光电技术及应用

DOI：10.16818/j.issn1001－5868.2024010903

面向无人机航拍图像的多尺度目标检测研究

贾　亮，林铭文*，戚丽瑾，谈　瑾

（沈阳航空航天大学 电子信息工程学院，沈阳 110136）

摘　要： 针对无人机航拍图像背景复杂、小目标占比高且分布不均导致的现有算法精度不佳等问题，提出了一种面向无人机航拍图像的多尺度目标检测网络VTO-YOLOv8。首先，采用WIoU v3作为边界框回归损失函数，并使用明智的梯度分配策略，这一策略将使网络更加关注普通质量样本，从而提高其定位能力；其次，设计四层T-BiFPN结构，加强浅层特征和深层特征的融合；此外，设计C2f-DBB多分支模块，在不增加计算量的前提下，提升检测性能；同时，使用聚焦调制模块，加强不同尺度信息的交互。实验结果表明，网络在Visdrone2019数据集上相较基准模型在mAP50和mAP指标上分别提高了9.0%和5.9%，同时参数降低了22.6%，可更好地应用于无人机航拍目标检测中。

关键词： 目标检测；无人机图像；特征融合；多分支结构；多尺度目标检测

中图分类号： TP391　**文章编号：** 1001－5868(2024)03－0501－07

Multiscale Target Detection for UAV Aerial Images

JIA Liang，LIN Mingwen，QI Lijin，TAN Jin

（College of Electronic Information Engineering，Shenyang Aerospace University，Shenyang 110136，CHN）

Abstract： A multiscale target detection network，VTO-YOLOv8，for unmanned aerial vehicle（UAV）images is proposed to address the low accuracy of existing algorithms caused by complex backgrounds，a high proportion of small targets，and uneven distributions. First，wise intersection over union（WIoU）v3 was used as the bounding-box regression loss，and a wise gradient allocation strategy was employed for the network to focus more on regular quality samples and improve localization ability. Second，a four-layer target bi-directional feature pyramid network（T-BiFPN）structure was designed to strengthen the integration of shallow and deep features. Furthermore，a faster implementation of CSP bottleneck with diverse branch blocks（C2f-DBB）module was designed to improve the detection performance of the network without increasing computational complexity. In addition，a focal modulation module was used to enhance the interaction of information at different scales. The experimental results demonstrated that the proposed network improved the mean average precision（mAP）and mAP50 by 5.9% and 9.0%，respectively，compared with those of the baseline network on the Visdrone2019 dataset. Moreover，the network parameters were reduced by 22.6%. The proposed method can be applied to target detection in UAV aerial photography.

Key words： object detection；UAV image；feature fusion；multibranch module；multiscale target detection

收稿日期：2024－01－09.

基金项目：国家自然科学基金项目(61671310)；航空科学基金项目(2019ZC054004).

* **通信作者：**林铭文

0 引言

近年来,无人飞行器(UAV)技术得到了迅速发展,其中,无人机图像目标检测作为其核心技术,已广泛应用于电力巡检[1]、生态保护[2]、城市交通监测[3]等场景。然而,无人机图像目标检测不同于常规图像的目标检测。首先,无人机的飞行高度变化较大,使得无人机捕获图像中的物体尺度变化剧烈,小目标的比例远远超过自然场景图像。其次,无人机飞行高度相对较低,空域环境复杂,从而导致无人机捕获的图像目标密度较大且存在物体之间的遮挡现象,这种遮挡现象会导致无人机捕获的图像目标呈现不连续和不完整的特点。最后,无人机具有更大的视场,对地理环境的覆盖更为广阔,但这种大背景不能突出目标的拍摄方式也会给检测任务带来噪声干扰。以上所述的三个问题对目标检测任务造成了极大的挑战,大大增加了目标检测的难度[4-6]。

随着深度学习技术的飞速发展,越来越多的学者将其应用于目标检测任务中,基于深度学习的目标检测方法已逐渐成为主流。基于深度学习的通用目标检测算法主要可分为双阶段算法和单阶段算法。双阶段算法如 R CNN[7-9]系列,具有先定位后识别的特点,在检测精度上有一定的优势,但也存在训练困难、检测速度慢、难以优化等缺点。而单阶段检测算法如 YOLO[10-12],SSD[13-14]系列有着端到端的性能优势,但是其在小目标检测的定位和识别上精度偏低。

