

DOI: 10.13382/j.jemi.B2407473

# 无人机对地小目标检测方法研究\*

苏雨蕾<sup>1</sup> 黄丹丹<sup>1</sup> 刘智<sup>1,2</sup> 田成军<sup>1</sup>

(1. 长春理工大学电子信息工程学院 长春 130022;

2. 长春理工大学空间光电技术国家地方联合工程研究中心 长春 130022)

**摘要:** 无人机图像中的小目标检测是研究的重难点之一。与大目标相比,小目标特征较少,更易受到遮挡和复杂背景的干扰,针对该问题,提出一种基于YOLOv7-tiny的多模型融合目标检测网络YOLO-DA。首先,增加小目标和极小目标检测层,提高网络对小目标特征的学习能力;其次,引入空间自适应特征融合ASFF-L检测头,通过学习空间过滤冲突信息来抑制不同尺度特征的不一致性,实现多尺度特征的自适应融合;最后,引入DCNS可变形卷积并设计了扩展变形建模范围的调制机制,增强模型的建模能力,降低遮挡重叠等对检测的影响。经试验验证,提出的方法在Visdrone2019数据集上实现了44.7%的平均精度及71 fps的推理速度,平均精度较基线算法提高了9.7%,模型内存为63.8 M,能够实现实时检测。通过消融、对比实验表明YOLO-DA在无人机航拍图像检测方面明显减少了误检和漏检问题,具有更高的检测性能,且算法参数数量和计算量可以满足无人机等边缘设备的实时检测需求。

**关键词:** 无人机;目标检测;YOLOv7-tiny;ASFF-L模块;DCNS卷积

**中图分类号:** TP391.4;TN919.5 **文献标识码:** A **国家标准学科分类代码:** 520.20

## Research on detection methods of small targets on ground by UAV

Su Yulei<sup>1</sup> Huang Dandan<sup>1</sup> Liu Zhi<sup>1,2</sup> Tian Chengjun<sup>1</sup>

(1. School of Electronics and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China;

2. National and Local Joint Engineering Research Center of Space Photoelectric Technology, Changchun University of Science and Technology, Changchun 130022, China)

**Abstract:** Small object detection in drone images is one of the key and difficult research areas. Compared with large targets, small targets have fewer features and are more susceptible to interference from occlusion and complex backgrounds. To address this issue, a multi model fusion object detection network YOLO-DA based on YOLOv7 tiny is proposed. Firstly, add layers for detecting small and extremely small targets to enhance the network's ability to learn small target features; Secondly, the spatial adaptive feature fusion ASFF-L detection head is introduced to suppress the inconsistency of features at different scales by learning spatial filtering conflict information, achieving adaptive fusion of multi-scale features; Finally, DCNS deformable convolution was introduced and a modulation mechanism was designed to expand the range of deformable modeling, enhance the modeling ability of the model, and reduce the impact of occlusion overlap on detection. Through experimental verification, the proposed method achieved an average accuracy of 44.7% and an inference speed of 71 fps on the Visdrone2019 dataset. The average accuracy was improved by 9.7% compared to the baseline algorithm, and the model memory was 63.8 M, enabling real-time detection. Through ablation and comparative experiments, it has been shown that YOLO-DA significantly reduces false positives and false negatives in drone aerial image detection, and has higher detection performance. Moreover, the algorithm parameters and computational complexity can meet the real-time detection requirements of edge devices such as drones.

**Keywords:** UAV; target detection; YOLOv7-tiny; ASFF-L module; DCNS convolution

收稿日期: 2024-04-27 Received Date: 2024-04-27

\* 基金项目: 国家自然科学基金(62127813)项目资助

## 0 引言

随着科学技术的快速发展,无人机以其广阔的视角、灵活的特点,被越来越多的应用于各个行业。其中目标检测是视觉信息处理的基础,因此,基于无人机航拍图像的目标检测成为当前的研究热点<sup>[1]</sup>。然而,在无人机对地面目标检测方面,俯视视角容易导致图像细节的丢失,并且由于成像面距离目标较远所以图像中存在大量小目标,这些目标的像素占比更少<sup>[2]</sup>、分辨率低,且常常被遮挡,传统的检测方法难以有效识别,极易造成漏检和误检<sup>[3-4]</sup>。随着深度学习技术在物体检测中的重要应用,基于神经网络的物体识别方法在准确率和速度上均有较大提高。

基于深度学习的目标检测算法可分为两类,分别是两阶段方法和单阶段方法。两阶段算法如 R-CNN<sup>[5]</sup>、Fast R-CNN<sup>[6]</sup>、Faster R-CNN<sup>[7]</sup>等,虽然检测的精度较高,但需要进行两次特征提取处理,因此检测速度较慢,不适合实时检测。而单阶段算法如 YOLO 系列<sup>[8-10]</sup>、SSD<sup>[11]</sup>等,只需一次特征提取即可实现目标检测,在保证一定精度的前提下,检测效率大幅提升,更适合搭载在无人机平台进行实时目标检测。

无人机图像中的小目标检测是研究的重点之一,由于小目标具有特征少、定位精度要求高、易受复杂背景干扰等特点,因此建立一个准确可靠、鲁棒性高的无人机图像小目标检测算法具有重要的研究价值。近年来,许多研究者提出了改进方案和模型来应对这一任务。陈欣等<sup>[12]</sup>对 SSD 算法进行了改进,将分辨率高的浅层特征图与具有丰富语义信息的深层特征图进行融合,对先验框的比例进行调整,提高了遥感图像多尺度目标的检测精度,但对密集分布的小目标检测效果不理想。Felzenszwalb 等在 HOG 的基础上提出另一重要的方法 DPM(deformable part models)检测算法,通过引入零件的概念,对目标的局部部分进行建模,当行人的某些部位被遮挡时,未遮挡部分的检测器不会受到影响,提高了遮挡情况下的检测性能。但在行人被严重遮挡的情况下,DPM 检测算法<sup>[13]</sup>很容易造成漏检。刘芳等<sup>[14]</sup>提出了一种在特征网络中引入多尺度卷积的方法,但改进后算法的计算复杂度增加,并且无法处理目标被遮挡的情况。Akyon 等<sup>[15]</sup>提出了 SAHI 方法,采用切片辅助推理,将图像分割为多个重叠的切片以增加小目标在切片图像中的检测占比,然后对切片图片进行非极大值抑制后处理,得到最终的检测结果。

在 YOLO 算法的基础上,也有许多学者提出了相应的改进方法。邝先验等<sup>[16]</sup>通过引入上下文解耦检测头的方式来改进网络对目标特征的提取能力,由于上下文

