北京林业大学学报
JOURNAL OF BEIJING FORESTRY UNIVERSITY

基于改进CenterNet的无人机森林病虫害图像目标检测方法

梁大双 刘文萍 赵玉刚 宗世祥 骆有庆

UAV-based forest pest image object detection using improved CenterNet

Liang Dashuang, Liu Wenping, Zhao Yugang, Zong Shixiang, Luo Youqing

引用本文:
梁大双, 刘文萍, 赵玉刚, 宗世祥, 骆有庆. 基于改进CenterNet的无人机森林病虫害图像目标检测方法[J]. 北京林业大学学报, 2026, 48(3):140–151. doi: 10.12171/j.1000–1522.20240018
Liang Dashuang, Liu Wenping, Zhao Yugang, Zong Shixiang, Luo Youqing. UAV–based forest pest image object detection using improved CenterNet[J]. Journal of Beijing Forestry University, 2026, 48(3):140–151. doi: 10.12171/j.1000–1522.20240018

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20240018

您可能感兴趣的其他文章
Articles you may be interested in

基于Mask R–CNN和迁移学习的无人机遥感影像杉木单木树冠提取
UAV remote sensing image extraction of single tree crown of Chinese fir based on Mask R–CNN and transfer learning
北京林业大学学报, 2024, 46(3): 153–166. https://doi.org/10.12171/j.1000–1522.20210343

改进YOLOv8n的林业害虫检测方法
Forestry pest detection method based on improved YOLOv8n
北京林业大学学报, 2025, 47(2): 119–131. https://doi.org/10.12171/j.1000–1522.20240326

基于改进RRT与GA的多目标路径规划
Multi–objective path planning based on improved RRT and GA: taking UAV forest inspection as an example
北京林业大学学报, 2025, 47(4): 129–141. https://doi.org/10.12171/j.1000–1522.20240356

基于无人机激光雷达的天然林林隙识别
Forest gap identification in natural forest based on UAV LiDAR
北京林业大学学报, 2022, 44(6): 44–53. https://doi.org/10.12171/j.1000–1522.20210524

面向钻蛀振动实时侦听的深度学习模型压缩
Deep learning model compression for real–time listening of boring vibration
北京林业大学学报, 2021, 43(6): 92–100. https://doi.org/10.12171/j.1000–1522.20200100

基于深度学习和上采样的杉木单木枯损模型构建
High–precision individual tree mortality prediction for Chinese fir using deep learning and oversampling
北京林业大学学报, 2025, 47(12): 39–47. https://doi.org/10.12171/j.1000–1522.20250103

DOI:10.12171/j.1000−1522.20240018

基于改进 CenterNet 的无人机森林病虫害图像目标检测方法

梁大双[1,2] 刘文萍[1,2] 赵玉刚[1,2] 宗世祥[3] 骆有庆[3]
（1. 北京林业大学信息学院，北京 100083；2. 国家林业草原林业智能信息处理工程技术研究中心，北京 100083；
3. 北京林业大学林学院，北京 100083）

摘要:【目的】为提高无人机图像中受害木的检测精度，在 CenterNet 目标检测方法基础上进行了改进，提出一种改进的检测方法——ECenterNet。**【方法】**(1)针对无人机图像中受害木外接框多为长方形的特点，用二维椭圆高斯替换原 CenterNet 方法中的二维圆形高斯来生成受害木中心点周围的惩罚区域，使得生成的惩罚区域与实际受害木的外接框形状更加匹配。(2)在大规模无标签的无人机图像数据集上，采用自监督学习方法训练 ResNet101 网络以获得用于受害木目标检测网络的预训练模型，为受害木目标检测模型提供丰富的先验信息，从而提升受害木检测精度。(3)在 CenterNet 原有损失函数的基础上，借鉴自监督学习的思想，设计了对比学习损失函数。该函数可使模型提取的受害木特征在类内更聚集、类间更可分，从而提升了模型类别判定的精度。(4)在目标检测网络的类别层分支中，利用中心差分卷积替换传统卷积，使得模型的类别分支层不仅能提取到图像的语义信息，还能提取到当前像素与周围像素的差分信息，为模型的分类提供了更多的有用信息。**【结果】**实验结果表明，ECenterNet 模型在几乎不增加推理耗时的情况下，mAP@[0.5，0.95]从原始 CenterNet 的 0.498 提升至 0.543，精度提高了 4.5 个百分点。其中，自监督预训练与对比学习损失函数的引入贡献了 2.4% 的性能增益，充分验证了所提优化算法及自监督学习策略的有效性。**【结论】**本文所提出的 4 点改进在几乎不增加模型推理计算成本的前提下，有效提升了模型精度，验证了 ECenterNet 方法的有效性，为森林病虫害的精准监测提供了有力的技术支撑。

关键词：无锚框检测器；CenterNet；受害木检测；深度学习；无人机
中图分类号：S771.8；TP751.1 **文献标志码：**A **文章编号：**1000−1522(2026)03−0140−12
引文格式：梁大双，刘文萍，赵玉刚，等. 基于改进 CenterNet 的无人机森林病虫害图像目标检测方法［J］. 北京林业大学学报，2026，48(3)：140−151. Liang Dashuang, Liu Wenping, Zhao Yugang, et al. UAV-based forest pest image object detection using improved CenterNet［J］. Journal of Beijing Forestry University, 2026, 48(3): 140−151.

UAV-based forest pest image object detection using improved CenterNet

Liang Dashuang[1,2] Liu Wenping[1,2] Zhao Yugang[1,2] Zong Shixiang[3] Luo Youqing[3]
(1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China;
2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China;
3. College of Forestry, Beijing Forestry University, Beijing 100083, China)

Abstract:［**Objective**］To enhance the detection accuracy of infested individual trees in Unmanned Aerial Vehicle (UAV) imagery, this study proposes an improved detection method termed ECenterNet, built upon the foundational CenterNet architecture.［**Method**］(1) According to the characteristics that most boundary boxes of the infested trees in UAV forest images are rectangular rather than square, a 2D elliptical Gaussian

收稿日期: 2024−01−15 **修回日期:** 2025−12−23
基金项目: 国家重点研发计划(2021YFD1400900)，松材线虫病防控关键技术研究与示范(ZD202001)。
第一作者: 梁大双，博士。主要研究方向：林业图像分析技术。Email：liangdas1986@163.com 地址：100083 北京市海淀区清华东路 35 号北京林业大学信息学院。
责任作者: 刘文萍，教授，博士生导师。主要研究方向：数字图像处理与模式识别。Email：wendyl@vip.163.com 地址：同上。
本刊网址: https://j.bjfu.edu.cn；https://journal.bjfu.edu.cn

rather than the 2D circular Gaussian is used to generate the penalty area for pixels near the center of infested trees because 2D elliptical Gaussian can better adapt to the shape of the infested trees's boundary box. (2) A ResNet101 pre-training model is trained on a large number of unlabeled UAV forest image data sets by a self-supervised learning method. As a pre-training model of the infested trees detection network, this model provides rich prior information for the detection model, which can at last improve the accuracy of the infested trees detection. (3) Besides the existing loss functions of the original CenterNet method, a contrastive learning loss function is also designed based on the idea of a self-supervised learning method. This function can make the infested trees characteristics extracted from the model gather within the class, and separate between classes as much as possible, which improves the classification accuracy of the model. (4) In the branch of the category layer in the network, the central difference convolution is used to replace the original ordinary convolution, so that the category branch layer of infested trees detection model can not only extract the semantic information of the image but also extract the difference information, which can provide more information for the classification of models. [**Result**] After the above improvements, the mAP@[0.5, 0.95] accuracy of the proposed model is improved from 0.498 to 0.543, with a total increase of 4.5% when compared to the original CenterNet model, while the inference time only increase a little. [**Conclusion**] The improved method can effectively improve the detection accuracy of the model and provide technical support for accurate monitoring and control of forest pests and diseases.

