

GDE-YOLO: 一种用于无人机图像的轻量级高精度小目标检测器

李 森¹, 硕良勋^{1,2,3*}

¹河北地质大学信息工程学院, 河北 石家庄

²智能传感物联网技术河北省工程研究中心, 河北 石家庄

³河北省地质环境智能感知与数据处理重点实验室, 河北 石家庄

收稿日期: 2026年6月19日; 录用日期: 2026年7月17日; 发布日期: 2026年7月24日

摘 要

针对无人机遥感图像中目标尺度小、背景复杂、密集遮挡导致检测精度低的问题, 提出一种基于YOLOv11s改进的小目标检测算法GDE-YOLO。该算法引入C3k2_StarCross模块增强小目标边缘与多尺度特征表达, 设计动态选择频率融合模块(DSFA)抑制复杂背景噪声, 采用尺度动态自适应融合(SDAF)提高跨尺度特征融合与空间对齐能力, 并通过动态因果跨门控检测头(DCPA_Head)提升密集目标定位精度。在VisDrone2019数据集上进行实验, 结果表明, GDE-YOLO在VisDrone2019数据集上的mAP50达到44.9%, 较YOLOv11s提高8.0个百分点模型参数量为5.39 M, 具有较好的检测精度和应用价值。

关键词

无人机遥感图像, 小目标检测, YOLOv11s, 特征融合, VisDrone2019

GDE-YOLO: A Lightweight and High-Precision Small Object Detector for UAV Imagery

Miao Li¹, Liangxun Shuo^{1,2,3*}

¹Hebei GEO University, Shijiazhuang Hebei

²Intelligent Sensor Network Engineering Research Center of Hebei, Shijiazhuang Hebei

³Hebei Province Key Laboratory of Intelligent Sensing and Data Processing for Geo-Environment, Shijiazhuang Hebei

Received: June 19, 2026; accepted: July 17, 2026; published: July 24, 2026

*通讯作者。

文章引用: 李森, 硕良勋. GDE-YOLO: 一种用于无人机图像的轻量级高精度小目标检测器[J]. 计算机科学与应用, 2026, 16(7): 140-153. DOI: 10.12677/csa.2026.167248

Abstract

To address the low detection accuracy caused by small object scales, complex backgrounds, dense distributions, and occlusions in UAV remote sensing images, this paper proposes GDE-YOLO, an improved small object detection algorithm based on YOLOv11s. The proposed algorithm introduces the C3k2_StarCross module to enhance edge information and multi-scale feature representation for small objects. A Dynamic Selection Frequency Fusion Module (DSFA) is designed to suppress complex background noise. In addition, Scale Dynamic Adaptive Fusion (SDAF) is adopted to improve cross-scale feature fusion and spatial alignment. Furthermore, Dynamic Causal Prism Adapter Head (DCPA_Head) is employed to enhance localization accuracy for densely distributed objects. Experiments conducted on the VisDrone2019 dataset demonstrate that GDE-YOLO achieves an mAP50 of 44.9%, which is 8.0 percentage points higher than that of YOLOv11s. Meanwhile, the model contains only 5.39 M parameters, indicating good detection accuracy and practical application value.

Keywords

UAV Remote Sensing Images, Small Object Detection, YOLOv11s, Feature Fusion, VisDrone2019

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着无人机(Unmanned Aerial Vehicle, UAV)技术的迅猛发展,低空遥感、灾害救援及交通巡检等领域的视觉感知需求日益增长。在这些应用场景中,目标检测技术扮演着不可或缺的核心角色。已有研究围绕无人机拍摄场景中的目标检测开展 YOLOv5s 改进实验,证明了 YOLO 系列模型在该类任务中的应用价值[1]。在 YOLOv11s 无人机航拍小目标检测方面,已有研究通过改进骨干网络与特征融合结构提升检测精度[2]。SMA-YOLO 面向无人机遥感图像中的多尺度小目标检测进行了优化[3]。RLRD-YOLO 则从无人机视角改进了 YOLOv8 小目标检测性能[4]。此外,YOLOv5 也已被用于车牌识别等一般视觉检测系统,说明 YOLO 系列方法具有较好的工程适配性[5]。然而,与常规地面平视视角的图像不同,无人机航拍图像通常具有拍摄距离远、视场广阔等特点,导致图像中的感兴趣目标(如行人、车辆)呈现出极端的微观尺度,通常占据不足 16×16 像素。这种微小目标往往缺乏清晰的纹理与结构信息,且极易被复杂多变的地物背景(如植被、阴影、光照突变)所淹没,导致现有目标检测算法频繁出现严重的漏检与误检现象。因此,设计一种兼顾边缘设备轻量化部署与高精度感知的无人机小目标检测算法,是当前计算机视觉领域面临的重大挑战。

目前,以 YOLO (You Only Look Once)系列为代表的单阶段目标检测算法,因其卓越的推理速度与均衡的精度,在工业界得到了广泛应用。YOLO 首次将目标检测建模为端到端的单阶段回归问题,显著提升了检测速度[6]。YOLOv3 通过多尺度预测结构进一步增强了不同尺度目标的检测能力[7]。YOLOv9 通过可编程梯度信息机制改进了网络学习过程[8]。YOLOv10 则进一步探索了实时端到端目标检测框架[9]。作为该系列的最新迭代版本,YOLOv11s 在通用目标检测任务中表现优异。然而,其原生网络架构主要针对通用尺度目标设计,直接迁移至无人机场景时存在明显的局限性:一方面,常规卷积在深层下采样

过程中会抹平微小目标的边缘细节; 另一方面, 传统特征金字塔难以有效过滤低频环境噪声, 导致分类与回归任务在极小尺度下出现空间错位。

为突破上述瓶颈, 本文在 YOLOv11s 基线模型的基础上, 提出了一种轻量化且高精度的网格细节增强小目标检测器——GDE-YOLO (Grid-Detail-Enhanced YOLO)。本文的主要贡献如下:

(1) 重构骨干网络以保护脆弱细节: 提出 C3k2_StarCross 模块, 通过将高频空间梯度与低频语义上下文显式解耦, 有效防止了微小目标形态细节在深层网络中的流失。

(2) 频域自适应融合与背景除噪: 引入动态选择频率融合(DSFA)模块, 利用渐进式大核注意力机制在频域内自适应过滤低频环境噪声, 提取纯净的多尺度特征。

(3) 消除极小目标的任务错配: 在特征颈部部署尺度动态自适应融合(SDAF)机制, 并搭配动态因果跨门控检测头(DCPA-Head), 实现了极小目标分类置信度与边界框回归的严格空间对齐。

实验表明, GDE-YOLO 仅需 5.39 M 的极低参数量, 便在 VisDrone2019、TinyPerson 及 AI-TOD 三大极具挑战性的无人机数据集上取得了优异的检测精度, 实现了算力开销与检测性能的完美平衡。

