

基于多尺度特征融合的树木检测算法——DDC-YOLO

张志杰, 王庆*

(东南大学 仪器科学与工程学院, 南京 210096)

摘 要: 针对树木检测存在的遮挡干扰和光照不充分等问题, 提出一种基于改进 YOLOv10 的检测算法——DDC-YOLO, DDC 是由 dynamic、DAF (dynamic align fusion) 与 CG-FPN (context guided-feature pyramid network) 首字母拼写。首先, 设计动态卷积混合模块 (dynamic convolutional mix block, DCMB), 通过自适应动态卷积增强多尺度特征融合能力, 解决传统卷积核单一性问题; 其次, 提出双主干动态特征融合网络, 结合 RT-DETR (real-time detection transformer) 和 YOLOv10 的主干结构, 并利用动态对齐融合 (dynamic align fusion, DAF) 模块调整特征权重, 提升模型对不同特征的适应性; 进一步引入金字塔上下文特征提取和空间特征重建技术优化颈部网络, 实现多层次语义信息的深度融合。试验基于自建数据集 TreeImages (包含 7 475 张图像) 进行验证, 结果表明, DDC-YOLO 的 mAP50 达到 46.7%, 较原 YOLOv10 模型提升 5.0 个百分点, 参数量由 2.27 M 降至 2.26 M (减少 0.44%), 检测速度 (FPS) 由 202 帧/s 提升至 254 帧/s (提升 25.4%)。改进后的模型在复杂场景下表现出更高的鲁棒性和实时性, 为森林资源调查提供了高效解决方案。

关键词: YOLOv10; 目标检测; 计算机视觉; 双主干式动态融合; 多尺度特征重建

中图分类号: S758

文献标识码: A

DOI: 10.7525/j.issn.1006-8023.2026.01.016

Tree Detection Algorithm Based on Multi-scale Feature Fusion——DDC-YOLO

ZHANG Zhijie, WANG Qin*

(School of Instrumentation Science and Engineering, Southeast University, Nanjing 210096)

Abstract: This paper proposes a detection algorithm DDC-YOLO based on improved YOLOv10 to address the issues of occlusion interference and insufficient lighting in tree detection. Firstly, a dynamic convolutional mix block (DCMB) was designed to enhance the multi-scale feature fusion capability through adaptive dynamic convolution, solving the problem of singularity in traditional convolution kernels; Secondly, a dual backbone dynamic feature fusion network was proposed, combining the backbone structures of RT-DETR and YOLOv10, and utilizing the dynamic alignment fusion (DAF) module to adjust feature weights and enhance the model's adaptability to different features; Further introduced pyramid context feature extraction and spatial feature reconstruction techniques to optimize the neck network and achieve deep fusion of multi-level semantic information. The experiment was validated based on the self built dataset TreeImages (containing 7475 images), and the results showed that the mAP50 of DDC-YOLO reached 46.7%, which was 5.0 percentage points higher than the original YOLOv10 model. The parameter size decreased from 2.27 M to 2.26 M (a decrease of 0.44%), and the detection speed (FPS) increased from 202 to 254 (an increase of 25.4%). The improved model exhibits higher robustness and real-time performance in complex scenarios, providing an efficient solution for forest resource surveys.

Keywords: YOLOv10; object detection; computer vision; dual-backbone dynamic fusion; multi-scale feature reconstruction

收稿日期: 2025-04-28

基金项目: 国家自然科学基金 (42074039)。

第一作者简介: 张志杰, 硕士研究生。研究方向为计算机视觉。E-mail: 220233677@seu.edu.cn

* **通信作者:** 王庆, 博士, 教授。研究方向为智慧城市时空信息智能采集。E-mail: w3398a@263.net

引文格式: 张志杰, 王庆. 基于多尺度特征融合的树木检测算法——DDC-YOLO[J]. 森林工程, 2026, 42(1): 170-183.

ZHANG Z J, WANG Q. Tree detection algorithm based on multi-scale feature fusion——DDC-YOLO[J]. Forest Engineering, 2026, 42(1): 170-183.

0 引言

在森林资源管理中,树木调查方法仍以传统人工测量和目测估算为主,这种方式耗时长且效率低。近年来,随着相机与激光雷达等多传感器数据融合技术的发展,树木资源调查逐渐走向自动化与智能化方向^[1],基于计算机视觉和深度学习方法的树木检测和识别成为研究的重点,尤其在复杂自然场景中的树木检测任务中展现出巨大潜力。深度学习检测方法相较于传统基于人工检查的方法,其具备强大的特征表达能力和端到端学习机制,能够自动从大规模数据中提取判别性特征,显著提升目标识别的精度与鲁棒性。借助深度学习在林业资源调查、城市绿化评估以及智慧林业管理等领域来实现树木的高效检测,不仅可以大幅减少人力成本,还能提高检测结果的一致性与可扩展性。因此,研究基于深度学习的高效树木目标检测算法在实际应用中具有广阔的前景。

基于深度学习的目标检测算法,如Faster R-CNN^[2]、YOLO^[3]和SSD^[4]等,被广泛应用于自然场景中的目标检测任务。现有树木检测方法,如通过图像预处理和空间注意力模块的引入^[5]、改进动态检测头^[6]等技术来进行识别。窦安亮等^[7]引入融合注意力机制的双向加权特征金字塔结构,减少冗余语义特征,来增加树木检测精度,但在实时性方面存在问题。周焱等^[8]在无人机上使用轻量级网络来对小目标树进行检测,减少了参数量,但泛化性不强。YOLO系列算法在目标检测中表现出优越的实时性和准确性,尤其适用于无人机航拍、地面监测和智能巡检等应用场景。近年来,YOLO算法从最初的YOLOv1^[3]发展至最新的YOLOv10^[9],其网络结构不断更新,检测能力不断优化。YOLOv10相较于前代版本,引入了更高效的主干网络、特征融合模块和目标检测头,显著提升了对不同尺度目标的检测能力。然而YOLOv10模型由于多尺度特征融合不足及特征融合拼接简单等问题,在树木检测任务中仍存在一定的局限性,例如对树木冠层的边界识别不精准、树干检测易受光照条件与背景干扰等问题。因此针对YOLOv10模型的改进研究,对提高树木检测精度、增强模型鲁棒性具有重要意义。

朱谨等^[10]通过在C2f模块中引入具有指数移动平均注意力机制的Faster-Block替代原有的残差模块,大幅度减少了计算量和浮点运算量,提高了YOLOv10算法的检测效率。Yu等^[11]提出的InceptionNeXt将Inception模块的多尺度特征提取能力与ConvNeXt的深度可分离卷积机制相结合,使模型能够更有效地捕捉全局与局部的多尺度特征信息。Tan等^[12]提出具有动

态学习权重特征的双向特征金字塔与EfficientNet模块机制结合来优化网络结构,不仅能保持高精度,同时还显著降低了计算成本。Liu等^[13]通过级联分组注意力,使得模型在持有Transformer全局特征建模能力的同时,大幅提升了计算效率,使其更适用于如移动端等资源受限的设备。Chen等^[14]挑战了传统“通过降低FLOPs实现模型加速”的理念,提出应合理提升FLOPs,以更充分地发挥现代硬件的并行计算能力。通过提升计算密集度并缓解访存瓶颈,作者设计了具有更大卷积核的结构以适配当代硬件架构。该策略在实现更快推理速度的同时,也为高性能神经网络的优化设计提供了新的研究方向。Woo等^[15]在ConvNeXt模块机制的基础上进行了优化,通过Masked Autoencoders预训练、自监督学习、优化卷积结构、改进残差连接以及提升模型扩展性等改进,使得卷积神经网络提高计算效率并保证其泛化能力,同时在如移动端和云端等多种计算平台上均能提供高效且高精度的特征提取能力。Li等^[16]针对实时目标检测模型的高效性,提出了一种基GSConv的轻量级颈部设计,减少了目标检测模型的计算复杂度,同时保持了检测精度,该方法提升了实时目标检测的效率,特别适用于低功耗设备和高吞吐量场景,在自动驾驶、智能监控、无人机检测等领域中具有广阔的应用前景。Wang等^[17]通过Gather-and-Distribute Mechanism和Gold-Neck设计,优化了特征提取和融合过程,在减少计算量的同时提高了检测性能。Yang等^[18]针对目标检测任务的精度提升和优化计算效率等问题,提出了一种基于多分支辅助融合机制,并结合重参数化异构卷积技术来提升目标检测的特征表达能力,并优化了计算效率。

