 拼接;HAB 为混合注意力块,具体结构如 HAB 箭头所指;OCAB 为重叠交叉注意力块;CAB 为通道注意力块;W-MSA 为基于窗口的多头自注意力;OCA 为重叠交叉注意力;LayerNorm 为层归一化;MLP 为多层感知机

图 2 HAT_Block 结构图

Fig. 2 HAT_Block structure




Permutation 为转置操作;MLP 为多层感知机;Reverse Permutation 为逆转置操作;sig 为激活函数;Conv 为卷积

图 3　GAM_Att 结构图

Fig. 3　GAM_Att structure

效能。主要思想是通过建模将边界框构造为二维高斯分布,期间需要使用表示框的中心点坐标($c_x$,$c_y$)和宽 $w$ 以及高 $h$ 的参数。水平框($c_x$,$c_y$,$w$, $h$)到二维高斯分布 $N(\mu,\Sigma)$,其中,$\mu$ 为高斯分布的均值,$\Sigma$ 为高斯分布的方差。

建模过程可由式(2)进行定义。

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

然后使用 NWD 来计算高斯分布的相似性。计算 Wasserstein 距离的计算公式为

$$n_1 = N(\mu_1,\Sigma_1) \tag{3}$$

$$n_2 = N(\mu_2,\Sigma_2) \tag{4}$$

$$W_2^2(n_1,n_2) = \|\mu_1-\mu_2\|_2^2 + \mathrm{tr}[\Sigma_1+\Sigma_2-2(\Sigma_1^{\frac{1}{2}}\Sigma_1\Sigma_2^{\frac{1}{2}})^{\frac{1}{2}}] \tag{5}$$

式中:$N$ 为高斯分布;$n_1$、$n_2$ 均服从高斯分布;$\mu$ 为高斯分布的均值;$\Sigma$ 为高斯分布方差;tr 为矩阵的迹;有序的 Wasserstein 距离定义为 $W_2^2$。

将 $W_2^2$ 简化后得到式 4,式中的 $\Sigma^{\frac{1}{2}}$ 为标准差,将式(3)中矩阵迹的计算简化为了 Frobenius 范数形式。

$$W_2^2(n_1,n_2) = \|\mu_1-\mu_2\|_2^2 + \|\Sigma_1^{\frac{1}{2}}-\Sigma_2^{\frac{1}{2}}\|_F^2 \tag{6}$$

在此基础上,再通过计算 NWD 作为新的度量,NWD 可嵌入损失函数以及非极大值抑制中取代常用的 IoU 度量指标。利用对边界框 a 和 b 建立的高斯分布模型 $n_a$ 和 $n_b$,求得 NWD。在嵌入过程中,可利用 iou_ratio 的参数调节实现对小目标友好的损失计算方式。它的值靠近 0 方向时,将更适用于数据集中小目标居多的情况。因此可根据数据集的小目标比例调整该参数值。考虑到所使用的数据集中小目标占比,在实验过程中将其值设为 0.4。

特别地,当边界框 a 和 b 是以中心坐标($c_x$,$c_y$)、宽 $w$ 和高 $h$ 来表示时,分别用 $n_a$ 和 $n_b$ 表示边界框 a 和 b 服从的高斯分布,先求出其 Wasserstein 距离,再根据 Wasserstein 距离求 NWD,其表达式为

$$W_2^2(n_a,n_b) = \left\| \left[c_{xa},c_{ya},\frac{w_a}{2},\frac{h_a}{2}\right]^{\mathrm{T}},\left[c_{xb},c_{yb},\frac{w_b}{2},\frac{h_b}{2}\right]^{\mathrm{T}} \right\|_2^2 \tag{7}$$

$$\mathrm{NWD}(n_a,n_b) = \exp\left[\frac{\sqrt{W_2^2(n_a,n_b)}}{C}\right] \tag{8}$$

