

复杂背景下基于 YOLOv7-tiny 的图像目标检测算法

薛 珊^{1,2*}, 安宏宇¹, 吕琼莹¹, 曹国华²

(1. 长春理工大学 机电工程学院, 吉林 长春 130022;
2. 长春理工大学 重庆研究院, 重庆 400000)

摘要:“黑飞”无人机一旦带有炸弹等物品,会对人们带来威胁。对在公园、游乐场、学校等复杂背景下“黑飞”的无人机进行目标检测是十分必要的。前沿算法 YOLOv7-tiny 属于轻量级网络,具有更小的网络结构和参数,更适合检测小目标,但在识别小目标无人机时出现特征提取能力弱、回归损失大、检测精度低的问题;针对此问题,提出了一种基于 YOLOv7-tiny 改进的无人机图像目标检测算法 YOLOv7-drone。首先,建立无人机图像数据集;其次,设计一种新的注意力机制模块 SMSE 嵌入到特征提取网络中,增强对复杂背景下无人机目标的关注度;然后,在主干网络中融入 RFB 结构,扩大特征层的感受野,丰富特征信息以增强特征提取的鲁棒性;然后,改进网络中的特征融合机制,通过新增小目标检测层,增加对小尺度目标的检测精度;然后,改变损失函数提高模型的收敛速度,减少损失以增强模型的鲁棒性;最后,引入可变形卷积(Deformable convolution, DCN),更好的根据目标本身形状进行特征提取,提升了检测精度。在 PASCAL VOC 公共数据集上进行对比实验,结果表明改进后的算法 YOLOv7-drone 相比于 YOLOv7-tiny,平均精度(mAP@0.5)提升了 6%;在自制无人机数据集上进行实验,结果表明 YOLOv7-drone 与原算法相比,平均精度(mAP@0.5)提高了 6.1%,并且检测速度为 72 帧/s;与 YOLOv5l、YOLOv7 目标检测算法进行对比实验,结果表明改进后的算法在平均精度(mAP@0.5)上分别高于对比算法 4%、3.1%,验证了文中算法的可行性。

关键词: 目标检测; 复杂背景; 注意力机制; 小目标检测

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20230472

0 引言

随着科技发展,民用无人机给人们生活带来便利,但是无人机“黑飞”事件的发生^[1-5],给人们也带来了极大的危害。尤其是在公园、游乐场、学校等公共区域,如果“黑飞”无人机携带炸弹等危险品,将对公共安全区域带来极大威胁。在公园、大型游乐场等公共区域进行无人机的检测和跟踪是十分必要的。目标检测是跟踪的前提。无人机属于小目标;公园、游乐场等属于复杂背景;在复杂背景下进行无人机小目标的检测,并且希望能够更准确和更快速,成为亟待解决的问题。传统的无人机目标检测方法存在窗口冗余,特征鲁棒性较差等问题。自深度学习出现之

后^[6],目标检测取得了巨大的突破,包括以 Faster R-CNN^[7]为代表的两阶段算法和以 YOLO^[8]、SSD^[9]系列为代表的单阶段目标检测算法。在实际目标检测中,运用传统方法对在复杂场景下的无人机检测会出现无法检测、漏检的问题。采用基于深度学习的目标检测算法进行无人机的检测成为研究热点。例如,吴曼佳根据近距离空域背景复杂以及无人机目标小的特点,提出了一种基于改进的 YOLOv3 的小目标检测网络,在复杂背景下的小目标检测性能上有明显提升;刘朋飞等人鉴于低空复杂背景下无人机尺度多变、背景复杂的特点,提出一种基于 SSD 的算法。虽然上述基于深度学习的检测算法有了很大提升,但是

收稿日期:2023-07-30; 修订日期:2023-10-15

基金项目:吉林省科技厅重点科技研发项目(20210203055SF);吉林省教育厅科学技术研究项目

作者简介:安宏宇,男,硕士生,主要从事现代检测理论与技术方面的研究。

导师(通讯作者)简介:薛珊,女,教授,博士生导师,博士,主要从事现代检测理论与技术方面的研究。

在复杂背景下小目标检测仍然有漏检、无法检测、检测精度不高的问题。

针对无人机在复杂背景下目标较小存在难检测,易漏检的问题和无人机样本较小的问题,文中基于 YOLOv7-tiny^[10] 的目标检测算法,将注意力机制、特征融合、扩大感受野的思想融入到 YOLOv7-tiny 网络中,提出一种改进的无人机目标检测算法 YOLOv7-drone 算法。此算法可以解决在复杂背景下检测小目标,无人机小样本的问题,并且具有更高的准确率和速度。

1 无人机数据集的建立

文中以复杂背景和以小目标为背景建立数据集,图像背景多以复杂的公共安全区域为主,例如,学校、游乐场、公园,且无人机目标均是远距离拍摄,进而建立了包含大量复杂背景小尺度图片的无人机数据集。

采用自制光学设备采集大量复杂背景、小尺度的无人机图片,无人机型号为大疆御 Mavic2 和大疆御 Mavic Air 两款机型,光学设备能清晰看到 1000 m 高空,而民用无人机限飞高度为 120 m。主要采集低空小无人机的图片。拍摄时,控制无人机飞行高度和距离,分别从各个角度拍摄不同背景下的小尺度无人机图片,构建无人机数据集,部分无人机图片如图 1 所示。



图 1 无人机数据集部分图片

Fig.1 Partial picture of drone dataset

为了获取更多无人机图片组成数据集,用于训练和评估。文中引入大连理工大学提出的无人机数据集 DUT-ANTI-UAV^[11],数据集中包含不同背景的无人机图片,且多为小尺度的无人机图片,部分图片如图 2 所示。与自制的无人机数据集合并,构建文中的无人机数据集。