Key words: Anchor-free detector; CenterNet; infested trees detection; deep learning; UAV

森林病虫害是严重威胁森林生长和林业可持续发展的重要因素，是森林健康和林业生产的大敌[1]。传统的森林病虫害监测手段主要依赖人工实地调查，成本较高，存在时间滞后性，且易受主观因素影响。在地势较险且行动受限的地区，病虫害灾情的实时监测更具挑战性[2]。随着遥感技术的发展，航空航天遥感影像已被广泛应用于林业病虫害监测[3−5]。相较于高空遥感技术，无人机(unmanned aerial vehicle，UAV)作为一种低空遥感工具，具有成本低、成像精度高、操作简单和机动灵活等独特优势[6]，因而在森林病虫害监测中得到广泛应用[7−8]。

与计算机视觉中的目标检测任务类似，无人机森林病虫害图像目标检测旨在从图像中检测出病虫害相关信息(如病虫害植株数量、空间分布及其危害程度等)，并准确回归病虫害植株在图像上的外接框坐标，是实现森林病虫害监测的关键技术之一[9−11]。目前，基于计算机视觉技术的森林病虫害图像目标检测的方法分为基于传统机器视觉方法和深度学习方法两类。传统机器视觉方法一般是通过提取森林病虫害图像的颜色、形状以及纹理等浅层特征，并借助随机森林、人工神经网络和支持向量机等机器学习模型检测出病虫害植株的方法[12]。但这类方法对特征的依赖性强，鲁棒性不足，易受到无人机飞行姿态、光照条件、拍摄角度及植株个体差异等因素干扰，导致检测精度下降。

相较之下，深度学习方法具备自动学习图像多层特征的能力，能够同时提取浅层细节特征信息与深层语义特征，有效克服了传统机器视觉方法特征表征能力不足的问题，近年来被广泛应用于玉米(*Zea mays*)[13−14]、水稻(*Oryza sativa*)[15−16]、苹果(*Malus pumila*)[17]等作物病虫害的图像识别与监测任务中。深度学习在视觉领域的应用主要包括图像分类、目标检测和图像分割 3 大任务。其中，在目标检测任务中，大致可以分为两大类：第一类是基于锚框方法，如区域卷积神经网络(faster region-based convolutional neural networks，Faster RCNN)系列[18−20]、一次性检测(You Only Look Once，YOLO)系列[21−22]、单次多框检测(single shot multibox detector，SSD)系列[23−24]、基于区域的全卷积网络(object detection via region-based fully convolutional networks，RFCN)[25]和焦点损失用于密集物体检测(focal loss for dense object detection，RetinaNet)[26]等，这些方法需要事先预定义一组锚框；第二类是无锚框方法，如角点网络(detecting objects as paired keypoints，CornerNet)[27]、中心点检测网络(objects as points，CenterNet)[28]和极值点网络(bottom-up object detection by grouping extreme and center points，ExtremNet)[29]等，这些方法将目标检测问题看成是一个标准的关键点估计问题。无锚框方法避免了预设锚框大小、比例、个数等超参数设计问题，并减少了大量锚框和预测框之间交并比(Intersection over Union，IoU)计算带来的训练时间过长的问题。该方法不需要使用锚框，而是采用关键点回归的思路：先定义描述一个物体所需要的关键点集合，然后对物体的各个关键点进行回

归。在一系列无锚框目标检测算法中，CenterNet 以其结构简单、推理快速、准确且无需任何非极大值抑制等后处理步骤，成为当前高效的目标检测算法之一。

本文以受红脂大小蠹(*Dendroctonus valens*)危害的油松(*Pinus tabuliformis*)林为研究对象，选用 CenterNet 作为基础模型，提出了一种改进的无人机图像红脂大小蠹受害木目标检测模型——ECenterNet，以进一步提升无人机图像中受害木目标检测的精度。

1 研究区概况

研究区位于辽宁省凌源市，地处辽宁省西部，与河北省和内蒙古自治区接壤。研究区南北长 93.3 km，东西宽 66.1 km，总面积 3 278 km^2，占辽宁省总面积的 2.2%。主要植被为油松，还有少量杨树(*Populus* spp.)零星分布。近年来，红脂大小蠹病虫害已在该地区造成一定的树木损害和死亡。

为获取足量的无人机图像样本用于模型训练，基于当地林场及病虫害分布特点，在凌源市选取了 6 个样地，依次编号为样地 1 至样地 6，其具体位置分布如图 1 所示。

8 月份是当地油松的生长季节，也是捕捉叶片信息的最佳时间窗口。因此，本试验的无人机数据采集工作于 2017 年 8 月 11 日和 8 月 12 日进行。使用的无人机的型号为四旋翼 DJI Inspire2，搭载了 DJI x5 专业相机，其有效分辨率为 2 × 107 像素，采集图像的空间分辨率为 5 280 × 3 956 像素。为避免图像过度曝光或过暗，采集时避开了阳光过曝或过暗的时间段。

2 研究方法

2.1 基础框架及改进

CenterNet 是一种简单、快速且准确的无锚框目标检测器。对于一幅输入图像，CenterNet 输出 3 个分支：(1)类别热力图，包含 C 个通道(C 为类别数)，每个通道上的像素值表示该位置存在目标中心点的置信度；(2)中心点偏移量预测，包含 2 个通道，分别表示目标中心点在 x 和 y 方向上的偏移；(3)目标尺寸预测，包含 2 个通道，分别表示目标边界框的宽度和高度。整体网络结构如图 2a 所示。

对于类别分数，采用焦点损失函数(focal loss)进行学习，其计算公式如式(1)所示。

$$L_k = -\frac{1}{N}\sum_{\mathrm{xyc}}\begin{cases}\left(1-\hat{Y}_{\mathrm{xyc}}\right)^{\alpha}\ln\left(\hat{Y}_{\mathrm{xyc}}\right) & \text{if } Y_{\mathrm{xyc}}=1\\ \left(1-Y_{\mathrm{xyc}}\right)^{\beta}\left(\hat{Y}_{\mathrm{xyc}}\right)^{\alpha}\ln\left(1-\hat{Y}_{\mathrm{xyc}}\right) & \text{otherwise}\end{cases} \tag{1}$$

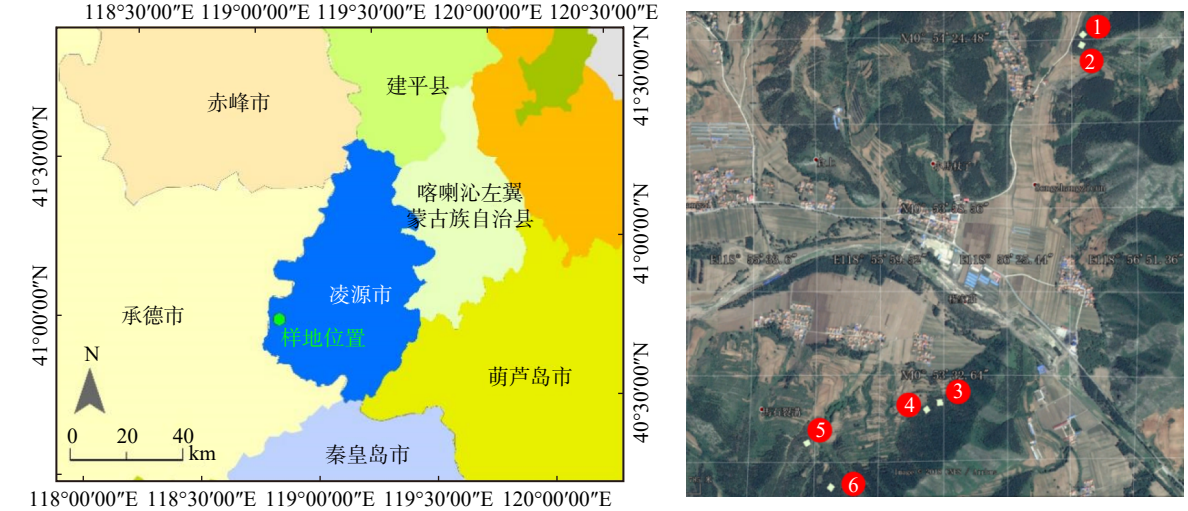
式中：N 是一张图上目标物体框的个数，$\hat{Y}_{\mathrm{xyc}}$ 是网络预测的类别分数；α 和 β 是超参数，一般取 $\alpha=2$，$\beta=4$；Y_{xyc} 是对不属于物体中心(背景)像素的惩罚值，定义如式(2)所示。

$$Y_{\mathrm{xyc}} = \exp\left(-\frac{(x-\tilde{p}_x)^2+\left(y-\tilde{p}_y\right)^2}{2\delta_p^2}\right) \tag{2}$$

