DOI:10.19651/j.cnki.emt.2518168

改进 YOLOv11 的无人机航拍图像检测算法*

李 珺 丁彬彬 史维娟 杨 琳
(兰州交通大学电子与信息工程学院 兰州 730000)

摘 要:针对无人机航拍图像检测任务中,存在目标尺寸微小且背景环境复杂,往往会导致漏检和误检的问题,本文提出了一种基于 YOLOv11 的航拍图像小目标检测算法 WT-YOLO。首先,考虑到无人机航拍图像普遍为小目标的问题,调整了 YOLOv11 颈部网络的结构,改变了输出特征图的尺寸,提高了算法对小目标的检测能力。其次,结合 WTConv,重新设计了 Bottleneck 和 C3k2 模块的结构,命名为 C3k2-WT,来实现特征的高效提取。再次,引入 Focal-Modulation 来替换 SPPF,通过在不同的空间尺度上聚焦和调制特征,使得模型在处理复杂场景时更具鲁棒性;最后,设计了共享卷积检测头,通过卷积共享机制,减少了模型的参数量,同时增强了特征图之间的全局信息融合能力。改进后的算法在 VisDrone2019 数据集上的实验表明,相较于基础 YOLOv11s 模型,准确率(P)、召回率(R)和检测精度(mAP50)分别提升了 5.6%,5.9%和 7.5%,并且参数量下降了约 1/4,对比其他算法表现出了良好的性能。

关键词:航拍图像;目标检测;YOLOv11;无人机;小目标
中图分类号:TP391.4;TN919.8 **文献标识码**:A **国家标准学科分类代码**:520.6

Improved UAV aerial image detection algorithm for YOLOv11

Li Jun Ding Binbin Shi Weijuan Yang Lin
(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730000, China)

Abstract: Aiming at the UAV aerial image detection task, there are problems of tiny target size and complex background environment, which often lead to leakage and misdetection, this paper proposes a small target detection algorithm WT-YOLO based on YOLOv11 for aerial images. First of all, taking into account the problem that UAV aerial images are generally small targets, the structure of the YOLOv11 necking network is adjusted, and the output feature map is changed size, which improves the algorithm's ability to detect small targets. Secondly, the structure of Bottleneck and C3k2 module, named C3k2-WT, is redesigned in combination with WTConv to realize the efficient extraction of features. Again, Focal-Modulation is introduced to replace SPPF, which makes the model more robust in dealing with complex scenes by focusing and modulating the features at different spatial scales; finally, the shared convolution detection head is designed to reduce the number of parameters of the model through the convolution sharing mechanism, while enhancing the global information fusion capability between feature maps. The experiments of the improved algorithm on the VisDrone2019 dataset show that compared with the base YOLOv11s model, the accuracy (P), recall (R), and detection precision (mAP50) are improved by 5.6%, 5.9%, and 7.5%, respectively, and the number of params decreases by about one-fourth, which shows a good performance compared with other algorithms.

Keywords: aerial images; object detection; YOLOv11; UAV; small target

0 引 言

随着无人机技术的快速发展,航拍图像已成为获取地面信息的重要手段,被广泛应用于农业监测、城市规划、环境保护以及灾害管理等领域[1]。在这些应用中,目标检测是实现数据解读与场景理解的核心任务,通过自动识别航拍图像中的物体,可以极大地提高数据分析的效率。然而,航拍图像的特性使得目标检测面临诸多挑战,包括目标尺度的多样性、复杂的背景以及目标密集分布等问题[2]。

此外,航拍图像常因无人机拍摄角度、光照条件和飞行高度的变化而呈现出质量差异,并且无人机的应用环境通常对算法的实时性和算力效率提出严格要求,由于硬件性

收稿日期:2025-02-25
*基金项目:国家自然科学基金(62241204)项目资助

能、功耗和飞行时间的限制，检测模型需在高分辨率图像下保持实时性，同时尽可能降低对计算资源的依赖[3]。

现阶段目标检测算法主要分为以区域卷积网络(region-based convolutional neural networks，R-CNN)系列为代表的双阶段检测，和以 SSD(single shot multibox detector)、YOLO(you only look once)系列为代表的单阶段检测。双阶段目标检测算法通过候选区域生成和候选区域分类与定位两个阶段的处理，表现出了优越的性能[4]。在双阶段航拍图像检测中，苗茹等[5]采用 Swin Transformer 来替代 Faster R-CNN 的骨干网络，增强了模型的特征提取能力；Chen 等[6]在 Faster R-CNN 主干网络的顶层设计了一个跨尺度层次感知模块，用于整合来自不同层次和尺度的上下文信息，从而优化微小对象在特征尺度上的表示。

双阶段检测算法虽然表现出了较高的检测精度，但是相比于单阶段目标检测算法，伴随着计算复杂度的增加，这对于一些对实时性要求较高的场景来说是一个不利因素。因此，本文选取单阶段 YOLO 系列检测算法展开研究，以满足计算资源有限和实时性检测的双重需求。基于此类算法研究者们提出了一系列的改进策略。

Liu 等[7]在 YOLOv5 中引入 SPD-Conv 模块，通过在下采样过程中保留图像细粒度特征，提升了模型对小尺度目标的识别能力，但在处理航拍场景中较大尺寸目标时检测效果一般。邹振涛等[8]在 YOLOv7 的特征提取阶段引入三维注意力机制，使模型能更聚焦于空间关键信息，但增加了计算复杂度，影响了检测效率。Bu 等[9]提出的检测细化模块将可变形卷积融入 YOLOv8 骨干网络，增强了模型对几何形变目标和模糊对象的特征提取能力，但在复杂背景环境下表现出适应性不足的缺陷。为进一步提升复杂背景下的检测性能，涂育智等[10]基于 YOLOv11 设计了融合通道-空间注意力与多尺度卷积的轻量级特征提取模块 C2SCSA 和 C2MCA，在保持较低计算开销的同时增强了特征表达能力。李彬等[11]则通过膨胀特征金字塔卷积替代 YOLOv11 中的 SPPF 层，利用多尺度膨胀卷积提升无人机视角下的小目标细节捕捉能力，不过改进后的网络结构稍显复杂。王晓峰等[12]提出的全局-局部特征增强模块，通过融合特征图的局部细节与全局上下文信息，强化了深层特征的表达能力，但模块设计引入了额外的计算负担。

现有改进方案虽在不同层面提升了航拍目标检测性能，但仍存在以下共性问题：复杂背景下的目标漏检率高、多尺度目标检测精度不均衡、计算效率与检测精度的矛盾突出。YOLOv11 算法通过优化网络结构和引入高效特征提取模块，在保持较高检测精度的同时确保了更快的推理速度。因此，本文在 YOLOv11 的基础上提出 WT-YOLO 算法，用以提升航拍图像小目标检测精度，主要工作如下：