为了利用双阶段算法的精度优势,Yang 等[15]提出了 ClusDet 网络,该网络将聚类和检测集成在一个框架内,这种集成显著减少了检测区域的数量,提高了检测效率。Duan 等[16]为解决密集小目标分布不均的问题,提出粗粒度密度图方法 CDMNet 生成密度图,并基于密度图将图像分为多个区域,合理调整密集小目标的尺度。此外,QueryDet[17]通过在低分辨率特征上进行预测,估计小目标的大致位置,然后利用粗略位置在高分辨率特征上进行精确检测,但粗略位置的错误预测可能会导致检测头无法准确定位目标。

针对单阶段算法在小目标检测上的精度劣势,Yang 等[18]对 YOLOX[19]算法进行了改进,引入 TAM 三重关注模块,重新设计了网络的颈部,在物体密集的场景和覆盖面积较大的图像中寻找注意区域,从而增强骨干网提取的特征。Wang 等[20]设计了一个 FFNB 特征处理模块,并在此基础上提出了两种新的检测尺度,使浅层特征和深层特征完全融合,大大提高了模型的检测性能,降低了小目标的漏检率。Pandey 等[21]针对 VisDrone 数据集使用了基于密度图的方法,并引入了增大感受野的结构,显著提升了模型对小物体的检测能力。

综合考虑了航拍图像的特点和单阶段算法端到端的优势,本文选择当前最新的单阶段 YOLO 算法 YOLOv8,针对航拍图像目标检测精度低和小目标漏检的问题,提出了一种基于 YOLOv8 算法改进的网络 VTO-YOLOv8,主要包括以下几点:

1)将 WIoU v3[22]纳入边界盒回归损失中,采用动态非单调机制设计更合理的梯度增益分配策略。WIoU v3 有效降低了高质量样本和低质量样本的梯度增益,增强了模型的定位性能和泛化能力。

2)设计四层的 T-BiFPN 结构,加强多尺度特征的融合,提高了模型对小目标的感知能力和准确性。

3)设计 C2f-DBB 分支采样模块代替颈部的 C2f 模块,在不增加模型计算量的前提下,提升模型的检测性能。

4)引入聚焦调制模块,增强不同粒度级别特征的语义信息交互。

5)在 VisDrone 数据集上对模型的性能进行验证,并与历代 YOLO 系列模型以及其他主流检测模型比较,实验结果证明了本文方法的优越性。

1 VTO-YOLOv8 算法设计

YOLOv8 虽然在目标检测领域有显著的检测效果,准确率和速度均超过了其他的 YOLO 系列模型,但在无人机目标检测上仍然存在小目标漏检的问题。因此本文对 YOLOv8 进行改进,提出了针对无人机目标检测的网络模型 VTO-YOLOv8,改进后的网络结构如图 1 所示。

1.1 WIoU v3 损失函数

YOLOv8 使用 DFL Loss[23]和 CIoU Loss[24]来计算边界盒的回归损失,但 CIoU 有以下缺点:首先,CIoU 不考虑难易样本的平衡;其次,CIoU 将纵横比作为损失函数的罚项之一,如果实际盒体和预测盒体的纵横比相同,但宽度和高度的值不同,则罚项不能反映两个盒体之间的真实差异。CIoU 公式可表示为:

$$L_{\mathrm{CIoU}}=1-IoU+\frac{\rho^2(b,b^{\mathrm{gt}})}{(c_{\mathrm{w}})^2+(c_{\mathrm{h}})^2}+\frac{4}{\pi^2}\left(\tan^{-1}\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\tan^{-1}\frac{w}{h}\right)\tag{1}$$

式中，IoU 为预测框和真实框的交并比，$\rho(b,b^{\mathrm{gt}})$ 为真实框与预测框中心之间的欧氏距离，c_{h} 和 c_{w} 分别表示由预测框和实际框组成的最小围框的高度和宽度，w 和 h 表示预测框的宽和高，w^{gt} 和 h^{gt} 表示真实框的宽和高。

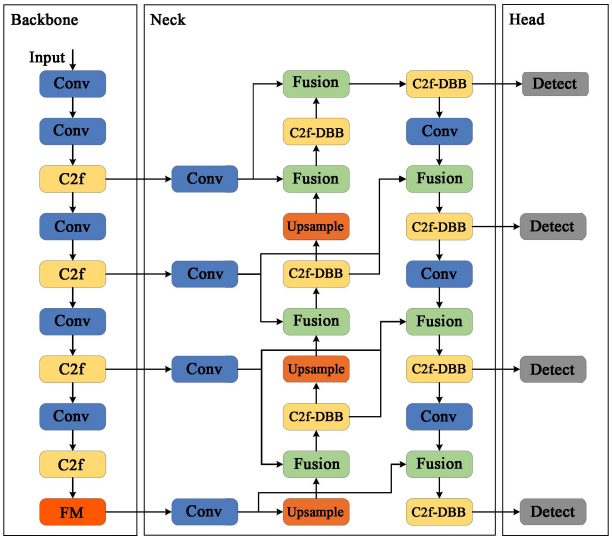


图 1　VTO-YOLOv8 结构示意图

CIoU 采用静态调焦机制，而 WIoU 不仅考虑了面、质心距离和重叠面积，还引入了动态非单调调焦机制。WIoU v3 采用合理的梯度增益分配策略来评价锚盒的质量。

