

面向无人机边缘端的轻量化航拍小目标检测方法

徐鹤^{1, 2}, 任治阳¹, 季一木^{1, 2}, 李鹏^{1, 2}

(1. 南京邮电大学计算机学院, 江苏 南京 210023; 2. 南京邮电大学江苏省高性能计算与智能处理工程研究中心, 江苏 南京 210023)

摘要: 针对无人机航拍图像中小目标占比高、密集遮挡严重、浅层细节易随下采样衰减, 以及机载边缘设备资源受限等问题, 提出一种面向无人机边缘端的轻量化航拍小目标检测方法。该方法以 YOLO11s 为基线, 将原始 P3/P4/P5 检测结构重构为 P2/P3/P4 结构, 使高分辨率浅层特征直接参与微小目标预测; 同时移除对无人机小目标场景收益有限的 P5 分支, 并将高层语义建模模块前移至 P4 层, 以降低冗余参数开销。进一步设计选择性引导模块、P2 引导的跨层再注入机制和自适应融合模块, 以增强浅层细节表达并缓解多源特征融合冲突。实验结果表明, 在 VisDrone 验证集上, 该方法较 YOLO11s 基线模型的平均精度均值 (mAP, mean average precision) @0.5 和 mAP@0.5: 0.95 分别提升 4.40 和 3.22 个百分点, 参数量由 9.46M 降至 3.82M; 在 Jetson Orin NX 16 GB 上采用 TensorRT 32 位浮点精度部署时, 推理速度达到 52.63 帧/s。结果表明, 该方法在提升无人机小目标检测精度的同时, 兼顾了模型规模和边缘端实时推理能力。

关键词: 边缘智能; 无人机小目标检测; YOLO11s; 特征增强

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-3750.

A lightweight method for small object detection in aerial imagery on UAV edge devices

XU He^{1,2}, REN Zhiyang¹, JI Yimu^{1,2}, LI Peng^{1,2}

1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2. Jiangsu HPC and Intelligent Processing Engineer Research Center, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: To address the problems of high small-object proportion, severe dense occlusion, shallow-detail degradation during downsampling, and limited onboard edge resources in aerial images captured by unmanned aerial vehicles (UAVs), a lightweight small-object detection method for UAV edge devices was proposed. YOLO11s was adopted as the baseline. The original P3/P4/P5 detection structure was reconstructed into a P2/P3/P4 structure, so that high-resolution shallow features could directly participate in tiny-object prediction. Meanwhile, the P5 branch with limited benefit for UAV small-object scenes was removed, and the high-level semantic modeling module was moved to the P4 layer to reduce redundant parameter overhead. A selective guidance block, a P2-guided cross-layer re-injection mechanism, and an adaptive fusion module were further designed to enhance shallow details and alleviate multi-source feature fusion conflicts. Experimental results on the VisDrone validation set showed that mAP (mean average precision)@0.5 and mAP@0.5:0.95 were improved by 4.40 and 3.22 percentage points, respectively, compared with the YOLO11s baseline, while the number of parameters was reduced from 9.46M to 3.82M. When deployed with TensorRT using 32-bit floating-point precision (FP32)

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 季一木, jiyim@njupt.edu.cn

基金项目: 江苏省研究生科研实践创新计划项目 (SJCX26)

Foundation Items: Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX26)

on a Jetson Orin NX 16 GB platform, an inference speed of 52.63 frames per second (FPS) was achieved. The results indicate that the proposed method improves detection accuracy while maintaining a compact model size and real-time edge inference capability.

Key words: edge intelligence, uav small object detection, yolo11s, feature enhancement

0 引言

近年来,随着无人机平台、机载传感器、边缘计算硬件以及低空物联网相关技术的发展,无人机已广泛应用于城市安防、交通巡检、应急救援、灾害监测、电力巡检和环境感知等低空应用场景^[1]。低空物联网强调低空空域中通信、感知、计算与智能决策的协同融合,而无人机作为典型的低空移动感知节点和边缘执行终端,能够承担场景数据采集、目标识别、状态感知与任务反馈等功能。相较于地面视觉系统,无人机具有部署灵活、观测范围大、俯视视角广和获取数据效率高等优势,能够在复杂开放环境中快速获取大范围场景信息。因此,面向无人机边缘设备的实时目标检测不仅是计算机视觉与智能无人系统研究中的重要方向,也直接关系到低空物联网中态势感知、任务决策与协同交互能力的实现^[2]。

然而,与自然场景图像相比,无人机航拍图像在成像视角、目标尺度分布和背景构成方面具有更强的特殊性^[3]。一方面,受拍摄高度、视场范围和俯视角度的影响,地面目标在图像中的面积占比通常较低,大量目标仅对应少量像素,表现出典型的小目标甚至微小目标特征^[4];另一方面,无人机场景中目标往往呈现密集分布、局部遮挡、尺度差异以及背景干扰复杂等现象,容易造成目标纹理信息弱化、轮廓边界模糊和类别间可分性下降^[5]。与此同时,航拍图像还常伴随光照变化、运动模糊和成像视角变化等问题,进一步增加了检测任务的难度^[6]。对于此类场景,常规目标检测器在深层下采样过程中容易造成浅层细节信息衰减,使小目标的定位与分类性能明显下降^[7]。

从目标检测方法的发展来看,现有研究大体可分为两阶段方法与单阶段方法两类^[8]。相关综述表明,两类方法在精度、速度与部署代价之间存在显著权衡^[9]。两阶段方法通过候选区域生成和精细分类回归获得较高检测精度,但其计算流程较长、推理开销较大,在实时无人机任务中部署受限。相比

之下,以 YOLO 系列为代表的单阶段方法因结构紧凑、推理速度快、工程实现成熟,已成为无人机实时目标检测的重要技术路线^[10]。近年来,围绕无人机小目标检测问题,研究者从多尺度特征融合、注意力建模、上下文增强、检测头重构和轻量化设计等方面开展了许多研究。张杨等基于改进 Faster R-CNN^[11]提升了无人机小目标的候选区域表达能力,在一定程度上改善复杂背景下的检测精度,但两阶段检测流程推理开销较大,难以满足边缘端实时部署需求。UAV-YOLO^[12]针对无人机视角下目标尺度小、分布密集的问题对 YOLO 结构进行改进,提高了小目标检测性能,但对浅层高分辨率细节的持续利用仍然有限。SOD-YOLO^[13]面向 UAV 图像改进 YOLOv8,通过增强特征表达提升了小目标检测效果,但检测尺度配置仍主要沿用通用场景设计,对无人机场景中目标占主导的尺度分布适配不够充分。MSUD-YOLO^[14]采用多尺度建模加强了 UAV 航拍图像中的小目标感知能力,但复杂特征融合结构带来了额外计算负担。LCM-YOLO^[15]进一步从轻量化和上下文建模角度改善检测性能,但在精度、参数规模和实时性之间仍需权衡。因此,如何在避免显著增加边缘端部署负担的前提下,提高浅层细节特征利用率并优化检测尺度配置,仍是无人机小目标检测中的关键问题。

尽管已有方法在无人机场景目标检测中取得了较好进展,但针对边缘端小目标检测仍存在若干突出问题。首先,小目标可用特征有限,经过多次卷积与下采样后,其纹理、边缘和位置信息极易被削弱,造成检测器对小目标的响应不稳定^[16]。其次,现有多尺度特征融合方法虽然能够在一定程度上提升检测性能,但部分方法对不同尺度特征采用近似一致的处理方式,未能充分考虑无人机场景中“小目标占主导”的尺度分布特征,导致高层语义信息与浅层细节信息之间的协同不足^[17]。再次,部分改进方法通过叠加注意力模块或复杂特征增强结构获得精度提升,但往往伴随着参数量、计算量和延迟的上升,对于在机载嵌入式平台上部署不友好^[18]。

对于实际无人机边缘视觉任务而言，检测模型不仅需要具备较高精度，还应兼顾实时性、存储开销和功耗约束，因此“精度—复杂度—速度”之间的平衡仍是当前研究中的核心问题^[19]。

结合 VisDrone 等典型无人机数据集可以发现，图像中的目标多集中于小尺度区间，大目标样本占比较低，这意味着通用检测器中面向大目标设计的深层检测分支未必能够在该类任务中持续带来有效收益^[20]。相反，浅层高分辨率特征对小目标边缘结构、空间位置信息和局部纹理细节的保持更为关键^[21]。如果仍沿用面向通用场景设计的检测头尺度配置，容易出现计算资源分配与数据分布特征不匹配的问题，进而限制模型在小目标场景中的性能释放。因此，针对无人机边缘端小目标检测任务，有必要从检测尺度配置、浅层特征增强和跨层特征流重构等角度出发，对现有 YOLO 结构进行面向场景特性的轻量化设计。

本文选择 YOLO11s 作为基线模型出于三方面考量：YOLO11s 在实时目标检测、参数规模和工程部署方面具有较成熟的实现基础，适合作为无人机边缘端轻量化改进的稳定基线；YOLO11s 原始检测头是 P3/P4/P5 三尺度结构，和本文设计的 P2/P3/P4 尺度重构层有明确对应关系；本文研究重点在于针对无人机小目标检测中的浅层细节衰减、尺度分布不匹配和跨层融合冲突问题，探索可迁移的尺度重构与特征再注入机制。

