

# 基于 FIDT 图的稠密小目标树木检测架构

冯戈<sup>1</sup> 霍光煜<sup>2</sup> 许新桥<sup>1</sup> 闫瑞华<sup>1</sup> 周庆宇<sup>1</sup> 乐林暄<sup>1</sup> 操东林<sup>1</sup>

(1. 国家林业和草原局信息中心, 北京 100714; 2. 北京林业大学信息学院(人工智能学院), 北京 100083)

**摘要:** 在高密度区域, 树木冠层重叠和遮挡严重, 传统方法难以准确识别单棵树木, 影响成活率和绿化覆盖率的评估精度, 因此, 亟需一种高精度和强鲁棒性的检测方法, 以提升城市树木成活监测的准确性和效率。提出基于焦距反比变换图通过计算目标点到最近边界的距离, 对稠密树木小目标检测方法, 采用一种多层次的局部极大值检测策略, 能够有效提取树木目标中心点信息, 并区分相邻目标, 从而降低重叠区域目标检测误差。此外, 为了增强模型对局部结构信息的学习能力, 引入独立结构性损失。选用一个高分辨率林地遥感数据集, 并对稠密小目标树木进行检测与定位分析。为了验证本文方法的有效性, 与多种现有深度学习方法进行对比分析, 评估不同方法在精确率、召回率、定位误差等方面表现。试验结果表明, 本文方法在城市林地遥感数据集上展现了较高的定位和计数性能。在计数任务中, 本文方法平均绝对误差 (MAE) 和均方误差 (MSE) 分别为 7.87 和 10.23, 相较于现有方法 CSRNET 分别降低 45.1% 和 48.3%; 在定位任务中, 本文方法在 Claremont、Long\_beach、Palm\_springs 和 Riverside 4 个区域的 F1 值分别达到 75.2%、72.7%、74.8% 和 71.5%, 较现有基于密度图法至少提升 19 个百分点, 同时在精确率和召回率等指标上也实现了全面超越。研究结果为林业资源管理、生态监测和环境保护提供了一种高效、可靠的技术手段。

**关键词:** 焦距反比变换图; 稠密树木; 小目标检测; 深度学习

中图分类号: Q958.15; S154.5 文献标识码: A 文章编号: 1000-1298(2026)02-0143-09

OSID:



## Dense and Small Tree Detection Framework Based on FIDT Map

FENG Ge<sup>1</sup> HUO Guangyu<sup>2</sup> XU Xinqiao<sup>1</sup> YAN Ruihua<sup>1</sup> ZHOU Qingyu<sup>1</sup> LE Linxuan<sup>1</sup> CAO Donglin<sup>1</sup>

(1. Information Center, National Forestry and Grassland Administration, Beijing 100714, China

2. School of Information Science and Technology (School of Artificial Intelligence),  
Beijing Forestry University, Beijing 100083, China)

**Abstract:** In high-density areas, the severe overlap and occlusion of tree canopies pose significant challenges for traditional methods in accurately identifying individual trees, thereby compromising the precision of evaluations related to survival rates and greening coverage. Consequently, there is an urgent need for a detection approach with high accuracy and strong robustness to enhance the effectiveness and reliability of urban tree survival monitoring. The detection method proposed was based on the focal inverse distance transform (FIDT), which computed the distance from each target point to its nearest boundary and performed an inverse transformation. On this basis, a multi-level local maxima detection strategy was introduced to effectively extract the center point information of tree targets and distinguish adjacent targets, thereby reducing detection errors in overlapping regions. Additionally, an independent structural loss was incorporated to enhance the model's ability to learn local structural information. Experiments were conducted on a high-resolution forest remote sensing dataset to detect and localize dense small tree targets. To validate the effectiveness of the proposed method, comparative analyses were performed against various existing deep learning approaches, evaluating their performance in terms of detection accuracy, recall, and localization error. Experimental results showed that the proposed method achieved strong localization and counting performance on the urban forest dataset. For counting, it

收稿日期: 2025-06-06 修回日期: 2025-08-13

基金项目: 国家重点研发计划项目(2023XAGG0065)和国家林业和草原局林草资源数字化治理能力提升技术研究项目(XTZD2024-05-21)

作者简介: 冯戈(1987—),女,高级工程师,主要从事林草行业网络安全和信息化发展研究,E-mail: 341802378@qq.com

通信作者: 许新桥(1972—),男,教授级高级工程师,博士,主要从事林草信息化、生态经济和森林经营研究,E-mail: xxxz3@forestry.gov.cn

attained MAE and MSE of 7.87 and 10.23, reducing errors by 45.1% and 48.3% compared with that of CSRNET. For localization, F1-scores in Claremont, Long\_beach, Palm\_springs, and Riverside were 75.2%, 72.7%, 74.8%, and 71.5%, averaging over 19 percentage points higher than that of existing density map-based methods, with overall improvements also observed in precision and recall. The research result can provide an efficient and reliable technical approach for forestry resource management, ecological monitoring, and environmental protection. In the future, integrating multi-source remote sensing data and time series analysis could further improve the accuracy and efficiency of dynamic forest resource monitoring.