1)优化了 YOLOv11 模型的颈部网络结构，通过调整输出特征图的尺寸，使用 160×160 的小目标检测层，从而显著提高了模型对小目标的检测能力。

2)考虑到处理效率和模型轻量化的需求，设计了高效特征提取模块 C3k2-WT，优化了特征提取过程，并且减少了计算资源的消耗。

3)采用了 Focal-Modulation 机制替换了之前的 SPPF 模块。通过在不同的空间尺度上聚焦和调制特征，能够更精确地处理图像中的关键信息。

4)为了整合并强化模型中的信息流，设计了共享卷积检测头。这种设计通过卷积共享的方式，不仅减少了模型的参数量，还提高了不同特征层之间的信息融合能力。

1　相关工作

1.1　YOLOv11

YOLOv11 是 Ultralytics 团队开发的 YOLO 系列的版本，与 YOLOv8 相比，YOLOv11 在结构设计上引入了多项改进来提升模型性能。首先，使用 C3k2 模块替换了 C2f 模块，优化了特征提取能力，C3k2 模块引入了两种配置结构：C3k 块和瓶颈块。当参数设置为 True 时，采用的是 C3k 块，集成额外的卷积来增强局部特征提取；当参数设置为 False 时，使用的是瓶颈块，保持传统的卷积结构。其次，在 SPPF 模块之后新增了跨阶段局部自注意 C2PSA 模块，使模型能够更加聚焦图像中的关键区域，过滤无关细节。此外，YOLOv11 的检测头整合了轻量的深度可分离卷积(DWConv)，降低了计算成本。这些结构优化不仅使 YOLOv11 保持了卓越的检测精度，还显著提高了计算效率。YOLOv11 的网络结构如图 1 所示。

1.2　算法改进

YOLOv11 算法在面对无人机航拍图像检测时，虽然其高准确性得到了广泛认可，但仍旧存在一些不足：对于分辨率较小的目标，容易丢失位置和细节信息，导致小目标漏检；其次，对复杂背景的适应性不足，无人机航拍图像背景复杂，干扰信息多，并且航拍图像中目标大小变化大，小目标细节特征易丢失。所以本文在 YOLOv11s 模型的基础上做了如下改进：在主干网络中，使用本文设计的 C3k2-WT 模块替换原来的 C3k2，增强模型的特征表达能力；其次，引入了 Focal-Modulation 来代替传统的 SPPF，在多个空间尺度上对特征进行精细调节，提升了模型在复杂环境中的表现稳定性；然后使用 160×160 的小目标检测层，增强模型对小尺寸目标的识别能力，提升检测精度；最后通过共享卷积检测头，不仅降低了模型的参数总量，而且通过共享机制强化了特征图间的信息融合，进一步优化了检测效果。改进后的 YOLOv11 网络结构如图 2 所示。

2　算法实现

2.1　C3k2-WT

小波变换(wavelet transform，WT)是一种信号分析方法，通过特定的小波基将信号分解成不同的频率分量，并实现多尺度分析。与传统傅里叶变换不同，小波变换具有

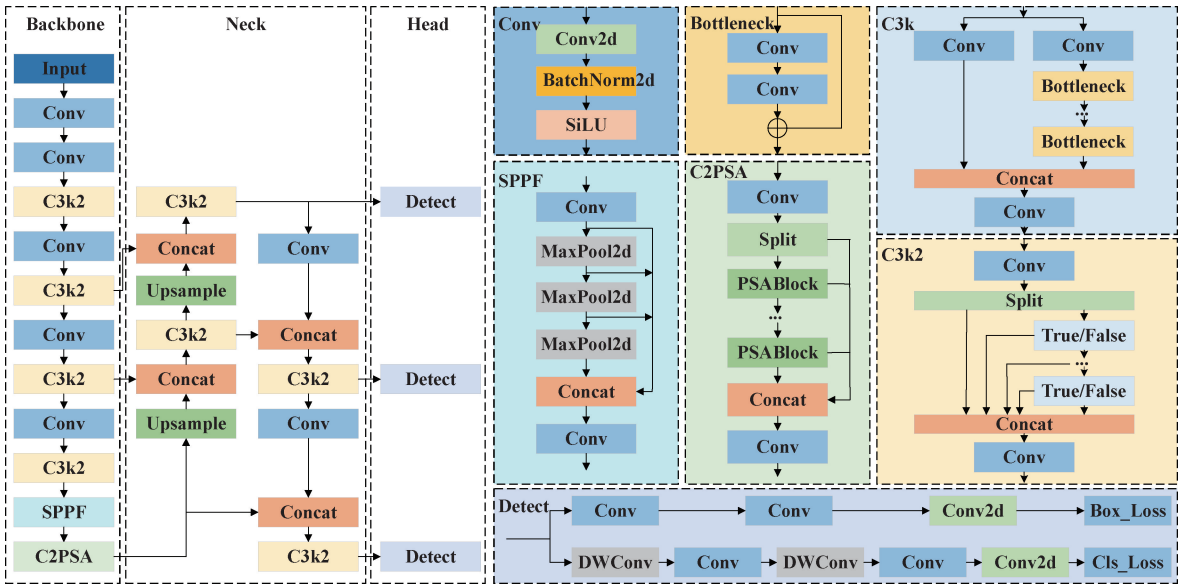

图 1　YOLOv11 网络结构

Fig. 1　Structure of YOLOv11 network

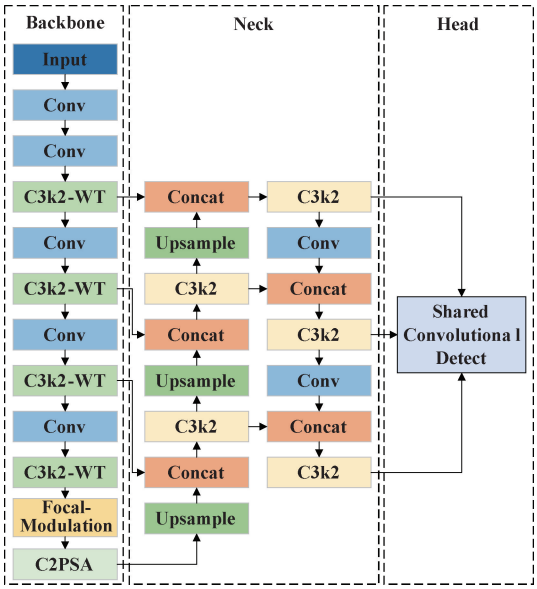

图 2　WT-YOLO 网络结构

Fig. 2　Structure of WT-YOLO network

时频局部化的特点,能够同时捕捉信号在时间域和频率域的信息。卷积是一种数学运算,用于从输入信号中提取特定模式,CNN 利用多层卷积操作,逐步提取从低级特征到高级特征的信息。然而,CNN 的传统卷积操作缺乏对频率域特征的建模能力,这正是小波变换的优势所在。