基于上述分析，本文以 YOLO11s 为基线，围绕无人机边缘端小目标检测中的尺度分布适配、浅层特征增强和跨层特征流重构开展改进。无人机小目标检测的核心矛盾在于：小目标依赖浅层高分辨率特征中的边缘、纹理和位置信息，但浅层特征语义表达较弱，且在特征金字塔网络（FPN, feature pyramid network）和路径聚合网络（PAN, path aggregation network）的多次下采样、融合和重构过程中，容易被深层语义特征稀释。为此，本文首先将原始 P3/P4/P5 检测结构重构为 P2/P3/P4，使浅层高分辨率特征直接参与微小目标预测；随后利用选择性引导模块（SGBlock, selective guidance block）对浅层 P2 特征进行选择性增强，在保留定位细节的同时抑制背景纹理干扰；进一步通过 P2 引导的跨层再注入机制，使增强后的浅层细节继续参与 P3 中间尺度特征构建；最后通过自适应融合

模块（ADF, adaptive fusion module）对增强 P2 路径、P3 语义路径和骨干 P2 先验路径进行动态权重分配，缓解多源特征直接融合带来的冲突。上述设计形成尺度适配—浅层筛选—细节回流—动态融合的闭环机制，在保证模型轻量化的前提下，提升无人机边缘端小目标检测的精度与实时部署适应性。本文的主要工作包括：

1) 针对无人机航拍场景中小目标占比较高、大目标样本占比较低的特点，对 YOLO11s 原有 P3/P4/P5 检测头尺度配置进行重构。通过引入 P2 高分辨率检测分支、移除对该场景收益相对有限的 P5 分支，并将高层语义建模模块重构至 P4 层，构建更适配无人机小目标分布的 P2/P3/P4 三尺度检测框架。

2) 针对浅层特征语义表达不足、复杂背景下小目标细节易被削弱的问题，设计 SGBlock 浅层特征增强模块，并构建 P2 引导的跨层再注入机制。该机制将增强后的 P2 特征及骨干网络中的 P2 先验重新引入 P3 特征构建过程，使浅层细节信息不再仅用于一次性检测，而是持续参与中间尺度特征表达，从而强化小目标细节表征和多层特征协同。

3) 针对多来源特征在不同场景下贡献不固定的问题，引入自适应融合模块 ADF，对增强 P2 路径、P3 语义路径和骨干 P2 先验路径进行动态权重分配。通过在再注入前自适应调整不同分支的特征贡献，提升跨层特征融合的稳定性和针对性，从而在检测精度、参数规模和实时性之间取得较好平衡。

1 方法设计

1.1 整体网络结构

针对无人机航拍场景中目标尺度小、密集分布且边缘端设备计算资源受限的问题，本文提出一种面向无人机边缘端小目标检测的轻量化 YOLO11s 改进方法。本文围绕小目标检测对高分辨率特征的需求以及边缘端部署对模型规模的约束，对 YOLO11s 的检测头尺度配置和跨层特征流进行重构^[22]。

原始 YOLO11s 采用 P3/P4/P5 三尺度检测头，其中 P3、P4 和 P5 分别对应不同下采样倍率的特征图，适用于通用目标检测场景。由于无人机视角下目标通常尺寸较小，浅层高分辨率特征对于目标

边缘、纹理和位置信息的表达更为关键，而深层大尺度检测分支对当前场景的收益相对有限。基于此，本文首先在基线模型中引入 P2 检测分支，使浅层高分辨率特征直接参与检测；进一步移除原始 P5 检测分支，并将高层语义增强模块重构至 P4 层，构建 P2/P3/P4 三尺度检测框架。随后，在 P2 路径引入选择性引导模块 SGBlock，对浅层特征进

行细节增强与背景抑制；在此基础上，设计 P2 引导的跨层再注入机制，将增强后的 P2 特征及骨干网络中的 P2 先验持续回流至 P3 层，并结合 ADF 提升多源特征聚合能力，最终得到适用于无人机边缘端部署的轻量化小目标检测网络^[23]，改进轻量化 YOLO11s 网络整体结构如图 1 所示。

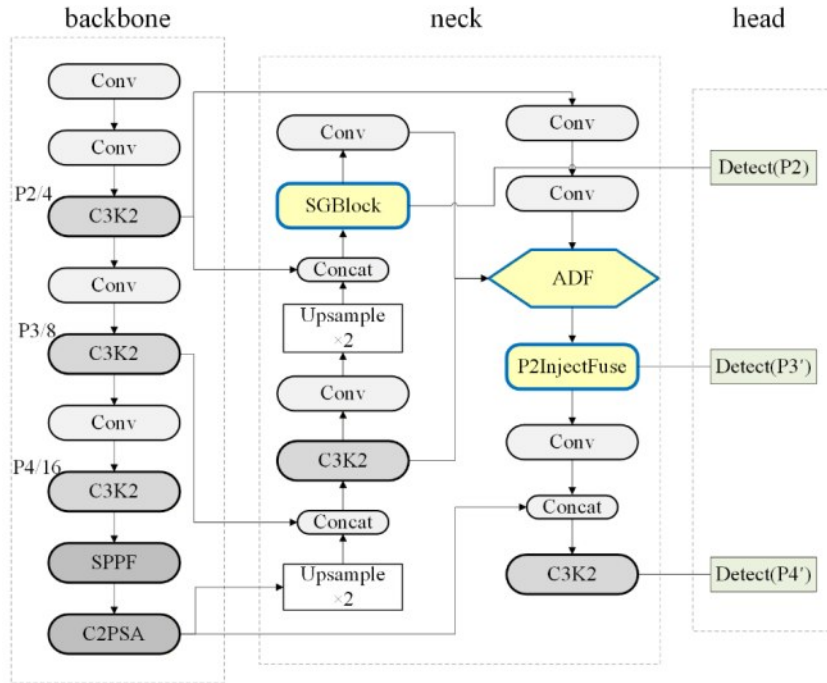


图1 改进后轻量化 YOLO11s 网络整体结构图

原始检测头如式 (1) 所示，改进检测头如式 (2) 所示：

$$S_{\text{base}} = \{ P3, P4, P5 \} \quad (1)$$

$$S_{\text{ours}} = \{ P2, P3', P4' \} \quad (2)$$

式中，P2 表示引入的浅层高分辨率检测特征，P3' 表示经过 P2 再注入与融合增强后的中间尺度特征，P4' 表示最终输出的高层检测特征， S_{base} 和 S_{ours} 分别表示基线模型和所提模型的检测尺度集合。

整体机制上，P2 分支、SGBlock、P2InjectFuse 和 ADF，围绕无人机小目标检测中的细节保留—语义补偿—跨层回流—冲突抑制形成闭环。首先，P2/P3/P4 尺度重构将检测重心前移，使小目标边缘、纹理和位置信息在浅层高分辨率特征中得到直接监督；其次，SGBlock 对 P2 浅层特征进行局部细节保留和全局门控筛选，在增强小目标相关响应的同时抑制背景纹理噪声；随后，P2InjectFuse 将增强后的 P2 细节重新注入 P3 构建过程，使浅层

细节不再停留于一次性检测，而是持续参与中间尺度语义表达；最后，ADF 在多源特征融合前进行动态权重分配，缓解增强 P2 路径、P3 语义路径和骨干 P2 先验之间的信息冲突。通过上述过程，网络实现了由尺度适配到特征筛选、由细节回流到自适应融合的协同增强。

1.2 基于小目标尺度适配的检测头轻量化重构

在无人机航拍场景中，目标通常只占据图像中极少的像素区域，若仍沿用常规检测器中以 P3/P4/P5 为核心的检测头设计，则浅层高分辨率细节信息无法被充分利用，容易在多次下采样与跨层融合过程中被削弱，从而影响微小目标的定位与识别。为此，本文首先在 YOLO11s 基线模型中引入 P2 检测分支，使浅层特征直接参与检测，以提升网络对微小目标轮廓和细节纹理的建模能力^[24]。

然而，直接增加 P2 分支虽然有助于提升小目标检测精度，但也会带来额外的参数数量和计算量开

销，不利于无人机边缘端实时部署^[25]。考虑到 Vis-Drone 场景中大目标占比相对较低，而深层大尺度检测分支对小目标的贡献有限，本文进一步移除原始 P5 分支，并将快速空间金字塔池化模块（SPPF, spatial pyramid pooling-fast）和跨阶段部分空间注意力模块（C2PSA, cross-stage partial spatial attention）等高层语义建模模块挂载到 P4 层，从而在保留必要语义感受野的同时，减少对低收益大尺度检测的计算投入^[26]。经重构后，网络最终形成 P2/P3/P4 三尺度检测框架。本文根据无人机航拍场景中的目标尺度分布，对检测头进行结构再分配。P2 特征图具有更高空间分辨率，能够保留更多边缘、角点和局部纹理信息，适合承担微小目标和密集小目标的初始定位任务；P3 特征在空间细节与语义表达之间具有较好的折中，可用于处理尺度略大但仍然容易受背景干扰的小目标；P4 特征则保留一定的高层语义感受野，用于增强类别判别能力和复杂场景下的鲁棒性。

