

基于旋涡流卷积的轻量化航拍小目标检测算法

田乔梅^{1,2}, 刘 杨³, 刘春红²

(1. 新乡学院 物理与电子工程学院, 河南 新乡 453003; 2. 河南师范大学 计算机与信息工程学院, 河南 新乡 453007;
3. 中国农业科学院农田灌溉研究所, 河南 新乡 453000)

摘要: 航拍影像中的小目标检测因目标尺度小、背景复杂和类内差异显著, 对卷积神经网络的特征提取能力提出了严峻挑战。这种类内差异主要表现为同一类目标在不同成像条件下的外观变异, 导致特征学习更具挑战性。传统目标检测算法在处理航拍小目标时, 往往因感受野不足或背景噪声干扰而导致定位精度下降。基于YOLOv11网络提出了一种新型旋涡流卷积模块VFConv, 通过旋涡核心提取、外围流场抑制和流场动态调谐机制, 有效增强了小目标的特征显著性并抑制背景干扰。VFConv集成到VortexBottleneck模块中, 结合流场密度调谐的残差连接, 进一步提升了特征表达的稳定性和鲁棒性。实验结果表明, YOLO11-VFConv在mAP50指标上达到74.40%, 其相较YOLOv11n、YOLOv8n、YOLOv10n和YOLOv12n的提升幅度分别为7.7%、6.5%、10.6%和6.3%。YOLO11-VFConv在定位精度、整体性能和推理效率上的表现优异, 为航拍小目标检测提供了高效、实用的解决方案。

关键词: 卷积神经网络; 小目标检测; YOLOv11; 旋涡流卷积

DOI: 10.11907/tj.251533

扫描二维码阅读全文:



中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2026)008-0091-08

Lightweight Aerial Small Object Detection Algorithm Based on Vortex Flow Convolution

TIAN Qiaomei^{1,2}, LIU Yang³, LIU Chunhong²

(1. School of Physics and Electronic Engineering, Xinxiang University, Xinxiang 453003, China;

2. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China;

3. Farmland Irrigation Research Institute, Chinese Academy of Agricultural Sciences, Xinxiang 453000, China)

Abstract: The detection of small targets in aerial images poses severe challenges to the feature extraction capability of convolutional neural networks due to the small scale of the targets, complex backgrounds, and significant intra-class variations. Such intra-class variations are mainly manifested as the appearance changes of targets belonging to the same class under different imaging conditions, making feature learning more challenging. Traditional object detection algorithms often suffer from reduced localization accuracy when handling aerial small objects, primarily due to insufficient receptive fields or interference from background noise. This paper proposes a novel Vortex Flow Convolution (VFConv) module based on the YOLOv11 network. VFConv effectively enhances the feature saliency of small objects and suppresses background interference through mechanisms such as vortex core extraction, peripheral flow suppression, and dynamic flow modulation. Integrated into the VortexBottleneck module and combined with flow density-tuned residual connections, VFConv further improves the stability and robustness of feature representation. Experimental results demonstrate that YOLO11-VFConv achieves an mAP50 of 74.40%, outperforming YOLOv11n by 7.7%, YOLOv8n by 6.5%, YOLOv10n by 10.6%, and YOLOv12n by 6.3%. With superior performance in localization accuracy, overall detection capability, and inference efficiency, YOLO11-VFConv provides an efficient and practical solution for aerial small object detection.

Key Words: convolutional neural network; small object detection; YOLOv11; vortex flow convolution

收稿日期: 2025-09-10

基金项目: 网络与交换技术全国重点实验室开放课题(SKLNST-2024-1-11); 河南省科技攻关项目(242102321148)

作者简介: 田乔梅(1984-), 女, 新乡学院物理与电子工程学院助教, 河南师范大学计算机与信息工程学院硕士研究生, 研究方向为深度学习、智慧农业、云计算、边缘计算; 刘杨(1978-), 男, 博士, 中国农业科学院农田灌溉研究所副研究员, 研究方向为智慧农业、节水灌溉技术、滴灌技术; 刘春红(1969-), 女, 博士, 河南师范大学计算机与信息工程学院教授, 研究方向为人工智能、云计算、边缘计算。本文通讯作者: 刘杨。

0 引言

航拍影像的目标检测作为计算机视觉领域一个研究方向,在城市监控、农业监测、灾害评估、无人机导航以及环境遥感等领域发挥着至关重要的作用。这些应用场景要求检测算法具备高精度、实时性和鲁棒性,以适应复杂多变的实际环境。然而,航拍影像中的小目标检测面临诸多严峻挑战:首先,目标尺度通常较小,仅占图像的几十个像素,导致特征信息稀疏且易被忽略;其次,背景环境高度复杂,包含建筑物、植被、道路、水体等多种地物,这些干扰因素会引入大量噪声,从而降低检测的准确性;最后,目标类内差异显著,同一类目标在不同光照条件、拍摄视角或天气影响下可能呈现出迥异的外观特征,这使得模型难以捕捉一致的类内特征表示,进一步增加了检测难度。

