

基于改进 YOLOv10n 的无人机航拍目标检测算法

冉同霄¹
RAN Tongxiao

摘要

针对无人机图像中目标尺寸微小、背景环境复杂的检测难题,文章提出一种基于 YOLOv10n 的无人机航拍目标检测算法 SSM-YOLOv10n。为保留更多小目标特征,采用共享特征卷积模块替换基准模型中的 SPPF 模块,减少噪声对小目标特征的干扰;为更好适配无人机小目标检测任务,对颈部网络进行重构,删除贡献度较低的深层检测头,保留高分辨率小目标检测分支;为实现多尺度特征的信息互补,设计多尺度特征检测头,借助参数共享机制提升检测头的特征表达能力。在 VisDrone2019 数据集上的实验结果表明,SSM-YOLOv10n 的精确率、召回率及 mAP 指标相较于 YOLOv10n 分别提升 1.9%、3.2% 和 3.2%,有效改善了无人机航拍目标检测中的漏检与误检问题。

关键词

深度学习; YOLOv10n; 无人机航拍; 小目标检测; 复杂背景

doi: 10.3969/j.issn.1672-9528.2026.05.006

0 引言

无人机因其轻便灵活、拍摄覆盖面广、成本较低等优势,在军事侦察、农业监测、城市规划与管理等领域发挥了重大作用^[1]。但是,无人机图像中的待检测目标像素低,背景复杂,容易受到噪声干扰,导致已有的算法对无人机航拍图像的检测结果不理想。因此,设计一个针对无人机航拍的小目标检测模型,具有重要的应用价值。

目前基于深度学习的目标检测算法按照是否生成候选框,划分为一阶段和两阶段算法两大类。双阶段检测器需要先生成候选框,再进行分类或检测任务,例如 R-CNN 和 Faster R-CNN^[2] 系列。单阶段检测器则直接提取特征进行检测或分类,例如 SSD^[3] 系列和 YOLO^[4-6] 系列。单阶段检测算法检测速度更快,更容易部署到无人机设备上,因而在工业应用中更受欢迎。史涛等^[7] 将大核可分离注意力机制 LSKA 插入 SPPF 模块中,使深层神经网络输出的特征图仍具有充足的小目标特征,提高小目标的特征表示,但是会出现一定的漏检和误检。吴明杰等^[8] 在 YOLO 模型的检测头中引入自注意力机制,当检测头拥有注意力机制后,有利于提供更多的对应目标特征,提高模型检测层的表示能力,解决特征丢失问题。Zhang 等^[9] 利用空洞卷积设计了专注于小目标检测的特征增强模块(feature enhanced module, FEM)减少了复杂背景对小目标的影响。Bai 等^[10] 引入一个细粒度信息提取模块以替代标准卷积操作,更好地保留细节信息,并在双向特

征金字塔网络的基础上,额外增加了一个跳跃连接分支,提升多尺度特征的融合效果。

综合上述研究,针对无人机图像检测任务中小目标特征提取不足的问题进行了多方面改进,但背景复杂,目标重叠的无人机图像检测仍存在检测精度低,召回率不足的问题。为此,本文提出基于 YOLOv10n 的无人机航拍目标检测算法 SSM-YOLOv10n。

1 模型概述

YOLOv10 由清华大学研究者于 2024 年提出,作为 YOLO 系列的新成员,整体结构在延续骨干网络、颈部网络与检测头框架的基础上,通过“无 NMS”训练策略和整体性结构优化,解决了传统检测器依赖后处理的瓶颈,显著提升了精度与速度的平衡,并提供了多种尺度的模型。