特征路径上，重构后的检测头首先由 P4 特征上采样至 P3，并与骨干网络中的 P3 特征进行融合，得到中间尺度特征；随后继续上采样至 P2，并与骨干网络中的 P2 特征进行融合，获得适用于微小目标检测的高分辨率特征图。之后，通过下采样路径重新构建 P3 和 P4 的检测特征，最终形成 P2、P3 和 P4 三尺度输出。该设计一方面使检测尺度前移，更适配无人机视角下的小目标分布；另一方面通过裁剪 P5 分支降低模型冗余，提升边缘端部署友好性^[27]。仅靠尺度前移虽然提升了浅层特征利用率，但浅层特征本身语义弱、噪声多，因此仍需进一步增强其判别性。虽然引入 P2 检测分支能够增强浅层细节表征，但 P2 特征通常语义信息不足，且在复杂背景下容易混入大量无关纹理，导致小目标与背景之间的区分能力下降。为此，本文在 P2 路径设计选择性引导模块 SGBlock，其结构如图 2 所示。该模块通过局部分支与全局分支的联合建模，实现浅层特征的细节增强与噪声抑制。

设 SGBlock 的输入浅层特征为 $\mathbf{X} \in \mathbf{R}^{B \times C \times H \times W}$ ，其中 B 、 C 、 H 、 W 分别表示批次大小、通道数、特征图高度和宽度。为分别建模局部细节信息与全局引导信息，本文首先在通道维度上将 \mathbf{X} 划分为局部分

支输入 \mathbf{X}_l 和全局分支输入 \mathbf{X}_g 。如式 (3)

所示：

$$\mathbf{X} = [\mathbf{X}_l, \mathbf{X}_g] \quad (3)$$

其中， $\mathbf{X}_l, \mathbf{X}_g \in \mathbf{R}^{B \times C/2 \times H \times W}$ 。 \mathbf{X}_l 用于保留浅层高分辨率特征中的边缘、纹理和位置信息， \mathbf{X}_g 用于提供更大感受野下的辅助语义引导。

随后，分别对两分支进行线性投影，得到局部分支特征 \mathbf{F}_l 和全局分支特征 \mathbf{F}_g ，如式 (4) 所示：

$$\mathbf{F}_l = \phi_l(\mathbf{X}_l), \quad \mathbf{F}_g = \phi_g(\mathbf{X}_g) \quad (4)$$

式中， $\phi_l(\cdot)$ 和 $\phi_g(\cdot)$ 标识 1×1 卷积映射操作。若投影过程保持分支通道数不变，则 $\mathbf{F}_l, \mathbf{F}_g \in \mathbf{R}^{B \times C/2 \times H \times W}$ 。

为提升全局分支中有效语义的选择能力，本文在全局分支后引入轻量通道门控。具体而言，首先对 \mathbf{F}_g

进行全局平均池化（GAP, global average pooling），获得通道统计信息；随后经过两层卷积、修正线性单元（ReLU, rectified linear unit）和 Sigmoid 映射生成通道权重 \mathbf{W}_c ，如式 (5) 所示；随后利用 \mathbf{W}_c 对

全局分支特征进行通道重标定，得到通道筛选后的全局特征 \mathbf{F}_g^c ，如式 (6) 所示：

$$\mathbf{W}_c = \sigma(\text{Conv}_2(\delta(\text{Conv}_1(\text{GAP}(\mathbf{F}_g)))))) \quad (5)$$

$$\mathbf{F}_g^c = \mathbf{F}_g \odot \mathbf{W}_c \quad (6)$$

式中， $\text{GAP}(\cdot)$ 表示全局平均池化， $\delta(\cdot)$ 表示 ReLU 激活函数， $\sigma(\cdot)$ 表示 Sigmoid 函数， \odot 表示逐元素

相乘。 $\mathbf{W}_c \in \mathbf{R}^{B \times C/2 \times 1 \times 1}$ ，在与 \mathbf{F}_g 相乘时沿空间维度进行广播。

在完成通道维度的特征筛选后，空间门控进一步对 \mathbf{F}_g^c 进行空间响应建模。该过程首先沿通道维度

分别执行平均池化和最大池化，并将二者拼接形成空间描述图 \mathbf{M}_s ，如式 (7) 所示，其中， $\text{AvgPool}_c(\cdot)$ 和 $\text{MaxPool}_c(\cdot)$ 分别表示沿通道维度执行的平均池化和最大池化操作， $\mathbf{M}_s \in \mathbf{R}^{B \times 2 \times H \times W}$ ；随后通过深度卷积（DWConv, depthwise convolution）和 1×1 卷积生成空间权重 \mathbf{W}_s ，并利用 Sigmoid 函数完成归一化，如

式 (8) 所示；最后利用 \mathbf{W}_s 对 \mathbf{F}_g^c 进行空间重标定，得到经通道—空间联合筛选后的全局分支特征 $\tilde{\mathbf{F}}_g$ ，

如式 (9) 所示。

$$\mathbf{M}_s = [\text{AvgPool}_c(\mathbf{F}_g^c), \text{MaxPool}_c(\mathbf{F}_g^c)] \quad (7)$$

$$W_s = \sigma(\text{Conv}_{1 \times 1}(\text{DWConv}(M_s))) \quad (8)$$

$$\tilde{F}_g = F_g^c \odot W_s \quad (9)$$

式中， $\text{DWConv}(\cdot)$ 表示深度卷积操作， $W_s \in \mathbf{R}^{B \times 1 \times H \times W}$ ，在与 F_g^c 相乘时沿通道维度进行广播。

最后，利用可学习参数 θ 对全局分支进行加权注入，并与局部分支融合，再经混合模块得到 SGBlock 增强后的 P2 输出特征，表示为式 (10)：

$$F_{P2}^{\text{sg}} = \text{Mix}(F_1 + \theta \tilde{F}_g) \quad (10)$$

其中， $\text{Mix}(\cdot)$ 对应 SGBlock 后续的 C3k2 混合结构， θ 表示 SGBlock 中控制全局分支注入强度的可学习系数。 F_{P2}^{sg} 表示经过 SGBlock 增强后的浅层 P2 特征，并作为后续 P2 检测分支及 P2InjectFuse 再

注入机制的输入特征。

上述公式描述了 SGBlock 的前向计算过程。浅层 P2 特征虽然具有较高空间分辨率，但其通道响应中同时包含目标边缘、局部纹理、背景纹理和噪声干扰。若直接将其送入检测头或跨层融合模块，容易导致背景区域与小目标区域产生相似响应。SGBlock 通过通道维度划分将浅层特征分解为局部细节路径和全局引导路径，其中局部路径保留高分辨率定位信息，全局路径先通过通道门控建模“哪些语义通道更重要”，再通过空间门控建模“哪些空间位置更可能对应有效目标区域”。因此，SGBlock 实质上构成一种轻量级浅层特征筛选机制，在不显著增加模型复杂度的前提下增强小目标响应，并抑制复杂航拍背景中的无关纹理。

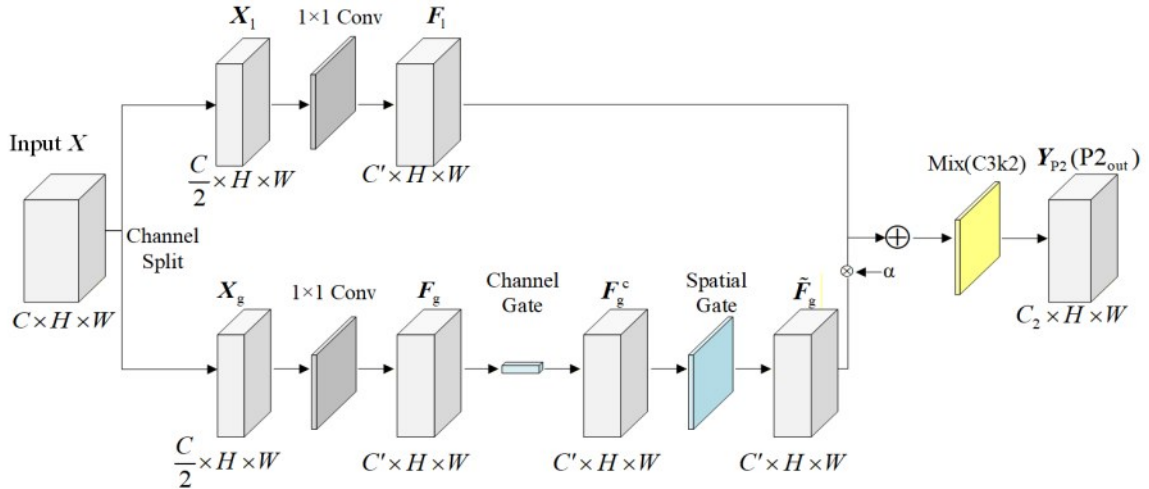


图2 SGBlock 结构图

1.3 P2 引导的跨层再注入机制

在传统 FPN 与 PAN 的多尺度特征融合结构中，浅层特征通常在一次融合后直接进入检测头，随后深层语义特征主导后续表示构建。这种方式虽然有利于形成多尺度语义表达，但浅层细节信息在后续路径中往往逐步衰减，不利于无人机视角下密集小目标的持续感知^[28]。为缓解这一问题，本文提出 P2 引导的跨层再注入机制，使增强后的 P2 特征在完成浅层检测后继续参与 P3 层特征构建。下文将该机制对应的实现模块命名为 P2 引导特征注入融合模块 (P2InjectFuse, P2-guided feature injection and fusion)。