小波卷积(wavelet convolutions,WTConv)[13]将小波变换的多分辨率分析能力与卷积的特征提取能力相结合,能够更加高效地捕捉信号的多尺度特征。其核心思想是:首先通过小波变换分解输入信号或图像信息,将其分解为多个频段(低频和高频分量)。其次在各个频段上应用卷积操作,提取对应尺度的特征。最后进行特征融合或逆变换,用于构建输出。

对于输入信号 $x(t)$,wavelet convolutions 可以表示为:

$$y(t)=\sum_{k}\phi_k(t)*(x*\psi_k)(t) \tag{1}$$

式中:ψ_k 代表小波基函数,用于多尺度分解;ϕ_k 代表多尺度卷积核;$*$ 表示卷积操作。

小波变换的核心优势是多分辨率分析能力,而 WTConv 继承了这一特性。首先是捕捉多尺度特征:小波卷积能够将信号或图像分解为不同频率的分量(如高频分量表示局部细节,低频分量表示全局特征),从而提取多尺度特征。其次适应非平稳信号:多分辨率分析对于非平稳信号具有极大的优势,因为它可以动态调整分辨率以适应局部变化。

本文利用 WTConv 重新设计了 YOLOv11 的 Bottleneck 结构,具体而言就是使用 WTConv 替换了原始的 Conv,构造了 WT_Bottleneck 的结构,如图 3 所示,WT_Bottleneck 利用小波卷积可以将输入特征分解为低频和高频子带,在高频子带提取细节特征,在低频子带提取全局结构特征。

同时利用 WT-Bottleneck 优化了 YOLOv11 网络中的 C3k 和 C3k2 模块,得到了 C3k-WT 和 C3k2-WT,将输入特征分解为低频和高频子带,用在 YOLOv11 骨干网络中,实现特征的更高效提取,从而增强网络的学习能力。C3k2-WT 在网络中的具体添加位置如图 3 所示。

2.2　Focal-Modulation

在 YOLOv11 的主干网络中,SPPF 模块通过不同尺度的池化操作来获取多尺度的特征信息,这种多尺度的特

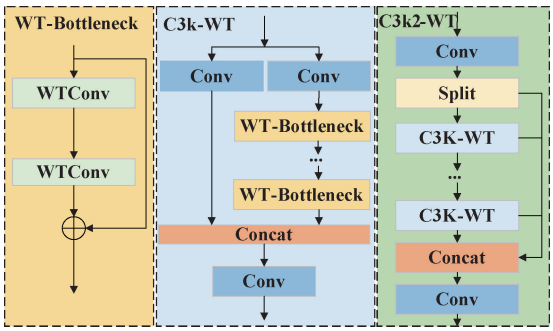

图 3　C3k2-WT 模块结构

Fig. 3　C3k2-WT module structure

征表示有助于模型更好地捕捉目标的形状、大小和位置信息，从而提高检测的准确性。尽管 SPPF 模块提高了计算能力，但在处理大规模数据集或复杂任务时，仍然需要足够的硬件资源来支持模型的训练和推理过程。为了进一步提高检测精度，本文选择将 Focal-Modulation[14] 引入 YOLOv11 替换原有的 SPPF 模块。其结构如图 4 所示。

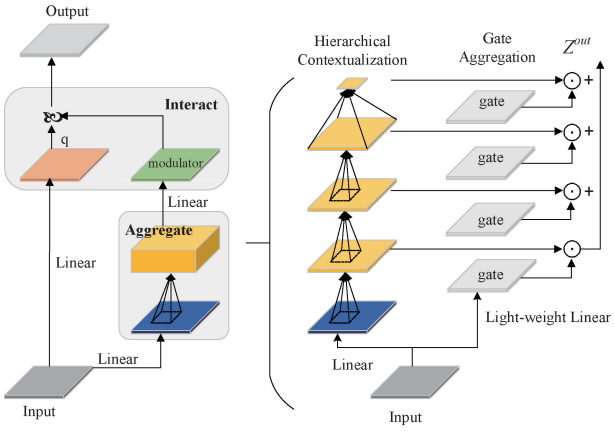

图 4　Focal-Modulation 结构

Fig. 4　Focal-Modulation structure

Focal-Modulation 的原理是通过捕捉图像中的长距离依赖和上下文信息，并与查询特征交互，增强特征表示的多样性和丰富性，从而更加有效地捕捉局部和全局信息，提高模型的识别能力。首先利用深度可分离卷积层对图像进行分层语境化，这一步骤通过不同粒度水平对短程到长程的视觉语境进行编码，减少了模型的参数数量和计算量，同时保持了良好的特征提取能力。接下来采用门控聚合策略，根据每个图像区域的内容，有选择性地收集语境信息。这一步骤通过调制器实现，调制器根据查询内容来聚合相关语境信息，从而优化了对视觉任务中标记交互的建模能力。门控聚合策略使得模型能够更准确地理解图像中的重要区域，为后续的特征提取和目标检测提供有力支持。最后通过逐元素的仿射变换，将调制器中的语境信息注入到查询中。这一步骤实现了对图像特征的优化提取，使得模型能够更准确地关注和分析图像中的重要部分。逐元素仿射变换的应用，不仅提高了目标检测的准确性，还增强了模型的鲁棒性和泛化能力。

使用 Focal-Modulation 替换 YOLOv11 中的 SPPF 模块，可以提供比固定尺度池化更强的全局上下文理解能力，动态调整局部特征的权重，突出重要区域，从而提升网络处理复杂目标或场景的鲁棒性。

2.3　小目标检测层

由于航拍图像中小目标的尺寸普遍较小，其主要的特征信息往往集中在浅层的特征图中。然而，原始 YOLOv11 网络的设计侧重于对多尺度目标的检测，导致初始特征图经过多次下采样后，小目标的特征信息逐渐丢失或被深层网络的高语义特征所掩盖，难以在最终检测头中完整保留。这种局限性降低了模型在航拍数据场景中对小目标检测的精度。因此，本文在 YOLOv11 网络的基础上增加 160×160 的小目标检测层，并且去除 20×20 的大目标检测层，以提升模型对浅层特征的利用能力。

改进后的网络结构在保持原 YOLOv11 核心模块的基础上，对颈部结构进行了调整。当通过上采样操作生成 80×80 的特征图后，额外增加了一次上采样操作，生成 160×160 特征图。为了进一步增强小目标的特征表达能力，生成的 160×160 特征图与主干网络中第一个 C3k2 模块提取的浅层特征图进行了特征融合操作，能够有效保留浅层特征中的空间细节信息，这些信息对于小目标检测尤为关键。融合后的特征图被直接输入到头部结构中进行目标分类和边界框回归，从而实现了对浅层特征图中小目标信息的深度挖掘和有效利用，改进后的检测层如图 5 所示。

