

改进YOLOv5的无人机影像小目标物体识别算法

窦同旭, 曾勇, 杨冲, 李昂

盐城工学院机械学院, 江苏 盐城

收稿日期: 2023年9月27日; 录用日期: 2023年11月10日; 发布日期: 2023年11月17日

摘要

无人机图像目标检测是流行的图像识别任务之一。然而, 由于无人机航拍影像具有目标尺度变化大、背景复杂等特点, 现有的检测器难以准确检测小目标。为了解决这个问题, 本文提出了改进的算法模型YOLO-ADOP, 用于提高在无人机影像中小目标的检测效果。首先, 通过调整检测分支以提高模型在小目标下的检测能力。其次, 将分类任务与回归任务解耦, 设计了解耦头结合辅助训练头的预测头。最后, 使用EIOU优化损失函数, 使用OTA改进标签分配策略。通过VisDrone数据集的实验结果表明, 改进后的YOLO-ADOP模型较YOLOv5模型在640分辨率和1536分辨率上的AP50指标分别提升了7.18%和4.65%。对比其他主流模型对于小目标拥有更好的检测效果, 能够有效完成无人机航拍影像的小目标检测任务。

关键词

无人机影像, YOLOv5, 小目标, 辅助训练头

Improved Algorithm for Small Target Object Recognition in Drone Images Based on YOLOv5

Tongxv Dou, Yong Zeng, Chong Yang, Ang Li

College of Mechanical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

Received: Sep. 27th, 2023; accepted: Nov. 10th, 2023; published: Nov. 17th, 2023

Abstract

Unmanned aerial vehicle (UAV) image object detection is a popular task in image recognition.

文章引用: 窦同旭, 曾勇, 杨冲, 李昂. 改进 YOLOv5 的无人机影像小目标物体识别算法[J]. 建模与仿真, 2023, 12(6): 5395-5407. DOI: 10.12677/mos.2023.126490

However, existing detectors face difficulties in accurately detecting small objects in UAV aerial images due to their large scale variations and complex backgrounds. To address this issue, this paper proposes an improved algorithm model called YOLO-ADOP, aiming to enhance the detection performance of small objects in UAV images. Firstly, the detection branch is adjusted to enhance the model's detection capability for small targets. Secondly, the classification task and regression task are decoupled, and a decoupled head with an auxiliary head is designed for prediction. Finally, the EIOU optimization loss function and OTA improved label assignment strategy are employed. Experimental results on the VisDrone dataset demonstrate that the improved YOLO-ADOP model achieves a 7.18% and 4.65% improvement in AP50 metric at resolutions of 640 and 1536, respectively, compared to the YOLOv5 model. It outperforms other mainstream models in detecting small targets and can effectively accomplish the task of small target detection in drone aerial images.

Keywords

Drone Images, YOLOv5, Small Targets, Auxiliary Head

Copyright © 2023 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, 卷积神经网络(CNN, Convolution Neural Network)在计算机视觉领域取得了巨大突破。基于深度神经网络的模型在目标检测任务中扮演着重要角色, 这些模型按不同标准可分为不同类型。按照是否需要生成候选, 其中 YOLO 系列模型[1]-[7]和 SSD [8]是典型一阶段算法, 而 Fast R-CNN [9]和 Faster R-CNN [10]是二阶段算法。基于是否有预定义锚点, 可将模型分为基于锚点检测器和无锚点检测器, 其中基于锚点的方法包括 Faster R-CNN 和 YOLOv5, 而 CornerNet [11]、CenterNet [12]和 YOLOX [13]则是著名的无锚点检测器。另外, 基于 Transformer [14]模型架构的 Swin-transformer [15]算法也在图像检测任务中取得了相当不错的成绩。

与此同时, 受益于近年来出现的大规模标准的无人机拍摄图像数据集如 VisDrone [16]、UAVDT [17]等, 利用深度神经网络的方法进行识别无人机图像目标的工作已经取得了显著的突破。

尽管基于 CNN 的检测算法在常规目标检测任务中取得了巨大成功, 但在无人机场景下对小目标的检测效果较差。主要原因是无人机航拍影像中物体尺寸变化巨大, 而单层特征图在卷积神经网络中的表征能力有限。同时, 航拍影像的背景复杂多变, 小目标容易受到背景干扰, 从而影响检测效果。因此, 需要针对无人机航拍图像中小目标的检测算法进行有针对性的研究, 并探索优化和改进卷积神经网络的方法, 以提高算法在无人机场景下的检测能力。为了应对无人机拍摄图像中车辆检测所面临的多样性挑战, AdNet [18]寻求在不同视角、光照、天气和背景之间对齐特征, 遵循域适应的思路。GLSAN [19]通过添加一个自适应区域, 选择全局 - 局部检测网络的算法, 找到高密度区域, 并准确地检测具有大尺寸变化的对象。Gallo 等人[20]利用 YOLOv7 模型解决了无组织作物条件和杂草生物变异性带来的挑战。为了应用深度学习算法在实时场景中的无人机上, 许多研究致力于设计快速、轻量级的模型以实现高质量的无人机目标检测。UAV-Net [21]全面分析了不同的骨干网络结构、预测头和模型修剪方法对实现快速目标检测的影响, 并构建了更好的组合。