该机制由三路输入构成：

- 1) 增强后的 P2 输出特征，即 $P2_{\text{out}}$ 下采样后

得到的浅层细节路径；

- 2) 自顶向下路径生成的 P3 特征即 X_{p3} ；
- 3) 骨干网络中的 P2 先验特征 X_{p2b} ，经轻量投影和下采样后作为补充浅层纹理路径。

设三路输入特征分别经通道对齐与投影后表示为式 (11)：

$$\begin{aligned} F_{p2h} &= G_h(\phi_h(X_{p2h})), \\ F_{p3} &= \phi_3(X_{p3}), \\ F_{p2b} &= G_b(\phi_b(X_{p2b})) \end{aligned} \quad (11)$$

$\phi_h(\cdot)$ 、 $\phi_3(\cdot)$ 、 $\phi_b(\cdot)$ 表示不同分支上的 1×1 卷积投影， $G_h(\cdot)$ 和 $G_b(\cdot)$ 表示相应门控操作。

随后，用可学习参数 η_h 和 η_b 分别控制两路浅层特征对中间尺度特征的注入强度，表示为式 (12)：

$$Y_{p3} = F_{p3} + \eta_h F_{p2h} + \eta_b F_{p2b} \quad (12)$$

最后，经混合模块对融合结果进行细化，得到增强后的中间尺度特征，如式（13）所示：

$$P3' = \text{Mix}(Y_{p3}) \quad (13)$$

其中， $\text{Mix}(\cdot)$ 对应 P2InjectFuse 中的 C3k2 混合结构。与传统 PAN 路径相比，该机制使 P2 特征不再仅用于一次性检测，而是以“细节回流”的方式持续参与 P3 表达构建，从而增强中间尺度特征对密集小目标的响应能力。需要指出的是，上述再注入过程虽然实现了浅层细节信息向中间尺度特征的持续回流，但来自增强 P2 分支、当前 P3 分支以及骨干浅层先验的多路特征在不同场景下贡献并不恒定。若仍采用固定权重或直接叠加方式进行融合，容易削弱各分支之间的互补性，进而限制再注入机制的作用发挥。

除前向信息回流外，P2InjectFuse 还改变了浅层特征的监督路径，使增强后的 P2 特征同时受到 P2 检测头和 P3 构建过程的约束。传统 PAN 结构中浅层特征通常只通过一次横向连接参与融合，之后主要由深层语义路径主导特征构建，导致浅层特征受到的监督信号较弱。本文引入 P2 检测头后，浅层特征首先受到 P2 检测损失的直接监督；进一步通过 P2InjectFuse 将增强后的 P2 特征重新注入 P3 构建过程，使其同时受到 P3 检测损失的间接约束。设总损失为式（14）：

$$L = L_{p2} + L_{p3} + L_{p4} \quad (14)$$

式中， L_{p2} 、 L_{p3} 和 L_{p4} 分别表示 P2、P3 和 P4 检测头对应的检测损失。增强后的浅层特征为 F_{p2}^{sg} ，再注入后的中间尺度特征为 F'_{p3} ，则浅层特征的梯度可表示为式（15）：

$$\frac{\partial L}{\partial F_{p2}^{sg}} = \frac{\partial L_{p2}}{\partial F_{p2}^{sg}} + \frac{\partial L_{p3}}{\partial F'_{p3}} \frac{\partial F'_{p3}}{\partial F_{p2}^{sg}} \quad (15)$$

其中，第一项来自 P2 检测头的直接监督，第二项来自 P3 检测头经再注入路径反向传播的监督。相比仅将 P2 特征用于一次性检测，该机制为浅层特征建立了额外的优化路径，使其在网络训练过程中持续参与中间尺度表征学习，从而缓解小目标细节在 PAN 路径中的衰减。

1.4 自适应多分支融合机制

在上述 P2 引导的跨层再注入机制中，中间尺度特征的构建依赖于增强后的 P2 细节路径、自顶向下的 P3 语义路径以及骨干浅层先验三路信息。虽然这种多源输入能够有效补充小目标相关细节，

但不同来源特征在不同场景、不同目标密度下的重要性并不相同。若仍采用固定权重或简单叠加方式进行融合，容易造成分支间信息竞争，削弱浅层细节与中层语义之间的协同表达。基于此，本文进一步引入自适应融合模块 ADF，对再注入前的多源特征进行动态加权建模，从而提升跨层特征聚合的针对性与稳定性。设输入分支数为 N ，第 i 路输入特征为 X_i 。首先，如式（16）所示，通过 1×1 卷积对各分支特征进行通道对齐。

$$F_i = \phi_i(X_i), \quad i = 1, 2, \dots, N \quad (16)$$

式中， $\phi_i(\cdot)$ 表示第 i 路特征的 1×1 卷积投影， F_i 表示通道对齐后的第 i 路特征。随后，通过全局平均池化和轻量感知机为每一路特征生成响应分数 s_i ，如式（17）所示：

$$s_i = \psi_i(F_i) \quad (17)$$

其中 $\psi_i(\cdot)$ 表示第 i 路轻量感知机。再通过 Softmax 函数将其归一化为分支权重，如式（18）所示：

$$w_i = \frac{e^{s_i}}{\sum_{j=1}^N e^{s_j}} \quad (18)$$

其中 w_i 表示经 Softmax 归一化后的分支权重。最终，多分支特征进行加权求和并经卷积细化，得到融合输出，如式（19）所示：

$$F_{\text{out}} = \text{Conv}\left(\sum_{i=1}^N w_i F_i\right) \quad (19)$$

其中 F_{out} 表示多分支融合输出。通过上述机制，网络能够根据输入场景自适应调整不同来源特征的贡献度，从而增强 P2 细节路径、P3 语义路径与骨干 P2 先验之间的互补表达能力。在实际结构中，ADF 位于三路特征再注入之前，先对不同来源特征的响应强度进行自适应分配，再由后续 P2InjectFuse 完成融合细化与表达重构，二者共同构成完整的 P2 引导跨层增强过程，P2InjectFuse 与 ADF 的协同融合结构如图 3 所示。增强 P2 路径、P3 语义路径和骨干 P2 先验路径，从多源特征融合角度来看，分别携带不同类型的信息：增强 P2 路径侧重小目标边缘和局部纹理，P3 路径则侧重中层语义和上下文，骨干 P2 先验路径保留未经检测头扰动的原始浅层细节。三类特征在不同图像中的有效性并不固定。例如，在密集小目标场景下，P2 细节路径应获得更高权重；在背景复杂或尺度稍大的目标场景下，P3 语义路径的贡献应相应增强。

若采用直接拼接或固定权重相加，容易造成浅层纹理噪声与中层语义信息之间的冲突。因此，ADF 通过 Softmax 归一化权重实现分支级信息选择，如式 (20) 所示：

$$\mathbf{F}_{\text{ADF}} = \phi \left(\sum_{i=1}^N w_i \hat{\mathbf{F}}_i \right), \sum_{i=1}^N w_i = 1 \quad (20)$$

其中， $\phi(\cdot)$ 表示融合后的卷积细化操作， \mathbf{F}_{ADF} 表示 ADF 模块的输出特征， $\hat{\mathbf{F}}_i$ 表示通道对齐后的分支特征。Softmax 约束使不同分支之间形成竞争式权重分配，从而突出与当前输入场景更相关的特征来源，抑制冗余或噪声分支。该机制可视为轻量级的信息瓶颈调节过程，即在多源特征进入 P2InjectFuse 前先完成有效信息筛选，降低无关背景纹理和冲突语义对融合结果的干扰。

1.5 检测头输出与损失函数

经过上述检测头重构、选择性引导增强以及

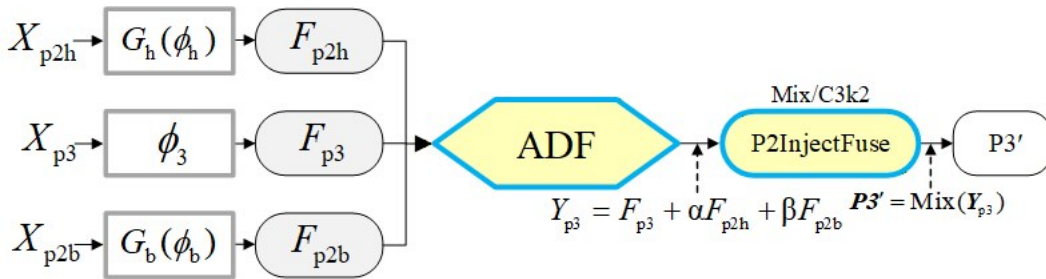


图3 P2InjectFuse 与 ADF 融合结构图

2 实验

2.1 实验数据集

本文实验采用无人机航拍场景下具有代表性的目标检测数据集 VisDrone。该数据集由无人机平台在不同地区、不同高度和不同拍摄视角下采集获得，场景覆盖城市道路、行人区域、车辆密集区域等复杂环境，具有小目标占比高、目标密集、遮挡严重及背景干扰强等特点，因此能够有效检验模型在复杂无人机场景中的目标检测性能。VisDrone 数据集共包含 10,209 张图像，其中 6,471 张用于训练，548 张用于验证，其余为测试集图像。本文后续实验中的精确率 (P, precision)、召回率 (R, recall)、平均精度均值 (mAP, mean average precision) @0.5 及 mAP@0.5:0.95 均在 VisDrone 验证集上计算，测试集不参与模型训练和消融实验指标统计。