WIoU v1 的计算公式如式(2)～(4)所示。

$$L_{\mathrm{WIoU\ v1}}=R_{\mathrm{WIoU}}\times L_{\mathrm{IoU}^{\omega}}\tag{2}$$

$$R_{\mathrm{WIoU}}=\exp\left(\frac{(x-x_{\mathrm{gt}})^2+(y-y_{\mathrm{gt}})^2}{(W_{\mathrm{g}}^2+H_{\mathrm{g}}^2)^2}\right)^{\omega}\tag{3}$$

$$L_{\mathrm{IoU}}=1-IoU_{\omega}\tag{4}$$

式中，x，y 和 x_{gt}，y_{gt} 分别是预测框和真实框的中心点坐标，W_{g} 和 H_{g} 是预测框和真实框最小围框的宽和高。

WIoU v3 定义了一个离群度 β 以描述锚框的质量，锚框质量越高，离群度越高。利用 β 构造了一个非单调聚焦系数 r，对离群度较大的锚框分配较小的梯度增益，有效防止低质量示例产生较大的有害梯度。将 r 应用于 WIoU v1 就得到了具有动态非单调聚焦系数的 WIoU v3。WIoU v3 采用合理的梯度增益分配策略，动态优化损失中高质量和低质量锚盒的权重，使模型专注于平均质量样本，提高了模型的整体性能。WIoU v3 公式如式(5)～(7)所示。式(7)中的 β 和 δ 为超参数，可适当调整以适应不同的模型。

$$\beta=\frac{L_{\mathrm{IoU}}^{*}}{L_{\mathrm{IoU}}}\in[0,+\infty)\tag{5}$$

$$r=\frac{\beta}{\delta\alpha^{\beta-\delta}}\tag{6}$$

$$L_{\mathrm{WIoUv3}}=rL_{\mathrm{WIoUv1}}\tag{7}$$

1.2　T-BiFPN 加权双向特征金字塔结构

YOLOv8 在特征融合网络中采用了 PANet[25] 结构，PANet 引入了双向通路结构，使用自下而上和自上而下的双向路径实现不同层次特征信息的融合。然而，不同层次的特征具有不同的分辨率，其对输出特征的贡献也不尽相同，仅靠简单的拼接无法准确区分语义信息的重要性，难以有效提高小目标检测的精度。为此，本文对 BiFPN[26] 进行改进，设计了四层的 T-BiFPN 结构。

BiFPN 通过引入更高效的双向跨尺度连接和动态加权特征图融合，实现了更有效的特征融合。然而，BiFPN 在特征融合网络的底层结构对浅层特征不友好，减少了浅层特征在融合时的权重，而浅层的高分辨率特征对于小目标检测尤为重要。T-BiFPN 加入小目标层，并在小目标的浅层特征路径中添加中间节点，提升浅层特征的利用率，更有益于无人机航拍图像中的小目标检测。T-BiFPN 结构如图 2 所示。

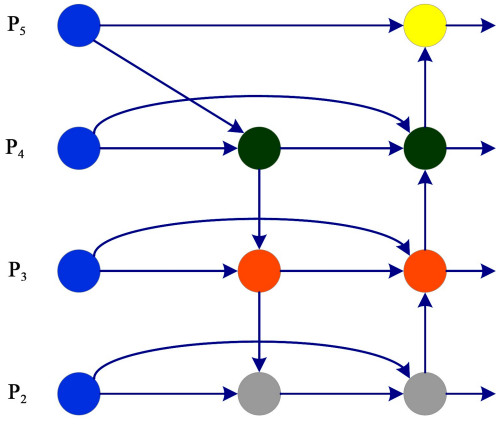


图 2　T-BiFPN 结构示意图

融合过程以图 2 中 P_4 特征融合为例，具体计算公式为：

$$P_4^{\mathrm{td}}=\mathrm{Conv}\left(\frac{w_1\cdot P_4^{\mathrm{in}}+w_2\cdot Resize(P_5^{\mathrm{in}})}{w_1+w_2+\dot{o}}\right)\tag{8}$$

$$P_4^{\mathrm{out}}=\mathrm{Conv}\left(\frac{w'_1\cdot P_4^{\mathrm{in}}+w_2'\cdot P_4^{\mathrm{td}}+w'_3\cdot Resize(P_3^{\mathrm{out}})}{w'_1+w'_2+w'_3+\dot{o}}\right)\tag{9}$$

