

基于改进 YOLOv8 的无人机航拍检测算法

邓世强, 简季

(成都理工大学 地理与规划学院, 成都 610059)

摘要: 针对当前的目标检测算法对无人机影像存在背景复杂、目标较小而造成的漏检误检等问题, 提出了一种基于 YOLOv8 的无人机航拍检测改进算法 YOLOv8-A; 为了提高算法对于小目标的检测能力, 使用全局通道-空间注意力模块来提高输入特征图的表达能力; 为了进一步提高算法对于捕获输入数据的上下文信息和内在特征的能力, 在颈部结构更换了 RT-DETR 算法的 AIFI 模块, 有助于提取上下文的信息; 为了提高算法对于困难样本的定位能力, 采用了 WIoU 函数来提高算法对于困难样本的定位稳定性; 在 Visdrone2019 数据集上的实验表明, 所改进的算法与原有算法相比, mAP 提升了 2.1%, 相比其他主流的算法, 该算法有更好的检测效果。

关键词: 目标检测; 无人机; YOLOv8; 注意力机制; AIFI; WIoU

Aerial Photography Detection Algorithm for Drones Based on Improved YOLOv8

DENG Shiqiang, JIAN Ji

(School of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China)

Abstract: Aiming at the problems such as missed detection and false detection caused by complex background and small targets in current object detection algorithms for unmanned aerial vehicle (UAV) images, an improved UAV aerial photography detection algorithm based on YOLOv8, YOLOv8-A, is proposed; To improve the detection ability of the algorithm for small targets, a global channel-spatial attention module is used to enhance the expressive ability of input feature maps. To further improve the algorithm's ability to capture the context information and intrinsic features of input data, the AIFI module of the real-time detection transformer (RT-DETR) algorithm is replaced in the neck structure, which is helpful for extracting contextual information. To enhance the algorithm's positioning ability for difficult samples, the WIoU function is adopted to improve the algorithm's positioning stability for difficult samples. Experiments on the visdrone2019 dataset show that, the improved algorithm increases the mAP by 2.1% compared to the original algorithm, and exhibits superior in detection performance to other mainstream algorithms.

Keywords: object detection; UAV; YOLOv8; attention mechanism; AIFI; WIoU

0 引言

配备光学遥感相机的无人机具有操作简单、成本低、限制少等特点, 它们可以轻松采集低空遥感图像, 并已广泛应用于智慧城市、灾后救援、农业监测和交通管理等多个领域。目前, 随着无人机和光学相机的快速发展, 无人机图像中捕捉到的物体细节越来越丰富, 这对无人机图像的理解和解读提出了更高的要求^[1]。目标检测是计算机视觉的一项基本任务, 它有助于估计场景中对象的类别并标记它们的位置。由此可见, 开展无人

机影像目标检测的研究, 具备显著的理论意义和实践意义。

当前的目标检测领域主要有两种方法: 基于深度学习的目标检测方法和传统的检测方法。传统方法包括 HOG、SUF 和 DPM 等, 这些方法需要手动设置滑动窗口遍历图像的所有可能位置和尺度, 然后对每个窗口提取特征并分类。但传统的方法大多是依靠人工进行特征提取, 无法应用于复杂的场景并且效率较低, 已经逐渐被基于深度学习的目标检测方法所替代。与传统检测

收稿日期: 2025-08-28; 修回日期: 2025-10-20。

作者简介: 邓世强(2001-), 男, 硕士研究生。

通讯作者: 简季(1972-), 男, 博士后, 教授。

引用格式: 邓世强, 简季. 基于改进 YOLOv8 的无人机航拍检测算法[J]. 计算机测量与控制, 2026, 34(8): 98-104.

方法相比, 基于深度学习的方法能够实现从端到端的检测和识别, 可以自动提取特征信息, 能够大幅度提升检测的精度和效率。基于深度学习的检测方法中 YOLO 系列是最典型的代表, 例如: YOLOv5^[2]、YOLOv6^[3]、YOLOv7^[4]、YOLOv8^[5]等。

近年来, 越来越多的研究人员正在采用深度卷积神经网络来检测和分类无人机影像。尽管一些基于 SSD^[6]的目标检测方法和用于评估焦损有效性的 RetinaNet^[7], 在一定程度上提升了小目标的检测能力, 但无法满足当前数据量大、精度要求高的应用需求。而 YOLOv8-nano^[8]检测以及基于 Swin Transformer^[9]和 YOLOv5 的改进框架, 虽然它们能够显著降低模型的参数和大小、提升小目标检测精度, 但其在复杂环境下的鲁棒性还有待提升。近期, 许多研究人员开始聚焦于改进 YOLO 算法来进行目标检测, 一些基于 YOLO 的改进算法不断涌现, 它们采用各种方法来提高小目标检测能力或解决小目标漏检问题, 但小目标漏检、计算冗余以及复杂背景干扰等问题仍未得到彻底解决^[10-13]。