综合上述研究可以看出,尽管现有算法在目标检测领域已经取得了一定的进展,但是传统多尺度特征融合多依赖固定权重或简单拼接策略,未能根据输入特征动态调整融合方式,因此本研究用以下方法来提升模型对多尺度特征的适应性。在YOLOv10基础上提出了针对树木检测任务的改进方法,主要包括以下3个方面。

1)设计可自适应平均池化的动态卷积混合模块(dynamic convolutional mix block, DCMB)来替换颈部的C2f模块,以增强模型对不同尺度特征的融合能力。

2)提出双主干式动态特征融合机制来优化主干网络,使用2个主干网络去融合输入的特征,然后根据输入特征图来动态调整2条路径的权重,再去对双主干进行融合拼接,从而增强模型对不同特征的适应能力,提高检测的鲁棒性和精度。

3)引入多层次尺度特征融合机制优化颈部部分,在颈部部分使用金字塔上下文特征提取和空间特征重

建技术。通过动态插值和多层级间信息融合,充分整合高层次的语义信息和低层次的信息,提高了模型的信息整合度和对复杂背景下目标的识别能力。

1 研究方法

1.1 DDC-YOLO 整体结构

本研究提出的基于YOLOv10算法改进模型DDC-YOLO,如图1所示,其结构主要包括3个部分:主干、颈部和检测头。在DDC-YOLO网络中,主干负责从输

入图像中提取低层次的特征信息;颈部负责将来自不同层次的特征进行融合;而检测头层负责最终的目标检测任务。利用自适应平均池化、多尺度特征融合和动态卷积等技术,在颈部将C2f模块替换为本研究设计的DCMB模块,用双主干式动态特征融合结构替换传统YOLOv10的主干网络,并且在颈部引入了金字塔上下文特征提取和空间特征重建结构,来充分整合不同层次的语义信息,提高模型的多尺度特征表示能力。

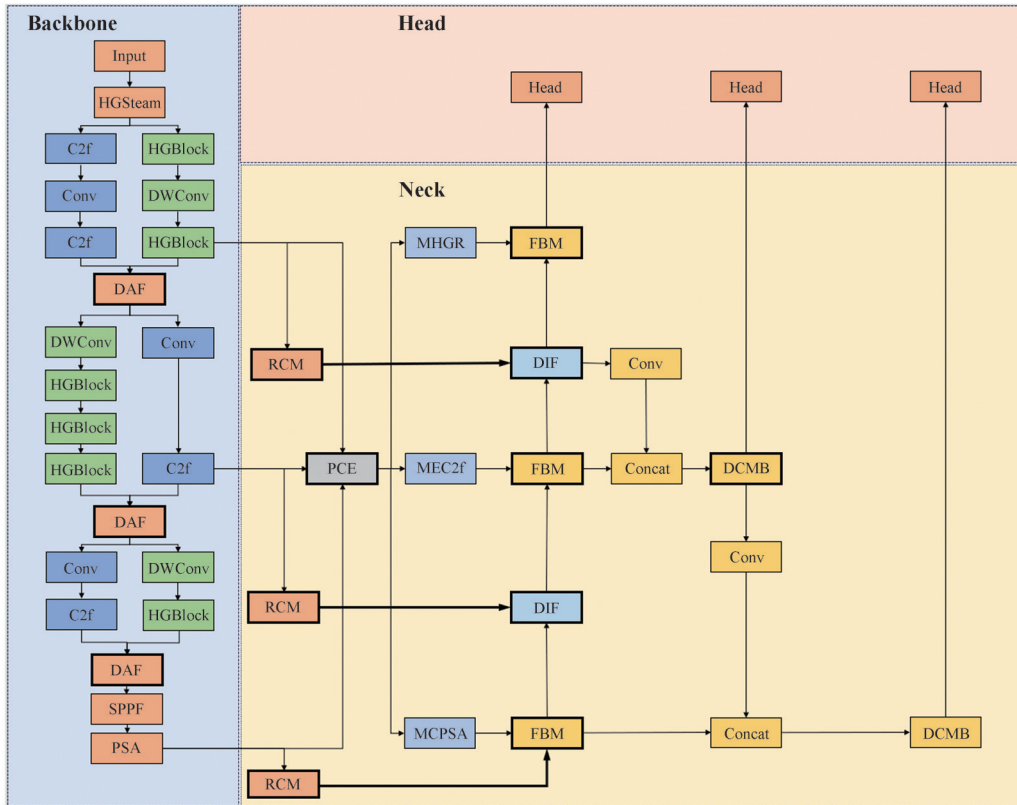


图1 DDC-YOLO 的整体结构
Fig. 1 The overall structure of DDC-YOLO

1.2 DCMB 模块

C2f模块因其卷积核的单一性,无法处理多尺度特征输入。其结构是静态的,不具备根据输入图像内容调整卷积核策略的能力,在面对不同种类或环境下的树木目标时表现出适应性不足的问题。C2f仅通过简单拼接进行特征融合,未显式考虑来自不同尺度的特征融合策略,导致高层语义信息与底层细节信息交互不充分,特别是在光照不足的条件下容易产生特征遗漏或弱化。因此设计DCMB模块,如图2所示,其中包含多个动态初始化混合模块(dynamic inception mixer block, DIMB),通过多个层级的特征处理增强网络的表达能力,解决原有网络结构的信息丢失问题。

DIMB 模块能够根据输入的特征自动调整融合策

略,使得特征融合更加灵活,如图3所示。其输入的特征图,先经过BN层进行归一化操作,再通过动态初始化混合器(dynamic inception mixer, DIM)模块进行通道混合和特征提取,最终与原始输入进行残差连接,如图4所示。然后再使用BN(batch normalization)层进行正则化。随后,特征图再经过卷积门控线性单元(convolutional gated linear units, CGLU^[19])进行非线性变换和通道间交互后,再次残差连接。DIM 模块通过通道拆分与融合,不仅降低了计算复杂度,还能融合不同尺度的特征,增强信息交互能力。