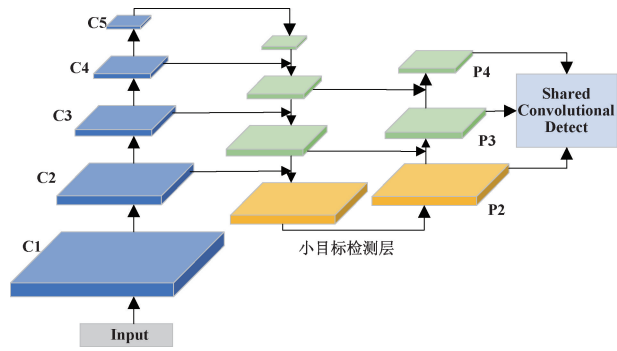

图 5　小目标检测层设计

Fig. 5　Small target detection layer design

2.4　共享卷积检测头

YOLOv11 的检测头结构中加入了两个深度可分离卷积(DWConv)，旨在减少计算量的同时保持网络性能。深度可分离卷积通过分别进行深度卷积和逐点卷积，有效降低了计算复杂度，使得模型在保持精度的同时，能够更快地运行。但在小目标检测任务时，仍面临着一些不足。因此，本文设计了一种轻量化的共享卷积检测头(shared convolutional detect，SCDetect)，如图 6 所示。SCDetect 保

留了 YOLOv11 原本的检测头结构,加强对小目标信息的提取,其核心思想是使用共享卷积在分类分支和回归分支共享参数,减少各分支使用标准卷积的次数,达到减少参数量目的。并且使用 Group Norm 代替 Batch Norm,因为 Group Norm 具有更好的适应性、能够保持通道之间的相对关系、提高模型的鲁棒性、在一些场景下性能更佳以及减少对小批量大小的依赖等优点。此外,Tian 等[15]已经证明在处理一些目标检测任务中,Group Norm 可以提升检测头定位和分类的性能。因此 SCDetect 在卷积结构中使用 GN 归一化。

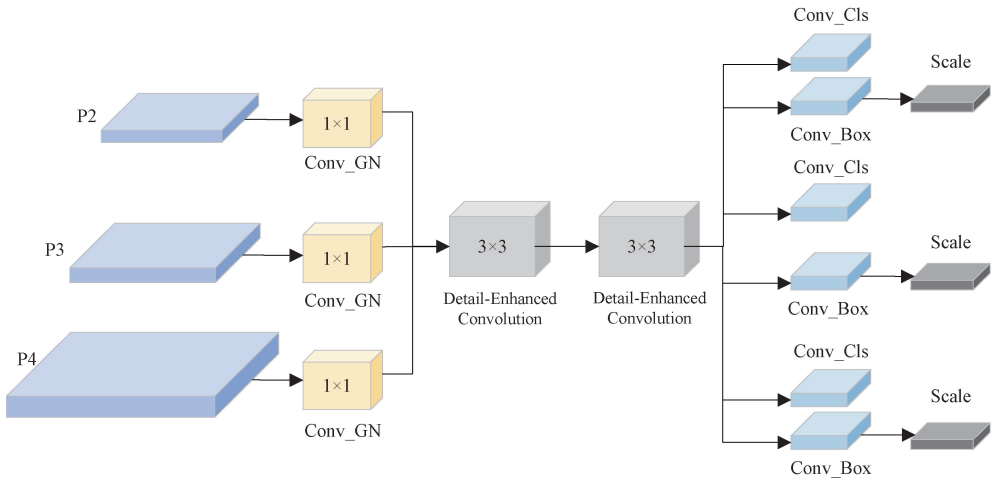

图 6　共享卷积检测头

Fig. 6　Shared convolutional detection head

SCDetect 的结构如图 6 所示,接收 P2、P3、P4 三个层级输入的特征后,首先在三层级分别使用卷积核大小为 1×1 的 GN 归一化标准卷积,增加通道维度的信息交换,丰富目标信息量。其次使用 2 个卷积核大小为 3×3 的共享细节增强卷积(detail-ehanced cnvolution,DEConv)[16],DEConv 是一种结合了传统卷积和差分卷积的卷积技术,它通过并行部署多个卷积层(包括差分卷积和普通卷积)来提取特征,并将梯度先验编码到卷积层中,从而增强模型的表示和泛化能力,差分卷积用于增强梯度级信息,而普通卷积则用于获得强度级信息。

最后,把共享细节增强卷积提取出的信息输入到分类和回归头中,同时为应对每个检测头所检测目标尺寸不一致的问题,预防小目标特征信息减少的情况,使用 Scale 层对回归检测头中的特征进行缩放,加强对多尺度特征的保留能力。Scale 层含有一个可学习的缩放因子 X,其与任意尺度的输入相乘,并且随着训练的进行,X 将根据梯度下降算法更新,从而实现对输入特征的缩放。

3　实　　验

3.1　实验数据集和实验环境

本文实验数据采用 VisDrone2019 数据集,它由天津大学 AISKYEYE 团队收集,广泛应用于各种计算机视觉任务,特别是目标检测。这些数据来源于各类无人机摄像头,覆盖了中国的 14 个不同城市,有不同环境和密度场景下的图像。包括了行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳蓬三轮车、巴士及摩托车 10 个类别,数据集中有 6 471 张图片用于训练,548 张图片用于验证。

本文所使用实验环境和配置信息如表 1 所示。

表 1　实验环境和配置信息

Table 1　Experimental environment and configuration information

配置类型	配置名称	配置信息
软件配置	操作系统	Windows11
	Python 版本	3.8
	深度学习框架	Pytorch 2.0
	CUDA	11.8
	Cudnn	8.9.5.30
硬件配置	CPU	Intel(R)Core(TM) i9-13900HX
	GPU	NVIDIA GeForce RTX 4060
	内存大小	16 GB
	显存大小	16 GB

相关参数设置:输入图像大小设置为 640×640 的统一尺寸;并且开启 Mosaic 数据增强,最后 10 轮关闭;batch_size 大小设置为 8,epoch 设置为 200;使用 SDG 优化器;训练过程中,初始学习率为 0.01。