YOLOv10 如今版本趋于稳定,且在各个领域中广泛使用,YOLOv10n 非常适合无人机图像目标检测任务对实时性和轻量化要求极高的边缘部署场景。因此本文采取 YOLOv10n 作为基准模型。但是,YOLOv10n 在针对无人机视角的小目标检测任务中,面临着目标噪声干扰严重,深层卷积导致目标特征丢失,检测头特征表示不足等问题,导致检测效果不理想。本文提出无人机航拍检测模型 SSM-YOLOv10n,其结构如图 1 所示。SSM-YOLOv10 模型在主干网络中设计了特征共享卷积(sharefeatureconv, SFConv),替换基准模型中的 SPPF 模块,保证浅层的位置信息与深层的语义信息更好地融合;重构 Neck 网络,替换更有利于小目标检测的高分辨

1. 重庆移通学院 重庆 401420

率检测头；构造多级特征检测头（multi-level feature detection head, MFDH），提供丰富的多尺度特征。

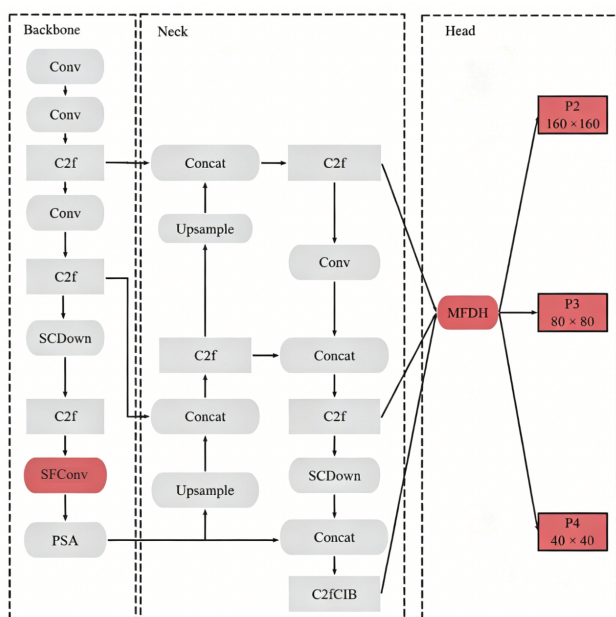


图 1 SSM-YOLOv10

1.1 SFCov

在 YOLOv10 中，采用 SPPF 模型对不同尺度特征进行池化，达到多层次特征融合的作用，但是连续池化操作会显著削弱小目标的微弱特征，导致其细节信息丢失，同时过大的感受野可能引入无关背景干扰，从而降低小目标检测的精度。因此，本文设计了共享特征卷积 SFCov，具体结构如图 2 所示。

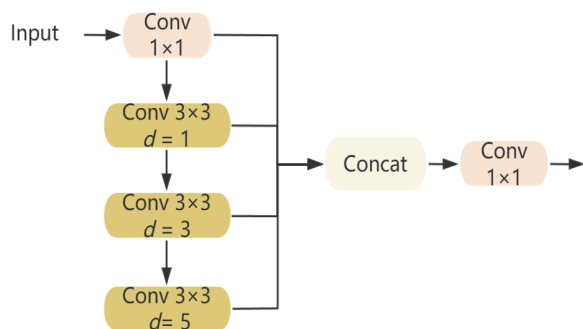


图 2 SFCov 结构图

SFCov 将输入特征先经 1×1 卷积降维，随后同步送入三条并行路径，分别使用膨胀率为 1、3、5 的 3×3 卷积，以同时捕获局部细节与不同范围的上下文信息；三条路径的输出在通道维度拼接，并与通过跳跃连接保留的原始输入细节融合，最后再经 1×1 卷积整合输出。使用 SFCov 替换 SPPF，在维持分辨率的前提下融合了多尺度特征，减少背景噪声对小目标特征的干扰，有效增强了模型对微小目标的辨识与定位能力。