DIM 模块中的动态深度可分离卷积(dynamic depthwise separable convolution, DDSConv)模块如图5所示。

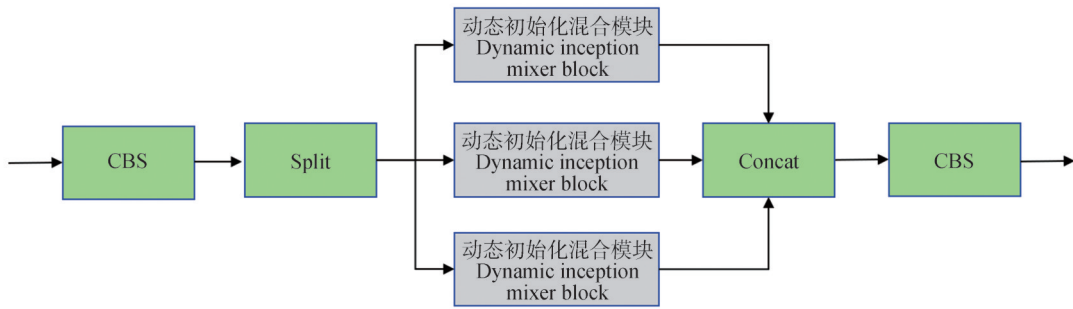


图 2 DCMB 结构

Fig. 2 DCMB structure diagram

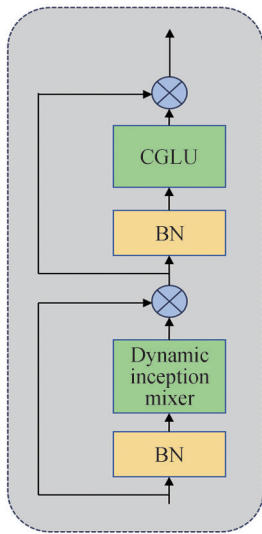


图 3 DIMB 结构

Fig. 3 DIMB structure diagram

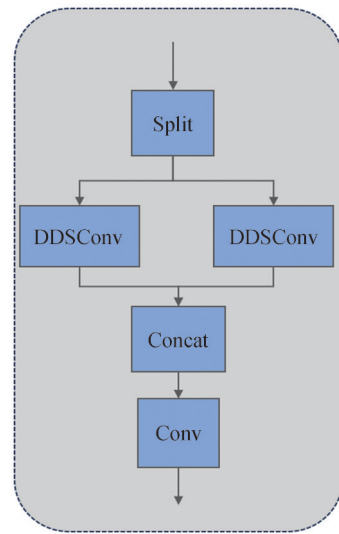


图 4 DIM 结构

Fig. 4 DIM structure diagram

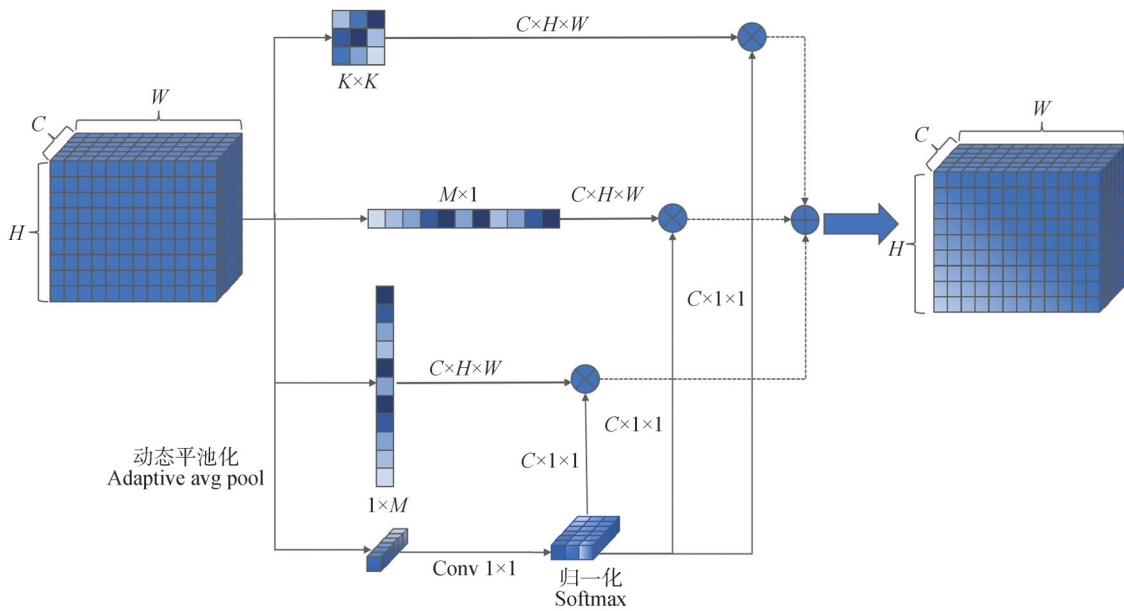


图 5 DDSConv 结构

Fig. 5 DDSConv structure diagram

DDSCConv是一个动态深度卷积模块,该模块能通过不同尺度卷积核和动态自适应调整,实现不同卷积核类型的组合,以应对多种不同的输入特征。具体地,先将输入的三维特征图 $X_i \in \mathbf{R}^{(C \times H \times W)}$ 经过多尺度深度可分离卷积,卷积核的类型包括: $1 \times M$ 横向卷积核、 $M \times 1$ 纵向卷积核和 $K \times K$ 方形卷积核。其中 M 和 K 的关系为: $M = 3 \times K + 2$ 。

在此基础上还引入了自适应动态卷积核权重,三维特征图 X_i 通过全局平均池化计算每个通道的全局特征,公式为

$$X_{\text{GAP}} = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W X_{c,h,w}, X_{\text{GAP}} = \mathbf{R}^{C \times 1 \times 1}. \quad (1)$$

式中: X_{GAP} 为特征图全局平均池化结果; H 为Height,代表图像或特征图的垂直尺寸(行数); W 为Width,代表图像或特征图的水平尺寸(列数); C 为Channels,表示图像或特征图的通道数; h 代表行数累加变量; w 代表列数累加变量。

再通过 1×1 卷积计算每个卷积分支的权重,以获得全局的统计信息,接着使用Softmax激活函数对得到的3个权重系数 α_i 进行归一化,公式为

$$\alpha_i = \frac{\exp(W_d^i)}{\sum_{j=1}^3 \exp(W_d^j)}, i \in \{1, 2, 3\}. \quad (2)$$

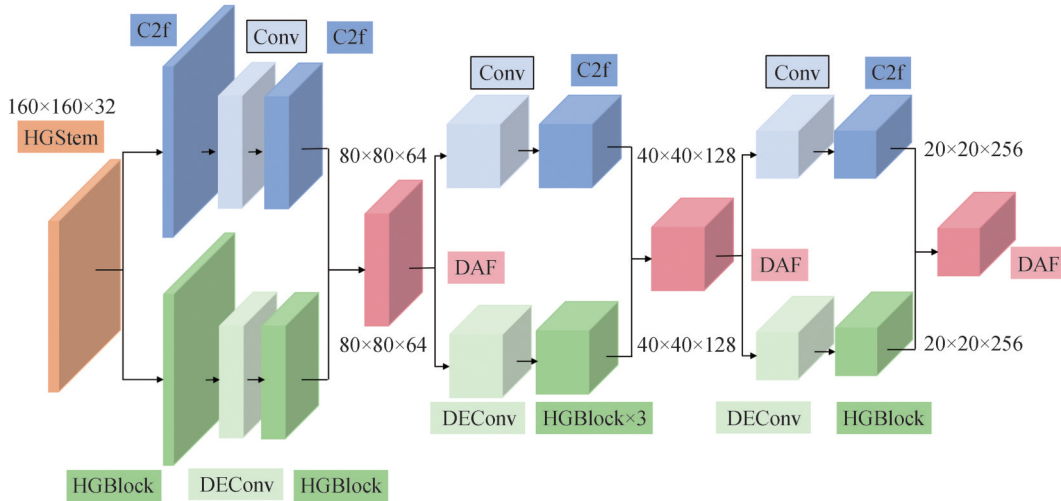


图6 双主干式动态融合网络

Fig. 6 Dual-backbone dynamic fusion network

HGStem模块^[20]首先将输入的特征图进行轻量化卷积操作,然后采用最大池化来提取信息,并减少计算量,最后将2条分支进行融合拼接后,再同时使用2条主干通道。上层延续了原YOLOv10的主干网络结构,下层借鉴了RT-DETR^[21]的网络结构,并在此基础上融入了DAF模块,如图7所示,融合不同层级的特征。DAF的

核心是将2个输入特征图 X'_1, X'_2 进行对齐和动态加权融合。先对输入的2个特征图进行特征对齐操作,过程为

$$X'_1 = \text{Conv}_{\text{align1}}(X_1) \in \mathbf{R}^{C_{\text{out}} \times H \times W}. \quad (4)$$

$$X'_2 = \text{Conv}_{\text{align2}}(X_2) \in \mathbf{R}^{C_{\text{out}} \times H \times W}. \quad (5)$$

其次,拼接 X'_1 和 X'_2 并进行 3×3 卷积,公式为

式中: W_d^i 表示输入特征 x 在第 i 个分支上的权重评分。最后,利用动态的权重系数对卷积核进行加权累加来得到最终的输出特征,公式为

$$Y = \sum_{i=1}^3 \alpha_i Y_{i \circ} \quad (3)$$

本模块通过自适应动态卷积机制,使模型能够根据不同输入图像的特征响应自动调整卷积策略,尤其在对图像中的树木存在光照不足的问题时,能强化关键区域的语义表达,避免因局部特征弱而导致目标漏检。多尺度特征融合结构还可增强对树木冠层和枝叶细节的捕捉能力,提高在复杂背景下的检测稳定性。