传统目标检测算法,如基于区域提议的 Faster R-CNN^[1]和单阶段的 YOLO^[2]系列,在通用场景下表现出色,但应用于航拍小目标时往往暴露不足。这些算法的卷积模块通常采用固定感受野设计,导致对小目标的上下文信息捕获能力有限,同时,背景抑制机制不完善,容易造成定位精度下降和误检漏检的问题。近年来,YOLO 系列模型凭借其高效的端到端框架和实时推理能力,成为目标检测的主流选择。从 YOLOv8 到 YOLOv10^[3]、YOLOv11^[4]以及 YOLOv12^[5],这些模型通过优化网络结构、引入注意力机制和改进损失函数,逐步提升了整体性能,尤其在参数效率和计算速度上取得了显著进步。

尽管如此,现有的 YOLO 模型卷积模块仍主要针对大规模通用目标设计,忽略了航拍小目标的独特需求。具体而言:①小目标检测需要更大的感受野来整合上下文信息,以弥补像素级特征的不足;②背景噪声的自适应抑制至关重要,以突出目标显著性并减少干扰;③特征表达过程应支持动态调谐机制,以平衡目标与背景信息,避免信息混杂或丢失。特别是针对类内差异引起的特征变异,该机制能自适应调整以提升泛化能力。上述局限性导致现有模型在航拍场景下的 mAP 指标和鲁棒性,难以满足实际应用要求。

针对上述挑战,本文提出了一种新型旋涡流卷积模块(VortexFlowConv, VFConv),并将其集成到 YOLOv11 网络中,形成 YOLO11-VFConv 模型。VFConv 模块借鉴流体力学中旋涡流的能量分布特性,通过旋涡核心提取机制增强小目标的高能量聚集,并借助外围流场抑制与动态调谐机制自适应衰减背景噪声,从而显著提升特征表达的精确性和稳定性。该方法不仅保留了 YOLO 系列的轻量化优势,还在航拍小目标检测任务中实现了更高的定位精度和推理效率。

1 相关工作

卷积神经网络(CNN)是目标检测的核心技术,其基本操作是卷积,用于从输入图像中提取空间特征^[6-9]。传统卷积通过局部感受野提取特征,适合捕获图像中的低级模式(如边缘、纹理)。然而,3x3 卷积的感受野难以捕获航拍小目标的上下文信息,无法自适应区分目标和背景,导致复杂场景下特征表达混杂。卷积核越大,参数量和计算量会呈平方增长,从而限制了它在轻量化模型中的应用。为克服传统卷积的局限性,研究者提出了多种卷积变体,如扩张卷积^[10]、分组卷积^[11]、深度可分离卷积^[12]和注意力机制增强的卷积^[13]。其中,扩张卷积通过在卷积核中插入间隙(扩张率 r)扩大感受野,以较低参数量实现大感受野,但稀疏采样可能丢失局部细节;分组卷积将输入通道分为 g 组,每组独立卷积,降低了参数量(约为标准卷积的 $1/g$),但组间信息交互受限;深度可分离卷积将标准卷积分解为深度卷积和点卷积,以降低计算复杂度,但特征表达能力较弱;注意力机制(如 SE、CBAM)通过通道或空间加权增强特征显著性,但其对背景噪声的抑制效果有限,难以适配航拍小目标场景。

有研究者针对 YOLOv11 的卷积模块进行了改进,旨在提升其在航拍图像小目标检测场景下的性能。例如,Zhang 等^[14]提出了一种增强的 YOLOv11 模型,用于河道航拍图像检测,通过引入注意力机制和 C3K2_PPA 模块,优化卷积层,以提升小目标的检测精度。但 C3K2_PPA 模块的复杂性增加了计算负担,且对背景噪声的动态抑制不足,容易在多变河道环境中出现误检的情况。Liu^[15]开发了 MEIS-YOLO,用于高效航拍物体检测,通过轻量设计优化卷积模块,降低了复杂度,但其主要关注整体效率,对小目标上下文捕获的专属机制有限,泛化能力在高噪声航拍场景下受限。Qiu 等^[16]提出 YOLO-Air 一种高效深度学习网络,用于 UAV 航拍小物体检测,通过增强特征模块改进 YOLOv11 的卷积路径。该方法在 UAV 影像中提升了小目标性能,但其特征增强主要依赖于静态卷积变体,缺乏动态调谐,导致在复杂背景下背景干扰抑制不充分。这些改进虽在航拍领域取得了进展,但是缺乏对小目标能量聚集和背景衰减的模拟机制,导致特征显著性增强不充分;同时,计算开销较高或泛化能力弱,无法兼顾轻量化和航拍场景的实时需求。此外,未引入动态调谐以平衡目标-背景信息,易受类内差异影响。这种类内差异主要源于同一类别目标在不同成像条件下的表现多样性,从而导致其特征表示不够稳定。文献[16-17]在航拍小目标检测中对比了 YOLO 变体与非 YOLO 算法(如 Faster R-CNN、SSD 和 RetinaNet),结果显示,YOLO 系列在 mAP 上提升 5-8%,推理速度提高 2 倍以上,这验证了 YOLO 基线模型的优势。本文提出的 VFConv 通过旋涡流模拟机制,针对上述问题