图 2 DUT-ANTI-UAV 数据集

Fig.2 DUT-ANTI-UAV dataset

2 基于 YOLOv7-tiny 的目标检测算法设计

目前, YOLOv7 作为先进算法较原来的目标检测算法有很大改进,既减少了网络参数,又提升了检测速度。YOLOv7 是一种无锚框的单阶段目标检测算法,有较高的检测精度,同时也有良好的推理速度。YOLOv7 设置了三个版本,面向三个不同的环境为边缘 GPU、普通 GPU 和云 GPU,分别为 YOLOv7-tiny、YOLOv7 和 YOLOv7-W6。文中既要保证较高的准确度,又要有较高的实时性,因此选择 YOLOv7-tiny 为研究对象。

YOLOv7-tiny 网络主要包含了 Input(输入)、Backbone(骨干网络)、Neck(颈部)、Head(头部)这四个部分。首先,在输入端图片经过 mosaic^[12] 数据增强,采用随机裁剪、缩放将四张图片拼接成一张图片,对图片进行预处理,同时又在 mosaic 的基础上融入 mixup^[13] 数据增强后,送入到主干网络当中,图片进入主干网络进行特征提取,随后将特征通过颈部的特征融合处理后得到大、中、小三种尺寸特征。最后,经过检测头中的卷积处理后进行输出。如图 3 所示,为 YOLOv7-tiny 算法结构图。

在主干网络部分,主要由卷积、ELAN^[14] 模块、SPP 模块构建而成。其中,网络使用高效聚合网络 ELAN,一改之前的反向传播,而是通过梯度传播路径来设计网络结构,提升梯度最短路径。当输入图片信息通过很多层之后,可能会消失或过度膨胀,所以引入 ELAN 模块可以加深网络,提高网络的精度。空间金字塔池化 (Spatial Pyramid Pooling, SPP)^[15] 通过不同大小的卷积核池化操作,可以增大感受野,融合不同尺度特征图信息,完成特征融合。

在颈部部分,依旧采用路径聚合网络 (Path Aggregation Network, PANet)^[16] 作为网络的特征融合部分。通过在特征金字塔 (Feature Pyramid Network,

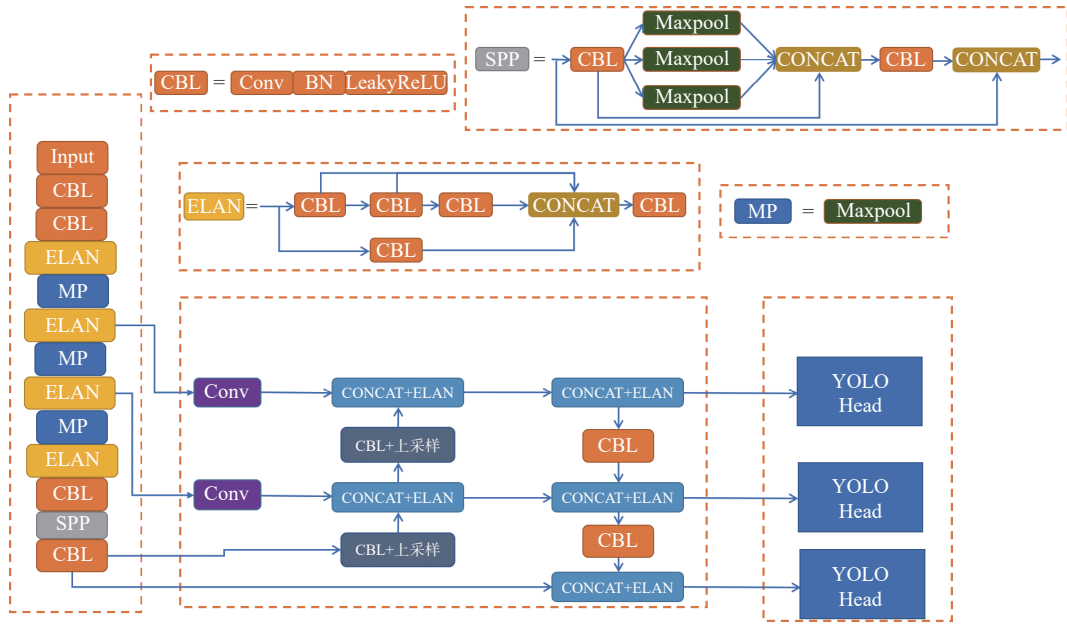


图 3 YOLOv7-tiny 算法结构图

Fig.3 Network chart of YOLOv7-tiny

FPN)^[17]的基础上增加自底向上的路径增强,利用准确的低层定位信号增强整个特征层次,从而缩短了低层与顶层特征之间的信息路径。

在预测端部分, YOLOv7-tiny 用三个卷积分别输出三个尺度的特征图,用于检测大、中、小物体。

3 基于改进YOLOv7-tiny的目标检测算法设计

由于 YOLOv7-tiny 网络在针对复杂背景小尺度

无人机检测时会出现检测效果差、平均精度低、以及出现误检和漏检的问题。所以,文中基于 YOLOv7-tiny 算法设计了一种引入多尺度通道注意力机制和小目标检测层的 YOLOv7-drone 无人机目标检测算法,网络结构如图 4 所示。文中设计算法分别从 5 个方向改进: 1) 设计多尺度通道注意力机制来增强对小尺度无人机的定位能力; 2) 加入 RFB 结构来增强浅层小目标无人机的特征的提取能力; 3) 在特征融合部

