DOI:10.19651/j.cnki.emt.2518387

# 一种复杂场景无人机图像多尺度目标检测方法

詹雨飞
(复杂航空系统仿真全国重点实验室 成都 610036)

**摘 要**:针对无人机数据集图像中目标尺度小、特征弱、背景干扰多等不利因素造成的目标检测算法总体精度不高,漏检误检严重的问题,提出了一种用于复杂场景下无人机图像多尺度目标检测的新算法。该算法通过 DConv、AIFI 和 Dyhead 等模块的引入,改进了原网络在多尺度目标检测能力上的不足;同时,通过采用 DIoU 损失函数,提高了模型的收敛能力。在公开数据集 VisDrone-DET2019 上对多尺度目标进行检测识别,与原网络相比,精确率提升了 3.7%;召回率提升了 1.2%;平均精度提升了 2.3%。同时,通过大量的实验验证,结果显示本文算法具有较强的鲁棒性,综合性能优秀,具有一定的工业应用价值。

**关键词**:YOLOv8;动态卷积;注意机制;尺度内特征交互;多尺度目标检测
**中图分类号**:TP391;TN01 **文献标识码**:A **国家标准学科分类代码**:510.4050

# A multi-scale target detection algorithm for unmanned aerial vehicle(UAV) images in complex scenarios

Zhan Yufei
(National Key Laboratory of Complex Aviation System Simulation, Chengdu 610036, China)

**Abstract**: To deal with the challenges faced by target detection algorithms due to the small scale, weak features, and high background interference in images of a drone dataset, a multi-scale target detection algorithm for unmanned aerial vehicle images in complex scenarios is proposed. This algorithm enhances the overall accuracy, reduces false negatives and positives, through the incorporation of modules such as DConv, AIFI, and Dyhead. These components address the limitations of the original network in handling multi-scale targets. Furthermore, the use of the DIoU loss function improves the model's convergence capability. The effectiveness of this approach is demonstrated through its application in detecting multi-scale targets on the VisDrone-DET2019 dataset. Compared to the original network, there is a 3.7% increase in precision, a 1.2% increase in recall rate, and a 2.3% improvement in average accuracy. Moreover, extensive experiments demonstrated that the proposed algorithm exhibits strong robustness and excellent overall performance, suggesting significant industrial application potential.

**Keywords**: YOLOv8; dynamic convolution; attention mechanism; intra-scale feature interaction; multi-scale target detection

## 0 引 言

近年来,目标检测技术对整个计算机视觉领域产生了深远影响,成为该领域中最重要和最具挑战性的分支之一。无人机是一种具有自主飞行能力的飞行器,兼具高机动性与便携性,能够轻松实现精确定位与导航。随着卷积神经网络及相关深度学习理论的持续突破与创新,无人机视角下的多尺度目标检测,受到了国内外研究者们愈发广泛的关注。在目标尺度的量化过程中,通常通过计算目标实例所覆盖区域的像素数与图像整体面积的比值,并对该比值进行平方根运算,这一操作被称为“尺度”指标。通过这一度量方式,可以得出在不同图像中,目标实例的相对尺度会呈现显著的差异。同时,即使在同一幅图像中,多个目标的尺寸也可能存在明显差异,这一现象揭示了“尺度问题”的复杂性。因此,尺度差异不仅体现在不同图像之间,还可能在同一图像中的多个目标之间表现得尤为突出,这对目标检测算法的精度和鲁棒性提出了更高要求。

相较于早期的 YOLO 系列,YOLOv8s 在保持较高精度的同时,显著减少了模型的参数量和计算复杂度;此外,该算法通过改进的网络结构设计和优化的训练策略,提高

收稿日期:2025-03-19

了模型的推理速度,适合运用于对无人机多尺度目标的检测进行针对性改进。本文为解决复杂背景下多尺度目标检测的瓶颈,选用 VisDrone-DET2019 公开数据集作为研究对象,基于 YOLOv8s 模型提出了一种检测性能更加优异无人机图像多尺度目标检测算法。通过实验发现:与原算法相比,该算法精确率提升了 3.7%;召回率提升了 1.2%;平均精度提升了 2.3%,是一种高效可靠的目标检测算法。本文的主要贡献有:

1) 在骨干网络和颈部网络中引入 DConv,使得模型为不同尺度的输入样本动态地调整卷积核的权重,以适应不同目标的特征,从而提升检测性能;

2) 修改骨干网络中的 SPPF 为 AIFI,旨在提高模型在处理复杂场景时的灵活性和精确度,尤其针对本文存在的目标大小、形状多样的情况下,这种基于注意的尺度内特征交互机制使得模型能够更有效地处理和融合重要的特征信息,从而提高检测性能;

3) 采用 Dyhead 检测头,通过统一尺度感知、空间感知和任务感知使得模型不仅能灵活地适应于不同尺度大小、形状的目标,还能自适应于不同的任务;

4) 优化损失函数为 DIoU,它不仅依赖于目标框的重叠度,还综合考虑了目标框的中心距离,使得模型在多尺度的目标检测中能够更加精确地优化目标框的位置,同时加快训练收敛。

## 1 相关工作

在深度学习兴起之前,传统目标检测方法如:尺度不变特征变换、支持向量机和方向梯度直方图等背景建模方法虽然被广泛使用,但是应用在复杂背景下多尺度目标检测的效果并不理想。随着深度学习的崛起,目标检测技术在神经网络的持续发展推动下取得了显著进展。当前,目标检测算法的研究与应用已取得了显著的突破,主要可划分为两类。这些进步不仅提升了算法的检测准确性和效率,还为处理更加复杂的视觉任务提供了新的思路和方法。

