文章编号：1003－1251(2026)05－0077－08

基于 Swin-T 和 YOLOv8 的遥感小目标检测算法

张云龙[1]，葛长青[2]，张东阳[1]，张亚超[1]

(1. 沈阳理工大学 装备工程学院，沈阳 110159；2. 中国人民解放军 32752 部队 湖北 襄阳 441100)

摘 要：针对遥感小目标检测中 YOLO 系列模型特征提取不足及检测精度受限的问题，提出一种融合 Swin Transformer(Swin-T)的 YOLOv8 改进模型。以 YOLOv8 为基础框架，首先将 C2f 模块中的 Bottleneck 进行改进，在第 2 个卷积后面引入 Swin-T，构建特征融合模块 C2f-ST，在骨干网络中增加一个 Swin-T 层以获取多样化图像特征，增强全局上下文信息提取能力；其次在颈部网络引入注意力模块 CBAM 用以增强对目标区域的检测；最后采用 WIoU 损失函数作为回归损失，同时新增一个检测头，以增强多尺度特征捕获能力。在 VisDrone2019 数据集上的实验结果表明改进模型比原模型在精确率、召回率及平均精度均值上分别提升 3.7%、3.6%和 4.6%，表明 C2f-ST 模块可以有效整合局部特征与全局信息，提升小目标检测精度。改进模型兼顾轻量化与实时性，适用于检测资源受限场景。

关 键 词：小目标检测；Swin Transformer；注意力机制；YOLOv8

中图分类号：TP391.41 **文献标志码**：A **DOI**：10.3969/j.issn.1003－1251.2026.05.011

Remote Sensing Small Target Detection Algorithm Based On Swin-T And Yolov8

ZHANG Yunlong[1]，GE Changqing[2]，ZHANG Dongyang[1]，ZHANG Yachao[1]

(1. Shenyang Ligong University，Shenyang 110159，China；2. Unit 32752 of the People's Liberation Army of China，Xiangyang 441100，China)

Abstract：To address the issues of insufficient feature extraction and limited detection accuracy in YOLO series models for remote sensing small target detection，an improved YOLOv8 model integrating Swin Transformer (Swin-T) is proposed. Based on the YOLOv8 framework，the Bottleneck in the C2f module is firstly improved by introducing Swin-T after the second convolution to construct the C2f-ST feature fusion module. An additional Swin-T layer is added to the backbone network to obtain diverse image features and enhance the ability to extract global context information. Then，a CBAM attention module is introduced in the neck network to enhance the detection of target areas. Finally，the WIoU loss function is used as the regression loss，and an additional detection head is added to enhance the ability to capture multi-scale features. The experimental results on the VisDrone2019 dataset show that the improved model achieves increases of 3.7%，3.6%，and 4.6% in precision，recall rate，and the mean average precision，respectively，compared to those with the original model，indicating that the C2f-ST module can successfully integrate local features and global information to improve the accuracy of small target detection. The improved model is lightweight and real-time，making it suitable for scenarios with limited detection resources.

Key words：small target detection；Swin Transformer；attention mechanism；YOLOv8

收稿日期：2025－09－04

基金项目：国家重点研发计划课题(2023YFB3209704)

作者简介：张云龙(2000—)，男，硕士研究生；葛长青(1983—)，通信作者，男，讲师。

光学遥感图像因其信息丰富,在军事侦察、城市规划、环境监测及灾害预警等领域具有重要应用。然而,该类图像中目标尺度差异大,小目标数量多且易受光照、天气及复杂背景干扰,导致传统检测方法精确率不足。近年来,卷积神经网络(CNN)的引入有效提升了遥感目标检测的准确性与适应性,成为应对复杂场景和小目标识别的主流算法。

CNN 依靠多层卷积结构逐步提取图像中的复杂特征,利用并行计算提升训练与推理速度,成为目标检测的研究热点。早期的目标检测方法通常由多层特征提取并结合区域候选网络(RPN)生成候选框,从而在一定程度上提高了检测精度。但以上技术运算量大、检测速度慢,难以满足实时应用的需求。相比之下,单阶段检测框架,如 YOLO 系列,具备较快的推理速度和较高的精度。冯迎宾[1]等提出了一种改进 YOLOv8n 算法,应用于夜间环境下的车辆检测;李佳兴[2]等提出了一种使用非局部注意力机制的改进 YOLOv8 算法用于检测细长物体。

目前,许多学者在遥感图像目标检测领域进行了研究。Hui 等[3]提出了 DSAA-YOLO 算法,引入超分辨率数据增强策略与密集残差结构,结合特征对齐模块与多目标优化策略,有效改善了小目标的细节表征,提升了无人机影像中小目标的识别精度。张腾泽等[4]在 YOLOv5 中引入 CA 注意力机制,设计多尺度 FRM 模块,并采用 Copy-reduce-paste 数据增强,提高了小目标在复杂背景中的检出率。秦伦明等[5]基于 YOLOv8 设计了一种双向加权特征金字塔结构,增加浅层特征提取分支,并引入 EMA 注意力机制与 SWIoU 损失函数,进一步优化了对多尺度目标的检测性能。万辉耀等[6]针对细粒度遥感图像进行检测,在 YOLOv8 中嵌入 FCBAM 模块及判别信息提取模块 DIEM,结合多策略数据增强,缓解了长尾分布带来的类别不平衡问题。侯林静等[7]则注重模型轻量化,使用 EfficientNet 替换 YOLOv7 原有结构,设计了 SPPAFC 模块,采用 WIoU 损失函数,在保持较低参数量的同时可维持较高的检测精度。