进行了改进。该方法通过能量聚集增强微弱特征的捕获, 利用动态衰减抑制背景信息, 并以调谐方式平衡特征表达, 从而在保持轻量化兼容性的同时实现更有效的特征调控。

2 旋涡流卷积

在航拍小目标检测任务中, 传统卷积操作往往难以有效捕捉小目标的微弱特征, 且容易受到复杂背景干扰。主要源于以下 3 个方面的因素: ①传统卷积的固定感受野 (如 3×3 核) 无法充分整合上下文信息, 导致小目标的稀疏特征被忽略; ②背景抑制机制不完善, 造成噪声混杂; ③特征表达静态, 难以适应类间差异和环境变化。

为解决上述问题, 本文提出旋涡流卷积 (Vortex Flow Convolution, VFConv), 一种受流体力学旋涡流动启发的卷积模块。该模块模拟旋涡流的动态特性, 通过增强目标区域的特征聚合和抑制外围背景噪声, 以提升小目标的显著性和检测精度。具体改进包括: ①针对感受野不足, 使用旋涡核心提取机制扩大并自适应增强核心特征; ②针对噪声干扰, 引入外围流场抑制实现动态衰减; ③针对表达不平衡, 采用流和谐调制机制平衡目标与背景信息。旋涡流卷积的核心思想是将卷积过程分解为旋涡核心提取、外围流场抑制以及动态流和谐调制 3 个阶段, 从而实现对特征流的定向调控。

旋涡流卷积的整体架构如下: 首先, 输入特征图 $x \in \mathbb{R}^{B \times C_{in} \times H \times W}$ (其中, B 为批次大小, C_{in} 为输入通道数, H 和 W 分别为高度和宽度) 经过多个子模块处理; 其次, 输出融合后的特征图 $out \in \mathbb{R}^{B \times C_{out} \times H' \times W'}$ (其中, C_{out} 为输出通道数, H' 和 W' 取决于步幅)。

2.1 旋涡核心提取器

旋涡核心提取器 (Vortex Core Extractor, VCE) 旨在通过扩张卷积增强小目标的感受野, 模拟旋涡中心的高密度流动聚集效应。该模块使用核大小为 3、扩张率为 2 的卷积操作, 提取核心特征, 同时, 引入旋涡流生成器来强化旋涡能量。

首先, 旋涡核心特征通过卷积提取, 如公式 (1) 所示:

$$VCE_{feat} = \text{Conv}_{3 \times 3, d=2}(x) \quad (1)$$

其中, $VCE_{feat} \in \mathbb{R}^{B \times C_{out} \times H \times W}$ 表示旋涡核心特征, $\text{Conv}_{3 \times 3, d=2}$ 为核大小 3×3 、扩张率 (dilation) 为 2 的卷积操作, 输入为 x , 无偏置。

其次, 旋涡流生成器生成旋涡流向量场, 如公式 (2) 所示:

$$\text{Vortex}_{flow} = \tanh(\text{BN}(\text{Conv}_{3 \times 3}(x))) \quad (2)$$

其中, $\text{Vortex}_{flow} \in \mathbb{R}^{B \times 4 \times H \times W}$ 表示旋涡流向量场 (4 个通道代表丰富旋涡信息), $\text{Conv}_{3 \times 3}$ 为核大小 3×3 的卷积, BN 为批标准化 (Batch Normalization), $\tanh$ 为双曲正切激活函数, 用于归一化流场。

然后, 旋涡能量通过范数计算, 如公式 (3) 所示:

$$\text{Vortex}_{energy} = \|\text{Vortex}_{flow}\|_2 \quad (3)$$

其中, $\text{Vortex}_{energy} \in \mathbb{R}^{B \times 1 \times H \times W}$ 表示旋涡能量图, $\|\cdot\|_2$ 为通道维上的 L2 范数。

最终, 增强后的旋涡核心特征, 如公式 (4) 所示:

$$VCE_{feat}^{enhanced} = VCE_{feat} \cdot (1 + \sigma(\text{Vortex}_{energy}) \cdot 0.7) \quad (4)$$