式中：(x,y)为物体中心点周围某个像素的 x 轴，y 轴方向的坐标值；$\left(\tilde{p}_x,\tilde{p}_y\right)$为物体中心点 x 轴，y 轴方向的坐标值；δ_p 为高斯半径，与物体的外接框的宽和高(w,h)大小相关的超参数，其取值如式(3)所示。

$$\sigma_p = \frac{w+h}{c} \tag{3}$$

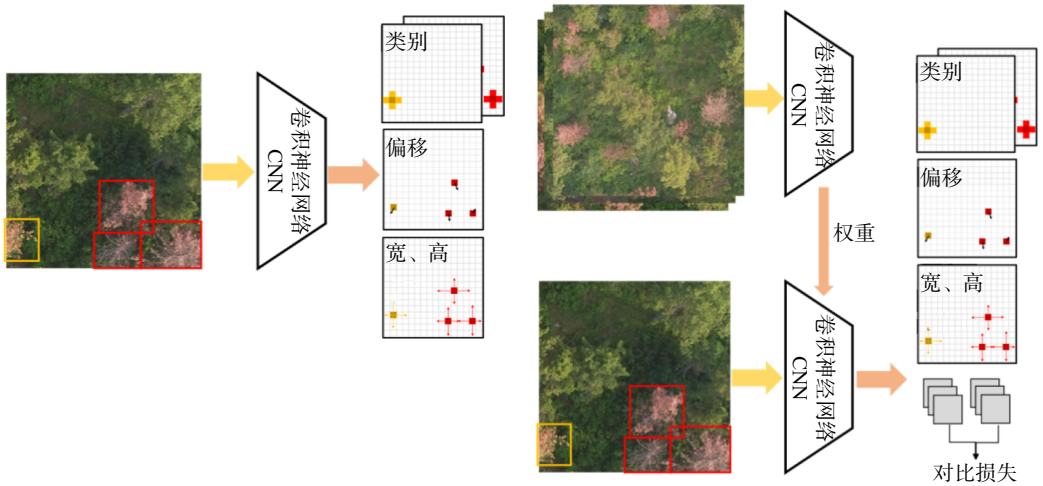
式中：w 是物体外接框的宽；h 是物体外接框的高；c



1~6 表示样地编号。

图 1 样地位置

Fig. 1 Location of the sample plot



a CenterNet 网络结构　　b ECenterNet 网络结构图

图 2　CenterNet 和 ECenterNet 网络结构图

Fig. 2　CenterNet and ECenterNet network architecture diagram

为常数，原始取值为 12。

在 CenterNet 模型中，物体中心点映射到特征图上浮点坐标和特征图上实际像素的整型坐标之间的偏移量是 $\left(\frac{x}{S_x}-\left\lfloor\frac{x}{S_x}\right\rfloor,\frac{y}{S_y}-\left\lfloor\frac{y}{S_y}\right\rfloor\right)$，CenterNet 是用 L_1 损失函数进行学习的，而且所有的类别都共用同一个偏移量，偏移量的损失函数具体表达式如下

$$L_{\text{off}}=\frac{1}{N}\sum_{p}\left(\left|\left(\hat{O}_{\tilde{p}_x}-\left(\frac{x}{S_x}-\left\lfloor\frac{x}{S_x}\right\rfloor\right)\right)\right|+\left|\left(\hat{O}_{\tilde{p}_y}-\left(\frac{y}{S_y}-\left\lfloor\frac{y}{S_y}\right\rfloor\right)\right)\right|\right) \tag{4}$$

式中：$\hat{O}_{\tilde{p}_x}$和$\hat{O}_{\tilde{p}_y}$为某个物体真实外接框的中心点映射到特征图上的浮点坐标。

对于物体的宽和高(w,h)的回归，CenterNet 也是用 L_1 损失函数来监督学习的，而且所有的类别也是共用同一个宽高的值，具体回归公式如下

$$L_{\text{size}}=\frac{1}{N}\sum_{p}\left(\left|\hat{w}_p-w_p\right|+\left|\hat{h}_p-h_p\right|\right) \tag{5}$$

式中：$\hat{w}_p$和$\hat{h}_p$为某个物体真实的宽高值，w_p和h_p为模型预测的宽高值。

总体上，整个 CenterNet 的目标函数的表达式如下

$$L_{\text{det}}=L_k+\lambda_{\text{size}}L_{\text{size}}+\lambda_{\text{off}}L_{\text{off}} \tag{6}$$

式中：L_{det}是 CenterNet 的总的损失函数，L_k是 CenterNet 的类别相关的损失函数，L_{size}是物体宽高相关的损失函数，L_{off}为物体中心坐标偏移量的损失函数，λ_{size}和λ_{off}分别是物体的宽高和物体中心坐标偏移量损失函数的权重。

本文提出了一种改进的 ECenterNet 架构(图 2b)，通过 4 项关键策略提升受害木检测性能：首先，针对受害木外接框的长宽比特征，引入二维椭圆高斯分布替代传统的圆形高斯核，以生成几何形态更匹配的惩罚区域；其次，采用自监督预训练与迁移学习相结合的策略，利用大规模无标签数据构建富含先验信息的特征提取器，以增强模型对有监督任务的表征能力；再次，设计了一种对比学习损失函数，通过优化特征空间分布，有效提升了类内特征的紧凑性与类间特征的可分性；最后，在类别预测分支中集成中心差分卷积，在保留语义信息的同时捕获像素级的局部差分细节，从而为精细分类提供更具判别力的特征表示。

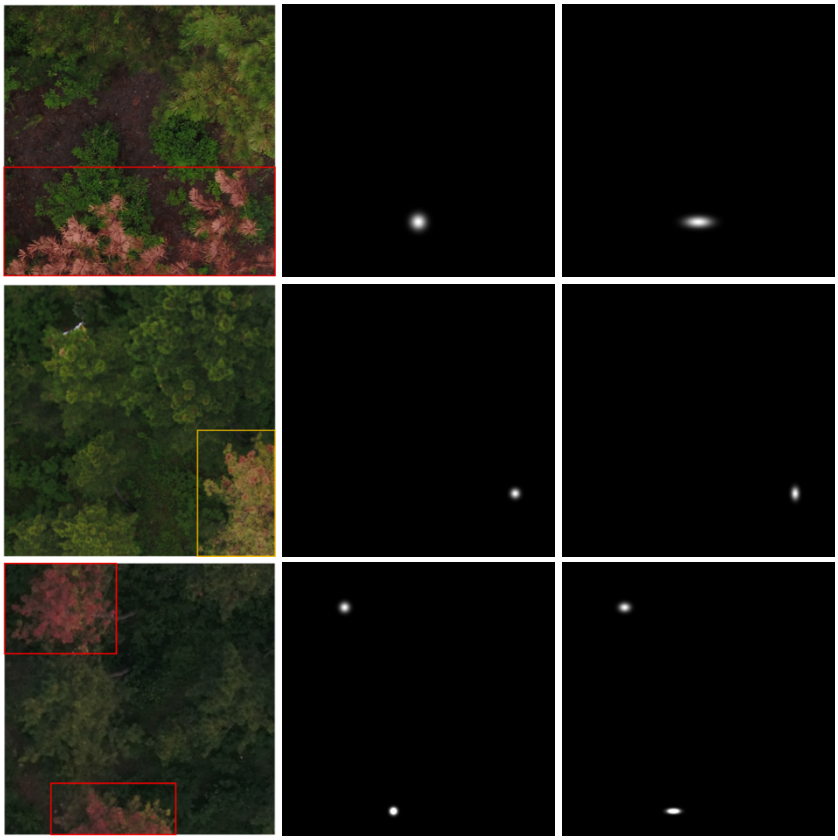
2.2　类别损失的惩罚函数优化

在原始 CenterNet 架构中，惩罚项 Y_{xyc} 被定义为二维圆形高斯分布(图 3b)，该假设隐含了目标外接框为正方形的先验条件，然而在实际无人机森林病虫害监测场景中，受个体受害木形态异质性及图像边缘截断效应的影响，受害木的外接框往往呈现显著的长宽比差异而非标准正方形。为解决这一几何失配问题并提升惩罚热力图的适应性，本研究对惩罚函数进行了优化重构，使其能够动态贴合实际受害木外接框的几何特征，改进后的数学表达如式(7)所示。