一类是基于区域推荐策略的目标检测算法,也称两阶段目标检测算法。这类算法在起始阶段采用特定的区域生成策略,以生成预期包含目标的候选区域集合。候选区域的处理过程中,算法通常会进行特征提取、分类以及边界框回归等一系列操作。特征提取是第一步,负责从图像中提取有意义的视觉信息;随后,分类步骤确定目标的种类;最后,边界框回归操作则精确地调整框的坐标,以更好地匹配目标的实际位置。通过上述过程,旨在精确判断目标是否存在,并准确估算目标的位置与类别。接着,通过实施后处理流程,剔除冗余的检测结果,从而确保输出的检测结果具有高度的精确性且无重复。经典的两阶段目标检测算法包括区域卷积神经网络(region based convolutional neural network,R-CNN)[1]、Fast R-CNN[2]、Faster R-CNN[3]等。另一种是基于回归的目标检测算法,也称为单阶段目标检测算法。这些算法无需预设候选框,极大地简化了计算流程,提升了检测速度。检测过程大致如下:首先,将原始图像输入预设的网络,通过一系列卷积运算及其他相关操作提取多层次、多维度的特征信息;随后,在生成的特征图上,进行分类与边界框回归任务,旨在精准定位图像中潜在目标的位置及识别其所属类别;最后,通过一些后处理步骤(如:阈值筛选和非极大值抑制等),进一步优化检测结果的同时有效去除冗余预测。经典的单阶段目标检测算法包括 YOLO(you only look once)[4]、SSD(single shot multi-box detector)[5]、RetinaNet[6]等。

近年来,无人机航拍技术的流行进一步推动了该领域相关理论的快速发展。陈占龙等[7]提出了 Correg-YOLOv3 算法,通过引入嵌入角点回归机制来扩展模型输出维度,增加偏移量额外损失项,达到精准定位的密集分布目标的效果。赵耘彻等[8]提出了一种新型的 YOLOv4 目标检测算法,该算法将轻量级网络 MobileNetv3 应用于特征提取任务,替代了传统 YOLOv4 中使用的标准卷积层。同时,用深度可分离卷积替换 3×3 常规卷积。该方法在保持高效检测精度的同时,显著减少了计算复杂度和模型参数的数量,从而提高了模型的运行速度和计算资源的利用效率。徐光达等[9]以 YOLOv5 为基础模型,设计出多层级特征融合层,用于整合不同感受野的特征信息。采用解耦检测头,有效提升了微小目标的检测性能。同年,Zhang 等[10]基于 YOLOv5 模型,采用坐标注意力和 SPD-Conv(space to depth convolution)模块改善了特征提取网络的性能。实验结果表明,该方法可以增强对无人机图像中不同尺度,尤其是小尺度目标的检测能力。随着研究的进一步深入,李安达等[11]基于 YOLOv7 引入 Focal NeXt block 和 RepLKDeXt 模块,强化输出小目标的特征质量和提高输出特征包含的上下文信息含量。张徐等[12]在其研究工作中对 YOLOv7 模型进行了创新性扩展,引入了一种专门用于小目标检测的网络层。通过结合余弦注意力机制(增强网络对细粒度特征的关注)和后正则化技术(提高网络的泛化能力),提出了名为 cosSTR 的模块,旨在有效应对目标尺度变化幅度大以及目标特征信息不足的挑战。然而,尽管该方法在很大程度上提升了模型对常规目标的检测性能,对于弱特征目标,依然存在着一定程度的漏检问题。Wang 等[13]通过引入 BiFormer 的注意力机制来优化 YOLOv8 模型,使用 WiseIoUv3 来改善梯度分配的策略,有效地改进了模型的小目标检测能力。然而,此类方法并不能解决无人机数据分布不均造成的分类问题。张河山等[14]在现有研究的基础上,提出了一种优化的 YOLOX 网络架构,专为无人机航拍图像的检测任务而设计。该网络在特征融合阶段创新性地引入了自适应空间特征融合模块,该模块通过自适应调整不同空间尺度的特征表示,促进了多尺度信息的有效融合,进而提升了网络对复杂输入数据的处理能力。此外,为了强化网络对正样本的学习,该方

法用变焦距损失(varifocal loss,VFL)替换了传统的二元交叉熵损失(binary cross entropy loss,BCE loss),以期更有效地捕捉和区分目标特征,进而增强模型的检测精度。

## 2　方法

### 2.1　YOLOv8 算法

YOLOv8,作为 YOLOv1 系列的延续与革新,代表了一种先进的端到端无锚框目标检测框架,其设计旨在实现高效性和精确性之间的平衡,特别适配于资源受限情况下移动设备应用。该算法提供了包括 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x 在内的多种不同规模的模型选择,以适应不同的应用场景需求。从网络整体架构来看,YOLOv8 可被大致划分为输入端、骨干网络、混合特征层和预测层 4 个关键组成部分,如图 1 所示。具体而言,输入端包含马赛克数据增强(增强数据多样性)、自适应锚框计算(动态调整锚框的尺寸和位置以提升模型在多尺度目标检测中的表现)、自适应图片缩放(增强模型鲁棒性)等功能模块。骨干网络部分,则采用了专注于提取图像特征的结构,其中包含注意模块、跨阶段局部网络(cross stage partial network,CSPNet)和空间金字塔池化(spatial pyramid pooling,SPP)等关键组件。注意模块在图像进入骨干网络之前,通过将每张图像分割为间隔取值的方式生成 4 张互补图像,并通过卷积操作确保无信息丢失的同时实现 2 倍下采样,从而获得更加紧凑且丰富的特征表示。CSPNet 中的 C2f 模块可以提供丰富的梯度流信息,而 SPP 模块的主要任务是将不同尺寸的输入特征图转换为固定尺寸的特征向量,可以有效捕捉图像中从细节到全局的多层次信息,以适应不同尺度的目标检测需求。在混合特征层,YOLOv8 采取了路径聚合网络和特征金字塔网络的融合策略,旨在实现多尺度特征的有效整合,确保图像信息在不同层次间的流畅传递。预测层则通过解耦检测头设计,考虑到分类任务和定位任务的不同特点,针对这两种任务分别使用不同的分支进行处理与计算。这种解耦策略有效地减少了任务之间的干扰,保证了每个任务能够在其特定的目标上得到优化,从而提升了整体模型的精度与效率。在损失函数的设计上,YOLOv8 引入了动态正负样本分配机制,结合了分类损失、回归损失和分布焦点损失,旨在优化模型的综合性能,特别是在复杂场景下的目标识别能力。