最近针对小目标的目标检测算法中, TPH-YOLOv5 [22]在基于 YOLOv5 检测模型上添加基于

transformer 的预测头，以提高网络的预测回归能力，同时使用注意力机制提高对小目标关注度，实现了在大尺寸变化和高密度场景中的显著性能提升。在 TPH-YOLOv5++ [23]中，设计了 CA-Trans 来替换额外的预测头，同时保留该预测头的知识。通过使用稀疏局部注意力(SLA)模块，可以有效地捕获额外预测头和其他头之间的非对称信息，丰富其他头的特征。TPH-YOLOv5 和 TPH-YOLOv5++虽然在识别精度上有明显的提升，但其计算量也成倍增加，对硬件性能要求大，不宜部署在普遍运算性能小的无人机上。

为此，本文提出了一种面向无人机影像的目标检测算法，针对现有目标检测算法在无人机场景下难以检测小目标的问题，重新进行网络设计。本文的贡献如下：

(1) 提出了适应小目标的 YOLO-ADOP (YOLO-AUX-DHead-OTA-P2)检测模型，通过添加 P2 识别层来检测小目标，同时去除了 P5 识别层用于大型目标检测。

(2) 创新性地使用解耦头和辅助训练头来分离分类和回归任务，提高目标检测精度，同时加速权重更新并减少训练批次之间的影响。

(3) 优化了损失函数计算，使用 EIOU [24]替代 CIOU 来更好地反映预测框与目标框的宽高差异，加速网络收敛。引入 OTA (Optimal transport assignment) [25]标签分配原则以提高准确度。

本文代码以公开：<https://github.com/doctor-zai/YOLO-ADOP.git>。

2. YOLO-ADOP 算法概述

2.1. YOLOv5 概述

YOLOv5 共有 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 以及 YOLOv5x 五个模型，其在结构上相同，只是在宽度和高度上不同。YOLOv5 采用 CSPDarknet53 和快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)作为主干网络，采用 PANet 作为颈部网络，头部由三个耦合预测头组成。其模型结构简图如图 1 所示。

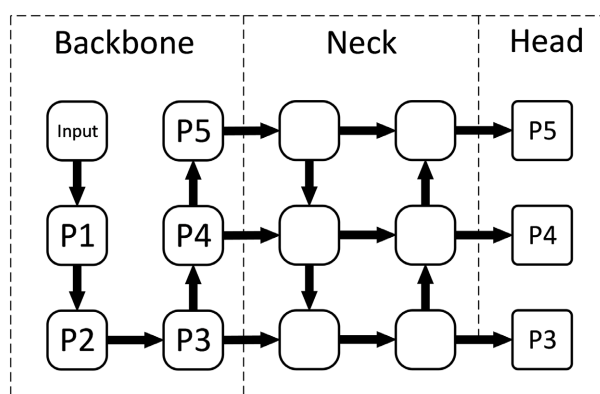


Figure 1. Simplified diagram of the YOLOv5 model architecture
图 1. YOLOv5 模型结构简图

2.2. YOLO-ADOP 模型

本文 YOLO-ADOP 模型网络结构如图所示，其中红色虚线框中是新增的检测分支 P2，同时减去了检测分支 P5；蓝色部分是使用的解耦头；虚线箭头为新增的辅助训练头部分。

图 2 中 CBS 模块是包含了卷积、归一化处理和激活函数的一个基础模块。C3 模块则是一个用于特征提取的模块，此模块不改变特征图的大小，采用了 ResNet 里的残差结构，其内部用于特征提取的 BottleNeck 数量可以进行调节。Upsample 为上采样，Concat 为拼接操作。

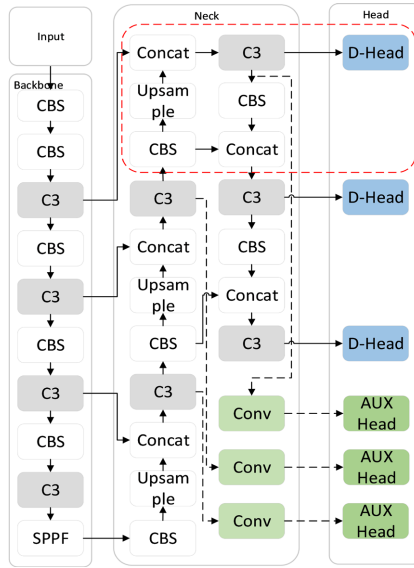


Figure 2. YOLO-ADOP model architecture diagram
图 2. YOLO-ADOP 模型结构图

2.2.1. 改进检测分支

如图所示，红色虚线框内为新增的 P2 检测部分，其作用是用于检测极小的目标。在卷积网络中，深层卷积会损失许多小目标信息，造成小目标的漏检误检；而 P2 检测分支的输入大部分来自于浅层主干网络，浅层网络保留了很多小目标的形状、大小和颜色等信息。所以引入 P2 检测分支可以更有效的提取到小目标信息，以此增加小目标识别精度。

于此同时，鉴于无人机拍摄图片中的大目标较为稀少，而使用 P5 检测分支来检测大目标不仅需要更多算力，还会在反向传播更新权重时引进误差，造成总体检测精度降低。因此，本文在新增 P2 检测分支的同时，也在算法中剔除了原本存在的 P5 检测分支。

2.2.2. 辅助训练头

如图 3 所示。辅助训练头(Auxiliary Head)的输入通道是上采样过程中 3 个 C3 的输出，然后通过卷积改变其通道数。辅助训练头在训练过程中能够起到加快中间层权重更替、减少训练批次的作用。