其中, $VCE_{feat}^{enhanced}$ 表示增强旋涡核心特征, σ 为 Sigmoid 函数, 0.7 为旋涡强度增强系数, 用于放大目标区域的显著性。该步骤通过旋涡能量自适应地强化核心特征, 避免小目标特征被稀释。

2.2 外围流场抑制器

外围流场抑制器 (Peripheral Flow Suppressor, PFS) 通过大核卷积捕捉外围背景, 并利用散度估计器抑制噪声流动, 模拟旋涡外围的低密度扩散效应。该模块有助于减少背景干扰, 提升小目标的信噪比。

外围特征通过卷积提取, 如公式 (5) 所示:

$$\text{PFS}_{feat} = \text{Conv}_{5 \times 5}(x) \quad (5)$$

其中, $\text{PFS}_{feat} \in \mathbb{R}^{B \times C_{out} \times H \times W}$ 表示外围流场特征, $\text{Conv}_{5 \times 5}$ 为核大小 5×5 的卷积, 步幅与 VCE 相同。

散度估计器计算流场散度, 如公式 (6) 所示:

$$\text{Divergence} = \sigma(\text{BN}(\text{Conv}_{3 \times 3}(x))) \quad (6)$$

其中, $\text{Divergence} \in \mathbb{R}^{B \times 1 \times H \times W}$ 表示散度图, 直接作为抑制权重 σ 为 Sigmoid 函数, 用于归一化输出 $[0, 1]$ 范围的权重。

抑制后的外围特征, 如公式 (7) 所示:

$$\text{PFS}_{feat}^{suppressed} = \text{PFS}_{feat} \cdot (1 - \text{Divergence} \cdot 0.9) \quad (7)$$

其中, $\text{PFS}_{feat}^{suppressed}$ 表示抑制外围特征, 0.9 为抑制强度系数。该公式通过散度权重自适应地减弱背景区域的特征响应, 确保旋涡外围流动不干扰核心目标。

2.3 流和谐调制器

流和谐调制器 (Flow Harmony Modulator, FHM) 动态调谐 VCE 和 PFS 的输出, 平衡目标增强与背景抑制的矛盾, 模拟旋涡流的整体和谐流动。该模块采用通道注意力机制, 实现特征的精细调控。

首先, 组合特征, 如公式 (8) 所示:

$$\text{Combined}_{feat} = VCE_{feat}^{enhanced} + \text{PFS}_{feat}^{suppressed} \quad (8)$$

其中, $\text{Combined}_{feat} \in \mathbb{R}^{B \times C_{out} \times H \times W}$ 表示初步组合特征。

其次, 和谐映射通过两层 1×1 卷积生成, 如公式 (9) 所示:

$$\text{Harmony}_{map} = \sigma\left(\text{Conv}_{1 \times 1}\left(\text{ReLU}\left(\text{BN}\left(\text{Conv}_{1 \times 1}(\text{Combined}_{feat})\right)\right)\right)\right) \quad (9)$$

其中, $\text{Harmony}_{map} \in \mathbb{R}^{B \times C_{out} \times H \times W}$ 表示和谐映射, 第一层卷积将通道减半 ($C_{out}/2$), 第二层卷积恢复通道, ReLU 为整流线性单元激活函数。

最后, 平衡特征, 如公式 (10) 所示:

$$\text{Balanced}_{\text{feat}} = \text{VCE}_{\text{feat}}^{\text{enhanced}} \cdot (1 + \text{Harmony}_{\text{map}} \cdot 1.2) - \text{PFS}_{\text{feat}}^{\text{suppressed}} \cdot \text{Harmony}_{\text{map}} \quad (10)$$

其中, $\text{Balanced}_{\text{feat}}$ 表示平衡特征, 1.2 为目标显著性增强系数。该公式通过和谐映射放大核心特征的同时, 进一步抑制外围干扰, 实现流场的动态和谐。

2.4 最终融合与输出

最终融合使用卷积整合平衡特征, 以提升整体感受野, 如公式(11)所示:

$$\text{Out}_{\text{pre}} = \text{Conv}_{3 \times 3}(\text{Balanced}_{\text{feat}}) \quad (11)$$