2. 相关理论研究

2.1. YOLO 系列目标检测算法

目标检测算法主要分为双阶段与单阶段两类。FasterR-CNN 通过候选区域机制提升了检测精度, 但推理流程相对复杂, 难以完全满足无人机边缘端实时检测需求[10]。由于无人机边缘计算平台(如 Jetson 系列)算力受限, 以 YOLO 为代表的单阶段算法因其端到端的推理时效性成为首选。最新的 YOLOv11 进一步优化了骨干网络中的 C3k2 等模块, 提升了特征提取效率。已有研究将线性注意力机制融入改进 YOLOv11, 用于提升无人机目标检测性能[11]。然而, YOLOv11 默认的各向同性卷积核在处理无人机航拍微小目标时, 极易将目标仅存的像素特征与大面积的背景语义混合, 导致特征表示能力急剧退化。

2.2. 无人机小目标检测优化策略

针对上述问题, 现有研究主要从检测头改进、注意力机制、跨尺度融合和轻量化结构等方面提升无人机小目标检测性能。TPH-YOLOv5 通过 Transformer 预测头增强了无人机拍摄场景下的目标检测能力[12]。TPH-YOLOv5++引入跨层非对称 Transformer, 进一步改善了密集无人机场景中的检测表现[13]。ALSS-YOLO 采用自适应轻量通道拆分与重排网络, 提高了 UAV 红外图像检测效率[14]。DAU-YOLO 面向 UAV 图像提出了轻量且有效的小目标检测结构[15]。

在特征融合与复杂场景适应方面, MFEL-YOLO 针对 UAV 航拍图像中的小目标进行了多层特征增强[16]。FFCA-YOLO 通过特征融合和上下文建模提升了遥感图像小目标检测效果[17]。YOLO-Remote 将 YOLO 框架适配到遥感目标检测任务中, 体现了 YOLO 系列在遥感目标识别中的拓展能力[18]。SSCW-YOLO 进一步探索了无人机场景下轻量化与高精度之间的平衡[19]。YOLO-RAW 则关注恶劣天气条件下 UAV 检测的鲁棒性问题[20]。

基于此, 本文设计的 GDE-YOLO 摒弃了计算昂贵的外部图像切片操作, 转而在网络内部进行密集的特征解耦与频域净化。在确保 5.39 M 轻量级参数量的前提下, 从根本上缓解了无人机视觉任务中轻量化与微小目标高精度捕获之间的固有矛盾。

3. 改进的 GDE-YOLO 算法

针对无人机航拍场景中微小目标特征易在深层网络中退化、且容易被复杂背景噪声淹没的问题, 本文在 YOLOv11s 的基础上提出了一种网格细节增强小目标检测算法(GDE-YOLO)。改进后的网络总体架

构如图 1 所示。该模型主要通过四大核心模块进行重构：1) 在骨干网络中引入 C3k2_StarCross 模块以保护高频边缘特征；2) 设计动态选择频率融合(DSFA)模块滤除环境噪声；3) 在特征颈部提出跨尺度自适应融合(SDAF)机制消除语义混叠；4) 采用动态检测头(DCPA_Head)缓解极小目标的分类与回归错配。

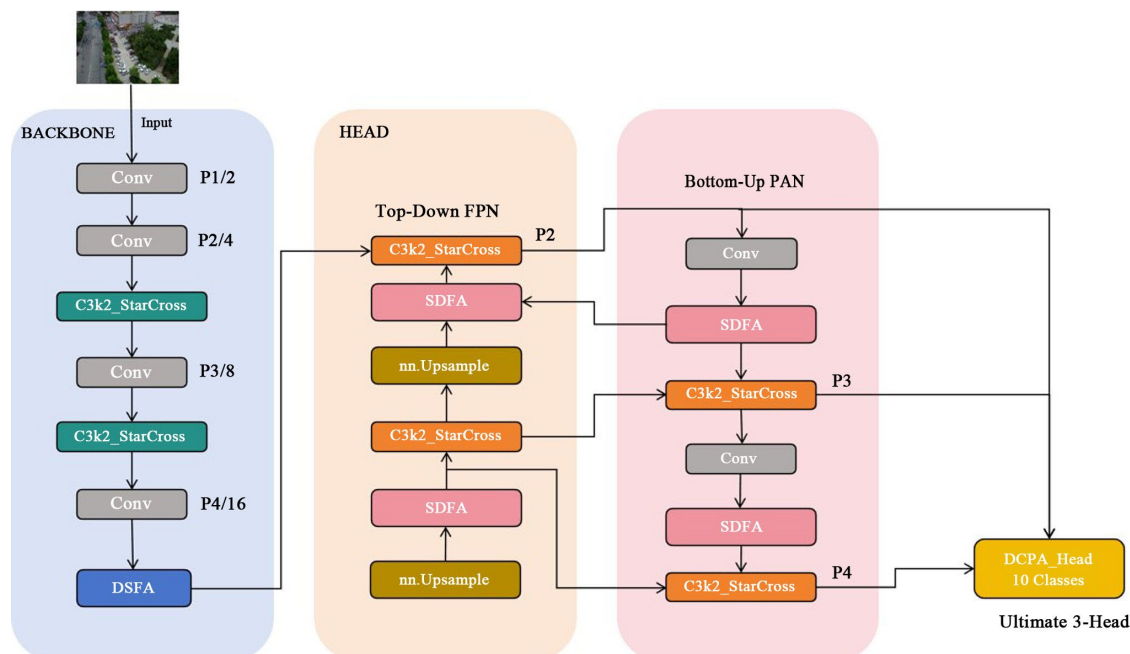


Figure 1. Overall network architecture of the proposed GDE-YOLO

图 1. 本文提出的 GDE-YOLO 总体网络结构

3.1. C3k2_StarCross 模块

传统深度卷积神经网络通常采用标准的 3×3 卷积核进行特征提取。这种各向同性的感受野在处理占据图像大部分像素的中大目标时非常有效，但对于无人机图像中仅有几个像素的行人或车辆，连续的标准卷积和池化操作极易导致目标特征在深层网络中被周围的背景语义所稀释甚至完全湮灭。

受 Res2Net 细粒度多尺度表示以及特征解耦思想的启发，本文重构了骨干网络的核心提取组件，提出了 C3k2_StarCross 模块。该模块的内部核心单元为 StarDrone_Bottleneck，其采用了“分而治之”的特征处理策略。BF-YOLO 通过改进 YOLOv8 提升了小目标检测效果，说明特征结构优化对小目标识别具有重要意义[21]。CF-YOLO 跨层特征融合框架强调了多层特征交互对小目标检测的促进作用[22]。另一项基于 YOLOv11 的 CF-YOLO 研究也表明，面向无人机图像的小目标检测需要充分利用跨层语义信息[23]。与上述方法相比，本文进一步强调边缘细节保护与频域背景抑制的协同作用。