解耦检测头需要独立的训练数据和标签,所以需要更多的训练时间进行收敛。戚玲珑等<sup>[17]</sup>则提出了改进的 YOLOv7 目标检测模型,其工作主要集中在改进池化模块、引入 ACmix 注意力机制和更换损失函数,这些措施对小目标的检测性能有很大改进。但是由于该算法需要处理较大尺寸的图片(720×720),且基准模型的参数量、计算量较大,导致计算速度较慢。刘毅等<sup>[18]</sup>提出一种基于注意力机制的 UAST-RCNN 网络,在 Faster R-CNN 网络基础上,采用 Swin-Transformer 作为骨干网络,并接入层级重采样模块,提高了模型对遮挡行人的检测精度,但复杂场景下的检测效果一般。梁礼明等<sup>[19]</sup>通过引入 GSConv 替换 Conv 的方式提高模型的平均精度,GSConv 采用形状注意力机制,增加了计算复杂度和内存占用。陈卫彪等<sup>[20]</sup>采用了 3 种方式改进 YOLOv5s 模型,分别是加深特征融合网络深度、增加额外的小目标检测头以及将普通卷积模块替换为深度可分离卷积模块。

以上文献大多采用轻量化和引入注意力机制的方式来提升无人机平台下的小目标检测。与传统的算法相比,轻量级算法在参数、计算量等方面都得到了极大的优化,但算法的精度却受到了很大的影响;而注意力机制的引入,使得有效信息的比重有所提高,可以更好地适应对弱小目标的探测,但是却没有考虑到网络中存在的参数量多、计算量大等问题。针对上述问题,采用适合于移动设备的轻量化模型 YOLOv7-tiny 为基本框架,使用 YOLO 算法与可变形卷积调节(deformable convolution networks scale, DCNS)模块、自适应特征融合层(adaptive structure feature fusion layer, ASFF-L)结构相结合方式,设计了一种无人机对地小目标的检测方法 YOLO-DA。

YOLO-DA 只保留了 YOLOv7-tiny 的 Backbone 部分,Head 部分则针对小目标特征提取不全的问题,新增小目标和微小目标检测层,取消原网络的大尺寸检测层,因此能有效提高小目标的检测率。此外,针对小目标特征信息少,容易受图像背景信息干扰等问题,YOLO-DA 引入改进后的 DCNS 模块,通过增加可变形卷积的数量和引入调节机制,使前景对应的特征点更多地覆盖整个物体,背景对应的特征点更多的包含语义信息,丰富特征信息的提取能力。最后,针对无人机在不同飞行高度,目标尺度变化较大的问题,引入改进后的 ASFF-L,抑制不同尺度特征在融合空间上可能存在的冲突信息,提高 FPN 的融合效果,进而提高目标检测的效果。

## 1 改进 YOLOv7-tiny 算法网络结构

### 1.1 YOLOv7-tiny 网络结构

YOLOv7-tiny 是一种面向嵌入式设备和移动终端的

轻量级网络架构,以 YOLOv7 为基础,通过对网络架构的简化,实现了一种基于 YOLOv7 的轻量级网络模型,它由输入端 (Input)、骨干网络 (Backbone)、颈部网络 (Neck) 和预测头 (Head) 4 部分组成。

相较于 YOLOv7, YOLOv7-tiny 在结构上的简化使得模型参数量和计算量有了很大程度的降低,但是模型精度也相应的下降。引起这种情况的主要原因包括如下两点:一方面 YOLOv7-tiny 网络中大、中尺度目标较为完善,但小目标像素占比小、包含信息少等特点,使得该网络缺乏针对小目标的尺寸检测;另一方面由于目标尺寸变换跨度较大,原网络中的特征融合并不能满足当前的需求。基于以上分析,本文主要从提高小目标检测能力和增加特征网络中多尺度融合能力两个方面对基准网络 YOLOv7-tiny 进行改进。

### 1.2 YOLOv7-tiny 改进方法

改进后的网络结构如图 1 所示,整体的处理流程与 YOLOv7-tiny 基本类似,其中虚线框内所示部分为本文主要的改进点,集中在特征提取部分和预测头部分。首先,将待检测的图像输入到骨干网络部分对输入数据进行卷积和池化操作,从而提取出数据的高质量特征表示。随后,将提取到的特征信息根据不同的深度得到不同尺寸的特征图,其中较深的网络特征图含有丰富的语义信息,有利于检测大尺度目标;浅层的网络特征图含有输入数据的低级特征表示,如边缘、纹理等,有利于检测小尺度目标。将浅层特征信息通过可变形卷积结构,深层的特征信息通过 ELAN 结构传入 Head 部分进行自适应特征融合操作,提取更加丰富的特征信息。最后由四层检测头输出无人机的目标检测结果。

