Transactions on Computer Science and Technology
Volume 10 Issue 2, 2023, PP. 1-6

Research on Anti-UAV Visual Detection Method Based on Deep Learning

Yinggang Liang [1], Shaobo Wu [1†]

1. Beijing Information Science and Technology University, Beijing, 100010, China

Email: wushaobo@bistu.edu.cn

Abstract

The illegal intrusion of drones poses a significant threat to daily life and societal security. Existing methods that rely on microwave radar and machine learning are not effective in detecting low, slow, and small drone targets. For this reason, this paper proposes a deep learning-based anti-drone visual detection method. The DETR model is employed to identify high-altitude multi-scene drone targets. Specifically, a deformable attention module is introduced into the DETR model to enhance the detection accuracy of small target drones. To reduce the model's parameter count while maintaining detection accuracy and fulfilling the real-time requirements for unmanned aerial vehicles, the ShuffleNet V2 model is utilized to optimize the DETR backbone network. Furthermore, a mixed attention mechanism is incorporated. Experimental results demonstrate that the improved model achieves an average accuracy increase of 2.4% compared to the original DETR model (mAP@0.5), with a 5.9% enhancement in small target detection accuracy. The lightweight improvements made to the model backbone network reduce the parameter count from 66.5Mb to 43.8Mb, resulting in improved detection speed and meeting the real-time and deployment requirements of the drone detection model.

Keywords: Anti-UAV; Deep Learning; Target Detection; DETR; Lightweight

基于深度学习的反无人机视觉检测方法研究

梁迎港 [1]，吴韶波 [1*]

1. 北京信息科技大学，北京 100010

摘 要：无人机的非法入侵对日常生活乃至社会治安造成了严重威胁。本研究基于微波雷达和机器学习的方法针对低、慢、小目标无法达到很好的检测效果，提出一种基于深度学习的反无人机视觉检测方法。利用检测器 DETR（DEtection TRansformer）对高空多场景无人机目标进行检测：在 DETR 模型中引入可变形注意力模块，提高对小目标无人机检测精度；利用 ShuffleNet V2 模型对 DETR 主干网络进行轻量化改进，并加入混合注意力机制，在不明显降低检测精度的同时减小模型参数量，满足无人机实时性检测需求。实验证明，改进后的模型与原 DETR 模型相比，平均精度（mAP@0.5）提高了 2.4%，小目标检测精度提升 5.9%，对模型主干网络进行轻量化改进后，模型参数量从 66.5Mb 降到 43.8Mb，并且提升了检测速度，满足了无人机检测模型的实时性与部署需求。

关键词：反无人机；深度学习；目标检测；DETR；轻量化

引言

无人机在勘探、摄影、监测和军事[1]等领域的广泛应用得益于其低成本和高机动性等优势，但这也使得无人机数量在近些年来急剧增多，由此产生了一系列问题，如无人机的非法飞行和滥用[2]。这些问题的存在对人们的日常生活乃至社会治安造成了严重威胁。因此，亟需使用更加先进的技术手段应对无人机的各种威胁，对无人机进行有效地监控与反制。反无人机技术是指反制无人机的反无人机系统和防御系统相关技术，主要分为探测系统和阻截系统两大类。其中，无人机精准探测是进行无人机阻截的必要前提，也是反无人系

统实现的重难点。基于光电传感器的反无人机视觉检测具有低成本、无辐射、适用多种场景的优点，同时也是对雷达、音频、射频等探测技术的重要补充探测手段。传统机器学习视觉检测方法主要依赖于人工选择或构建特征并选取模型进行训练，代表性方法包括：基于哈尔特征（Haar-like features）的级联分类器维奥拉-琼斯检测（Viola–Jones Detection，VJ Det）[3]、定向梯度检测直方图（Histogram of Oriented Gradients Detection，HOG Det）[4]、采用部件组合生成模型来对目标进行建模的可形变部件模型（Deformable Part Model, DPM）[5]。传统机器学习方法在目标检测领域往往需要进行更多的人工特征工程和模型调优，难以应对复杂的场景变化。相比之下，深度学习方法能够自动提取特征和训练模型，具有更高的泛化能力和更好的鲁棒性。