式中，P_4^{td} 为自顶向下路径第 4 层的中间特征，P_4^{out} 为自底向上路径第 4 层的输出特征，w_i 为可学习的权重，ò 是一个很小的参数，以防止数值不稳定，数值为 0.000 1。

1.3 C2f-DBB 模块

DBB(Diverse Branch Block)[27] 是一种重参化技术，在训练阶段以多分支结构丰富卷积块的特征空间。这些分支结构包括平均池化、多尺度卷积等，同时新增了卷积串联结构，从而提供了更多的选择。在推理阶段，所有的多分支结构被重参化为一个主分支，这也是结构化模型重参化的核心思想。这种等价变换可作为一个嵌入式模块，灵活地应用于现有的网络结构中。在训练阶段，使用多个分支替代传统的卷积操作，为网络引入更多的分支结构，从而提高模型的精度。而在推理阶段，将 DBB 模块转换为等效的卷积操作，实现了模型精度的提高而不引入额外的计算消耗。DBB 模块结构如图 3 所示。

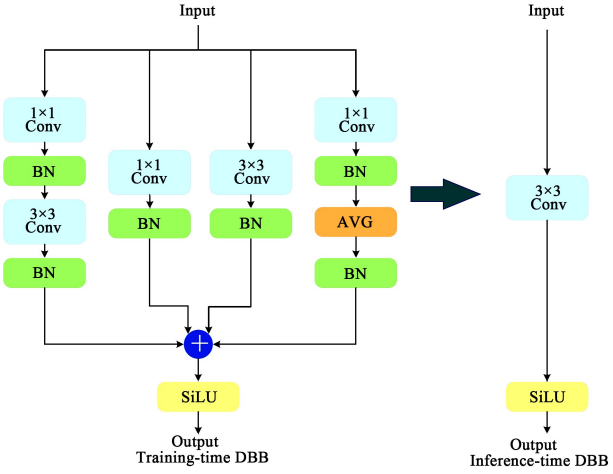


图 3 DBB 模块结构示意图

借此，本文利用 DBB 的思想对 C2f 的结构进行重新构造，设计了 C2f-DBB。C2f 模块中含有多个 bottleneck 块，使用 DBB 多分支结构代替 bottleneck 中的 3×3 卷积，在训练阶段使模型拥有更强的特征获取能力，在推理阶段时将多分支结构重参化，等价转换为单个卷积层，简化模型结构，保持与 C2f 相同的结构，从而在不增加任何参数和推理时间的情况下，实现性能的提升。C2f-DBB 结构如图 4 所示。