对于输入特征图 $X \in \mathbb{R}^{C \times H \times W}$ ，首先经过 1×1 逐点卷积进行通道降维与映射，随后在通道维度上将其等分为两个子张量：边缘张量 $X_{\text{edge}} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 与上下文张量 $X_{\text{ctx}} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 。

(1) 高频边缘特征提取支路：

针对 X_{edge} ，本文设计了 RepCross 结构。考虑到无人机视角下的车辆和建筑物大多具有强烈的人造正交边界(水平与垂直边缘)，RepCross 摒弃了常规的方形卷积，转而采用 3×3 、 1×3 和 3×1 的非对称卷积并行处理机制。为了使网络能够根据目标形态自适应聚焦，引入了可学习参数权重 $W \in \mathbb{R}^3$ ，并通过 Softmax 函数进行归一化。该支路的前向计算可表示为：

$$F_{\text{edge}} = \sigma_{\text{SiLU}} \left(\mathcal{B} \left(\sum_{i=1}^3 \frac{\exp(w_i)}{\sum_{j=1}^3 \exp(w_j)} * \text{Conv}_i(X_{\text{edge}}) \right) \right) \quad (1)$$

其中, $\mathcal{B}(\cdot)$ 表示批量归一化(Batch Normalization), $\sigma_{\text{SiLU}}(\cdot)$ 为 SiLU 激活函数。这种非对称交叉结构以极少的参数代价, 显著提升了模型对微小目标几何边界的敏感度。

(2) 低频上下文特征提取支路:

针对 X_{ctx} , 为了在不增加庞大计算量(GFLOPs)的前提下获取足够大的感受野以感知周围环境上下文, 本文结合了局部分组卷积与空洞卷积。特征首先经过 5×5 的深度分组卷积获取局部语义, 随后叠加一个膨胀率(Dilation)为 2 的 5×5 空洞分组卷积。该过程可形式化为:

$$F_{\text{ctx}} = \mathcal{B} \left(\text{Conv}_{5 \times 5, g=\frac{C}{2}}(X_{\text{ctx}}) + \text{Conv}_{5 \times 5, d=2, g=\frac{C}{2}}(X_{\text{ctx}}) \right) \quad (2)$$

最终, 如图 2 所示, 将 F_{edge} 与 F_{ctx} 在通道维度上重新拼接, 通过 1×1 卷积进行跨通道信息交融, 并乘以缩放因子 γ (初始化为 0.1) 与原始输入 X 进行残差连接。这种结构不仅有效抑制了深层网络的梯度消失问题, 还完美保留了极小目标的网格级细节。

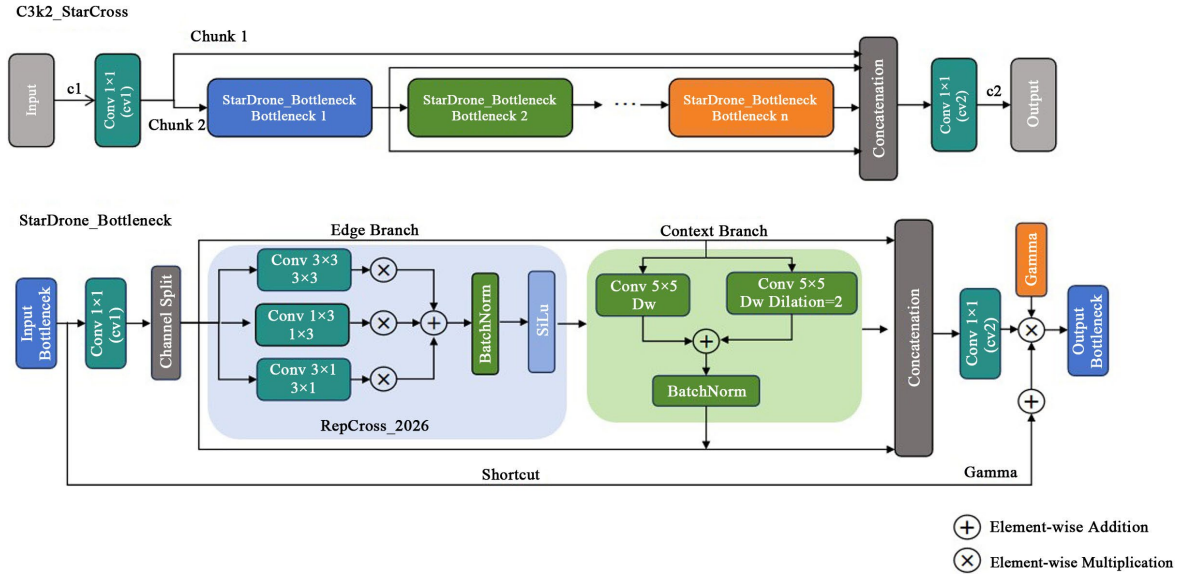


Figure 2. Structure of the C3k2_StarCross module

图 2. C3k2_StarCross 模块的结构

3.2. DSFA 动态空间特征聚合模块

YOLOv11 原生网络在骨干网络末端采用了空间金字塔池化(SPPF)模块, 通过级联的 5×5 最大池化来聚合多尺度特征。然而, 最大池化操作本质上会导致细粒度空间信息的不可逆丢失, 粗暴的降维会导致微小目标的细粒度特征产生严重的混叠。

为此, 本文提出了动态空间特征聚合模块(DSFA, Dynamic Spatial Feature Aggregation)模块, 具体如图 3 所示。DSFA 构建了一种纯粹基于数据驱动的多尺度特征选择机制。对于输入特征 F_{in} , 首先利用池化生成 4 个尺度的基础金字塔特征序列 $Y = \{y_0, y_1, y_2, y_3\}$ 。随后, 本文引入了渐进式大核注意力(LKA,

Large Kernel Attention)机制。为了捕捉从极微小目标到宏观复杂背景的宽泛上下文, 设计了四个并行的 LKA 分支, 分别采用感受野递减(膨胀率依次为 4, 3, 2, 1)的深度可分离卷积进行多尺度频率特征提取, 得到特征集 $F_{LKA} \in \mathbb{R}^{B \times 4 \times C \times H \times W}$ 。

为了自适应滤除如海浪、茂密树林等冗余背景噪声对检测结果的干扰, DSFA 设计了动态门控机制。通过沿通道空间维度对 F_{LKA} 求和并进行全局自适应平均池化(AdaptiveAvgPool), 生成具有全局视野的描述符, 再由两层全连接网络计算各感受野分支的动态注意力权重 $A \in \mathbb{R}^{B \times 4 \times C \times 1 \times 1}$:

$$A = \text{Softmax} \left(W_2 \delta \left(W_1 \cdot \text{GAP} \left(\sum_{i=1}^4 F_{LKA_i} \right) \right) \right) \quad (3)$$