**Key words:** focal inverse distance transform; high-density tree; small target detection; deep learning

## 0 引言

随着遥感技术和计算机视觉的发展,基于遥感图像的目标检测为城市绿化树木的自动识别与监测提供了新的可能。基于遥感、无人机图像的树木检测技术能够在较大范围内快速识别成活树木,评估绿化覆盖率,显著提升绿化管理效率<sup>[1-4]</sup>。然而,在实际遥感影像中,城市树木往往以高密度、小目标形式存在,且常伴有遮挡与冠层重叠,导致传统检测方法难以准确定位和计数<sup>[5]</sup>。为此,需引入更具针对性和鲁棒性的检测方法,有效提升对城市绿化树木的精细化监测水平,服务于城市生态管理和可持续发展目标。

早期的树木检测方法主要依赖于手工设计的特征和传统的机器学习算法,如定向梯度直方图(Histogram of oriented gradients, HOG)<sup>[6]</sup>和尺度不变特征变换(Scale-invariant feature transform, SIFTq)<sup>[7]</sup>等,此类方法基于人工提取特征并训练分类器识别目标,限制了其在复杂场景下的应用。随着深度学习技术的不断发展,出现了基于深度学习的树木检测方法。一般来说,基于深度学习的目标密度估计的方法有3种:基于检测的目标密度估计<sup>[8]</sup>、基于回归的目标密度估计<sup>[9]</sup>和基于密度图的目标密度估计<sup>[10]</sup>。早期基于深度学习的树木检测方法包括Faster-RCNN<sup>[11]</sup>和YOLO<sup>[12]</sup>,已经能够实现一定的检测精度和速度,但大多方法仍存在一定局限性,特别是多数方法依赖于密度图回归进行目标计数。这些方法通过生成目标密度图并对其进行积分,以估算目标数量,但主要问题是无法提供每个目标的精确位置信息,其根本原因在于常用的密度图中高斯块(Gaussian blobs)在目标密集区域容易重叠,导致局部最大值未必对应于真实目标的位置,因此,尽管这些方法虽然能够估计目标的总数,但无法精确定位每个目标。位置和大小信息对稠密树木检测至关重要,尽管基于部分尺度感知深度检测网络<sup>[13]</sup>和局部自我校正卷积神经网络<sup>[14]</sup>的方法通过目标中心距离生成伪真值框以替代点标注,能提升

树木检测的定位精度,但由于伪真值框难以准确反映真实目标,性能提升仍有限,研究者通过设计二元类图<sup>[15]</sup>、分割类图<sup>[16]</sup>、三值图<sup>[17]</sup>、拓扑图<sup>[18]</sup>和独立实例地图<sup>[19]</sup>实现目标定位。对密集树木的检测,研究者们设计了多种深度学习网络结构,广泛应用于遥感图像处理。常见的网络架构包括基于卷积神经网络的模型,如用于图像分割的U-Net<sup>[20]</sup>和全卷积网络(Fully convolutional network, FCN)<sup>[21]</sup>,以及针对特定需求设计的网络模型,如树形网<sup>[22]</sup>和树形长短期记忆网络模型<sup>[23]</sup>等。针对遥感图像中树木具有不同尺度和形状的特点,研究者提出了多尺度特征融合的方法增强网络对树木的检测能力,利用多层次的特征金字塔或者多尺度感受收集树木信息,以确保在不同尺度下都能全面获取特征<sup>[24]</sup>。此外,也有研究者们致力于运用各种数据增强技术,提升模型的泛化性能和鲁棒性<sup>[25]</sup>。通过这些手段,模型能够更好地适应各种场景和数据变化,同时减轻模型对输入图像变换的敏感性,从而提高模型在实际应用中的性能表现。

尽管深度学习在遥感图像中的树木检测取得了显著进展,但现有方法多侧重于标准目标的检测,对于稠密树木小目标的检测效果较差,尤其是在高分辨率遥感图像中,小树木的检测精度较低,小目标在图像中所占的像素较少,特征信息也较为稀缺,这使得传统方法难以有效提取和识别这些关键信息。因此,如何提高小目标检测的精度和效率,成为亟待解决的难题。此外,在树木密集的情况下,树木之间的相互遮挡也是一个不可忽视的挑战,严重影响了树木检测的准确性。

因此,本文提出一种基于焦距反比变换图(Focal inverse distance transform, FIDT)对稠密小目标树木检测方法,以解决目标重叠问题。在此基础上,采用多层次局部极大值检测策略,从预测的FIDT图中精确提取每个树木目标的中心点。为进一步提升检测性能,基于深度学习技术引入独立性损失,使检测模型更加关注并学习局部结构信息,以准确识别局部最大值。

## 1 基于焦距反比变换图的稠密小目标树木检测模型

检测模型工作流程如图 1 所示。首先,通过在训练时引入高分辨率网络(HRNet)生成预测的焦距反比变换图(FIDTM),并利用均方误差损失和改进的独立结构性损失(1-SSIM Loss)度量预测结果与真实值之间的差异。然后在测试时将焦距反比变换图利用局部最大值检测策略(LMDS)获取定位。

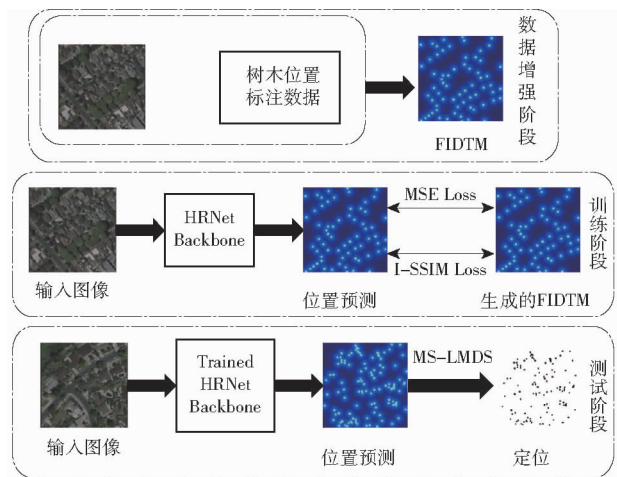


图 1 基于焦距反比变换图的稠密小目标树木检测模型

Fig. 1 Dense small object tree detection model based on FIDTM map

在 FIDTM 中,每个像素值表示其距离周围目标的反比权重。具体地,对于图像中每个目标点, FIDTM 通过计算该目标点与其邻域内其他目标点之间的距离,并将该距离的倒数作为每个像素权重。此种基于反比距离的计算方式有助于突出目标的集中区域,并抑制背景噪声干扰<sup>[26]</sup>。此外, FIDTM 的计算公式中引入距离的倒数和自适应参数,使得目标密度估计不仅考虑了像素之间的距离关系,还能够根据目标的局部聚集性和目标与背景的关系进行动态调整。其计算公式为