1.2 小目标检测网络

在无人机图像小目标检测任务中，基准模型 YOLOv10n 采取的网络结构是经过深层的主干网络提取特征，使用 P3、P4、P5 三层检测头进行目标检测，对应特征图尺寸为 $80 \text{ px} \times 80 \text{ px}$ 、 $40 \text{ px} \times 40 \text{ px}$ 、 $20 \text{ px} \times 20 \text{ px}$ ，如图 3（a）所示。这种神经网络在小目标检测任务中的主要缺点是深层特征图分辨率过低，导致微小目标的特征信息在传播过程中严重丢失，会出现漏检的问题。常见的解决方案是直接增加一个 160×160 的 P2 浅层检测头，形成四层检测网络。虽然能保留更多浅层细节，但也会显著增加模型参数数量与计算复杂度。考虑到最深的 P5 检测头对小目标检测的作用有限，本文优化主干网络结构，去除了深度卷积层，仅保留 P2、P3、P4 三个高效检测头，结构如图 3（b）所示。这样更浅的神经网络可以保留更多的小目标特征，高分辨率的 P2 层增强了小目标特征提取能力，大幅降低了模型复杂度，实现了精度与效率的更好平衡。

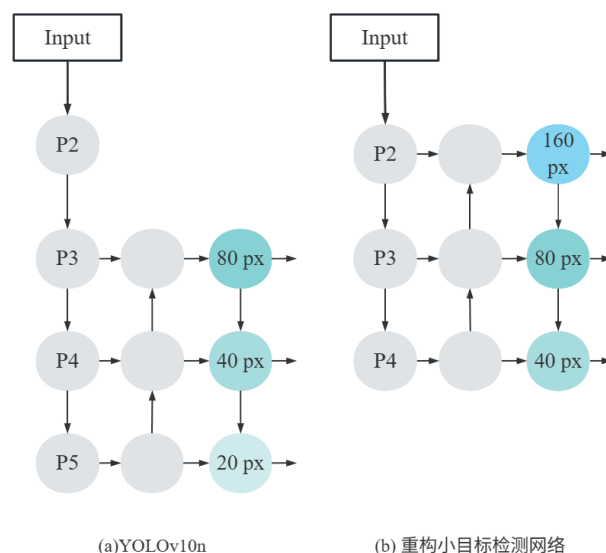


图 3 改进小目标检测网络结构

1.3 共享卷积检测头

在 YOLO 系列的检测头设计中，虽然针对不同尺度的目标设计了不同尺度的检测头，但是检测头本质上学习的特征都是同样的目标特征，导致不同检测头所提取特征之间存在一定的相似性。尽管 YOLOv10 在消除 NMS 冗余和整体架构设计上取得了显著进步，但是各个检测头仍然保持相互独立的设计范式，并分别进行一系列独立的卷积运算来生成最终的预测。独立检测头忽视了多尺度特征之间天然的互补性，每个检测头在封闭的路径中处理信息，无法有效利用其他层级更丰富的语义上下文或更精细的空间细节来辅助本层的判断，造成了信息利用的不足。为此，本文设计了多级特征检测头（MFDH），其结构如图 4 所示。

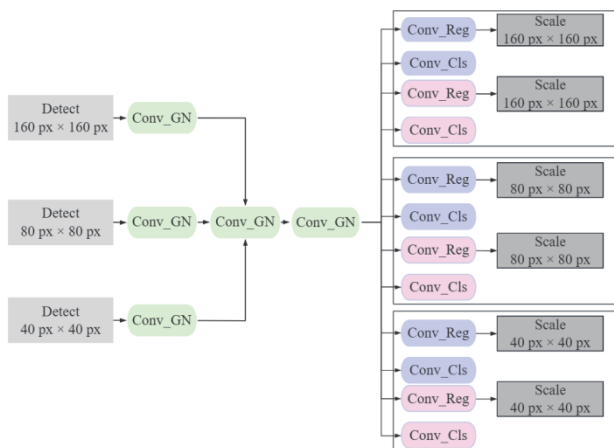


图 4 MFDH 检测头结构

MFDH 引入了多级特征参数共享机制，不同层级的特征图经过一个独立卷积进行初步适配，然后输入共享卷积模块中进行处理。通过共享卷积可以实现不同层级特征间的信息交互与融合，使每个检测头都能拥有丰富的多尺度特征信息。处理后的特征再通过尺度缩放层恢复至各检测头所需的特定尺寸。