### 2.2　本文算法

本文模型总体结构主要可分为输入端、骨干网络、混合特征网络层和预测层 4 个部分,如图 2 所示。与 YOLOv8 模型相比,主要的改进点有:第 1 步,将骨干网络及混合特征网络层中的 C2f 模块替换为 C2f_DConv,该模块能在保证模型为不同尺度的输入样本动态地调整卷积核的权重,从而提升检测性能;第 2 步,将原始 SPPF 模块替换为 AIFI,通过这种基于注意的尺度内特征交互机制使模型能够更有效地处理和融合重要的特征信息;第 3 步,修改检测头为 Dyhead,使模型能够自适应于多尺度的目标检测;最后,采用 DIoU 损失函数,它综合考虑了边界框之间的重叠区域和目标框的中心距离,使回归过程专注于高质量锚框,同时加快了训练收敛。

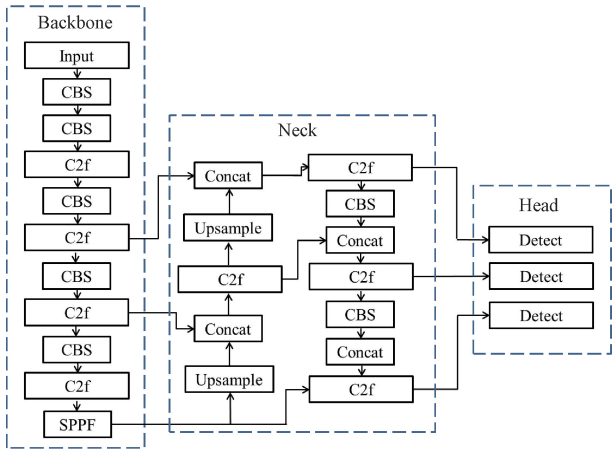



图 1　YOLOv8 模型结构图

Fig. 1　The overall structure of YOLOv8

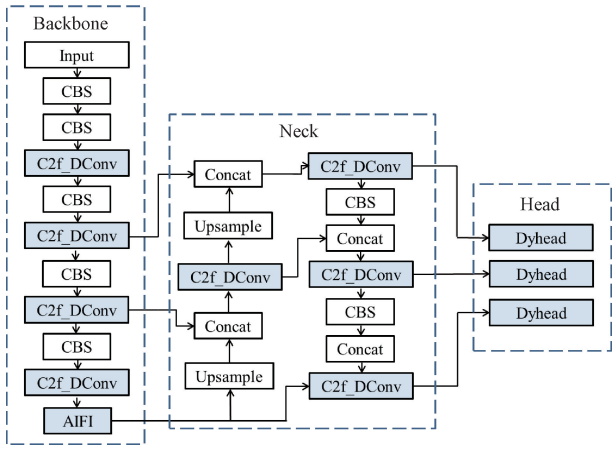



图 2　本文模型结构图

Fig. 2　The overall structure of our model

### 2.3　动态卷积

动态卷积(dynamic convolution,DConv)的核心思想是动态地生成卷积核,而非使用固定的卷积核。通过引入更多的计算灵活性和适应性来增强卷积操作的表达能力,进而提升模型的性能。

首先,需要定义动态感知器,后将其应用于卷积。设静态感知器为:

$$y = g(\boldsymbol{W}^{\mathrm{T}}x + \boldsymbol{b}) \tag{1}$$

其中,$\boldsymbol{W}$ 和 $\boldsymbol{b}$ 分别代表权重矩阵和偏置向量,$g$ 为激活函数。

则可以通过聚合多个线性函数 $\{\tilde{\boldsymbol{W}}_k^{\mathrm{T}}x + \tilde{\boldsymbol{b}}_k\}$ 来定义动态感知器,如下:

$$y = g(\tilde{\boldsymbol{W}}^{\mathrm{T}}(x)x + \tilde{\boldsymbol{b}}(x)) \tag{2}$$

$$\widetilde{\boldsymbol{W}}(x)=\sum_{k=1}^{K}\pi_k(x)\widetilde{\boldsymbol{W}}_k,\widetilde{\boldsymbol{b}}(x)=\sum_{k=1}^{K}\pi_k(x)\widetilde{\boldsymbol{b}}_k \tag{3}$$

$$s.t.\ 0\leqslant\pi_k(x)\leqslant 1,\sum_{k=1}^{K}\pi_k(x)=1 \tag{4}$$

其中，$\pi_k$ 代表第 $k$ 个线性函数 $\widetilde{\boldsymbol{W}}_k^{\mathrm{T}}x+\widetilde{\boldsymbol{b}}_k$ 的注意力权重。

动态卷积层与动态感知器在机制上具有相似性，其关键在于采用 $K$ 个拥有相同核尺寸且输入输出维度一致的卷积核进行计算。在这一过程中，注意力机制发挥着重要作用，通过为卷积核分配权重来实现特征的聚合。聚合后的结果通过批归一化和激活函数进一步处理，以增强模型的非线性表达能力和训练稳定性，如图 3 所示。这一系列操作共同作用，有效提升了模型在特征提取过程中的灵活性与针对性，使得网络能够更好地适应复杂场景下的任务需求。

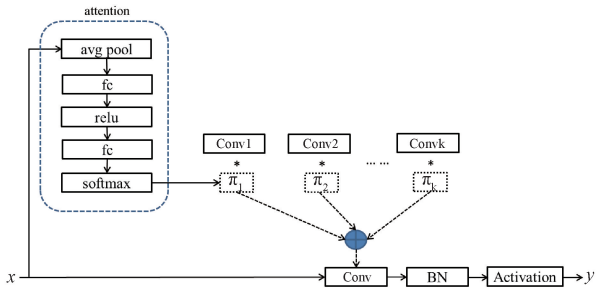



图 3　动态卷积结构图

Fig. 3　The structure of dynamic convolution

## 2.4　基于注意机制的尺度内特征交互

引入基于注意力机制的尺度内特征交互(AIFI)模块，旨在提升目标检测过程中特征提取的效率和效果。其重点在同一尺度内部的特征融合，这有助于捕捉更细粒度的信息，进而促进更丰富的特征融合，并提高特征融合的计算效率，以提升模型整体性能。具体而言，受 Transformer 中编码器结构的启发，AIFI 的核心在于集成自注意力机制，以高效地解析和理解图像中的高层次特征。

该过程的数学实现首先依赖于多头自注意力机制，该机制通过并行处理不同的注意力头，有效捕捉输入序列中元素之间的依赖关系，并增强模型在不同上下文信息处理上的能力；随后，数据通过前馈神经网络进行非线性变换，从而进一步提取和优化特征表示；再将输出 Reshape 回二维，如图 4 所示。