式(8)中:$C$ 为常数,与数据集相关,本文数据集中将其值设为 2。

## 2　实验结果与分析

### 2.1　实验数据集及对比算法

实验采用的两个遥感数据集均包含有大量的小目标,能够很好地验证模型的目标检测效果。一个是由武汉大学标注的 RSOD 数据集,共包含 4 类数据,分别是飞机、操场、立交桥和油桶。数据集一共有 935 张图像,其中飞机实例 4 993 个,操场 191 个,立交桥 180 个,油桶 1 586 个。另一个数据集是 NWPU VHR-10 数据集。这些是由西北工业大学标注的,分别指飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁和车辆,而这 10 类实例分布于该数据集的 800 张图像之中,具体而言,背景图像 150 张,图像含目标的则有 650 张。在实验数据设计中进行了细致的分配,训练集、测试集、验证集之比为 7:2:1,也就是说,训练集占据了数据集的 70%,数据集的 20% 视为测试集,剩余部分为验证集。为了验证本文算法的有效性,选取有代表性的部分双阶段和单阶段算法进行比较,包括检测精度高但速度慢的 Faster R-CNN 算法、以 VGGNet 作为骨干网络的 SSD(single shot multibox detector)算法[23]、YOLO 系列模型中的 YOLOv5 算法和

YOLOv7 算法。

### 2.2 实验评价指标

实验结果的优劣主要通过平均检测精度 mAP 来进行判定。同时,还涉及其他指标,如准确率 $P$、召回率 $R$ 和单个类别的平均精度 mAP,其计算公式分别为

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{n} \int_0^1 P(R)\,\mathrm{d}R}{n} \tag{11}$$

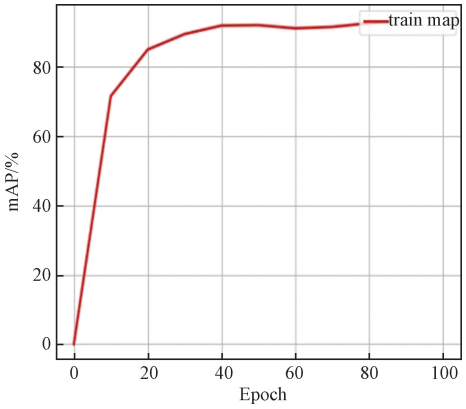
式中:TP 为检测结果为正例,实际也是正例;FP 为被错误地判断为正例的负例的样本;FN 为实际为正例,却被检测为负例,但这一情况不常见,因此值一般小一些;$n$ 为数据集总的类别数。

### 2.3 实验参数及实验结果

实验所使用的操作系统为 64 位 Windows 10 系统,显卡是 NVIDIA GeForce RTX 3070Ti 8GB,CUDA 版本为 11.7,CUDNN 版本为 11.0,PyTorch 版本为 1.7.1,python 版本为 3.7.1。训练过程中的 mAP(IoU =0.5)变化如图 4 所示。该数据集在训练过程中在 40 轮基本收敛,后续逐渐趋于平稳。

RSOD 数据集上的对比检测结果如表 1 所示。本文算法相比 Faster R-CNN,mAP 提升 7.64%;相比 SSD,提升了 17.39%;相比 YOLOv5,提高了 4.98%。相对于 YOLOv7,修改后的本文算法在 RSOD 数据集的 mAP 提升了 2.9%,其中飞机、操场和立交桥分别在增加了 1.3%、0.4% 和 10.02%,油罐检测结果几乎无变化。

根据 NWPU VHR-10 数据集上的对比检测结果表 2 中呈现了相关数据。修改后的算法训练所得 mAP




Epoch 为轮次

图 4 RSOD 数据集 mAP 变化图

Fig. 4 ThemAP variation graph of RSOD dataset

表 1 部分两阶段和单阶段算法在 RSOD 数据集上训练对比结果

Table 1 Comparison results of some two-stage and single-stage algorithms trained on the RSOD dataset

| 算法 | 单类别检测精度/% | | | | mAP/% |
|---|---|---|---|---|---|
| | 飞机 | 操场 | 立交桥 | 油罐 | |
| Faster R-CNN | 63.51 | 99.26 | 88.59 | 92.82 | 86.04 |
| SSD | 45.42 | 99.06 | 68.73 | 91.95 | 76.29 |
| YOLOv5 | 93.09 | 98.15 | 66.26 | 98.15 | 88.70 |
| YOLOv7 | 92.29 | 98.77 | 75.17 | 96.88 | 90.78 |
| 本文算法 | 93.59 | 99.10 | 85.15 | 96.87 | 93.68 |