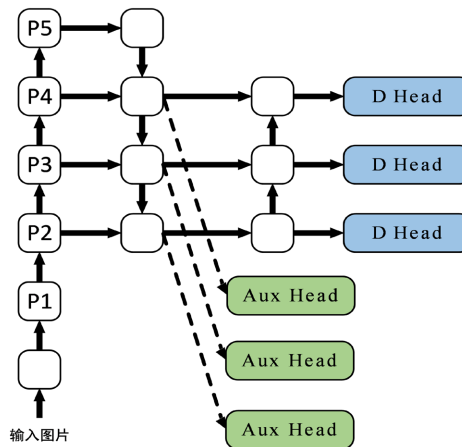


Figure 3. Simplified diagram of the YOLO-ADOP model
图 3. YOLO-ADOP 模型简图

2.2.3. 解耦头

YOLOv5 的检测头中分类任务与回归任务是通过权重耦合在一起的，这种设计会造成两种任务之间的冲突。YOLOX 证明如果将分类任务与回归任务解耦，可以提高模型检测精度。本文参考了 YOLOX 设计的解耦头，并对其解耦头进行简化，解耦头设计如图 4 所示。

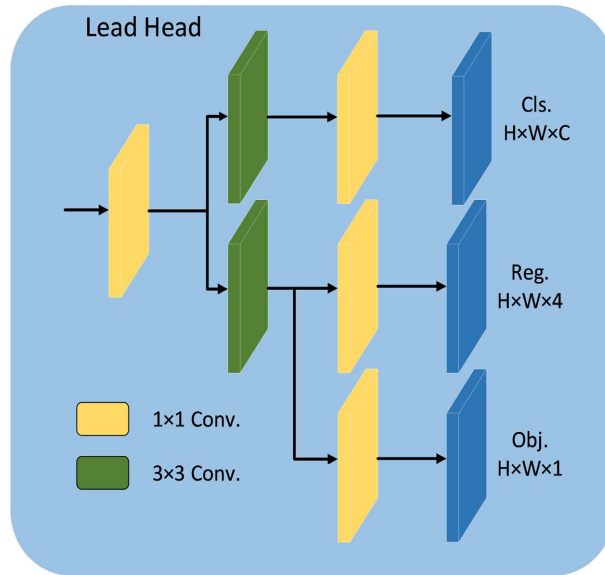


Figure 4. Diagram of the decoupled head structure

图 4. 解耦头结构图

如图 4 所示，解耦头的输入依次通过 1×1 、 3×3 、 1×1 的卷积后，分别得到三组，从而实现将分类任务与回归任务分离的目的。检测头解耦后减少了任务之间的差异带来的预测偏差，从而提高了模型检测的精度。

2.3. 改进损失函数(EIOU, OTA)

YOLOv5 使用的损失函数为 CIOU 损失函数，公式如下，其中， w ， h ， w^{gt} ， h^{gt} ，分别表示预测框和真实框的宽高， b 和 b^{gt} 表示预测边界框和真实边界框的中心点， ρ 表示预测值和真实值的欧式距离， w^c ， h^c 分别表示预测框和真实框的最小外接矩形的宽高，IOU 表示交并比。

$$L_{CIOU} = 1 - IOU + \frac{\rho^2 b, b^{gt}}{(w^c)^2 + (h^c)^2} + \alpha v \tag{1}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{2}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{3}$$

在回归时，当预测框和真实框的纵横比成线性时，其惩罚项就失去了原本的作用；预测框的梯度中， w 和 h 值不能保持同步增加或者同步减少，这样就无法有效的描述回归目标，可能会导致收敛缓慢并且回归不准确。

为解决上述问题，Zhang 等人提出了 EIOU [24]，公式如下所示：

$$L_{\text{EIOU}} = 1 - \text{IOU} + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (4)$$

EIOU 延续了 CIoU 的中心距离损失和重叠损失，但调整了宽高损失，在惩罚过程中，可以更好的反应预测框与目标框之间的宽高差异，加速了网络收敛，有助于提升回归精度。

2.4. 标签分配原则

在本文中我们引入 OTA 标签分配原则[25]，OTA 从全局的角度创新性地重新审视了标签分配，并提出将分配过程作为一个最优输运问题来定义，这是优化理论中一个经过深入研究的主题。具体来说，将每个需求者(anchor)与供应者(gt)之间的单位运输成本定义为它们分类和回归损失的加权和，然后将寻找最佳分配方案转化为在最小运输成本下解决最优输运计划，这将有效的提高识别精度。

3. 实验与结果分析

本文采用的实验数据是 VisDrone 数据集[16]，它是由中国天津大学机械学习与数据挖掘实验室 AISKYYE 团队制作。VISDrone 作为历届比赛官方数据集，其图片数据由不同型号的小型无人机，在各种场景以及不同天气和光线情况下进行收集制作而成。数据集包含四个部分：训练集、验证集、测试集，以及测试挑战集，其中图像的最大分辨率为 2000 × 1500，训练集具有 6471 个图片及其对应的用于训练模型的注释；验证集和测试集分别有 548 个和 1610 个图片，它们都有相应的注释，并且这两个集合用于评估模型；测试挑战集有 1580 个图像，并且不提供注释。训练集各类物体数量如图 5 所示。