其中, δ 为 ReLU 激活函数。接着, 利用该权重对 4 个尺度的空间特征进行加权聚合:

$$F_{\text{dyn}} = \sum_{i=1}^4 (F_{LKA_i} \odot A_i) \quad (4)$$

最后, 将其与经过全局池化和双线性插值上采样的全局正交语义分支 F_{global} 拼接, 实现了多尺度特征在空间与语义维度上的高效融合与纯化。

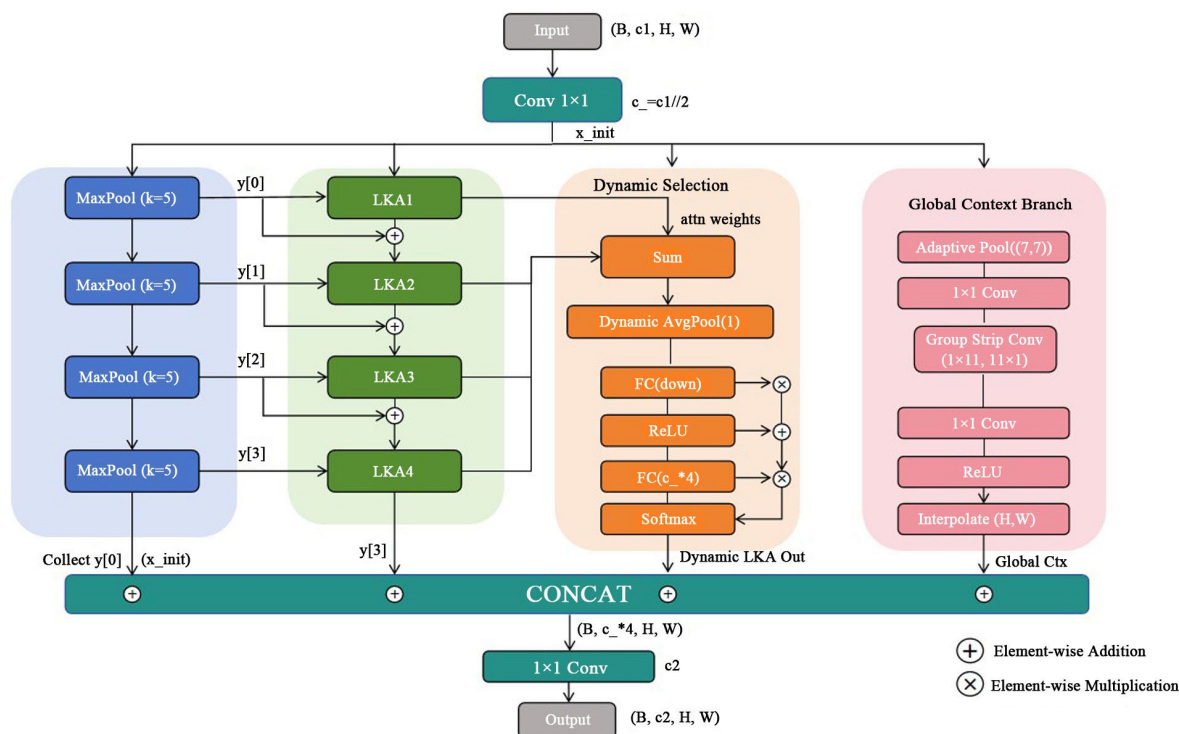


Figure 3. Structure of the proposed Dynamic Selection Frequency Fusion (DSFA) module

图 3. 本文提出的动态选择频率融合(DSFA)模块架构

3.3. SDAF 跨尺度自适应融合机制

在 Neck 层构建特征金字塔时, 传统的 PANet 或 FPN 架构通常采用最近邻上采样后直接将浅层特征与深层特征在通道上拼接(Concat)。然而, 由于网络不同层级的感受野差异巨大, 高分辨率的浅层特征(富含物理纹理)与低分辨率的深层特征(富含抽象语义)之间存在固有的空间不对齐(Spatial Misalignment), 强制拼接会引发严重的语义混叠现象。

为了根除这一问题, 本文提出了尺度动态自适应融合(SDAF)机制如图4所示。该机制由空间形变对齐(SDABlock)与全局-局部去噪融合(GLD-Fusion)两个级联阶段构成。

(1) SDABlock 亚像素空间对齐阶段:

该模块旨在利用流动特征(Flow Feature)与参考特征(Ref Feature)联合预测空间偏移量, 将特征强行扭曲至统一的坐标系下。设输入特征分别为 X_{flow} 与 X_{ref} , 网络首先通过卷积预测每个空间位置的 2D 偏移量 Δp 与调制掩码 Δm 。

特别需要指出的是, 考虑到无人机边缘计算平台(如 Jetson Nano)在进行混合精度(AMP)推理时极易因极端的像素形变引发底层寻址越界与内存崩溃, 本文在算法底层引入了截断保护机制。通过 `torch.clamp` 函数对偏移量 Δp 施加 $[-100.0, 100.0]$ 的硬性物理约束。经过安全保护的可形变卷积(Deformable Conv)对齐过程定义如下:

$$X_{\text{aligned}}(p_0) = \sum_{k=1}^K w_k \cdot X_{\text{flow}}(p_0 + p_k + \text{Clamp}(\Delta p_k, -100, 100)) \cdot \sigma(\Delta m_k) \quad (5)$$

(2) GLD-Fusion 全向感知去噪阶段:

在实现空间对齐后, GLD-Fusion 首先进行局部高频锐化。提取浅层特征的高频拓扑残差 $E_{\text{shallow}} = X_{\text{shallow}} - \text{AvgPool}(X_{\text{shallow}})$, 并通过可学习参数 α 将其注入深层特征中以增强目标轮廓; 同时, 利用 Sigmoid 通道门控网络对浅层特征本身进行背景除噪。