P2 引导的跨层再注入与自适应融合后，网络最终输出三个尺度的检测特征，即 P2、P3' 和 P4'，分别面向微小目标、小目标和中等尺度目标的检测。其中，P2 特征主要负责极小目标和高密度小目标的定位，P3' 特征在浅层细节与中层语义之间实现折中，P4' 特征则提供更强的语义判别能力，以保证整体检测框架的稳定性。

考虑到本文工作的重点在于检测尺度配置与特征流重构，损失函数部分不对 YOLO11 基线做额外修改，而是沿用目标检测任务中的多任务联合优化方式。总损失函数如式 (21) 所示：

$$L = \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{dn}} L_{\text{dn}} \quad (21)$$

式中， L_{box} 表示边界框回归损失， L_{cls} 表示分类损失， L_{dn} 表示分布焦点损失， λ_{box} 、 λ_{cls} 、 λ_{dn} 为相应权重系数。通过上述联合优化，网络能够在保持较低模型复杂度的前提下，学习更适合无人机边缘端小目标检测场景的判别特征^[29]。

2.2 评价指标

为了评价模型性能，本文采用精确率 (P)、召回率 (R)、mAP@0.5 和 mAP@0.5:0.95 作为评价指标。其中，mAP@0.5 为交并比 (IoU, intersection over union) 阈值取 0.5 时各类别平均精度 (AP, average precision) 的均值；mAP@0.5:0.95 为 IoU 阈值从 0.5 到 0.95、步长为 0.05 时 AP 的平均值。AP 为精确率—召回率曲线 (PR curve, precision-recall curve) 下的面积，其计算如式 (22) 所示：

$$\text{AP} = \int_0^1 P(R) dR \quad (22)$$

令真正例 (TP, true positive)、假正例 (FP, false positive) 和假负例 (FN, false negative) 分别表示正确检测目标数量、误检目标数量和漏检目标数量，则精确率和召回率分别由式 (23) 和式 (24) 计算：

$$P = \frac{TP}{TP + FP} \quad (23)$$

$$R = \frac{TP}{TP + FN} \quad (24)$$

为考察模型在不同尺度目标上的表现，本文进一步按照标注框像素面积进行划分。具体而言，根据 YOLO 格式标签中的归一化宽高和原始图像尺寸，将目标框宽高还原为像素尺度，并计算该目标的面积

$\text{area} = w_{\text{box}} \times h_{\text{box}}$ ，area 单位为像素的平方。其

中， mAP_{Tiny} 表示 $\text{area} < 16^2$ 的目标检测性能， mAP_{S} 表

示 $\text{area} < 32^2$ 的小目标检测性能，包含 Tiny 目标； mAP_{M} 表示 $32^2 \leq \text{area} < 96^2$ 的中目标检测性能； mAP_{L} 表示 $\text{area} \geq 96^2$ 的大目标检测性能。统计结果如表 1 所示，训练集中 Small+Tiny 目标占全部目标的 60.49%；Large 目标为 18,980，占 5.53%。这表明 VisDrone 数据集目标尺度分布明显偏向小尺度目标，为本文将检测尺度由 P3/P4/P5 前移至 P2/P3/P4 提供了数据依据。

表 1 VisDrone 训练集不同类别目标尺度分布统计

类别编号	名称	总数	Tiny	Small-only	Medium	Large	Tiny+Small 占比/%
0	pedest	79,337	34,337	30,901	13,793	306	82.23
1	people	27,059	12,989	10,521	3,450	99	86.88
2	bicycle	10,480	2,393	4,710	3,250	127	67.78
3	car	144,867	25,462	44,385	63,084	11,936	48.21
4	van	24,956	2,929	7,822	11,523	2,682	43.08
5	truck	12,875	1,057	2,927	6,724	2,167	30.94
6	tricycle	4,812	656	1,523	2,423	210	45.28
7	aw-tricy	3,246	392	973	1,695	186	42.05
8	bus	5,926	422	1,235	3,269	1,000	27.96
9	motor	29,647	8,627	13,344	7,409	267	74.11
合计	-	343,205	89,264	118,341	116,620	18,980	60.49

pedest 表示 pedestrian, aw-tricy 表示 awning-tricycle, 表 6 同理。

2.3 实验设置

本文实验在 Linux 系统下完成。模型训练使用一张 24 GB 显存的 RTX 3090 GPU；模型每秒处理帧数 (FPS, frames per second) 的评估是在 Jetson Orin NX 16 GB 边缘计算平台上完成的。软件环境方面，实验基于 Python 3.13 和 CUDA 12.7 环境完成模型训练与测试。

本文将网络的输入图像尺寸统一调整为 640×640 像素，单次参与训练的样本批次大小设定为 8，最大训练轮数设为 200，采用早停策略，实际在第 120 个训练轮次停止，训练曲线统计至停止轮次。训练采用 AdamW 优化器，初始学习率设为 0.001，并使用线性衰减策略调整学习率。此外，整个训练流程关闭混合精度训练功能。详细的软硬件配置清单均已汇总于表 2 中。除文中特别声明的改进之外，其余各项未提及的模型超参数均默认。

2.4 实验结果分析

1) 整体性能与收敛分析

表 2 实验软硬件环境与训练参数设置

参数名称	实验设置
操作系统	Linux
GPU	RTX 3090
显存	24 GB
Python 版本	3.13
CUDA 版本	12.7
输入尺寸	640×640
batch size	8
epoch	最大训练轮数为 200, 早停于第 120 轮
优化器	AdamW
初始学习率	0.001
学习率衰减策略	线性衰减
混合精度训练	关闭
FPS 测试设备	Jetson Orin NX 16 GB

为验证本文方法，首先对不同模型的训练过程和整体检测性能进行分析。图 4 给出了基线模型与改进模型在训练过程中的 mAP 变化曲线。可见各模型在训练前期均表现较快的性能提升，在中后期

逐渐趋于收敛。相较于基线 YOLO11s，本文方法在训练中后期保持更高的 mAP 水平，说明改进后

的检测尺度配置和特征增强结构能够提升模型在 VisDrone 数据集上的检测能力。

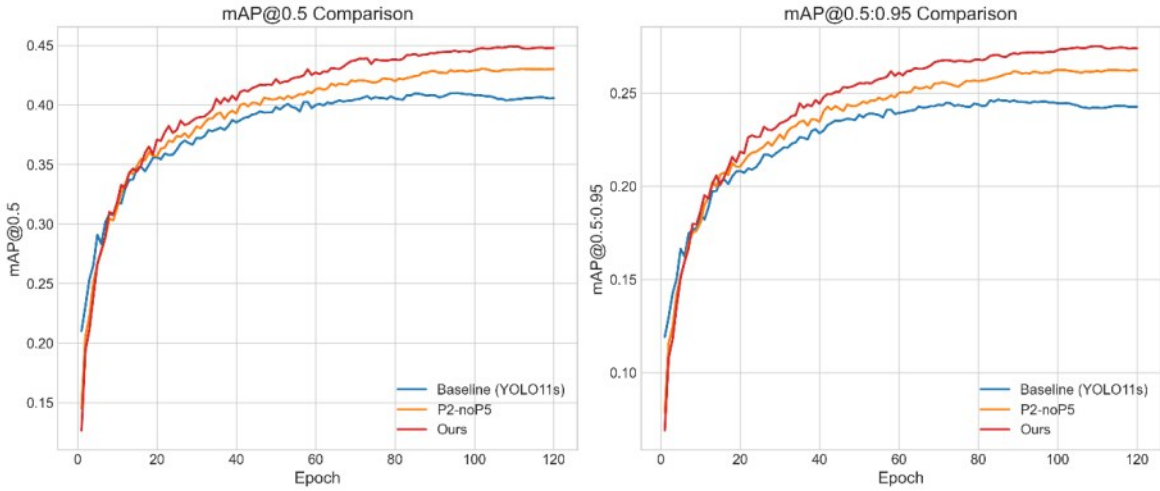


图4 不同模型训练过程中的 mAP 变化曲线

由表 3 可知，基线 YOLO11s 的 mAP@0.5 和 mAP@0.5:0.95 分别为 40.50% 和 24.24%；本文最终模型分别达到 44.90% 和 27.46%。同时，精确率由 51.67% 提升至 55.22%，召回率由 38.99% 提升至 42.86%。上述结果表明，本文方法不仅提升了整体检测精度，也在一定程度上改善了模型的检出能力。