目前，部分学者对基于深度学习的反无人机视觉检测方法进行了研究，例如：2019 年，Mrunalini 等人，利用基于 ResNet-101 的 SSD 模型在监管视频中检测小型无人机[6]。2020 年，Garcia 等人使用了基于 ResNet-101 的 Faster R-CNN，在 SafeShore 项目的数据集上进行反无人机检测[7]。2021 年，Xun 等人将带有预先训练权值的目标检测器 YOLOv3 进行迁移学习，以训练 YOLOv3 检测无人机[8]。2023 年，薛珊等人提出了一种融入注意力机制和尺度自适应特征融合的 YOLOv5 反无人机检测算法[9]。以上基于卷积神经网络的无人机检测模型都需要进行一系列复杂的前处理、后处理操作来剔除冗余的预测框，如生成候选框、非极大值抑制（Non-Maximum Suppression，NMS）等操作。为简化检测流程，本文将 Transformer 在目标检测的开山之作 DETR 运用于无人机端到端检测，并进行相应的改进：引入可变形注意力模块替代 DETR 中的自注意力，提高网络对小目标无人机检测效果，加快网络收敛速度；利用 ShuffleNet V2 网络体积较小的优点，对 DETR 的主干网络进行轻量化改进，并在轻量化网络中加入混合注意力机制，提高检测精度的同时，满足无人机检测模型的实时性与部署需求。

1 DETR 模型

检测器 DETR（DEtection TRansformer）是 Facebook 团队于 2020 年在 ECCV 上发表的一种端到端的目标检测算法，其主体基于 Transformer 模型实现[10]。传统基于锚框（anchor-based）的目标检测算法通常是首先确定目标可能所在的锚框位置，然后对锚框内的目标进行分类和位置回归，而 DETR 则将目标检测任务视为一个集合预测问题。DETR 的集合预测思路是先设定一个远大于图像中实际目标数量的固定值 N，然后使用 Transformer 解码器一次性预测出 N 个对象，最后利用匈牙利算法将预测对象与真实对象匹配。这种思路可以避免锚框不匹配、数量不足等带来的问题，并且更加高效和精确。

相比于传统的目标检测器，DETR 使用 Transformer 模型可以计算每个像素和其他所有像素之间的相关性，从而提高感受范围和检测的效果。DETR 的网络架构主要由三个部分组成：主干卷积神经网络（Convolutional Neural Network，CNN）网络、Transformer 编码器-解码器和前馈神经网络（Feed Forward Network，FFN）。

2 基于改进 DETR 的反无人机目标检测模型

基于改进 DETR 的反无人机目标检测网络架构如图 1 所示。

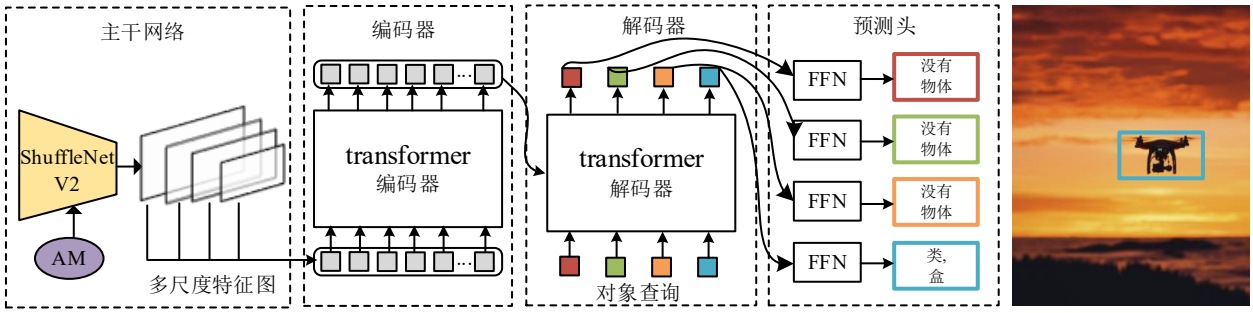


图 1 基于改进 DETR 的反无人机目标检测网络架构

基于改进 DETR 的反无人机目标检测网络架构与 DETR 检测模型类似，网络架构仍然主要由三个部分组成：轻量化主干网络、Transformer 可变形编码器-解码器和前馈神经网络。改进的模型工作流程如下：首先，