其中, Out_{pre} 表示预融合输出, $\text{Conv}_{3 \times 3}$ 为核大小 3×3 的卷积。

经过标准化和激活后得到最终输出, 如公式(12)所示:

$$\text{Out} = \text{ReLU}(\text{BN}(\text{Out}_{\text{pre}})) \quad (12)$$

其中, Out 表示旋涡流卷积的最终输出。该步骤确保了特征的稳定性和非线性表达。

2.5 VortexBottleneck 模块

为将旋涡流卷积集成到 YOLOv11 骨干网络中, 本文设计了 VortexBottleneck 模块, 一种轻量级残差结构。该模块通过嵌入 VFConv, 并引入基于流场密度的残差调谐, 以提升网络的稳定性和适应性。这一集成有效解决了传统 YOLO 卷积在航拍小目标上的局限性。通过 VFConv 的动态机制, 弥补了感受野和抑制不足。同时, 残差调谐根据流场密度自适应调控贡献, 避免特征丢失, 从而确保了在复杂背景下的鲁棒性提升。VortexBottleneck 模块结构主要包括: ① 1×1 卷积降维: $\text{Out}_1 = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(x)))$; ② VFConv 核心: $\text{Out}_2 = \text{VFConv}(\text{Out}_1)$; ③ Dropout (步幅为 1 时应用, 率为 0.15): $\text{Out}_2 = \text{Dropout}(\text{Out}_2)$; ④ 1×1 卷积恢复: $\text{Out}_3 = \text{BN}(\text{Conv}_{1 \times 1}(\text{Out}_2))$ 。

流场密度调谐权重如公式(13)所示:

$$\text{Density}_{\text{weight}} = \sigma \left(\text{Conv}_{1 \times 1} \left(\text{ReLU} \left(\text{BN} \left(\text{Conv}_{1 \times 1}(\text{Out}_3) \right) \right) \right) \right) \quad (13)$$

其中, $\text{Density}_{\text{weight}} \in \mathbb{R}^{H \times 1 \times H \times W}$ 表示密度权重, 第一层卷积通道为 $C/4$ (C 为当前特征的通道数), 第二层输出为 1 通道。

最终残差输出如公式(14)所示:

$$\text{Out}_{\text{final}} = \text{ReLU}(\text{Out}_3 + \text{Identity} \cdot \text{Density}_{\text{weight}} \cdot \sigma) \quad (14)$$

其中, Identity 为捷径连接(若需下采样可进行调整), δ 为残差权重微调系数。该机制根据输出特征的流场密度自适应地调控残差贡献, 以确保小目标检测的鲁棒性。

旋涡流卷积通过上述组件的协同作用, 在理论上模拟了旋涡流动的聚合—扩散—平衡动态, 用以提升航拍场景下小目标的检测性能。

VortexBottleneck 模块将 VFConv 集成到 YOLOv11 的骨干网络中, 构建了 YOLO11-VFConv 网络结构, 如图 1 所

示。通过 1×1 卷积降维、VFConv 特征提取和 1×1 卷积升维, 实现了高效的特征处理。VFConv 通过替换部分原卷积层, 增强了对小目标的上下文捕获能力和背景抑制能力。同时, 引入的残差连接确保了梯度流动的稳定性, 从而有效避免了 YOLOv11 在航拍场景下的感受野不足问题。

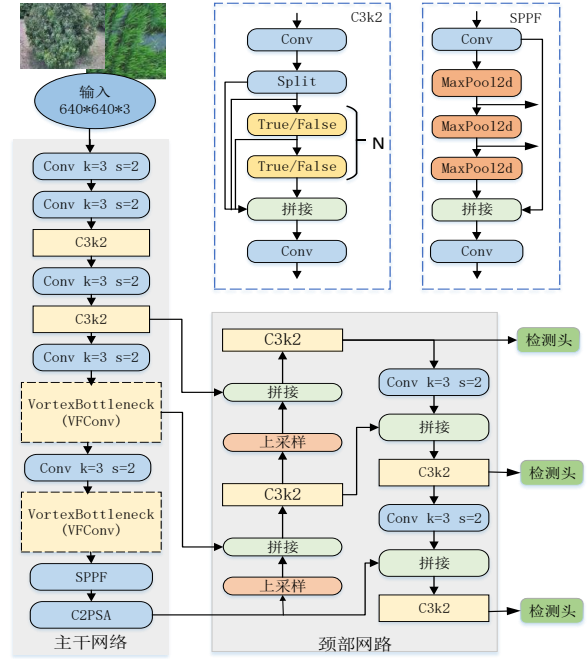


Fig. 1 Network structure of YOLO11-VFConv

图1 YOLO11-VFConv 网络结构

3 实验结果与分析

3.1 数据集

实验使用的数据集(三叶草数据集: <https://github.com/liu-zey/public-dataset>; 芒果数据集: https://universe.roboflow.com/levi-udoq1/mango_fruit-wkohw/dataset/1) 涵盖芒果、三叶草和三叶草花 3 个目标类别, 图像总计 8 888 张, 其中芒果图像 8 675 张, 三叶草及三叶草花图像 213 张。数据集被划分为包含 877 张图像的验证集和 112 张图像的测试集。数据集具有复杂背景, 适于评估航拍小目标的性能检测^[18-21]。尽管类别数量较少且样本分布不均衡(芒果类主导), 这可能影响模型的整体泛化性, 但该数据集针对航拍小目标的特定挑战(如稀有小物体在复杂背景中)进行了定制, 选择少数类别以聚焦方法的核心改进验证, 为避免泛化过度导致注意力分散。基于当前实验设置, 后续结果将验证所提方法的有效性, 尤其是其在提升少数类别性能上所具备的显著优势。