尽管上述方法在不同层面提升了遥感小目标检测效果、但在高密度场景下仍存在漏检和误检问题,特征提取不足、检测精度较低,且部分模型计算复杂度较高,在实际部署中面临挑战。针对上述问题,本文利用 Vision Transformer(ViT)[8]在计算机视觉领域的优势,以 YOLOv8s 作为基础框架进行改进,将 C2f 模块中的 Bottleneck 进行改进,构建特征融合模块 C2f-ST,与其他改进方法仅引入注意力机制不同的是,本文在骨干网络中增加一个 Swin Transformer(Swin-T)层以获取多样化图像特征,将原有卷积结构替换为 C2f-ST 的组合结构,从而提高捕获全局和上下文信息的能力;在颈部网络引入注意力模块 CBAM,用以增强对目标区域的检测从而减少对目标的漏检;采用 WIoU 损失函数作为回归损失,并在头部新增一个检测头,用以提升遥感小目标的检测精度和检出率。

1 Vision Transformer 相关原理

ViT 将 Transformer 架构首次直接应用于图像分类任务,通过对非重叠的中等尺寸图像块进行处理替代传统卷积结构。其通常需要先在大规模数据集上完成预训练,再针对小规模下游任务进行微调,才能达到较好的适应性与泛化能力,模型结构如图 1 所示。其将输入的整幅图像划分为固定尺寸的图像块,并通过线性映射后叠加位置编码输入到标准的 Transformer 编码器中。在生成的嵌入序列中,额外加入一个可训练的类别标记以实现图像分类任务。

1.1 注意力机制

在计算机视觉领域,注意力机制应用于建立空间维度上的长程依赖,以克服卷积核感受野受限的不足[9]。

ViT 使用的自注意力机制(SA)是神经网络架构中的核心,同时也是一种缩放点积注意力,防止在高维场景下的梯度消失,其计算过程如图 2 所示。SA 层通过计算序列内元素间的相关性权重,动态聚合上下文信息。注意力权重 A 是由序列中两个元素的成对相似度及其对应的查询和键生成,计算过程如下。

$$[\boldsymbol{q},\boldsymbol{k},\boldsymbol{v}]=\boldsymbol{z}U_{qkv},U_{qkv}\in\mathrm{R}^{D\times 3D_h} \tag{1}$$

式中:$\boldsymbol{q},\boldsymbol{k},\boldsymbol{v}$ 分别为查询、键和值矩阵;$\boldsymbol{z}$ 为输入序列特征矩阵;U_{qkv} 为线性投影矩阵;$\boldsymbol{D}$ 为特征维度;D_h 为第 h 个注意力头的特征维度。

$$\boldsymbol{A}=\mathrm{Softmax}(\boldsymbol{q}\boldsymbol{k}^{\mathrm{T}}/\sqrt{D_h}),\boldsymbol{A}\in\mathrm{R}^{N\times N} \tag{2}$$

式中:$\boldsymbol{A}$ 为注意力权重矩阵;$\boldsymbol{q}\boldsymbol{k}^{\mathrm{T}}$ 为一个 $N\times N$ 的相关性矩阵,用于计算 $\boldsymbol{q}$ 与 $\boldsymbol{k}$ 之间的点积相似度;$\sqrt{\boldsymbol{D}_h}$ 为缩放因子。

加权聚合过程可表示为

$$\boldsymbol{S}\mathrm{A}(\boldsymbol{z})=AV \tag{3}$$

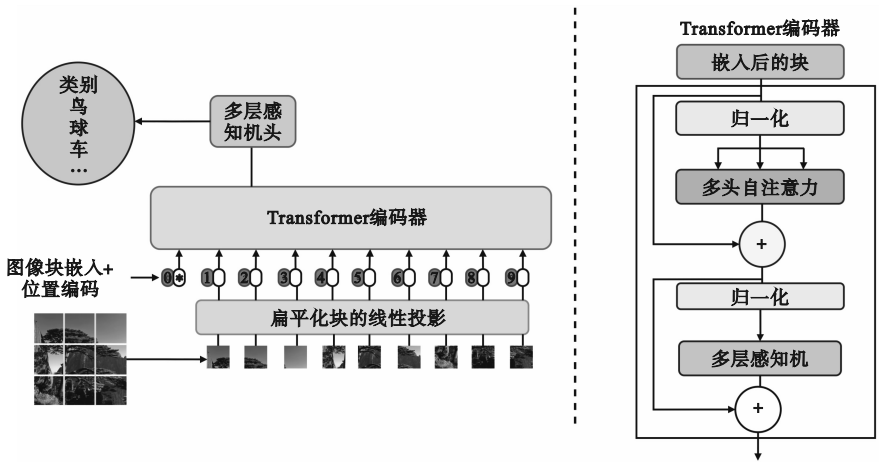


图 1　ViT 模型结构

Fig. 1　ViT model architecture

式中:$SA(z)$为自注意力层的输出;V 为值矩阵。

1.2　多头自注意力

为了获得更丰富的特征表示,ViT 采用了如图 2 所示的多头自注意力机制(MSA),该机制并行执行 h 个自注意力头,再将各头的结果拼接并映射到目标空间。为了在改变头数 h 时保持计算量和参数规模不变。通常将 $\boldsymbol{D}_h$ 设置为 $\boldsymbol{D}/h$。多头注意力计算过程如下。