1.3 主干改进

YOLOv10的主干网络依赖于传统的上采样和拼接技术来处理多尺度特征的融合拼接问题,然而不同层的特征由于感受野与语义深度存在分布差异,并且浅层网络结构简单,在树木存在遮挡问题时,融合特征容易出现空间或语义信息失真。因此固定加权或简单拼接容易带来信息冗余或冲突,缺乏表达能力,导致高分辨率带来的优势并不能充分发挥。本研究设计了双主干式动态特征融合网络,如图6所示,使用双路径增强特征表达能力,并且利用动态对齐融合(dynamic align fusion, DAF)模块来处理2条主干上的信息差异。

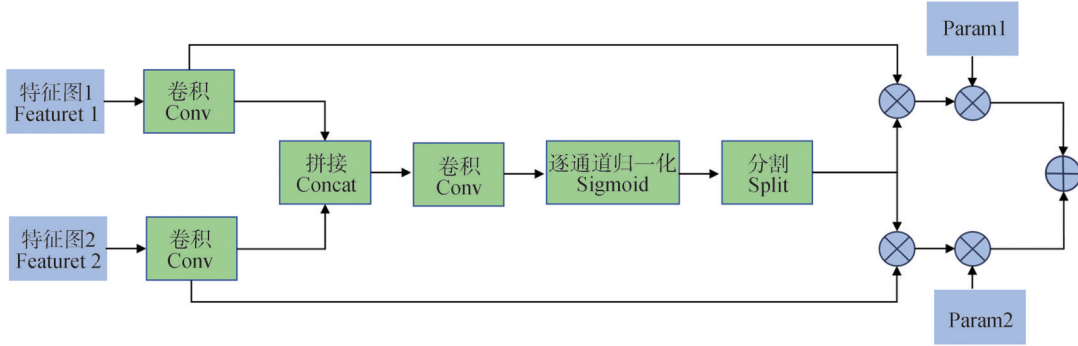


图 7 DAF 结构

Fig. 7 DAF structure diagram

$$X_{concat} = Conv_{concat}(Concat(X'_1, X'_2)) \quad (6)$$

随后,分割权重再计算加权特征,公式为

$$W_1, W_2 = Chunk(W, 2) \quad (7)$$

$$X''_i = X'_i \times W_i, i \in \{1, 2\} \quad (8)$$

最后,引入 2 个可学习的动态参数用于控制融合后特征图的加权比例,分别于 X'_1 和 X'_2 相乘累加后输出最终特征,公式为

$$Y = Conv_{final}(X''_1 \times P_1 + X''_2 \times P_2) \quad (9)$$

式中: X'_1 和 X'_2 分别代表 2 条通路上卷积后的特征图; X''_1 和 X''_2 代表 2 条通路上卷积后的特征图乘以相应权重的结果; P_1 和 P_2 代表引入的参数。

在森林场景中,树木在不同层级上具有显著的尺度与纹理差异,例如远处树干与近处枝叶的纹理特征

差异较大,DAF 模块通过对双主干输出的特征进行动态加权融合,能够自适应调整特征权重,有效应对树木目标在尺度变化和遮挡情况下的特征不一致问题,从而提升整体检测的鲁棒性和准确性。

1.4 颈部改进

针对 YOLOv10 模型颈部特征融合机制,缺乏高效的跨层信息流交互机制,并且信息流动主要是自上而下,未提供额外的横向连接来改善跨层信息交互融合而导致的跨层信息交互效率不足问题。本研究提出基于金字塔上下文提取(pyramid context extraction, PCE)的特征引导金字塔网络,其核心模块 PCE 如图 8 所示。该网络通过多层次特征融合增强策略,显著提升了跨尺度特征的交互能力。

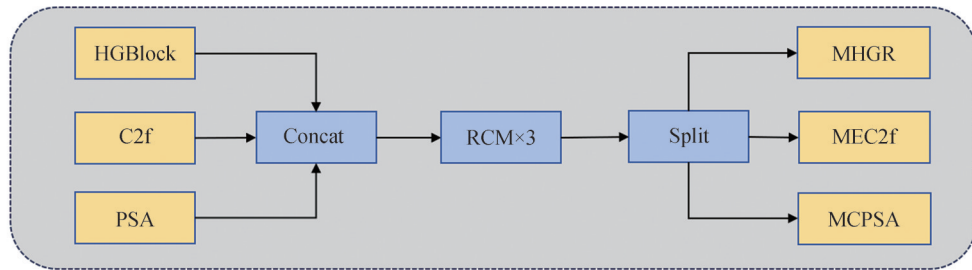


图 8 PCE 结构

Fig. 8 PCE structure diagram

将来自主干部分的 HGBlock、C2f 和 PSA 的 3 个模块进行融合拼接,再借鉴 CGRSeg^[12] 中的设计,引入矩形自校准模块(rectangular self-calibration module, RCM) 机制,如图 9 所示。RCM 采用双路径结构:首先通过深度可分离卷积提取局部空间特征,随后并行执行扩张率为 2 和 4 的扩张卷积以扩展感受野,最后通过通道拆分生成 3 个优化分支——HGBlock 精炼模块(HGR)、增强型 C2f 模块(EC2f)和上下文感知 PSA 模块(CPSA)。经多重 RCM 迭代处理后,进一步生成多级增强特征 MHGR、MEC2f 和 MCPSA。RCM 机制扩大感受野,能捕获遮挡区域外围的上下文信息,弥补局部细节的缺失。

多分支融合模块(fusion block multi-branch, FBM) 用以实现多特征混合,如图 10 所示。首先对 CPSA 与 MCPSA 特征分别进行 1×1 卷积降维,然后采用 h_sigmoid 函数生成 $[0, 1]$ 区间的空间注意力权重,通过逐元素相乘完成通道自适应加权。最后,使用动态插值融合模块(dynamic interpolation fusion, DIF) 对加权特征进行双线性上采样,如图 11 所示,并与 MEC2f 特征进行通道维度相加,实现跨分辨率特征融合。DIF 模块则通过动态插值提升低分辨率特征的表达能,从而增强对树冠、枝叶边界的识别能力,有效提升模型在遮挡和低光环境下的检测能力。

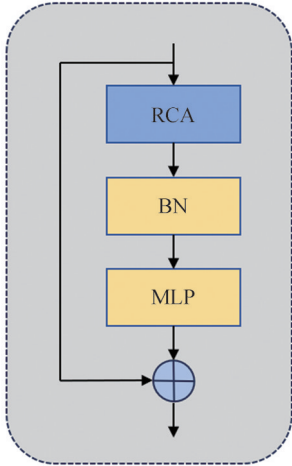


图9 RCM结构

Fig. 9 RCM structure diagram

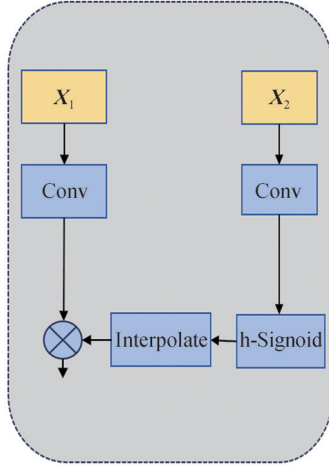


图10 FBM结构

Fig. 10 FBM structure diagram

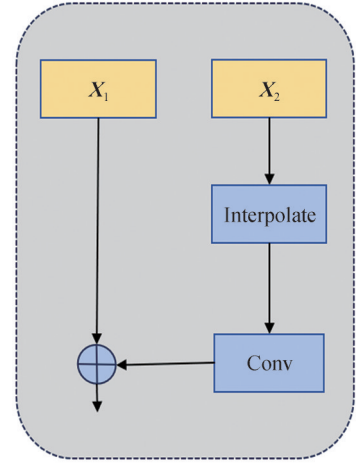


图11 DIF结构

Fig. 11 DIF structure diagram

上下文引导的特征金字塔网络(context guided-FPN, CG-FPN)采用自底向上的级联优化策略:初级融合特征经DIF模块与底层特征拼接后,通过DCMB深度卷积模块进行通道重组;次级融合过程引入残差连接,将DCMB输出与初始特征二次拼接,进一步增强特征表