2) 消融实验分析

为进一步分析各改进模块的贡献，本文按照“P2 分支引入、P5 分支移除、SGBlock 浅层增强、P2InjectFuse 跨层再注入、ADF 自适应融合”的顺序进行消融实验，结果如表 3 所示。

表3 不同改进阶段模型检测性能对比

模型 (Model)	<i>P</i>	<i>R</i>	mAP @0.5	mAP@0.5:0.95
YOLO11s	51.67	38.99	40.50	24.24
+P2	52.66	41.43	42.61	25.89
+P2-P5	54.50	40.65	43.09	26.28
+SGBlock	53.93	42.22	43.66	26.52
+P2InjectFuse	54.20	42.13	44.02	27.03
+ADF	55.22	42.86	44.90	27.46

在基线 YOLO11s 上加入 P2 检测分支后，模型 mAP@0.5 由 40.50% 提升至 42.61%，mAP@0.5:0.95 由 24.24% 提升至 25.89%。该结果说明，在无人机航拍图像中引入浅层高分辨率检测分支，有助

于补偿小目标在深层下采样过程中损失的空间细节信息，从而提升小目标检测效果。

在此基础上移除 P5 分支后，+P2-P5 模型的 mAP@0.5 和 mAP@0.5:0.95 分别达到 43.09% 和 26.28%，较 +P2 模型继续提升。该结果表明，对于 VisDrone 这类小目标占比较高的数据集，削减面向大尺度目标的 P5 检测分支，并将检测资源向 P2/P3/P4 尺度集中，能够进一步提高模型对小目标场景的适配性。

加入 SGBlock 后，模型 mAP@0.5:0.95 由 26.28% 提升至 26.52%。虽然整体增幅相对有限，但结合

表 5 可以看出，mAP_{Tiny} 由 5.72% 提升至 5.81%，mAP_s 由 14.05% 提升至 14.83%。这说明 SGBlock

主要通过增强浅层有效细节并抑制背景干扰，改善极小目标和密集小目标的检测效果。

继续引入 P2InjectFuse 后，模型 mAP@0.5:0.95 提升至 27.03%。从尺度结果看，mAP_M 由 34.39% 提升至 35.73%，同时 mAP_{Tiny} 和 mAP_s 仍保持在较高水平。该结果表明，P2InjectFuse 能够缓解常规 PAN 路径中浅层细节信息在后续特征传递过程中逐步衰减的问题，使增强后的 P2 特征继续参与 P3 层表达构建，从而提升中间尺度特征对小目标的感知能力。

加入 ADF 后，模型取得最优结果，mAP@0.5

和 $mAP@0.5:0.95$ 分别达到 44.90% 和 27.46%。相较于 P2InjectFuse 阶段，ADF 使 $mAP@0.5:0.95$ 进一步提升。该结果说明，不同来源特征在复杂无人机场景中的贡献并不固定，通过自适应权重分配能够提高 P2 细节路径、P3 语义路径和骨干 P2 先验之间的融合稳定性。

3) 模型复杂度与实时性分析

为分析不同改进阶段对模型规模和推理效率的影响，本文统计了各模型的参数量、浮点运算量 (FLOPs, floating-point operations) 和 FPS，结果如表 4 所示。所有 FPS 测试均在 Jetson Orin NX 16 GB 上完成，输入尺寸为 640×640 ，batch size 为 1。为更贴近边缘端部署场景，本文将各阶段模型分别导出并构建 TensorRT 32 位浮点精度 (FP32, 32-bit floating-point precision) 推理引擎，统计 TensorRT FP32 引擎的单帧推理时间，并换算得到 FPS。

由表 4 可知，基线 YOLO11s 的参数量为 9.46M，FLOPs 为 21.72G，FPS 为 67.24。引入 P2 分支后，模型参数量略增至 9.62M，FLOPs 增至 29.44G，FPS 降至 53.31，说明高分辨率浅层检测分支会带来明显计算开销。

进一步移除 P5 分支后，参数量大幅降至 2.89M，FPS 回升至 63.11。需要说明的是，该操作并非仅删除 Detect 中的 P5 输出，而是整体裁剪 P5/32 高通道语义路径，包括 P5/32 下采样层、1024 通道 C3k2 模块及 head 中 P4 到 P5 的下采样融合路径，并将 SPPF、C2PSA 由 P5 层重构至 P4 层，因此能够显著减少模型参数规模。在后续加入 SGBlock、P2InjectFuse 和 ADF 后，模型参数量逐步增加，最终达到 3.82M，仍显著低于基线 YOLO11s。该结果表明，本文方法的轻量化优势主要体现在参数量和模型存储开销方面，而非理论计算量降低。由于 P2 高分辨率特征图参与检测和融合，浅层操作带来了额外计算负担。在 TensorRT FP32 部署测试中，最终模型在 Jetson Orin NX 16 GB 上达到 52.63 FPS，虽然低于基线 YOLO11s 的 67.24 FPS，但仍满足无人机边缘端实时检测需求。综合来看，本文方法在参数量由 9.46M 降至 3.82M 的同时，将 $mAP@0.5$ 从 40.50% 提升至 44.90%，体现出较好的检测精度、模型规模与 TensorRT FP32 推理效率折中。

进一步从边缘端部署角度看，参数量、FLOPs

和 FPS 分别反映了模型开销的不同侧面。参数量主要影响模型存储、加载和显存占用，FLOPs 反映理论计算复杂度，而 FPS 更直接体现模型在具体硬件平台和推理框架下的实际运行效率。本文方法虽然由于引入 P2 高分辨率特征导致 FLOPs 高于基线模型，但最终参数量仅为基线模型的约 40.38%，说明其在模型存储和权重传输方面具有更明显优势。对于无人机边缘设备而言，模型体积降低有助于减轻存储压力，并为多任务模型部署和边缘端协同推理预留更多资源。

表 4 不同改进阶段模型复杂度与推理速度对比

模型 (Model)	Params	FLOPs	TensorRT
	(M)	(G)	FP32 FPS
YOLO11s	9.46	21.72	67.24
+P2	9.62	29.44	53.31
+P2-P5	2.89	22.18	63.11
+SGBlock	2.93	23.74	59.77
+P2InjectFuse	3.12	24.51	55.14
+ADF(本文方法)	3.82	28.31	52.63

4) 尺度敏感性分析

为进一步分析本文方法对不同尺度目标的影响，表 5 给出了各改进阶段在 Tiny、Small、Medium 和

Large 目标上的检测结果。相较于基线 YOLO11s，本文最终模型的 mAP_{Tiny} 由 3.44% 提升至 6.16%；

mAP_s 由 11.61% 提升至 15.35%； mAP_M 由 32.76% 提升至 35.69%。结果表明，本文方法对极小目标、小目标和中等尺度目标均具有改善作用，其中对小尺度目标的提升尤为明显。

与此同时， mAP_L 由 49.60% 下降至 43.09%。其主要原因在于，本文为强化小目标检测能力，将检测尺度重心前移至 P2/P3/P4，并移除了偏向大尺度目标建模的 P5 检测分支，使模型对大目标全局结构和高层语义感受野的建模能力有所削弱。该结果表明，本文方法在提升 Tiny、Small 和 Medium 等主要目标尺度检测性能的同时，对 Large 目标存在一定精度损失，因而更适用于小目标和密集目标占主导的无人机航拍场景。对于大型车辆、建筑设施、桥梁构件等中大尺度目标占比较高，或需要同时稳定检测大目标的业务，可考虑保留 P5 分支，构建 P2/P3/P4/P5 四尺度检测结构，或进一步引入

尺度自适应检测头，以兼顾小目标精度和大目标鲁棒性。

从各阶段变化看，P2 分支显著提升了 mAP_{Tiny} 和 mAP_S ，说明浅层高分辨率特征对小目标检测具

有直接作用；SGBlock 进一步改善了小尺度目标结果，表明浅层特征增强能够提高目标与复杂背景之间的区分度；P2InjectFuse 和 ADF 则主要提升了中等尺度目标和整体 mAP ，说明跨层再注入和动态融合有助于增强多尺度特征协同表达。

综上，本文方法更适用于目标尺度分布明显偏向 Tiny/Small 的无人机航拍检测任务，尤其是多数目标框面积低于 32^2 平方像素、目标密集且边缘端模型存储受限的场景。对于中大目标占比较高、需稳定检测大型车辆、桥梁构件或基础设施目标的任务，本文的 P2/P3/P4 结构可能因缺少 P5 大感受野分支而带来 Large 目标精度损失，此时建议采用 P2/P3/P4/P5 四尺度检测头或尺度自适应检测头。

5) 类别检测结果分析

表 6 给出了基线 YOLO11s 与本文方法在 Vis-Drone 各类别上的检测结果。总体来看，本文方法在多数类别上均取得了更高的检测精度。其中，pedestrian 的 $mAP@0.5:0.95$ 由 20.7% 提升至 24.5%，people 由 13.3% 提升至 17.7%，bicycle 由 5.8% 提升至 8.2%，motor 由 21.0% 提升至 24.5%。这些类别通常具有目标尺寸小、遮挡频繁、外观纹理弱和背景干扰强等特点，对浅层高分辨率特征依赖较高，因此与本文方法的改进方向相吻合。