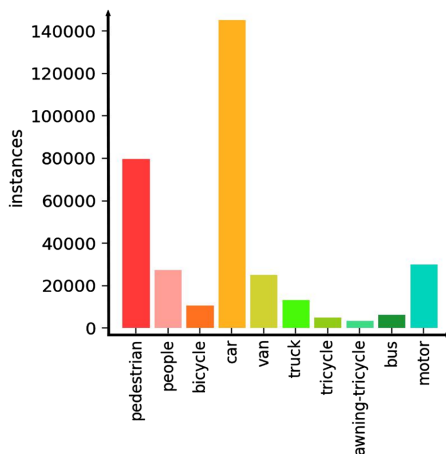


Figure 5. Label distribution of VisDrone dataset
图 5. VisDrone 数据标签分部图

3.1. 实验环境

本文实验在 Linux 操作系统下进行，采用基础模型为 7.0 版本的 YOLOv5 [7]，使用 NVIDIA GeForce RTX 3090 显卡进行训练和测试。为了减少训练时间，本文实验在训练阶段都使用了相对应 COCO 数据集下的预训练权重。软件环境采用 Python 3.9、PyTorch 1.13.1 作为实验框架，使用 CUDA 11.6 进行 GPU 加速，学习率为 0.01，采用带动量的随机梯度下降法(SGD)作为优化器，动量设置为 0.937，权重衰减系数为 0.0005。

3.2. 实验评价指标

本文采用 AP₅₀、mAP、精确度(Precision)、召回率(Recall)作为评估指标。相应的计算公式如下：

$$\text{precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{6}$$

其中 TP 为正确分类的正样本数，FP 为误报的负样本数，FN 为漏报的正样本数。

$$AP_i = \int P(R) dR \tag{7}$$

$$AP = \frac{1}{C} \sum_{i=1}^C AP_i \tag{8}$$

mAP 是在区间[0:50, 0:95]区间内，以 0:05 为步长 10 个 AP 值的平均值。

$$mAP = \frac{1}{10} \sum_{i=1}^{10} AP \tag{9}$$

3.3. 实验结果与分析

3.3.1. 检测分支对检测精度影响实验分析

本文首先对检测分支的位置和数量进行实验。以 YOLOv5s 为基础模型，对从 Neck 中引出的 4 个检测分支 P2~P5 进行实验。实验中通过增加或舍弃不同的检测分支以及其对应的 neck 层来判断检测分支对最终结果的影响。实验数据为 VisDrone 数据集，epoch 为 300，测试图片大小固定为 640 × 640，实验结果如表 1 所示。AP₅₀^{val}，mAP^{val}，AP₅₀^{test} 和 mAP^{test} 分别表示验证集和测试集上的 AP₅₀ 和 mAP。

Table 1. Detection results of different detection branch combinations

表 1. 不同检测分支组合的检测结果

保留的预测层				AP ₅₀ ^{val}	AP ₅₀ ^{test}	mAP ^{val}	mAP ^{test}	计算量 GFLOPs
P2	P3	P4	P5					
	√	√		34.89	28.93	19.33	15.37	14.6
	√	√	√	34.18	28.83	18.84	15.42	16.0
√	√	√		39.15	31.02	21.87	16.45	17.5
√	√	√	√	38.05	29.08	21.22	15.96	18.9

由表 1 中的第 3、4 和第 1、2 项项对比可看出，增加 P2 检测分支可以有效提高检测效果，原因在于 P2 检测分支包含更多的浅层特征，包括形状、大小和位置等信息，因此它能更准确地定位小目标，这也是它能有效提升小目标检测效果的原因。对比表 1 中的第 1、3 项和第 2、4 项可以看出，剔除掉 P5 检测分支后，检测效果不仅没有下降反而有所提升，原因在于大物体(像素占比大的物体)在 VisDrone 数据集里所占比重很小，而 P5 检测分支主要检测大物体，不适合检测小目标物体。综上，为提高检测效果，本文选用 P2、P3、P4 三个分支进行检测。

3.3.2. 各类 IOU 对比实验

本文对各类 IOU 进行实验，研究目前常见的 CIOU、DIOU、EIOU、GIOU 和 SIOU 以及其相关 α 变种对 VisDrone 数据集结果影响。其中 CIOU 为 YOLOv5 默认 IOU，本文 α 取值为 3，epoch 为 100，batch-size 为 64。实验结果如下表 2 所示。

Table 2. Impact of different IOU metrics on experimental results
表 2. 不同种类 IOU 对实验结果的影响

IOU 种类	AP_{50}^{test}	mAP^{test}	P^{test}	R^{test}
CIOU	27.69	14.76	38.87	30.55
DIOU	26.85	14.47	40.01	29.86
EIOU	28.29	14.86	39.37	31.64
GIOU	26.81	14.34	39.24	29.77
WIOU	26.85	14.08	38.67	29.73
α -CIOU	27.76	15.51	38.84	31.28
α -EIOU	28.29	14.86	39.37	31.64

由表可以看出，相较于 CIOU，使用 EIOU 的 AP_{50} 会有 0.60% 的提升，mAP 的提升为 0.10%。综合考虑本文使用 EIOU 作为算法。

3.3.3. 解耦头对实验的影响分析

为验证解耦头的检测效果，本文设计了耦合头与解耦头的对比实验。实验数据采用 VisDrone 数据集，实验图片固定大小为 640×640 ，基准模型 Baseline 为 YOLOv5s 模型，D-Head 为使用了解耦头的模型，实验结果如表 3 所示。

Table 3. Experimental results of coupled head and decoupled head comparison
表 3. 耦合头与解耦头对比实验结果