$$\boldsymbol{M}_{\mathrm{SA}}(z)=[\boldsymbol{S}\mathrm{A}_1(z);\boldsymbol{S}\mathrm{A}_2(z);\cdots\boldsymbol{S}\mathrm{A}_h(z);\cdots;\boldsymbol{S}\mathrm{A}_j(z)]\boldsymbol{U} \tag{4}$$

式中:$\boldsymbol{M}_{\mathrm{SA}}(z)$为输出结果;$\boldsymbol{SA}_h(z)$为第 h 个注意力头的输出;$[\boldsymbol{S}\mathrm{A}_1(z);\boldsymbol{S}\mathrm{A}_2(z);\cdots;\boldsymbol{S}\mathrm{A}_k(z)]$为将 i 个注意力头的输出在特征维度上进行拼接的结果;$\boldsymbol{U}\in\mathbf{R}^{hD_h\times D}$,为线性投影矩阵。

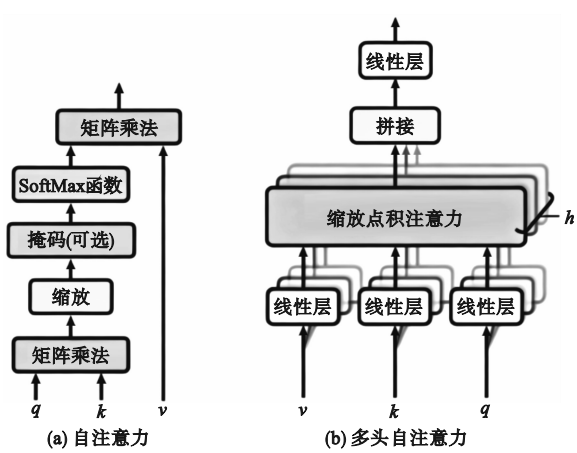


图 2　注意力机制

Fig. 2　Attention mechanism

2　Swin Transformer

ViT 需在大规模数据集训练下才有较高的检测水平,但高效视觉 Transformer [10] 运用多种训练策略使得 ViT 在较小的 ImageNet-1K 数据集上可进行有效训练。但其架构因特征图分辨率低且计算复杂度随图像尺寸呈二次增长,在视觉任务比较密集时难以发挥通用性。

2.1　输入局部化

为减小计算量,Swin-T 模块在保持其他层级结构不变的前提下,采用了一种基于窗口的多头注意力机制,Swin-T 与 ViT 分割检测对比如图 3[11] 所示,将图像划分成若干窗口,仅在窗口内部进行交互。与传统的 ViT 分割检测相比,有效限制了注意力的作用范围,使计算复杂度从传统 Transformer 的 $O(N^2)$降至 $O(N)$。此改进可实现密集图像分类的任务。

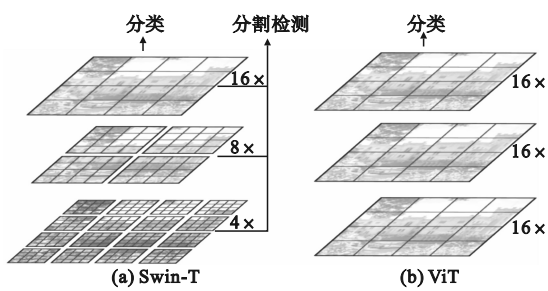


图 3　Swin-T 与 ViT 分割检测对比

Fig. 3　Comparison of Swin-T and ViT for segmentation and detection

2.2　Swin-T 模型结构

Swin-T 架构的整体流程如图 4(a)所示。该架构首先通过图像块划分模块将输入 RGB 图像分割为 4×4 的非重叠图像块,图中 H,W 分别表示输入图像的高度和宽度。每个块的特征由其原始 RGB 像素值直接拼接生成。随后,线性嵌入层

将原始特征映射至预设维度 C,完成从低维像素空间到高维语义空间的转换[12]。在阶段 1 中,多个采用局部窗口 SA 的 Swin-T 块被依次应用于这些块令牌,保持令牌数量恒定,同时通过层级化特征提取逐步捕获图像的语义信息。此阶段输出的特征图将作为后续层级(阶段 2 ~4)的输入,通过块合并操作逐步构建多尺度特征金字塔。