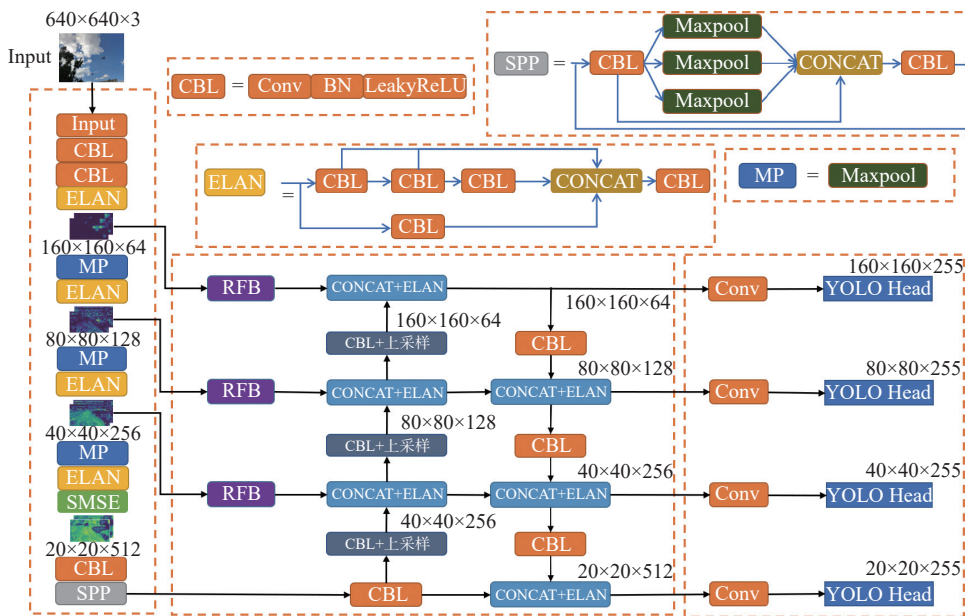


图 4 YOLOv7-drone 算法结构图

Fig.4 Network chart of YOLOv7-drone

分继续进行上采样增加一个新的小目标检测层,以便与检测更小的无人机目标;4)采用 *SIoU* 损失函数作为算法的定位损失函数,加强算法对无人机框的优化能力;5)引入可变形卷积来更加贴近小目标形状。

3.1 多尺度通道注意力机制的改进

SENet (Squeeze-and-Excitation Networks)^[18] 是由 Momenta 和牛津大学的 Jie Hu 等人提出的一种新的网络结构,目的是通过建模卷积特征通道之间的相互依赖关系来提高网络的表达能力。SENet 重点在于关注通道之间的关系,希望模型可以自动学习到不同通道特征的重要程度。由于每一个通道的权重都是不同的,如果能够把这些信息捕获出来,那么就可以获得更多的通道信息,进而提高模型的准确率。

SENet 结构主要由压缩 (Squeeze) 和激励 (Excitation) 两部分组成。其中压缩操作通过全局平均池化,将每个通道的二维特征 (HW) 压缩为一个实数,将特征图从 (H,W,C) 变为 (1,1,C),得到通道级的全局特征,激励操作通过给每个特征通道生成一个权重值,用两个全连接层构建通道间的相关性,输出的权重值数目和输入特征图的通道数相同;最后将前面得到的归一化权重加权到每个通道的特征上。

文中针对的是小目标检测问题,所以在 SENet 的基础上进行设计,构建多尺度通道注意力模块 (SiLU-Multi SENet, SMSENet),增强了复杂背景下的关注度和小目标的特征表达能力,多尺度通道注意力机制模块结构图如图 5 所示。首先,采用 3×3 和 5×5 两个不同大小的卷积核进行多尺度的特征提取,获得多种尺度的特征图,方便后续的特征提取。其次,对 SENet 的压缩部分进行改进,在原有的全局平均池化层的基础上,增加一个最大池化的分支。由于之前只关注了全局特征信息,而忽视了局部特征信息,在检测小目标的时候,会造成小目标的特征信息丢失,所以引入两种池化去提取特征,可以更好的提取小目标的局部特征信息,增强了复杂背景下小目标的关注度。最后,经过两种池化后,分别经过全连接层、激活函数和全连接层,再经过 *Sigmoid* 激活函数后输出后与输入特征进行加权后得到改进后的通道注意力输出。其中,在经过第一次的全连接层后的激活函数上,将 SENet 原有的 ReLU 激活函数替换成 SiLU 激活函数。SiLU 激活函数是一个非单调、有下界、无

上界、正则化的平滑激活函数,它允许一部分负梯度流入保证信息流动,以防使用 ReLU 函数导致部分神经元死亡,丢失过多有效特征,同时,有助于提高非线性关系。可以看出,增加了多尺度卷积核和最大池化的操作较 SENet 相比,提升了不同尺度的小目标无人机检测,通过引入最大池化增加局部特征提取的能力以便于观察到更小的目标。

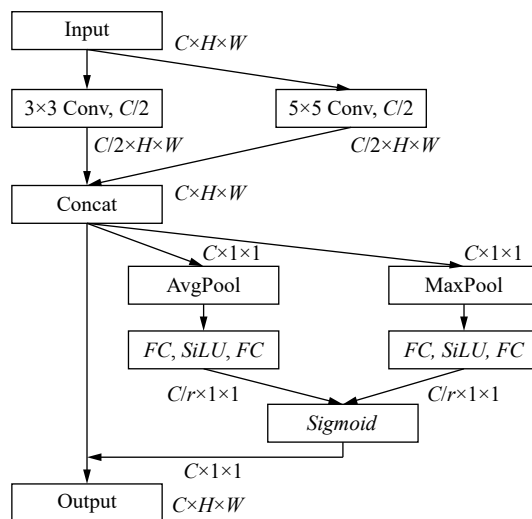


图 5 多尺度通道注意力机制模块结构图

Fig.5 Multi-scale channel attentional mechanism module

首先,对输入特征进行不同尺寸卷积提取特征的操作:

$$X_s = K_{3 \times 3} X + K_{5 \times 5} X \quad (1)$$

式中: K 代表使用不同卷积核对此输入特征进行卷积操作; X_s 为经过卷积后的输入特征。

然后,分别对输入特征进行全局平均池化和最大池化操作,接着分别对平均池化分支 X_a 和最大池化分支 X_m 进行特征提取,之后将它们转化为线性空间,并激活,如公式 (2)、(3)、(4) 所示。

$$X_a = FC(SiLU(FC(X_{avg}))) \quad (2)$$

$$X_m = FC(SiLU(FC(X_{max}))) \quad (3)$$

$$X_z = Sigmoid(X_a + X_m) \quad (4)$$

式中: X_{avg} 和 X_{max} 分别是全局平均池化和全局最大池化操作; FC 为全连接函数; $SiLU$ 为非线性激活函数; $Sigmoid$ 为归一化函数。