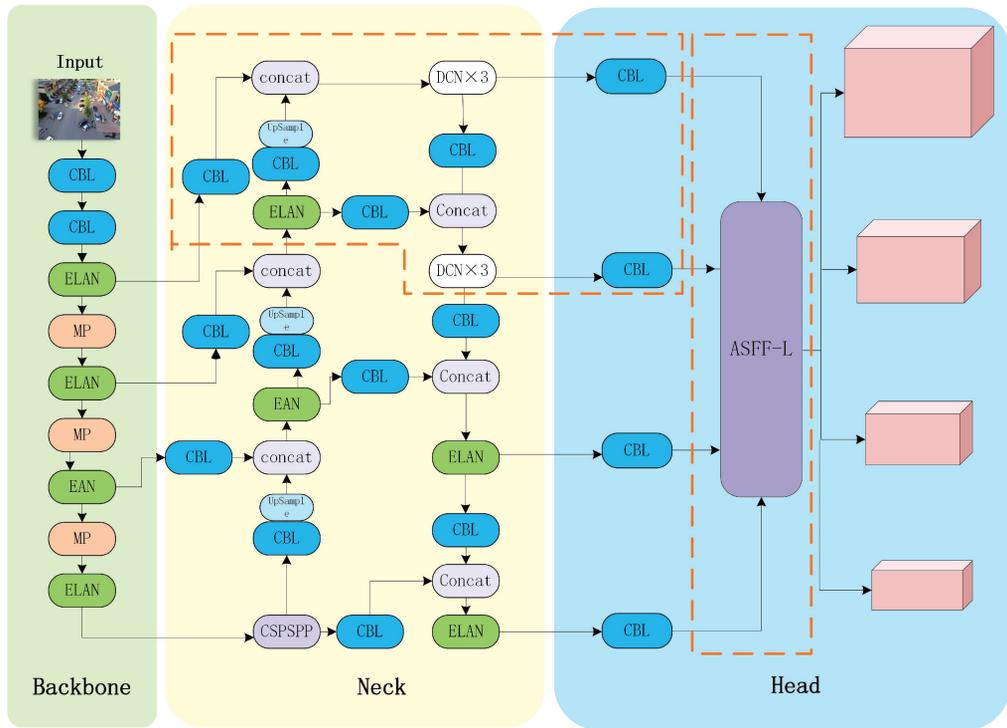


图 1 YOLO-DA 网络模型  
Fig. 1 YOLO-DA network model

#### 1) 优化目标检测层

原 YOLOv7-tiny 网络采用金字塔网络 (feature pyramid network, FPN) 结构,在 Head 部分输出 3 种不同尺度的目标检测层,分别用于检测大、中、小尺度目标的检测。

上述的方法在针对常规尺度目标时有显著效果,但由于无人机在高空飞行,目标尺度普遍较小,会出现小目标特征检测不全的问题。针对以上问题,本文对原网络检测层进行了重构改进。首先,在 Neck 部分由两次下采

样结构变为 3 次下采样结构 (图 2 左侧虚线框所示),引出小目标和微小目标检测层 (图 2 右侧虚线框所示),取消原网络中的极大目标检测层。由于浅层网络的特征图分辨率相对较高,因此更能捕捉到图像的细节特征,从而解决了随着网络深度增加和特征图分辨率降低导致的小目标定位不准确的问题。优化后的目标检测层使模型更好地捕获全局信息以及上下文小目标信息,增强了模型对目标的特征提取能力。具体模型结构如图 2 所示。

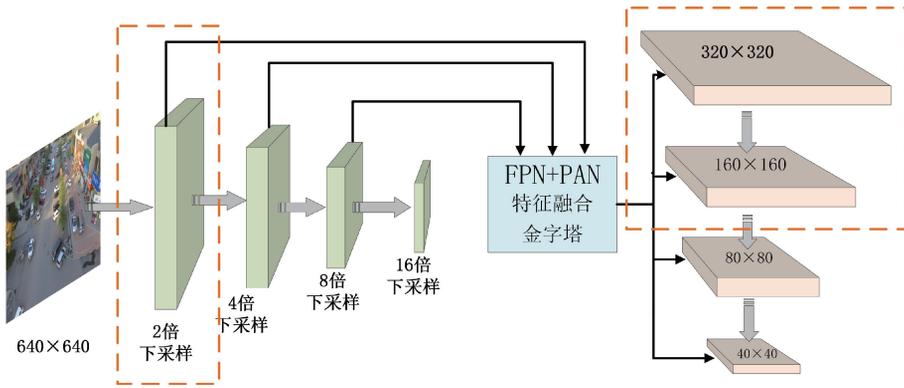


图2 增加小目标尺寸后的特征提取网络模型结构

Fig. 2 The structure of feature extraction network model after increasing the size of small target

2) ASFF-L 模块

锥体特征表示是解决目标检测中尺度变化问题的常用方法。然而,不同特征尺度之间的一致性是基于特征金字塔单次检测的主要限制。针对这一问题,本文提出一种新的四检测头数据驱动的金字塔特征融合方法,

称为自适应空间特征融合层 (ASFF-L)。该算法通过空域中对冲突信息进行过滤,可以消除多源数据之间特征不一致性,在保持特征尺度不变的同时,提升整体的特征表达能力。ASFF-L 网络结构如图 3 所示。

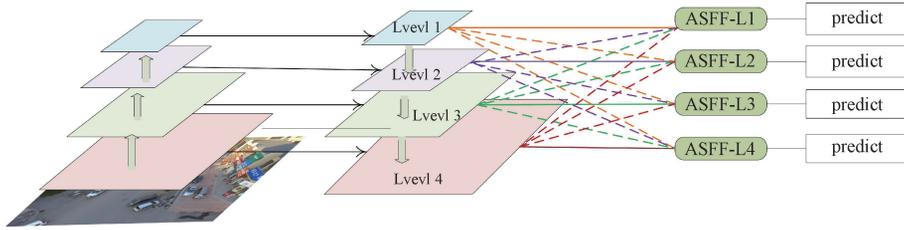


图3 AFSS-L 网络模型

Fig. 3 AFSS-L network model

网络模型中 Detect 部分使用了 ASFF-L 来解决一副图像中同时有大目标和小目标时,出现不同层之间特征不一致性的问题。ASFF-L 与以往的基于元素和或级联的多层次特征融合方法不同,该方法的核心思路是通过对不同尺度的特征进行自适应学习,从而实现不同尺度特征的融合。从图 3 可知,AFSS-L 模型首先通过 FPN 产生 level1 ~ level4 不同尺度的特征图,然后将 level1 ~ level4 尺度图分别再融合成 4 个对应尺度的特征图。以 ASFF-L4 为例,首先将 4 个尺度的特征图的尺寸都重设到 level4 尺度大小,然后学习一个融合权重,这样可以更好地学习不同特征尺度对于预测特征图的贡献。融合流程如式 (1) 所示。再由层级  $n$  缩放到的  $l$  特征图上,位置  $(i, j)$  的特征向量记做  $\mathbf{x}_{ij}^{n \rightarrow l}$ 。融合相应层级  $l$  的特征:

$$\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l, \theta_{ij}^l \in [0, 1].$$

下面对 ASFF 实现特征融合的过程进行分析,讨论未经尺寸变化时的 level 1 中位置  $(x, y)$  处向量的梯度为:

$$\frac{\partial L}{\partial \mathbf{x}_{ij}^1} = \frac{\partial y_{ij}^1}{\partial \mathbf{x}_{ij}^1} \cdot \frac{\partial L}{\partial y_{ij}^1} + \frac{\partial \mathbf{x}_{ij}^{1 \rightarrow 2}}{\partial \mathbf{x}_{ij}^1} \cdot \frac{\partial y_{ij}^2}{\partial \mathbf{x}_{ij}^{1 \rightarrow 2}} \cdot \frac{\partial L}{\partial y_{ij}^2} + \frac{\partial \mathbf{x}_{ij}^{1 \rightarrow 3}}{\partial \mathbf{x}_{ij}^1} \cdot \frac{\partial y_{ij}^3}{\partial \mathbf{x}_{ij}^{1 \rightarrow 3}} \cdot \frac{\partial L}{\partial y_{ij}^3} + \frac{\partial \mathbf{x}_{ij}^{1 \rightarrow 4}}{\partial \mathbf{x}_{ij}^1} \cdot \frac{\partial y_{ij}^4}{\partial \mathbf{x}_{ij}^{1 \rightarrow 4}} \cdot \frac{\partial L}{\partial y_{ij}^4} \quad (2)$$