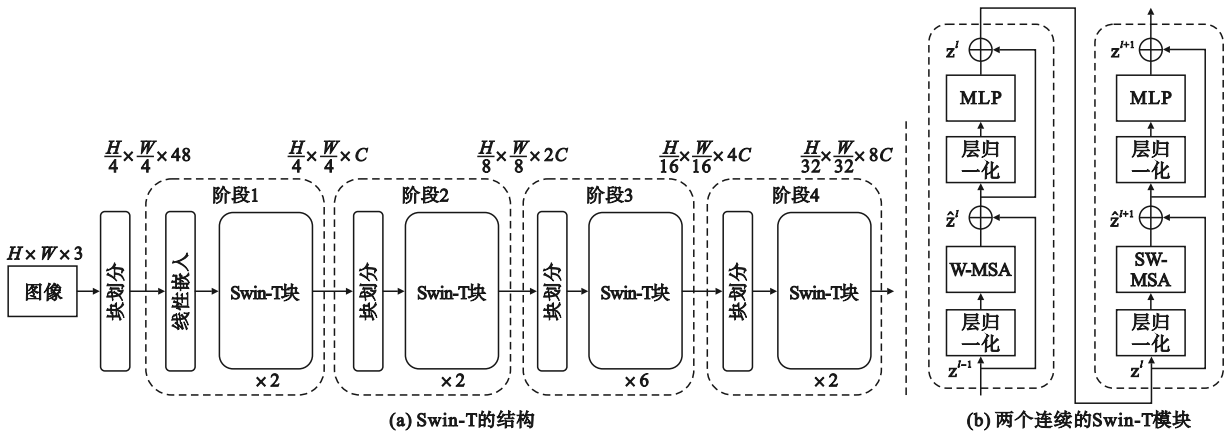


图 4 Swin-T 模型结构

Fig. 4 Swin-T model structure

两个连续的 Swin-T 模块结构如图 4(b)所示,该模块采用级联结构,由基于窗口的多头注意力(W-MSA)、基于移位窗口的多头注意力(SW-MSA)、多层感知机(MLP)与 GeLU 非线性激活函数组成。每个 MSA 模块与 MLP 前均进行层归一化[13]处理,模块间通过残差连接确保梯度高效传递。

3 基于 YOLOv8 引入 Swin-T 的改进模型

YOLOv8 目标检测框架采用模块化结构设计,其整体由主干网络(Backbone)、特征融合层(Neck)以及检测头(Head)三部分构成。在特征提取阶段,输入图像依次经过卷积层(Conv)、跨阶段特征融合模块(C2f)和空间金字塔池化(SPPF),以获取多尺度特征图。不同分辨率的特征被送入改进型特征金字塔网络(PA-FPN),实现多层级的特征融合。该颈部网络通过上采样(Upsample)、拼接(Concat)和特征增强模块构建特征交互通道[14-15]。检测模块采用解耦式无锚框设计,通过分离分类与定位分支有效提升模型鲁棒性,同时实现了参数规模与计算效率的优化。

为应对实际应用中对计算资源与检测精度的差异化需求,YOLOv8 系列提供 N、S、M、L、X 五种不同规格的模型。其中 YOLOv8s 是一个较小的模型,提供的参数数量较少,内存使用量较小。权衡检测目标的准确性与系统实时性能要求,本文选择 YOLOv8s 作为基础框架。

3.1 YOLOv8 改进模型

如图 5 所示为基于 YOLOv8s 模型的改进模型结构,集成了 YOLOv8s、卷积注意力模块(CBAM)、Swin-T 模块,以增强小目标检测的鲁棒性。在骨干网络中增加一个 Swin-T 层以获取多样化图像特征,将原有卷积结构替换为 C2f-ST 的组合结构,在颈部网络引入 CBAM,采用 WIoU 损失函数作为回归损失,并在头部新增一个检测头(Detect)。

3.2 CBAM

为提高模型在复杂背景下对遥感小目标的特征提取能力,在 YOLOv8s 的颈部网络中引入了如图 6 所示的 CBAM 模块。CBAM 是一种轻量级的前馈注意力机制,其核心原理是顺序式地在通道和空间两个独立维度上生成注意力图。通道注意力模块通过对特征图进行全局平均池化和最大池化,利用共享多层感知机生成各通道的权重,从而突出关键通道;空间注意力模块则对通道维度进行池化拼接,经由卷积获得空间权重分布,使模型更加关注目标区域。将 CBAM 集成于颈部网络的目的在于提升多尺度特征表达能力,从而缓解漏检问题。