3.2 实验环境与评估指标

实验环境如下: Debian 12、vCPU 80、Mem 64GB、NVIDIA RTX 3060、PyTorch 2.6.0、CUDA 12.8、Python 3.12.9。训练参数如下: 训练 300 个 epoch, 批量大小 16, 图像尺寸 640×640 , 数据加载使用 32 个工作线程, 启用自动混合精

度(AMP)训练。评估指标如下:重点评估 mAP50 以突出定位精度,辅助评估 mAP50-95 以反映整体性能,同时记录训练时间、推理速度和类间检测精度。

3.3 性能对比

3.3.1 模型性能对比

尽管未与非 YOLO 算法进行直接对比,但近期文献[16-17]表明,YOLO 变体在小目标检测的 mAP 指标上已优于 Faster R-CNN(提升约 6%~8%)和 SSD,同时具备更快的速度。这表明,本文 YOLO11-VFConv 的轻量化设计在类似场景下具备潜在优越性。YOLO11-VFConv 与 4 种基线模型 YOLOv8n、YOLOv10n、YOLOv11n 和 YOLOv12n 的性能对比如表 1 所示。

在 mAP50 指标上,YOLO11-VFConv 达到 74.40%,其相较 YOLOv11n、YOLOv8n、YOLOv10n 和 YOLOv12n 的提升幅度分别为 7.7%、6.5%、10.6% 和 6.3%。这表明,YOLO11-VFConv 在目标定位能力上显著优于基线模型,尤其是在航拍小目标检测中,成功检测目标的能力更强。YOLO11-VFConv 在 mAP50 上的显著提升表明,其在航拍小目标定位精度上的优势,尤其在稀有类别(如三叶草和三叶草花)中表现突出。

在 mAP50-95 指标上,YOLO11-VFConv 为 34.58%,优于 YOLOv11n(33.70%)、YOLOv10n(33.26%)和 YOLOv12n(32.70%),但略低于 YOLOv8n(35.90%)。这可能是因为小目标的 IoU 在高阈值(如 0.95)时难以达到较高值。此外,数据集的类别少和样本不均衡,可能加剧了对少数类别的泛化挑战,导致整体 mAP50-95 受限,但其 mAP50 的显著提升证明了 VFConv 在定位精度上的优势。

Table 1 Comparison of model performance
表 1 模型性能对比

模型	mAP50/ %	mAP50- 95/%	训练时 间/h	验证 box_loss	验证 cls_loss	验证 dfl_loss	参数 量/M	FLO Ps/G
YOLOv8n	69.87	35.90	6.099	2.206 9	1.082 5	1.045 0	3.01	8.2
YOLOv10n	67.25	33.26	8.351	4.369 8	2.242 6	2.008 4	2.71	8.4
YOLOv11n	69.10	33.70	6.367	2.110 4	1.060 6	1.032 4	2.59	6.4
YOLOv12n	70.00	32.70	8.725	2.239 6	1.116 9	1.053 0	2.57	6.5
YOLO11- VFConv	74.40	34.58	7.448	2.157 3	1.055 5	1.029 2	3.89	8.1

在训练时间方面,YOLO11-VFConv 耗时 7.448 h,略高于 YOLOv11n(6.367 h),但低于 YOLOv12n(8.725 h)和 YOLOv10n(8.351 h)。

在验证损失方面,YOLO11-VFConv 的 box_loss、cls_loss 和 dfl_loss 均低于 YOLOv11n,表明其收敛性更优。

3.3.2 类间性能对比

鉴于数据集的样本不均衡(芒果类远多于其他类别),模型在少数类别(如三叶草和三叶草花)上的性能可能存在泛化风险。但通过类间对比可见,YOLO11-VFConv 在这些稀有类上仍取得了显著提升,证明了 VFConv 对小目

标显著性的增强作用。类间性能对比如表 2 所示。

Table 2 Performance comparison among classes
表 2 类间性能对比

模型名称	类名	精确率/P	召回率/R	mAP50	mAP50- 95
YOLOv8n	芒果	0.699	0.512	0.590	0.242
	三叶草	0.928	0.926	0.979	0.648
	三叶草花	0.513	0.622	0.528	0.189
YOLOv10n	芒果	0.840	0.418	0.614	0.269
	三叶草	0.830	0.821	0.906	0.577
	三叶草花	0.539	0.455	0.492	0.148
YOLOv11n	芒果	0.775	0.474	0.607	0.253
	三叶草	0.878	0.786	0.936	0.588
	三叶草花	0.593	0.545	0.531	0.169
YOLOv12n	芒果	0.755	0.470	0.588	0.238
	三叶草	0.999	0.857	0.933	0.522
	三叶草花	0.511	0.524	0.577	0.216
YOLO11-VFConv	芒果	0.796	0.489	0.623	0.263
	三叶草	0.928	0.929	0.985	0.593
	三叶草花	0.669	0.591	0.625	0.187