因为第一层特征图上位置  $(x, y)$  处的特征向量会产生 4 条不同的输出,如图 3 所示,这里采用链式法则将偏导形式展开。特征图的上采样通常通过插值实现,下采样通过池化形式。式(2)可以简化为:

$$\frac{\partial L}{\partial \mathbf{x}_{ij}^1} = \frac{\partial L}{\partial y_{ij}^1} + \frac{\partial L}{\partial y_{ij}^2} + \frac{\partial L}{\partial y_{ij}^3} + \frac{\partial L}{\partial y_{ij}^4} \quad (3)$$

$y_{ij}^l = \alpha_{ij}^l \cdot \mathbf{x}_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot \mathbf{x}_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot \mathbf{x}_{ij}^{3 \rightarrow l} + \theta_{ij}^l \cdot \mathbf{x}_{ij}^{4 \rightarrow l}$  (1)  
式中:  $y_{ij}^l$  是输出特征图  $y^l$  上第  $(i, j)$  个向量。  $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l, \theta_{ij}^l$  是特征图上 4 个层级相对于层级  $l$  的空间权重,由网络自适应地学习得到。  $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l, \theta_{ij}^l$  可以是简单的标量变量,在所有通道之间共享。要求  $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l + \theta_{ij}^l = 1$ ,

假设 level1  $(i, j)$  对应位置特征图上有物体并且为正样本,那其他 level 中可能将相同的位置标记为负样本,这种情况下会导致反向传播时梯度包含了正负两个采样,造成不连续性,影响梯度结果并降低训练时的效率。

因此, SFF 中采取的方式是在每项梯度前面添加一个可学习的参数, 得到如式(4)所示。

$$\frac{\partial L}{\partial x_{ij}^1} = \alpha_{ij}^1 \cdot \frac{\partial L}{\partial y_{ij}^1} + \alpha_{ij}^2 \cdot \frac{\partial L}{\partial y_{ij}^2} + \alpha_{ij}^3 \cdot \frac{\partial L}{\partial y_{ij}^3} + \alpha_{ij}^4 \cdot \frac{\partial L}{\partial y_{ij}^4} \quad (4)$$

其中,  $\alpha_{ij}^1, \alpha_{ij}^2, \alpha_{ij}^3, \alpha_{ij}^4 \in [0, 1]$ 。如果  $\alpha_{ij}^2 \rightarrow 0, \alpha_{ij}^3 \rightarrow 0, \alpha_{ij}^4 \rightarrow 0$ , 则梯度内的一致性就会缓解。由于融合参数可以通过标准的反向传播算法学习, 所以能够得到有效的系数, 从而实现特征的有效融合。

### 3) DCNS 模块

由尺度、姿态、视角和部分形变因素产生的几何形变是目前小目标检测的一大挑战。本文在网络的 Neck 部分引入 DCNS 模块作为检测头, 它能够对非矩形区域位置采样, 从而使得模型很好地应对目标外观等因素。通过对其自适应行为的分析可知, 尽管与传统的卷积神经网络相比, 神经网络在空间上的支持更加接近目标, 但支持范围往往过大, 且易受背景信息干扰。针对这一问题, 本文采用可变形卷积调节网络架构, 从增强模型的建模能力、训练强度等方面提升关联区域感受野的性能, 使得模型更高效地拟合感兴趣区域, 从而减少背景信息对目标特征的影响。

本文对原有的可变形卷积结构进行如下两方面的改进, 以增强网络的建模能力。首先, 在原模块中叠加更多的可变形卷积层, 具体实现方法是: 将原模块中的  $3 \times 3$  卷积层使用可变形卷积替换, 目的是可变形卷积网络可以通过学习变形参数, 在更好的感受野下捕捉目标的细节信息, 尤其在处理小目标时具有明显优势。其次的创新是通过引入一种可变形卷积模块中的调制机制, 这种调制机制增强了可变形卷积神经网络对空间支持区域的控制能力。该方法不仅可以调节被检测到的输入信息的偏移量, 还能在不同位置上调整输入信号的幅值。在极端情况下, 通过将特征振幅值设为零的方式, 模块可以选择不接受来自特定位置的信号。这使得模块可以在不同的空间支持区域进行适当调整, 以减少局部图像内容对输出结果的影响, 并为网络模块的调节提供了更多的自由度来调整其空间支持区域。这种调制机制可通过下述过程进行计算。

给定  $k$  个样本点的卷积核函数, 记  $\omega_k$  是  $k$  个位置权重,  $p_k$  是预先指定的偏移量, 分别将输入特征映射  $x$  和输出特征映射  $y$  表示位置  $P$  的特征, 调制的可变形卷积可以表示为:

$$y(p) = \sum_{k=1}^K \omega_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (5)$$

其中,  $\Delta p_k$  和  $\Delta m_k$  分别是第  $k$  个位置的可学习偏移量和调制标量。调制标量  $\Delta m_k$  在  $[0, 1]$  范围内,  $\Delta p_k$  为实数, 范围不受约束。为了解决特征提取时引入无关区域的问题, 本文在 DCNS 中不只添加每一个采样点的偏移,

还添加了一个权重系数, 来区分引入的区域是否为感兴趣的区域, 假如这个采样点所处的区域是不感兴趣区域, 则把权重学习为 0。具体如式(6)所示。

$$y(k) = \sum_{j=1}^{n_k} x(p_{kj} + \Delta p_k) \cdot \Delta m_k / n_k \quad (6)$$

其中,  $\Delta p_k$  和  $\Delta m_k$  的值是通过在相同的输入特征图上应用单独卷积层获得, 该卷积层具有与当前卷积层相同的空间分辨率。通过以上的改进, 当在小物体的感受野上时, DCNS 的采样范围过大, 除了关注的小物体外, 会出现其他和小物体不相关的部分, 而可变形卷积的目的是利用空间冗余, 根据局部空间的相关性来实现分类等任务。改进前的模块对于这种过大的偏移会影响到最后的性能, 改进后的 DCNS 显然就控制好了偏移的幅度, 避免提取不相关的特征信息, 更好的展现了模型的表现力。