$$FIDTM(x, y) = \sum_{i=1}^N \frac{1}{d_i(x, y)^{(\alpha d_i(x, y) + \beta)} + C} \quad (1)$$

式中  $d_i(x, y)$ ——图像中目标点  $i$  与相邻目标点的距离

$\alpha, \beta$ ——调节距离衰减速率的常数

$C$ ——常数,避免计算中除零错误并调整变换的平移量

式(1)计算过程将每个目标点与其邻域内其他目标点之间的距离通过倒数进行加权,并根据预设参数  $\alpha, \beta$  调节距离的影响。这种方法不仅使得目标区域的响应更加突出,同时有效抑制了远离目标的背景噪声,增强了前景区域的定位精度。

与传统使用高斯核的密度图方法相比, FIDTM

具有明显的优势<sup>[27]</sup>。高斯核通常为固定大小和形状,在目标密集的场景中可能会有目标重叠现象,影响目标计数和定位的准确性,而 FIDTM 通过自适应的距离衰减机制,解决目标重叠的问题。

### 1.1 HRNet 基础网络

受到 LIANG 等<sup>[28]</sup>工作的启发,采用高分辨率网络(HRNet)<sup>[29]</sup>作为基础网络,添加 1 个卷积层和 2 个转置卷积层,构建新的表示头。这些改动使网络能回归焦距反比变换图,并确保目标中心与局部最大值之间的精确对应。为了进一步优化目标的分割和计数效果,结合独立实例地图分割(IIM)<sup>[19]</sup>方法不仅能有效提取目标高分辨率特征,还能确保 FIDTM 的响应与目标中心之间的匹配。

### 1.2 基于多尺度卷积操作的局部最大值检测策略

在传统的局部最大值检测策略(LMDS)中,局部最大值被用于识别图像中的关键特征点,这些特征点通常表示目标的存在位置。然而,传统方法在处理不同尺寸和形态的目标时,尤其是在树木分布不均的复杂森林环境中,常常面临尺度适应性差、鲁棒性不足的问题。因此,本文基于多尺度卷积操作的局部最大值检测策略(Multi-Scale convolutional local maximum detection strategy, MS-LMDS),通过引入不同尺度的卷积操作,增强对树木的多尺度特征提取能力,提高树木检测的精度和鲁棒性。

在 MS-LMDS 中,通过应用不同大小的卷积核提取多尺度的特征图,每种卷积核通过改变感受野的大小,能有效捕捉不同尺度下的树木特征。具体地,使用 3 种不同尺寸卷积核,  $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$ , 每种卷积核对应着不同的感受野,从而能提取细节特征和全局特征。在网络中引入这些多尺度卷积操作,能够同时处理不同大小、密度的树木目标,并在多尺度空间中进行特征融合。卷积操作可表示为

$$I_k = \text{Conv}_k(I) \quad (2)$$

式中  $I$ ——输入图像

$\text{Conv}_k(\cdot)$ ——应用大小为  $k \times k$  的卷积核,得到对应尺度  $k$  的特征图  $I_k, k \in \{3, 5, 7\}$

为了提取局部最大值,在每个尺度的特征图上使用了最大池化操作。最大池化能够在每个尺度下保留局部区域中的最大响应,从而有效捕捉树木的关键特征点。MS-LMDS 方法将每个尺度下的最大池化结果与原始输入进行比较,通过选择每个尺度上的最大值确定潜在的树木位置。具体来说,对于每个尺度的特征图,首先应用  $3 \times 3$  感受野的最大池化操作,以捕捉局部的最大响应,然后使用最大化函数将不同尺度的特征图进行融合,选择最大响应的

位置作为最终的树木定位点。最大池化操作公式为

$$keep_k = \text{MaxPool}\left(I_k, (3, 3), \text{stride}1, \text{padding} = 1\right) \quad (3)$$

式中  $\text{MaxPool}(\cdot)$ ——最大池化操作,步长和填充都为 1

$$keep = \max(keep_3, keep_5, keep_7) \quad (4)$$

式中  $keep_3, keep_5, keep_7$ ——来自  $3 \times 3, 5 \times 5$  和  $7 \times 7$  卷积核的池化结果

$\max(\cdot)$ ——对多个尺度特征图进行逐元素最大化

为了进一步优化局部最大值检测过程,MS-LMDS 引入阈值设定策略,将图像中低于一定像素值的区域设为零,以过滤背景噪声或误检测点。在 MS-LMDS 中,局部最大值的像素值低于预定阈值(如  $100.0/255.0 \times \text{input\_max}$ )时,被视为背景,置为零;同时,为了处理负样本,当图像的最大值低于设定阈值,将输入图像中的所有像素值置为零,避免误检测。