为了在特征融合后建立长程拓扑关联, GLD-Fusion 引入了硬件友好的全局线性自注意力。不同于传统 Transformer 中具有 $O((HW)^2)$ 复杂度的标准自注意力, 本文利用通道维度的点积特性, 将拼接后的特征映射为 Q, K, V , 其自注意力计算公式优化为:

$$F_{\text{global}} = \frac{Q}{\sqrt{d_k}} (\text{Softmax}(K^T V)) \quad (6)$$

由于矩阵乘法的结合律改变, 该模块的时间复杂度被成功降维至 $O(HW)$, 从而在满足无人机实时帧率(FPS)要求的前提下, 完成了跨尺度的深度融合。

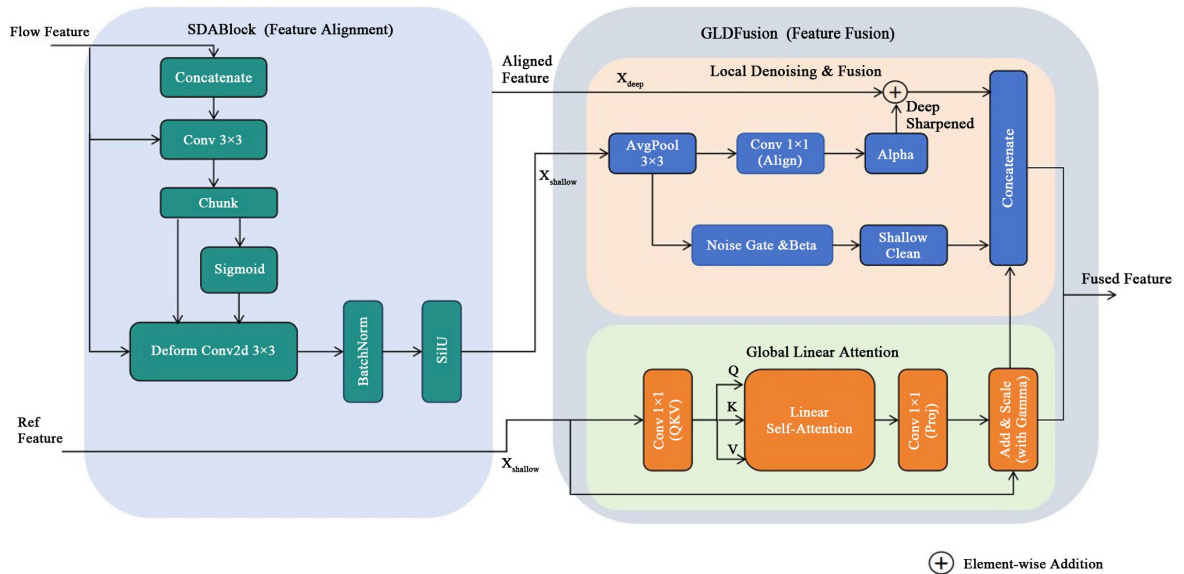


Figure 4. Structure of the proposed Scale-Dynamic Adaptive Fusion (SDAF) mechanism

图4. 本文提出的尺度动态自适应融合(SDAF)机制的结构

3.4. DCPA_Head 动态任务交互门控检测头

目标检测领域存在一个固有的矛盾: 分类任务需要特征具备平移不变性(Translation Invariance), 而边界框回归任务需要极高的平移可变性(Translation Variance)。在无人机视角密集的微小目标场景下, 传统 YOLO 检测头将双任务独立并行处理, 极易出现“高分置信度但定位偏差极大”的空间错配现象。

为此, 本文设计了动态任务交互门控检测头(DCPA_Head, Dynamic Cross-gated Task Adapter Head), 其结构如图 5 所示。该模块在将特征解耦为回归分支 X_{reg} 与分类分支 X_{cls} 后, 首先使两路特征分别通过基于大核注意力融合机制的 Dual_LKA 适配器, 以建立更广范围的空间上下文依赖, 进一步扩大目标所在区域的感受野。

随后, 模型引入了双向任务交互门控机制(RFD_InteractionPrism), 进行深层的跨任务信息协同与纠偏: 第一步, 进行语义引导定位。利用包含丰富类别信息的分类特征 X_{cls} , 经过批归一化与 Sigmoid 激活生成空间掩码 $Mask_{cls}$, 将其点乘作用于回归分支, 从而引导边界框的回归过程显式聚焦于网络认定的前景实体区域, 其公式为:

$$X_{reg}^{aligned} = X_{reg} + \max(\gamma, 0.005) \cdot (X_{reg} \odot Mask_{cls}) \quad (7)$$

第二步, 进行边界细化分类。利用已经对齐后的高精度回归张量 $X_{reg}^{aligned}$ 反向生成掩码 $Mask_{reg}$, 作用于分类分支, 以此有效惩罚和抑制那些超出了目标形态物理边界的分类预测像素。

通过这种底层特征张量级别的细粒度任务协同与空间对齐, DCPA_Head 有效实现了“判断目标是什么”与“定位目标在哪里”的高度空间同步, 最大化地降低了复杂背景下的假阳性(False Positive)误检率。

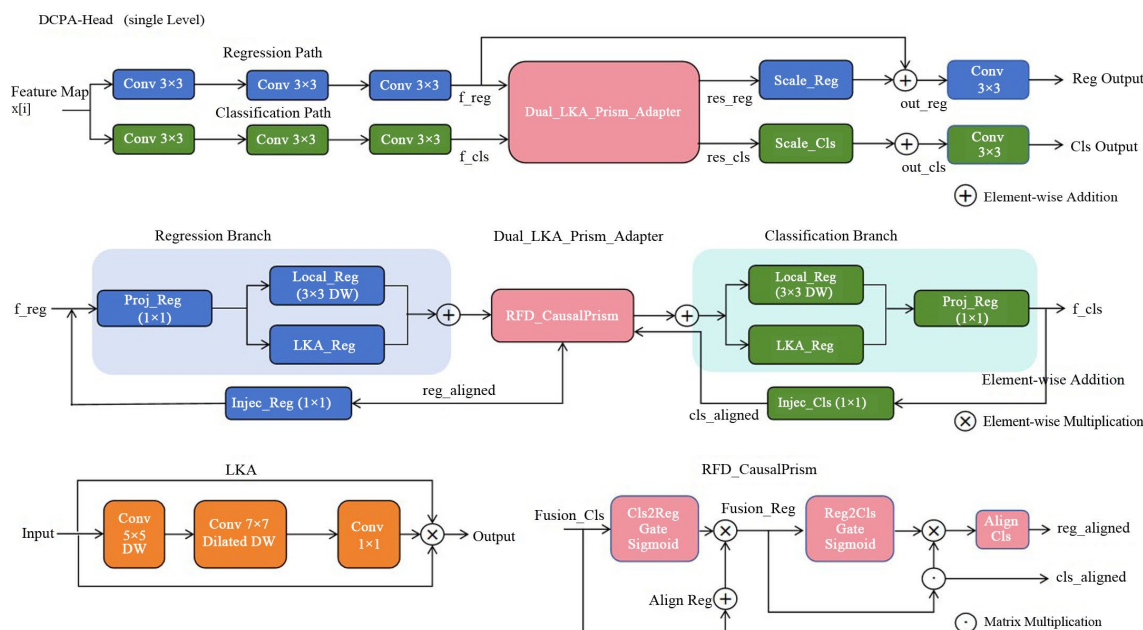


Figure 5. Structure of the proposed Dynamic Causal Prism Attention Head (DCPA_Head)