## 2 实验数据与分析

### 2.1 数据集与评价指标

本文使用 VisDrone2019 数据集进行算法的检测性能测试, 此数据集由天津大学机器学习和数据挖掘实验室 AISKYEYE 团队共同收集而成。VisDrone2019 是一个无人机航拍数据集, 主要针对复杂环境下无人机对地的目标采集而成, 有着地理覆盖范围广、场景复杂且繁多、目标类别多等特点。该数据集包含 288 个视频片段, 由 261 908 帧视频序列和 10 209 幅静态图像组成, 共标记了 10 个预定义类别 (pedestrian, person, car, van, bus, truck, motor, bicycle, awning-tricycle, andtricycle), 涵盖了无人机对地目标检测的大部分常见类别, 并且涵盖了目标检测问题中的各种难点, 如小目标、目标重叠、复杂背景、舰船检测等。除此之外, 该数据集充分考虑了环境因素的重要属性, 例如场景可见性, 对象类别和遮挡等, 因此尤其适用于无人机视角下的目标检测问题。

本文以精确率 (precision, P)、召回率 (recall, R)、平均精度 (mean average precision, mAP)、模型参数量 (Params)、模型复杂度 (FLOPs) 和每秒传输帧速率 (FPS) 作为算法性能的主要评估指标。其中多类别平均精度 mAP 是评价算法的重要指标, 用以评估检测算法对目标分类的综合性能。

上述主要指标的计算公式定义如式(7)~(10)所示。

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 PRdr \quad (9)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (10)$$

## 2.2 实验环境与配置

本文所采用的实验环境是基于 Windows-10, GPU 为 24 G 显存的 NVIDIA GeForce RTX 3090, 深度学习框架为 Pytorch, 并采用 CUDA11.3 作为加速训练架构。在模型训练前, 需要对各阶段的参数进行优化, 分别有每批训练样本数量 batch-size, 初始学习率 lr (learning rate), 权重衰减值 weight-decay, 训练轮数 epoch 以及图像输入尺寸 Image-size。具体参数配置如表 1 所示。

表 1 实验参数配置

Table 1 Experimental parameter configuration

batch-size	learning rate	Weight decay	epoch	Image-size
2	0.01	0.0005	200	640×640

在以上参数中, 学习率直接影响模型的收敛状态, 具有最优值, 过大则导致模型不收敛, 过小则导致模型收敛特别慢或者无法学习。本文针对不同的初始学习率 lr 进行了对比实验, 如表 2 所示。设置 200 个训练轮数, 实验结果显示, 当初始学习率设置为 0.05 时, 模型的训练速度变慢, 需要更多的迭代次数才能达到最佳值; 设置为 0.1 时, 梯度过大, 模型参数出现不稳定的数值, 最终导致梯度爆炸出现 NaN 值; 而初始学习率设置为 0.01 时, 模型收敛拟合状态好, 精度值最佳。

表 2 初始学习率

Table 2 Initial learning rate

Learning rate	epoch	mAP@0.5/%
0.005	200	0.434
0.01	200	0.447
0.1	200	NaN

权重衰减值 weight-decay 主要为了让模型加入正则项, 加入正则项后将使模型参数不会过大或过小, 从而控制模型的复杂度, 防止模型过拟合。本文中 weight-decay 设置为 0.0005, Loshchilov 等<sup>[21]</sup> 研究发现, 较小的衰减系数对模型的学习很重要。

## 2.3 改进模块的消融实验

为验证本文所提出 YOLO-DA 的有效性, 首先对文中提出的小目标检测层、ASFF-L 模块和 DCNS 卷积进行了消融实验。以平均精度值 (mAP), 模型参数量 (Parameters/M), 算法计算量 (FLOPs/G) 作为评估指标, 检测在原有模型上分别添加各种方法对模型的影响程度。本次消融实验在 VisDrone2019 数据集上进行, 通过 200 个 epoch 训练, 并在输入网络进行训练前, 将输入图像尺寸全部调节为 640×640 大小, 最终获得的训练结

果如表 3 所示。

在同等实验条件下进行了 6 组消融实验对比, 详细介绍如下:

第 1 组实验选用 YOLOv7-tiny 作为基准模型进行了实验, 其在 Visdrone2019 数据集的 mAP 值为 35%, 以此为标准对其余 6 组数据做对比实验。

第 2 组实验是在基准网络的基础上添加了 P2、P3 小目标检测层。对比原模型, mAP 值提升了 6.2%, 精确率提升了 4.4%, 召回率提升了 5.4%。提升的主要原因如下两点: 1) 针对小目标和极小目标新增两支小目标检测层, 取消极大目标检测层, 新增的小目标检测层与浅层网络直接相连, 其优势在于浅层的底层特征图能够更好地提取图像的分辨率和细节信息, 从而提升弱小目标的表达能力。2) 由原先的 3 个检测头扩展为 4 个检测头, 对上下文的信息提取能力更广泛, 给模型带来了更好的特征融合机制, 可以更好的预测和辨别小目标的位置与类别, 从而提高检测率。

第 3 组实验是在原基准网络上加入了 ASFF-L 特征融合模块。对比原网络, mAP 值上升了 1.3%, 精确率提升了 1.5%。改进后的特征融合模块在降低召回率的情况下很好的提高了原模型的精度问题。分析认为改进的 ASFF-L 融合模块能很好地与基准网络结合, 通过学习过滤冲突信息的方法来抑制尺度的特征不一致性, 从而提高了特征尺度的不变性。通过将 ASFF-L 模型与小目标微小目标检测层的融合, 小目标的检测性能得到了显著提高, 不同特征尺度对于预测特征图的学习能力也得到了更好地提升。

第 4 组实验在原模型检测层上增加了 DCNS 可变形卷积。对比原模型 mAP 值提升了 2.3%, 精确率提升了 1.3%, 召回率提升了 0.4%。在增加较小的参数量计算量的情况下, 引入多层可变形卷积层和调制机制的 DCNS 模块, 增强了建模能力, 使模型更高效的拟合感兴趣区域, 避免了由于目标尺寸小混入背景信息导致漏检的情况。

第 5 组实验在 4 层检测层的基础上添加了 ASFF-L 模块。改进后的自适应特征融合模块更加适应不同尺度的目标变换情况, 从表 3 中可知, 该组实验结果的 mAP 值较基准网络提升了 7.6%, 精确率提升了 9.6%; 同时 mAP 值较添加小目标检测层后的网络提升了 1.4%, 精确率提升了 5.2%。