通过轻量化主干网络进行特征提取得到多尺度特征图，将特征图展开，得到特征序列作为编码器的输入；然后，编码器中进行多尺度可变形注意力映射后输出 N 个向量，与对象查询一同输入解码器；最后，将解码器输出通过 FFN 进行预测。改进的 DETR 模型主要包括以下变化：（1）用多尺度可变形注意力模块替换原本注意力模块，自然地扩展到聚合多尺度特征。（2）使用添加混合注意力机制的轻量化网络改进模型主干网络，使模型参数量减少，实现无人机实时检测。下面分别介绍这两部分改进。

2.1 可变形注意力模块

引入可变形注意力模块（Deformable Attention Module）将解码器中交叉注意力模块替换为多尺度可变形注意力模块，自注意力模块保持不变，缓解 DETR 收敛缓慢和高复杂性的问题[11]。可变形注意力模块受 Deformable Convolution 启发，只关注参考点附近少量关键采样点，通过为每个查询（Query）分配较少数量的计算权重（键，Key），从而缓解了收敛速度和特征空间分辨率的问题。该模块可以自然地扩展到聚合多尺度特征，而无需特征金字塔网络 (Feature Pyramid Networks, FPN)的帮助[12]。使用多尺度可变形注意模块来代替 Transformer 注意力模块处理特征映射，如图 2 所示。

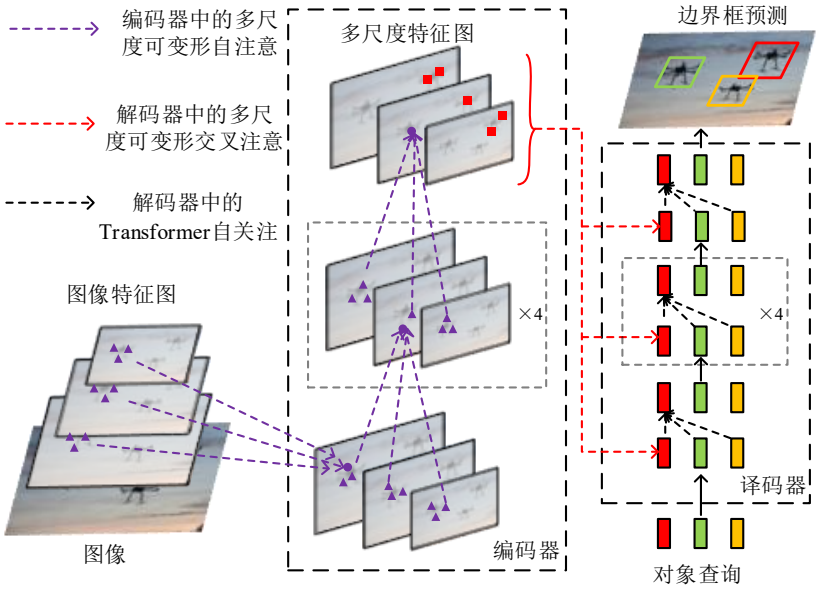


图 2 可变形的 DETR 检测器

可变形注意力特征的计算公式如式(1)所示。其中，输入特征图$x \in \mathbb{R}^{C\times H\times W}$，$q$作为一个 Query 元素的索引，查询内容特征为$z_p$、二维参考点坐标为$p_q$，C 为特征图通道数，H 为特征图的高，W 为特征图的宽。

$$\mathrm{DeformAttn}\left(z_q,\mathrm{p}_q,x\right)=\sum_{\mathrm{m}=1}^{M}W_m\left[\sum_{k=1}^{K}A_{mqk}\cdot W_m^{'}x\left(p_q+\Delta p_{mqk}\right)\right] \tag{1}$$