表 2 部分两阶段和单阶段算法在 NWPU VHR 数据集上训练对比结果

Table 2 Comparison results of some two-stage and single-stage algorithms algorithms trained on NWPU VHR dataset

| 类别 | 不同算法检测结果/% | | | | |
|---|---|---|---|---|---|
| | Faster R-CNN | SSD | YOLOv5 | YOLOv7 | 本文算法 |
| 飞机 | 98.31 | 90.40 | 99.95 | 100.00 | 99.99 |
| 棒球场 | 99.55 | 89.90 | 97.73 | 98.07 | 98.10 |
| 篮球场 | 95.35 | 80.60 | 83.18 | 89.86 | 96.05 |
| 桥梁 | 86.33 | 76.70 | 75.61 | 71.29 | 87.89 |
| 田径场 | 99.95 | 98.31 | 100.00 | 99.19 | 99.90 |
| 港口 | 95.75 | 73.40 | 89.39 | 91.35 | 94.56 |
| 舰船 | 75.60 | 60.90 | 82.44 | 86.64 | 85.40 |
| 油罐 | 65.60 | 79.80 | 98.33 | 93.12 | 95.51 |
| 网球场 | 81.80 | 82.60 | 91.11 | 90.48 | 93.89 |
| 车辆 | 47.82 | 52.10 | 73.93 | 81.97 | 86.83 |
| mAP | 84.61 | 78.40 | 89.17 | 90.20 | 93.81 |

在 Faster R-CNN 基础上提升了 9.2%;相比 SSD 算法,提高了 15.41%;相较 YOLOv5 算法,mAP 提升 4.64%;与 YOLOv7 相比,mAP 增加了 3.61%。其中,篮球场、桥梁、港口、网球场、车辆的 AP 值提升较为明显。

为了直观展示目标检测效果,下面以 RSOD 数据集中的图像为例,在保持实验参数一致的基础上,根据图 5 所示,其中 YOLOv7 的检测结果如图 5(a)所示,对于其中两个飞机的实例未检测出,本文算法的检测结果如图 5(b)所示。可以看出,在相同的情况下,YOLOv7 算法出现了漏检飞机实例的现象,而本文算法能够将更多小目标检测出来。

### 2.4 消融实验

为了验证所修改各个部分是否有效,在 RSOD 数据集上进行了消融实验,并在表 3 中展示实验结果。修改加入 HAT_Block 模块的 mAP 增长主要是由立交桥的值提高所引起的。在仅添加 HAT_Block 模块时,mAP 提高了 0.31%;仅添加 GAM_Att 模块时,mAP 仅提高了 0.36%;仅修改度量方式时,mAP 提升较少,只有 0.2%。当所提三部分都进行修改后,检测结果相较原 YOLOv7 算法提升了 2.9%,在


(a) YOLOv7算法


(b) 本文算法

图 5　检测飞机对比图

Fig. 5　Comparison of test airplanes

**表 3　消融实验结果**

**Table 3　Results of ablation experiment**

| 算法 | 改进 1 | 改进 2 | 改进 3 | mAP/% |
|---|---|---|---|---|
| YOLOv7 | | | | 90.78 |
| HAT_Block | √ | | | 91.07 |
| GAM_Att | | √ | | 91.02 |
| NWD 度量 | | | √ | 90.90 |
| 本文算法 | √ | √ | √ | 93.68 |

注:“√”为添加某一改进算法模块。

RSOD 数据集上,改进后的算法显示出对各块的改进是有效的。

## 3　结论

在遥感图像目标检测中,由于小目标信息量较少,有效鉴别特征提取困难,导致整体检测精度下降。为了增加小目标检测的准确性,在基于 YOLOv7 模型框架基础上,进行改进。通过实验得出以下结论。

(1)引入混合注意力机制提取深层特征,并融合浅层特征以增强多层局部特征的丰富性,进一步有效提升目标检测的准确性。

(2)利用上下文信息添加全局注意力机制,进一步实现全局特征表达能力的提升。

(3)NWD 度量融入到边界框损失函数的计算过程中,以减弱原模型中 IoU 度量对小目标位置偏差敏感的缺陷,提高目标检测准确率。

(4)进行了一系列实验,其中包括 RSOD 数据集和 NWPU VHR-10 数据集。本文方法的平均检测精度分别达到了 93.68% 和 93.81%。相比其他几类代表性算法,具有明显提升。与传统 YOLOv7 算法相比,本文算法在 RSOD 和 NWPU VHR-10 两个数据集上均表现出优越性,特别地,在 RSOD 数据集上均值平均精度提升了 2.90%,在 NWPU VHR-10 数据集上均值平均精度实现了 3.61% 的提升。此外,本文方法相对 YOLOv7 增加了模型参数量和计算压力,并在一定程度上降低了检测的时间效率,因此后序将针对模型轻量化展开进一步研究。

## 参　考　文　献

[1] 马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标检测[J]. 光电工程, 2022, 49(4): 49-65.
Ma Liang, Gou Yutao, Lei Tao, et al. Small target detection in remote sensing images based on multi-scale feature fusion[J]. Photoelectric Engineering, 2022, 49(4): 49-65.