### 1.3 损失函数设计

因为树木通常较小且密集,传统的回归损失函数(如均方误差 MSE)可能不足以有效捕捉这些目标的复杂特征。为此,本文融合 MSE 损失和独立结构相似性损失(Independent structural similarity loss, ISSL)提升模型在小目标检测的性能。

独立结构相似性损失 ISSL 是一种度量图像结构相似性的新损失函数,能够有效捕捉目标的局部结构信息,该损失函数基于结构相似性度量(SSIM)通过计算图像的亮度、对比度和结构信息评估 2 幅图像之间的相似度。对于小目标树木计数和定位任务,ISSL 能够在保持全局信息的同时,更加注重局部区域的结构相似性,从而提升小目标检测精度。对于真实图像  $I$  和预测图像  $\hat{I}$ ,在局部区域上,SSIM 计算式为

$$SSIM(I, \hat{I}) = \frac{(2\mu_I \mu_{\hat{I}} + c_1)(2\sigma_{II} + c_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + c_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + c_2)} \quad (5)$$

式中  $\mu_I, \mu_{\hat{I}}$ ——图像  $I$  和预测图像  $\hat{I}$  在区域上均值(即亮度信息)

$\sigma_I^2, \sigma_{\hat{I}}^2$ ——图像  $I$  和预测图像  $\hat{I}$  在区域上方差(即对比度信息)

$\sigma_{II}$ ——图像  $I$  和预测图像  $\hat{I}$  在区域上协方差(即结构信息)

$c_1, c_2$ ——用于稳定分母的常数,避免分母为零

ISSL 通常是通过计算预测图像和真实图像之间的 SSIM 相似度获得的,为使得损失函数具有优化意义,通常将其转化为“损失”形式,即最大化

SSIM 相似度。而本文需要最小化损失,因此 ISSL 损失定义为

$$L_{ISSL} = 1 - SSIM(I, \hat{I}) \quad (6)$$

该损失函数在训练过程中鼓励模型产生具有更高结构相似性的预测图像,从而提高图像的空间结构保真度。为了更好地对图像中小目标树木检测,ISSL 通常会考虑局部区域的结构相似性。因此,SSIM 在图像的每个局部窗口计算,在整个图像中求取平均值,计算式为

$$L_{ISSL} = 1 - \frac{1}{N} \sum_{i=1}^N SSIM(I_i, \hat{I}_i) \quad (7)$$

式中  $I_i, \hat{I}_i$ ——图像  $I$  和预测图像  $\hat{I}$  中第  $i$  个局部窗口区域

$N$ ——局部窗口总数

此种基于局部窗口的计算方式,ISSL 损失能够捕捉到图像中细小区域的结构信息,有助于小目标树木的准确检测和定位,特别是对复杂的树木计数。

为了在小目标树木计数和定位任务中有效地融合 MSE 损失和 ISSL 损失,采用加权求和的方式,计算式为

$$L_{\text{total}} = \partial L_{MSE} + \varphi L_{ISSL} \quad (8)$$

式中  $\partial, \varphi$ ——超参数

超参数控制 MSE 损失和 ISSL 损失在最终损失函数中的权重。通过调整 2 个权重,在训练过程中平衡回归精度和结构相似性,从而提高小目标树木计数和定位精度。

## 2 试验

### 2.1 数据集

数据集采用来自加州 8 个城市(Bishop、Chico、Claremont、Eureka、Long Beach、Palm Springs、Riverside 和 Santa Monica)的 2016、2018、2020 年 NAIP 的 256 像素  $\times$  256 像素 60 cm 遥感图像,共 1 651 幅图像和 95 972 个已标注树木,对图像中可见的所有树木进行点标注。将这些数据集整合为 Urban Tree 数据集。这些数据集包含来自城市、公园、河流和乡村的高分辨率航拍图像,涵盖了不同季节和天气条件的树木分布情况。将数据集划分为训练集和测试集,划分比例为 17:3。

### 2.2 数据预处理

本文所用的焦距反比变换模型对其所训练的图像数据集有一定的需求,因此需要对数据集进行预处理,为模型训练提供良好的基础。数据集涵盖了不同地区和不同季节的树木影像,为了保证数据的一致性,对所有图像进行标准化处理,统一调整图像大小为 256 像素  $\times$  256 像素,消除不同图像之间的

差异,提高模型的泛化能力。

## 2.3 环境配置及参数设置

采用 Pytorch 深度框架进行代码开发,版本为 2.0.0;试验运行平台为 Windows 操作系统和 Ubuntu 20.04 操作系统,版本为 11.0, GPU 采用 24 G 显存的 NVIDIA RTX 4090,并行计算架构(CUDA)版本为 11.8,深度神经网络训练加速库 CuDNN 版本为 7.3.1。

## 2.4 评价指标

使用精确率( $P$ )、召回率( $R$ )和 F1 值评估模型的定位性能。精确率是预测正确的正例数据占预测为正例数据的比例;召回率是预测为正确的正例数据占实际为正例数据的比例;F1 值是精确度和召回率的调和平均值。

使用平均绝对误差(Mean absolute error, MAE)和均方误差(Mean squared error, MSE)评估模型的计数性能。均方误差 MSE 是一种常用的衡量模型预测值与实际观测值之间差异的指标,用于评估模型在给定数据上的拟合程度, MSE 是通过计算预测值与实际观测值之间差异平方的平均值得到。平均绝对误差 MAE 是一种应用于回归模型的损失函数,通过计算真实值与预测值之差的绝对值之和获得,其取值范围为 $[0, +\infty)$ 。计算式为

$$MAE = \frac{1}{M} \sum_{i=1}^M |P_i - G_i| \quad (9)$$

$$MSE = \sqrt{\frac{1}{M} \sum_{i=1}^M |P_i - G_i|^2} \quad (10)$$

式中  $M$ ——测试所使用图像总数

$P_i$ ——第  $i$  幅图像预测值(Pred)