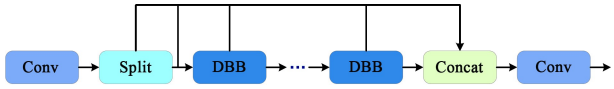


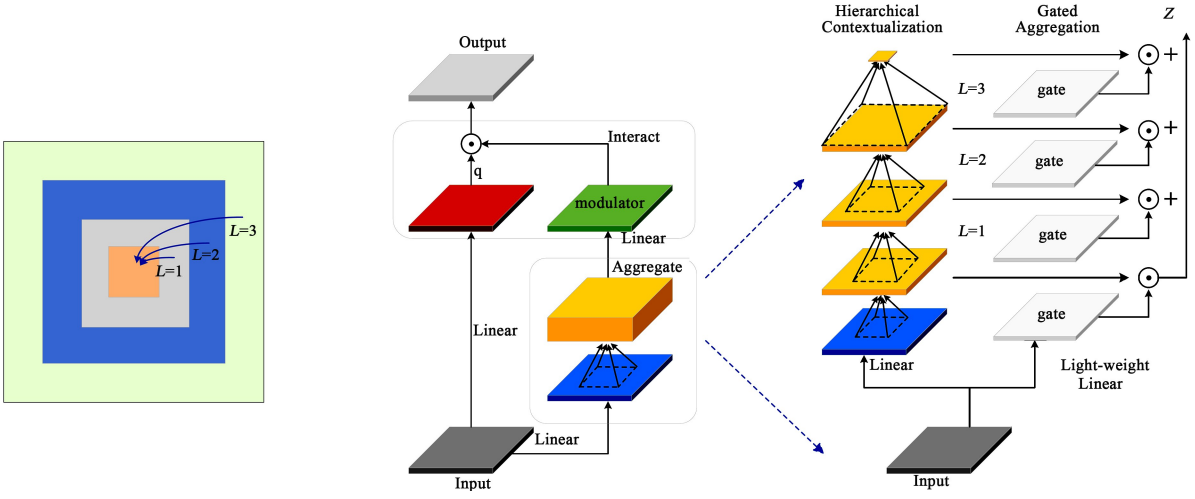
图 4 C2f-DBB 结构示意图

1.4 聚焦调制模块

SPPF 模块采用多个小尺寸池化核生成不同感受野的特征图，通过级联操作将不同感受野的特征图进行融合，实现局部特征和全局特征的融合。然而，这种交互方式存在以下不足之处：获得的空间信息不够丰富，并且简单拼接融合方式无法准确区分目标检测任务中不同语义特征的重要性。

针对于此，使用聚焦调制模块（Focal Modulation[28]，FM）替代 SPPF 模块。FM 模块用于空间的语义信息交互，通过多层次的上下文聚合自然地捕捉了短距离和长距离之间的关系，从而取得了更好的准确性和效率的平衡。

FM 的结构相对 SPPF 模块，聚合的方式可以获得更为丰富的空间信息，动态自适应的融合方法更为高效，可提高检测的准确性和鲁棒性。聚焦调制结构和原理如图 5 所示。



(a) 结构图 (b) 原理图

图 5 Focal Modulation 结构和原理图

FM 模块通过门控聚合将不同粒度级别的上下文特征浓缩为单个特征向量，具体来说，使用线性层来获得空间和级别感知的权重，然后，通过元素相乘执行加权和，以获得与输入大小相同的单个特征映射，整体的焦点调制公式为：

$$y_i = q(x_i) \odot h\left(\sum_{l=1}^{L+1} g_i^l, z_i^l\right) \tag{10}$$

式中，$q(\cdot)$是一个查询映射函数，$\odot$是按元素的乘法运算符，$h(\cdot)$为调制函数。g_i^l 和 z_i^l 分别为位置 i 处第 l 级的感知权重和对应的视觉特征。

2　实验设计

2.1　数据集及数据分析

实验采用天津大学机器学习与数据挖掘实验室收集的 VisDrone2019 数据集。VisDrone 无人机数据集由多架无人机倾斜俯视在不同场景、天气和光照条件拍摄而成，拍摄地点涵盖了中国 14 个不同城市。该数据集包含了 10 209 张图片，其中训练集有 6 480 张，验证集有 548 张，测试集有 3 190 张，包含 10 种目标类别共 260 万标注信息，其中目标小、密度高、类别易混淆、分布不均的特点对算法的设计提出了极大挑战。

2.2　实验环境和实验策略

所采用的服务器配置如下：Intel(R) Xeon(R) Platinum 8255C CPU，内存 43 GB，显卡为 NVIDIA RTX3090 (24 GB)，服务器配置操作系统为 Ubuntu20.04，运行环境为 PyTorch 1.11.0，CUDA 11.3，Python 3.8。

模型参数配置如下：实验采用 Mosaic 数据增强，总共训练 300 个 epoch，前 3 个 epoch 进行热身，使用 SGD 梯度下降优化器，初始学习率为 0.01，动量为 0.937，批大小为 16。

2.3　评价指标

为了准确评估模型的检测性能，本文采用准确率 P、召回率 R、平均精度均值(mAP)、模型参数、计算量和单张图片推理时间等作为评估指标。以下参数用于上述一些评价指标的公式中：TP 为正确检测数量，TP 为错误检测数量，FN 为漏检数量，相应的公式如式(11)～(14)所示。

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 P \mathrm{d}R \tag{13}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{14}$$

在上述公式中，准确率 P 指的是在所有检测结果中正确检测数量所占比例，召回率 R 表示数据集目标中实际检测出的比例，平均精度 AP 可表示为 P-R 曲线的面积，通过对各类别的值求平均得到平均精度均值(mAP)。$mAP50$，$mAP75$ 和 mAP 分别为检测框与真实框的交并比为 0.5，0.75 和 0.5∶0.95(即从 0.5 至 0.95，步长为 0.05)的总类别平均检测准确率。

2.4　实验结果

2.4.1　损失函数对比实验

为了验证 WIoU v3 对于航拍图像目标检测的优越性，在保持其他训练条件一致的前提下，使用 WIoU v3 和一些主流损失函数对 YOLOv8n 进行了对比实验。实验结果如表 1 所示。

表 1 实验结果表明，以 WIoU v3 作为边界框回归损失时，模型的检测性能最好，准确率 P 和 $mAP50$ 均取得了最好的成绩，mAP 仅次于 GIoU。此外，该模型使用 WIoU v3 时的 $mAP50$ 相较于使用 CIoU 时高 1.1%，这证明了引入 WIoU v3 的有效性。