达能力;最终通过三级检测头实现多尺度目标检测,如图12所示。该架构通过横向连接与动态加权机制,构建了多方向信息流,有效解决了传统FPN单向信息传递的局限性。

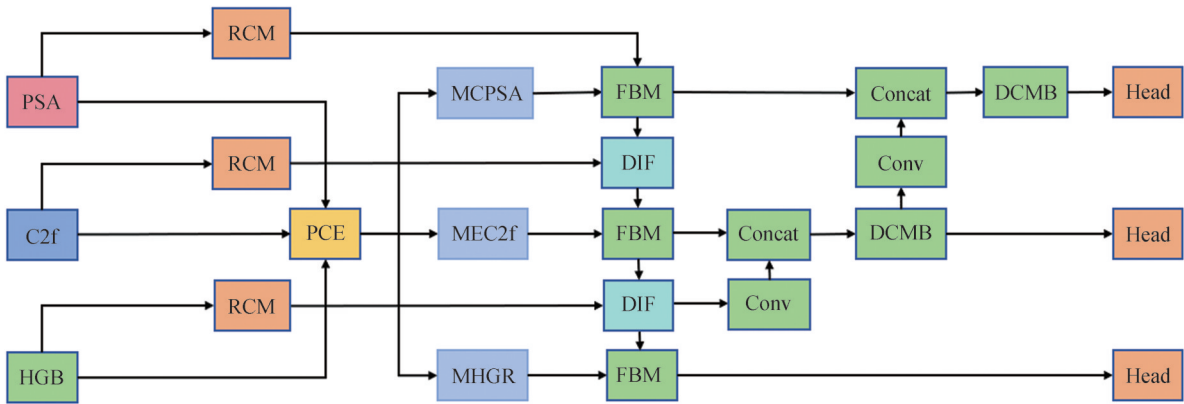


图12 CG-FPN结构

Fig. 12 CG-FPN structure

2 结果与分析

2.1 数据集构建

本研究在江苏省南京市老山森林公园利用相机拍摄了5 214张照片,其中包含不同光照及相互遮挡情况下的树木图像,如图13所示。为了准确标定每张照片中的树木,使用标记工具labelimg,通过矩形框选出目标,并为每个框选的树木标注了类别信息,这些标记数据以txt格式保存。

并且还采用了Open-Images-v7公开数据集中的2 261张树木分类图像来共同构建数据集TreeImages,

共计7 475张图像。其中训练集5 474张、验证集1 301张、测试集700张,划分比例约为7:2:1。通过这些详细步骤和严格的数据管理,确保了模型在实际应用中能够达到高效准确地检测和识别树木能力。

2.2 试验平台及训练参数

本试验所用的服务器系统为Ubuntu22.0系统,1张NVIDIA GeForce RTX 4090显卡,编译语言为Python3.10.15,深度学习框架为PyTorch2.3.0,CUDA版本为12.1。试验中,原始图片尺寸为640×640像素,初始学习率设置为0.01, batchsize设置为16,共训练200epochs,其他超参数设置遵循YOLOv10的默认设置。

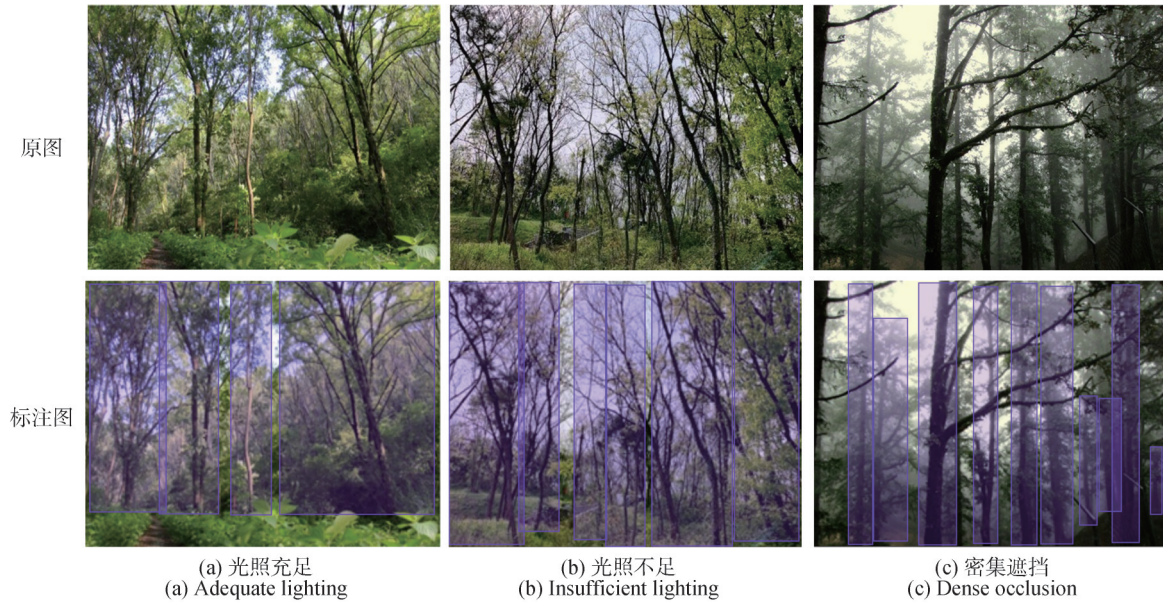


图 13 试验区域图像

Fig. 13 Map of the experimental area

2.3 评估指标

本研究采用精确率(AP)、召回率(Recall)、平均精度(mAP)包括 mAP50 与 mAP50-95、每秒十亿次浮点运算(GFLOPS)、参数量(Params)、模型权重大小(pt. size)和检测速度每秒帧数(FPS)作为评估模型检测性能的定量评估指标。

精确率指正样本中被预测为正样本占比;召回率指在预测为正的样本中确是正样本的占比。AP 指标反映目标检测对单个类别的准确性,mAP 表示求平均 AP 值,mAP50 指标代表了 IoU (intersection over union) 阈值为 50% 时的 mAP 值,mAP50-95 则为 IoU 从 50% 开始,以步长为 0.05 上升到 95% 时的平均 mAP 值,判

断 predicted bbox 和 ground truth bbox 之间的 IoU 大于 50 才被认定为正确的检测。检测速度每秒帧数(FPS)用于评估模型在批量大小为 1 的推理速度。

2.4 试验结果和分析

2.4.1 对比试验

首先是对比不同 C2f 模块的改进效果,结果见表 1。DCMB 模块与 C2f 模块相比,提升了精度并且大幅度降低了参数量,其 mAP50 最高,为 43.4%,模型参数量最低,为 1.81 M。这是因为 DCMB 模块通过动态卷积核选择与跨尺度特征融合,充分捕捉多尺度和多样化的特征信息,提升检测精度。并且构建了一种灵活且高效的卷积架构,因此整体参数较低。

表 1 对比不同 C2f 模块的改进

Tab. 1 Comparison of improvements of different C2f modules

模型 Model	精确率/% Precision	召回率/% Recall	平均精度均值/% mAP50	高平均精度均值/% mAP50-95	浮点运算量/G GFLOPS	参数量/M Params	权重体积/MB pt.size
YOLOv10	50.7	44.2	41.7	24.1	6.5	2.27	5.8
C2f-WTConv ^[22]	52.4	42.5	42.8	26.3	5.8	2.05	5.8
C2f-FMB ^[23]	51.5	41.1	42.3	24.9	6.3	2.22	5.7
C2f-Faster ^[24]	52.5	39.9	41.4	25.3	4.9	1.78	4.8
C2f-ODConv ^[25]	50.7	41.4	41.5	24.8	4.4	2.30	5.9
DCMB	52.5	43.8	43.4	27.0	4.9	1.81	5.0

其次是对比不同主干的改进效果,结果见表 2。DB-DAF (DB 为 Dual backbone, DAF 为 dynamic alignment) 主干能在精度提升的同时,保持参数量降低,其 mAP50 为 42.8%,参数量为 2.18 M。由于其引入了可

轻量化卷积的 HGStem 结构和基于多尺度特征融合 DAF 模块来融合来自不同模块的特征并降低计算量,因此增强了模型对不同特征的适应能力,从而提升整体网络的精度和鲁棒性。