图 5. 本文提出的动态因果跨门控检测头(DCPA_Head)架构

4. 实验设计

4.1. 实验环境及参数设置

本文实验基于 Ultralytics YOLO 框架实现, 深度学习框架为 PyTorch, 实验设备采用 NVIDIA Tesla

V100 GPU。为保证实验结果的稳定性, 训练过程中固定随机种子为 42, 并保持各模型的主要训练参数一致。

实验采用 VisDrone2019 数据集进行训练和验证, 输入图像尺寸统一设置为 640×640 , batch size 设置为 8, 优化器采用 AdamW, 初始学习率设置为 0.002, 权重衰减系数设置为 0.0005。训练过程中采用 Mosaic 数据增强, 并在训练后期关闭 Mosaic, 以提高模型收敛稳定性。

本文对比实验训练轮数设置为 300 epoch, 用于充分比较不同检测算法在 VisDrone2019 数据集上的检测性能; 消融实验训练轮数设置为 70 epoch, 用于验证各改进模块对模型性能的影响。除训练轮数和模型结构不同外, 其余主要训练参数保持一致, 以保证实验结果的公平性。

4.2. 数据集及评价指标

本研究主要依托权威的无人机航拍目标检测数据集 VisDrone2019 展开核心实验。该数据集由各类无人机平台在不同高度、不同光照条件下捕获, 场景涵盖了中国 14 个不同城市的城市与乡村环境。本实验严格遵循官方的数据划分协议, 采用 VisDrone2019-DET-train 包含的 6471 张图像作为训练集, 采用 VisDrone2019-DET-val 包含的 548 张图像作为验证集。

数据集中共包含 10 个常见的前景目标类别。如图 6 所示, 该数据集的类别实例数量呈现出显著的长尾分布(Long-tail Distribution)与极端的尺度变异特性。其中, 行人(pedestrian)与轿车(car)的实例数量占据绝对主导地位, 且在图像中往往以高度密集的微小像素簇群形式存在, 这对检测器的感知下限提出了极高的挑战。因此, VisDrone2019 数据集能够较为全面地反映无人机航拍场景下小目标检测所面临的尺度差异、背景干扰、密集分布与遮挡等关键难题, 为验证所提算法的有效性与鲁棒性提供了具有代表性的实验基础。同时, 该数据集在目标类别分布、成像视角、环境复杂度和目标密集程度等方面具有较强的多样性, 可有效检验模型在真实无人机遥感场景中的泛化能力。因此, 基于该数据集开展实验, 能够更有针对性地评估 GDE-YOLO 在复杂背景、小尺度目标和密集遮挡场景下的检测性能。

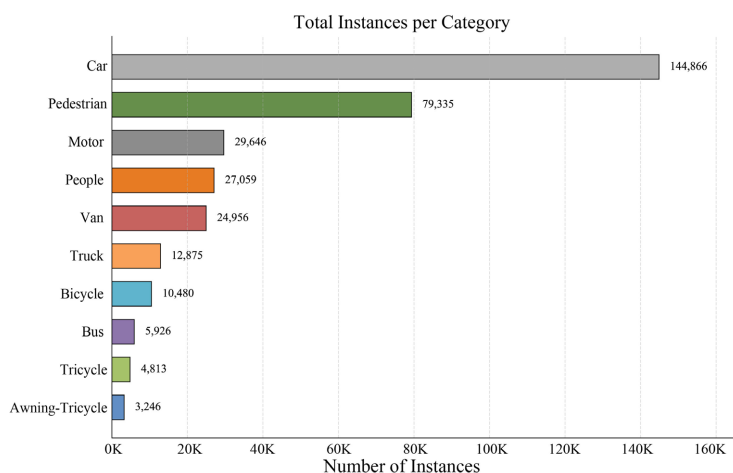


Figure 6. Representative samples and target distribution of the VisDrone2019 dataset
图 6. VisDrone2019 数据集的代表性样本与目标分布

5. 结果分析

5.1. 消融实验

为了验证各改进模块对模型性能的影响, 本文在 VisDrone2019 数据集上进行消融实验, 实验结果见

表 1. 需要说明的是，对比实验的 Baseline 为 YOLOv11s，消融实验中的 Baseline 并非原始 YOLOv11s，而是在 YOLOv11s 基础上去掉 P5 层并添加 P2 层后的模型。该设置能够更好地适应无人机图像中小目标占比较高的特点，并作为后续模块消融实验的基础模型。

由表 1 可知，Baseline 的 mAP50 为 37.6%，mAP50:95 为 21.2%。在单独加入 C3k2_StarCross 模块后，mAP50 提升至 38.7%，mAP50:95 提升至 21.7%，说明该模块能够增强主干网络对小目标结构特征的提取能力。单独加入 DSFA 模块后，mAP50 达到 38.2%，mAP50:95 达到 21.5%，表明动态特征选择与融合对复杂背景下的目标检测具有一定促进作用。单独加入 SDAF 模块后，mAP50 提升至 40.6%，mAP50:95 提升至 23.2%，提升幅度最为明显，说明跨尺度特征融合和空间对齐对无人机小目标检测具有重要作用。单独加入 DCPA-Head 后，mAP50 达到 39.5%，mAP50:95 达到 22.1%，说明改进检测头能够提高模型对密集目标的定位和分类能力。

从多模块组合结果来看，C3k2_StarCross、DSFA、SDAF 和 DCPA-Head 之间具有较好的互补性。当 SDAF 与 DCPA-Head 同时加入时，mAP50 达到 41.9%，mAP50:95 达到 23.9%；当 C3k2_StarCross、DSFA 和 SDAF 同时加入时，mAP50 达到 42.3%，mAP50:95 达到 24.3%。最终，完整模型 GDE-YOLO 的 mAP50 达到 42.4%，mAP50:95 达到 24.4%，相比 Baseline 分别提高 4.8 个百分点和 3.2 个百分点。实验结果表明，本文所提出的各改进模块均能提升模型检测性能，完整模型取得了最优的综合检测效果。

Table 1. Component-wise ablation analysis of the proposed modules on the VisDrone2019 dataset
表 1. 本文提出模块在 VisDrone2019 数据集上的逐项消融实验分析

C3k2_StarCross	DSFA	SDAF	DCPA-Head	mAP50	mAP50:90	AP ^s	AP ^m	AP ^l	AR ^s	AR ^m	AR ^l
				37.6	21.2	12.7	31.2	39.1	36.1	58.2	58.2
√				38.7	21.7	13.3	31.3	40.1	36.8	57.8	61.1
	√			38.2	21.5	13.2	31.2	38.9	36.1	58.1	61.7
		√		40.6	23.2	14.5	33.4	39.3	37.6	59.0	63.6
			√	39.5	22.1	13.7	32.5	39.6	36.6	58.7	58.3
√	√			39.2	21.8	13.2	31.6	38.9	36.5	58.7	59.2
√		√		40.7	22.8	14.6	32.8	39.6	37.6	58.7	58.0
√			√	39.7	22.2	13.8	31.7	40.7	37.1	58.1	60.2
	√	√		40.6	23.3	14.6	33.9	41.4	36.9	59.4	62.9
	√		√	39.6	22.4	13.7	32.6	39.1	36.5	59.1	60.3
		√	√	41.9	23.9	15.6	34.4	42.0	38.7	60.3	62.8
√	√	√		42.3	24.3	14.9	35.5	43.7	38.7	59.8	61.6
√	√		√	39.8	22.4	13.9	32.5	39.8	37.3	59.4	60.3
√		√	√	42.1	24.0	15.3	34.3	43.8	38.6	60.5	64.5
	√	√	√	42.3	24.3	14.8	35.1	44.4	38.5	60.4	65.1
√	√	√	√	42.4	24.4	15.8	34.8	44.1	38.6	59.5	63.2