	AP_{50}^{val}	AP_{50}^{test}	mAP^{val}	mAP^{test}	计算量 GFLOPs
Baseline	34.18	28.83	18.84	15.43	16.0
D-Head	35.50	28.91	19.79	15.66	28

从表 3 可以看出，使用解耦头后模型的检测效果要优于基准模型，在验证集上的 AP_{50} 提升了 1.32%。证明了解耦头设计的有效性。

3.3.4. OTA 损失函数和辅助训练头对实验影响分析

为验证本文使用的 OTA 机制和辅助训练头对实验结果的影响，本文通过增加 OTA 和辅助训练头，来探究辅助训练头和 OTA 对实验结果的影响，实验结果如表 4 所示。

Table 4. Improvement effects of OTA and auxiliary heads
表 4. OTA 和辅助训练头的提升效果

	AP_{50}^{test}	mAP^{test}	P^{test}	R^{test}	计算量 GFLOPs	参数量/M
Baseline	28.83	15.43	39.72	32.05	16.0	7.1
+OTA	32.32	17.23	44.11	35.15	16.0	7.1
+AUX	28.87	15.45	40.14	32.17	21.8	10.2
+OTA+AUX	33.61	17.84	46.05	35.58	21.8	10.2

从表 4 中可以看出, 使用了 OTA 之后, 检测效果提高了 3.49%; 只添加辅助训练头的检测效果提高并不明显, 但其加快了训练速度, 训练时长由 6.1 h 缩短到 3.3 h; 两者结合后, 检测效果增加了 4.78%, 证明了辅助训练头的有效性。

3.3.5. 综合对比实验

最后, 为了综合测试本文模型的检测效果, 本文在对比基准模型的同时也对比了目前主流模型进行对比。所有模型的训练数据是 Visdrone 训练集, 并最终在 VisDrone 测试集上进行验证, batch-size 设置为 1, 验证结果如下表 5 所示。

Table 5. Comparative experimental results of various models on the VisDrone test dataset

表 5. 各种模型在 VisDrone 测试集对比实验结果

算法	Input size	$AP_{50}^{test}/AP_{50}^{val}$	mAP^{test}/mAP^{val}	推理时间/ms	计算量 GFLOPs	参数量/M
YOLOv5s	640	28.83%/34.18%	15.43%/18.84%	7.6	16.0	7.1
YOLOv5m	640	31.85%/38.83%	17.77%/22.44%	11.1	48.3	20.9
YOLOv5l	640	33.54%/40.83%	19.17%/24.39%	13.5	108.4	46.1
YOLOv5x	640	34.61%/42.28%	19.76%/25.14%	16.1	204.8	86.2
YOLOv7-tiny	640	30.29%/35.80%	15.36%/18.61%	7.2	13.3	6.2
YOLOv8s	640	32.84%/41.20%	19.02%/24.60%	8.6	28.7	11.2
TPH-YOLO	640	36.63%/44.60%	20.93%/26.33%	27.7	271.3	112.9
YOLO-ADOP (ours)	640	36.01%/45.01%	19.47%/25.84%	10.7	35.4	12.8
YOLOv5s	1536	42.89%/53.29%	25.19%/37.32%	8.4	16.0	7.1
YOLO-ADOP* (ours)	1536	47.05%/57.70%	27.05%/35.31%	15.4	42.8	37.0

(注: *表示该方法没有剔除 P5 检测分支)。

由表 5 所示, YOLO-ADOP 模型在 VisDrone 数据集上的检测效果要优于同级别先进算法。本文算法在 640 分辨率下, 较基准模型 AP_{50} 提升了 7.18%、 mAP 提升了 4.04%, 而推理时间为 10.7 ms, 计算量只增加了 19.4 G, 参数量更是只增加了 5.7 M, 可以满足实际的运用需求。训练过程如图 6 所示。

由图 6 所示, 横轴为训练批次, 设定训练批次为 300, YOLO-ADOP 和 YOLOv5s 均在 200 epoch 左右达到最优模型。在验证集上 AP_{50} 提升了 10.83%, 有效证明本文算法改进的有效性。

如图 7 所示, 更为直观的可以看出, 本文算法在检测效果上超过了所有 YOLOv5 模型, 以及同级别 YOLOv8s 和 YOLOv7-tiny 算法。本文算法与 TPH-YOLO [22]模型相比, 在检测效果相似的情况下, 本文算法的推理时间和计算量只是 TPH-YOLO 模型的 38.6%和 13.0%, 参数量更是只是有其 11.3%, 可见 YOLO-ADOP 对于硬件要求更低、表现更优, 更适合部署在无人机上。

此外, 本文还分别对比了不同小目标、中等目标以及大目标的 AP_{50} 值, 测试结果如表 6 所示。

由表 6 可见, 本文改进算法在小、中、大目标检测方面相较于基准模型均有提升, 证明本文算法更适合于物体变化大的无人机拍摄场合。

3.3.6. 消融实验

为了更好的说明本文改进模块与方法对模型检测能力的提升, 以及对计算量、推理时间和模型体积的影响, 本文最后进行了消融实验。本消融实验是在 YOLOv5s 模型上逐个添加本文改进模块与方法所得

出的实验结果，首先添加的是设计的解耦头，其次改变检测分支，然后使用 EIOU 代替 CIOU 优化损失函数，然后添加 OTA，最后新增了辅助训练头(AUX)。消融实验的测试结果基于 VisDrone 测试集，测试图片分辨率大小固定为 640，测试结果如表 7 所示。