YOLOv8 作为 YOLO 系列中的最具代表性的算法, 在检测精度和速度上相比于其他系列都具有很大的优势, 该算法很适合对无人机影像目标检测进行改进。因此, 本文选择了对 YOLOv8s 算法进行了优化, 提出了 YOLOv8-A 算法, 以此来进一步提高算法的准确性, 具体工作包括: 使用了全局通道-空间注意力模块 (GCSA, global channel-spatial attention) 来提高特征图的空间信息表达能力; 将 RT-DETR 算法中的 AIFI 模块整合到 YOLOv8 的颈部, 替换了原始的 SPPF 模块, 增强网络对于上下信息的提取能力, 提高对于小目标检测的准确度; 使用 WIoU 函数来提高困难样本的强化加权, 捕捉长距离的特征信息, 在不增加算法参数数量的同时提高了检测的准确度。最后, 将改进的算法在 Vis-drone2019 数据集上进行实验, 结果表明, 相比于 YOLOv8s, 本文所提出的改进算法在平均精度 mAP 上提升了 2.1%, FPS 只降低了 1.9 ms, 该结果表明了所提出算法的有效性。

1 YOLOv8 算法

YOLO 系列作为单阶段检测的代表, 在检测精度和速度之间保持了良好的平衡, 广泛地受到业界的青睐。YOLOv8 是 YOLO 系列中的最经典的基线算法。YOLOv8 采用了无锚点的思想, 使得算法不再过分依赖人工设计, 在检测多尺度、小目标时更加灵活^[14]。YOLOv8 一共有 5 个版本: YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x, YOLOv8s 兼顾了检测精度和内存小的特点。YOLOv8s 算法由输入、骨

干网络、颈部结构和头部这 4 个部分组成, 如图 1 所示。

1) 输入: YOLOv8 的输入部分会将输入图像的大小调整为典型的 $640 \times 640 \times 3$, 随后对图像进行 Mosaic 增强、张量转换等预处理操作。

2) 主干网络 (Backbone): 用来提取特征的重要部分, 使用了 CSPDarkNet53 作为特征提取网络, 该网络借鉴了 CSPDarkNet 的思想而改进生成, 其主要由跨阶段部分融合模块 (C2F, faster implementation of CSP bottleneck with two convolutions)、空间金字塔池化快速模块 (SPPF, spatial pyramid pooling-fast) 和标准卷积块 (CBS, standard convolution with Silu activation) 组成。

3) 颈部结构 (Neck): 颈部结构位于主干网络和头部之间, 负责高效融合主干网络提取的多尺度特征, 并增强对不同尺寸目标的检测能力。通过 FPN+PAN 的组合可以将不同层的特征进行融合, 提高了对目标的检测能力。

4) 头部 (Head): 头部是目标检测的核心部分, 负责将颈部结构输出的多尺度特征解码后进行分类和预测。头部由 3 个检测模块构成, 每个模块由两个分支构成, 每个分支包含两个 CBS、一个二维卷积和边界框损失函数。3 个检测模块有多个输出层, 可以输出不同分辨率的特征图, 在实现了轻量化的同时, 检测的精度也得到了提升。

2 YOLOv8-A 整体结构

在这项研究中, 本文构建了一个基于 YOLOv8s 的检测算法—YOLOv8-A。该算法主要有 3 项改进: 首先, 针对无人机影像中的小目标使用了全局通道-空间注意力模块 (GCSA) 来提高算法对小目标检测的精度; 然后, 将颈部结构换成 AIFI 模块来提高算法提取特征图中上下文信息的能力; 最后对损失函数进行了改进来提高算法对于小目标和困难样本定位的稳定性。

2.1 全局通道-空间注意力模块 (GCSA)

在传统的卷积神经网络架构中, 通道之间的依赖关系很容易被忽略, 这会导致无法充分利用特征图中的信息。为此, 本文使用了全局通道-空间注意力模块 (GCSA), 旨在强化通道间的依赖关系, 增强 YOLOv8s 的识别性能。该模块的原理如图 1 所示。

从结构上来看, 该模块主要由输入特征、通道注意力模块、通道洗牌机制、空间注意力模块和输出特征图 5 个部分构成。在这个过程中, 输入的特征图首先被输送到通道注意力模块中, 被转置操作处理后, 特征图的维度从 $C \times H \times W$ 转换为 $W \times H \times C$, 随后 MLP 多层

