

# 基于改进YOLOv13的小目标检测研究

范西程, 李征宇

沈阳建筑大学计算机科学与工程学院, 辽宁 沈阳

收稿日期: 2026年3月3日; 录用日期: 2026年3月27日; 发布日期: 2026年4月3日

## 摘要

针对图像中小目标分布密集、易发生重叠且受背景干扰, 以及类别不平衡导致检测性能下降等问题, 文章提出了一种基于改进YOLOv13的目标检测算法。首先, 在骨干网络和颈部网络中, 引入上下文引导块CGblock模块并替换所有C3K2模块, 形成CG-C3K2模块, 使检测更聚焦小目标区域, 自适应地增强关键信息的表达能力, 克制感受野受限问题, 加强对小目标的特征增强; 其次, 在颈部结构中融合通道-空间协同注意力机制(Channel-Spatial Collaborative Attention Mechanism, CSCAM), 使特征融合过程更加聚焦于有效目标区域, 提升对小目标和密集目标的检测效果; 最后, 采用改进的Wise-PIoU复合损失函数替代原YOLOv13模型中的损失设计, 使得小目标能够更好地回归到真实标注框, 有效提高了训练稳定性与检测精度。基于VOC2007/2012数据集的实验结果表明, 所提出的改进算法较原始YOLOv13模型的mAP@0.5提升了6.6%, 精确率提升了4.7%, 召回率提升了5.8%。研究结果验证了改进的YOLOv13算法在复杂场景下对小目标的检测性能得到了有效提升。

## 关键词

目标检测, YOLOv13, CG-C3K2, 损失函数, 注意力机制

# Research on Small Object Detection Based on Improved YOLOv13

Xicheng Fan, Zhengyu Li

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang Liaoning

Received: March 3, 2026; accepted: March 27, 2026; published: April 3, 2026

## Abstract

To address the challenges of dense small object distribution, severe occlusion, background interference, and category imbalance that degrade detection performance, this paper proposes an improved

**YOLOv13-based object detection algorithm. First, the Contextual Guided Block (CGblock) is introduced into both the backbone and neck networks, replacing all C3K2 modules to form CG-C3K2 modules, which adaptively enhance the representation of critical small-object regions, mitigate receptive field limitations, and strengthen feature expression for small targets. Second, a Channel-Spatial Collaborative Attention Mechanism (CSCAM) is integrated into the neck structure to focus feature fusion on effective target areas, thereby improving detection accuracy for small and densely packed objects. Finally, an improved Wise-PIoU composite loss function replaces the original loss design in YOLOv13, enabling more accurate regression of small objects to ground truth boxes and enhancing both training stability and detection precision. Experimental results on the VOC2007/VOC2012 datasets demonstrate that the proposed algorithm achieves a 6.6% increase in mAP@0.5, along with 4.7% and 5.8% improvements in precision and recall, respectively, compared to the original YOLOv13 model. The results validate the effectiveness of the improved algorithm in enhancing small-object detection performance in complex scenarios.**

## Keywords

Object Detection, YOLOv13, CG-C3K2, Loss Function, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

作为计算机视觉领域的重要研究方向之一, 目标检测旨在从图像或视频中自动识别并定位特定类别的对象。随着计算机硬件性能的提升和深度学习框架的普及, 该技术已在自动驾驶、智能安防、医疗影像分析、智慧家居等领域得到广泛应用。在这一研究体系中, 小目标检测作为关键分支, 关注于对尺寸较小、特征不明显的目标进行精确识别与定位, 旨在解决小目标易被背景干扰、信息表达不足等问题, 从而进一步提升检测算法在复杂场景下的鲁棒性与实用价值。

传统的目标检测算法[1]主要依赖人工设计的图像特征和分类器, 缺乏对复杂场景和多尺度目标的自适应检测能力。近年来, 基于深度学习的目标检测算法[2]逐渐取代传统的目标检测算法。基于深度学习的目标检测算法主要分为两阶段目标检测和单阶段目标检测, 两阶段目标检测分为候选区域生成、特征提取、目标识别等步骤。典型的两阶段目标检测算法有 R-CNN [3]、Fast R-CNN [4]、Faster R-CNN [5]等。单阶段目标检测算法核心是将目标检测任务直接简化为端到端的回归问题, 直接在图像上应用卷积神经网络(CNN)进行特征提取和目标检测, 特点是检测精度高。目前的代表性算法包括 SSD [6]、YOLO [7]、RetinaNet [8]、CenterNet [9]等。由于单阶段目标检测算法不需要预先生成候选区域, 去除了繁琐的数据计算, 减少了计算成本, 加快了处理速度, 相比于两阶段目标检测算法应用更广泛。