## 2.5　动态卷积头

本文采用一种名为“动态头”(dynamic head，Dyhead)的新型检测头，将目标检测头的输入视为一个具有层级×空间×通道 3 个维度的张量，用于统一尺度感知(scale-awareness)、空间感知(spatial-awareness)和任务感知(task-awareness)。这意味着 Dyhead 能够处理在图像中共存的多种不同尺度的对象，同时考虑到对象通常在不同视角下呈现出不同的形状、旋转和位置；也可以根据任务的具体需求自适应地选择。

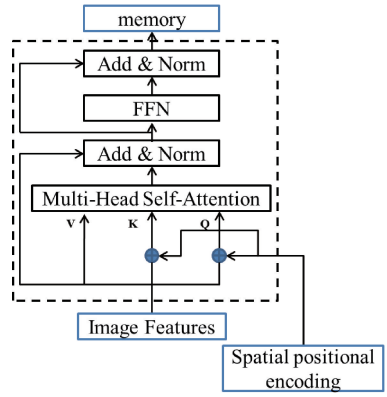



图 4　基于注意机制的尺度内特征交互结构图

Fig. 4　The structure of attention-based intra-scale feature interaction

具体而言，尺度感知注意力模块专注于特征层级($\boldsymbol{L}$)，通过将特征金字塔缩放到相同的尺度来形成一个三维张量 $\boldsymbol{F}\in\boldsymbol{R}^{L\times S\times C}$，然后将其作为动态头的输入。接着，多个动态头块(Dyhead blocks)被依次堆叠，每个头块整合了尺度感知、空间感知和任务感知注意力机制。这个模块专注于空间位置($\boldsymbol{S}$)，通过对特征进行聚合来关注图像中的区别性区域。最终，动态头的输出适应并服务于目标检测领域多种任务需求，它不仅限于单一功能，而是能够灵活支持分类任务以及中心点或边界框的回归任务等。这个模块专注于通道($\boldsymbol{C}$)，动态地开启或关闭特征通道以支持不同的任务。图 5 中描述了每个注意力模块的详细实现。尺度感知注意力($\pi_L$)，空间感知注意力($\pi_S$)和任务感知注意力($\pi_C$)分别对应着不同的模块，每个模块针对的是特征张量 $\boldsymbol{F}$ 的不同维度(层级 $\boldsymbol{L}$、空间 $\boldsymbol{S}$、通道 $\boldsymbol{C}$)。$\pi_L$ 模块使用了平均池化、1×1 卷积、ReLU 激活函数和 hard sigmoid 函数；$\pi_S$ 模块包括偏移量学习和 3×3 卷积；$\pi_C$ 模块则通过全连接层、ReLU 激活函数、以及归一化操作来处理，如图 5 所示。通过上述操作，本文将广义形式的注意力函数($\boldsymbol{W}(\boldsymbol{F})=\pi(\boldsymbol{F})\cdot\boldsymbol{F}$)转换为了 3 个序列注意力，每个注意力专注于其中的一个维度：

$$\boldsymbol{W}(\boldsymbol{F})=\pi_C(\pi_S(\pi_L(\boldsymbol{F})\cdot\boldsymbol{F})\cdot\boldsymbol{F})\cdot\boldsymbol{F} \tag{5}$$

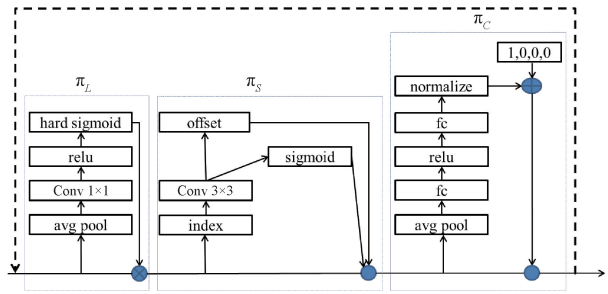



图 5　动态检测头结构图

Fig. 5　The structure of dynamic head

## 2.6　损失函数

DIoU 损失函数(distance intersection over union loss，

DIoU)是一种在目标检测任务中常用的损失函数,用于优化边界框的位置。这种损失函数是 IoU 损失函数的改进版,其不仅考虑了边界框之间的重叠区域,还考虑了它们中心点之间的距离,从而提供更加精确的位置优化。计算步骤如下:

首先,计算 IoU:

$$IoU = \frac{I}{U} \tag{6}$$

其中,$I$ 是两个边界框 A 和 B 的交集面积;$U$ 是两个边界框 A 和 B 的并集面积。

然后,计算框的中心距离,设预测框的中心为 $(x_p, y_p)$,真实框的中心为 $(x_g, y_g)$,则中心点距离 $d$ 的计算公式为:

$$d = \sqrt{(x_p - x_g)^2 + (y_p - y_g)^2} \tag{7}$$

再计算包围预测框和真实框的最小闭合矩形(称为最小闭合框),并求出其对角线长度。归一化中心距离为 $\frac{d}{c}$,以确保距离的比例适应不同大小的边界框。

最后,DIoU 可以通过式(8)计算:

$$DIoULoss = 1 - IoU + d^2/c^2 \tag{8}$$

其中,$d^2/c^2$ 表示中心点距离的归一化平方,这样确保了距离项在损失函数中占有合适的权重。

## 3　实验与结果分析

### 3.1　实验环境与参数配置

本文实验采用相同的软硬件环境,其中,软件环境为 Python 3.11.7、PyTorch 2.1.2、CUDA12.1;硬件环境为 Nvidia A800 80 GB 显存显卡。实验使用相同的参数设置,如表 1 所示。

**表 1　实验参数配置**

**Table 1　Experimental parameter configuration**

| 参数名 | 数值 |
|---|---|
| 训练轮次 | 200 |
| 批大小 | 16 |
| 图像增强方法 | Mosaic |
| 优化器 | SGD |
| 初始学习率 | 0.01 |
| 动量 | 0.937 |
| 权重衰减 | 0.000 5 |