考虑到共享卷积可能会削弱对特定尺度特征的差异化提取能力，本文将 MFDH 检测头中的归一化操作由批归一化（batch normalization, BN）替换为组归一化（group normalization, GN）。因为文献 [11] 指出 BN 操作依赖于较大的批次大小，在小批量训练时性能不稳定；而 GN 操作通过对通道进行分组处理来归一化，摆脱了对批次规模的依赖，能够更稳定地提升模型的表征能力。

2 实验及结果分析

2.1 数据集

Visdrone 数据集由天津大学的 AISKEYEYE 团队构建，包含共 10 209 张静态图片，提供的实例标注框总数约为 260 万个。数据集中的图片均是无人机在多样化的实际场景中采集获取，涵盖了不同城市、多种天气条件与光照变化。数据集中标注的物体类别包括行人、汽车、公交车等类别。

2.2 实验环境

实验基于 Windows 11 系统，使用 NVIDIA GeForce RTX 3070 显卡进行训练。训练环境选择的解释器为 Python 3.10 和 PyTorch 1.13.1，CUDA 版本为 11.7。训练过程中的超参数设置如下：输入图像尺寸固定为 640 px×640 px，初始学习率为 0.01，优化器选择 SGD，动量为 0.937，批次大小设置为 4，模型迭代次数为 200。

2.3 评价指标

本文使用精确率（Precision, P ）、召回率（Recall, R ）、全类平均正确率 mAP@0.5（mean average precision, mAP）作为模型的评价指标。

精确率表示模型预测正确的正样本准确度，计算方法为：

$$P = \frac{TP}{TP+FP} \quad (1)$$

召回率表示预测正确的正例数量占总体正例数量的比例，计算方法为：

$$R = \frac{TP}{TP+FN} \quad (2)$$

mAP@0.5 表示 IoU 阈值为 0.5 时，对所有类别的平均精确率（average precision, AP）取平均值，计算方法为：

$$mAP = \frac{1}{N} \sum_i^N \int_0^1 P(R) dR \quad (3)$$

2.4 实验结果与分析

2.4.1 消融实验

为了验证本文各个模块的有效性，在 YOLOv10n 的基础上依次加入各个模块进行消融实验，实验结果如表 1 所示。

表 1 消融实验

Method	$P/\%$	$R/\%$	mAP@0.5/%
YOLOv10n	42.7	32.2	31.7
+SFConv	43.3	34.8	33.4
+小目标网络	44.1	32.1	34.2
+MFDH	43.2	33.1	32.2
SSM-YOLOv10n	44.6	35.4	34.9

由消融实验可得，与基准模型 YOLOv10n 相比较，SFConv 模块能显著提升模型的召回率与 mAP@0.5，分别提升了 1.6% 和 1.7%，证明 SFConv 模块增强了模型对小目标的感知能力。替换小目标检测网络，精确率提升显著，达到 44.1%，验证了高分辨率特征图对小目标定位精度的积极影响。MFDH 检测头则在促进多尺度特征交互的同时，对模型性能起到了良好的平衡作用。最终，融入三个模块得到 SSM-YOLOv10n 模型，取得了最优的综合性能，其精确率、召回率与 mAP@0.5 分别达到 44.6%、35.4% 与 34.9%，充分证明了本文所提出的各个模块在协同作用下，能有效解决无人机图像中小目标特征易丢失、检测精度不足的问题。

3.4.2 对比实验

为了验证 SSM-YOLOv10n 的有效性，将其与其他算法进行对比实验，包含以往主流的 YOLO 目标检测系列进行对比实验，结果如表 2 所示。

表 2 对比实验

Model	$P/\%$	$R/\%$	mAP@0.5/%
YOLOv5n	42.1	32.1	31.6
YOLOv8n	42.4	32.8	31.2
YOLOv10n	42.7	32.2	31.7
YOLOv11n	42.3	32.4	32.6
WEMD-YOLOv11 ^[12]	44.3	33.1	34.4
SSM-YOLOv10n	44.6	35.4	34.9