从召回率变化看，people 类别由 30.3% 提升至 38.2%，bicycle 由 16.9% 提升至 21.4%，motor 由 47.1% 提升至 50.6%，说明本文方法对降低小目标漏检具有积极作用。对于 car、van、bus 等相对较大的类别，本文方法也取得了一定提升。例如 car 的 $mAP@0.5:0.95$ 由 52.3% 提升至 60.4%，bus 由 41.7% 提升至 46.1%。这表明移除 P5 分支后，模型虽然在大尺度目标总体指标上有所下降，但并未破坏常见车辆类目标的基本检测能力。

需注意，truck 类别的精确率由 58.0% 下降至 54.8%，但召回率由 36.5% 提升至 39.6%， $mAP@0.5:0.95$ 由 27.3% 小幅提升至 27.8%。这说明模型在部分大尺度或外观差异较大的类别上存在精确率与召回率之间的权衡。后续可结合类别重采样、损失权重调整或尺度自适应检测头进一步优化该类目

标的检测稳定性。

表 5 不同改进阶段在不同尺度目标上的检测结果对比

改进项	mAP_{Tiny}	mAP_S	mAP_M	mAP_L
YOLO11s	3.44	11.61	32.76	49.60
+P2	5.51	14.11	33.68	44.09
+P2-P5	5.72	14.05	34.25	42.98
+SGBlock	5.81	14.83	34.39	41.88
+P2InjectFuse	5.68	14.67	35.73	44.01
+ADF	6.16	15.35	35.69	43.09

表 6 不同类别目标检测结果对比

Class	YOLO11s				本文方法			
	<i>P</i>	<i>R</i>	$mAP@0.5:0.95$	$mAP@0.5:0.95$	<i>P</i>	<i>R</i>	$mAP@0.5:0.95$	$mAP@0.5:0.95$
all	51.7	39.0	40.5	24.2	55.2	42.9	44.9	27.5
pedest	54.7	42.1	45.0	20.7	61.1	45.4	51.0	24.5
people	56.1	30.3	34.9	13.3	59.0	38.2	42.2	17.7
bicycle	28.5	16.9	13.7	5.8	32.3	21.4	17.6	8.2
car	73.1	78.4	75.6	52.3	74.2	81.4	83.7	60.4
van	54.4	42.7	45.0	32.1	53.5	46.9	48.4	35.4
truck	58.0	36.5	40.4	27.3	54.8	39.6	41.1	27.8
tricycle	45.3	29.0	30.2	17.2	48.9	30.3	31.6	18.8
aw-tricy	32.5	17.9	16.1	10.2	35.3	19.5	18.5	11.8
bus	68.7	51.4	57.1	41.7	71.7	57.6	62.6	46.1
motor	52.7	47.1	46.8	21.0	58.4	50.6	52.0	24.5

6) 横向对比与可视化分析

为验证本文方法的综合性能，本文在 Vis-Drone 数据集上与多种典型检测模型进行横向对比，结果如表 7 所示，带“*”项为相同配置下的复现实验。

从参数效率看，本文方法参数量仅为 3.8M，明显低于 YOLOv8l、YOLOv10l、YOLO11l 以及多数改进模型。虽然 Drone-YOLO 和 LE-YOLO 的参数量更低，但其 $mAP@0.5$ 分别比本文方法低 6.8 和 5.6 个百分点。上述结果表明，本文方法在检测精度与参数规模之间取得了较优折中。

基线模型与本文方法在典型无人机航拍场景下的检测效果对比如图 5 所示。相较于基线模型，本文方法能够检出更多低置信度小目标，并在一定程度上减少密集场景中的漏检现象。

综合表 4 至表 7 的结果可以看出，本文方法的性能提升来源于检测尺度重构、浅层特征增强、跨

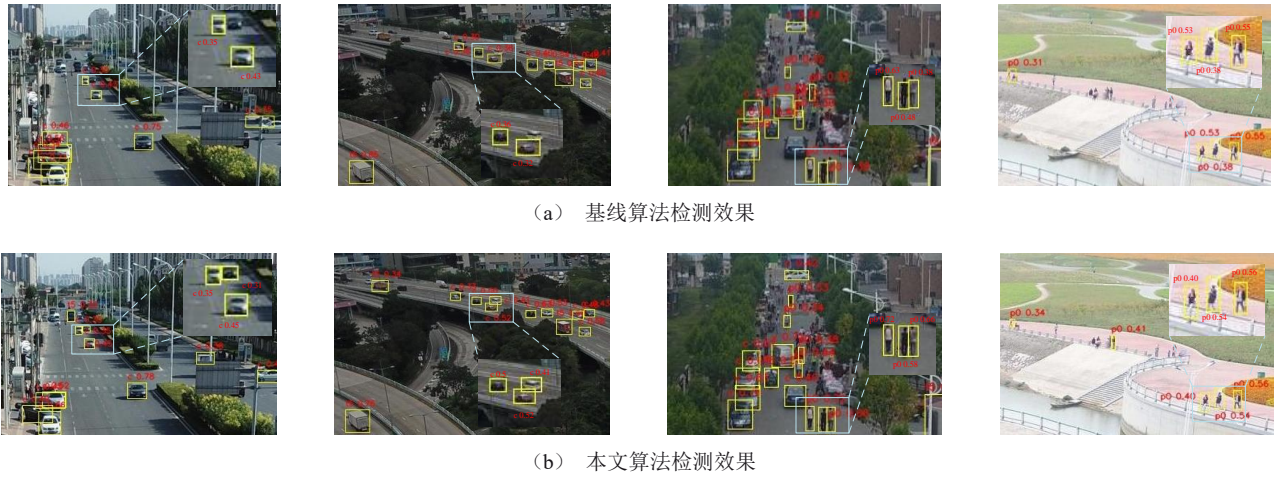


图5 本文方法与基线模型在典型场景下的检测结果对比

图中 c 表示 car, p0 表示 pedestrian, 数字表示检测置信度

层再注入和自适应融合的协同作用，能够在控制参数规模的同时提升无人机小目标检测性能。

从可视化结果看，本文方法在密集行人、小车辆和远距离目标区域具有更高的检出率，说明 P2 高分辨率检测分支和跨层再注入机制能够有效缓解浅层细节信息丢失问题。与此同时，部分大目标检测置信度下降以及个别密集区域仍存在漏检现象，表明模型在小目标优先优化后，对大目标全局结构和严重遮挡目标的处理仍有提升空间。

表7 VisDrone 数据集上与其他检测模型的性能对比

模型 (Model)	mAP@0.5/%	Params (M)
YOLOv8l	43.4	43.6
YOLOv10l ^[30]	41.7	25.7
YOLO11l	45.9	25.3
YOLOv12s* ^[31]	39.8	9.3
YOLO26s*	40.2	10.0
YOLOv13s*	40.7	9.6
RT-DETR-R18* ^[32]	44.7	19.9
HIC-YOLOv5* ^[33]	43.4	9.3
PDWT-YOLO* ^[7]	41.8	6.4
Drone-YOLO ^[34]	38.1	3.1
LRDS-YOLO ^[35]	43.6	4.2
MSUD-YOLO ^[14]	43.4	6.8
LCM-YOLO ^[15]	40.7	8.2
LE-YOLO ^[36]	39.3	2.1
本文方法	44.9	3.8

YOLOv13s 和 YOLO26s 为近期公开的开源实时检测模型，本文基于其公开实现或官方配置在相同训练设置下复现实验，仅作为工程基线参考。

3 结束语

针对无人机边缘端场景下小目标检测中精度、实时性与轻量化难以兼顾的问题，对 YOLO11s 模型进行了针对性改进，提出了一种面向无人机边缘端小目标检测的轻量化检测方法。针对无人机航拍图像中目标尺度小、分布密集、易遮挡以及浅层细节信息易在下采样过程中衰减等特点，该方法首先在基线网络中引入 P2 检测分支，以强化高分辨率浅层特征对微小目标的表征能力；进一步结合 VisDrone 数据集中目标尺度分布特征，移除对大目标检测收益有限的 P5 分支，并将高层语义建模重构至 P4 层，构建了更适配无人机小目标场景的 P2/P3/P4 三尺度检测框架。在此基础上，设计选择性引导模块对浅层关键信息进行增强，同时构建 P2 引导的跨层再注入与自适应融合机制，以提升中间层特征对小目标的持续感知和多层协同表达能力。从作用机制角度看，该方法的核心在于对浅层细节信息的持续利用与选择性融合。P2 分支解决小目标直接监督不足的问题，SGBlock 解决浅层特征语义弱且背景噪声多的问题，P2InjectFuse 解决浅层细节在 PAN 路径中逐步衰减的问题，ADF 则解决多来源特征直接融合时贡献不一致和语义冲突的问题。因此，本文方法形成了尺度前移提供细节入口、门控筛选增强有效浅层响应、跨层再注入延长细节传播路径、自适应融合抑制多源冲突的完整机制链条。