近年来, Chi 等人[10]针对多尺度特征融合过程中信息冗余与失真问题, 提出了自适应空间特征融合(Adaptive Spatial Feature Fusion, ASFF)机制。该方法通过学习不同尺度特征的融合权重, 自适应地选择有效特征区域, 从而提升了网络对小目标的感知能力与检测精度。实验结果表明, ASFF 在小尺度目标检测任务中取得了优异的表现。Zhu 等人[11]提出的 Deformable DETR 模型在 Transformer 检测框架中引入了多尺度可变形注意力机制, 使模型能够在关键区域集中注意力, 避免对背景无效信息的冗余计算。该方法有效改善了小目标的定位误差与收敛速度, 成功解决了原 DETR 在小目标检测中性能不足的问题。Wang 等人在 YOLO 系列模型基础上提出 YOLOv7, 采用扩展高效层聚合网络结构(E-ELAN)以增强特征

重用能力, 并结合动态标签分配机制提升训练稳定性。该改进使模型在检测速度不变的情况下有效提升小目标检测精度, 为实时检测算法提供了新的优化方向。Zhao 等人[12]提出了实时检测网络 RT-DETR, 在端到端检测框架中实现了高效特征聚合与轻量级解码结构。该模型在保持 Transformer 检测精度的同时大幅降低了推理时间, 尤其在小目标和密集目标检测中表现出更好的定位精度与特征分辨率。Chen 等人针对实时检测中“速度与精度难以兼顾”的问题, 提出了 YOLOv10 [13]模型。该模型通过高效的主干网络设计与任务一致性优化策略, 实现了结构轻量化与特征表达能力的平衡。在小目标检测任务中, YOLOv10 有效提升了特征解析度与回归稳定性, 为工业级实时检测提供了新的方案。

然而, 针对图像中小目标分布密集、易发生重叠且受背景干扰, 以及类别不平衡导致检测性能下降等问题, 本文提出三项改进措施:

1) 在 YOLOv13 [14]骨干网络和颈部网络中, 引入 CG-C3K2 模块并替换所有 C3K2 模块, 使检测更聚焦小目标区域, 有效缓解了小目标检测中常见的细节丢失与上下文缺失问题, 从而增强对小目标的特征表达能力。

2) 在颈部结构中, 融合注意力机制(CSCAM), 通过建模通道与空间特征间的交互关系, 使融合特征更具判别性与聚焦性, 改善提升复杂背景下的小目标识别精度。

3) 在损失函数设计方面, 引入改进的 Wise-PIoU 复合损失函数, 以优化分类与定位之间的权衡, 使小目标预测框能够更准确地拟合真实边界框, 从而提高模型训练的稳定性与收敛速度。

## 2. 相关工作

### 2.1. YOLOv13 模型

本文以 YOLOv13 为基本模型进行改进, 其主要结构如图 1 所示。

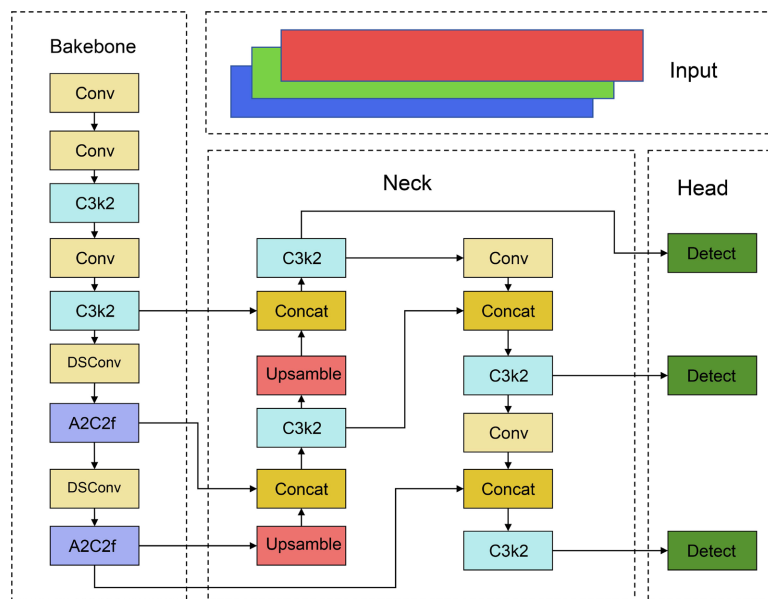


Figure 1. YOLOv13 network architecture

图 1. YOLOv13 网络结构

YOLOv13 模型的整体架构由输入层(Input)、特征提取主干(Backbone)、特征融合层(Neck)以及预测层(Head)组成。在输入阶段, 模型利用 Mosaic 数据增强策略, 使训练过程能够适应不同尺寸的图像, 从而