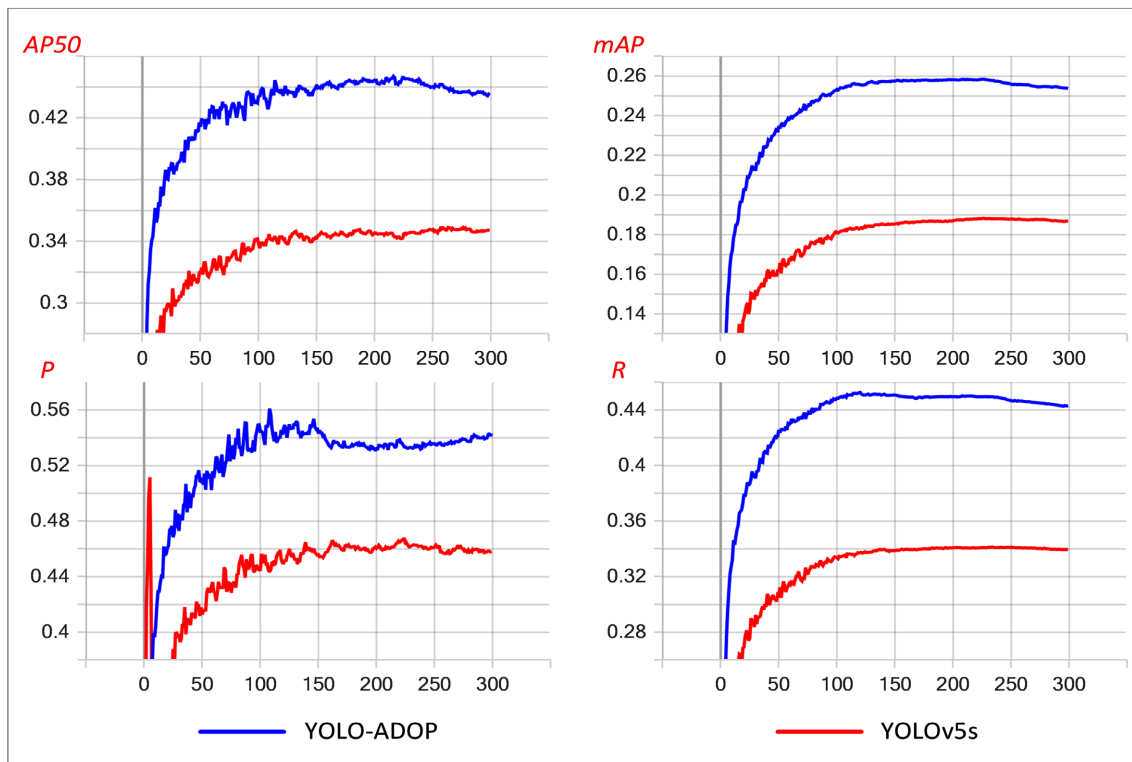


Figure 6. Comparison of the training process
图 6. 训练过程对比图

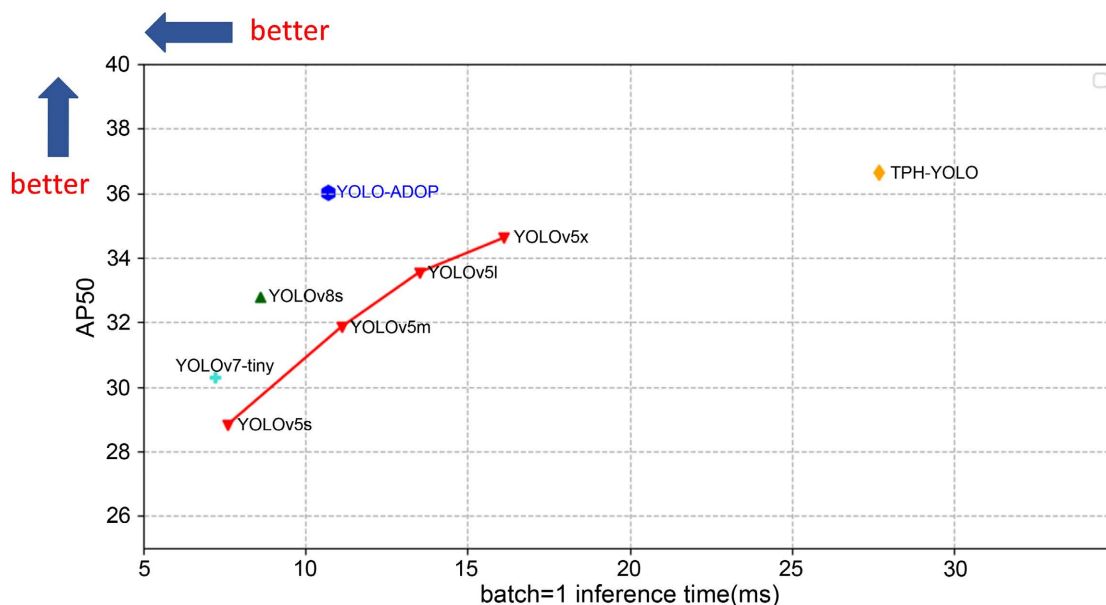


Figure 7. Comparison of model results
图 7. 模型结果对比图

Table 6. Detection performance of YOLO-ADOP and YOLOv5 for objects of different scales
表 6. YOLO-ADOP 和 YOLOv5 对于不同尺度物体的检测效果