其中，M为注意力头的数量，m为对应注意力头的索引，K为所有采样点k的数量（$K \ll HW$），$W_m \in \mathbb{R}^{C\times C_v}$和$W_m' \in \mathbb{R}^{C_v\times C}$为全连接层 FC 可学习权重（$C_v = C/M$，为各注意力头处的特征维数）。$A_{mqk}$为第$m$个注意力头中第$k$个采样点的注意力权重，$\Delta p_{mqk}$表示第$m$个注意力头中第$k$个采样点的预测偏移量。注意力权重$A_{mqk}$通过 softmax 函数实现归一化，取值位于[0,1]之间，Δp_{mqk}为取值范围没限制的二维实数。$p_q + \Delta p_{mqk}$为小数，所以采用双线性插值方法计算，即$x\left(p_q + \Delta p_{mqk}\right)$。对 Query 的特征$z_p$进行线性映射得到注意力权重$A_{mqk}$和预测偏移量$\Delta p_{mqk}$。

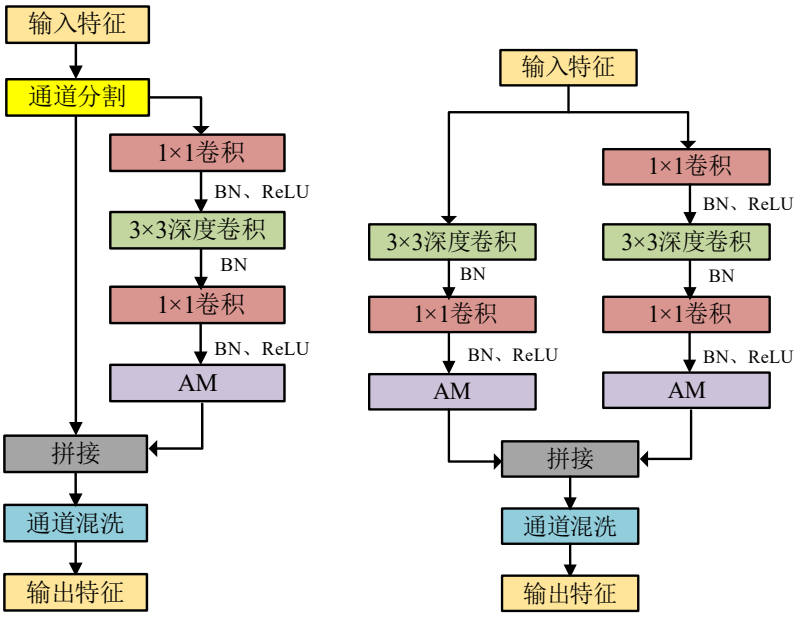
可变形注意力可以代替 FPN 实现多尺度特征图输入。在多尺度特征图$\{x^l\}_{l=1}^{L}$中，令$x^l \epsilon \mathbb{R}^{C\times H_l\times W_l}$，$\hat{p}_q$为 query 元素$q$的参考点归一化坐标，则多尺度可变形注意力模块如式(2)所示。

$$\mathrm{MSDeformAttn}\left(z_q,\hat{p}_q,\{x^l\}_{l=1}^{L}\right)=\sum_{\mathrm{m}=1}^{M}W_m\left[\sum_{l=1}^{L}\sum_{k=1}^{K}A_{mlqk}\cdot W_m^{'}x^l\left(\phi_l\left(\hat{p}_q\right)+\Delta p_{mlqk}\right)\right] \tag{2}$$

多尺度可变形注意力模块从多尺度特征图中采样LK个点。其中，l为特征图层级的索引，$\hat{p}_q\in[0,1]^2$是归一化坐标，$(0,0)$为输入图像左上角坐标，$(1,1)$为右下角坐标点。使用函数$\phi_l(\hat{p}_q)$对归一化坐标$\hat{p}_q$缩放至符合第l个特征图层级的大小。