表 1　不同损失函数下 YOLOv8n 网络精度对比

方法	P/%	R/%	$mAP50$/%	mAP/%
CIoU	45.2	33.9	34.0	19.6
DIoU	46.2	34.0	34.5	20.1
GIoU	46.7	34.1	34.8	20.2
WIoU v1	44.8	34.9	34.4	19.8
WIoU v2	44.6	34.8	34.3	19.6
WIoU v3	46.7	34.4	35.1	20.1

2.4.2　消融实验

为了进一步验证本文所提出的每种改进方法的有效性，使用 YOLOv8n 作为基线模型进行了消融实验，对应的实验内容和结果如表 2 所示。

对表 2 消融实验的结果进行分析可知，各项改进方法对基线模型性能均有不同程度的提升。使用 WIoU v3 作为损失函数利用动态的权重分配策略，提高了模型的分类能力，使 $mAP50$ 提高了 1.1%。BiFPN 的特征融合网络结构使模型的参数减少了 33.9%，计算量降低 12.3%，除此之外模型的精度也有小幅的提升。C2f-DBB 的引入在不增加模型参数和计算负担的前提下，使 mAP 和 $mAP50$ 均提

高了 0.5%。当前使用的数据含有大量的小目标，使用四层的 T-BiFPN 结构大大增强了模型的小目标检测能力，*mAP* 和 *mAP*50 分别提升了 3.7%和 5.3%。FM 模块的引入加强了局部特征和全局特征的交互，*mAP* 和 *mAP*50 进一步提升了 1.0%和 1.8%。

表 2 消融实验结果

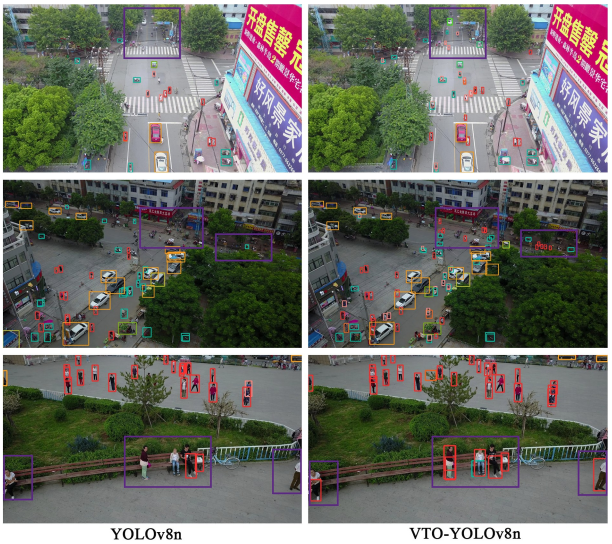
样品	baseline	WIoU v3	BiFPN	C2f-DBB	T-BiFPN	FM	*mAP*50/%	*mAP*/%	参数量/MB	GFLOPs	Infer/ms
A	√						34.0	19.6	3.01	8.1	10.7
B	√	√					35.1	20.1	3.01	8.1	10.8
C	√	√	√				35.4	20.3	1.99	7.1	13.1
D	√	√	√	√			35.9	20.8	1.99	7.1	13.1
E	√	√		√	√		41.2	24.5	2.22	18.2	15.0
F	√	√		√	√	√	43.0	25.5	2.33	18.3	15.4

本文提出的改进网络有效的提升了模型的检测性能，相较于原始的 YOLOv8n 检测模型，VTO-YOLOv8n 的 *mAP*50 提高了 9.0%，*mAP* 也提高了 5.9%。同时，模型的参数降低了 22.6%，尽管计算量和推理时间有所增加，但相较于模型性能的大幅提升，这是完全值得的。

2.4.3 可视化分析

为了直观的对本文所提出改进方法进行效果评估，使用 YOLOv8n 和 VTO-YOLOv8n 对十字路口、市场路口、公园三个具有代表性的场景进行检测实验对比，对比 YOLOv8n 显著改进的区域使用紫色框图进行标注，检测对比效果如图 6 所示。

由图 6 对比效果可知，在第一张十字路口和第二张市场路口的检测中，YOLOv8n 存在对远端小尺度目标漏检十分严重的问题，而本文提出的 VTO-YOLOv8n 相对基准模型针对该区域小目标漏检的问题有了极大改善。在第三张公园场景下的检测中，YOLOv8n 存在对图像边缘和座椅等复杂背景下行人的漏检问题，VTO-YOLOv8n 则成功检测出以上漏检行人。