[2] 程塨, 周培诚, 韩军伟. 基于旋转不变卷积神经网络的高分辨率光学遥感图像目标检测[J]. 科学观察, 2020, 15(6): 75-76.
Cheng Gong, Zhou Peicheng, Han Junwei. Target detection in high-resolution optical remote sensing images based on rotationally invariant convolutional neural networks[J]. Scientific Observation, 2020, 15(6): 75-76.

[3] 院老虎, 常玉坤, 刘家夫. 基于改进 YOLOv5s 的雾天场景车辆检测方法[J]. 郑州大学学报(工学版), 2023, 44(3): 35-41.
Yuan Laohu, Chang Yukun, Liu Jiafu. Vehicle detection method based on improved YOLOv5s in foggy scene[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(3): 35-41.

[4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2014: 580-587.

[5] Girshick R. Fast R-CNN[C]//IEEE International Conference on Computer Vision(ICCV). Piscataway: IEEE, 2016: 1440-1448.

[6] 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.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.

[8] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, H. I.: IEEE, 2017: 6517-6525.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 1-17.

[10] 李启明, 阙祖航. 基于改进 YOLOv5 的 X 射线图像危险品检测[J]. 科学技术与工程, 2023, 23(4): 1598-1606.
Li Qiming, Que Zuhang. Detection of dangerous objects in X-ray images based on improved YOLOv5[J]. Science Technology and Engineering, 2023, 23(4): 1598-1606.

[11] 周孟然, 王昊男, 高立鹏, 等. 基于 YOLOv5s-FCS 的钢材表面缺陷检测[J]. 科学技术与工程, 2024, 24(14): 5901-5910.
Zhou Mengran, Wang Haonan, Gao Lipeng, et al. YOLOv5s-FCS based steel surface defect detection study[J]. Science Technology and Engineering, 2024, 24(14): 5901-5910.

[12] 郭华玲, 刘佳帅, 郑宾, 等. 融合 RepVGG 的 YOLOv5 交通标志识别算法[J]. 科学技术与工程, 2024, 24(9): 3869-3875.
Guo Hualing, Liu Jiashuai, Zheng Bin, et al. YOLOv5 traffic sign recognition algorithm combined with RepVGG[J]. Science Technology and Engineering, 2024, 24(9): 3869-3875.

[13] 蒋启超, 余成波, 宣以国, 等. 基于轻量级主干的 YOLOv5 驾驶员疲劳检测算法[J]. 科学技术与工程, 2024, 24(16): 6766-6774.
Jiang Qichao, Yu Chengbo, Xuan Yiguo, et al. Driver fatigue detection algorithm based on lightweight YOLOv5[J]. Science Technology and Engineering, 2024, 24(16): 6766-6774.

[14] Li C, Xu R, Lü Y, et al. Edge realtime object detection and DPU-based hardware implementation for optical remote sensing images[J]. Remote Sensing, 2023, 15(16): 3975.

[15] 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 Chaoyang, Zhang Shang, Wang Hengtao, et al. Remote sensing of small targets with multiple attention at multiple scales force detection [J]. Computer Engineering and Applications, 2023, 59(8): 227-238.

[16] Li Y, Hou Q, Zheng Z, et al. Large selective kernel network for remote sensing object detection[J]. arXiv Preprint, 2023: https://arxiv.org/pdf/2303.09030.pdf.

[17] Rabbi J, Ray N, Schubert M, et al. Small objectdetection in remote sensing images with end-to-end edge-enhanced GAN and object detector network[J]. Remote Sensing, 2020, 12: 1432.

[18] Zhang J, Lei J, Xie W, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.

[19] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies setsnew state-of-the-art for realtime object detectors [J]. arXiv Preprint, 2022: https://arxiv.org/abs/2207.02696.

[20] Chen X Y, Wang X T, Zhou J T, et al. Activating more pixels in image super-resolution transformer[J]. arXiv Preprint, 2022: https://arxiv.org/pdf/2205.04437.pdf.

[21] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[J]. arXiv Preprint, 2021: https://arxiv.org/pdf/2112.05561.pdf.

[22] Wang J W, Xu C, Yang W et al. A normalized gaussian wasserstein distance for tiny object detection[J]. arXiv Preprint, 2021: https://arxiv.org/abs/2110.13389.

[23] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi box detector[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.