$G_i$ ——第  $i$  幅图像真实值

## 2.5 不同模型性能试验

为了验证和评估本文提出的计数方法的有效性和可行性,选取 9 种传统的深度学习模型进行计数预测对比试验。这些模型涵盖了不同网络架构,旨在全面评估本文模型的表现。

ALEXNET<sup>[30]</sup>:通过增加网络深度并使用 ReLU 激活函数、数据增强和 Dropout 技术,显著提高图像分类准确率。

Inception-v3<sup>[31]</sup>:通过采用不同大小的卷积核和精巧的网络设计,能够在较少的计算资源下实现高效的特征提取。其在图像分类任务中表现卓越,并在多个计算机视觉任务中取得了优异成绩。

MCNN<sup>[32]</sup>:采用多列结构,每列网络分别处理不同的输入区域,进而汇聚信息进行计数预测,有效解决密集场景下的计数问题。

CSRNET<sup>[33]</sup>:引入社交关系模型捕捉个体之间

的相互关系,从而提高计数精度,有效解决传统方法在复杂场景下的误差问题。

EFFICIENTNET<sup>[34]</sup>:通过结合深度、宽度和分辨率的复合尺度优化模型架构,在图像分类任务中展现了极高的效率和准确度,能够在较少的计算资源下实现更好的性能。

Vision Transformer(ViT)<sup>[35]</sup>:一种将图像切分为小块并通过自注意力机制处理的模型,在图像分类等视觉任务中表现出色。

METAMSNET<sup>[36]</sup>:通过引入多源时空信息进行更加精准的计数预测,在处理具有时序特征和多模态数据的计数任务中,能有效提升预测稳定性与精度。

DSNET<sup>[37]</sup>:一种密集连接的多尺度卷积网络,专为密集计数任务设计,通过引入不同尺度的信息流,增强对不同对象尺度的感知能力,有效提升计数精度。

BCNET<sup>[38]</sup>:一种基于双层卷积结构的网络,通过深层和浅层特征的结合提高计数精度,不仅能捕捉局部细节特征,还能有效捕捉全局信息,优化计数性能。

对不同模型在训练集上进行树木计数训练,在测试集进行测试,测试结果如表 1 所示。

表 1 不同模型在数据集上的测试结果

Tab.1 Results of proposed model with other existing models on dataset

模型	MAE	MSE
ALEXNET	16.24	22.20
METAMSNET	25.87	31.35
BCNET	25.87	31.35
ViT	25.87	31.35
Inception-v3	25.87	31.35
EFFICIENTNET	17.75	23.71
CSRNET	14.34	19.80
MCNN	15.79	20.54
DSNET	18.68	23.03
本文	7.87	10.23

## 3 结果分析

### 3.1 目标计数结果分析

从表 1 可以看出,本文模型 MAE 和 MSE 均优于其他模型,其 MAE 为 7.87, MSE 为 10.23,相较于现有最佳方法 CSRNET 分别降低 45.1% 和 48.3%,表明在计数任务中具有较高精度和较低误差。

尽管 ALEXNET(MAE 为 16.24, MSE 为 22.20)和 EFFICIENTNET(MAE 为 17.75, MSE 为 23.71)在某些传统视觉任务中表现良好,但在计数任务中表现较差。ALEXNET 的较低层次结构无法有效处

理复杂的密集计数问题,而 EFFICIENTNET 虽有所优化,但在处理大范围尺度差异和复杂场景时,仍无法达到本文模型的精度,本文模型的 FIDTM 和多尺度卷积操作有效弥补了这些传统模型在多尺度特征提取和局部精度上的不足。CSRNET (MAE 为 14.34, MSE 为 19.80) 和 MCNN (MAE 为 15.79, MSE 为 20.54) 是为人群计数设计的网络,尽管在计数任务上已有一定优化,但在面对密集或复杂场景时,无法与本文模型竞争。本文模型局部最大值检测策略和 FIDTM 能够更好地捕捉到图像中各个物体之间的微小差异,特别是在高密度场景下,准确性和鲁棒性都大大增强。ViT 和 Inception-v3 虽在某些图像分类任务上取得了显著成绩,但自注意力机

制和卷积操作之间的协同效果在计数任务中并不如本文模型高效,ViT 模型本身虽在长程依赖关系建模上有优势,但在处理局部细节和多尺度问题时,未能展现出相应的优势。本文模型多尺度卷积操作和 FIDTM 使得模型能够更好地融合局部特征和全局信息,从而提升了计数精度。

表 2 为本文模型在被检测个体疏密度不同场景的定量计数结果,Claremont 和 Palm\_springs 为个体低密度场景,Long\_beach 和 Riverside 为个体高密度场景。由表 2 可以看出,该方法在 2 种场景下均表现出优秀的目标计数性能,在显著的幅度上超过了使用密度图回归方法 (Density),表明本文模型既能对稀疏场景的目标计数,也能对密集场景的目标计数。

表 2 不同场景下模型计数结果

Tab. 2 Counting performance evaluation indexes in different scenarios

方法	Claremont		Long_beach		Palm_springs		Riverside	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
FIDT(本文)	6.676	9.304	7.500	10.236	5.355	7.483	7.870	10.076
Density	8.112	11.952	17.349	23.548	7.833	10.287	14.953	26.869