提升检测的鲁棒性。Backbone 部分基于改进的残差高效层聚合网络(R-ELAN), 该结构在原有 ELAN 的基础上引入残差分支与缩放因子, 有效提升了训练稳定性并降低了额外计算开销。同时, YOLOv13 在将区域注意力机制融入特征提取过程中, 将特征划分为多个子区域进行建模, 以较低的复杂度获取更大感受野并增强全局依赖建模能力。在 Neck 部分, YOLOv13 结合改进的 FPN 与 PAN 结构, 实现多层次特征的高效融合, 使模型能够兼顾语义信息与空间细节, 提升对不同尺度目标的感知能力。Head 部分则采用 Anchor-Free 检测策略, 直接预测目标的中心点及边界框参数, 避免了传统锚框设计带来的额外超参数依赖。YOLOv13 更加侧重于在计算效率与检测性能之间取得平衡。相较于前代模型(如 YOLOv8 与 YOLOv10), 其通过引入轻量化的 R-ELAN 结构与区域注意力机制, 在保证特征表达能力的同时有效控制模型的参数规模与计算复杂度; 同时, Anchor-Free 检测头进一步简化了检测流程, 提高了模型的泛化能力与部署灵活性。然而, 与 YOLOv10 相比, YOLOv13 在基线性能上略有下降, 这主要源于其在网络宽度与深度上的压缩以及轻量化设计对高层语义表达能力的限制。此外, YOLOv13 在标签分配策略与训练机制方面进行了更为系统的优化, 使其在复杂场景下具备更强的特征建模能力与检测精度能力。

## 2.2. 改进 YOLOv13 模型

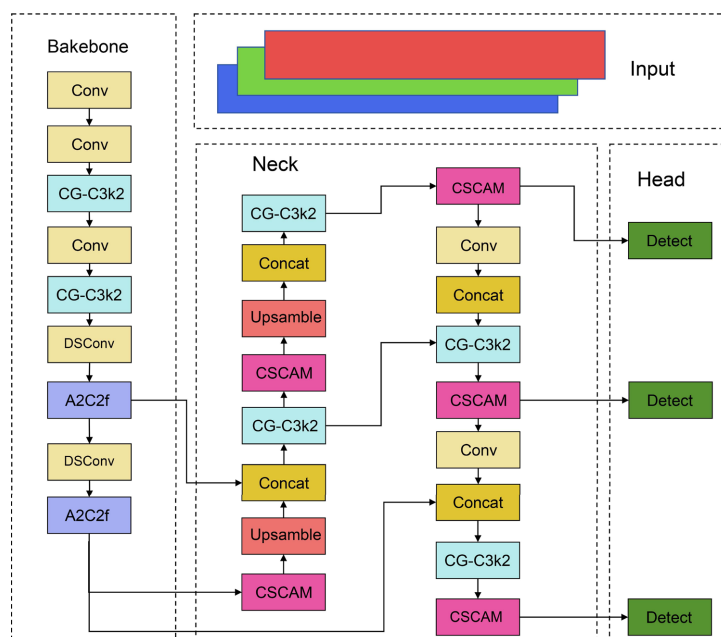


Figure 2. Improved network architecture of YOLOv13

图 2. 改进的 YOLOv13 网络结构

为提高模型对小目标的检测性能, 整体结构由 Backbone-Neck-Head 三部分组成, 如图 2 所示。在 Backbone 部分, 首先通过多层卷积对输入图像进行初步特征提取, 并逐步降低特征图分辨率。在此基础上, 引入 CG-C3K2 模块以增强特征表达能力, 同时结合 DSCov 和 A2C2f 模块提取更丰富的深层语义信息, 在保证计算效率的同时提升网络的特征提取能力。在 Neck 部分, 采用上采样与特征拼接实现不同尺度特征的融合, 并在关键位置引入 CSCAM 注意力机制, 通过通道与空间信息的联合建模增强重要特征响应, 从而提高多尺度特征融合效果, 融合后的特征进一步通过 CG-C3K2 模块进行优化。在 Head 部分, 模型采用多尺度检测结构, 在不同尺度特征图上进行目标分类与边界框回归, 从而减少信息冗余并增强特征表示能力。

### 2.2.1. CG-C3K2 模块

在小目标检测任务中, 因目标尺寸微小、形态多样且在特征图中分布零散, 传统卷积模块常面临特征提取不充分、上下文关联薄弱的困境。固定结构的卷积单元仅能提取局部特征, 难以建立小目标与周边环境、全局场景的语义联系; 同时, 单一的特征融合方式无法兼顾小目标的细粒度细节与多尺度场景信息, 导致检测过程中易出现细节丢失、背景干扰严重的问题, 尤其对于密集场景中的小目标, 传统模块的特征判别性不足, 直接影响检测性能。

为缓解上述问题, 本文在 YOLOv13 网络的骨干与颈部结构中引入 CG-C3K2 模块, 以强化模型对小目标的细节保留能力与上下文关联能力。CG-C3K2 模块如图 3 所示, 其以 C3k2 模块为基础架构, 在瓶颈单元中嵌入上下文引导块(CGblock)。CGblock 模块如图 4 所示, 其通过“局部细节提取(3 × 3 标准卷积) + 周围环境捕捉(3 × 3 低空洞率卷积) + 全局语义融合(全局平均池化 + 多层感知器)”的三层架构, 为不同尺度、不同形态的小目标提供自适应的特征提取策略。与传统 C3k2 模块仅依赖固定卷积核提取特征不同, CG-C3K2 模块通过 CGblock 的动态特征加权机制, 在保留轻量化优势的前提下, 实现局部细节与多尺度上下文的深度关联, 有效提升特征的判别性与鲁棒性。