表2 对比不同主干的改进

Tab. 2 Comparison of improvements of different backbones

模型 Model	精确率/% Precision	召回率/% Recall	平均精度均值/% mAP50	高平均精度均值/% mAP50-95	浮点运算量/G GFLOPS	参数量/M Params	权重体积/MB pt.size
YOLOv10	50.7	44.2	41.7	24.1	6.5	2.27	5.8
efficientViT ^[13]	53.5	39.9	42.1	26.1	8.1	3.50	6.8
fasternet ^[14]	51.6	41.8	41.5	25.3	9.4	3.76	8.8
timm ^[26]	51.1	42.1	42.0	25.3	33.8	1.29	27
convnextv2 ^[15]	51.8	40.0	40.7	25.5	12.8	5.24	11.7
DB-DAF	51.0	43.1	42.8	25.5	6.4	2.18	5.7

最后是对比不同颈部的改进效果,见表3。CG-FPN机制与原YOLOv10颈部相比,其精度与轻量化程度都有改善。其mAP50为43.5%,并且参数量为2.19M,两者相对平衡。该机制通过空间特征重建和金字塔上下文提取,在水平方向和垂直方向上捕获全局上下文,

同时显式建模矩形关键区域的轴向全局信息,用以提升特征表达能力。此外,通过金字塔结构提取多尺度上下文信息,实现不同层级特征的有效整合,增强模型的上下文感知能力。

表3 对比不同颈部的改进

Tab. 3 Comparison of improvements of different necks

模型 Model	精确率/% Precision	召回率/% Recall	平均精度均值/% mAP50	高平均精度均值/% mAP50-95	浮点运算量/G GFLOPS	参数量/M Params	权重体积/MB pt.size
YOLOv10	50.7	44.2	41.7	24.1	6.5	2.27	5.8
BiFPN ^[27]	54.7	42.4	43.2	26.9	5.9	1.72	4.1
Slimneck ^[16]	51.5	42.2	42.1	26.7	5.9	2.40	6.0
GoldYOLO ^[17]	52.8	43.0	44.0	27.8	8.9	5.40	12.2
MAFPN ^[28]	54.6	41.1	42.2	26.5	7.9	3.05	7.4
CG-FPN	52.8	41.8	43.5	27.3	6.7	2.19	5.7

2.4.2 消融试验

为了全面评估本研究提出的DDC-YOLO算法在复杂环境中对树木检测的性能,设计了6组消融试验,见表4,以严格评估每个模块的性能及各个模块间的

兼容性,消融试验结果对比如图14所示,所有模型在训练过程中在150轮次就保持收敛状态。DCMB模块作用于颈部部分,以增强全局信息聚合,DB-DAF主要针对于主干部分,来优化不同路径的特征对齐。前者

表4 消融实验

Tab. 4 Ablation experiment

YOLOv10	DCMB	DB-DAF	CG-FPN	精确率/% Precision	召回率/% Recall	平均精度均值/% mAP50	高平均精度均值/% mAP50-95	浮点运算量/G GFLOPS	参数量/M Params	权重体积/MB pt.size
√				50.7	44.2	41.7	24.1	6.5	2.27	5.8
√	√			52.5	43.8	43.4	27.0	4.9	1.81	5.0
√		√		51.0	43.1	42.0	25.5	6.4	2.18	5.7
√			√	52.8	41.8	43.5	27.3	6.7	2.19	5.7
√	√	√		54.4	43.8	44.2	27.5	6.5	2.24	5.9
√	√	√	√	56.4	43.8	46.7	29.0	6.5	2.26	5.8

注:√代表包含该组件。

Note: √ represents that this component is included.

通过动态调整不同尺度的特征提取方式,并利用层间残差提升语义一致性,后者通过自适应权重调整不同主干的信息,使其融合更一致,提高浅层和深层特征的协同能力。两者在计算流程上无冲突,并且互相增强模型整体提取与融合特征能力,因此具有极高兼容性。并且由于 DB-DAF 主要通过 DAF 来减少特征融合差异,而 CG-FPN 的上下文提取模块可以增强模型对场

景背景和目标区域的理解,有效弥补模型在大尺度和全局特征建模方面的不足,使得不同尺度的特征能够高效融合,降低融合时的信息损失,并且 DCMB 在颈部部分增强了局部细节提取,与 CG-FPN 形成互补,因此三者在设计上互补,能够在不同尺度、不同背景和不同输入特征下自适应调整,使得目标检测精度、检测鲁棒性和计算效率均有所提升,证明了其兼容性与有效性。

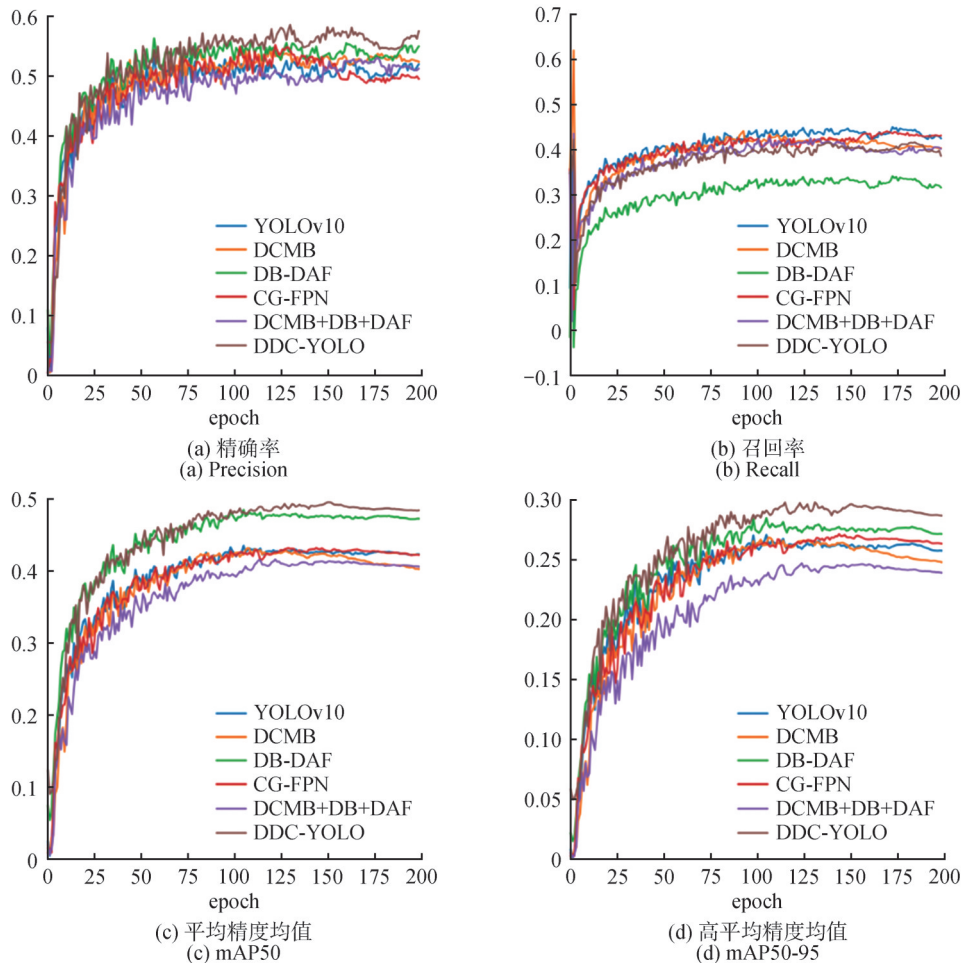


图 14 消融试验结果对比

Fig. 14 Comparison of ablation experiment results

2.4.3 对比不同整体模型

本实验对比了 Faster-RCNN、YOLOv5n^[29]、YOLOv8n^[30]、YOLOv10n 以及 DDC-YOLO 在相同数据集场景下的目标检测性能。通过对 YOLOv10n 的改进,DDC-YOLO 在多个关键指标上取得了更优表现。