$$\begin{cases} Y_{\text{xyc}}=\exp\left(-\dfrac{(x-\tilde{p}_x)^2}{2\delta_w^2}-\dfrac{(y-\tilde{p}_y)^2}{2\delta_h^2}\right) \\ \delta_w=\dfrac{w}{6} \\ \delta_h=\dfrac{h}{6} \end{cases} \tag{7}$$

式中：w、h 分别是物体的外接框的宽和高。如此，$\pm3\delta_w$和$\pm3\delta_h$的范围分别可覆盖整个受害木外接框的宽度(w)和高度(h)。

从式(7)可以看出，它是一个二维椭圆高斯表达式，其在平面上所产生的二维分布图如图 3c 所示。



a 原图 b 圆高斯生成的热力图 c 椭圆高斯生成的热力图

图 3 不同高斯函数生成的类别损失函数的惩罚图

Fig. 3 Penalty graphs generated by different Gaussian functions

对照图 3 中的 a、b、c 可以发现：二维椭圆高斯表达式生成的惩罚区域更加贴合原图中受害木外接框形状，从而其对惩罚区域的表达也更加准确。

2.3 自监督预训练模型

在目标检测模型训练之前，通常需要一个预训练模型来初始化网络参数。该预训练模型通常是在公开的 ImageNet[30]数据集或者 COCO[31]数据集上训练得到的。然而，由于公开数据集大多是来自互联网的通用图片，这些数据与无人机森林病虫害图像数据集之间存在一定的域差异。因此，将基于公开数据集构建的预训练模型用于无人机图像受害木目标检测网络训练并不是最优选择。另一方面，随着外业数据采集工作的推进，无人机森林病虫害图像数据不断累积，图像数量也在不断增加。

为了在不增加人力标注成本的前提下，充分利用大量无标注的无人机图像来提高受害木目标检测模型的精度，本文提出了一种基于自监督表征学习的检测方案。该方案首先在无标注的无人机图像数据集上，通过 MoCoV2[32] 算法对 ResNet101 进行自监督预训练(图 4)，构建具有强鲁棒性的预训练模型。随后，将该预训练模型作为 CenterNet 的特征提取主干，并在带标注的无人机图像数据集上进行迁移训练。实验流程有效结合了自监督学习在大规模数据挖掘上的优势与有监督学习在目标检测上的精确性。

2.4 对比学习损失函数和目标检测信号同时监督网络的学习

除了利用自监督学习方法提供预训练模型外，本文基于自监督学习思想，设计了一个对比学习损失函数，并将其嵌入到当前无人机图像受害木目标检测模型 ECenterNet 的训练任务中。这样，除了常规的目标检测损失函数作为监督信号以外，本文引入自监督对比学习损失函数作为辅助监督信号，共同监督无人机图像受害木目标检测网络的学习。该损失函数旨在提升相同类别受害木中心点特征的相似度，并促进类内特征聚集，从而增强 CenterNet 模型的目标检测精度。具体策略是在 CenterNet 网络类别分数层前一层特征图中，选取与受害木中心对应的通道向量作为中心点特征表示，并通过损失函数引导这些中心特征向量在同类间高度相似并聚类。自定义的自监督对比学习损失函数的表达式如下

$$L_{\text{contrast}} = \frac{\sum_i^{M^+} S_{ik}}{\sum_i^{M^+} S_{ik} + \sum_j^{M^-} S_{jk}} \tag{8}$$

式中：k 为训练的一个批次(Batch)中选中的某一个

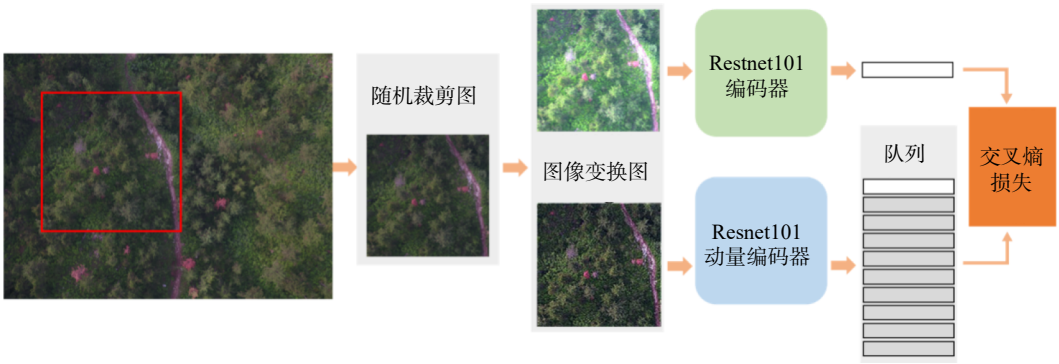


图 4　基于 ResNet101 的无人机森林病虫害图像自监督学习流程图

Fig. 4　Self supervised learning flow chart of UAV forest pest image based on ResNet101

中心点的向量，M^+为和向量 k 同类别的M^+个中心，M^-为和向量 k 不同类别的M^-个中心，S_{ik} 为从M^+个向量中选取的向量 i 与向量 k 的相似度，S_{jk} 为从M^-个向量中选取的向量 j 与向量 k 的相似度。

结合 L_{contrast} 和 CenterNet 原有的目标函数（公式 6），则带自监督的无人机森林病虫害图像目标检测模型的损失函数变为

$$L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} + \lambda_{\text{contrast}} L_{\text{contrast}} \tag{9}$$

式中：L_{det} 是 CenterNet 的总的损失函数，L_k 是类别相关的损失函数，L_{size} 是物体宽高相关的损失函数，L_{off} 为物体中心坐标偏移量的损失函数，L_{contrast} 为自定义的自监督损失函数，λ_{size}、λ_{off} 分别是物体宽高相关和物体中心坐标偏移量损失函数的权重，$\lambda_{\text{contrast}}$ 为自监督损失函数的权重。

2.5　中心差分卷积提取类别信息

中心差分卷积是在标准卷积的基础上引入中心差分操作的卷积方法。以 3 × 3 卷积核为例，传统卷积通过卷积权重与特征图上的对应 9 个像素进行逐元素相乘并累加，如图 5a 所示。中心差分卷积由两个分支构成：第一分支对 3 × 3 卷积核对应的 9 个像素与中心点像素值逐一相减，生成 9 个中心差分像素值，再与卷积核权重逐元素相乘并累加；第二分支为常规卷积操作。最终，将两个分支计算所得的特征值加权合并，如图 5b 所示。

中心差分卷积不仅能像普通卷积一样获取特征图上的语义信息，还能获取局部梯度信息。对于图 2b 中，ECenterNet 网络类别分数分支输出的特征图层中，只有当某个像素的类别分数值都大于其 8 邻域内所有像素的类别分数值时，才能确定该像素代表某个物体。该像素值与其 8 邻域内像素值比较的过程就是计算中心差分的过程。因此，引入中心差分卷积有利于类别分支更好地提取差分信息，从而判断输出特征图层中的某个像素是否能代表物体以及其类别。