在 YOLOv13 中, 骨干网络负责多尺度特征的初步提取, 颈部网络承担多尺度特征的聚合融合。通过在骨干与颈部的 C3k2 模块中同步嵌入 CG-C3K2 模块, 网络在浅层可精准捕获小目标的纹理、边缘等细粒度信息, 在深层能将小目标特征与全局场景语义深度融合。经 CG-C3K2 模块处理后的多尺度特征图, 既包含小目标的精准细节, 又具备丰富的上下文关联信息, 为检测头实现小目标的精确分类与定位提供了优质特征支撑。

综上所述, CG-C3K2 模块通过 CGblock 模块“局部 - 周围 - 全局”的三层上下文引导机制, 打破了传统模块特征关联性弱、小目标细节易丢失的局限, 使网络具备更强的小目标特征表达与多尺度关联能力, 从而在小目标检测任务中有效提升检测精度与场景鲁棒性。

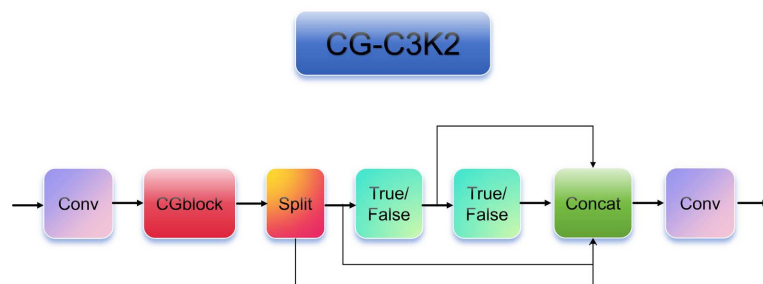


Figure 3. CG-C3K2 module  
图 3. CG-C3K2 模块

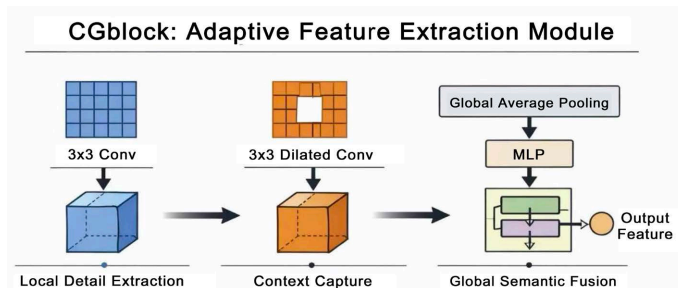


Figure 4. CGblock module  
图 4. CGblock 模块

### 2.2.2. CSCAM 注意力机制

CSCAM 网络结构图如图 5 所示。CSCAM 注意力模块主要由通道注意力模块和空间注意力模块两部分组成，通道注意力模块如图 6 所示，空间注意力模块如图 7 所示。其自适应地细化中间特征图，将输入的图像特征依次遍历特征通道模块和特征空间模块，经过卷积计算输出优化后的图像特征，模型通过串联特征通道和特征空间分配权重，能获取更多有用信息，增加目标识别的准确性。对于输入形状为  $C \times H \times W$  的特征图  $F$  (其中  $C$  表示图像的通道数、 $H$  表示图像的高度、 $W$  表示图像的宽度)，首先分别使用两种全局池化层对其通道进行压缩，获得两个  $1 \times H \times W$  的特征图，并在通道维度上进行拼接，得到形状为  $2 \times H \times W$  的特征图；其次利用感受野较大的  $7 \times 7$  卷积(通道数为 1)，对特征进一步融合；最后使用 Sigmoid 激活函数获得最终形状为  $1 \times H \times W$  的注意力权重，并将输入特征图和注意力权重在空间维度相乘，获得形状为  $C \times H \times W$  的输出特征图，其计算表达式为：

$$\begin{aligned}
 M_s(F) &= F \times \sigma\left(f^{(7 \times 7)}(\text{AvgPool}(F); \text{MaxPool}(F))\right) \\
 &= \sigma\left(f^{(7 \times 7)}\left[F_{\text{Avg}}^S; F_{\text{Max}}^S\right]\right)
 \end{aligned} \tag{1}$$

式中： $M_s(F)$  表示空间注意力输出结果； $F$  表示输入特征图； $\text{AvgPool}$  表示全局平均池化； $\text{MaxPool}$  表示全局最大池化； $f^{(7 \times 7)}$  表示卷积核尺寸为  $7 \times 7$  的卷积； $\sigma$  表示 Sigmoid 激活函数； $F_{\text{Avg}}^S$  和  $F_{\text{Max}}^S$  表示经过全局平均池化和全局最大池化后的特征图。