根据表 2 的对比实验结果可知,SSM-YOLOv10n 模型在无人机目标检测任务中展现出最优的综合性能,精确率为 44.6%,召回率为 35.4%,mAP@0.5 为 34.9% 均是最优表现。与主流 YOLO 系列模型相比召回率的显著提升,验证了改进模型有效增强了对小目标的检测能力,降低漏检的发生率。与同样针对小目标改进的 WEMD-YOLOv11n 相比,虽然精确率相差不大,但是获得了更高召回率和 mAP@0.5,证明了所提改进方案的有效性。

3.4.3 对比结果可视化

为了更加直观地展示本文模型的检测性能,部分检测结果如图 5 所示,左侧是基准模型 YOLOv10n 的检测效果,右侧是 SSM-YOLOv10n 的检测效果。



图 5 检测结果图

通过对比可视化检测结果图可以发现,在日常检测环境下 SSM-YOLOv10n 检测到更多的目标对象,证明在背景复杂,目标堆叠的航拍图像检测任务中可以提高漏检误检率,更好地应用在实际场景中。

3 总结

本文提出 SSM-YOLOv10n 模型用于改善无人机航拍图像检测任务出现漏检、误检的现象。设计 SFConv 模块,通过多层级特征共享,减少背景噪声对特征图的干扰;重构小目标检测网络,保证神经网络可以提取更多有利于小目标检测的特征信息,更加关注小目标特征;构造 MFDH 检测头,使每个检测头都拥有丰富的多尺度特征信息。实验结果表明,SSM-YOLOv10n 模型在 visdrone 数据集上取得较优的检测结果,为无人机航拍目标检测任务提供了一种有效的算法。今后研究工作将专注高精度模型的轻量化,平衡检测精度与检测速度。

参考文献:

- [1] 姜筱巍,吴一全.无人机航拍图像拼接方法研究进展[J].航空学报,2025,46(17):213-243.
- [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] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision (ECCV). Berlin:Springer,2016: 21-37.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.
- [5] WANG C Y, YE H I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]//European conference on computer vision. Berlin: Springer, 2024: 1-21.
- [6] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[PP/OL]. (2024-10-30)[2026-04-15]. <https://doi.org/10.48550/arXiv.2405.14458>.
- [7] 史涛,崔杰,李松.优化改进 YOLOv8 实现实时无人机车辆检测的算法[J].计算机工程与应用,2024,60(9):79-89.
- [8] 吴明杰,云利军,陈载清,等.改进 YOLOv5s 的无人机视角下小目标检测算法[J].计算机工程与应用,2024,60(2):191-199.
- [9] ZHANG Y, YE M, ZHU G Y, et al. FFCA-YOLO for small object detection in remote sensing images[J/OL]. IEEE transactions on geoscience and remote sensing, 2024[2026-04-15]. <https://ieeexplore.ieee.org/abstract/document/10423050>.DOI: 10.1109/TGRS.2024.3363057.
- [10] BAI C X, ZHANG K X, JIN H Z, et al. SFFEF-YOLO: small object detection network based on fine-grained feature extraction and fusion for unmanned aerial images[J]. Image and vision computing, 2025, 156: 105469.
- [11] TIAN Z, SHEN C H, CHEN H, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 44(4): 1922-1933.
- [12] 闫建红,王嘉乐.WEMD-YOLOv11n: 融合频域增强和动态感知的小目标检测算法[J].信息技术与信息化,2026(2):1-5.

【作者简介】

冉同霄(1998—),男,重庆人,硕士研究生,研究方向:目标检测。

(收稿日期:2026-03-16 修回日期:2026-04-17)