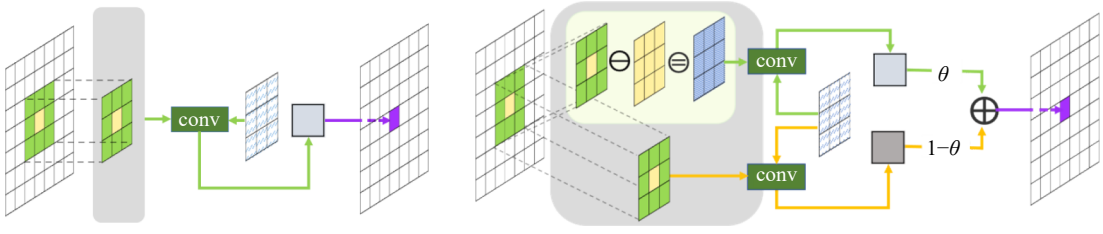
2.6　测试指标

最终的检测结果，将使用官方 COCO API[33]进行评价指标的计算，它是在 IoU 阈值从 0.50 到 0.95 之间以 0.05 的步长来计算目标检测模型的平均精确率（mAP），简单地表示为 mAP@[0.5，0.95]。其中，IOU 计算如式（10）所示。

$$\text{IOU}(A, B) = \frac{A \cap B}{A \cup B} \tag{10}$$

式中：A 为真实的标记框（ground truth box），B 为网络的预测框（predicted box）。

平均精度（AP）反映的是模型在某一个类别上的效果。首先，以召回率（Recall）为横坐标、以精确率（Precision）为纵坐标绘制单个类别的 Precision-Recall（P-R）曲线，然后对召回率分别从 0 到 1 以步长为 0.1 取 11 个采样点，再通过公式（11）计算得到平均精度。



a 普通卷积操作　　b 中心差分卷积操作

图 5　普通卷积和中心差分卷积结构图

Fig. 5　Structure of ordinary convolution and center difference convolution

$$\mathrm{AP}=\frac{1}{11}\sum_{r\in\{0,0.1,\cdots,1\}}P(r)$$
$$P(r)=\max_{\tilde{r}:\,\tilde{r}\geqslant r}p(\tilde{r}) \tag{11}$$

式中：r 表示召回率；$P(r)$表示在召回率 r 处的精确率，它是通过在一定区间内取最大值得到的一个值；$p(\tilde{r})$表示在召回率大于 r 的区间内的最大精确率的值。

如果一个研究中有多个类别，则用上述方法分别计算每个类别的 AP 值之后，再将所有类别的 AP 值求平均，得到均值平均精度 mAP 指标，其数学计算如公式(12)所示。

$$\mathrm{mAP}=\frac{1}{M}\sum_{j=1}^{M}\mathrm{AP}_i \tag{12}$$

式中：M 是类别的个数，AP_i表示第 i 个类别的 AP 值。

不同的 IoU 阈值对应着不同的 P-R 曲线，不同的 P-R 曲线又对应着不同的 AP 值，所以，不同的 IoU 阈值对应着不同的 AP 值，也就对应着不同的 mAP 值。一般的，mAP@[0.5] 表示在 IoU 为 0.5 的阈值条件下统计得到的 mAP 值，mAP@[0.75] 表示在 IoU 为 0.75 的阈值条件下统计得到的 mAP 值。为了统计指标的稳定，常常在 IoU 为 0.5~0.95 区间内以步长 0.05 取不同的 IoU 值对应的不同的 mAP 值，然后将这些 mAP 值求均值得到 mAP@[0.5, 0.95] 指标。

3 结果与分析

3.1 自监督模型的数据处理与训练

在自监督模型训练中，使用约 600 张无人机拍摄的原始森林病虫害图像，每张图像分辨率为 5 280 × 3 956 像素。不同于有监督模型训练，自监督模型无需人工标注图像。为与下游有监督受害木目标检测模型参数一致，所有模型输入尺寸均设为 512 × 512 像素。对于每张 5 280 × 3 956 像素的原始图像，先用 1 000 ~ 2 000 像素的随机窗口进行裁剪，生成一系列正方形小图，再将其统一缩放至 512 × 512 像素，输入自监督模型进行训练。为便于迁移学习，选择的主干网络与有监督受害木目标检测模型保持一致，均采用经典的 ResNet101 网络结构。

自监督模型训练过程的准确率变化曲线和损失函数值的下降曲线如图 6 所示。从图中可以看出，经过 0 ~ 2 000 次迭代，准确率从 0 逐渐升高到接近 100%，而损失函数值则逐渐下降至 2 左右，表明该自监督模型训练过程收敛。

3.2 有监督模型的数据处理与训练

在实地调查中，绿叶植株为健康植株，黄叶植株为感染期受害木，红叶植株则为枯死受害木。本研究仅关注感染受害木和枯死受害木，因此，在有监督训练任务人工标记数据时，仅标注了黄色和红色受害木。在本文中，样地 1、3、4 和 6 中的无人机图像被划分到训练和验证数据集中，而样地 2 和 5 中的图像则被划分为测试数据集。

由于 PyTorch 是目前著名且高效的深度学习框架之一，本试验选用 PyTorch 进行模型训练。在有监督的无人机图像受害木目标检测任务中，网络输入大小固定为 512 × 512 像素，类别数为 2，包括“死亡”和“感染”两类。无论原始图像大小如何，输入网络之前都会经过仿射变换将其缩放到 512 × 512 像素，然后输入 ECenterNet 网络进行训练。训练是在一张 12 GB 显存的 NVIDIA Titan GPU 上进行，从第 0 个开始，持续训练到第 120 个 epoch，训练时间约为 3 d，学习策略用的是 Adam。详细的训练参数如表 1 所示。

3.3 各个模型的结果对比

基于相同训练集，分别训练经典的有锚框与无

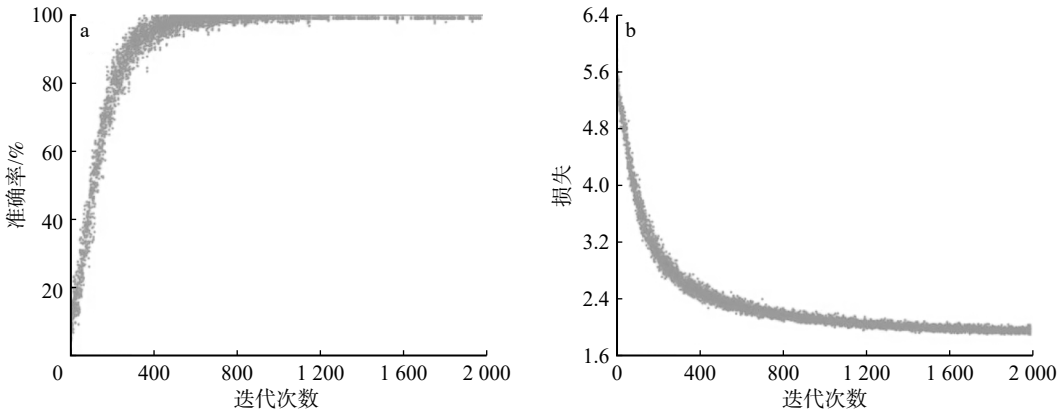


图 6 准确率和损失的随模型训练迭代步数的变化

Fig. 6 The change of accuracy and loss with the number of model training iterations

表 1　网络训练参数

Tab. 1　Network training parameters

参数	值
批次大小	8
训练轮次	120
学习策略	Multistep
步数	60, 80
初始学习率	1.25^{-4}
动量项系数	0.1

锚框目标检测模型，并在统一测试集上对比评估（表 2）。由表 2 可知，在所有方法中，本文所提出的 ECenterNet 方法的精度最高，mAP@[0.5，0.95] 达到了 0.543，比 CenterNet 精度 mAP@[0.5，0.95] 高 4.5 个百分点。在基于锚框方法中，RetinaNet 的精度最高，其 mAP@[0.5，0.95] 达到了 0.480，分别比 SSD 方法和 Faster RCNN 高出 2.9 和 0.8 个百分点。在无锚框方法中，CornerNet 的精度 mAP@[0.5，0.95] 最低，但它比基于锚框方法中的最好精度的 RetinaNet 高出 1.1 个百分点。