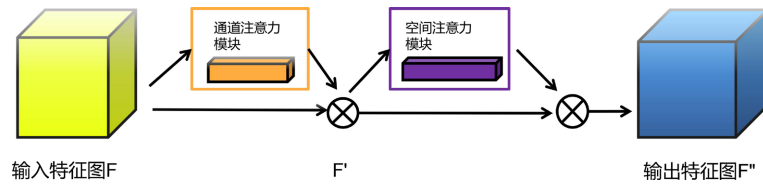


Figure 5. Network structure of CSCAM  
图 5. CSCAM 网络结构图

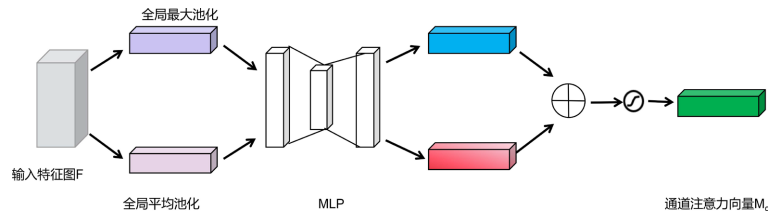


Figure 6. Structure of the channel attention module  
图 6. 通道注意力模块结构图

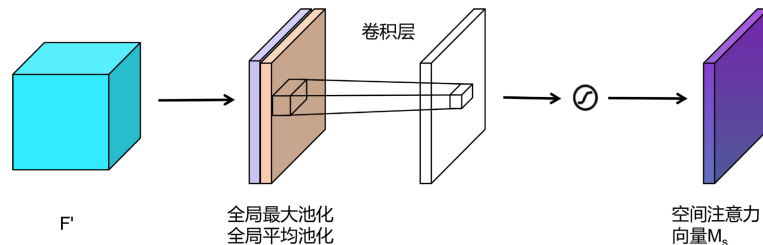


Figure 7. Structure of the spatial attention module  
图 7. 空间注意力模块结构图

在本文中, CSCAM 被嵌入 YOLOv13 的结构中, 用于加强多尺度特征融合阶段的表示能力。通过引入通道和空间的双重注意力机制, 模型能够在特征融合过程中突出小目标与关键目标区域的特征, 从而提升在复杂背景和多尺度场景下的检测精度。实验结果表明, 融合 CSCAM 的改进模型, 其检测性能较基准 YOLOv13 模型有所提升。

### 2.2.3. Wise-PIoU 损失函数

在目标检测任务中, 模型的检测精度在很大程度上取决于损失函数的设计合理性。损失函数不仅决定了模型在训练过程中对分类误差与回归误差的关注权重, 也直接影响模型在复杂场景下的收敛速度与稳定性。YOLOv13 原始损失函数在平衡分类与回归任务时存在一定局限性, 尤其在目标尺寸差异较大或样本分布不均的情况下, 容易出现梯度不稳定和定位精度下降等问题。针对这一不足, 本文在分析现有损失函数特点的基础上, 引入了 Wise-IoUv3 [15]改进型边界框回归损失, 并将其融合形成一种新型复合 Wise-PIoU 损失函数。该损失函数在保持边界框重叠度量准确性的同时, 进一步优化了对小目标与高重叠目标的敏感性, 有效提升了模型在复杂场景下的检测精度与泛化能力。

在目标检测中, 交并比(Intersection over Union, IoU)常用于衡量预测框与真实目标之间的重合程度, 以评估检测结果的准确性。记锚框为  $B = [x \ y \ w \ h]$ , 目标框为  $B_{gt} = [x_{gt} \ y_{gt} \ w_{gt} \ h_{gt}]$ , 如图 8 所示。IoU 定义如式(2)和式(3)所示:

$$S_u = wh + w_{gt}h_{gt} - W_iH_i \tag{2}$$

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{W_iH_i}{S_u} \tag{3}$$

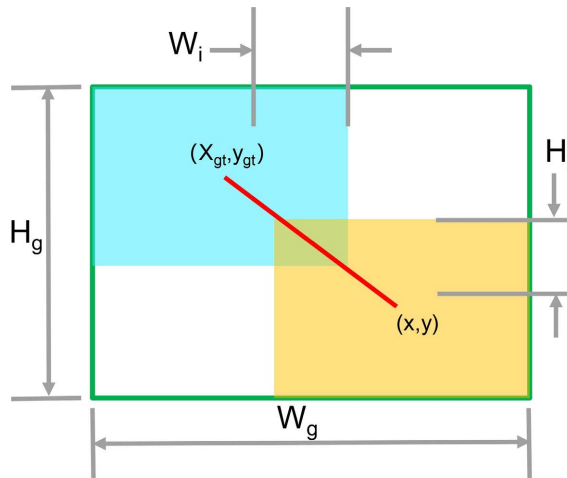


Figure 8. Schematic diagram of IoU  
图 8. IoU 示意图

Wise-IoU 在经典 IoU 的基础上引入了类别权重机制, 从而有效缓解了类别不平衡问题。WIoU 采用动态方法计算类别预测损失中的 IoU 损失, 定义如式(4)和式(5)所示。

$$\mathcal{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{4}$$

$$\mathcal{L}_{WIoU} = \mathcal{R}_{WIoU} \mathcal{L}_{IoU} \tag{5}$$

PIoU 对传统 IoU 中的交并运算进行了改进, 通过动态调整边界框回归损失, 减弱了对距离与纵横比等几何因素的过度惩罚。PIoU 的计算如式(6)所示:

$$\mathcal{L}_{\text{PIoU}} = \mathcal{R}_{\text{wIoU}} \times \mathcal{L}_{\text{IoU}} + \frac{1}{2} \left( \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2} \right) \quad (6)$$