2.2 基于注意力机的轻量化主干网络改进

将原 DETR 模型常用主干网络 ResNet50 改进为 ShuffleNet V2 轻量化网络[13]，并在网络有效层加入注意力机制，以少量的计算代价来赋予主干网络对卷积后特征图各通道或空间不同关注度的能力，筛选更有效的目标特征，从提高算法的精确度。加入注意力机制的 ShuffleNet V2 的模块结构如图 3 所示，称为 AM_ ShuffleNet V2，其中用 AM 代表混合注意力机制。混合注意力机制采用代表性的卷积块注意力模块（Convolutional Block Attention Module，CBAM）[14]，CBAM 是由通道注意力模块（Channel Attention Module，CAM）和空间注意力模块（Spatial Attention Module，SAM）组成的结构。CAM 用于在通道域上为张量分配注意力权重，而 SAM 用于在空间域上为张量分配注意力权重。



（a）基本模块　　（b）卷积步长为 2 的空间下采样模块

图 3 AM_ ShuffleNet V2 模块结构

在 AM_ShuffleNet V2 模块中图 3（a）基本模块将图片特征进行通道分割，与 ResNet 网络不同之处在于通道分割将输入特征映射直接分割为两部分。右侧路径首先进行 1×1 卷积和批标准化，然后通过一个 3×3 深度卷积和批标准化来提取特征，接着进行 1×1 卷积和批标准化对通道特征信息进行整合。最后添加本文介绍的注意力机制模块，为神经网络特征图中代表无人机的特征通道和特征点分配较高的权重，同时给干扰信息较多的特征通道和特征点分配较低的权重。左侧路径不进行卷积处理，直接与右侧路径的特征进行拼接，然后进行通道混洗操作，这种操作减少了运算量并且增加通道间信息交流，通道数不变。在每个基本模块中，一半的特征通道不进行处理直接进入下一基本模块中，可以看作是一种特征重用。图 3（b）为步长为 2 的空间下采样模块，此模块不需要进行通道分割，图片特征分别经过左右两侧路径操作。左侧路径首先进行步长为 2 的 3×3 深度卷积，然后进行 1×1 卷积，最后经过注意力模块进行通道或空间特征权重分配；右侧路径分别经过 1×1 卷积、步长为 2 的 3×3 深度卷积、1×1 卷积和注意力模块。最后两侧路径特征进行拼接，使通道数加倍，降低计算复杂度与参数量。

3 实验

实验运行在 Ubuntu 20.04 操作系统上，深度学习框架为 PyTorch 1.12，硬件环境为 4 个 Intel(R) Xeon(R) Platinum 8338C CPU，4 个显存为 24G 的 RTX 3090 显卡。DUT Anti-UAV 反无人机数据集[15]为大连理工大学制作的反无人机数据集，它总共包含 10,000 张图像的检测数据集。包括一个训练集（5200 张图片）、一个验证集（2600 张图片）和一个测试集（2200 张图片）。考虑到一张图像包含多个对象的情况，检测对象总数为 10,109 个，其中训练集、测试集和验证集分别有 5243 个、2245 个和 2621 个对象。使用 Facebook 团队在 DETR 模型在 COCO 数据集上训练的模型作为预训练模型。迭代次数 epoch 为 100 个周期，batch-size 设为 8，初始学习率为 0.015，使用 SGD 模型优化器，动量（momentum）参数为 0.9，权重衰减系数（decay）为 10-4。模型中总共设置 6 个编码器层和 6 个解码器层，每层注意力头数$\mathrm{M}=8$、采样点数K为 4。分类损失函数为 Focal loss，回归损失函数为 L1 Loss 和 GIoU Loss。

为评估改进的 DETR 算法在无人机目标检测场景下的性能，本文使用的评估指标包括：平均精度均值（mean Average Precision，mAP）、每秒帧数 (Frames Per Second，FPS)、参数总量（Params）。设置真实框与预测框的交并比值（intersection over union，IoU）为 0.50。较小的目标框（$\mathrm{AP_S}$）像素边小于 32，较大的目标框（$\mathrm{AP_M}$）像素边大于 96，中等的目标框（$\mathrm{AP_L}$）在两者之间。无人机检测模型性能对比实验结果如表 1 所示，其中 DE_Res50 表示使用 ResNet-50 作为主干网络的 DETR 模型，CA_ DE_Res50 表示引入可变形注意力模块的 DETR 模型，CA_ DE_SN V2 表示使用 ShuffleNet V2 作为主干网络的 DETR 模型，CA_ DE_SN V2_AM 表示使用加入混合注意力机制的 AM_ShuffleNet V2 作为主干网络的 DETR 模型。