	Input size	AP_S^{test}	AP_M^{test}	AP_L^{test}
YOLOv5s	640	6.4	21.9	32.3
YOLO-ADOP (ours)	640	9.2	27.1	38.3
YOLOv5s	1536	14.2	34.9	46.4
YOLO-ADOP* (ours)	1536	16.3	36.5	49.3

(注：*表示该方法没有剔除 P5 检测分支)。

Table 7. Results of ablation experiments
表 7. 消融实验结果图

	AP_{50}^{test}	mAP^{test}	计算量 GFLOPs	推理时间/ms	模型体积/Mb
Baseline	28.83	15.43	16.0	7.6	14.4
+D-Head	28.91	15.66	28	9.1	27.5
+P2	31.37	16.92	34.9	11.8	28.6
-P5	32.02	17.23	29.5	10.6	14.8
+EIOU	33.07	17.85	29.5	11.3	14.8
+OTA	35.32	19.08	29.5	11.1	14.8
+AUX	36.01	19.47	35.4	10.7	25.8

表 7 从上到下依次添加了本文改进的方法，从表中可以看出，本文设计的组件对检测效果均有提升。其中，本文中添加的 P2 检测分支对 AP_{50} 作用效果最为明显，提升了 2.46%，说明 P2 检测分支对小目标检测的有效性，可以极大减少由于检测分支设定的不合理性造成的漏检情况。剔除掉 P5 检测分支后，计算量、推理时间和模型体积都下降的同时，检测效果不仅没有下降，反而上升了 0.65%，这表明 VisDrone 数据集不适合采用大目标检测分支，更进一步佐证了检测分支的设计应充分考虑数据集待检测物体大小的理念。

3.4. 实验效果

从上述的对比实验中可以看出本文改进的模型在 VisDrone 数据集上的检测效果要优于 YOLOv5 以及其他同等规模的主流模型。

图 8 展示了本文改进的 YOLO-ADOP 模型与使用的基准模型 YOLOv5s 在 VisDrone 测试集上的检测效果对比图。

从对比图结果 1 和结果 3 中可以看出，本文提出的 YOLO-ADOP 模型在识别远处的密集区域小目标时的检测效果要优于 YOLOv5s 模型，能够有效的识别到远处的行人、三轮车等目标，做到最佳的识别效果。从结果 2 可以看出，对于近处的密集目标识别任务，YOLOv5s 只识别到了一个人，而 YOLO-ADOP 能够准确的识别出电动自行车以及车上的两个人，做到识别精度更高、识别误差更小的优点。从结果 4 可以看出，YOLO-ADOP 没有做出对汽车的误判，证明其相较于 YOLOv5s 具有更佳的鲁棒性。综上，本文模型 YOLO-ADOP 拥有更良好的检测效果。