最后,将计算所得的注意力权重加权到第一步生成的特征图上作为最后的输出结果。

$$X_w = Scale(X_s, X_x) \quad (5)$$

3.2 融合 RFB 结构的 YOLOv7-tiny 模型

为提高鲁棒性,在模型的主干网络中融入 RFB 网络结构,增加多个不同尺度的特征信息,扩大有效特征层的感受野,提高特征提取能力。可以在浅层特征图中获得更多的高层信息,增强特征提取的鲁棒性,提高目标检测的精度。

RFB 一种新的特征提取模块,通过模拟人类视觉的感受野从而加强网络的特征提取能力,在结构上 RFB 借鉴了 Inception 的多分支网络结构,主要是在 Inception 的基础上加入了空洞卷积,卷积核越大所用的空洞卷积的膨胀率越大,从而有效增大了感受野。RFB 网络结构如图 6 所示,在三个通道上分别使用了 1×1、3×3、5×5 的卷积核进行特征提取,之后用三个 3×3 卷积对应的膨胀率 (rate) 分别为 1、3、5,然后将三个分支不同尺度的有效特征层进行拼接。最后,采用 1×1 的卷积层跨通道与输入的有效特征层进行残差连接,以融合不同大小的感受野,使网络结构通过扩大特征信息提取,进一步提升检测的准确率和速度。

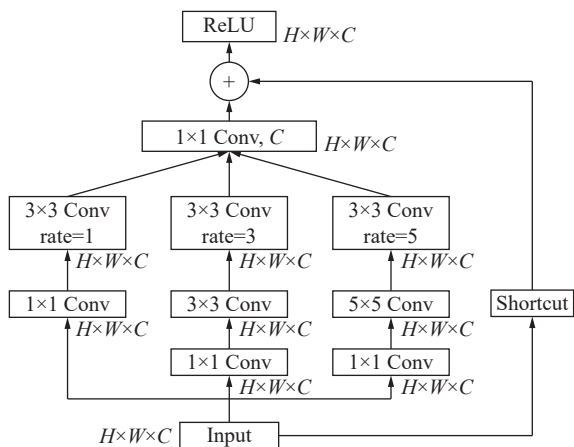


图 6 RFB 网络结构图

Fig.6 RFB network structure diagram

3.3 基于特征融合的小目标检测

检测头主要用于承担主干特征提取网络无法完成的定位任务,通过主干特征提取网络提取到的有效特征图来检测目标物体所在的位置和它的所属类别。因此,文中针对小目标检测的问题,相比于原网络 YOLOv7-tiny,新增了一个针对小目标的探测头,用于提升小目标检测精度。结合原有的检测头,可以更加有效、快速地对小目标物体进行识别检测。

不同检测头所检测到的效果和特征都有所不同,原始的 YOLOv7-tiny 在颈部部分采用了 PANet 特征融合结构,具有三个尺度的检测头,其分别在 20×20、40×40、80×80 大小的特征图上完成目标检测的任务。而文中研究的小目标检测效果并不是很好,对于输入图像尺寸为 640 pixel×640 pixel,若检测的目标在这张图像中占比不到 5% 甚至更小时,物体就会在特征图上消失,导致无法检测到目标物。而现有的网络检测头无法满足小目标检测的要求。所以,采用增加极小目标检测层的方法,虽然增加了参数量导致计算量增大、推理速度减慢,但是大大改善了小目标检测精度低的问题。在 YOLOv7-tiny 网络的基础上,继续增加一层上采样卷积层,并与主干网络中上一层特征提取进行横向连接,得到了 160×160 大小的特征图,这一大尺度特征图负责检测小目标。

3.4 损失函数的优化

YOLOv7-tiny 中的损失函数由置信度损失、分类损失和定位损失组成,其中定位损失采用的是 $CIoU$ 损失函数,其缺点是没有考虑到预测框与真实框只检测不匹配问题,导致收敛速度缓慢,而 $SIoU$ 非常适用于提高收敛速度,总损失计算公式。

$$Loss_{total} = Loss_{loc} + Loss_{conf} + Loss_{class} \quad (6)$$

定位损失则采用的是 $CIoU$ 来进行计算,计算公式如下所示:

$$L_{CIoU} = 1 - IoU + \frac{p^2(b, b^{gr})}{c^2} + \alpha v \quad (7)$$

式中: α 是权重函数; v 用来度量长宽比的相似性。如式所示, v 不变时, IoU 越大, α 越大,损失越大,说明高 IoU 时,更加关注长宽比,低 IoU 时,更关注 IoU 。当真实框和预测框的长宽比接近, v 越小。 v 的值为零时,长宽比的惩罚项没有起到任何作用, $CIoU$ 损失函数就无法稳定表达。

由于 $CIoU$ 没有考虑到所需真实框与预测框之间不匹配的问题。这种不足导致收敛速度缓慢且效率较低,因为预测框可能在训练过程中四处游荡并最终产生更差的模型,得出回归函数如下式所示:

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (8)$$

基于此,文中引用新的损失函数 $SIoU$ ^[19] 替代原网络的 $CIoU$ 损失函数。其中,考虑到所需回归之间

的向量角度,将角度问题纳入收敛方向的考量当中,使模型能够更快速的找到最优收敛方向,并重新定义了惩罚指标。相较于原有的 $CIoU$ 损失函数,新的损失函数显著提高了训练的速度和推理的准确性。

3.5 可变形卷积

可变形卷积是在普通卷积的基础上,增加了自适应学习的水平方向和垂直方向的偏移,使得对输入的特征图不再是在固定的位置进行采样,这样卷积核就能在训练过程中扩展到很大的范围。因此,可变形卷积在采样时可以更贴近物体的形状和尺寸,更具有鲁棒性。图 7 所示为可变形卷积示意图。

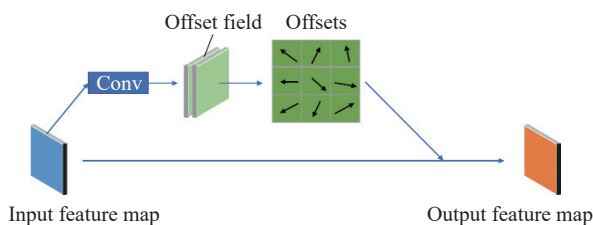


图 7 可变形卷积示意图

Fig.7 Deformable convolution diagram

对于普通卷积,以 3×3 卷积为例,对于每个输出 $y(p_0)$, $p_0=(0,0)$,都要从 x 上面采样九个位置,这九个位置都在中心位置 $x(p_0)$ 向四周扩散, $(-1,-1)$ 代表 $x(p_0)$ 的左上角, $(1,1)$ 代表 $x(p_0)$ 的右下角。采样区域包括:

$$R = \{(-1,-1),(-1,0),\dots,(1,1)\} \quad (9)$$

传统卷积的输出为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (10)$$

式中: p_n 为采样网格中的第 n 个点; $w(p_n)$ 为对应的卷积核权重系数。

可变形卷积在传统的卷积操作上加入一个偏移量 Δp_n , 偏移量使得卷积变形为不规则卷积,可变形卷积输出如下式,特征值通过双线性插值的方法来计算。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (11)$$

4 实验结果与分析

4.1 数据集的建立及实验环境配置

文中实验数据集是通过光学设备拍摄以及大连

理工无人机数据集合并成,总共 3800 张无人机图片,并使用 LabelImg 图像标注工具对数据集图片中的无人机进行标注。数据集按 7:3 的比例随机划分为训练集和测试集,其中训练集 2660 张图片,测试集 1140 张图片。为了获得更好训练结果,对网络训练的超参数进行设置,输入图片的尺寸为 640×640 ,优化器采用 SGD 优化器,SGD 动量 (Momentum) 设为 0.937,批量大小设为 8 和 16,训练轮次为 100 个 epoch,前 3 轮采用热身训练策略,热身训练动量为 0.8,初始学习率为 0.01,终止学习率 0.1,权重衰减 (Weight Decay) 系数为 0.0005,其余超参数均为默认。实验的环境:编程语言 python3.8,深度学习框架为 Pytorch,采用 Win10 操作系统,CPU 为 12 th Gen Intel(R) Core(TM) i5-12400 @ 2.50 GHz,内存 16 GB;GPU 为 8 G 的 NVIDIA GeForce RTX2060 SUPER。

4.2 评价指标

在目标检测的算法中,平均精度均值 (Mean Average Precision, mAP) 和检测速度 (Frame Per Second, FPS) 是评价检测性能的重要指标。为了验证模型的检测性能,文中采用 $mAP@0.5$, $mAP@0.5$ 表示 IOU 阈值为 0.5 时的平均精确度、检测速度 FPS,即每秒帧率,每秒内可以处理的图片数量、参数量 (Parameters) 和计算量 (GFLOPS)。

4.3 改进注意力机制实验

为了验证在主干网络中引入注意力机制模块对模型性能的影响,实验将常用到的 SENet、CBAM (Convolution Block Attention Module)^[20]、EMA 注意力模块和文中提出的多尺度通道注意力机制 SMSSEnet 模块进行四组对比实验,表 1 为 YOLOv7-tiny 引入不同注意力机制模块前后的检测性能对比表。实验表

表 1 引入不同注意力机制算法检测性能对比表

Tab.1 Comparison of detection performance of different attention mechanism algorithms

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
+SE	6.05	84.5	75	13.2
+CBAM	6.02	84.9	73	13.3
+EMA	6.06	85.9	75	13.5
+SMSE	8.99	86.8	71	15.6

明,引入不同注意力机制模块均会是网络模型的平均精度有所提高,但是 YOLOv7-tiny 在加入 SMSE 注意力机制模块后,无人机检测精度最高,图 8 为加入多尺度通道注意力机制前后检测结果图。



图 8 加入注意力机制前后检测结果对比图

Fig.8 Comparison chart of test results before and after adding attention mechanism

4.4 引入 RFB 结构实验

在 YOLOv7-tiny 的主干网络与特征融合网络 (PANet) 之间引入 RFB 结构,引入 RFB 结构后的算法与原 YOLOv7-tiny 算法相比,平均精度提高了 1.3%,虽然参数量有所增加,但是却大大的提高了模型的精度,说明了 RFB 结构通过扩大感受野,提高特征提取能力,有效的提升了网络的检测精度,改进算法前后检测性能对比表,如下表 2 所示。图 9 为加入 RFB 结构前后检测结果图。

表 2 引入 RFB 结构算法检测性能对比表

Tab.2 The RFB structure algorithm is introduced to detect the performance comparison