5.2. 对比实验

为了进一步验证本文算法的有效性，将 GDE-YOLO 与多种主流目标检测算法在 VisDrone2019 数据

集上进行对比, 结果见表 2。需要说明的是, 对比实验中的 YOLOv11s 为原始官方 YOLOv11s 模型, 与消融实验中的 Baseline 设置不同。由表 2 可知, 原始 YOLOv11s 的 mAP 为 21.2%, mAP50 为 36.9%, mAP75 为 20.0%; 本文模型的 mAP 达到 26.0%, mAP50 达到 44.9%, mAP75 达到 25.6%, 相比原始 YOLOv11s 分别提高 5.3、8.3 和 5.6 个百分点, 说明本文改进算法能够显著提升 YOLOv11s 在无人机小目标检测任务中的性能。

与其他 YOLO 系列模型相比, 本文模型同样取得了较好的检测效果。相比 YOLOv5-S、YOLOv8-S、YOLOv9-S 和 YOLOv10-S, 本文模型的 mAP50 均有明显提升; 相比参数量更大的 YOLOv11-M, 本文模型的 mAP50 提高 2.9 个百分点, 同时参数量由 20.06 M 降低至 5.39 M, 模型规模更小。与 RT-DETR 相比, 本文模型在 mAP、mAP50 和 mAP75 上均取得更优结果, 说明 GDE-YOLO 更适合 VisDrone2019 这种目标密集、小目标占比较高的无人机检测场景。

从模型规模来看, 本文模型参数量为 5.39 M, 低于 YOLOv11s 的 9.43 M, 也明显低于 YOLOv11-M、YOLOv9-M 和 RT-DETR 等模型, 说明本文模型具有较好的轻量化优势。虽然模型 GFLOPs 为 55.6, 相比原始 YOLOv11s 有所增加, 但其检测精度提升明显, 且参数量较低, 能够在精度和模型规模之间取得较好的平衡。综上所述, GDE-YOLO 在 VisDrone2019 数据集上具有较好的检测精度和轻量化特性, 能够有效改善无人机遥感图像中小目标漏检、误检和定位不准确的问题。

Table 2. Comparison of experimental metrics of different SOTA detectors on the VisDrone2019 dataset

表 2. 不同 SOTA 检测器在 VisDrone2019 数据集上的实验指标对比

Model	Params (M)	GFLOPs	mAP50:90	mAP50	mAP75	AP ^s	AP ^m	AP ^l	FPS
YOLOv5-S	9.13	24.1	21.1	37.5	20.0	11.4	32.4	43.9	122.2
YOLOv5-M	25.07	64.4	23.8	41.0	23.0	13.4	36.5	45.9	78.2
YOLOv8-S	11.14	28.7	21.2	37.4	20.4	12.0	31.9	44.4	156.7
YOLOv8-M	25.86	79.1	23.8	40.9	23.0	13.8	35.7	46.9	75.6
YOLOv9-S	7.29	27.4	21.3	37.7	20.6	11.5	33.4	41.5	57.7
YOLOv9-M	20.17	77.6	24.3	41.7	23.7	13.9	36.9	48.0	66.5
YOLOv10-S	8.07	24.8	22.3	38.4	20.0	11.5	31.5	42.5	113.5
YOLOv10-M	16.50	64.0	24.8	41.7	23.2	13.4	35.6	42.4	91.2
YOLOv11-S	9.43	21.6	21.2	36.9	21.9	11.4	32.0	42.2	104.3
YOLOv11-M	20.06	68.2	24.8	42.5	24.2	15.0	36.7	47.9	75.4
RTDETR	32.82	108.0	19.5	35.2	18.8	12.8	27.9	30.0	19.1

5.3. 可视化分析

如图 7~图 9 所示, 在第一列图像中, 基线模型对画面右下角阴影区域内密集的行人与非机动车出现了大面积的漏检, 而加入 C3k2_StarCross 与 SDAF 模块的模型则精准地捕获了这些微小目标, 未产生明显的遗漏。在第二列图像中, 基线模型漏检了画面顶端远景区域且受复杂背景干扰的微小车辆, 引入 C3k2_StarCross 和 SDAF 的模型则能准确识别并将这些边缘目标召回。上述对比结果表明, C3k2_StarCross 和 SDAF 模块有效增强了网络对高密度小目标的特征辨别能力, 显著减少了无人机复杂视角和阴影遮挡下的漏检与误检现象。



Figure 7. Detection results of the baseline model

图 7. 基线模型检测结果

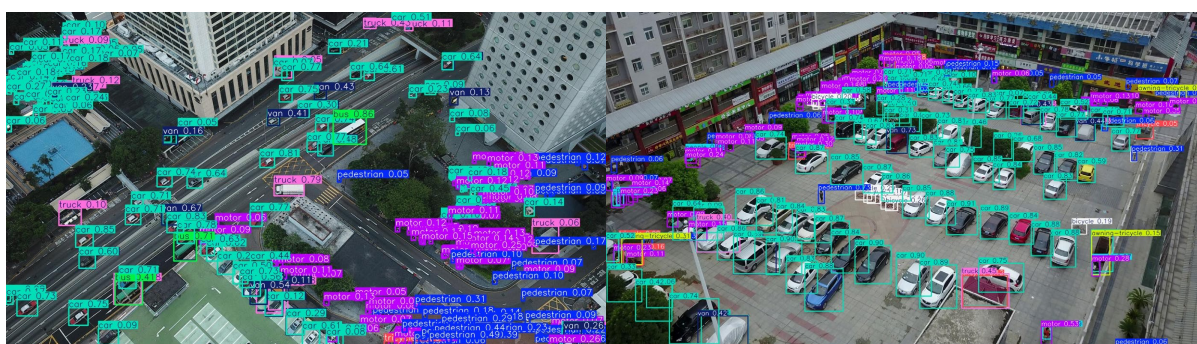


Figure 8. Detection results with C3k2_StarCross

图 8. 加入 C3k2_StarCorss 后的检测结果



Figure 9. Detection results with SDAF

图 9. 加入 SDAF 后的检测结果

6. 结论

本文针对无人机遥感图像中目标尺度小、背景复杂、目标密集以及遮挡严重导致检测精度不足的问题,提出一种基于改进 YOLOv11s 的小目标检测算法 GDE-YOLO。该算法以 YOLOv11s 为基础,引入 C3k2_StarCross 模块增强小目标边缘结构和多尺度特征表达能力;设计 DSFA 模块,提高复杂背景下有效特征的选择能力;采用 SDAF 模块增强跨尺度特征融合与空间对齐效果;同时构建 DCPA-Head 检测头,提高密集小目标场景下的分类和定位能力。