该方法在综合考虑预测框与真实框之间的重叠程度、位置偏移、尺度差异及形状匹配等多维特征的基础上, 更加全面地反映了目标间的空间关系, 从而有效提升了检测精度。同时, PIoU 对损失函数中的第二、三项进行了减半处理, 以较低强度干预模型对几何度量的惩罚, 使得训练过程更加平衡, 有助于提升模型的稳定性与泛化性能。

### 3. 实验结果与分析

#### 3.1. 数据集介绍

本文采用 PASCAL Visual Object Classes (VOC) [16]数据集, 该系列是目标检测、分类与语义分割领域最具影响力的基准之一。其 2007 与 2012 两个版本被学术界广泛采用, 合称 VOC07/12。VOC07 共 9963 张图像, 标注 24,640 个目标实例; VOC12 将训练验证集扩充至 11,540 张图像, 27,450 个实例, 并提供了更精确的边界框与像素级分割标注。该数据集具有常见的 20 个物体类别, 诸如汽车、人、飞机等, 以及一些不太常见的类别, 如椅子、鸟、瓶子等。数据集涵盖 20 类常见目标(人、动物、交通工具及室内物品), 实例分布均衡、视角多样, 已成为验证新检测算法泛化能力的标准测试数据集。

#### 3.2. 实验环境配置

本实验建立在 Windows 操作系统中, 所有实验是基于 Pytorch 和 Python 实现的, 实验环境 GPU 采用 RTX3090(24G), 环境为 Pytorch 2.5.4, python 3.12, CUDA 版本为 12.4, 模型初始图像大小为 640 × 640, 模型迭代次数为 300 轮, Batch Size 设置为 8, 学习率为 0.01。前 250 轮使用 Mosaic 数据增强, 后 50 轮不再使用。

#### 3.3. 评估标准

为了客观评估模型的性能, 检测精度评价指标选择平均精度(Average Precision, AP)和类别平均精度(Mean Average Precision, mAP)。其中 mAP 是衡量检测精度的重要指标, 能够综合反应模型在不同类别上的检测效果。对于检测结果, 本文采用以下规则进行正负样本的判定: 当检测框与真实标注框的 IoU 值大于或等于 0.5 时, 判定该检测框为真阳性(True Positive, TP); 当 IoU 值小于 0.5 时, 判定该检测框为假阳性(False Positive, FP)。同时, 未被检测到的真实标注框被记为假阴性(False Negative, FN)。基于上述定义, 可以计算出每个类别的精确率(Precision, P)和召回率(Recall, R), 其计算公式如式(7)、式(8)所示:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

平均精度(Average Precision, AP)是精确率 - 召回率曲线下的面积, 反映了模型在单一类别上的检测性能。AP 的计算公式如式(9)所示:

$$AP = \int_0^1 p(r) dr \quad (9)$$

最后, 通过对所有类别(假设共有  $B$  个类别)的 AP 值取平均, 即可得到模型的平均精度均值(mAP), 其计算公式如式(10)所示:

$$mAP = \frac{1}{B} \sum_{i=1}^B AP_i \quad (10)$$

其中,  $AP_i$  表示第  $i$  个类别的平均精度。mAP 值越高, 表明模型的整体检测性能越好。mAP@0.5 表示在 IoU 阈值为 0.5 时的平均精度值。此外, 为全面评估模型的复杂度与实际部署性能, 本文引入参数量(Parameters)和计算量(GFLOPs)作为辅助评价指标。其中, 参数量表示模型中所有可学习参数的总数, 用于衡量模型的存储开销与表示能力; 计算量以浮点运算次数(Floating Point Operations, FLOPs)表示, 并以 GFLOPs 为单位, 用于反映模型在前向推理过程中的计算复杂度。

### 3.4. 对比实验

**Table 1.** Comparison of experimental results

**表 1.** 对比实验结果

模型	精确率%	召回率%	mAP@0.5%	Params (M)	FLOPs (G)
Faster R-CNN	70.2	56.2	61.8	41.5	135.0
YOLOv5s	71.5	57.7	63.5	7.2	16.5
YOLOv6n	72.3	58.4	64.2	4.5	11.0
YOLOv7n	73.7	59.5	65.1	6.2	13.2
YOLOv8s	74.8	60.2	66.8	11.2	28.6
YOLOv10n	75.1	60.5	67.2	5.8	13.2
本文方法	76.7	61.1	68.6	6.0	12.5