3.4　不同主干网络的结果对比

模型运行时间不仅受模型方法影响，还受主干网络结构影响。为评估不同主干网络对 ECenterNet 模型运行时间的影响，分别使用 ResNet-101、Hourglass-104 和 ResNet-101-FPN 作为主干网络在相同训练集上训练模型，并在统一测试环境（英特尔 i7-8086K 处理器，英伟达 Titan Xp 显卡）下进行测试，结果如表 3 所示。主干网络为 ResNet-101 时，ECenterNet 和 CenterNet 的处理速度为每秒 45.0 帧；采用 ResNet-101-FPN 时，每秒处理帧数降至 28.0，但相比以 ResNet-101 为主干网络时，mAP@[0.5，0.95] 提高了 0.006；当主干网络为 Hourglass-104 和 Hourglass-104-FPN 时，mAP@[0.5，0.95] 分别为 0.551 和 0.559，处理帧数分别为 7.8 和 4.0。因此，为兼顾精度和速度，最终选择 ResNet-101 作为主干网络。

3.5　消融试验及相关结果

为验证本文提出的 4 个改进点的有效性，进行了系列补充试验，每个试验在前一试验基础上增加一个改进点，结果如表 4 所示。总体而言，4 个改进点均有效提升模型性能，其中“类别损失的惩罚函数优化”和“自监督预训练模型”效果显著，分别提高 mAP@[0.5，0.95]1.3% 和 1.4%；“对比学习损失函数”和“中心差分卷积”则分别提高 1.0% 和 0.8%。结合自监督相关的两项改进，“自监督预训练模型”和“对比学习损失函数”可使 mAP@[0.5，0.95] 总体提高 2.4%。

尽管中心差分卷积计算量是普通卷积的两倍，但本文仅在原始 CenterNet 网络的类别特征图分支的前一层引入此操作，整个网络的计算量（乘加累积操作数）仅从 100.021 9 G 增加至 100.022 0 G，增量

表 2　不同网络模型的结果对比

Tab. 2　Comparison of results of different network models

方法	类型	主干网络	均值平均精度（IoU阈值为0.5）mAP@[0.5]	均值平均精度（IoU阈值为0.75）mAP@[0.75]	均值平均精度（IoU阈值为0.5～0.95）mAP@[0.5,0.95]
Faster RCNN[20]	有锚框	ResNet-101	0.693	0.535	0.472
SSD[23]	有锚框	ResNet-101	0.573	0.489	0.451
RetinaNet[26]	有锚框	ResNet-101-FPN	0.724	0.546	0.480
CornerNet[27]	无锚框	Hourglass-104	0.722	0.559	0.491
CenterNet[28]	无锚框	ResNet-101	0.703	0.557	0.498
ECenterNet（本文）	无锚框	ResNet-101	0.762	0.604	0.543

表 3　不同主干网络模型的结果对比

Tab. 3　Comparison of results of different models with different backbone

方法	主干网络	帧每秒 FPS	均值平均精度（IoU阈值为0.5）mAP@[0.5]	均值平均精度（IoU阈值为0.75）mAP@[0.75]	均值平均精度（IoU阈值为0.5～0.95）mAP@[0.5, 0.95]
CenterNet[27]	ResNet-101	45.0	0.703	0.557	0.498
ECenterNet	ResNet-101-FPN	28.0	0.768	0.611	0.549
ECenterNet	Hourglass-104	7.8	0.770	0.615	0.551
ECenterNet	Hourglass-104-FPN	4.0	0.774	0.618	0.559
ECenterNet	ResNet-101	45.0	0.762	0.604	0.543

表 4 不同改进点的消融试验

Tab. 4 Ablation experimental results of different improvements

方法	均值平均精度（IoU阈值为0.5）mAP@[0.5]	均值平均精度（IoU阈值为0.75）mAP@[0.75]	均值平均精度（IoU阈值为0.5~0.95）mAP@[0.5，0.95]
中心点检测网络	0.703	0.557	0.498
+ 类别损失的惩罚函数优化	0.725	0.565	0.511
+ 自监督预训练模型	0.745	0.582	0.525
+ 对比学习损失函数	0.752	0.596	0.535
+ 中心差分卷积	0.762	0.604	0.543

注：“+”表示是在上一行的基础上，累积增加当前的优化方法。

相对于原始计算量可忽略不计。因此，本文提出的 4 个改进点在模型推理阶段几乎不增加计算量。

3.6 模型结果样例对比分析

图 7 呈现了优化前 CenterNet 模型与优化后 ECenterNet 模型的检测结果对比，图 7a 展示图像中部存在两株空间位置重叠的受害木，空间干扰为该场景的主要检测挑战；图 7b 为叶片已完全脱落的受害死亡木，目标因叶片缺失呈现形态异化的特性；图 7c 为处于图像边缘、仅部分区域处于图像视场内的受害木，目标信息不完整进一步增加检测难度；而图 7d 则呈现受害木枝条末梢光线条件较差、叶片存在脱落的场景，其枝条边缘模糊暗淡、与背景界限辨识度低，显著提升了检测任务的复杂度。对比分析结果表明，在上述 4 类复杂检测场景中，ECenterNet 模型的检测性能均优于原始 CenterNet 模型，验证了优化后模型对复杂检测场景的适应性显著提升。

3.7 对比学习损失函数对类别特征的聚集作用分析

为分析加入“对比学习损失函数”对模型效果的影响，分别训练得到未加入和加入该损失函数的两个模型，并在测试数据集上提取“对比学习损失函数”前一层对应物体中心点的 256 维特征。将这些特征进行高维可视化，结果如图 8 所示。图 8 左侧展示了受害木个体从原始图像经裁剪及空间映射后，在高维特征空间中的可视化图；右侧则是 256 维特征的高维可视化分布。

由图 8 可知，加入“对比学习损失函数”后，带有红色边框的“死亡”类受害木和带有黄色边框的“感染”类受害木均表现出更明显的聚集趋势，两类受害木之间的间距更为分明。这表明，“对比学习损失函数”能够增强相同类别受害木的聚集程度，并促进不同类别受害木的分离，从而实现类内聚集和类间分离的目的。

3.8 检测难点分析

虽然本文提出的方法提升了无人机图像受害木目标检测模型的精度，但仍有一些客观因素影响模型性能。一方面，“感染”期受害木的颜色差异较大，影响检测效果——感染初期受害木呈现绿偏黄色（图 9 第 1、2 张图所示），中期呈浅黄色（图 9 的第 3 张图所示），后期则为黄偏红色（图 9 的第 4 张图所示）。这种大幅度颜色变化增加了“感染”受害木的