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
+XMB	6.30	85.6	77	14.7

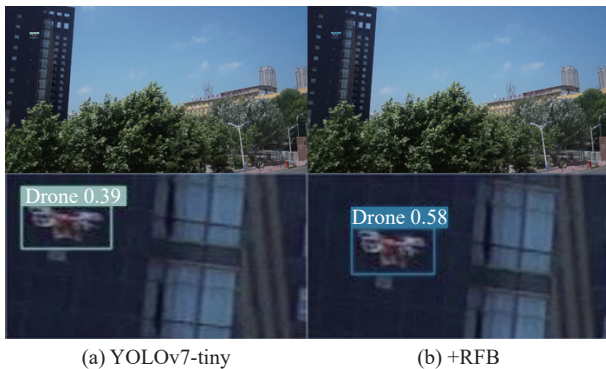


图 9 加入 RFB 结构前后检测结果图

Fig.9 Before and after adding RFB structure

4.5 引入小目标检测层实验

引入小目标检测层算法,借助 FPN 特征融合的思想构建小目标检测层,以 YOLOv7-tiny 为基础,在 FPN 基础上再次进行一个上采样并于主干网络提取层进行特征融合,增强了特征图表达,减缓了小目标漏检的问题。经过实验验证,改进后的算法 YOLOv7-tiny-XMB 与原算法 YOLOv7-tiny 相比,平均精度提高了 0.8%,检测速度为 70 帧/s,降低了 4 帧/s,同时参数量也有所增加。虽然检测速度有所减慢,但依旧可以满足实时性需求。改进特征融合层算法前后检测性能对比表,如下表 3 所示。图 10 为引入小目标检测层前后检测结果图。

表 3 引入小目标检测层算法检测性能对比表

Tab.3 The small target detection layer algorithm is introduced to detect the performance comparison

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
+XMB	6.10	85.1	70	15.5



(a) YOLOv7-tiny (b) +XMB

图 10 引入小目标检测层前后检测结果图

Fig.10 Before and after the introduction of small target detection layer detection results map

4.6 改进损失函数实验

改进损失函数算法前后检测性能对比表,如表 4 所示,改进后的 YOLOv7-tiny-SIoU 检测算法较原来的 YOLOv7-tiny 算法相比,平均精度提升了 0.7%,并且检测速度也比原来快了 4 帧/s,可见新的 Siou 损失函数较原来的 CIoU 增加了角度成本的计算以后,使

得网络收敛速度更快更平稳,并且显著的改善了回归损失的精度,降低了预测误差。图 11 为加入 *SIoU* 损失函数前后检测结果图。

表 4 改进损失函数算法检测性能对比表

Tab.4 Improved loss function algorithm detection performance comparison

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
+ <i>SIoU</i>	6.02	85.0	78	13.2



(a) YOLOv7-tiny (b) +*SIoU*

图 11 加入 *SIoU* 结构前后检测结果图

Fig.11 Before and after adding *SIoU* structure

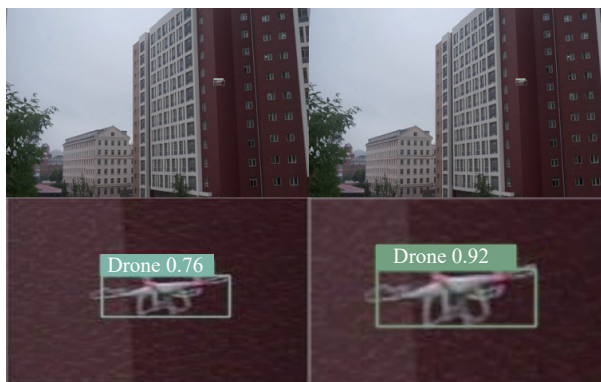
4.7 引入可变形卷积实验

引入可变形卷积前后检测性能对比表,如下表 5 所示,引入可变形卷积后的 YOLOv7-tiny-DCN 算法与原 YOLOv7-tiny 算法相比,虽然检测速度有所下降,计算量有所增大,但是平均精度提升了 2.3%。由此可见,引入可变形卷积后,增加了水平和垂直方向偏移量,使得采样更加贴近物体的形状和尺寸,模型更具有鲁棒性。图 12 为加入可变形卷积前后结果对比图。

表 5 引入可变形卷积前后检测性能对比表

Tab.5 A comparison of detection performance before and after deformable convolution is introduced

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
+DCN	6.08	86.6	71	14.8



(a) YOLOv7-tiny (b) +RFB

图 12 加入可变形卷积结构前后检测结果图

Fig.12 Before and after adding DCN structure

4.8 消融实验

消融实验应用了控制变量法的原理,在复杂的深度神经网络中,通过去除网络某些部分,以便更好的理解网络的行为。因此,文中通过逐个加入各个模块来实验验证改进的各个模块对算法检测性能的影响。实验结果如表 6 所示,将改进的四个部分逐个加入到原算法 YOLOv7-tiny 中,与原 YOLOv7-tiny 算法比较。最后,得到每组实验的参数量、平均精度、检测速度和计算量,虽然计算量和参数量有所提高,检测速度有所减慢,但是平均精度得到了显著的提高,改进后的最终算法 YOLOv7-drone 比原算法 YOLOv7-tiny 在平均精度上高了 6.1%,同时检测速度也略有提高。

表 6 逐步加入各个模块算法检测性能对比表

Tab.6 Gradually add each module algorithm detection performance comparison

SMSE	RFB	XMB	<i>SIoU</i>	DCN	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
					6.02	84.3	74	13.2
√					8.99	86.8	71	15.6
√	√				9.29	87.4	73	17.1
√	√	√			9.39	88.2	69	20.1
√	√	√	√		9.39	88.7	73	20.1
√	√	√	√	√	9.45	90.4	72	21.7