3.2　评价指标

为了体现本文算法改进的有效性,选取的评价指标有准确率(precision,P)、召回率(recall,R)、平均精度均值(mean average precision,mAP)、参数量(Params)、模型大小(model size)、计算量(GFLOPs)和帧率(FPS)。

mAP@50 代表当 IoU(交并比)阈值设定为 0.5 时的

平均精度，而 mAP@50:95 则指的是 IoU 从 0.5 开始，以 0.05 的步长逐步递增至 0.95 时的平均精度均值，mAP 值越高，意味着模型的整体精度越出色。Params 的数值大小反映了模型的轻便程度，Model Size 值代表训练出来的模型权重文件的大小，其数值越低，说明模型所占用的内存越小；GFLOPs 值体现了模型的计算复杂度，数值越低，意味着模型在运行时所需的计算量越小；FPS 代表模型每秒可处理的帧数，其数值越高，说明模型的推理速度越快。

评价指标的相关计算公式如下：

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{n}\int_{0}^{1} P(R)\mathrm{d}R \tag{4}$$

式中：TP 表示检测样本实际为真，预测为真；FP 表示样本实际为假，预测为真；FN 表示样本实际为真，预测为假。

3.3　检测头对比实验

为了验证 WT-YOLO 中所使用的小目标检测层改进的实际效果，在保持骨干网络不变的条件下对 YOLOv11s 模型进行了对比实验，实验选取两种使用小目标检测层的方式，如图 7 所示，图中左侧为 YOLOv11s 颈部结构，检测头尺寸为 P3、P4、P5；中间方法 1 为直接添加 P2 小目标检测层，右侧方法 2 是使用 P2 小目标检测层的同时，去除了 P5 大目标检测层，实验结果如表 2 所示。

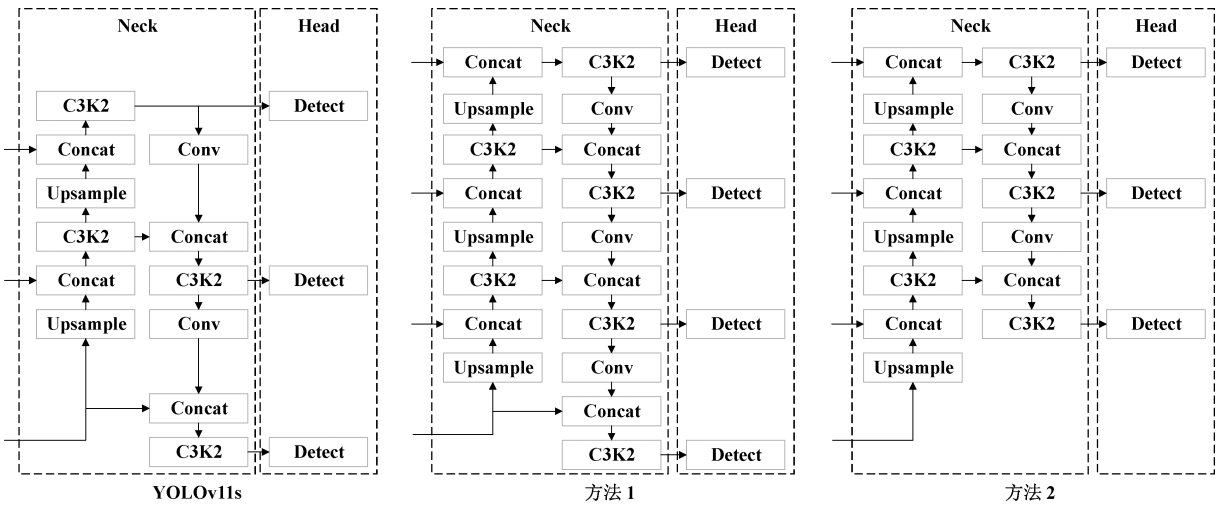

图 7　3 种检测头结构对比

Fig. 7　Comparison of three detection head structures

表 2　检测头结构对比试验

Table 2　Comparative test of detection head structure

P2	P3	P4	P5	P/%	R/%	mAP@50/%	Params/M	Model Size/MB	GFLOPs
	√	√	√	48.8	38.0	38.7	9.4	18.2	21.3
√	√	√	√	53.9	41.7	43.3	9.5	18.6	28.6
√	√	√		52.7	42.0	43.4	7.1	13.9	25.6

根据表 2 的数据可知，采用右侧的改进方案网络性能最佳，mAP@50 达到了 43.4，相比 YOLOv11 提升了 4.7%。同时，模型参数从 9.4 M 减少到 7.1 M，下降了约 1/4。中间方案虽然也显著提升了检测性能，但计算量有所增加，考虑到无人机平台的计算能力有限，模型的轻量化显得尤为重要，因此本文选择右侧的改进方案。

3.4　各个类别精度对比

本文对比了基准算法与 WT-YOLO 在 VisDrone2019 数据集上各类别的检测精度的差别，结果如表 3 所示。显示了 Pedestrian、People、Bicycle、Car、Van、Truck、Tricycle、Awning-tricycle、Bus、Motor10 个类别的 mAP50 数据值，以及总的 mAP50 数据，改进算法在待检测的 10 个类别上均超过了原 YOLOv11s 算法，且均超过 4.0% 的精度增长，这表现了改进算法对不同尺度目标检测的适应性。对 Pedestrian 和 People 这两个类别的检测精度提高尤为明显，分别提升了 11.3% 和 10.5%，且行人和人均是目标尺寸较小的类别，在数据集中多出现聚集的现象，表明了算法可以更好地关注到图像中的小目标。此外，对于汽车和卡车等尺寸较大的目标检测性能也有不小的提升，说明本文设计的 WT-YOLO 算法对于实际目标检测性

能的优异性。

表 3　各个类别检测精度对比

Table 3　Comparison of detection accuracy by category

类别	mAP/%		
	YOLOv11s	WT-YOLO	increase
Pedestrian	42.1	53.4	↑11.3
People	32.7	43.2	↑10.5
Bicycle	12.1	17.3	↑ 5.2
Car	79.5	84.9	↑ 5.4
Van	45.9	50.4	↑ 4.5
Truck	36.4	40.4	↑ 4.0
Tricycle	25.8	34.5	↑ 8.7
Awning-tricycle	14.7	20.3	↑ 5.6
Bus	55.1	63.3	↑ 8.2
Motor	42.8	53.7	↑10.9
all	38.7	46.2	↑ 7.5

3.5　消融实验

为了验证本文所采用每种改进策略对检测精度的提升,在 VisDrone2019 数据集上使用 YOLOv11s 作为基线模型进行消融实验,每组实验采用相同的训练策略,实验结果如表 4 所示。其中:1)实验 A 表示在骨干网络中,引入 C3k2-WT 模块来替换 C3k2,mAP@50 提升了 1.1%,证明了 C3k2-WT 模块的可行性;2)实验 B 表示使用 Focal-Modulation 来替换 SPPF,模型性能得到了小幅度提升;3)实验 C 表示使用设计的卷积共享检测头,mAP@50 提升了 0.8%;4)实验 D 表示添加小目标检测层,此项改动对模型性能有较大提升,mAP@50 提升了 4.7%,mAP@50:95 提升了 3.1%,同时使模型参数由 9.4 M 降至 7.1 M,大幅增强了网络的特征提取能力;5)实验 E 表示对 YOLOv11s 模型添加本文所有的改进措施。