芒果类别方面,YOLO11-VFConv 的 mAP50 为 0.623,优于 YOLOv11n(0.607)、YOLOv12n(0.588)和 YOLOv8n(0.590),表明其在主要类别上的定位精度更优。YOLOv10n 的精确率虽高(0.840),但召回率较低(0.418),导致 mAP50 仅 0.614。

三叶草类别方面,YOLO11-VFConv 的 mAP50 达到 0.985,优于 YOLOv12n(0.933)、YOLOv10n(0.906)和 YOLOv11n(0.936),与 YOLOv8n(0.979)相比也略有提升。其高召回率(0.929)表明模型对稀有小目标的检测能力极强。

三叶草花类别方面,YOLO11-VFConv 的 mAP50 为 0.625,显著高于 YOLOv10n(0.492)、YOLOv11n(0.531)、YOLOv8n(0.528)和 YOLOv12n(0.577),在稀有且尺寸极小的目标上展现出优异的定位能力。

此外,在三叶草类别中,YOLO11-VFConv 检测到 28 个实例中的大部分(召回率 0.929,mAP50 0.985),而 YOLOv12n 漏检部分目标(召回率 0.857,mAP50 0.933)。在三叶草花类别中,YOLO11-VFConv 的 mAP50(0.625)显著高于 YOLOv10n(0.492),展现出对极小目标的优异定位能力。

3.3.3 推理速度

推理速度(ms/每张图像)对比如表 3 所示。

YOLO11-VFConv 的总推理时间为 3.5 ms。其相较 YOLOv12n(4.2 ms)、YOLOv10n(3.9 ms)和 YOLOv11n(3.7

Table 3 Comparison of inference speed
表 3 推理速度对比

模型名称	预处理	推理	后处理	总时间
YOLOv8n	0.2	1.5	1.9	3.6
YOLOv10n	0.2	3.2	0.5	3.9
YOLOv11n	0.2	1.6	1.9	3.7
YOLOv12n	0.2	2.4	1.6	4.2
YOLO11-VFConv	0.2	2.4	0.9	3.5



Fig. 4 Detection effect of YOLO11-VFConv

图 4 YOLO11-VFConv 的检测效果

和推理效率上的优异表现,为无人机在农业监测、生态保护等领域的航拍小目标检测提供了高效解决方案。

尽管 YOLO11-VFConv 取得了显著改进,但仍存在以下缺陷:①参数量略有增加,可能在极低功耗设备上受限;②对极小目标(<5 像素)的鲁棒性依赖数据集多样性,在光照剧变或遮挡严重场景下可能出现漏检;③当前实验主要基于自定义数据集,其类别较少且样本不均衡,可能导致泛化性不足。例如对未见分布的适应性较弱,泛化能力需进一步验证。此外,虽未直接与非 YOLO 算法对比,但通过文献[16-17]的结果,可间接推断本文方法在精度和速度上的优势。

未来工作将从以下几个方面开展:①扩展到更多公开航拍数据集(如 VisDrone 或 UAVDT),通过增加类别数量和平衡样本分布以验证泛化性,并与非 YOLO 算法(如 Faster R-CNN)直接对比,以全面验证本文算法的优势;②进一步优化 VFConv 的计算复杂度,通过结构重参数化实现更轻量化;③探索多模态融合(如结合红外影像),以提升在复杂环境下的检测鲁棒性。

参考文献:

[1] 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 and Machine Intelligence, 2017, 39(6): 1137-1149.

[2] 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, 2016: 779-788.

[3] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[DB/OL]. <https://arxiv.org/abs/2405.14458>.

[4] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[DB/OL]. <https://arxiv.org/abs/2410.17725>.

[5] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors[DB/OL]. <https://arxiv.org/abs/2502.12524>.

[6] QIAO S C, DANG S S, HE H Z, et al. Research review on crop disease detection based on convolutional neural network [J]. Journal of Shanxi Agricultural University (Natural Science Edition), 2025, 45 (2): 113-127. 乔世成, 党珊珊, 何海祝, 等. 基于卷积神经网络的农作物病害检测研究综述 [J]. 山西农业大学学报(自然科学版), 2025, 45 (2): 113-127.

[7] ZHOU S L, LI R. Research review on crop pest and disease recognition based on convolutional neural network [J]. Smart Agriculture Guide, 2024, 4 (17): 39-45. 周善良, 李锐. 基于卷积神经网络的农作物病虫害识别研究综述 [J]. 智慧农业导刊, 2024, 4 (17): 39-45.