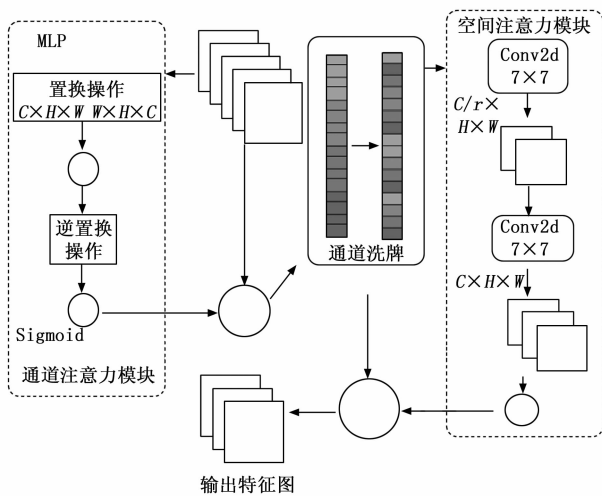


图 1 GCSA 原理图

感知器会加强特征图提取信息的能力，最后在经过逆转置操作将维度转换到原来的初始维度后，Sigmoid 激活函数会生成通道注意力图。通道注意力图会和输入的特征图进行相乘的操作，得到增强后的特征图：

$$F_{\text{channel}} = \sigma\{\text{MLP}[\text{Permute}(F_{\text{input}})]\} \odot F_{\text{input}} \quad (1)$$

其中： F_{channel} 为增强后的特征图， σ 为 Sigmoid 函数， $\odot$ 为逐元素相乘， F_{input} 为原始输入特征图。

增强的特征图随后进行通道洗牌的操作：首先将特征图分成 4 组，随后会对各组进行转置操作，并将特征图的通道进行随机分配，最后还原到初始的维度。通过该机制后，特征图的通道信息会得到进一步的增强，即：

$$F_{\text{shuffle}} = \text{ChannelShuffle}(F_{\text{channel}}) \quad (2)$$

其中： F_{shuffle} 为混洗后的特征图， F_{channel} 为输入特征图的通道数。

最终将特征图引入空间注意力模块进行处理：首先，输入特征图经过一个 7×7 卷积层，通道数被压缩为原有规模的 $(1/4)$ 倍；其次，通过 ReLU 激活函数对特征图进行非线性变换；随后通过第二个 7×7 卷积层将通道数还原至原始维度 C ；最后利用 Sigmoid 激活函数生成空间注意力图。将洗牌后的特征图和空间注意力图按元素相乘，得到最终的输出特征图^[16]，即：

$$F_{\text{spatial}} = \sigma\{\text{Conv}\{\text{BN}\{\text{ReLU}[\text{Conv}(F_{\text{shuffle}})]\}\}\} \odot F_{\text{shuffle}} \quad (3)$$

2.2 AIFI 模块

空间金字塔快速池化 (SPPF) 模块是 YOLOv8 中提高特征提取效率和性能的关键组件。通过汇集不同尺度的特征，SPPF 有效地捕获上下文信息，提高算法的空间不变性^[17]。AIFI 模块是 RT-DERT 实现尺度内特征交互的核心。AIFI 本质是一个 Transformer 编码器，主要通过多头注意力机制 (MHSA, multiheads atten-

tion) 捕获局部特征。MHSA 将输入序列分布在多个子注意力头上，并独立计算注意力图，这种方法在提高性能方面非常有效^[18]。AIFI 模块将自注意力机制应用于高级的特征，这能帮助网络更高效地获取图像中不同序列的目标的信息，有助于后续的模块更好的识别图像的信息。相比于 SPPF 模块，AIFI 模块利用 MHSA 机制避免了多次卷积而带来的数据丢失的问题，提高了检测精度的同时还增加了后续处理中有用的参数量。AIFI 模块的结构如图 2 所示。

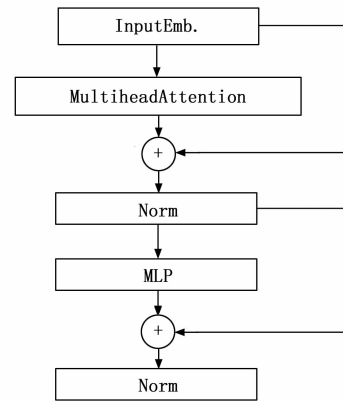


图 2 AIFI 模块结构图

AIFI 模块的特征融合过程如图 3 所示。算法首先将 S5 特征图线性化并排列成一个长向量，随后 MHSA 处理图像中复杂实体关系，通过层层的归一化来提取图像中各个物体的语义特征。然后，特征通过 MLP 进行非线性变换，最后，通过再次归一化生成特征信息的注意力分数并将其重塑为二维形式，表示为 F5。AIFI 模块使用其丰富的语义提取能力和捕捉长距离依赖关系的能力来减少计算的冗余，提高了特征的表达能力。