实验结果表明，所提方法在 VisDrone 数据集

上取得了较基线模型更优的检测性能, $mAP@0.5$ 和 $mAP@0.5:0.95$ 均实现了较为明显的提升; 与此同时, 模型参数量显著下降, 说明该方法在增强小目标检测能力的同时, 有效降低了参数规模和模型存储开销。需要指出的是, 由于高分辨率浅层特征参与检测与融合, 模型 FLOPs 有所增加, TensorRT FP32 推理速度相较基线模型有所下降, 但在 Jetson Orin NX 16 GB 上仍保持实时检测可接受水平。综合实验结果表明, 本文方法在无人机边缘端小目标检测任务中较好地兼顾了检测精度、模型规模与 TensorRT 部署推理效率, 具备进一步部署至资源受限边缘设备的应用潜力。

参考文献:

- [1] 杨永刚, 姜文韬, 高志云. 低空无人机实时目标检测算法[J]. 航空学报, 2025, 46(16): 331619.
YANG Y G, JIANG W T, GAO Z Y. Real-time object detection algorithm for low-altitude UAVs[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(16): 331619.
- [2] 吴一全, 童康. 基于深度学习的无人机航拍图像小目标检测研究进展[J]. 航空学报, 2025, 46(3): 30848.
WU Y Q, TONG K. Research progress on small object detection in UAV aerial images based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(3): 30848.
- [3] 高广帅, 尚云琦, 董燕. 光学遥感图像小目标检测深度学习算法综述[J]. 中国图象图形学报, 2025, 30(11): 3479-3505.
GAO G S, SHANG Y Q, DONG Y. Review of deep learning algorithms for small object detection in optical remote sensing images [J]. Journal of Image and Graphics, 2025, 30(11): 3479-3505.
- [4] 张琴, 郭为安. 深度学习小目标检测算法综述[J]. 计算机应用研究, 2025, 42(10): 2893-2904.
ZHANG Q, GUO W A. Survey of deep learning-based small object detection algorithms[J]. Application Research of Computers, 2025, 42(10): 2893-2904.
- [5] 徐鹤, 张恩俊, 谭萍. AGRF-Net: 基于边界感知与自适应门控残差融合的 RGB-D 语义分割网络[J/OL]. 南京邮电大学学报(自然科学版), 2026: 1-11.
XU H, ZHANG E J, TAN P. AGRF-Net: An RGB-D semantic segmentation network based on boundary-aware and adaptive gated residual fusion[J/OL]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2026: 1-11.
- [6] 刘尚东, 吴健, 徐鹤, 等. 基于改进 CycleGAN 的低光域图像增强模型[J/OL]. 数据采集与处理, 2026: 1-16.
LIU S D, WU J, XU H, et al. A low-light domain image enhancement model based on improved CycleGAN[J/OL]. Journal of Data Acquisition and Processing, 2026: 1-16.
- [7] ZHANG L, XIONG N, PAN X, et al. Improved object detection method utilizing YOLOv7-Tiny for unmanned aerial vehicle photographic imagery[J]. Algorithms, 2023, 16(11): 520.
- [8] NISA U, MOHD POZI M S, SAIP M A. A decade of research in small object detection: A comprehensive bibliometric analysis[J]. International Journal of Data Science and Analytics, 2025, 20(8): 7331-7355.
- [9] NIKOUEI M, BAROUTIAN B, NABAVI S, et al. Small object detection: A comprehensive survey on challenges, techniques and real-world applications[J]. Intelligent Systems with Applications, 2025, 27: 200561.
- [10] ZHAO Y, SUN H, WANG S. Small object detection in medium-low-resolution remote sensing images based on degradation reconstruction[J]. Remote Sensing, 2024, 16(14): 2645.
- [11] 张杨, 程正东, 朱斌. 基于改进 Faster R-CNN 的无人机小目标检测算法[J]. 量子电子学报, 2022, 39(3): 354-363.
ZHANG Y, CHENG Z D, ZHU B. UAV small object detection algorithm based on improved Faster R-CNN[J]. Chinese Journal of Quantum Electronics, 2022, 39(3): 354-363.
- [12] LIU M, WANG X, ZHOU A, et al. UAV-YOLO: Small object detection on unmanned aerial vehicle perspective[J]. Sensors, 2020, 20(8): 2238.
- [13] LI Y, LI Q, PAN J, et al. SOD-YOLO: Small-object-detection algorithm based on improved YOLOv8 for UAV images[J]. Remote Sensing, 2024, 16(16): 3057.
- [14] ZHAO X, ZHANG H, ZHANG W, et al. MSUD-YOLO: a novel multiscale small object detection model for UAV aerial images[J]. Drones, 2025, 9(6): 429.
- [15] LIU S, SHAO F, CHU W, et al. LCM-YOLO: a small object detection method for UAV imagery based on YOLOv5[J]. IET Image Processing, 2025, 19: e70051.
- [16] 胡惠娟, 秦一锋, 徐鹤, 等. 面向无人机航拍图像的 YOLOv8 目标检测改进算法[J]. 计算机科学, 2025, 52(4): 202-211.
HU H J, QIN Y F, XU H, et al. An improved YOLOv8 object detection algorithm for UAV aerial images[J]. Computer Science, 2025, 52(4): 202-211.
- [17] 王胜科, 任鹏飞, 吕昕, 等. 基于中心点和双重注意力机制的无人机高分辨率图像小目标检测算法[J]. 应用科学学报, 2021, 39(4): 650-659.
WANG S K, REN P F, LV X, et al. Small object detection algorithm for UAV high-resolution images based on center points and dual attention mechanism[J]. Journal of Applied Sciences, 2021, 39(4): 650-659.
- [18] 张伟, 庄幸涛, 王雪力, 等. DS-YOLO: 一种部署在无人机终端上的小目标实时检测算法[J]. 南京邮电大学学报(自然科学版), 2021, 41(1): 86-98.
ZHANG W, ZHUANG X T, WANG X L, et al. DS-YOLO: A real-time small object detection algorithm deployed on UAV terminals [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2021, 41(1): 86-98.
- [19] 张寅, 朱桂熠, 施天俊, 等. 基于特征融合与注意力的遥感图像

- 小目标检测[J]. 光学学报, 2022, 42(24): 2415001.
- ZHANG Y, ZHU G Y, SHI T J, et al. Small object detection in remote sensing images based on feature fusion and attention[J]. Acta Optica Sinica, 2022, 42(24): 2415001.
- [20] HUA W, CHEN Q. A survey of small object detection based on deep learning in aerial images[J]. Artificial Intelligence Review, 2025, 58: 162.
- [21] RABBI J, RAY N, SCHUBERT M, et al. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network[J]. Remote Sensing, 2020, 12(9): 1432.
- [22] MAHAUR B, MISHRA K K. Small-object detection based on YOLOv5 in autonomous driving systems[J]. Pattern Recognition Letters, 2023, 168: 115-122.
- [23] YUAN Z, GONG J, GUO B, et al. Small object detection in UAV remote sensing images based on intra-group multi-scale fusion attention and adaptive weighted feature fusion mechanism[J]. Remote Sensing, 2024, 16(22): 4265.
- [24] LEE Y W, KIM B G. Attention-based scale sequence network for small object detection[J]. Heliyon, 2024, 10(12): e32931.
- [25] 郭磊, 王邱龙, 薛伟, 等. 基于改进 YOLOv5 的小目标检测算法[J]. 电子科技大学学报, 2022, 51(2): 251-258.
- GUO L, WANG Q L, XUE W, et al. Small object detection algorithm based on improved YOLOv5[J]. Journal of University of Electronic Science and Technology of China, 2022, 51(2): 251-258.
- [26] 杨海舟, 李丹. 基于改进 SPPNet 的 YOLOv4 目标检测[J]. 电子制作, 2021(22): 52-54.
- YANG H Z, LI D. YOLOv4 object detection based on improved SPPNet[J]. Practical Electronics, 2021(22): 52-54.
- [27] 常青, 韩文, 王清华, 等. 改进 YOLO 轻量化网络的行人检测算法[J]. 光学技术, 2022, 48(1): 80-85.
- CHANG Q, HAN W, WANG Q H, et al. Pedestrian detection algorithm based on improved lightweight YOLO network[J]. Optical Technique, 2022, 48(1): 80-85.
- [28] MA Z, ZHOU L, WU D, et al. A small object detection method with context information for high altitude images[J]. Pattern Recognition Letters, 2025, 188: 22-28.
- [29] MA Y C, MA X, HAO T R, et al. Knowledge distillation via hierarchical matching for small object detection[J]. Journal of Computer Science and Technology, 2024, 39(4): 798-810.
- [30] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2024, 37: 107984-108011.
- [31] TIAN Y, YE Q, DOERMANN D. YOLOv12: attention-centric real-time object detectors[C]//Advances in Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2025, 38: 78433-78457.
- [32] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 16965-16974.
- [33] TANG S, ZHANG S, FANG Y. HIC-YOLOv5: improved YOLOv5 for small object detection[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2024: 6614-6619.
- [34] ZHANG Z X. Drone-YOLO: an efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8): 526.
- [35] HAN Y, WANG C, LUO H, et al. LRDS-YOLO enhances small object detection in UAV aerial images with a lightweight and efficient design[J]. Scientific Reports, 2025, 15: 22627.
- [36] WANG J, HUANG Z, DONG Y, et al. LE-YOLO: a lightweight and effective YOLO model for remote sensing image object detection[J]. Journal of Real-Time Image Processing, 2026, 23: 27.