YOLOv8n VTO-YOLOv8n

图 6 YOLOv8n 与 VTO-YOLOv8n 检测效果对比

2.4.4 对比实验

为了对模型进行全面的评估，进行了两组对比实验以证明所提改进方法的优越性和有效性。

在第一组对比实验中，将 VTO-YOLOv8n 模型与主流的 YOLOv5，YOLOv7 和 YOLOv8 系列的轻量模型进行了对比，相应的对比结果如表 3 所示。在第二组对比实验中，将所提出的一系列改进方法应用于更大规模的 YOLOv8s 中得到 VTO-YOLOv8s，并将其与其他优秀的先进算法进行比较，结果如表 4 所示。

表 3 与 YOLO 系列算法的对比实验结果

模型	*mAP*50/%	*mAP*/%	Param/MB	GFLOPs
YOLOv5n	25.7	12.9	1.77	4.2
YOLOv5s	32.9	17.8	1.99	7.1
YOLOv7tiny	34.8	18.2	6.03	13.1
YOLOv8n	34.0	19.6	3.01	8.1
YOLOv8s	40.1	23.9	11.13	28.5
VTO-YOLOv8n	43.0	25.5	2.33	18.3

表 4 与其他先进算法的对比实验结果

模型	*mAP*50/%	*mAP*75/%	*mAP*/%
Cascade R-CNN	39.9	23.4	23.2
Faster R-CNN	40.0	20.6	21.5
QueryDet	48.1	28.8	28.3
CDMNet	49.5	29.8	29.2
CLusDet	50.6	24.7	26.7
VTO-YOLOv8s	50.9	32.5	31.2

由对比实验结果表 3 可知，本文所提出的 VTO-YOLOv8n 在主流的 *mAP*50 和 *mAP* 指标上均优于以上 YOLO 系列模型。值得一提的是，VTO-YOLOv8n 是本文在 YOLOv8n 上进行改进的检测模型，经过实验所提出的一系列改进方法，VTO-YOLOv8n 在以上所述指标上对更大规模的 YOLOv8s 模型实现了全面的超越。其中，VTO-YOLOv8n 在参数和计算量分别仅为 YOLOv8s 的 20.9%和 64.2%的情况下，在 *mAP*50 和 *mAP* 指标上相较于 YOLOv8s 分别提升了 2.9%和 1.6%，这可以充分证明本文所提出改进方法的优越性和有效性。

由对比实验结果表 4 可知，本文将 VTO-YOLOv8s 与其他先进算法进行了比较，VTO-YOLOv8s 在 mAP，$mAP50$，$mAP75$ 三个指标上均大幅领先于 Cascade R-CNN 和 Faster R-CNN。与 QueryDet，CDMNet 和 CLusDet 相比，VTO-YOLOv8s 在 $mAP50$ 指标上领先幅度较少，但在 mAP 和 $mAP75$ 两个更严苛的指标上有显著的精度优势，这进一步证明了所提出改进方法对无人机航拍目标检测任务的适用性。

3　结论

针对无人机航拍场景下的目标检测任务，现有多数模型存在检测精度较低、小目标漏检严重等问题，本文在 YOLOv8 基础上引入 WIoU v3 损失函数和 FM 模块，设计 T-BiFPN 的四层金字塔结构和 C2f-DBB 模块，提出一种无人机航拍目标检测网络 VTO-YOLOv8。在 VisDrone2019 数据集中的实验结果表明，VTO-YOLOv8n 相对基准模型 YOLOv8n 和其他 YOLO 系列轻量模型有着显著的优势。此外，将改进方法应用于 YOLOv8s 得到的 VTO-YOLOv8s 也表现出良好的性能，检测精度优于对比的先进算法，可更好应用于无人机航拍场景下的目标检测任务。

在下一步的研究重点中，模型对于小目标的检测精度还有进一步提高的空间，另一方面，将考虑通过剪枝蒸馏和更换卷积等方式在保持较高精度的前提下，降低模型的参数和计算消耗。

参考文献：

[1] Wang Q, Wang W, Li Z, et al. Close-range transmission line inspection method for low-cost UAV: Design and implementation[J]. Remote Sens., 2023, 15(19): 4841.

[2] Assiri M, Sartori A, Persichetti A, et al. Leaf area index and aboveground biomass estimation of an alpine peatland with a UAV multi-sensor approach[J]. GISci. Remote Sens., 2023, 60(1): 2270791.

[3] Zhang G. 6 G enabled UAV traffic management models using deep learning algorithms[J]. Wireless Netw., 2023, 1-11.