图 2 为预测的焦距反比变换图和密度图在目标计数任务上的可视化结果,在目标计数方面,相比于密度图,焦距反比变换图由于能够更准确地保留目标的位置信息,即使在密集区域也能提供几乎精确的定位,使得焦距反比变换图有着良好的目标计数性能。在高密度场景下,由于密度图存在严重的位置信息丢失和重叠现象,所以密度图在稠密区域无

法给出可靠的结果,而焦距反比变换图在密集区域的计数效果远高于密度图,为每个场景提供更可靠的技术性能。这一对比凸显了焦距反比变换图在目标计数任务中的优势,验证了其在处理密集目标场景时的有效性和准确性。

### 3.2 目标定位结果分析

表 3 为在被检测个体疏密度不同的场景下本文

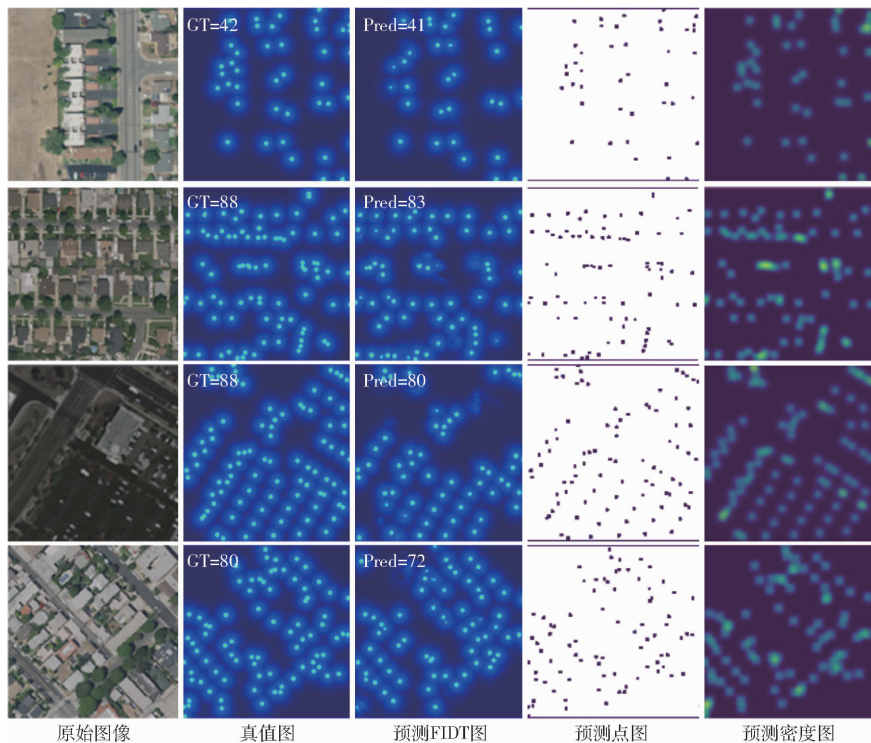


图 2 不同场景下的计数性能可视化

Fig. 2 Visualization of counting performance in different scenarios

方法与密度图回归方法定位性能结果。由表 3 可以看出,本文方法在多种场景下均表现出优秀的目标定位性能,在显著的幅度上超过了使用密度图回归

方法,表明本文方法既能应对稀疏场景,也能满足密集场景的目标定位需求。

图 3 为预测的焦距反比变换图 (FIDT) 和密度

表 3 不同场景下的定位性能结果

Tab.3 Location performance evaluation indexes in different scenarios

%

方法	Claremont			Long_beach			Palm_springs			Riverside		
	精确率	召回率	F1 值	精确率	召回率	F1 值	精确率	召回率	F1 值	精确率	召回率	F1 值
FIDT (本文)	80.1	73.5	75.2	77.3	70.6	72.7	79.9	71.3	74.8	75.2	69.3	71.5
Density	58.9	54.1	56.2	56.5	49.6	51.3	57.6	53.5	50.3	55.1	48.7	50.4

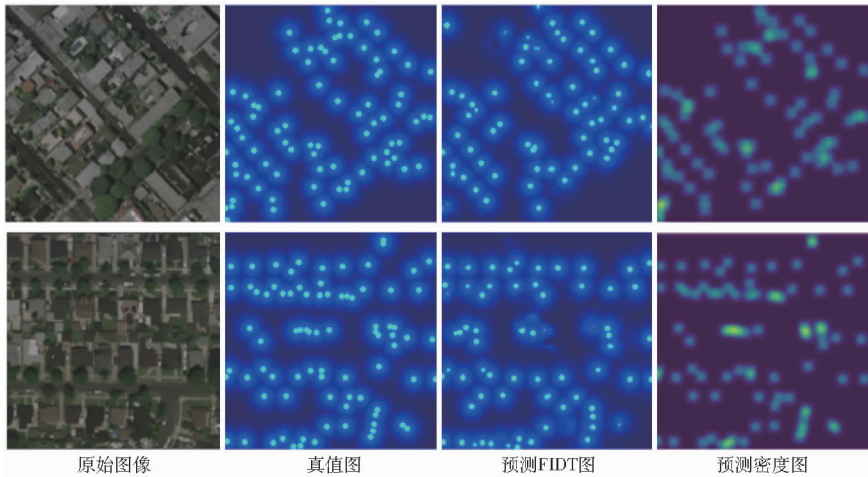


图 3 不同场景下定位性能可视化

Fig. 3 Visualization of location performance in different scenes

图 (Density) 在目标定位任务上的可视化结果,结合表 3 可知,在不同目标密度的场景中,FIDT 方法均显著优于密度图法。在 4 个城市数据集上,FIDT 平均 F1 值分别为 Claremont 75.2%, Long\_beach 72.7%, Palm\_springs 74.8%, Riverside 71.5%,相比之下,密度图方法对应的 F1 值仅为 Claremont 56.2%, Long\_beach 51.3%, Palm\_springs 50.3%, Riverside 50.4%,本文方法提升幅度为 19~24.5 个百分点。