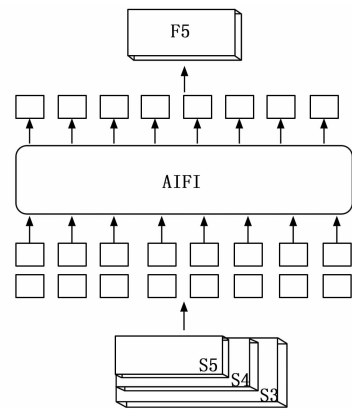


图 3 AIFI 模块融合过程

2.3 Wise-IoU (WIoU)

在 YOLOv8 的检测器中，CIoU 损失函数用于计算位置回归损失。在实际工业生产中，CIoU 函数会对微

小的靶标敏感, 可能导致阳性样品分布不足, 检测精度降低^[19]。在无人机影像中存在许多目标边界不清晰、物体相互遮挡、图像质量较差的场景, CIoU 函数会导致低质量的样本优化不足, 对于困难样本和小目标样本的定位能力低, 从而出现漏检和误检等问题。为了优化损失函数, 本文采用了 Wise-IoU^[20] 损失函数替换 CIoU。

Wise-IoU 引入了基于传统 IoU 的类别加权, 以最大限度地减少类别之间的差异并减少其对测试结果的影响^[21]。这包括为每个类别分配权重, 并在计算 IoU 时使用不同的权重来权衡不同类别之间的重叠, 从而获得更准确的评估结果^[22]。WIoU 有 3 个版本, 其中 WIoUv1 构建基于注意力的边界框损耗, WIoUv2 和 WIoUv3 是在 v1 的基础上对焦点机制添加梯度增益得到的。本文选取 WIoUv1 作为改进的损失函数。

WIoUv1 的计算公式如下所示:

$$L_{IoU} = 1 - \frac{W_i H_i}{wh + w_{gt} h_{gt} - W_i H_i} \overline{IoU}$$

(4)

$$R_{WIoU} = \exp\left[\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)}\right]$$

(5)

$$L_{WIoUv1} = R_{WIoU} L_{IoU}$$

(6)

式中的一些参数如图 4 所示, x 和 y 表示预测框的坐标值。 x_{gt} 和 y_{gt} 为真实值的坐标值, 相应的 H 和 W 值分别为两个框的宽度和高度。

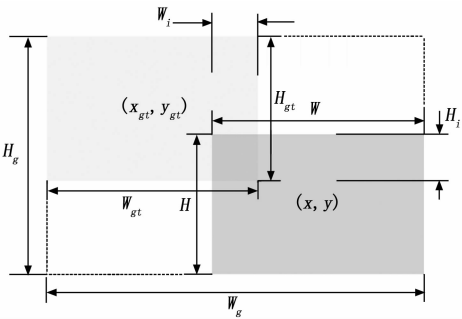


图 4 WIoU-IoU 示意图

3 实验及结果分析

3.1 数据集

本文使用的数据集是 visdrone2019。该数据集是由天津大学等团队开发的一个大型无人机影像数据集, 数据集被分为 3 个部分: 训练集有 6 471 张图像、测试集有 1 580 张图像、验证集有 548 张图像, 包含了多种复杂环境下的视觉数据^[23]。其中超过 60.1% 的注释样本是小目标, 34.2% 是中等的, 5.7% 是大的注释目标, 并且由于数据集中的小目标偏多, 加大了算法对于目标的标注与检测, 给数据集的标注带来了一定的挑战性, 非常适合进行基于深度学习的目标检测任务。

3.2 实验设置及评价指标

visdrone2019 数据集中的目标含有大量的小目标, 为了提高检测过程中的准确性和实时性, 将样本的尺寸统一为 640×640 像素, 这一尺寸不会破坏图像中的有效信息。在硬件配置上的信息: CPU 采用的是 16 v, CPU Intel (R) Xeon (R) Platinum 8481C, GPU 采用的是 RTX 4090D (24 GB 显存), 编译环境是 PyTorch 1.11.0+Python 3.8+CUDA 11.3。设置的 YOLOv8 的超参数如下: 总训练轮次为 200 轮, 初始学习率为 0.01, 动量为 0.937, 批量大小为 16。此外, 为了避免训练中出现过拟合的情况, 实验引入了早停机制: 若验证集损失在连续 20 个 epochs 未出现下降的趋势, 训练过程将会停止, 并且保持当前的最优模型作为输出。其他的重要参数如表 1 所示。