第 6 组实验在增加的小目标检测层上使用了改进后的 DCNS 卷积模块。对比原模型 mAP 值提升了 9.2%, 精确率提升了 6.3%, 召回率提升了 7.3%。对比增加微小目标检测层后的网络, mAP 值提升了 3%, 精确率提升了 1.9%, 召回率提升了 1.9%。

第 7、8 组实验分别是在 YOLO-DA 上添加小目标检

测层和未添加小目标检测层的结果。实验显示,在基准模型上添加了 P2、P3 分支、DCNS 可变形卷积以及 ASFF-L 特征融合模块的 YOLO-DA 算法较原算法 mAP 值提升了 9.7%,精确率提升了 5.4%,召回率提升了 9.2%。而缺少 P2、P3 检测分支的 YOLO-DA 算法 mAP 值减少了 7%。

根据上述实验结果的分析可得出结论如下:首先,在特征模型上添加小目标和微小目标检测分支对于提升小目标检测的性能至关重要;其次,文中提出的 YOLO-DA 算法在无人机平台上面对小目标、背景混淆和多尺度变换等情况,能够更好地提高小目标检测的性能,取得更优的检测精度。

表 3 消融实验

Table 3 Ablation experiment

Baseline	+P2、P3 分支	ASFF-L	DCNS 卷积	P	R	mAP0.5/%	mAP0.95/%	Parameters/M	FLOPs/G	FPS
√				45.6	38.2	35	18.5	6.04	13.3	126
√	√			50	43.6	41.2	23.3	35.04	85.7	84
√		√		47.1	36.8	36.3	18.7	11.4	21.5	100
√			√	46.9	38.6	37.3	18.9	10.5	15.8	105
√	√	√		55.2	41.9	42.6	23.7	59.2	134.6	72
√	√		√	51.9	45.5	44.2	25.1	39.7	98.9	69
√		√	√	45.7	39.7	37.7	19.2	15.9	24.2	89
√	√	√	√	51	47.4	44.7	25.3	63.8	147.9	71

为了验证所提模型对不同尺度不同类别目标的检测性能,给出了数据集中的各类别目标的检测效果,如表 4 所示。从表中可知,本文所提模型在 Pedestrian 上的检测效果提升最为明显,mAP 值提升了 14.4%;在 Tri、Bus 以及 Motor 上的提升也相对较好,mAP 值分别提升了 11.4%、11.7% 和 12.7%;由于 Pedestrian 类别的目标相对于其他类别的目标的尺寸更为小,因此能够充分证明本文算法在小目标检测方面的优势。此外,本文算法在 Car 类别的 mAP 检测值达到了 84.6%,在尺度变化跨度和分布密度都较大的情况下,仍然取得了很好的检测效果。文中算法对其余各类别目标的检测效果也都较基准网络有较好的提升。

表 4 各类别对比实验

Table 4 Different types of comparative experiments

Name	YOLOv7-tiny (mAP50)	YOLO-DA (mAP50)
Pedestrian	37.6	52(+14.4)
People	36.9	44.8(+7.9)
Bicycle	9.3	17.7(+8.4)
Car	75.6	84.6(+9)
Van	38.8	47.7(+8.9)
Truck	31.5	39.6(+8.1)
Tri	19.6	31(+11.4)
Awn-tri	11.9	16.1(+4.2)
Bus	47	58.7(+11.7)
Motor	42.4	55.1(+12.7)
Mean	35	44.7(+9.7)

## 2.4 对比实验

为验证文中所提出算法的有效性,将本文所提算法

与当前主流算法在 VisDrone2019 数据集上进行对比实验。所选取的主流算法有 DroneEye2020、Cascade R-CNN、CenterNet、SDS-YOLO 等各种无人机航拍目标检测算法,实验结果如表 5 所示,表中最优的检测结果用加粗字体标注。从表 5 中可知,文中的算法在 Pedestrian、People、Car、Bus 和 Motor 5 个类别上效果提升显著,其中 mAP 指标的值提升最为明显,分别是 52%、44.8%、84.6%、58.7% 和 55.1%。在 Bicycle、Truck 检测方面略领先于当前最优算法,分别是 17.7% 和 39.9%。在 Tri、Awn-tri 检测类别上,仅次于 DroneEye2020 算法,分别是 31% 和 16.1%。

根据以上分析,本文所提算法所提升的检测精度对比双阶段主流算法 Fast R-CNN 以及 Cascade R-CNN 在平均精度上分别提升了 23% 和 21.5%,其中 Pedestrian、People 和 Car 的提升最为明显,分别提升了 30.6%、29.2% 和 32.9%。对比主流的单阶段算法 YOLOv5s,平均精度提升了 11.5%,在任何一个检测类别上 YOLO-DA 都领先于 YOLOv5s 算法。综合可知,文中提出的小目标检测算法极大提升了小目标检测精度,在推理速度上,YOLO-DA 能够满足实时检测的需求,并且优于多数的对比网络。由上可知,YOLO-DA 算法在检测无人机航拍图像时具有显著的优势。

## 2.5 算法效果分析

为了直观验证本文算法 YOLO-DA 在无人机小目标数据集上的有效性,选取数据集 VisDrone2019 测试集中复杂场景下的无人机拍摄图片进行可视化测试,检测效果如图 4 所示,图 4(a)~(e) 分别给出 5 组无人机视角下目标检测常见的挑战场景,其中左侧图像为 YOLOv7-

表5 不同检测算法的对比实验

Table 5 Comparative experiments of different detection algorithms

算法	Pedestrian	People	Bicycle	Car	Van	Truck	Tri	Awn-tri	Bus	Motor	mAP50	FPS
DREYE2020 <sup>[22]</sup>	35.7	18.3	14.0	56.5	42.9	37.6	<b>35.4</b>	<b>25.9</b>	50.4	28.9	34.6	-
CASCADE R-CNN <sup>[23]</sup>	22.2	14.8	7.6	54.6	31.5	21.6	14.8	8.6	34.9	21.4	23.2	15
CENTERNET <sup>[24]</sup>	22.6	20.6	14.6	59.7	24.0	21.3	20.1	17.4	37.9	42.0	26.2	-
SDS-YOLO <sup>[25]</sup>	46.7	35.4	14.4	82.0	45.1	35.8	26.5	12.7	54.3	47.4	40.0	40
FAST R-CNN <sup>[26]</sup>	21.4	15.6	6.07	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7	20
YOLOV5S <sup>[27]</sup>	38.2	30.6	12.0	72.4	35.2	30.8	19.3	10.2	45.5	37.8	33.2	<b>85</b>
MIXEDYOLOV3-LITE <sup>[28]</sup>	34.5	23.4	7.9	70.8	31.3	21.9	15.3	6.20	40.9	32.7	28.5	-
YOLO-DA	<b>52.0</b>	<b>44.8</b>	<b>17.7</b>	<b>84.6</b>	<b>47.7</b>	<b>39.6</b>	31	16.1	<b>58.7</b>	<b>55.1</b>	<b>44.7</b>	71

tiny 检测效果,右侧图为 YOLO-DA 算法检测效果,效果图中的虚线框标与红色指示箭头标记 YOLOv7-tiny 与 YOLO-DA 检测结果的不同之处。