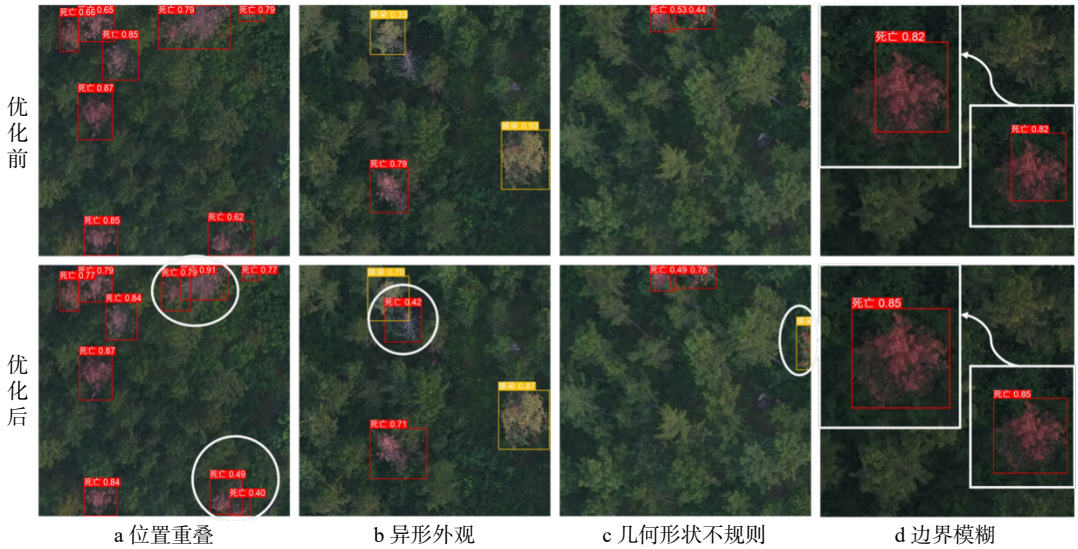


图 7 优化前后模型的结果图

Fig. 7 Result images of the model before and after optimization

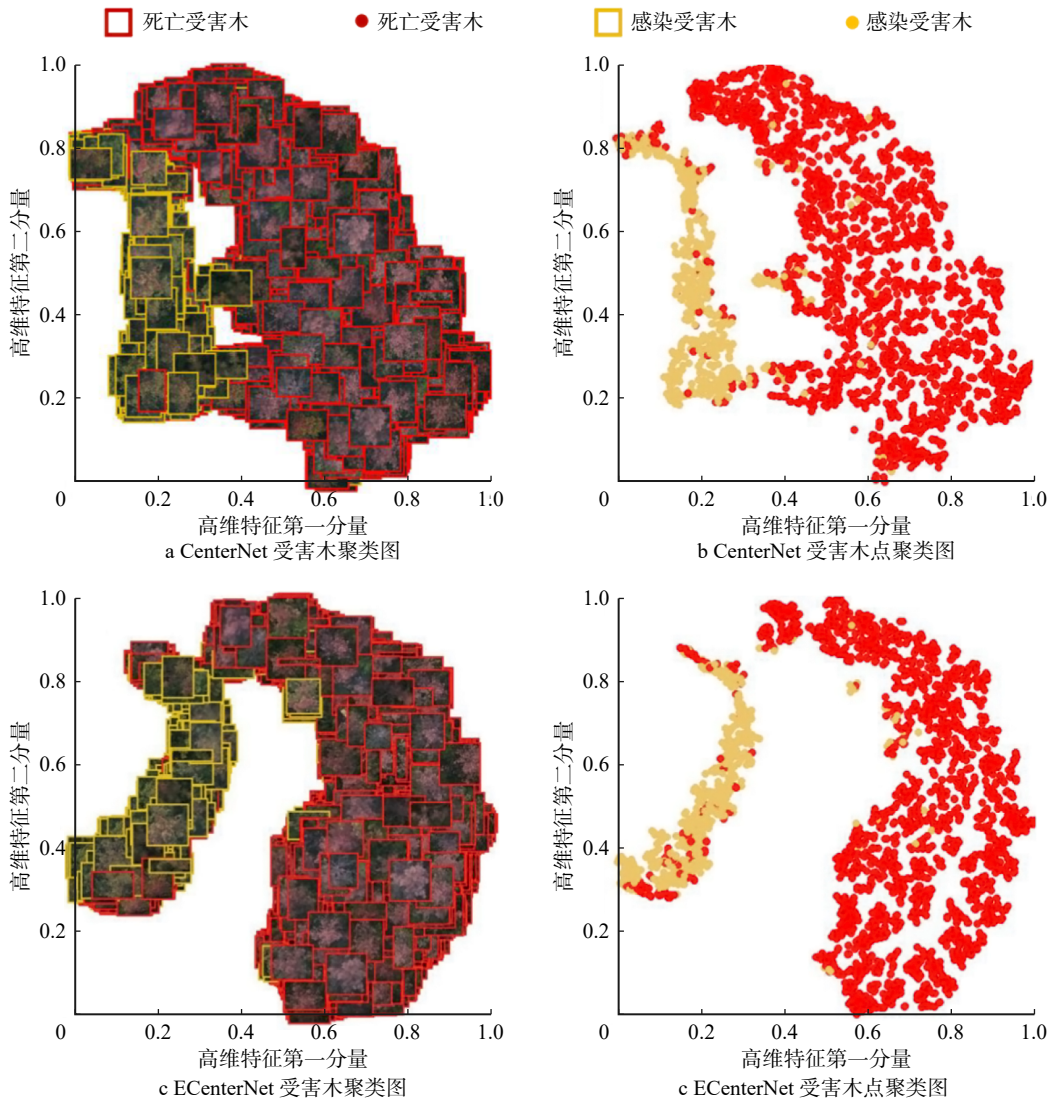


图 8　加入对比学习损失函数前后的分类特征高维可视化图

Fig. 8　High-dimensional visualization of classification features before and after adding self-contrast loss

检测难度，感染初期受害木与健康植株颜色差异小，导致检测困难，而感染后期受害木容易被误报为红色“死亡”类受害木。另一方面，图像边缘处的受害木由于仅部分位于图像内，形状不完整(图 9 第一张图中的“死亡”受害木)，进一步加大了检测难度。因此，后续工作可考虑引入多光谱等多数据源以丰富数据，提高受害木目标检测精度。

3.9　研究样地的受害木类别占比统计

为评估 6 块样地中受害木的类别分布，将各样地的无人机遥感图像输入 ECenterNet 模型进行检测；基于模型输出结果，统计了“感染”与“死亡”两类受害木的外接框像素面积占整幅图像总面积的比例，并与实地调查所得的真实面积比例进行了对比分析，结果如表 5 所示。结果显示，样地 3、样地 5 和样地 6 受灾严重，“死亡”加“感染”受害木面积占比分别达到 11.63%、9.53% 和 8.32%；样地 1、样地 2 和样地 4 受灾较轻，“死亡”加“感染”受害木面积占比分别为 1.04%、2.09% 和 1.25%。此外，每个样地内“死亡”

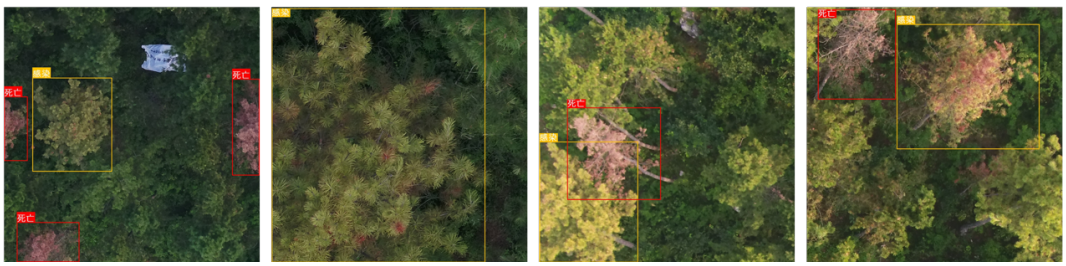


图 9　“感染”植株的图片样例

Fig. 9　Samples of ‘infected’ plant

表 5 研究样地的森林受害木面积占比

Tab. 5 The statistics of infested trees in study sample plot %

类别		样地1	样地2	样地3	样地4	样地5	样地6
模型	死亡	1.04	1.50	9.55	1.15	7.28	5.77
	感染	0.00	0.59	2.08	0.10	2.25	2.55
真实	死亡	1.18	1.65	10.72	1.39	7.64	5.71
	感染	0.00	0.48	2.32	0.09	2.17	2.65

受害木面积普遍大于“感染”期受害木，可能原因是数据采集前实施了病虫害防治措施，遏制了病虫害蔓延，减少了感染期受害木的数量。

4 结 论

本文提出了一种基于改进 CenterNet 的无人机图像受害木目标检测方法——ECenterNet。通过使用二维椭圆高斯生成受害木外接框中心附近像素的惩罚区域、借助自监督学习构建预训练模型、增加自定义“对比学习损失函数”来监督模型学习、以及引入中心差分卷积获取更细节的差分特征等方法，对原始 CenterNet 算法进行了改进并提升了精度。