4.9 不同目标检测算法的对比实验

为了进一步验证文中提出的算法的检测性能,文中将改进后的算法 YOLOv7-drone 与 YOLOv7-tiny、

YOLOv7、YOLOv5l 三种目标检测算法进行对比实验。实验结果如表 7 所示,可以看出,文中提出的 YOLOv7-drone 算法的平均精度达到 90.4%,检测速度达到 72 帧/s,与其他目标检测算法 YOLOv7-tiny、YOLOv7、YOLOv5l 相比,平均精度分别提升了 6.1%、3.1% 和 4%,验证了文中算法的可行性,对其中于一张图片用不同算法检测,所得不同算法检测性能对比图如图 13 所示。

表 7 不同目标检测算法检测性能对比表

Tab.7 Comparison of detection performance of different target detection algorithms

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	84.3	74	13.2
YOLOv7	37.2	87.3	57	104.8
YOLOv5l	46.1	86.4	42	107.9
YOLOv7-drone	9.45	90.4	72	21.7

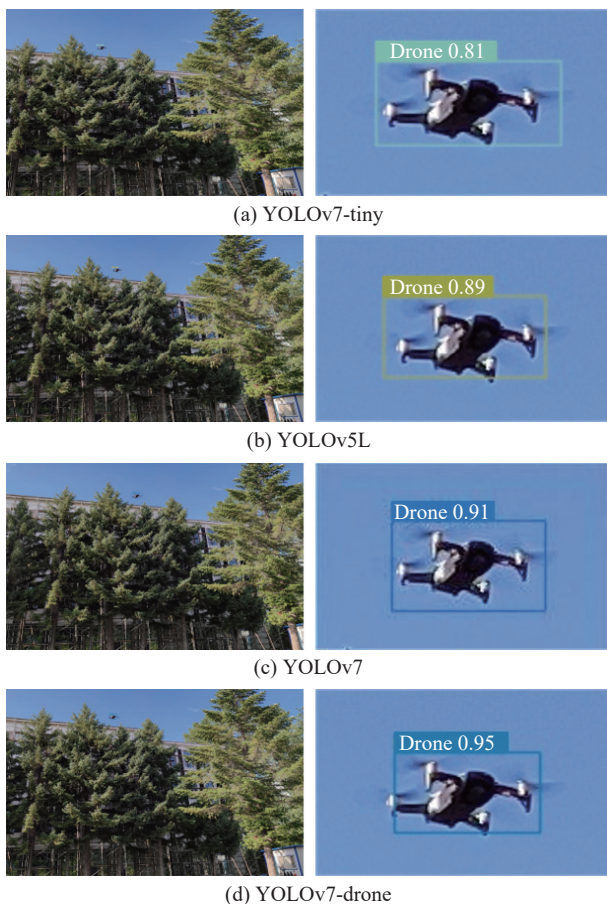


图 13 不同算法检测性能对比图

Fig.13 Comparison of detection performance of different algorithms

为了更进一步的验证改进后的算法对复杂背景下小尺度无人机的检测性能,引入了 CAM (Class Activation Mapping) 热力图可视化模块。其中,红色的区域为网络模型着重关注区域,颜色越深代表关注度越高。算法改进前后热力图对比如图 14 所示,观察发现改进后热力图中无人机与周围环境区分更加明显。

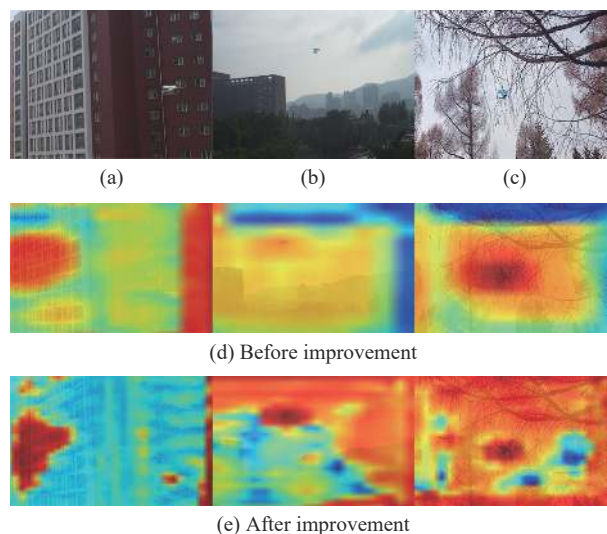


图 14 算法改进前后 CAM 对比图

Fig.14 Before and after the algorithm improvement CAM contrast chart

4.10 VOC 公共数据集实验

为了验证改进的算法针对不同数据集依旧又很好的检测性能,文中引入了 PASCAL VOC 公共数据集,它是一项视觉对象类挑战赛,对提供的二十类图像进行分类。应用提出的算法对 PASCAL VOC 数据集进行训练,同时用测试集对其进行验证。文中将改进后的 YOLOv7-drone 算法和 YOLOv7-tiny 算法在 PASCAL VOC 数据集上训练后的平均精度对比表如表 8 所示,实验结果表明改进后的算法 YOLOv7-drone 较原算法 YOLOv7-tiny,在平均精度上高了 6%,证明了文中改进算法的有效性。

表 8 PASCAL VOC 数据集检测性能对比表

Tab.8 PASCAL VOC dataset detection performance comparison

Model	Params/M	mAP@0.5	FPS/frame·s ⁻¹	GFLOPS
YOLOv7-tiny	6.02	65.0	74	13.2
YOLOv7-drone	9.45	71.0	72	21.7

5 结束语

1) 针对公园、游乐场、体育场等复杂背景下小尺度无人机的检测会出现误检、漏检、检测精度不高和检测速度不够快的问题,设计了一种基于多尺度通道注意力机制和小目标特征融合层的 YOLOv7-drone 无人机目标检测算法。