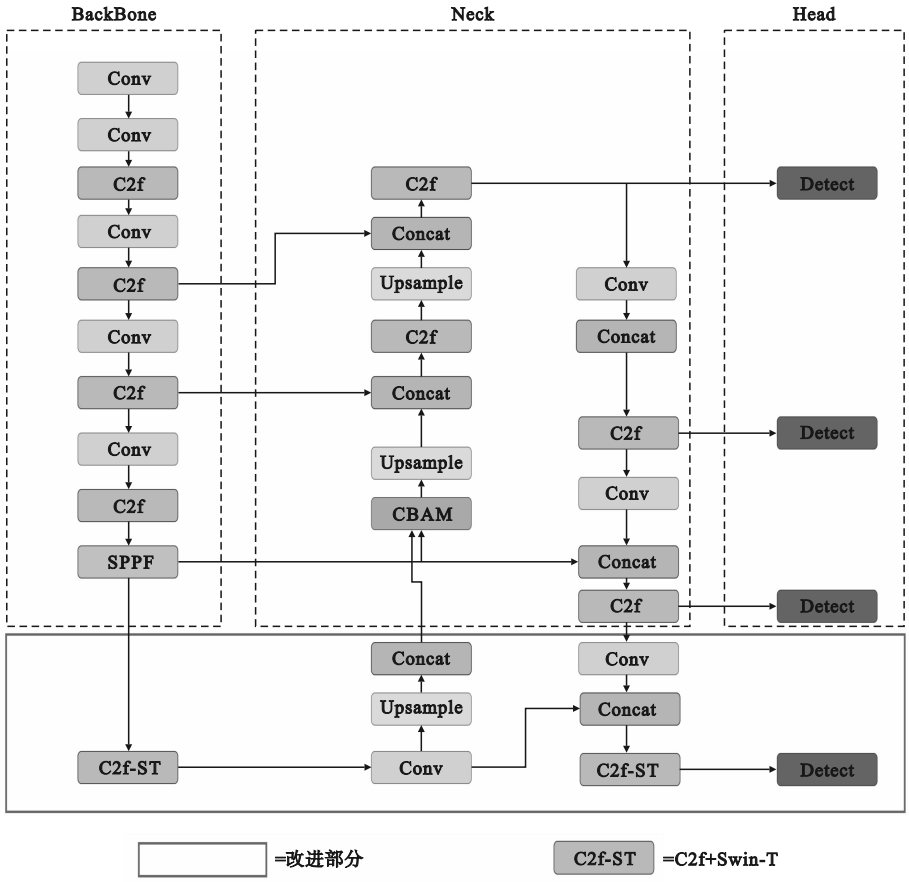


图 5　Yolov8 改进模型网络结构

Fig. 5　Network structure of improved Yolov8 model

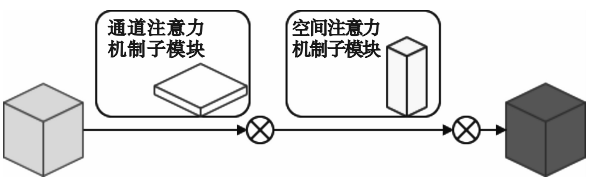


图 6　CBAM 模块结构

Fig. 6　CBAM module structure

3.3　改进的 C2f-ST 结构

对 C2f 模块中的 Bottleneck 进行改进,在第 2 个卷积后面引入 Swin-T,构建 Swin-TBottleneck 模块,结构如图 7 所示。



图 7　Swin-TBottleneck 模块结构

Fig. 7　Structure of the Swin-T Bottleneck Module

为在模型容量与计算效率间取得平衡,将注意力头数设置为 12;窗口大小设置为 7;MLP 扩展比例为 4,与标准 Swin 架构保持一致;激活函数采用 GeLU。

用 Swin-TBottleneck 模块替换 C2f 模块中的 Bottleneck,其结构如图 8 所示。

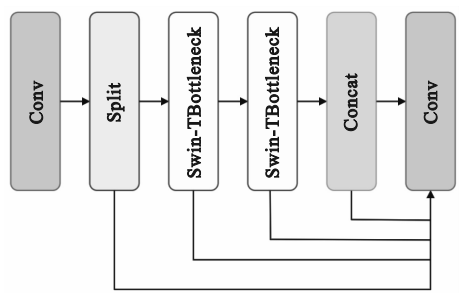


图 8　C2f-ST 模块结构

Fig. 8　Structure of C2f-ST module

该模块构建了 YOLOv8 目标检测框架的多尺度特征融合单元,其核心设计在于整合图像多区域特征提取与信息关联建模能力,充分发挥 Swin-T 的架构优势。C2f 层主要负责基础特征抽取,而 Swin-T 模块专注于建立数据元素间的长程依赖关系。通过二者的协同工作,C2f-ST 模块不仅能高效提取图像特征,还能深度解析特征间的信息关联,从而提升目标检测性能。

C2f-ST 采用层归一化与残差连接机制保障训练稳定性,有效缓解深度网络中的梯度消失问题。在多尺度处理方面,该模块通过接收并融合不同分辨率特征图,构建尺度鲁棒性特征表达,确保对不同尺寸目标具有均衡检测能力。其结构头部通过下采样操作与多尺度特征拼接,经 Swin-T 处理后可获得跨尺度的上下文关联特征。在功能特性上,C2f-ST 展现出显著的位置独立性优势,基于 Transformer 的全局注意力机制,无需依赖目标的空间排列顺序即可有效捕获位置模式,这对复杂场景下的目标定位尤为重要。通过将 Swin-T 的全局关系建模能力与 CSPDarknet 的局部特征提取优势相结合,C2f-ST 模块构建出兼顾局部细节与全局语义的特征学习框架。