表 1 训练参数

参数	设置
优化器	SGD
权重衰减	5×10^{-4}
马赛克	1.0
α (Wise-Iou)	1.9
δ (Wise-Iou)	3
关闭马赛克	10

为了准确验证改进算法的优越性, 本文的评价指标主要是准确率 (P, precision)、召回率 (R, recall)、平均精密度 (mAP, mean average precision)、参数量 (Number of Parameters)、每秒浮点运算次数 (FLOPS, floating-point operations per second) 和每秒帧数 (FPS, frames per second)。

1) 准确率, 主要表示检测结果是否正确, 即样本中检测为正例中实际也为正例的所占的比例, 计算公式如下:

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

(7)

2) 召回率, 主要表示检测结果是否全面, 即在检测为正例的样本中正确预测为正例的比例, 计算公式如下:

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

(8)

其中: TP 为检测结果中被成功检测为正确目标的数量, FN 为样本中所有检测正确的目标中未被检测到的数量, FP 为检测结果中被错误检测为正确目标的数量。

3) 平均精密度, 是检测所有种类的精度的平均值, 是对 m 个类别的准确度 (P) 和召回率 (R) 综合考虑, 较全面反映网络的性能指标, 如公式 (9) 所示:

$$mAP = \frac{1}{m} \sum_{i=1}^m \int_0^1 P(R) dR$$

(9)

3.3 消融实验

本文提出的无人机航拍检测算法对 YOLOv8s 的主干部分和颈部网络进行了改进，并使用 Wise-IoU 替换 CIoU 作为最终的边界框损失函数。为了系统的分析各个部分对于算法的提升效果，将 YOLOv8s 定义为 A，YOLOv8s+GCSA 为 A+B，YOLOv8s+GCSA+AIFI 为 A+B+C，YOLOv8s+GCSA+AIFI+WIoU 为 A+B+C+D，YOLOv8s+AIFI 为 A+C，YOLOv8s+WIoU 为 A+D，通过探究几组算法的评价指标来验证该算法的优越性，实验结果和 FPS 对比结果如表 2 和表 3 所示。

表 2 消融实验结果 %

类别	指标	A	A+B	A+C	A+D	A+B+C	A+B+C+D
全部	<i>P</i>	49.1	49.4	49.8	50.8	49.8	50.1
	<i>R</i>	36.6	37.8	37.8	37.2	37.4	37.9
	<i>mAP</i>	37.3	38.9	38.1	38.5	38.3	39.4
行人	<i>P</i>	51.4	51.1	51.2	55.4	52.8	51.2
	<i>R</i>	37.2	39.3	38.5	38.5	39.3	39.9
	<i>mAP</i>	40.4	42	40.8	42.9	42.6	42.8
人	<i>P</i>	53.6	55.6	53.7	57.3	54.3	55.8
	<i>R</i>	26.5	28	26.7	25.7	25.6	27.7
	<i>mAP</i>	31.2	32.5	31.4	30.9	30.4	32.7
自行车	<i>P</i>	24.4	29.4	26	27	27	26.9
	<i>R</i>	15.1	13.3	16.1	15.5	17.1	14.8
	<i>mAP</i>	11.4	12.3	11.5	11.9	12.9	12.2
汽车	<i>P</i>	71.7	72.5	72.	73.1	71.5	71.6
	<i>R</i>	73.3	77	76.2	76	76.5	77.3
	<i>mAP</i>	76.4	79.3	79.2	78.6	78.7	79.8
面包车	<i>P</i>	51.6	50.1	51.9	52.6	52.9	54.2
	<i>R</i>	42.9	44.4	44.5	42.6	42.5	44.1
	<i>mAP</i>	42.4	44.3	44.4	44.5	44.1	45.3
卡车	<i>P</i>	51.7	47.8	52.2	51.1	51.9	52.5
	<i>R</i>	34.7	36	36	35.5	34.3	35.6
	<i>mAP</i>	35.2	35.5	35	36	34.9	37
三轮车	<i>P</i>	39.5	41.9	41.6	41	41.5	41
	<i>R</i>	29.4	30.3	28.3	27.8	28.3	29.8
	<i>mAP</i>	26.1	27.7	27	26.5	27	27.2
带遮阳篷的三轮车	<i>P</i>	29.4	27.5	29.	30.3	28.8	3
	<i>R</i>	14.8	16.9	18	15.6	16.4	16.6
	<i>mAP</i>	14.1	15.2	14.7	14.2	14	14.8
公共汽车	<i>P</i>	67.1	63	67.9	69.6	69.1	60.9
	<i>R</i>	49	48.8	50.6	52.6	51.8	50.9
	<i>mAP</i>	53.3	55.5	54.3	58.1	56.2	52.3
摩托车	<i>P</i>	50.4	54.8	52.1	50.5	48.5	51.5
	<i>R</i>	42.6	43.9	42.7	42%	42.1	44.5
	<i>mAP</i>	41.9	45.1	42.7	41.8	41	43.8