为了评估本文方法在目标检测任务中的性能, 本研究将其与当前主流检测算法, 包括 Faster R-CNN、YOLOv5 [17]、YOLOv6 [18]、YOLOv7 [7]、YOLOv8 [19] 以及 YOLOv10, 进行了对比实验, 比较指标包括精确率(Precision)、召回率(Recall)、mAP@0.5、参数量(Parameters)和计算量(GFLOPs)。实验结果如表 1 所示。可以看出, 本文方法在参数量与计算量基本保持轻量化的前提下, 实现了更优的检测精度。在精确率、召回率和 mAP@0.5 三个指标上均取得了优异的成绩, 其中精确率达到 76.7%, 召回率为 61.1%, mAP@0.5 达到 68.6%, 均高于其他对比方法。与 YOLOv10n 相比, 本文方法在 mAP@0.5 上提升约 2 个百分点, 说明本文方法在检测精度和对少量样本的泛化能力上具有明显优势。同时, 本文方法在保持高精度的同时, 召回率也表现稳定, 进一步证明其对目标的覆盖能力良好。整体实验结果表明, 本文方法能够在保证检测精度的前提下有效提升目标检测性能, 尤其适用于少样本或小目标检测场景, 具有较强的实用价值和推广潜力。

### 3.5. 消融实验

为验证本文所提改进算法的有效性, 对各改进模块进行消融实验。实验之间严格控制模型参数一致, 其中“√”表示在模型中加入该模块。实验结果如表 2 所示。YOLOv13 算法的 mAP@0.5 为 62%, 精确率为 72%, 召回率为 55.3%。单独在 Backbone 中引入 CG-C3K2 模块后, mAP@0.5 提升至 62.8%, 精确率和召回率分别提升至 73.2% 和 56.3%; 在此基础上使用 Wise-HIoU 损失函数后, mAP@0.5 提升至 63.7%, 精确率下降 1.1%, 召回率提升 0.5%; 同时, 在 Backbone 部分引入 CG-C3K2 模块和 Neck 部分引入 CSCAM

模块后,  $mAP@0.5$  提升至 66.5%, 精确率和召回率分别提升至 76% 和 59.3%; 单独引入 CSCAM 模块并且使用 R Wise-HIoU 损失函数后,  $mAP@0.5$  提升 1.2%, 精确率下降 0.8%, 召回率提升 1.1%; 最后, 当同时引入 CG-C3K2 模块和 CSCAM 模块并且使用 Wise-HIoU 损失函数后,  $mAP@0.5$  提升 0.9%, 精确率提升 1.5%, 召回率提升 0.7%。由此看见, 改进后的 YOLOv13 算法较原算法精度  $mAP@0.5$  提升 6.6%, 达到 68.6%。

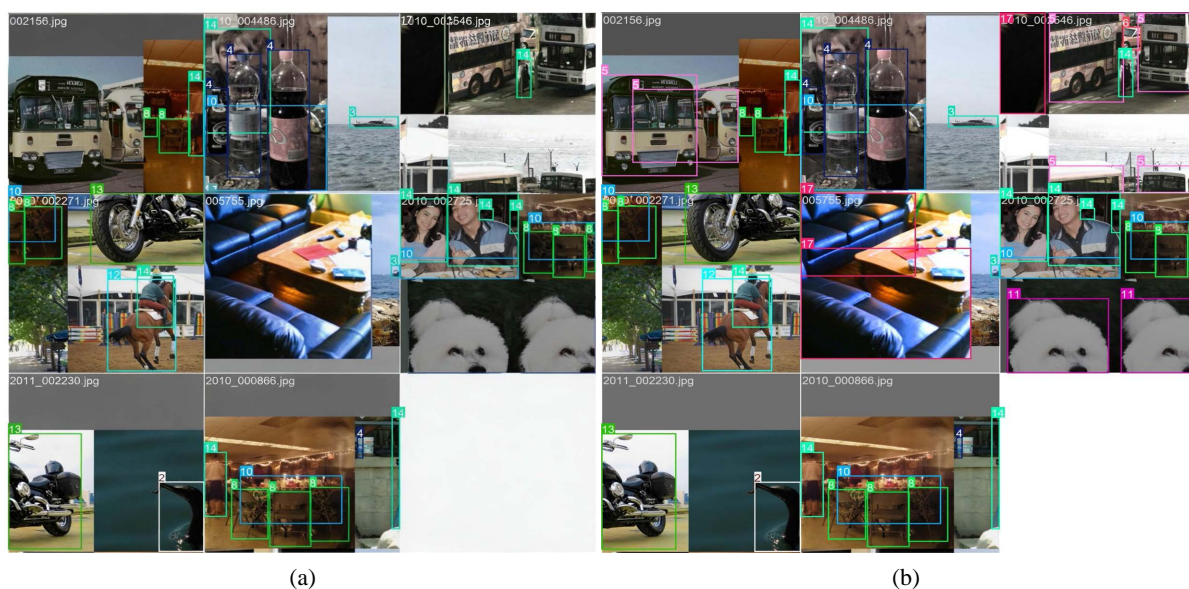
**Table 2.** Results of ablation experiments

**表 2.** 消融实验结果

模型	CG-C3K2	Wise-HIoU	CSCAM	精确率%	召回率%	$mAP@0.5$
1				72	55.3	62
2	✓			73.2	56.3	62.8
3	✓	✓		72.1	56.8	63.7
4	✓		✓	76	59.3	66.5
5		✓	✓	75.2	60.4	67.7
6	✓	✓	✓	76.7	61.1	68.6