由表 5 可知,Faster-RCNN 的参数量与计算量远大于其他模型,并且推理速度最慢,但其精度高于 YOLO 系列原始模型。DDC-YOLO 的精确率 (AP) 达到 56.4%,mAP50 值为 46.7%,其精度远高于所有模型,表明该模型在较低 IoU 阈值下也能更准确地检测目标。DDC-YOLO 的 mAP50-95 值为 29.0%,在所有模

型中表现最佳,说明其在不同 IoU 阈值下的检测能力更稳定,具备更强的泛化能力。召回率 (Recall) 为 43.5%,仅略低于 YOLOv10n (44.2%),仍然保持在可接受范围内。

DDC-YOLO 的 GFLOPS 与 YOLOv10n 相同,表明其计算复杂度较低,适用于轻量化部署。该模型参数量与权重,仅略高于 YOLOv5n,但低于 YOLOv8n,说明其模型复杂度低,有利于轻量化部署。更重要的是,DDC-YOLO 在推理速度方面表现最佳,检测速度 (FPS) 值达到 254 帧/s,远超其他模型,表明其在保证高精度的同时,具备更快的推理速度,能够满足实时应用需求。

表 5 对比不同的模型在数据集中的表现
Tab. 5 Comparison of the performance of different models in the dataset

模型 Model	精确率/% Precision	召回率/% Recall	平均精度均值/% mAP50	高平均精度均值/% mAP50-95	浮点运算量/G GFLOPS	参数量/M Params	权重体积/MB pt.size	检测速度/(帧·s ⁻¹) FPS
Faster-RCNN	50.0	43.1	43.5	26.2	56.1	3.91	12.4	89
YOLOV5n	48.4	42.1	42.4	24.2	4.5	1.76	2.5	200
YOLOV8n	54.2	45.2	41.7	24.4	8.7	2.87	6.2	186
YOLOV10n	50.7	44.2	41.7	24.1	6.5	2.27	5.8	202
DDC-YOLO	56.4	43.5	46.7	29.0	6.5	2.26	5.8	254

2.4.4 结果可视化分析

对来自 Open-Images-v7 数据集的随机样本进行特征图可视化操作。为了避免背景部分的干扰,首先对热力图进行归一化操作,使得热力图只作用于目标检测框内,如图 15 所示。在图 15 中颜色越深的区域表示对网络的响应度越高,即对预测的影响越大。结果表明,DDC-YOLO 网络结构在提取有相关特征和准确

定位树木位置方面体现出优良的性能。具体而言,在图 15 的第 1 列中,DDC-YOLO 能检测到位于图片边缘的树木并提取特征。在第 2 列图片中对存在多个目标密集区域进行检测时也能不产生遗漏问题,并且在第 3 列第 3 行中,DDC-YOLO 相较于原基准模型 YOLOv10n 能避免错误预测。在第 4 列有明显遮挡,且特征提取不足的情况下也能检测目标,证明其性能优良。

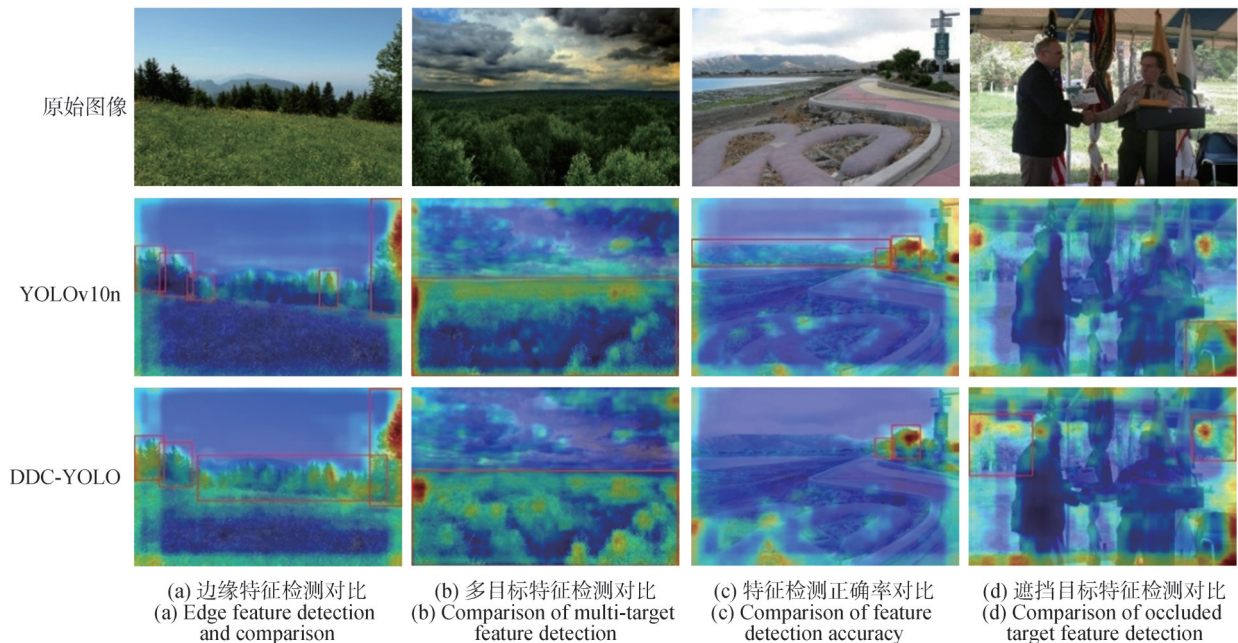


图 15 热力图可视化

Fig. 15 Heat map visualization

根据图 16 中的检测结果,DDC-YOLO 在昏暗和明亮的场景下与在密集目标和相互遮挡的情况下都能够有效地检测出目标。这一表现证明了 DDC-YOLO 模型具有广泛的适用场景和强大的模型泛化能力,能够在各种复杂环境中保持较高的检测精度和鲁棒性。相比于其他目标检测模型,DDC-YOLO 展示了在多变光