如图所示,图 4(a)为夜晚小目标的检测效果,且在图中分布有大量不同尺度的目标,YOLO-DA 中在夜晚路灯及车灯照射下,行人及摩托车的漏检及误检率降低,且在商店门口的行人及自行车的检测效果也相对提高,这是由于 ASFF-L 自适应多尺度融合模块,发挥了不同特征尺度对于预测特征图的学习效果。当存在遮挡和运动模糊造成小目标特征信息不全的情况下,由图 4(b)、(d)的检测效果可以看出,所提算法在日光及路灯下均能够正确检测出部分误检目标和漏检目标。图 4(c)、(e)为密

集小目标的检测效果分析,通过左右两图对比可以看出在密集场所时,针对三轮车及驾驶员的误检现象有了很大的改善;图 4(e)关于小目标的复杂背景下行人漏检率也有很好的提升。

通过算法效果对比可得,经改进过的算法在白天、黑夜、强曝光场景及模糊目标和集市等复杂环境密集目标镜框下,对人、车等小目标能够有效抑制背景和环境的干扰,正确定位且分类每个目标。综上所述,本文提出的基于 YOLO-DA 的多目标检测方法,在不同背景、不同目标分布和不同光照条件等实际场景下,对各类小目标检测的漏检率及误检率有了很大的提升,能够满足无人机航拍图像中的目标检测任务。



(a) 夜晚小目标  
(a) Small night goals



(b) 遮挡目标  
(b) Concealed object



(c) 密集目标  
(c) Intensive targeting

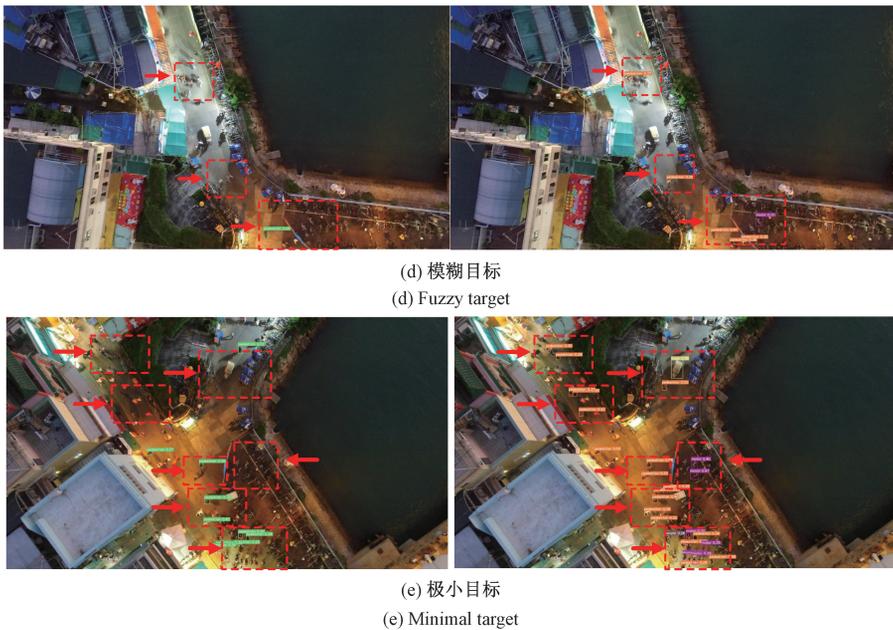


图 4 检测效果对比

Fig. 4 Comparison of detection effect

### 3 结 论

为解决无人机在高空中由于尺度不一致、复杂环境下目标遮挡密集以及极小目标下由于先验框过大导致特征信息丢失等问题,提出了 YOLO-DA 算法。在特征提取部分,改进目标检测层加入小目标和微小目标检测层,对小目标的改进效果非常显著,增强了特征信息的提取能力。在特征提取部分引入可变形卷积调节模块,替换可变形卷积层以及增加调制机制的方式,使得模型更高效地拟合感兴趣区域,避免了由于目标尺寸小背景信息复杂导致漏检的问题。在检测阶段,引入自适应特征融合层,通过学习过滤冲突信息用来抑制尺度的不一致性的方法,提高了由于不一致性出现的小目标漏检问题。YOLO-DA 算法在 VisDrone 无人机数据集上对比原算法,在准确率、召回率和 mAP 值均有提升,mAP50 值提升了 9.7%,mAP50-95 值提升了 6.8%,此外图像推理速度为 71 fps,能够达到目标的实时处理。在未来的工作中,将尝试将 YOLO-DA 算法进行嵌入式端的植入,满足无人机小目标的日常场景检测需求。

### 参考文献

- [ 1 ] 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 target 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.

- [ 2 ] 闫钧华,张琨,施天俊,等. 融合多层级特征的遥感图像地面弱小目标检测[J]. 仪器仪表学报,2022,43(3):221-229.  
YAN J H, ZHANG K, SHI T J, et al. Ground dim small target detection in remote sensing images by fusing multi-level features [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.
- [ 3 ] ZHANG Y. Detection and tracking of human motion targets in video images based on camshift algorithms[J]. EE Sensors Journal,2020,20(20):11887-11893.
- [ 4 ] CHEN X L, YU X H, HUANG Y, et al. Adaptive clutter suppression and detection algorithm for radar maneuvering target with high-order motions via sparse fraction ambiguity function[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020,13:1515-1526.
- [ 5 ] GIRSHICK R, DONAHUE J, DARRELL T, et al. Richfeature hierarchies for accurate object detection and semantic segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [ 6 ] GIRSHICK R. Fast R-CNN [C]. IEEE International Conference on Computer Vision, 2015:1440-1448.
- [ 7 ] RUN J, HE K, GIRSHICK R, et al. Faster R- CNN;