在 VisDrone2019 数据集上的实验结果表明,本文算法的 mAP50 达到 44.9%,mAP50:95 达到 26.0%,相比原始 YOLOv11s 分别提高 8.0 和 4.8 个百分点,说明改进模型能够有效提升无人机遥感图像小目标

检测精度。消融实验结果也表明, 各改进模块均能在不同程度上改善模型检测性能, 证明了本文算法设计的有效性。

与其他主流目标检测算法相比, GDE-YOLO 在检测精度和模型参数量之间取得了较好的平衡, 能够减少复杂无人机场景中的漏检和误检现象。但本文模型的计算量仍有进一步优化空间。后续研究将继续从网络轻量化、特征融合结构优化和边缘设备部署等方面进行改进, 以进一步提升模型在实际无人机应用场景中的检测效率和部署能力。

参考文献

- [1] 李新晨, 李晓飞. 基于 YOLOv5s 的无人机拍摄场景下的目标检测[J]. 智能计算机与应用, 2025, 15(4): 151-157.
- [2] 李云龙. 基于改进 YOLOv11s 的无人机航拍图像小目标检测模型[J]. 计算机科学与应用, 2026, 16(2): 522-531.
- [3] Zhou, S., Zhou, H. and Qian, L. (2025) A Multi-Scale Small Object Detection Algorithm SMA-YOLO for UAV Remote Sensing Images. *Scientific Reports*, **15**, Article No. 9255. <https://doi.org/10.1038/s41598-025-92344-7>
- [4] Li, H., Li, Y., Xiao, L., Zhang, Y., Cao, L. and Wu, D. (2025) RLRO-YOLO: An Improved YOLOv8 Algorithm for Small Object Detection from an Unmanned Aerial Vehicle (UAV) Perspective. *Drones*, **9**, Article 293. <https://doi.org/10.3390/drones9040293>
- [5] 李世伟. 基于 YOLOv5 算法的目标检测与车牌识别系统[J]. 电子技术与软件工程, 2022(1): 138-141.
- [6] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 79-788. <https://doi.org/10.1109/cvpr.2016.91>
- [7] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement.
- [8] Wang, C.Y., Yeh, I.H. and Liao, H.Y.M. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In: *Lecture Notes in Computer Science*, Springer, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1
- [9] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., et al. (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems*, **37**, 107984-108011.
- [10] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [11] Zhou, S., Yang, L., Liu, H., Zhou, C., Liu, J., Zhao, S., et al. (2025) A Lightweight Drone Detection Method Integrated into a Linear Attention Mechanism Based on Improved YOLOv11. *Remote Sensing*, **17**, Article 705. <https://doi.org/10.3390/rs17040705>
- [12] Zhu, X., Lyu, S., Wang, X. and Zhao, Q. (2021) TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, 11-17 October 2021, 2778-2788. <https://doi.org/10.1109/iccvw54120.2021.00312>
- [13] Zhao, Q., Liu, B., Lyu, S., Wang, C. and Zhang, H. (2023) TPH-YOLOv5++: Boosting Object Detection on Drone-Captured Scenarios with Cross-Layer Asymmetric Transformer. *Remote Sensing*, **15**, Article 1687. <https://doi.org/10.3390/rs15061687>
- [14] He, A., Li, X., Wu, X., Su, C., Chen, J., Xu, S., et al. (2024) ALSS-YOLO: An Adaptive Lightweight Channel Split and Shuffling Network for TIR Wildlife Detection in UAV Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **17**, 17308-17326. <https://doi.org/10.1109/jstars.2024.3461172>
- [15] Wan, Z., Lan, Y., Xu, Z., Shang, K. and Zhang, F. (2025) DAU-YOLO: A Lightweight and Effective Method for Small Object Detection in UAV Images. *Remote Sensing*, **17**, Article 1768. <https://doi.org/10.3390/rs17101768>
- [16] Hou, T., Leng, C., Wang, J., Pei, Z., Peng, J., Cheng, I., et al. (2025) MFEL-YOLO for Small Object Detection in UAV Aerial Images. *Expert Systems with Applications*, **291**, Article 128459. <https://doi.org/10.1016/j.eswa.2025.128459>
- [17] Zhang, Y., Ye, M., Zhu, G., Liu, Y., Guo, P. and Yan, J. (2024) FFCA-YOLO for Small Object Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, **62**, 1-15. <https://doi.org/10.1109/tgrs.2024.3363057>
- [18] Fan, K., Li, Q., Li, Q., Zhong, G., Chu, Y., Le, Z., et al. (2024) YOLO-Remote: An Object Detection Algorithm for Remote Sensing Targets. *IEEE Access*, **12**, 155654-155665. <https://doi.org/10.1109/access.2024.3479320>
- [19] He, Z., She, R., Tan, B., Li, J. and Lei, X. (2026) SSCW-YOLO: A Lightweight and High-Precision Model for Small

-
- Object Detection in UAV Scenarios. *Drones*, **10**, Article 41. <https://doi.org/10.3390/drones10010041>
- [20] Munir, A., Siddiqui, A.J., Hossain, M.S. and El-Maleh, A. (2025) YOLO-RAW: Advancing UAV Detection with Robustness to Adverse Weather Conditions. *IEEE Transactions on Intelligent Transportation Systems*, **26**, 7857-7873. <https://doi.org/10.1109/tits.2025.3560792>
- [21] 蒲小莉, 赖惠成, 高古学. BF-YOLO: 基于 YOLOv8 改进的小目标检测算法[J]. 计算机工程与科学, 2025, 47(8): 1425-1434.
- [22] Wei, L., Afzal, I. and Guo, J. (2025) CF-YOLO: A Cross-Layer Feature Fusion YOLO Framework for Small Object Detection. *Journal of Electronic Imaging*, **34**, Article 053006. <https://doi.org/10.1117/1.jei.34.5.053006>
- [23] Wang, C., Han, Y., Yang, C., Wu, M., Chen, Z., Yun, L., *et al.* (2025) CF-YOLO for Small Target Detection in Drone Imagery Based on YOLOv11 Algorithm. *Scientific Reports*, **15**, Article No. 16741. <https://doi.org/10.1038/s41598-025-99634-0>