对表 4 中实验结果分析可知,相较于基线 YOLOv11s 模型,本文方法准确率提升了 5.6%,mAP@50 提升了 7.5%,mAP@50:95 提升了 5.0%,同时参数量下降了约 1/4,模型体积下降了约 15%。虽然本文改进方法牺牲了部分计算量和检测帧率,但是在实际应用中能够满足实时性的检测要求,并且保证检测结果的高度准确性。证明了本文改进措施的有效性。

表 4　消融实验结果

Table 4　Results of ablation experiments

方法	C3K2-WT	Focal-Modulation	SC-Detect	小目标检测层	P/%	mAP@50/%	mAP@50:95/%	Params/M	Model Size/MB	GFLOPs	FPS
YOLOv11s					48.8	38.7	23.2	9.4	18.2	21.3	131
A	√				50.5	39.8	23.8	8.9	17.5	22.5	122
B		√			49.4	39.1	23.5	9.8	20.5	21.7	155
C			√		50.7	39.5	23.8	9.0	18.9	23.8	151
D				√	52.7	43.4	26.3	7.1	13.9	25.6	119
E	√	√	√	√	54.4	46.2	28.2	6.9	15.4	27.6	109

3.6　不同算法对比实验

为验证所提算法在小目标检测上的性能优势,本文在 VisDrone2019 数据集上开展了对比实验,选取了双阶段 R-CNN 系列与单阶段 YOLO 系列中的先进检测算法作为基准,实验结果如表 5 所示。在包含 10 个目标类别的检测任务中,WT-YOLO 算法在各类别上均展现出最优检测性能,特别是在 Pedestrian 和 People 这类小目标检测上表现尤为突出,分别取得了 53.4% 和 43.2% 的 mAP@50 精度。对于检测难度较大的 Bicycle 类别,改进算法仍能达到 17.3% 的检测精度。在保持较大尺寸目标检测性能平衡方面,WT-YOLO 在 Car 和 Truck 类别上相比原始 YOLOv11s 模型分别提升 5.4% 和 4.0%。最终,算法在总 mAP@50 指标上达到 46.2%,实验结果表明,改进后的模型架构能有效增强网络对小目标的特征捕捉能力,在无人机航拍图像目标检测任务中具有显著优势。

3.7　其他改进算法对比试验

为了验证本文改进算法的优越性,在 VisDrone2019 数据集上,与其他先进航拍图像目标检测算法进行了对比。表 6 和 7 分别是本文在 YOLOv11s 模型和 YOLOv11n 模型的基础上与其他学者算法的对比实验,s 模型比 n 模型拥有更高的检测精度,实验结果如下,表中展示了 mAP@50 和 mAP@50:95 的对比数据。

在表 6 的 YOLOv11s 模型的对比中,LSOD-YOLO 在 mAP@50 和 mAP@50:95 上的表现分别为 45.2% 和 27.4%,而本文算法达到了 46.2% 和 28.2%。相比之下,文献[17]和 YOLO-RLDW 的检测结果和本文相近,mAP@50 数值分别为 46.1% 和 45.8%,但仍低于本文算法。

在表 7 的 YOLOv11n 模型对比中,文献[11]和文献[12]致力于小目标信息的特征提取,HPRS-YOLO 更加关注不同尺度特征信息的融合,而本文通过对 C3k2 的改进、使用小目标检测层和共享卷积检测头,兼顾了特征提

表 5　不同算法对比试验

Table 5　Comparison test of different algorithms　　%

模型	目标类别										mAP
	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning-tri	Bus	Motor	
SSD	18.7	9.0	5.0	63.2	30.0	33.1	11.7	15.5	47.2	19.1	25.3
Faster R-CNN	20.9	14.8	7.3	51.0	29.7	19.5	14.0	8.8	30.5	21.2	21.8
RetinaNet	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9
CornerNet	20.4	6.6	4.6	40.9	20.2	20.5	14.0	9.3	24.4	12.1	17.4
CenterNet	22.6	20.6	14.6	59.7	24.0	21.3	20.1	17.4	37.9	23.7	26.2
YOLOv5s	39.2	31.4	10.6	72.6	33.7	26.2	18.2	9.8	39.9	38.4	32.0
YOLOv6s	37.2	29.8	8.9	78.1	42.6	32.5	23.6	14.8	51.2	39.6	35.8
YOLOv8s	42.0	32.5	13.0	79.6	44.8	34.2	26.4	15.9	57.6	43.3	38.9
YOLOv9s	38.4	32.8	10.3	78.0	43.2	34.3	26.9	15.0	51.5	42.2	37.3
YOLOv10s	39.9	31.3	11.6	79.0	45.0	35.4	24.8	16.5	55.8	42.5	38.2
YOLOv11s	42.1	32.7	12.1	79.5	45.9	36.4	25.8	14.7	55.1	42.8	38.7
WT-YOLO	53.4	43.2	17.3	84.9	50.4	40.4	34.5	20.3	63.3	53.7	46.2

表 6　改进 YOLOv11s 模型对比试验

Table 6　Comparative experiment on improving YOLOv11s

模型	mAP@50/%	mAP@50:95/%
LSOD-YOLO[18]	45.2	27.4
改进 YOLOv8s[17]	46.1	27.9
SOD-YOLO[19]	45.1	26.6
YOLO-RLDW[20]	45.8	27.4
WT-YOLO	46.2	28.2

表 7　改进 YOLOv11n 模型对比试验

Table 7　Comparative experiment on improving YOLOv11n

模型	mAP@50/%	mAP@50:95/%
改进 YOLOv11n[11]	40.1	24.1
改进 YOLOv11n[12]	36.7	21.9
HPRS-YOLO[21]	38.4	22.7
WT-YOLO	40.3	24.2

取以及特征融合策略。在检测精度方面均优于其他几种改进算法，在小目标检测任务中表现出色，体现了本文 WT-YOLO 在航拍小目标检测中的优势。

3.8　模型泛化实验

为验证本文改进方法的泛化性能与鲁棒性，选取 RSOD 和 DIOR 两个遥感图像数据集开展模型对比实验。RSOD 数据集包含飞机、油罐、操场、立交桥 4 类目标，共 976 幅图像；DIOR 数据集作为大规模光学遥感检测基准，涵盖飞机、机场等 20 个类别，共 23 463 幅图像。两个数据集在目标类别、场景复杂度及数据量层面存在显著差异，构成多样化的挑战场景。本文采用 7∶2∶1的比例对 RSOD 数据集进行训练集/验证集/测试集划分，DIOR 数据集则遵循官方划分规则。选取 YOLOv5s、YOLOv8s、YOLOv10s、YOLOv11s 作为对比算法，从准确率、召回率和检测精度方面开展对比实验，结果如表 8 所示。