### 3.2　实验数据集

VisDrone-DET2019 基准数据集[15]包含 10 209 张静态图像,拍摄自多样化且无任何限制的场景,其在中国 10 个不同城市,通过多种无人机在多个角度、场景和背景下进行拍摄完成。包含 10 类数据(pedestrian、person、car、van、bus、truck、motor、bicycle、awning-tricycle 和 tricycle),数据整体分布不均,目标尺度差异较大。数据集被划分为 3 个部分:训练集包含 6 471 张图像,验证集由 548 张图像组成,测试集则包括 3 190 张图像。该数据集较为复杂,挑战性很强。其中,目标的密集分布特性、部分遮挡(如图 6(a)所示),光照条件的变化(如图 6(b)所示)以及目标物体尺寸的差异性(如图 6(c)所示)是主要难点。此外,无人机的不同高度以及相机方向的差异导致了观察视角的显著变化,进一步加剧了识别和定位的难度。

(a) 密集分布特性、部分遮挡示例
(a) Example of dense distribution and partial occlusion

(b) 光照条件的变化示例
(b) Example of changing in lighting condition

(c) 尺寸差异示例
(c) Example of size variability

图 6　数据集图片示例

Fig. 6　Dataset image illustration

### 3.3　评价指标

为了全面评估所提出算法在目标检测任务中的性能,本文综合采用了准确率(precision,P)、召回率(recall,R)以及平均精确度均值(mean average precision,mAP)等多项评价指标[16]。这些指标从不同的角度对模型的表现进行量化分析,以对所提出算法的优缺点做出客观评价。其中,P 量化了算法正确识别出的所有正样本与所有预测为正样本的数量之间的比率,以此反映其在预测正类时的准确性;R 则衡量了算法成功识别出所有实际正样本的能力,即正确预测的正类个数占所有真实正类样本的比例,体现了其在覆盖所有真正样本方面的表现;结合考虑上述两个指标

的平均值,并针对每个类别进行此操作后求平均,最终得到的 mAP 值能够全面反映整个检测网络在多类别情况下的整体性能。

P 和 R 的定义式如下:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

在上述公式中,定义了 4 个关键指标来评估分类模型的性能:TP(真正例),代表真实属于正类别且被正确预测为正类别的样本数量;TN(真反例),指真实属于负类别且被准确预测为负类别的样本数量;FP(假正例),表示实际属于负类别却被错误判断为正类别的样本数量;FN(假反例),则为真实属于正类别但被误判为负类别的样本数量。

在目标检测任务中,mAP@0.5 是用于衡量 IoU 阈值为 0.5 时,模型对目标类别的检测精度。此指标有效反映了网络在各种目标类别上的综合分类性能;mAP@0.5∶0.95 则是一个更为精确的性能评估标准。它通过在 0.5～0.95 范围内,步长为 0.05 的多个 IoU 阈值下,计算目标检测精度的平均值。该评估方式通过涵盖 10 个不同的 IoU 阈值,提供了一个对模型性能更为全面的考量。

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{11}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{12}$$

式中:$n$ 表示数据集的类别数。

### 3.4　实验结果

在训练集下训练 200 个轮次后,使用验证集验证的结果如图 7 所示。从图 7 中可以看出:3 个损失函数在开始训练后不久就迅速收敛,在 120 轮训练周期后趋于稳定。4 个性能指标(精度、召回率、mAP50、mAP50-95)则呈现相反趋势,在训练后不久开始迅速攀升,之后,逐步增长直至稳定。最终,本文算法精确率可达 60.4%,召回率 47.0%,mAP50 为 50.1%,mAP50-95 为 30.7%。此外,本文方法 PR 曲线十分平滑且大部分类别靠近图像右上角(如图 8 所示),进一步证明了本文算法的优越性能。