3.4 损失函数的改进

YOLOv8 模型在边界框回归中采用完全交并比(CIoU)作为损失函数,但其仅通过宽高比差异衡量预测框与真实框之间的相似性,难以准确反映位置关系,对不同质量样本的梯度贡献缺乏区分,限制了模型性能。为此,引入动态非单调聚焦机制的 WIoU,其通过离群度评估锚框质量,并据此自适应调节梯度权重,减弱高质量和低质量样本对优化过程的影响,使网络更聚焦于中等质量样本的回归,从而提升边界框损失函数的效果。表 2 为不同损失函数的性能对比,其结果显示 WIoU 损失函数可有效提高边界框的收敛质量,提高了模型的检测性能。

表 2 不同损失函数的性能对比

Table 2 Comparison of Different Loss Functions

模型	mAP@0.5	精确率
YOLOv8 + CIoU	0.382	0.496
YOLOv8 + WIoU	0.392	0.504

4 实验与结果分析

4.1 实验环境配置与训练使用的关键参数

本文实验用计算机 CPU 为 i7-9750H,GPU 为 NVIDIA GeForce GTX 1650,内存为 16 GB。实验环境在 Windows11 系统下配置,在 Anaconda 下搭建深度学习框架 Pytorch1.13 和 Python3.10;安装 CUDA11.6 用于提高 GPU 的并行计算能力。迭代次数设置为 200,学习率为 0.01,输入图像的分辨率大小为 640×640,并将分辨率小于等于32×32 的定义为小目标。训练批处理大小为 8,优化器为 SGD,权重衰减为 5×10^{-4},动量为 0.931。数据增强策略中的马赛克增强为 1,最后 10 轮,关闭马赛克增强。本文实验利用模型大小、每张图像的检测速度、精确率(Precision)、召回率(Recall)、平均精度均值(mAP)评估模型的检测能力。

4.2 数据集

为验证改进后模型的检测性能,本文使用开源的 Visdrone2019 数据集。该数据集包含行人、车辆等 10 个不同类别数据。VisDrone2019 数据集具有数据密集分布且包含大量极小目标,具有多尺度、多场景和多视角的独特属性。在实验中,将该数据集的 10 209 幅静态图像按比例划分为训练集、验证集和测试集,分别占 70%、20% 和 10%。

4.3 消融实验

在 VisDrone2019 数据集上进行消融实验以验证改进模块的有效性,实验结果如表 3 所示。

由表 3 可知,第 1 组作为基准实验,使用的是原始 YOLOv8 算法。第 2 组实验在 YOLOv8 颈部网络引入 CBAM 模块,模型的参数量仅略微增加,但 mAP@0.5、mAP@0.5:0.95 分别有 0.9% 和0.5% 的提升,说明了CBAM在小目标特征提

表 3 消融实验结果

Table 3 Results of the ablation experiments

模型	CBAM	C2f-ST	WIoU	mAP@0.5	mAP@0.5:0.95	参数量/10^6
1	—	—	—	0.382	0.221	11.1
2	√	—	—	0.391	0.226	11.2
3	—	√	—	0.411	0.236	13.6
4	—	—	√	0.392	0.227	11.1
5	√	√	—	0.419	0.241	13.7
6	√	√	√	0.428	0.243	13.7

取方面的有效性。第 3 组实验验证了 C2f-ST 模块的有效性,在 mAP@0.5、mAP@0.5:0.95 分别有 2.9% 和 1.5% 的提升。第 4 组实验使用 WIoU 函数,在不增加计算资源的情况下,降低了模型的错检、漏检问题。第 5 组是叠加实验,第 6 组实验是本文提出的模型(YOLOv8s-ST),在 mAP@0.5、mAP@0.5:0.95分别达到了最佳性能,分别为 42.8% 和 24.3% 。

4.4　改进模型与 YOLOv8s 对比实验

本文提出的 YOLOv8s-ST 与 YOLOv8s 在训练阶段关键评估指标的变化曲线如图 9 所示。实验结果表明,所提出的模型在训练轮数达到 120 左右时趋于稳定,在精确率、召回率和 mAP@0.5 检测指标上均优于 YOLOv8s。

4.5　不同算法对比实验

为验证所提出模型的有效性,将 YOLOv8s-ST 与 YOLOv8n、YOLOv8s、YOLOv8m 在 VisDrone2019 验证集上进行对比研究,表 4 给出了在验证集上的实验结果。改进后的模型与 YOLOv8s 相比,在 Precision、Recall 以及 mAP@0.5 三项指标上分别提升了 3.7% 、3.6% 和 4.6% 。与 YOLOv8m 相比,改进模型在保持较小参数量的情况下,获得了更优的检测性能。