表 8　模型泛化对比试验

Table 8　Model generalization comparison test

数据集	模型	P/%	R/%	mAP@50/%	mAP@50:95/%
RSOD	YOLOv5s	78.4	84.2	78.2	55.4
	YOLOv8s	81.2	81.4	78.4	55.4
	YOLOv10s	83.2	75.1	78.0	55.3
	YOLOv11s	81.4	84.0	80.2	56.7
	WT-YOLO	86.3	84.2	83.2	60.6
DIOR	YOLOv5s	84.8	71.2	77.0	55.9
	YOLOv8s	87.7	74.3	79.7	60.1
	YOLOv10s	87.0	72.8	79.2	59.1
	YOLOv11s	88.5	74.5	80.1	60.8
	WT-YOLO	88.9	77.5	83.6	62.0

实验数据表明，WT-YOLO 在两个数据集上均展现出最优检测性能。在 RSOD 数据集上，该算法相比 YOLOv11s 的 mAP@50 指标提升 3.0%，达到 83.2%；在 DIOR 数据集上，mAP@50 指标提升 3.5%，达到 83.6%。值得注意的是，WT-YOLO 在保持最高检测精度的同时，模型参数量低于其他对比算法。通过跨数据集的泛化性验证，充分证明了本文改进方法在处理遥感图像小目标检测任务时的有效性，展现出良好的算法适应性与鲁棒性。

3.9　检测结果对比试验

为了直观地比较本文改进算法与 YOLOv11s 在无人机航拍场景下的检测效果，在 VisDrone2019 数据集中选取了小目标密集分布、光照条件变化、多尺度目标和光线暗

淡几个复杂场景的航拍图像进行了检测结果对比实验,小目标聚集和光线复杂是航拍图像检测较为常见的复杂场景,多尺度目标场景指同一图像中同时存在显著尺寸差异的目标(如行人、汽车、卡车),此类场景要求模型兼顾浅层细节与深层语义特征,实验结果如图 8 所示。其中左侧为原图,中间为 YOLOv11s 模型在该场景下的检测效果图,右侧是采用本文改进算法 WT-YOLO 的检测效果图,与 YOLOv11s 的检测差别在图中以红色矩形框标出。

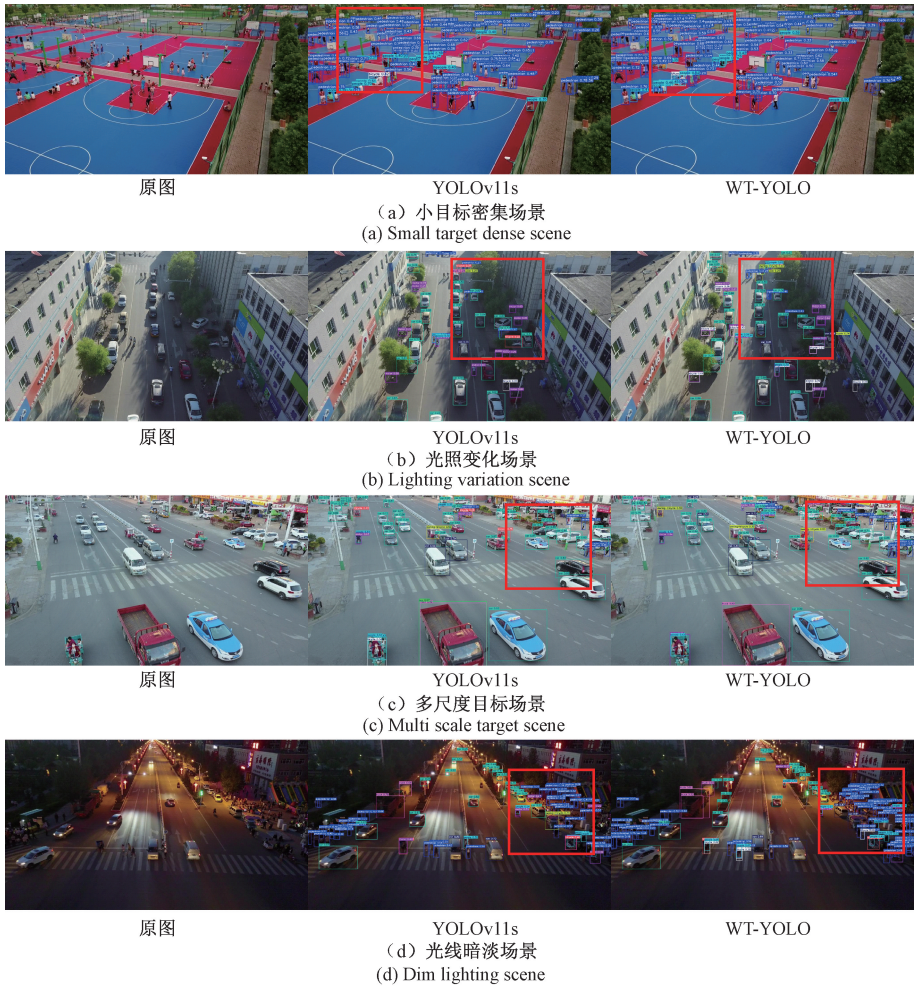

图 8　各类场景下图像检测效果对比

Fig. 8　Comparison of image detection effects in various scenarios

在小目标密集分布场景下,相较于基准算法,WT-YOLO 算法的漏检率更低,可以在高度重叠的环境下准确找到待检测实体,比基准算法多检测出 15 个人;在自然光照变化场景下,WT-YOLO 算法能够抑制背景噪声信息干扰并保留对目标决策更重要的特征信息,比基准算法多检测出 21 个物体;在多尺度场景下,WT-YOLO 算法能够有效地适应不同尺度下目标的变化,精确检测各类尺寸的物体,对行人和汽车有着更高的识别率;在夜间光线暗淡场景下,WT-YOLO 算法可以克服光线不足对目标识别的影响,相比基准算法多检测出 36 个行人、4 辆自行车。综上所述,WT-YOLO 算法在各种复杂场景均表现出了更好的检测效果,尤其在光照不稳定的情况下有着更好的鲁棒性。

4　结　　论

针对无人机航拍图像中目标尺度微小、背景环境复杂且目标间易重叠等挑战,本文从特征提取和融合策略,以及检测头设计等多个维度,对 YOLOv11s 模型进行了改进,提升了算法的检测性能,同时减少了模型的参数量。