通过分析表中的数据可以得到如下的结论：

1) YOLOv8s 的精度在所有改进算法中处于最低，

表 3 FPS 对比结果

指标	A	A+B	A+C	A+D	A+B+C	A+B+C+D
<i>FPS</i> /ms	33.03	29.34	26.19	30.04	22.74	31.13
参数量/ 10^6	11.12	17.68	10.57	11.13	17.13	17.13
<i>FLOPS</i>	28.5	33.7	27.7	28.5	33	33

但 *FPS* 处于最高。不同的改进使算法的 *FPS* 降低，说明改进的模块增加了网络的层数和推理时间。

2) 第二组加入了 GCSA 模块后的算法与 YOLOv8s 相比参数量增加了 600 多万的，但是整体 *mAP* 提升了 1.6%，在汽车、面包车和摩托车上 *mAP* 都得到了很大的提升，增强的主要原因：GCSA 模块能够捕捉特征图中的空间关系，提取复杂背景下的信息，从而提升了检测的性能。第三组整合了 AIFI 模块后与 YOLOv8s 相比 *mAP* 提升 0.8%，参数减少了 60 万，AIFI 模块提高了精度的同时还减少了参数量，主要原因：AIFI 模块相比 SPPF 模块避免了过多的卷积，使识别过程中参数量减少，并且提取了不同序列的目标的信息，提高了检测精度。第四组整合了 WIoU 函数，算法与 YOLOv8s 比较 *P*、*R* 和 *mAP* 都得到了很大的提升，其中行人和汽车是目标密集且处于背景复杂的目标，这两个种类的精度得到了最高的提升，主要原因：WIoU 可以减少各种类别间的差异，最小化了目标与锚框之间的高宽差异，使收敛速度更快，定位效果更好，提高了定位困难样本的效率和精度。第五组将 GCSA 模块和 AIFI 模块整合在一起，虽然相比第二组的结果精度下降了，但比 YOLOv8s 还是提高了 1%，下降的主要原因：GCSA 模块处理后的数据在经过 SPPF 模块的多次卷积后包含了更多有用的参数，导致精度下降了。第六组 YOLOv8-A 算法在 *R* 和 *mAP* 上得到了最大的提升，虽然 GCSA 模块和 AIFI 模块在一起没有产生较大的促进作用，但加入了 WIoU 函数后，函数可以帮助 AIFI 模块提取数据中的复杂空间信息，更好地处理来自 GCSA 模块的数据，使精度和计算速度得到了很大的提升。

3) 总体来看，分别整合 3 个模块精度都得到提升，并且 3 个整合在一起后精度提升最大。参数量虽然提升了许多，但 *FPS* 只降低了 5.7%，*FLOPS* 提升了 15.8%，这相对于精度的提升来说是满足实际部署的要求的，对于 YOLOv8s 的提升是可行且有效的。

为了更加直观地体现各模块间的协同作用，本文使用了热力图来对比模块改进前后特征提取效果，如图 5 所示。图 5 (a) ~ (f) 分别表示 YOLOv8s、YOLOv8s+AIFI、YOLOv8s+GCSA、YOLOv8s+WIoU、YOLOv8s+AIFI+GCSA 和 YOLOv8s+AIFI+GCSA+WIoU。图 5 (a) 中热力集中在中间部分的汽车和左边

的两辆车，而道路上的汽车和行人以及左上角的车辆都没有覆盖到；

图 5（b）中加入了 AIFI 模块后，热力集中在图片中间的车辆，并且右上角的小目标车辆也有了覆盖；图 5（c）中加入了 GCSA 模块后，热力覆盖和图 5（a）几乎一样，但是是图 5（a）的增强；图 5（d）中加入了 WIoU 模块后，热力图开始覆盖在左上角的不容易被识别的汽车上；图 5（e）中加入 AIFI 模块和 GCSA 模块后的图像，相比于图 5（b）和图 5（c），图片中的左侧的车辆加大了覆盖面积，中间部分车辆的覆盖面积也增大了，并且右上角的小目标车辆也被覆盖。图 5（f）为 YOLOv8-A 的图像，该图像与图 5（e）相比左上角的车辆被重点检测，中间部分的热力覆盖面积也增大。综上所述，随着不断加入模块后，模型的检测效果在不断提升，YOLOv8-A 得到了最佳的检测效果。