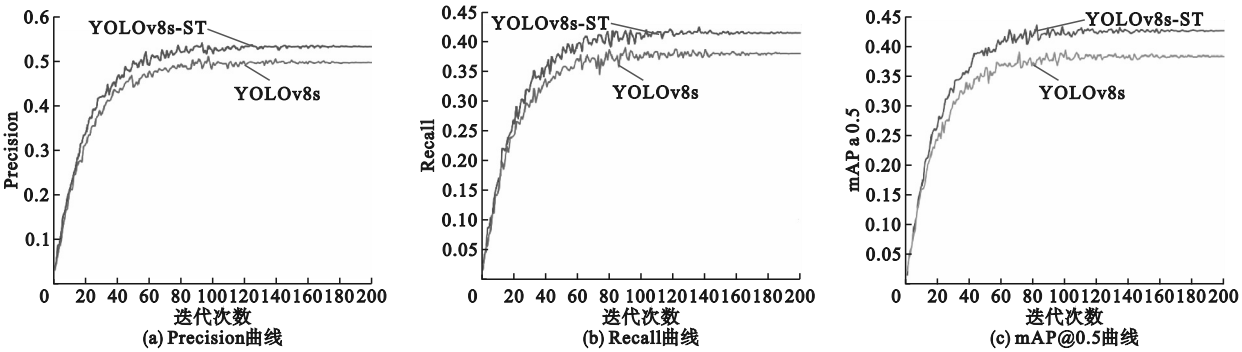


图 9　YOLOv8s-ST 与 YOLOv8s 评估指标变化曲线

Fig. 9　Curves of evaluation metrics changes for YOLOv8s-ST and YOLOv8s

表 4　四种模型检测结果对比

Table 4　Comparison of detection results of four models

模型	参数量/10^6	Precision	Recall	mAP@0.5	时间/ms
YOLOv8n	6.2	0.435	0.323	0.324	4.3
YOLOv8s	22.5	0.496	0.378	0.382	6.1
YOLOv8m	49.2	0.523	0.407	0.414	11.6
YOLOv8s-ST	26.3	0.533	0.414	0.428	6.5

4.6　可视化结果对比实验

图 10 为 YOLOv8s 与 YOLOv8s-ST 部分检测结果的可视化对比,可以看到改进模型相较于 YOLOv8s 在遥感小目标检测中更好地降低了误检与漏检的发生。实验结果表明本文提出的模型能够有效提升小目标的检测准确率。



图 10　YOLOv8s 与 YOLOv8s-ST 部分检测结果对比

Fig. 10　Comparison of partial detection results between YOLOv8s and YOLOv8s-ST

5　结论

本文在 YOLOv8 架构中引入 Swin-T 层,增强了模型识别图像中不同区域间关系的能力,同时在颈部引入 CBAM 模块,使模型更加关注目标区域,从而提升检测精度。Transformer 的自注意力机制使模型在进行目标预测时能够衡量图像不同区域的重要性,从而捕捉到更大范围的全局依赖关系,而不仅仅停留在局部特征层面。实验结果表明,改进后的模型在遥感小目标检测精度方面表现更优,并且在多目标及复杂背景的场景下也能实现更为准确的识别。改进模型的参数量较小,能够达到较好的实时检测效果,在计算资源有限的遥感检测领域有良好的应用前景。

参考文献(References):

[1] 冯迎宾,刘艾妮. 基于 YOLOv8n 的夜间车辆检测[J]. 沈阳理工大学学报,2025,44(2):1 -6,12.
Feng Y B, Liu A N. Night-time vehicle detection based on YOLOv8n[J]. Journal of Shenyang Ligong University,2025,44(2):1 -6,12. (in Chinese)

[2] 李佳兴,文峰. 基于改进 YOLOv8 的细长物体检测方法[J]. 沈阳理工大学学报,2024,43(4):13 -18,26.
Li J X, Wen F. Method of detecting slender object based on improved YOLOv8[J]. Journal of Shenyang Ligong University,2024,43(4):13 -18,26. (in Chinese)

[3] Hui Y M, Wang J, Li B. DSAA-YOLO:UAV remote sensing small target recognition algorithm for YOLOV7 based on dense residual super-resolution and anchor frame adaptive regression strategy[J]. Journal of King Saud University-Computer and Information Sciences,2024,36(1):101863.

[4] 张腾泽,李旭军,饶立明. 改进 YOLOv5 的遥感图像小目标检测算法[J]. 计算机时代,2023(12):89 -95.
Zhang T Z, Li X J, Rao L M. Improved YOLOv5 for remote sensing image small target detection[J]. Computer Era,2023(12):89 -95. (in Chinese)

[5] 秦伦明,梅温泉,崔昊杨,等. 改进 YOLOv8s 的遥感图像目标检测算法[J]. 激光与光电子学进展,2025,62(10):1028003.
Qin L M, Mei W Q, Cui H Y, et al. Improved YOLOv8s object detection algorithm for remote sensing image[J]. Laser & Optoelectronics Progress,2025,62(10):1028003. (in Chinese)

[6] 万辉耀,张朝琛,陈杰,等. 基于 VG-YOLO 的遥感图像车辆检测[J]. 现代应用物理,2025,16(3):195 -201.
Wan H Y, Zhang(C/Z)C, Chen J, et al. Vehicle detection in remote sensing image based on VG-YOLO algorithm[J]. Modern Applie Physics,2025,16(3):195 -201. (in Chinese)

[7] 侯林静,李芳. 基于改进 EAF-YOLO 的轻量化遥感图像检测[J]. 沈阳理工大学学报,2025,44(2):7 -12.
Hou L J, Li F. Lightweight remote sensing image detection based on improved EAF-YOLO[J]. Journal of ShenYang Ligong University,2025,44(2):7 -12. (in Chinese)

[8] Wu J F, Hu R, Xiao Z H, et al. Vision Transformer-based recognition of diabetic retinopathy grade[J]. Medical Physics,2021,48(12):7850 -7863.