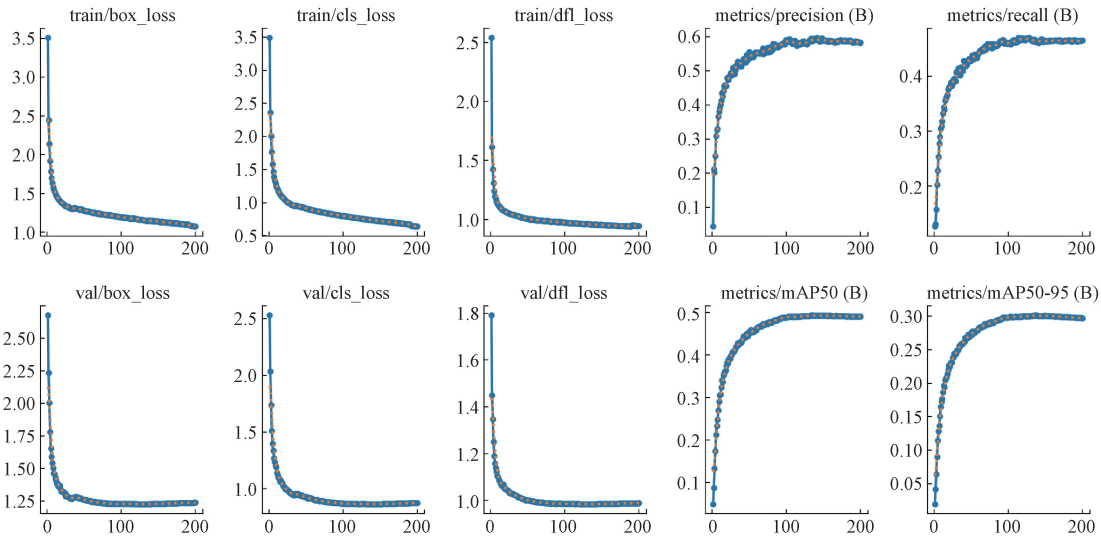



图 7　性能指标变化趋势图

Fig. 7　The graph of trends in performance metrics

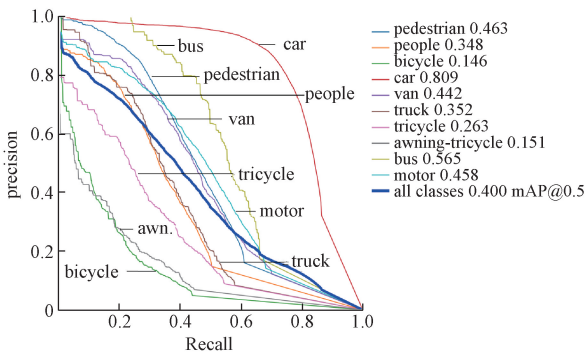


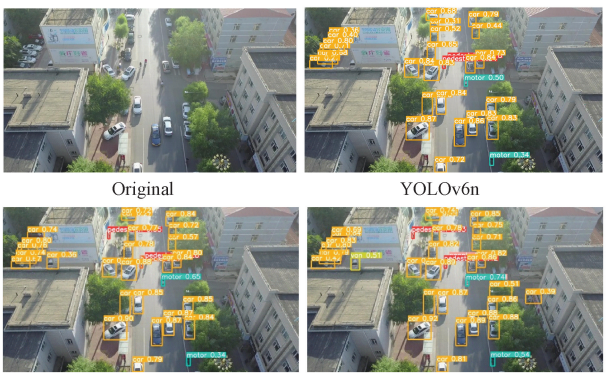

图 8　P-R 曲线

Fig. 8　Precision-recall curve

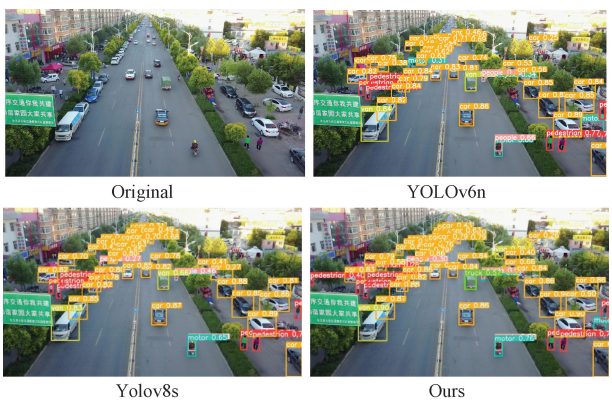
此外,使用测试集图片分别对 YOLOv5n、YOLOv8s 与本文模型做测试,部分结果可视化如图 9 所示。图 9 中选取了城市公路、暗光街道等具有代表性的场景作为测试照片。综合来看,在包含多个种类的目标街道复杂场景中,YOLOv5n 和 YOLOv8s 模型在识别街道远处部分行人与摩托车等小尺度或遮挡目标时易出现漏检误检情况,且对于一些密集排列目标的置信度较低;而改进后的模型不仅显著提升了检测框的定位精度,对于每个目标的分类置信度也得到了明显提高,这主要得益于本文采用的 DConv 模块,使得模型为不同尺度的输入样本动态地调整卷积核的权重,以适应不同目标的特征,也应证了改进后模型的特征提取能力更强,对无人机图像多尺度目标的检

测效果较原始模型得到了有效的改善。此外，对于一些有重叠、遮挡的目标，也能有效检测，这主要得益于本文采用的 AIFI 模块，使模型能够更有效地处理和融合重要的特征信息，同时证明了本文算法具有较强的鲁棒性和准确性。


(a) 示例一
(a) Example one




(b) 示例二
(b) Example two

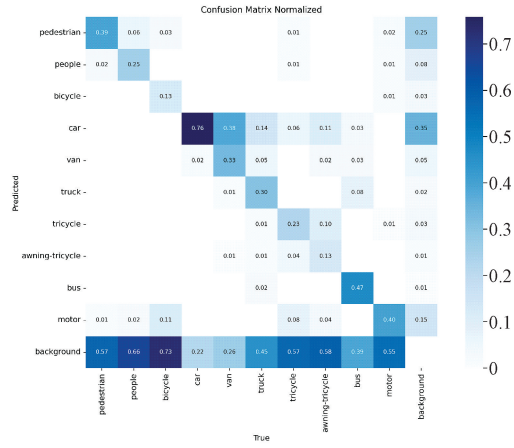



(c) 示例三
(c) Example three

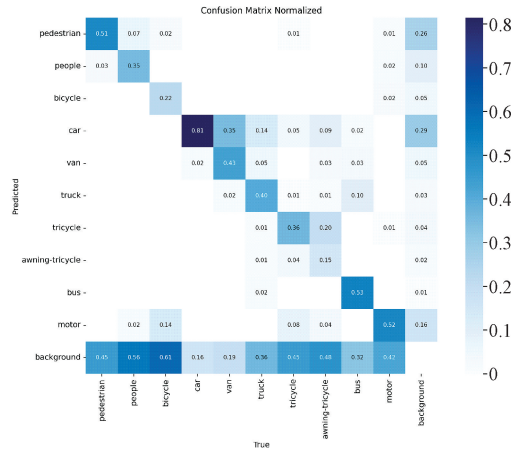
图 9　模型检测效果对比

Fig. 9　Comparison of model performance

进一步地，本文将模型输出的混淆矩阵与 YOLOv8s 进行了对比。由图 10 可知，数据集中 10 个类别的识别精度在本文模型上均有不同程度的提升。其中，Tricycle 类别的提升幅度最大(7%)。与其他目标相比，Tricycle 的尺寸和密度都较小，且很多特征不清晰，易出现遮挡情况，对这类目标的检测效果的显著提升，也表明了本文模型鲁棒性较强，可以有效地提高小目标的检测精度，进而提升多尺度目标检测整体性能。




(a) YOLOv8s
(a) YOLOv8s




(b) 本文模型
(b) Our model

图 10　混淆矩阵对比

Fig. 10　Comparison of confusion metrics

### 3.5　消融实验

为了验证本文所提出网络的先进性能，设计了消融实验来证明，选取两个具有代表性的指标(召回率 R 和平均精度 mAP50)为参照。实验结果如表 2 所示。可以看出，在基线网络 YOLOv8s 中加入 DConv、AIFI、Dyhead 后模型的 mAP50 分别提升了 0.3%、0.8%以及 2%。其中，加入 AIFI 模块后模型的 mAP50 提升了 0.8%。本文通过实验证明了注意力机制在深度学习模型中的应用能够显著优化卷积层输出的特征图，通过学习并赋予特征图中的各个元素以不同的权重，从而强化了模型对关键特征的选择与提取能力，同时削弱了干扰特征的影响，进而输出特征表达更好的特征图。这一过程对于提升模型的整体检测