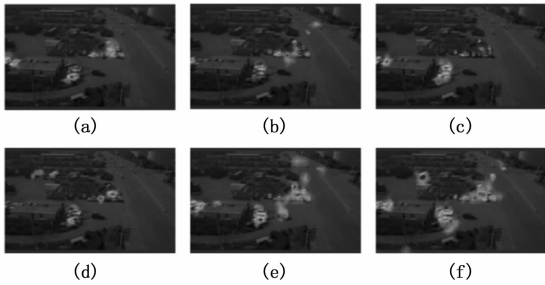


图 5 热力图对比

3.4 不同学习率对比实验

算法的学习率不同，会影响检查的速度和精度。为了得到算法的最优性能，本文选择在不同的学习率下来对比 YOLOv8-A 的检测精度，对比的结果如表 4 所示。由于 YOLOv8-A 使用的是 SGD 优化器，所以选择学习率在 0.001 和 0.01 到 0.1 之间。随着学习率的不断提高，YOLOv8-A 的检测精度都在提高，在 0.1 时 3 个评价指标都达到了最高。实验结果表明，YOLOv8-A 的学习率为 0.1 时可以达到最佳的性能。

表 4 不同学习率算法对比结果 %						
指标	0.001	0.01	0.03	0.05	0.07	0.1
<i>mAP</i>	28.5	39.4	40.3	40.8	40.9	41.2
<i>P</i>	39.3	50.1	51.9	51.2	52.3	52.4
<i>R</i>	29.6	37.9	38.8	39	39.1	39.5

3.5 深度学习算法对比实验

为了验证改进的无人机航拍检测算法的先进性和有效性，本文选择了该领域中更先进和通用的 YOLO 系列算法作为比较的对象，分别是：YOLOv4、YOLOv5s、YOLOXs、YOLOv11s。在表 5 中，将它们定义为 A、

B、C、D，并将 YOLOv8s 和 YOLOv8-A 定义为 E 和 F，结果如表 5 所示。与基线模型 YOLOv8s 相比，YOLOv8-A 在 *mAP* 上有着很大的提升，表明了 YOLOv8-A 的有效性，与 YOLOv4 相比 YOLOv8-A 的 *mAP* 提升了 0.6%，准确率和召回率也提升了 0.5% 和 0.1%。与 YOLOv5s、YOLOXs 和 YOLOv11s 相比，*mAP* 分别提升了 0.9%、6.9% 和 1.2%，虽然 YOLOXs 和 YOLOv11s 的准确率比 YOLOv8-A 高，但召回率和 *mAP* 不如 YOLOv8-A，总体上 YOLOv8-A 更加优异。实验结果表明，与目标检测领域中更先进的算法相比，YOLOv8-A 有着更高的检测精度。

表 5 深度学习算法对比结果						%
指标	A	B	C	D	E	F
mAP	38.8	38.5	32.5	38.2	37.3	39.4
P	49.6	49.4	81.7	50.3	49.1	50.1
R	37.8	37.9	18.2	37	36.6	37.9

3.6 自建数据集对比实验

为了证明所提出算法的普适性，本文基于城市道路上的行人建立了无人机目标检测数据集，该数据集包含 1 000 张不同道路上的行人影像，只包含行人一个检测类别。实验过程中，数据集被分为 800 张训练集、160 测试集和 40 张验证集。与 Visdrone2019 数据集实验相比，也采用了 640×640 的样本尺寸，学习率为 0.01，其他参数和表 1 相同。

从 Visdrone2019 数据集的对比实验来看，YOLOXs 有着最差的检测表现，因此本次实验使用了 YOLOv4、YOLOv5s、YOLOv8s、YOLOv11s 和 YOLOv8-A 进行对比。实验结果如表 6 所示，表中分别按顺序将对比赛法定义为 A、B、C、D 和 E。从结果来看，对比算法的检测精度都差距不大，而 YOLOv8-A 具有最高的检测精度，并且在 *mAP* 和 *P* 上高出 YOLOv8s，达到了 3% 和 5.7%。通过自建数据集的对比实验可以得出 YOLOv8-A 具有普适性。

表 6 自建数据集对比结果 %					
指标	A	B	C	D	E
<i>mAP</i>	59.3	59	59.1	58.9	62.1
<i>P</i>	56.1	55	57.1	57.2	62.8
<i>R</i>	58.8	61.6	59	59.2	58.6

4 结束语

为了提高当前目标检测算法对于无人机影像的检测精度，本文基于 YOLOv8s 提出了一种改进算法 YOLOv8-A，通过在主干网络加入 GCSA 模块，增强捕捉特征图中的全局依赖关系的能力。在颈部结构中将特