- Towards real-time object detection with region proposal net-works [ C ]. Advances in Neural Information Processing Systems, 2015:1137-1149.
- [ 8 ] REDMON J, FARHADI A. YOLOv3: An incremental improvement [ J ]. ArXiv Preprint arXiv: 1804.02767, 2018.
- [ 9 ] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint arXiv: 2004. 10934, 2020.
- [10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [ J ]. ArXiv Preprint arXiv: 2207. 02696, 2022.
- [11] LIU W, ANGUELOVD, ERHAN D, et al. SSD: Single shot multibox detector [ C ]. European Conference on Computer Vision, 2016:21-37.
- [12] 陈欣,万敏杰,马超,等.采用多尺度特征融合 SSD 的遥感图像小目标检测 [ J ]. 光学精密工程, 2021, 29(11): 2672-2682.  
CHEN X, WAN M J, MA CH, et al. Recognition of small targets in remote sensing image using multi-scale feature fusion-based shot multi-box detector [ J ]. Optics and Precision Engineering, 2021, 29(11): 2672-2682.
- [13] 石欣,卢灏,秦鹏杰,等.一种远距离行人小目标检测方法 [ J ]. 仪器仪表学报,2022,43(5):136-146.  
SHI X, LU H, QIN P J, et al. A long-range pedestrian small target detection method [ J ]. Chinese Journal of Scientific Instrument, 2022, 43(5):136-146.
- [14] 刘芳,韩笑.基于多尺度深度学习的自适应航拍目标检测 [ J ]. 航空学报,2022,43(5):471-482.  
LIU F, HAN X. Adaptive aerial object detection based on multi-scale deep learning [ J ]. Acta Aeronauticae Astronautica Sinica, 2022,43(5):471-482.
- [15] AKYON F C, ALTINUC S O, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection [ J ]. ArXiv Preprint arXiv: 2202. 06934,2022.
- [16] 邝先验,程福军,吴翠琴,等.基于改进 YOLOv7-tiny 的高效轻量化遥感图像目标检测方法 [ J ]. 电子测量与仪器学报,2024,38(7):22-33.  
KUANG X Y, CHENG F J, WU C Q, et al. Efficient and lightweight remote sensing image object detection method based on improved YOLOv7 tiny [ J ]. Electronic Measurement and Instrumentation,2024,38(7):22-33. .
- [17] 戚玲珑,高建瓴.基于改进 YOLOv7 的小目标检测 [ J ]. 计算机工程,2023,49(1):41-48.  
QI L L, GAO J L. Small object detection based on improved YOLOv7 [ J ]. Computer Engineering, 2023, 49(1):41-48.
- [18] 刘毅,于畅洋,李国燕,等. UAST-RCNN:遮挡行人的目标检测算法 [ J ]. 电子测量与仪器学报,2022,36(12): 168-175.  
LIU Y, YU CH Y, LI G Y, et al. UAST-RCNN: Object detection algorithm for blocking pedestrians [ J ]. Journal of Electronic Measurement and Instrumentation, 2022, 36(12):168-175.
- [19] 梁礼明,龙鹏威,冯耀,等.改进轻量化 VTG-YOLOv7-tiny 的钢材表面缺陷检测 [ J ]. 光学精密工程, 2024, 32(8): 1227-1240.  
LIANG L M, LONG P W, FENG Y, et al. Improved steel surface defect detection of lightweight VTG-YOLOv7-tiny [ J ]. Optics and Precision Engineering, 2024, 32(8): 1227-1240.
- [20] 陈卫彪,贾小军,朱响斌,等.基于 DSM-YOLO v5 的无人机航拍图像目标检测 [ J ]. 计算机工程与应用, 2023,59(18):226-233.  
CHEN W B, JIA X J, ZHU X B, et al. Target detection for UAV image based on DSM-YOLOv5 [ J ]. Computer Engineering and Applications, 2023, 59(18): 226-233.
- [21] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [ J ]. ArXiv Preprint arXiv: 1711.05101,2017.
- [22] CAO Y, HE Z, WANG L, et al. VisDrone-DET2021: The vision meets drone object detection challenge results [ C ]. IEEE International Conference on Computer Vision,2021.
- [23] YU W P, YANG T, CHEN CH. Towards resolving the challenge of long-tail distribution in UAV images for object detection [ C ]. IEEE Winter Conference on Applications of Computer Vision ( WACV ), 2021: 3257-3266.
- [24] ALBABA B M, OZER S. SyNet: An ensemble network for object detection in UAV images [ C ]. IEEE International Conference on Pattern Recognition, 2021.
- [25] 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 uav aerial target detection algorithm [ J ]. Journal of Electronic Measurement Technology, 2022, 45 ( 19 ): 167-174.
- [26] YU W P, YANG T J N, CHEN C. Towards resolving the challenge of long- tail distribution in UAV images for object detection [ C ]. IEEE Winter Conference on Applications of Computer Vision, 2021: 3258-3267.
- [27] 刘展威,陈慈发,董方敏.基于 YOLOv5s 的航拍小目标检测改进算法研究 [ J ]. 无线电工程,2023,53(10): 2286-2294.

LIU ZH W, CHEN C F, DONG F M. Research on improved algorithm for airborne small target detection based on YOLOv5s [J]. Radio Engineering, 2023, 53 (10): 2286-2294.

[28] ZHAO H, ZHOU Y, ZHANG L, et al. Mixed YOLOv3-LITE: A lightweight real-time object detection method[J]. Sensors, 2020, 20: 1861.

作者简介



苏雨蕾, 2021 年于青岛理工大学琴岛学院获得学士学位, 现为长春理工大学硕士研究生, 主要研究方向为图像处理与机器视觉。

E-mail: 1595438923@qq.com

**Su Yulei** received her B. Sc. degree from Qingdao College of Qingdao University of Technology in 2021. She is now a M. Sc. candidate in Changchun University of

Science and Technology. Her main research interests include image processing and machine vision.



**黄丹丹**(通信作者), 2007 年于长春理工大学获得学士学位, 2009 年于东北大学获得硕士学位, 2016 年于大连理工大学获得博士学位, 现为长春理工大学讲师, 主要研究方向为计算机视觉、模式识别、机器学习等方面的研究。

E- mail: hdd@ cust. edu. cn

**Huang Dandan**(Corresponding author), received her B. Sc. degree from Changchun University of Science and Technology in 2007, her M. Sc. degree from Northeastern University in 2009, and her Ph. D. degree from Dalian University of Technology in 2016. She is now a lecturer at Changchun University of Science and Technology. Her main research interests include computer vision, pattern recognition, machine learning and so on.