精度起到了至关重要的作用。引入 Dyhead 模块后,也观察到了显著的性能提升,具体表现为召回率的大幅增长(提升 1.6%),这充分证实了 Dyhead 模块能够有效增强网络对目标特征的捕获能力,使其对于不同尺度的目标检测均表现出优异性能,从而有效降低了漏检现象的发生。此外,模型的平均精度(mAP)在引入 Dyhead 后提高了 2%,这进一步证明了该模块在提升模型性能方面的有效性与重要性。在分别删减 DConv、AIFI、Dyhead 模块后,模型的 mAP50 均有不同程度的下降,分别为 1.0%、1.7%及 1.9%。降幅最为显著的是 Dyhead 模块,表明其在模型整体性能中占据了较为核心的地位,对模型的决策起到了关键作用。综合来看,各模块的删减均对模型性能产生了一定的负面影响,这些实验结果进一步验证了各个模块在优化目标检测模型中的不可或缺性及其相互之间的协同效应。最后,对模型同时使用 4 种改进策略训练,模型的召回率及平均精度值都达到了最优,验证了综合使用四种改进策略的有效性。在面临复杂场景下的多尺度目标检测任务时,该模型具备明显优势。

**表 2 消融实验**

**Table 2 Ablation experiment**

| DConv | AIFI | Dyhead | DIoU | mAP50/% | R/% |
|---|---|---|---|---|---|
| × | × | × | × | 47.8 | 45.8 |
| √ | × | × | × | 48.1 | 45.8 |
| × | √ | × | × | 48.6 | 45.9 |
| × | × | √ | × | 49.8 | 47.4 |
| √ | √ | × | √ | 48.2 | 46.1 |
| √ | × | √ | √ | 48.4 | 46.2 |
| × | √ | √ | √ | 49.1 | 46.0 |
| √ | √ | × | √ | 48.3 | 46.2 |
| √ | √ | √ | √ | **50.1** | **47** |

### 3.6 不同超参数设置下的对比实验

优化器的选择对于模型的训练效率和最终性能有着至关重要的影响。针对优化器和学习率设置,本文开展了几组对比试验,如表 3 所示,旨在全面考察不同优化器与学习率策略的优劣,并最终选择带动量的 SGD。在这一系列实验中,可以发现,虽然某些优化器如 Adam 和 AdamW 在理论层面上展示了较好的收敛性和稳定性,然而,在实际应用中,特别是对于需要高效处理大量数据、追求快速收敛的复杂场景时,带有动量项的 SGD 展现出了独特的优势。具体而言,带有动量项的 SGD 不仅能够在训练过程中加速收敛,显著缩短达到最优解所需的迭代次数,而且还能有效避免陷入局部最优解,这对于提升模型的泛化能力和最终性能至关重要。此外,动量项的存在有助于优化器在梯度方向上的累积,从而使得模型能够在复杂优化景观中更加稳健地移动,避免了随机梯度的剧烈波动,进而提高了训练的稳定性。

**表 3 超参数实验**

**Table 3 Hyper-parameter experiment**

| 优化器 | 学习率 | P/% | R/% | mAP50/% | mAP50-95/% |
|---|---|---|---|---|---|
| Adam | 0.001 | 58.6 | 46.5 | 49.6 | 30.5 |
| AdamW | 0.001 | 57.1 | 45.6 | 48.1 | 29.2 |
| Adamax | 0.001 | 58.9 | 44.5 | 48.9 | 27.5 |
| SGD+动量 | 0.01 | **60.4** | **47.0** | **50.1** | **30.7** |

### 3.7 对比实验

为了验证本文所提出网络的先进性能,设计了对比实验来证明,选取 4 个具有代表性的指标(精度 P、召回率 R、平均精度 mAP50 和 mAP50-95)为参照。实验结果如表 4、5 所示。从表 4 的对比实验结果来看,本文提出的模型相较于其他经典模型在上述 4 个评价指标下均具有最好的检测性能。相比于基础网络—YOLOv8s,本文算法在精确率上提升了 3.7%,在召回率上提升了 1.2%,mAP50 提升了 2.3%,mAP50-95 提升了 1.4%。这主要得益于本文综合使用了动态卷积模块、基于注意机制的尺度内特征交互以及动态检测头,使得模型能够根据不同输入样本尺寸自适应地调整,提高其在处理复杂场景时的灵活性和精确度。同时,损失函数的改进也加速了模型收敛,优化了训练过程。与 YOLOv5n 相比,本文模型在 R 和 mAP50 两个评价指标上涨幅明显,分别提升了 7%和 8.6%。此外,虽然本文模型较基线模型增加了一定程度的参数量,但对于验证集图片的平均推理时间为 15.6 ms,仍符合目标检测的实时性要求。另外,通过对调整 IoU 阈值的设置,可以看出,较高的 IoU 阈值可以提高准确率,但会导致召回率下降。在实际应用中,选择合适的 IoU 阈值需要根据具体的应用场景来决定,例如:在安全监控系统中,可能需要更高的准确率以避免误报;而在自动驾驶系统中,则可能需要更高的召回率以确保不漏掉任何潜在的危险物体。

**表 4 对比实验**

**Table 4 Comparative experiment**

| 方法 | P/% | R/% | mAP50/% | mAP50-95/% |
|---|---|---|---|---|
| YOLOv5n | 51.6 | 40.0 | 41.5 | 25.0 |
| YOLOv5s | 57.3 | 45.8 | 46.0 | 28.3 |
| YOLOv6n | 49.7 | 38.5 | 39.3 | 23.6 |
| YOLOv8n | 51.5 | 38.4 | 40.0 | 24.3 |
| YOLOv8s | 56.7 | 45.8 | 47.8 | 29.3 |
| Ours | **60.4** | **47.0** | **50.1** | **30.7** |