表 1 无人机检测模型性能对比实验结果

方法	mAP@0.5	$\mathrm{AP_S}$	$\mathrm{AP_M}$	$\mathrm{AP_L}$	Params/Mb	FPS
DE_Res50	69.3	30.5	55.8	67.3	66.5	24
CA_ DE_Res50	72.2	37.2	58.1	64.1	64.9	21
CA_ DE_SN V2	66.1	31.5	51.4	61.2	39.6	37
CA_ DE_SN V2_AM	71.7	36.4	54..2	66.5	43.8	34

实验结果表明，引入可变形注意力机制的模型平均精度（mAP@0.5）提升了 2.9%，小目标检测精度提升 6.7%，可见可变形注意力 DETR 模型小目标检测精度提升效果明显。使用 ShuffleNet V2 作为 DETR 主干网络后，无人机目标检测精度降低，但极大的减小了模型参数量，并且提升了模型推理速度，更适合反无人机目标检测实时性检测和部署的应用场景。使用加入混合注意力机制的轻量化网络作为 DETR 主干网络后，使轻量化后的网络检测精度得到改善，最终提出的模型平均精度（mAP@0.5）提升了 2.4%，小目标检测精度提升 5.9%，模型参数量从 66.5Mb 降到 43.8Mb，对于反无人机的检测效果更好，满足了无人机检测模型的实时性与部署需求。

4 结论

随着无人机普及率的增加，“黑飞”“扰航”等安全隐患日益凸显，视觉检测方法对于无人机反制具有重要意义，目前深度学习方法在这一领域表现出色。本文将自然语言处理（NLP，Natural Language Processing）领域的注意力模型 Transformer 应用于反无人机检测，对基于 DETR 的目标检测算法进行了改进并对其进行轻量化研究：引入可变形注意力模块，提高对小目标无人机检测精度；利用基于混合注意力机制的 ShuffleNet V2 模型对 DETR 主干网络进行轻量化改进。最终使模型满足反无人机高精度和实时性检测的应用场景需求。在未来工作中，可以研究多种传感器融合算法，通过与其他反无人机探测技术结合，实现全天候反无人机检测。

参考文献

[1] Fan Z, Gao X, Jin Y, et al. Research on Route Planning of Group UAV Cooperation for Deception Jamming to Radar Network[C]// IEEE Information Technology, Networking, Electronic and Automation Control Conference. IEEE, 2020.

[2] Jiang C, Fang Y, Zhao P, et al. Intelligent UAV Identity Authentication and Safety Supervision based on Behavior Modeling and Prediction[J]. IEEE Transactions on Industrial Informatics, 2020, PP (99):1-1.

[3] Viola P, Jones M J. Robust real-time face detection[J]. International journal of computer vision, 2004, 57: 137-154.

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893

[5] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 32(9): 1627-1645.

[6] Nalamati M, Kapoor A, Saqib M, et al. Drone detection in long-range surveillance videos[C]//2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2019: 1-6.

[7] Garcia A J, Lee J M, Kim D S. Anti-drone system: A visual-based drone detection using neural networks[C]//2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2020: 559-561.

[8] Xun D T W, Lim Y L, Srigrarom S. Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2[C]//2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP). IEEE, 2021: 1-6.

[9] 薛珊,张亚亮,吕琼莹等.复杂背景下的反无人机系统目标检测算法[J].吉林大学学报(工学版), 2023,53(03):891-901.DOI:10.13229/j.cnki.jdxbgxb.20221288.

[10] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

[11] Zhu X. Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection[J]. arXiv preprintarXiv:2010.04159.2020

[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[13] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[14] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[15] Zhao J, Zhang J, Li D, et al. Vision-based anti-uav detection and tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 25323-25334.