首先,面对无人机航拍图像中小目标检测的挑战,对 YOLOv11 的颈部结构进行了重构,通过调节输出特征图的维度,显著增强了模型对小尺寸目标的识别能力。进一步地,结合 WTConv 的轻量级特性,对 Bottleneck 和 C3K2 模块进行了改进,设计了 C3K2-WT 模块,提升了特征提取的效率和精度。此外,引入了 Focal-Modulation 来代替传统的 SPPF,有效提升了模型在复杂环境中的表现稳定性。

最后，通过实施共享卷积检测头，不仅降低了模型的参数总量，而且通过共享机制强化了特征图间的信息融合，进一步优化了检测效果。这一系列改进使模型在小目标检测领域表现出卓越的性能。

在 VisDrone2019 数据集上的实验结果显示，改进后的算法在检测效果上取得了显著的提升，具有参数量少、检测精度高等优点，基本可以满足实时性应用的需求。但是该算法在某些情况下仍存在漏检现象，未来的研究方向将聚焦于如何进一步提升网络的性能，并将本文所提的算法应用于更复杂多变的实际场景中。

参考文献

[1] MU A M, WANG H J, MEN W J, et al. Small target detection in drone aerial images based on feature fusion[J]. Signal, Image and Video Processing, 2024, 18(Suppl 1): 585-598.

[2] 袁玲玲，陈春梅，朱天鑫，等. 基于航拍图像的自适应感知目标检测网络[J]. 电子测量技术，2025，48(2)：57-65.
YUAN L L, CHEN CH M, ZHU T X, et al. Adaptive perception object detection network based on aerial photography [J]. Electronic Measurement Technology, 2025, 48(2): 57-65.

[3] BAI CH SH, BAI X F, WU K J. A review: Remote sensing image object detection algorithm based on deep learning[J]. Electronics, 2023, 12(24): 4902.

[4] 陈朋磊，王江涛，张志伟，等. 基于特征聚合与多元协同特征交互的航拍图像小目标检测[J]. 电子测量与仪器学报，2023，37(10)：183-192.
CHEN P L, WANG J T, ZHANG ZH W, et al. Small object detection in aerial images based on feature aggregation and multiple cooperative features interaction[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(10): 183-192.

[5] 苗茹，李祎，周珂，等. 一种改进的 Faster R-CNN 遥感图像多目标检测模型研究[J/OL]. 计算机工程，1-14 [2024-12-02]. https://doi.org/10.19678/j.issn.1000-3428.0068856.
MIAO R, LI Y, ZHOU K, et al. A study on an improved Faster R-CNN model for multi-object detection in remote sensing images[J/OL]. Computer Engineering, 1-14 [2024-12-02]. https://doi.org/10.19678/j.issn.1000-3428.0068856.

[6] CHEN P L, WANG J T, ZHANG ZH W, et al. CSPGNet: Cross-scale spatial perception guided network for tiny object detection in remote sensing images[J]. Digital Signal Processing, 2024, 154: 104674.

[7] LIU S Z, CAO L H, LI Y. Lightweight pedestrian detection network for UAV remote sensing images based on strideless pooling [J]. Remote Sensing, 2024, 16(13): 2331.

[8] 邹振涛，李泽平. 改进 YOLOv7 的航拍图像目标检测[J]. 计算机工程与应用，2024，60(8)：173-181.
ZOU ZH T, LI Z P. Improved YOLOv7 for UAV image object detection[J]. Computer Engineering and Applications, 2024, 60(8): 173-181.

[9] BU Y CH, YE H R, TIE ZH X, et al. OD-YOLO: Robust small object detection model in remote sensing image with a novel multi-scale feature fusion [J]. Sensors, 2024, 24(11): 3596.

[10] 涂育智，王法翔，吴春霖. 融合多注意力机制的轻量级无人机航拍小目标检测模型[J]. 计算机工程与应用，2025，61(11)：93-104.
TU Y ZH, WANG F X, WU CH L. A lightweight UAV aerial small object detection model integrating multi-attention mechanisms[J]. Computer Engineering and Applications, 2025, 61(11): 93-104.

[11] 李彬，李生林. 改进 YOLOv11n 的无人机小目标检测算法[J]. 计算机工程与应用，2025，61(7)：96-104.
LI B, LI SH L. Improved YOLOv11n small object detection algorithm in UAV view [J]. Computer Engineering and Applications, 2025, 61(7): 96-104.

[12] 王晓峰，黄俊俊，谭文雅，等. 基于深度特征强化与路径聚合优化的目标检测[J/OL]. 计算机科学，1-18[2025-03-24]. http://kns.cnki.net/kcms/detail/50.1075.tp.20250227.0933.005.html.
WANG X F, H J J, TAN W Y, et al. Object detection based on deep feature enhancement and path aggregation optimization[J/OL]. Computer Science, 1-18[2025-03-24]. http://kns.cnki.net/kcms/detail/50.1075.tp.20250227.0933.005.html.

[13] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields[C]. European Conference on Computer Vision, 2024: 363-380.

[14] YANG J W, LI CH Y, DAI X Y, et al. Focal modulation networks [J]. Advances in Neural Information Processing Systems, 2022, 35: 4203-4217.

[15] TIAN ZH, SHEN CH 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.

[16] CHEN Z X, HE Z W, LU ZH M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention [J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.

[17] 杨路，裴俊莹. 融合多尺度特征的航拍目标检测算

法[J]. 系统仿真学报,2025,37(6):1486-1498.
YANG L,PEI J Y. Aerial target detection algorithm fused with multi-scale features[J]. Journal of System Simulation,2025,37(6):1486-1498.

[18] LI H K, WU J. LSOD-YOLOV8s: A lightweight small object detection model based on YOLOv8 for UAV aerial images[J]. Engineering Letters, 2024, 32(11):2073-2082.

[19] KHALILI B, SMYTH A W. SOD-YOLOv8-Enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes[J]. Sensors, 2024, 24(19): 6209.

[20] ZHAO L J, LIANG G, HU Y M, et al. YOLO-RLDW: An algorithm for object detection in aerial images under complex backgrounds[J]. IEEE Access, 2024, 12: 128677-128693.

[21] 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(12): 331619.
YANG Y G, JIANG W T, GAO ZH Y. Algorithm for real-time target detection in low-altitude UAVs[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(12): 331619.

作者简介

李珺,博士,副教授,主要研究方向为智能计算和图像处理。

E-mail: Lijane@mail.lzjtu.cn

丁彬彬(通信作者),硕士,主要研究方向为目标检测。

E-mail: 12231963@stu.lzjtu.edu.cn

史维娟,硕士,主要研究方向为图像识别。

E-mail: 12231975@stu.lzjtu.edu.cn

杨琳,硕士,主要研究方向为图像和自然语言处理。

E-mail: 12240742@stu.lzjtu.edu.cn