在密集区域,FIDT 方法对极小的目标也能提供准确的定位,表现为精准率在 4 个城市均高于 75%,而密度图方法的精准度普遍低于 60%。这得益于 FIDT 在保留目标位置的同时,避免了密度图由于高斯核过滤产生的模糊化,导致位置信息丢失和目标重叠难以区分的缺陷。在低密度区域,虽 2 种方法都能给出目标位置,但 FIDT 的召回率和 F1 值仍然更高,表明其定位准确率和完整性更优。这一对比凸显了焦距反比变换图在目标定位任务中的优势,验证了其在处理密集目标场景时的有效性和准确性。

## 4 讨论

分析了焦距反比变换图 (FIDTM) 结合多尺度

卷积操作 (MS-LMDS) 在稠密小目标场景下的计数与定位性能,结果表明,本文方法相比于传统密度图回归方法具有显著优势。在 Urban Tree 数据集中,本文方法平均 F1 值提升 19~24.5 个百分点,同时 MAE 和 MSE 均优于 CSRNET、MCNN、EFFICIENTNET、ViT 等常用模型。试验结果表明本文方法能够更精准地捕捉目标的位置信息,尤其是在密集区域,避免了密度图方法因高斯核模糊导致的定位偏差与目标重叠,验证了其在复杂小目标检测任务中的适用性与有效性。

同时,本文仍存在一些局限性。本文方法的性能依赖于高质量输入图像,如果原始影像存在噪声或分辨率不足,可能会影响目标定位精度。未来,需要结合不同成像传感器(如航空影像、无人机近景影像)对模型进行迁移泛化能力评估。在国内的应用前景方面,本文提出的方法可用于街道行道树普查、果园高密度果实检测等场景,能够为林业资源普查和城市绿地信息化管理提供高效可靠的技术支撑。

## 5 结束语

提出的基于焦距反比变换图的方法在稠密小目标树木检测中展现了显著优势。与传统的密度图方

法相比, FIDT 更准确地描述了树木目标位置, 且有效避免了目标重叠问题。在此基础上, 多层次局部极大值检测策略能够精确提取树木的中心点, 而独立结构性损失的引入进一步优化了模型对局部结构

信息的学习, 提升了对树木位置和数量的准确识别。试验结果表明, 所提方法在林地遥感数据集上均表现出优越的树木定位和计数性能, 尤其在处理稠密小目标树木时, 能够大幅提高检测精度。

### 参 考 文 献

- [1] WALLACE L, LUCIEER A, WATSON C S, et al. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(12): 7619–7628.
- [2] FEKETE A, CSEREP M. Tree segmentation and change detection of large urban areas based on airborne LiDAR[J]. *Computers & Geosciences*, 2021, 156: 105156.
- [3] CHI D, DEGERICKX J, YU K, SOMERS B. Urban tree health classification across tree species by combining airborne laser scanning and imaging spectroscopy[J]. *Remote Sensing*, 2020, 12(15): 2435.
- [4] HARIKUMAR A, BOVOLO F, BRUZZONE L. A local projection-based approach to individual tree detection and 3-D crown delineation in multistoried coniferous forests using high-density airborne LiDAR data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(2): 1168–1182.
- [5] WANG X, ZHAO Q, JIANG P, et al. LDS-YOLO: a lightweight small object detection method for dead trees from shelter forest[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107035.
- [6] 汤高扬, 卢天利, 黄杰. 基于改进方向梯度直方图的行人检测[J]. *工业控制计算机*, 2019, 32(4): 87–88.  
TANG Gaoyang, LU Tianli, HUANG Jie. Pedestrian detection based on improved gradient direction histogram[J]. *Industrial Control Computer*, 2019, 32(4): 87–88. (in Chinese)
- [7] 李文举, 王子杰, 崔柳. 基于多特征融合和改进 SIFT 的目标跟踪算法[J]. *郑州大学学报(理学版)*, 2024, 56(1): 40–46.  
LI Wenju, WANG Zijie, CUI Liu. Object tracking algorithm based on multi-feature fusion and improved SIFT[J]. *Journal of Zhengzhou University (Natural Science Edition)*, 2024, 56(1): 40–46. (in Chinese)
- [8] 杜紫薇, 周恒, 李承阳, 等. 面向深度卷积神经网络的小目标检测算法综述[J]. *计算机科学*, 2022, 49(12): 205–218.  
DU Ziwei, ZHOU Heng, LI Chengyang, et al. A survey of small object detection algorithms for deep convolutional neural networks[J]. *Computer Science*, 2022, 49(12): 205–218. (in Chinese)
- [9] 蔺素珍, 郑瑶, 禄晓飞, 等. 基于多域卷积神经网络与自回归模型的空中小目标自适应跟踪方法[J]. *光学学报*, 2017, 37(12): 281–288.  
LIN Suzhen, ZHENG Yao, LU Xiaofei, et al. Airborne small target adaptive tracking method based on multi-domain convolutional neural network and autoregressive model[J]. *Acta Optica Sinica*, 2017, 37(12): 281–288. (in Chinese)
- [10] 马皓, 殷保群, 彭思凡. 基于特征金字塔网络的人群计数算法[J]. *计算机工程*, 2019, 45(7): 203–207.  
MA Hao, YIN Baoqun, PENG Sifan. Crowd counting algorithm based on feature pyramid network[J]. *Computer Engineering*, 2019, 45(7): 203–207. (in Chinese)
- [11] 张杰. 基于改进 Faster-RCNN 的小目标检测[J]. *现代计算机*, 2023, 29(14): 14–18.  
ZHANG Jie. Small object detection based on improved Faster-RCNN[J]. *Modern Computer*, 2023, 29(14): 14–18. (in Chinese)
- [12] 贵向泉, 秦庆松, 孔令旺. 基于改进 YOLO v5s 的小目标检测算法[J]. *计算机工程与设计*, 2024, 45(4): 1134–1140.  
GUI Xiangquan, QIN Qingsong, KONG Lingwang. Small object detection algorithm based on improved YOLO v5s[J]. *Computer Engineering and Design*, 2024, 45(4): 1134–1140. (in Chinese)
- [13] LIU Y, SHI M, ZHAO Q, et al. Point in, box out: beyond counting persons in crowds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 6462–6471.
- [14] SAM D B, PERI S V, SUNDARARAMAN M N, et al. Locate, size, and count: accurately resolving people in dense crowds via detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(8): 2739–2751.
- [15] LIU C, WENG X, MU Y. Recurrent attentive zooming for joint crowd counting and precise localization[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 1217–1226.
- [16] XU C, LIANG D, XU Y, et al. Autoscale: learning to scale for crowd counting[C]//IEEE International Conference on Computer Vision (ICCV), 2022: 1–30.
- [17] ARTETA C, LEMPITSKY V, ZISSERMAN A. Counting in the wild[C]//European Conference on Computer Vision, 2016: 483–498.
- [18] ABOUSAMRA S, HOAI M, SAMARAS D, et al. Localization in the crowd with topological constraints[C]//AAAI Conference on Artificial Intelligence, Virtual Event, 2021: 1–17.
- [19] SHEN Z Y, LI M. Domain adaptation via model-independent inverse mapping and model reuse[C]//2018 IEEE International Conference on Data Mining (ICDM), 2018: 1224–1229.
- [20] 张善文, 许新华, 齐国红. 基于空洞空间金字塔池化 U-Net 的遥感图像多目标检测方法[J]. *弹箭与制导学报*, 2023, 43(5): 1–8.