照和拥挤环境中的优越性能,进一步验证了其在不同场景下的可靠性。

3 结论

本研究针对当前森林资源普查中存在的树木检测精度不足以及在复杂场景下当前主流算法适应性差等

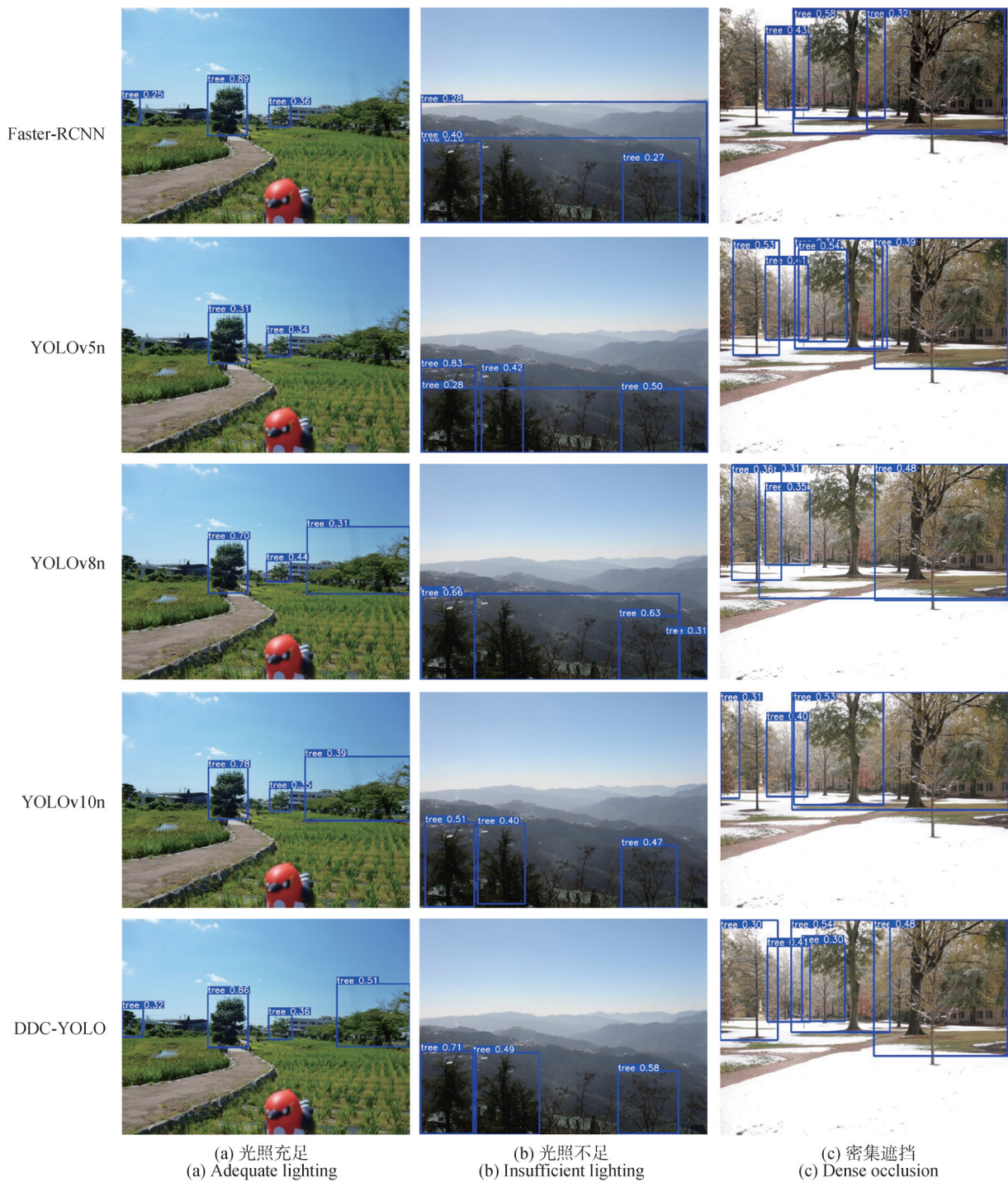


图 16 各个模型检测结果
 Fig. 16 Detection results of each model

问题,以 YOLOv10 为基础模型,通过设计 DCMB 模块来替换颈部的 C2f 模块、设计双主干式动态特征融合网络和引入金字塔上下文特征提取与空间特征重建模块机制,提升了树木检测的精度与速度,改善了模型在不同光照变化和相互遮挡场景下的特征提取能力,有效提高了检测精度。

通过消融试验验证了各改进后模块的有效性,DDC-YOLO 的 mAP 值达到 46.7%,相较于原基础 YOLOv10 网络结构提升了 5%,且参数量小幅降低,检测速度大幅度提升。对比试验表明,DDC-YOLO 在类似密集树木区域的复杂场景下的检测性能优于 YOLOv5、YOLOv8 等主流算法。并且对热力图进行可

视化操作,表明模型能够更精准地聚焦于树木冠层与树干的关键特征区域,体现了检测结果的可解释性。未来可将视觉与激光雷达点云数据融合来进行研究,并将其技术应用于其他植被检测,以推动智慧林业技术的全面落地。

参 考 文 献

- [1] 赵子琪,李丹丹,赵鼎,等. 基于深度学习的树冠分割及生物量估算[J]. 森林工程,2024,40(5):145-155.
ZHAO Z Q, LI D D, ZHAO D, et al. Canopy segmentation and biomass estimation based on deep learning [J]. Forest Engineering, 2024, 40(5): 145-155.
- [2] 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.
- [3] HUSSAIN M. YOLOv1 to v8: Unveiling each variant-A comprehensive review of YOLO [J]. IEEE Access, 2024, 12: 42816-42833.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [5] AKDOĞAN C, ÖZER T, OĞUZ Y. PP-YOLO: Deep learning based detection model to detect apple and cherry trees in orchard based on histogram and wavelet preprocessing techniques [J]. Computers and Electronics in Agriculture, 2025, 232: 110052.
- [6] 周宏威,纪皓文,吴羿轩,等. 基于YOLOv8算法改进模型检测梢斑螟虫蛀树木[J]. 森林工程, 2025, 41(1): 126-137.
ZHOU H W, JI H W, WU Y X, et al. Improve the detection model of tree decay by *Dioryctria* based on the YOLOv8 [J]. Forest Engineering, 2025, 41(1): 126-137.
- [7] 窦安亮,路红,杜一君,等. 基于改进YOLO v5的城市树木检测[J]. 南京工程学院学报(自然科学版), 2023, 21(3): 1-7.
DOU A L, LU H, DU Y J, et al. Urban tree detection based on enhanced YOLO v5 [J]. Journal of Nanjing Institute of Technology (Natural Science Edition), 2023, 21(3): 1-7.
- [8] 周焱,刘文萍,骆有庆,等. 基于深度学习的小目标受灾树木检测方法[J]. 林业科学, 2021, 57(3): 98-107.
ZHOU Y, LIU W P, LUO Y Q, et al. Small object detection for infected trees based on the deep learning method [J]. Scientia Silvae Sinicae, 2021, 57(3): 98-107.
- [9] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection [J/OL]. arXiv, 2024, arXiv: 2405. 14458.
- [10] 朱谨,俞伟浩,谭明兴,等. 基于YOLO v10的目标检测优化方法[J]. 计算机工程, 2024, 50(2): 123-132.
ZHU J, YU W H, TAN M X, et al. Object detection optimization method based on YOLO v10 [J]. Computer Engineering, 2024, 50(2): 123-132.
- [11] YU W H, ZHOU P, YAN S C, et al. InceptionNeXt: When inception meets ConvNeXt [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 5672-5683.
- [12] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 10778-10787.
- [13] LIU X Y, PENG H W, ZHENG N X, et al. EfficientViT: Memory efficient vision transformer with cascaded group attention [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 14420-14430.
- [14] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 12021-12031.
- [15] WOO S, DEBNATH S, HU R H, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 16133-16142.
- [16] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A lightweight-design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024, 21(3): 62.
- [17] WANG C C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [EB/OL]. arXiv preprint, 2023. arXiv: 2309. 11331.
- [18] YANG Z Q, GUAN Q, ZHAO K, et al. Multi-Branch auxiliary fusion YOLO with re-parameterization heterogeneous convolutional for accurate object detection [EB/OL]. arXiv preprint, 2024. arXiv: 2407. 04381.
- [19] SHI D. TransNeXt: Robust foveal visual perception for vision transformers [EB/OL]. arXiv preprint, 2024. arXiv: 2311. 17132.
- [20] LI B X, FANG J D, ZHAO Y. RTDETR-RefA: A real-time detection method for multi-breed classification of cattle [J]. Journal of Real-Time Image Processing, 2025, 22(1): 38.

- [21] ZHAO Y A, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16–22, 2024, Seattle, WA, USA. IEEE, 2024: 16965–16974.
- [22] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields [EB/OL]. arXiv preprint, 2024. arXiv: 2407. 05848.
- [23] LI H, WU A Y, JIANG Z Y, et al. Improving object detection in YOLOv8n with the C2f-f module and multi-scale fusion reconstruction [C]//2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). May 24–26, 2024, Chongqing, China. IEEE, 2024: 374–379.
- [24] ZHU J, HU T, ZHENG L H, et al. YOLOv8-C2f-faster-EMA: An improved underwater trash detection model based on YOLOv8[J]. Sensors, 2024, 24(8): 2483.
- [25] ZHAO Z L, DONG M. Channel-spatial dynamic convolution: An exquisite omni-dimensional dynamic convolution [C]//2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP). April 21–23, 2023, Xi'an, China. IEEE, 2023: 1707–1711.
- [26] WIGHTMAN R, TOUVRON H, JÉGOU H. ResNet strikes back: An improved training procedure in timm [EB/OL]. arXiv preprint, 2021. arXiv: 2110. 00476.
- [27] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13–19, 2020, Seattle, WA, USA. IEEE, 2020: 10778–10787.
- [28] ZHAO X D, CHEN J L, LIU M M, et al. Multi-scale attention-based feature pyramid networks for object detection [M]//Image and Graphics. Cham: Springer International Publishing, 2021: 405–417.
- [29] KHANAM R, HUSSAIN M. What is YOLOv5: A deep look into the internal features of the popular object detector [EB/OL]. arXiv preprint, 2024. arXiv: 2407. 20892.
- [30] VARGHESE R, SAMBATH M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness [C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). April 18–19, 2024, Chennai, India. IEEE, 2024: 1–6.