[4] 陈　旭，彭冬亮，谷　雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. 光电工程，2022，49(3)：210372-1-210372-13.
Chen Xu, Peng Dongliang, Gu Yu. Real-time object detection for UAV images based on improved YOLOv5s[J]. Opto-Electronic Engineering, 2022, 49(3): 210372-1-210372-13.

[5] 张云佐，武存宇，刘亚猛，等. 联合自注意力和分支采样的无人机图像目标检测[J]. 光学精密工程，2023，31(18)：2723-2735.
Zhang Yunzuo, Wu Cunyu, Liu Yameng, et al. Joint self-attention and branch sampling for object detection on drone imagery[J]. Optics and Precision Engineering, 2023, 31(18): 2723-2735.

[6] 李　程，车文刚，高盛祥. 一种用于航拍图像的目标检测算法[J]. 山东大学学报(理学版)，2023，58(9)：59-70.
Li Cheng, Che Wengang, Gao Shengxiang. An object detection algorithm for aerial images[J]. J. of Shandong University (Natural science), 2023, 58(9): 59-70.

[7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2014: 580-587.

[8] Girshick R. Fast R-CNN[C]// Proc. of the IEEE Inter. Conf. on Computer Vision, 2015: 1440-1448.

[9] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards realtime object detection with region proposal networks[J]. Adv. in Neural Inf. Process. Syst., 2015, 39(6): 91-99.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition, 2016: 779-788.

[11] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]// Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition, Piscataway, 2017: 7263-7271.

[12] 郁　强，王　宽，王　海. 一种多尺度 YOLOv3 的道路场景目标检测算法[J]. 江苏大学学报(自然科学版)，2021，42(6)：628-633，641.
Yu Qiang, Wang Kuan, Wang Hai. A multiscale YOLOv3 object detection algorithm[J]. J. of Jiangsu University: Natural Science Edition, 2021, 42(6): 628-633, 641.

[13] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// Computer Vision-ECCV 2016, 2016: 21-37.

[14] Zhang S, Wen L, Bian X, et al. Single-shot refinement neural network for object detection[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 4203-4212.

[15] Yang F, Fan H, Chu P, et al. Clustered object detection in aerial images[C]// Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 2019: 8311-8320.

[16] Duan C, Wei Z, Zhang C, et al. Coarse-grained density map guided object detection in aerial images[C]// Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 2021: 2789-2798.

[17] Yang C, Huang Z, Wang N. QueryDet: Cascaded sparse query for accelerating high-resolution small object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2022: 13668-13677.

（下转第 514 页）

algorithms[J]. Commun. ACM, 1989, 32(3): 359-373.
[10] Bae H, An Y K. Computer vision-based statistical crack quantification for concrete structures[J]. Measurement, 2023, 211: 112632.
[11] Tang Y, Huang Z, Chen Z, et al. Novel visual crack width measurement based on backbone double-scale features for improved detection automation[J]. Eng. Struct., 2023, 274: 115158.

作者简介：

代少升(1974—)，男，河南省潢川县人，博士，教授，主要从事红外成像系统及 SOPC 嵌入式系统设计与开发；

余自安(1974—)，男，河南省光山县人，高级工程师，主要从事机械设计及理论研究。

(上接第 507 页)

[18] Yang Y, Gao X, Wang Y, et al. VAMYOLOX: An accurate and efficient object detection algorithm based on visual attention mechanism for UAV optical sensors[J]. IEEE Sens. J., 2022, 23(11): 11139-11155.
[19] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. 2021: arXiv: 2107.08430. https://arxiv.org/abs/2107.08430.
[20] Wang G, Chen Y, 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.
[21] Pandey V, Anand K, Kalra A, et al. Enhancing object detection in aerial images[J]. Math. Biosci. Eng., 2022, 19(8): 7920-7932.
[22] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[EB/OL]. 2023: arXiv: 2301.10051. https://arxiv.org/abs/2301.10051.
[23] Li X, Wang W, Wu L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Adv. in Neural Inf. Process. Syst., 2020, 33: 21002-21012.
[24] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]// Proc. of the AAAI Conf. on Artificial Intelligence, 2020: 12993-13000.
[25] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 8759-8768.
[26] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020: 10781-10790.
[27] Ding X, Zhang X, Han J, et al. Diverse branch block: Building a convolution as an inception-like unit[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2021: 10886-10895.
[28] Yang J, Li C, Dai X, et al. Focal modulation networks[J]. Adv. in Neural Inf. Process. Syst., 2022, 35: 4203-4217.

作者简介：

贾　亮(1971—)，男，辽宁省大石桥市人，副教授，硕士生导师，主要从事信息获取与处理方面的研究；

林铭文(2000—)，男，福建省福州市人，硕士研究生，主要从事图像处理与目标检测方向的研究。