（1）改进的网络模型在测试集上的精度 mAP@[0.5，0.95] 由原始 CenterNet 的 0.498 提升至 0.543，提高了 4.5 个百分点，且模型计算量和推理时间基本未增加，充分证明了所提优化算法的优越性。

（2）与自监督学习相关的“构建自监督预训练模型”和“对比学习损失函数”两个改进点，共同贡献了 2.4 个百分点的精度提升，充分验证了自监督学习方法在本研究中的有效性。

（3）尽管本文方法提高了无人机图像受害木目标检测模型的精度，仍受“感染”期受害木颜色变化显著和图像边缘受害木不完整等因素影响。因此，后续研究可考虑引入多光谱图像及相关图像分析方法，以进一步提升模型精度。

参 考 文 献

[1] 张田, 张晓丽, 刘红伟, 等. 遥感在森林病虫害监测中的应用研究 [J]. 安徽农业科学, 2010, 38(21): 11604−11607.
Zhang T, Zhang X, Liu H, et al. Application of remote sensing technology in monitoring gorest diseases and pests [J]. Anhui Agricultural Science, 2010, 38(21): 11604−11607.

[2] Chiu S F. Investigations on botanical insecticides in south China - an update[C]. Botanical Pesticides in Integrated Pest Management, Rajahmundry: Indian Society of Tobacco Science, 1993: 134−137.

[3] Lehmann J R K, Nieberding F, Prinz T, et al. Analysis of unmanned aerial system-based CIR images in forestry: a new perspective to monitor pest infestation levels [J]. Forests, 2015, 6(3): 594−612.

[4] 高永刚, 刘丹, 张福娟, 等. 卫星遥感在森林病虫害监测上的研究进展 [J]. 中国农学通报, 2006, 22(2): 113−117.
Gao Y G, Liu D, Zhang F J, et al. The application development of satellite remote sensing technology in the assessment of forest damage [J]. Chinese Agricultural Science Bulletin, 2006, 22(2): 113−117.

[5] 安树杰. 应用遥感与 GIS 的松材线虫病预测模型的研究 [D]. 北京: 北京林业大学, 2006.
An S J. The research on predictive model of *Bursaphelenchus xylophilus* by RS and GIS [D]. Beijing: Beijing Forestry University, 2006.

[6] 唐晏. 基于无人机采集图像的植被识别方法研究 [D]. 成都: 成都理工大学, 2014.
Tang Y. Research on the vegetation identification method based on UAV image acquisition [D]. Chengdu: Chengdu University of Technology, 2014.

[7] 张瑞瑞, 夏浪, 陈立平, 等. 基于 U-Net 网络和无人机影像的松材线虫病变色木识别 [J]. 农业工程学报, 2020, 36(12): 61−68.
Zhang R R, Xia L, Chen L P, et al. Recognition of wilt wood caused by pine wilt nematode based on U-Net network and unmanned aerial vehicle images [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(12): 61−68.

[8] 童彤, 林思美, 李林源, 等. 联合微波与光学时间序列影像的马尾松林松材线虫病遥感识别 [J]. 北京林业大学学报, 2024, 46(3): 40−52.
Tong T, Lin S M, Li L Y, et al. Remote sensing recognition of pine wilt disease in *Pinus massoniana* forest combined with microwave and optical time-series images [J]. Journal of Beijing Forestry University, 2024, 46(3): 40−52.

[9] 黄丽明, 王懿祥, 徐琪, 等. 采用 YOLO 算法和无人机影像的松材线虫病异常变色木识别 [J]. 农业工程学报, 2021, 37(14): 197−203.
Huang L M, Wang Y X, Xu Q, et al. Recognition of abnormally discolored trees caused by pine wilt disease using YOLO algorithm and UAV images [J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(14): 197−203.

[10] 孙钰, 周焱, 袁明帅, 等. 基于深度学习的森林虫害无人机实时监测方法 [J]. 农业工程学报, 2018, 34(21): 74−81.
Sun Y, Zhou Y, Yuan M S, et al. UAV real-time monitoring for forest pest based on deep learning [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(21): 74−81.

[11] 李就好, 林乐坚, 田凯, 等. 改进 Faster R-CNN 的田间苦瓜叶部病害检测 [J]. 农业工程学报, 2020, 36(12): 179−185.

Li J H, Lin L J, Tian K, et al. Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(12): 179−185.

[12] 张军国, 冯文钊, 胡春鹤, 等. 无人机航拍林业虫害图像分割复合梯度分水岭算法 [J]. 农业工程学报, 2017, 33(14): 93−99.
Zhang J G, Feng W Z, Hu C H, et al. Image segmentation method for forestry unmanned aerial vehicle pest monitoring based on composite gradient watershed algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(14): 93−99.

[13] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别 [J]. 农业工程学报, 2017, 33(19): 209−215.
Sun J, Tan W J, Mao H P, et al. Recognition of multiple plant leaf diseases based on improved convolutional neural network [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(19): 209−215.

[14] 樊湘鹏, 周建平, 许燕. 基于改进区域卷积神经网络的田间玉米叶部病害识别 [J]. 华南农业大学学报, 2020, 41(6): 82−91.
Fan X P, Zhou J P, Xu Y. Recognition of field maize leaf diseases based on improved regional convolutional neural network [J]. Journal of South China Agricultural University, 2020, 41(6): 82−91.

[15] 刘婷婷, 王婷, 胡林. 基于卷积神经网络的水稻纹枯病图像识别 [J]. 中国水稻科学, 2019, 33(1): 90−94.
Liu T T, Wang T, Hu L. *Rhizocotonia solani* recognition algorithm based on convolutional neural network [J]. Chinese Journal of Rice Science, 2019, 33(1): 90−94.

[16] 骆润玫, 王卫星. 基于卷积神经网络的植物病虫害识别研究综述 [J]. 自动化与信息工程, 2021, 42(5): 1−10.
Luo R M, Wang W X. Review on plant disease and pest identification based on convolutional neural network [J]. Automation & Information Engineering, 2021, 42(5): 1−10.

[17] Jiang P, Chen Y, Liu B, et al. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks [J]. IEEE Access, 2019, 7: 59069−59080.

[18] 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. Santiago: IEEE, 2013: 580−587.

[19] Girshick R. Fast R-CNN[C]// Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440−1448.

[20] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137−1149.

[21] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas: IEEE and CVF, 2016: 779−788.

[22] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii: IEEE and CVF, 2017: 7263−7271.

[23] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multiBox detector[C]// European Conference on Computer Vision. Amsterdam: Springer, 2016: 21−37.

[24] Zhang S, Wen L, Lei Z, et al. RefineDet++: single-shot refinement neural network for object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(2): 674−687.

[25] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]// Advances in neural information processing systems, Barcelona: Proceedings of Machine Learning Research (PMLR), 2016: 379−387.

[26] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318−327.

[27] Law H, Deng J. CornerNet: Detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128(3): 642−656.

[28] Zhou X, Wang D, Krähenbühl P. Objects as points[PP/OL]. arXiv preprint arXiv: 1904.07850 [2024−10−01]. https://doi.org/10.48550/arXiv.1904.07850.

[29] Zhou X, Zhuo J, Krähenbühl P. Bottom-up object detection by grouping extreme and center points[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach, California: IEEE and CVF, 2019: 850−859.

[30] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database[C]// 2009 IEEE conference on computer vision and pattern recognition. Miami, Florida: IEEE, 2009: 248−255.

[31] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[C]// Computer Vision – ECCV 2014: 13th European Conference. Zurich: Springer, 2014: 740−755.

[32] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Virtual: IEEE and CVF, 2020: 9729−9738.

[33] Lin T Y, Dollar P. Ms Coco Api[DB/OL]. 2014. https://github.com/pdollar/coco.

（责任编辑　范　娟
责任编委　陈志泊）