征金字塔替换为 AIFI 模块, 进一步提高算法对于捕获输入数据的上下文信息和内在特征的能力。将 CIoU 替换为 WIoU, 降低了算法对小目标定位的漏检概率, 增强了对困难样本的定位稳定性, 提高了算法对小目标检测的精度。通过消融实验和对比实验表明, 所提出的 YOLOv8-A 算法在 Visdrone2019 数据集上实现了更高的精度, 与当前目标检测领域中的通用算法比较, 该算法仍具有优越性, 更是在自建的数据集上实现了 62.1% 的检测精度, 高出基线模型 3%。总体而言, 该算法具有较好的检测精度和通用性。

虽然, 本文所提出的改进算法 YOLOv8-A 精度得到了提升, 但增加了很多参数量和计算量, 导致算法处理的速度变慢了。考虑到当前领域的发展趋势, 未来的研究重点需要兼顾轻量化设计和高精度两个方面, 并且计算量也不能很大, 以此来满足部署到无人机等其他嵌入式设备的需求。

参考文献:

- [1] CAO J, BAO W, SHANG H, et al. GCL-YOLO: ag-hostconv-based lightweight YOLO network for UAV small object detection [J]. Remote Sensing, 2023, 15 (20): 1-16.
- [2] DONG X D, YAN S, DUAN C Q. A lightweight vehicles detection network model based on YOLOv5 [J]. Engineering Applications of Artificial Intelligence, 2022, 113: 1-10.
- [3] LING H, ZHAO T, ZHANG Y, et al. Engineering vehicle detection based on improved YOLOv6 [J]. Applied Sciences, 2024, 14 (17): 8054-8065.
- [4] IGNAZIO G, UR A R, HEIDARIAN R D, et al. Deep object detection of crop weeds: performance of YOLOv7 on a real case dataset from UAV images [J]. Remote Sensing, 2023, 15 (2): 539-550.
- [5] 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): 1-18.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [R]. Cornell University Library: arXiv, 2015: 1-15.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (2): 318-327.
- [8] YASIR M, LIU S, PIRASTEH S, et al. YOLOShipTracker: tracking ships in SAR images using lightweight YOLOv8 [J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 134: 104137-104148.
- [9] XUAN C, ZHANG Y W, LI S, et al. Swin-transformer-based YOLOv5 for small-object detection in remote sensing images [J]. Sensors, 2023, 23 (7): 1-16.
- [10] ZHANG Z. Drone-YOLO: an efficient neural network method for target detection in drone images [J]. Drones, 2023, 7 (8): 526-538.
- [11] LI Y, FAN Q, HUANG H, et al. A modified YOLOv8 detection network for UAV aerial image recognition [J]. Drones, 2023, 7 (5): 1-14.
- [12] XUAN C, WANG Y Q, LI S, et al. FE-YOLO: a feature enhancement network for remote sensing target detection [J]. Remote Sensing, 2021, 13 (7): 1311-1322.
- [13] LIU H, TAN F, JIN Y. Dual-YOLO: dual-path UAV aerial image target detection algorithm [J]. The Journal of Supercomputing, 2025, 81 (7): 809-820.
- [14] YAN Z, ZHANG H F, HUANG Q Q, et al. DsP-YOLO: an anchor-free network with DsPAN for small object detection of multiscale defects [J]. Expert Systems with Applications, 2024, 241: 1-12.
- [15] 陈鑫, 于云飞, 丁相玉, 等. 基于 GCSA-YOLOv8 的航空材料表面缺陷检测算法 [J]. 失效分析与预防, 2025, 20 (1): 39-47.
- [16] 于松正, 谭钦红. 基于多注意力机制的人体姿态检测模型 [J]. 激光杂志, 2025, 46 (5): 78-85.
- [17] ZOU L, LIU A. Intelligent detection algorithm for concrete bridge defects based on SATH-YOLO model [J]. Sensors, 2025, 25 (5): 1449-1460.
- [18] TANG S, BAO Q, JI Q, et al. Improvement of RT-DETR model for ground glass pulmonary nodule detection [J]. PloS One, 2025, 20 (3): 1-12.
- [19] XU J, YE D, ZHANG S, et al. Metallic surface defect detection via NWD-WIoU based on grayscale co-generation entropy gain [J]. Applied Intelligence, 2025, 55 (10): 752-765.
- [20] JIANG T, CHEN S. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [R]. Cornell University Library: arXiv, 2023: 1-10.
- [21] JIANG T, CHEN S. Alightweight forest pest image recognition model based on improved YOLOv8 [J]. Applied Sciences, 2024, 14 (5): 1-15.
- [22] GUO F, GUO X, GUO L, et al. Target detection of diamond nanostructures based on improved YOLOv8 modeling [J]. Nanomaterials, 2024, 14 (13): 1115-1126.
- [23] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results [C] // Seoul: IEEE, 2019: 1-6.