### 3.6. 定性分析

图 9 为 YOLOv13 和本文改进 YOLOv13 在 VOC 数据集上检测效果对比图。图中每个小框代表检测到的区域, 由此可以看出, 当面对目标密集的图像时, YOLOv13 常出现漏检问题; 经过改进的 YOLOv13 提升了检测框的置信度, 同时有效缓解了漏检, 展现出更强的检测性能。



**Figure 9.** Comparison of detection effects before and after improvement

**图 9.** 改进前后检测效果对比

## 4. 总结和未来工作

本文针对图像中小目标分布密集、易重叠且受背景干扰, 以及类别不平衡导致检测性能下降等问题,

提出了一种基于改进 YOLOv13 的高效目标检测算法。在改进过程中, 通过在骨干网络和颈部网络中, 引入上下文引导块 CGblock 模块并替换所有 C3K2 模块, 形成了 CG-C3K2 模块, 使网络能够根据目标的形态与尺度自适应调整感受野范围, 从而强化关键信息的表达能力, 提升对小目标的感知效果; 结果显示, mAP@0.5 提升至 62.8%, 精确率和召回率分别提升至 73.2% 和 56.3%。在颈部结构中融合通道-空间协同注意力机制(CSCAM), 使特征融合过程更加聚焦于目标有效区域, 增强了模型对背景复杂下目标的检测能力, mAP@0.5 提升至 66.5%, 精确率和召回率分别提升至 76% 和 59.3%。此外, 采用改进的 Wise-PIoU 复合损失函数, 使得小目标能够更好地回归到真实标注框, 提高了训练稳定性与预测精度。实验表明, 改进模型相比原 YOLOv13 算法, mAP@0.5 提升了 6.6%, 精确率提升了 4.7%, 召回率提升了 5.8%。这充分证明了引入 CG-C3K2 模块与 CSCAM 注意力机制的有效性, 以及改进损失函数对检测任务整体性能的促进作用。尽管本文在 VOC 数据集上取得了较好的实验结果, 但模型的泛化能力可能受限于特定的数据集, 未来工作将验证模型在其他数据集上的表现, 以进一步增强模型的鲁棒性和实用性, 提升模型的泛化能力。

## 参考文献

- [1] Zhao, Z., Zheng, P., Xu, S. and Wu, X. (2019) Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 3212-3232. <https://doi.org/10.1109/tnnls.2018.2876865>
- [2] 张阳婷, 黄德启, 王东伟, 等. 基于深度学习的目标检测算法研究与应用综述[J]. *计算机工程与应用*, 2023, 59(18): 1-13.
- [3] Gao, S. and Liu, H. (2022) Retinanet-Based Compact Polarization SAR Ship Detection. *IEEE Journal on Miniaturization for Air and Space Systems*, **3**, 146-152. <https://doi.org/10.1109/jmass.2022.3203214>
- [4] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/iccv.2015.169>
- [5] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., et al., Eds., *Computer Vision—ECCV 2016*, Springer International Publishing, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [7] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 18-22 June 2023, 7464-7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
- [8] Lin, T.Y., Goyal, P., Girshick, R., et al. (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://arxiv.org/abs/1708.02002v2>
- [9] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q. and Tian, Q. (2019) CenterNet: Keypoint Triplets for Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27-28 October 2019, 6569-6578. <https://doi.org/10.1109/iccv.2019.00667>
- [10] Zhang, S., Chi, C., Yao, Y., et al. (2020) Bridging the Gap between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 9759-9768. <https://arxiv.org/abs/1912.02424>
- [11] Zhu, X., Su, W., Lu, L., et al. (2020) Deformable DETR: Deformable Transformers for End-to-End Object Detection. *9th International Conference on Learning Representations, ICLR 2021*, 3-7 May 2021. <https://arxiv.org/abs/2010.04159>
- [12] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>
- [13] Chen, H., Chen, K., Ding, G., Han, J., Lin, Z., Liu, L., et al. (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems 37*, Vancouver, 10-15 December 2024, 107984-108011. <https://doi.org/10.52202/079017-3429>
- [14] Lei, M., Li, S., Wu, Y., et al. (2025) YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive

- 
- Visual Perception. <https://arxiv.org/abs/2506.17733>
- [15] Tong, Z., Chen, Y., Xu, Z., *et al.* (2023) Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. <https://arxiv.org/abs/2301.10051>
- [16] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J. and Zisserman, A. (2009) The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, **88**, 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [17] Khanam, R. and Hussain, M. (2024) What Is YOLOv5: A Deep Look into the Internal Features of the Popular Object Detector. <https://arxiv.org/abs/2407.20892>
- [18] Li, C., Li, L., Jiang, H., *et al.* (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. <https://arxiv.org/abs/2209.02976>
- [19] Reis, D., Kupec, J., Hong, J., *et al.* (2023) Real-Time Flying Object Detection with YOLOv8. <https://arxiv.org/abs/2305.09972>