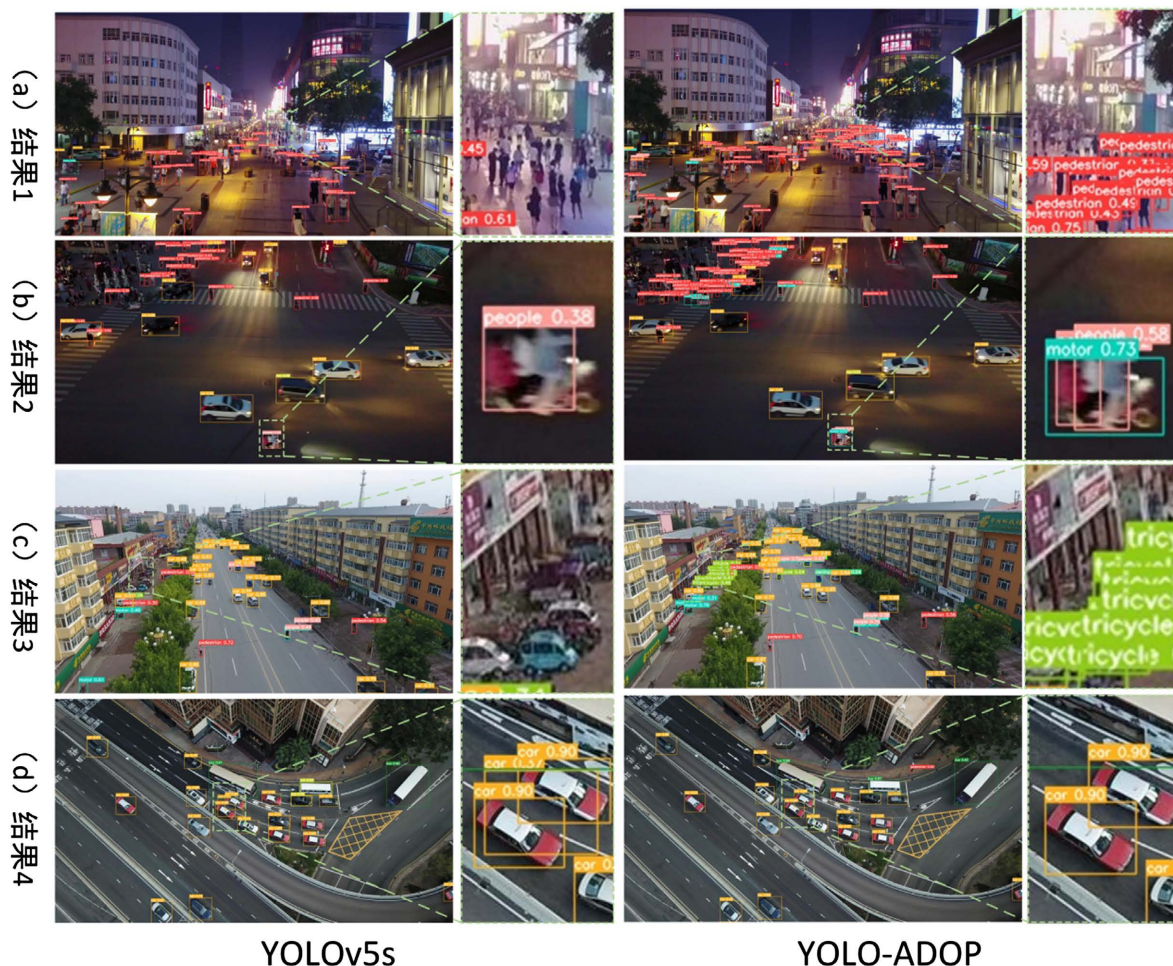


Figure 8. Comparison of detection results between YOLO-ADOP and YOLOv5
 图 8. YOLO-ADOP 和 YOLOv5 检测结果对比图

4. 结论

本文提出了 YOLO-ADOP 算法，用于解决无人机图像目标检测的三大挑战：尺寸变化、高密度和大范围的目标覆盖。通过调整网络识别层、使用解耦头、改进 IOU 计算方式以及 OTA 标签分配机制，我们成功改进了 YOLOv5s 模型，将 VisDrone 数据集上的检测结果提升了 7.18%。这对无人机应用的精确性和广泛应用领域具有重要意义。目前，无人机图像目标检测领域仍在不断探索和优化中，以实现更大的社会价值。

参考文献

- [1] Redmon, J., Divvala, S., Girshick, R., et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [2] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 7263-7271. <https://doi.org/10.1109/CVPR.2017.690>
- [3] Redmon, J. and Farhadi, A. (2018) Yolov3: An Incremental Improvement.
- [4] Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y.M. (2020) Yolov4: Optimal Speed and Accuracy of Object Detection.

- [5] Li, C., Li, L., Geng, Y., *et al.* (2023) YOLOv6 v3.0: A Full-Scale Reloading.
- [6] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/CVPR52729.2023.00721>
- [7] Jocher, G., Chaurasia, A., Stoken, A., *et al.* (2022) Ultralytics/yolov5: v7.0-YOLOv5 SOTA Realtime Instance Segmentation. Zenodo.
- [8] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) Ssd: Single Shot Multibox Detector. *Computer Vision-ECCV 2016: 14th European Conference*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [9] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [10] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149.
- [11] Law, H. and Deng, J. (2018) Cornernet: Detecting Objects as Paired Keypoints. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 734-750. https://doi.org/10.1007/978-3-030-01264-9_45
- [12] Duan, K., Bai, S., Xie, L., *et al.* (2019) Centernet: Keypoint Triplets for Object Detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 6569-6578. <https://doi.org/10.1109/ICCV.2019.00667>
- [13] Ge, Z., Liu, S., Wang, F., *et al.* (2021) Yolox: Exceeding Yolo Series in 2021.
- [14] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [15] Liu, Z., Lin, Y., Cao, Y., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 11-17 October 2021, 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [16] Du, D., Zhu, P., Wen, L., *et al.* (2019) VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, 27-28 October 2019, 213-226.
- [17] Du, D., Qi, Y., Yu, H., *et al.* (2018) The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 370-386. https://doi.org/10.1007/978-3-030-01249-6_23
- [18] Zhang, R., Newsam, S., Shao, Z., *et al.* (2021) Multi-Scale Adversarial Network for Vehicle Detection in UAV Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **180**, 283-295. <https://doi.org/10.1016/j.isprsjprs.2021.08.002>
- [19] Deng, S., Li, S., Xie, K., *et al.* (2021) A Global-Local Self-Adaptive Network for Drone-View Object Detection. *IEEE Trans Image Process*, **30**, 1556-1569. <https://doi.org/10.1109/TIP.2020.3045636>
- [20] Gallo, I., Rehman, A.U., Dehkordi, R.H., *et al.* (2023) Deep Object Detection of Crop Weeds: Performance of YOLOv7 on a Real Case Dataset from UAV Images. *Remote Sensing*, **15**, Article No. 539. <https://doi.org/10.3390/rs15020539>
- [21] Ringwald, T., Sommer, L., Schumann, A., *et al.* (2019) UAV-Net: A Fast Aerial Vehicle Detector for Mobile Platforms. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 16-17 June 2019, 544-552. <https://doi.org/10.1109/CVPRW.2019.00080>
- [22] Zhu, X., Lyu, S., Wang, X., *et al.* (2021) TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 11-17 October 2021, 2778-2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>
- [23] Zhao, Q., Liu, B., Lyu, S., *et al.* (2023) TPH-YOLOv5++: Boosting Object Detection on Drone-Captured Scenarios with Cross-Layer Asymmetric Transformer. *Remote Sensing*, **15**, Article No. 1687. <https://doi.org/10.3390/rs15061687>
- [24] Zhang, Y.-F., Ren, W., Zhang, Z., *et al.* (2022) Focal and Efficient IOU Loss for Accurate Bounding Box Regression. *Neurocomputing*, **506**, 146-157. <https://doi.org/10.1016/j.neucom.2022.07.042>
- [25] Ge, Z., Liu, S., Li, Z., *et al.* (2021) Ota: Optimal Transport Assignment for Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 303-312. <https://doi.org/10.1109/CVPR46437.2021.00037>