表 5　不同类别 AP 对比

Table 5　Comparison of AP in different categories　　%

| 方法 | Ped. | People | Bicycle | Car | Van | Truck | Tricycle | Awn. | Bus | Motor |
|---|---|---|---|---|---|---|---|---|---|---|
| MSA-YOLO[17] | 33.4 | 17.3 | 11.2 | 76.8 | 41.5 | 41.4 | 14.8 | 18.4 | 60.9 | 31.0 |
| SDS-YOLO[18] | 46.7 | 35.4 | 14.4 | 82.0 | 45.1 | 35.8 | 26.5 | 12.7 | 54.3 | 47.4 |
| YOLOv7-tiny | 39.6 | 36.2 | 9.6 | 77.5 | 38.3 | 30.3 | 19.4 | 10.2 | 49.6 | 44.5 |
| YOLOv8n | 40.0 | 46.3 | 34.8 | 80.9 | 44.2 | 35.2 | 26.3 | 15.1 | 56.5 | 45.8 |
| YOLOv8s | 55.5 | 43.2 | 23.2 | 85.4 | 52.0 | 43.9 | 35.1 | 19.0 | 64.7 | 56.4 |
| Ours(IoU=0.7) | **56.7** | **45.5** | **25.8** | **86.0** | **54.2** | **44.4** | **39.5** | **21.7** | **68.7** | **58.5** |
| Ours(IoU=0.5) | 45.6 | 45.8 | 20.5 | 84.8 | 49.9 | 45.0 | 42.6 | 19.7 | 64.2 | 57.3 |
| Ours(IoU=0.9) | 57.0 | 45.8 | 26.8 | 84.7 | 54.6 | 45.0 | 40.3 | 22.0 | 68.8 | 57.9 |

## 4　结　　论

本文针对复杂背景下的多尺度目标检测存在的精度低、漏检率严重等问题，提出了一种基于 YOLOv8s 算法改进的无人机图像多尺度目标检测方法。该方法综合使用了 DConv、AIFI 模块以及动态检测头，能够动态的调整不同尺度目标的特征感受视野，从而提升对不同尺度弱特征的感知能力，进而提升检测精确度并降低漏检和误检概率。同时，修改了损失函数，进一步提升了检测精度，优化了训练过程。在公开数据集 VisDrone2019-DET 上的实验结果表明，与原网络相比，本文优化后的模型精确率提升了 3.7%；召回率提升了 1.2%；平均精度 mAP50 提升了 2.3%，在各项指标中都展现出明显优势。未来的任务中，将深入探索在确保计算效率的前提下，提升多尺度目标检测的准确性与效率，特别是在资源约束严格的无人机航拍环境中的应用。具体而言，将重点分析并提出算法，在有效降低模型的计算复杂度的同时减少精度的损失，进而加速推理过程，满足实际应用场景中对实时性和资源利用效率的高要求。以推动无人机航拍技术在复杂环境监测、环境评估、安全巡查等领域的发展，实现高效、精准的目标检测与识别，为相关行业提供有力的技术支撑。

## 参考文献


[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014: 580-587.

[2] GIRSHICK R. Fast R-CNN [C]. 2015 IEEE International Conference on Computer Vision, 2015: 1440-1448.

[3] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi-box detector[C]. 14th European Conference on Computer Vision, Amsterdam, 2016: 21-37.

[6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327.

[7] 陈占龙，李双江，徐永洋，等. 高分影像密集建筑物 CorregYOLOv3 检测方法[J]. 测绘学报，2022，51(12)：2531-2540.
CHEN ZH L, LI SH J, XU Y Y, et al. CorregYOLOv3 detection method for high resolution image dense buildings[J]. Journal of Surveying and Mapping, 2022, 51(12): 2531-2540.

[8] 赵耘彻，张文胜，刘世伟. 基于改进 YOLOv4 的无人机航拍目标检测算法[J]. 电子测量技术，2023，46(8)：169-175.
ZHAO Y CH, ZHANG W SH, LIU SH W. UAV aerial target detection algorithm based on improved YOLOv4[J]. Electronic Measurement Technology, 2023, 46(8): 169-175.

[9] 徐光达，毛国君. 多层级特征融合的无人机航拍图像目标检测[J]. 计算机科学与探索，2023，17(3)：635-645.
XU G D, MAO G J. Multi-level feature fusion for unmanned aerial vehicle aerial image object detection[J]. Computer Science and Exploration, 2023, 17(3): 635-645.

[10] ZHANG J, WAN G Y, JIANG M, et al. Small object detection in UAV image based on improved YOLOv5[J]. Systems Science & Control Engineering, 2023,

11(1): 2247082.

[11] 李安达, 吴瑞明, 李旭东. 改进 YOLOv7 的小目标检测算法研究[J]. 计算机工程与应用, 2024, 60(1): 122-134.
LI AN D, WU R M, LI X D. Research on improving the small object detection algorithm of YOLOv7[J]. Computer Engineering and Applications, 2024, 60(1): 122-134.

[12] 张徐,朱正为,郭玉英,等.基于 cosSTR-YOLOv7 的多尺度遥感小目标检测[J]. 电光与控制, 2024, 31(4): 28-34.
ZHANG X, ZHU ZH W, GUO Y Y, et al. Multi scale remote sensing small target detection based on cosSTR-YOLOv7[J]. Electronics Optics & Control, 2024, 31(4): 28-34.

[13] WANG G, CHEN Y F, AN P, et al. UAV-YOLOv8: A small object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

[14] 张河山, 范梦伟, 谭鑫, 等. 基于改进 YOLOX 的无人机航拍图像密集小目标车辆检测[J]. 吉林大学学报(工学版), 2023(12): 1-13.
ZHANG H SH, FAN M W, TAN X, et al. Dense small target vehicle detection in drone aerial images based on improved YOLOX[J]. Journal of Jilin University(Engineering Edition), 2023(12): 1-13.

[15] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection challenge results[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshops(ICCVW), 2019: 213-226.

[16] 杨瑞君, 张浩, 叶璟. 改进 YOLOv8n 的轻量级遥感图像军用飞机检测算法[J]. 电子测量技术, 2025, 48(1): 154-165.
YANG R J, ZHANG H, YE J. Improved YOLOv8n lightweight remote sensing image military aircraft detection algorithm[J]. Electronic Measurement Technology, 2025, 48(1): 154-165.

[17] 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法[J]. 航空学报, 2023, 44(5):273-283.
MAO G T, DENG T M, YU N J. Object detection algorithm for unmanned aerial vehicle aerial images based on multi-scale segmentation attention[J]. Journal of Aeronautics, 2023, 44(5): 273-283.

[18] 王恒涛, 张上, 陈想, 等. 轻量化无人机航拍目标检测算法[J]. 电子测量技术, 2022, 45(19): 167-174.
WANG H T, ZHANG SH, CHEN X, et al. Lightweight drone aerial target detection algorithm[J]. Electronic Measurement Technology, 2022, 45(19): 167-174.

## 作者简介

**詹雨飞**(通信作者),硕士,助理工程师,主要研究方向为深度学习、目标检测。

E-mail:unayufei99@163.com