[8] LI W J, BAI J, PENG B, et al. Graph convolutional neural network and its application review in the field of image recognition [J]. Computer Engineering and Applications, 2023, 59 (22): 15-35. 李文静, 白静, 彭斌, 等. 图卷积神经网络及其在图像识别领域的应用综述 [J]. 计算机工程与应用, 2023, 59 (22): 15-35.

[9] DU Z W, ZHOU H, LI C Y, et al. Survey of small object detection algo-

- rithms for deep convolutional neural networks [J]. Computer Science, 2022, 49 (12): 205–218.
- 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.
- [10] LI K Y, ZHU X Y, MA J C, et al. Cucumber disease severity estimation method based on hybrid dilated convolution and attention [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54 (2): 231–239.
- 李凯雨,朱昕怡,马浚诚,等. 基于混合扩张卷积和注意力的黄瓜病害严重程度估算方法 [J]. 农业机械学报, 2023, 54 (2): 231–239.
- [11] LU D, ZHAO Q. Lightweight object detection algorithm with spatial grouped inner convolution [J/OL]. Control and Decision, 2025–05–12. <https://doi.org/10.13195/j.kzyjc.2025.0035>.
- 卢迪,赵庆. 空间分组内卷积轻量级目标检测算法 [J/OL]. 控制与决策, 2025–05–12. <https://doi.org/10.13195/j.kzyjc.2025.0035>.
- [12] CHEN H, YAN J G, YANG H, et al. Depthwise separable convolutional neural network based on structural reparameterization [J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2024–07–03. <https://doi.org/10.13700/j.bh.1001-5965.2024.0287>.
- 陈红,闫建国,杨华,等. 基于结构重参数化的深度可分离卷积神经网络 [J/OL]. 北京航空航天大学学报, 2024–07–03. <https://doi.org/10.13700/j.bh.1001-5965.2024.0287>.
- [13] LI Y J, ZHOU D M, LI M, et al. Low-light image enhancement combining smooth extended convolution network and attention mechanism [J]. Journal of Yunnan University (Natural Science Edition), 2022, 44 (5): 912–924.
- 李宇娇,周冬明,李森,等. 结合平滑扩展卷积网络与注意力机制的低照度图像增强 [J]. 云南大学学报 (自然科学版), 2022, 44 (5): 912–924.
- [14] ZHANG L, ZHENG A, SUN X, et al. Enhanced YOLOv11-based river aerial image detection research [J]. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 1–5.
- [15] LIU Y, WU J, CHEN L. MEIS-YOLO: improving YOLOv11 for efficient aerial object detection with lightweight design [J]. Intelligent and Converged Networks, 2025, 6(2): 151–163.
- [16] QIU J, CAI F, FU N, et al. YOLO-Air: an efficient deep learning network for small object detection in drone-based imagery [J]. IEEE Access, 2025, 13: 79718–79735.
- [17] YANG Y, FENG F, LIU G, et al. MEL-YOLO: a novel YOLO network with multi-scale, effective, and lightweight methods for small object detection in aerial images [J]. IEEE Access, 2024, 12: 194280–194295.
- [18] MA B J, QIU Y Y, CHEN B B, et al. Apple recognition method based on improved YOLO v8n lightweight model [J]. Journal of Shihezi University (Natural Science Edition), 2025, 43 (2): 143–151.
- 马保建,邱媛媛,陈棒棒,等. 基于改进 YOLO v8n 轻量化模型的苹果识别方法 [J]. 石河子大学学报 (自然科学版), 2025, 43 (2): 143–151.
- [19] HAN K L, WANG Z K, YU Y F, et al. Obstacle detection method in complex cotton field environment based on improved YOLO 11n model [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56 (5): 111–120.
- 韩科立,王振坤,余永峰,等. 基于改进 YOLO 11n 模型的棉花田间复杂环境障碍物检测方法 [J]. 农业机械学报, 2025, 56 (5): 111–120.
- [20] WANG Y, JIANG X G, QIN H Y, et al. WS-YOLO: small target detection method from aerial photography perspective [J]. Modern Electronics Technique, 2025, 48 (5): 68–74.
- 王尧,蒋行国,秦海洋,等. WS-YOLO: 航拍视角小目标检测方法 [J]. 现代电子技术, 2025, 48 (5): 68–74.
- [21] SUN J Y, XU M J, ZHANG J P, et al. Optimized and improved YOLOv8 target detection algorithm from UAV perspective [J]. Computer Engineering and Applications, 2025, 61 (1): 109–120.
- 孙佳宇,徐民俊,张俊鹏,等. 优化改进 YOLOv8 无人机视角下目标检测算法 [J]. 计算机工程与应用, 2025, 61 (1): 109–120.

(责任编辑:陈维捷)