[9] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用,2021,57(20):64 -72.
Zhang C J, Zhu L, Yu L. Review of attention mechanism in convolutional neural networks[J]. Computer Engineering and Applications,2021,57(20):64 -72. (in Chinese)

[10] Deng X, Fang X B, Huang G J, et al. Efficient vision transformer:application of data-efficient image transformer for aero engine bearing fault classification[J]. Signal, Image and Video Processing,2025,19(7):515.

[11] Liu Z, Lin Y T, Cao Y, et al. Swin transformer:hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision(ICCV). Montreal, QC, Canada:IEEE,2021:9992 -10002.

[12] Shuai Y Z, Shi J S, Li Y, et al. YOLO-SW:a real-time weed detection model for soybean fields using swin transformer and RT-DETR[J]. Agronomy,2025,15(7):1712.

[13] Kim R, Lee D, Kim J, et al. Hardware accelerator for approximation-based softmax and layer normalization in transformers[J]. Electronics,2025,14(12):2337.

[14] Lou H T, Duan X H, Guo J M, et al. DC-YOLOv8:small-size object detection algorithm based on camera sensor[J]. Electronics,2023,12(10):2323.

[15] Liu Z C, Zhang R S, Zhong H, et al. YOLOv8 for fire and smoke recognition algorithm integrated with the convolutional block attention module[J]. Open Journal of Applied Sciences,2024,14(1):159 -170.

(责任编辑:和晓军)

(上接第 76 页)

[5] 黄勇康,梁美玉,王笑笑,等. 基于深度时空残差卷积神经网络的课堂教学视频中多人课堂行为识别[J]. 计算机应用,2022,42(3):736 -742.
Huang Y K, Liang M Y, Wang X X, et al. Multi-person classroom action recognition in classroom teaching videos based on deep spatiotemporal residual convolution neural network[J]. Journal of Computer Applications,2022,42(3):736 -742. (in Chinese)

[6] 杨继海,裴晓芳. 多尺度与跨空间信息聚合网络的水下目标检测[J]. 激光与光电子学进展,2024,61(24):2437005.
Yang J H, Pei X F. Underwater object detection using a multi-scale and cross-spatial information aggregation network[J]. Laser & Optoelectronics Progress,2024,61(24):2437005. (in Chinese)

[7] 陈峤鹰,傅哲,张高翔,等. 基于 Transformer 的单轨吊多模态数据融合决策技术[J]. 煤矿机电,2024,45(3):33 -37.
Chen J Y, Fu Z, Zhang G X, et al. Multimodal data fusion decision technology for overhead monorail based on transformer[J]. Colliery Mechanical & Electrical Technology,2024,45(3):33 -37. (in Chinese)

[8] 苏佳,杨梦凡,张柏杨,等. 改进 RT-DETR 的无人机小目标检测算法[J]. 微电子学与计算机,2025,42(7):114 -125.
Su J, Yang M F, Zhang B Y, et al. Enhancing RT-DETR for small object detection in UAV[J]. Microelectronics & Computer,2025,42(7):114 -125. (in Chinese)

[9] Tang J D, Chen X, Fan L Y, et al. LN-DETR:an efficient Transformer architecture for lung nodule detection with multi-scale feature fusion[J]. Neurocomputing,2025,633:129827.

[10] Lin Z J, Zhu Z L, Guo L L, et al. Disease detection algorithm for tea health protection based on improved real-time detection transformer[J]. Applied Sciences,2025,15(4):2063.

[11] Yuan G J, Song J T, Li J J. IF-USOD:multimodal information fusion interactive feature enhancement architecture for underwater salient object detection[J]. Information Fusion,2025,117:102806.

[12] Ghose D, Desai S M, Bhattacharya S, et al. Pedestrian detection in thermal images using saliency maps[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). Long Beach, CA, USA:IEEE,2019:988 -997.

[13] 陈昊,郭文普,康凯. 通道门控 Res2Net 卷积神经网络自动调制识别[J]. 电讯技术,2023,63(12):1869 -1875.
Chen H, Guo W P, Kang K. Channel gated Res2Net convolutional neural network for automatic modulation recognition[J]. Telecommunication Engineering,2023,63(12):1869 -1875. (in Chinese)

[14] Qiu Y, Liu Y, Zhang L, et al. Boosting salient object detection with transformer-based asymmetric bilateral U-net[J]. IEEE Transactions on Circuits and Systems for Video Technology,2024,34(4):2332 -2345.

[15] Yu N N, Wang J, Shi H, et al. Degradation-removed multi-scale fusion for low-light salient object detection[J]. Pattern Recognition,2024,155:110650.

(责任编辑:和晓军)