- ZHANG Shanwen, XU Xinhua, QI Guohong. Remote sensing image multi-object detection method based on dilated spatial pyramid pooling U-Net[J]. *Journal of Missiles and Guided Weapons*, 2023, 43(5): 1-8. (in Chinese)
- [21] 周晓玲, 张朝霞, 鲁雅, 等. 基于改进 R-FCN 的 SAR 图像识别[J]. *系统工程与电子技术*, 2022, 44(4): 1202-1209. ZHOU Xiaoling, ZHANG Chaoxia, LU Ya, et al. SAR image recognition based on improved R-FCN[J]. *Systems Engineering and Electronics*, 2022, 44(4): 1202-1209. (in Chinese)
- [22] 陶旭. 面向点云数据的树种识别方法研究与应用[D]. 保定: 河北农业大学, 2023. TAO Xu. Research and application of tree species recognition methods for point cloud data[D]. Baoding: Hebei Agricultural University, 2023. (in Chinese)
- [23] 杨萌, 李培峰, 朱巧明. 一种基于 Tree-LSTM 的句子相似度计算方法[J]. *北京大学学报(自然科学版)*, 2018, 54(3): 481-486. YANG Meng, LI Peifeng, ZHU Qiaoming. A Tree-LSTM based sentence similarity computation method[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2018, 54(3): 481-486. (in Chinese)
- [24] SCHÜRHOLOZ D, CASTELLANOS-GALINDO G A, CASELLA E, et al. Seeing the forest for the trees: mapping cover and counting trees from aerial images of a mangrove forest using artificial intelligence[J]. *Remote Sensing*, 2023, 15(13): 3334.
- [25] CHEN L, WEI Y, YAO Z, et al. Data augmentation in prototypical networks for forest tree species classification using airborne hyperspectral images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-16.
- [26] YE J, GUO S, LI X, et al. PFIDT: pyramid focal inverse distance transformer for crowd localization[C]//2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS), 2024: 549-553.
- [27] HUANG W, WU X, WANG C, et al. Adaptive focal inverse distance transform maps for cell recognition[C]//International Conference on Neural Information Processing, 2023: 148-160.
- [28] LIANG D, XU W, ZHU Y, et al. Focal inverse distance transform maps for crowd localization[J]. *IEEE Transactions on Multimedia*, 2022, 25: 6040-6052.
- [29] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3349-3364.
- [30] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. Imagenet classification with deep convolutional neural networks[C]//Neural Information Processing Systems, 2012: 1097-1105.
- [31] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [32] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 589-597.
- [33] LI Y, ZHANG X, CHEN D. Csrnet: dilated convolutional neural networks for understanding the highly congested scenes[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1091-1100.
- [34] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning, 2019: 6105-6114.
- [35] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words; transformers for image recognition at scale[C]//International Conference on Learning Representations, 2020: 1-28.
- [36] FANG S, PAN X, XIANG S, et al. Meta-msnet: meta-learning based multi-source data fusion for traffic flow prediction[J]. *IEEE Signal Processing Letters*, 2020, 28: 6-10.
- [37] DAI F, LIU H, MA Y, et al. Dense scale network for crowd counting[C]//International Conference on Multimedia Retrieval, 2021: 64-72.
- [38] KE L, TAI Y W, TANG C K. Deep occlusion-aware instance segmentation with overlapping bilayers[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4019-4028.