2) 采用自制光学设备采集无人机图片,并与 DUT-Anti-UAV 数据集合并,共同构建文中数据集;引入多尺度通道注意力机制模块获得多尺度特征信息,并着重关注局部特征信息以便检测小尺度无人机;引入 RFB 结构增大感受野,提高浅层网络特征提取能力,以便在浅层结构获得高等语义信息;新增小目标特征融合检测层以适应更小尺度无人机的检测;将算法原有的损失函数替换为 *SIOU* 损失函数,改善回归损失精度,降低误差,加快网络的收敛速度;将普通卷积替换为可变形卷积以适应不同形状和大小的目标。

3) 将改进后的算法 YOLOv7-drone 与原算法 YOLOv7-tiny 进行对比实验,实验结果表明,对于在复杂背景下小尺度目标情况下的无人机检测效果更好;将其与不同目标检测算法 (YOLOv7、YOLOv5l、YOLOv7-tiny) 相比,具有更好的检测性能。同时,在视觉公共数据集 PASCAL VOC 上进行对比实验,实验结果表明,改进后的 YOLOv7-drone 算法相较于原算法仍有更优异的检测效果。

参考文献:

- [1] Xue Shan, Chen Yuchao, Lv Qiongying, et al. Image recognition method of anti drone system based on coordinate attention mechanism [J]. *Infrared and Laser Engineering*, 2022, 51(9): 20211101. (in Chinese)
- [2] Xue Shan, Lu Tao, Lv Qiongying, et al. The drone target detection algorithm based on multiscale fusion and lightweight network[J]. *Journal of Hunan University (Natural Sciences)*, 2023, 50(08): 82-93. (in Chinese)
- [3] Xue Shan, Zhang Yaliang, Lv Qiongying, et al. Anti-UAV system object detection algorithm under complex back ground [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2023, 53(3): 891-901. (in Chinese)
- [4] Xue Shan, Wang Yabo, Lv Qiongying, et al. Anti-occlusion target detection algorithm for anti-UAV system based on YOLOX-drone[J]. *Chinese Journal of Engineering*, 2023, 45(9): 1539-1549. (in Chinese)
- [5] Xue Shan, Wei Liwei, Gu Chenyu, et al. Drone identification method based on mixed domain attention mechanism [J]. *Journal of Xi'an Jiao Tong University*, 2022, 56(10): 141-150. (in Chinese)
- [6] Dai Xinxue, Fan Songtao, Zhou Yan. Speech enhancement method for laser microphone based on ResUNet and TFGAN networks [J/OL]. *Infrared and Laser Engineering*: 1-10[2023-09-20]. (in Chinese)
- [7] Gedmon J, Donahue J, Darrell T, et al. Region based convolutional Networks for accurate object detection and segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(1): 142-158.
- [8] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, Real-Time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2015: 779-788.
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21-37.
- [10] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Train-able bag-of-freebies sets new state-of-the-art for real-time object detectors [DB/OL]. (2022-06-06) [2023-07-30]. <https://arxiv.org/abs/2207.02696>.
- [11] Zhao J, Zhang J H, Li D D, et al. Vision-based anti-UAV detection and tracking[DB/OL]. (2022-05-22) [2023-07-30]. <https://arxiv.org/abs/2205.10851>
- [12] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection [DB/OL]. (2020-04-23) [2023-07-30]. <https://arxiv.org/abs/2004.10934>.
- [13] Zhang H Y, Cisse M, Dauphin Y N, et al. mixup: Beyond Empirical Risk Minimization[DB/OL]. (2018-04-27) [2023-07-30]. <https://arxiv.org/abs/1710.09412>.
- [14] Zhang X, Zeng H, Guo S, et al. Efficient longrange attention network for image superresolution [DB/OL]. (2022-03-13) [2023-07-30]. <https://arxiv.org/abs/2203.06697>.
- [15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//IEEE Trans Pattern Anal Mach Intell, 2015, 37(9): 1904.
- [16] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 8759.
- [17] 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, Honolulu, 2017:

2117. defect detection algorithm for metal bipolar plate [J]. *Acta Harbin Institute of Technology*, 2023, 55(12): 104-112. (in Chinese)
- [18] Hu J, Shen L G. Squeeze-and-excitation networks [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [19] Ling Qiang, Liu Yu, Wang Chunju, et al. DN-YOLOv5 surface [20] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich: ECCV, 2018: 3-19.

Image target detection algorithm based on YOLOv7-tiny in complex background

Xue Shan^{1,2*}, An Hongyu¹, Lv Qiongying¹, Cao Guohua²

(1. College of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China;

2. Chongqing Research Institute, Changchun University of Science and Technology, Chongqing 400000, China)

Abstract:

Objective Once the "black flying" drone carries items such as bombs, it can pose a threat to people. Target detection of "black flying" drones in complex backgrounds such as parks, amusement parks, and schools is the key to anti-drone systems in public areas. This paper aims to detect small-scale targets in complex background. Because the traditional manual image feature extraction methods are not targeted, time complexity is high, windows are redundant, the detection effect is poor, and the average accuracy is low. The problems of false detection and missing detection will occur when detecting small-scale UAVs in complex background. Therefore, this paper aims to develop a black flying UAV detection model based on deep learning, which is essential for the detection of unmanned aerial vehicles.

Methods YOLOv7 is a stage target detection algorithm without anchor frame, with high detection accuracy and good inference speed. YOLOv7-tiny belongs to the grain grabbing memory model, with fewer parameters and fast operation, making it widely used in industry. In the backbone network, the built multi-scale channel attention module SMSE (Fig.5) is introduced to enhance the attention of UAVs in complex backgrounds. Between the backbone network and the feature fusion layer, the RFB feature extraction module (Fig.6) is introduced to increase the Receptive field and expand the feature information extraction. In the feature fusion, the small target detection layer is added to improve the detection ability of small UAV targets. In terms of calculating losses, the introduction of *SIOU* Loss function redefines the penalty index, which significantly improves the speed of training and the accuracy of reasoning. Finally, the ordinary convolution is replaced by the deformable convolution (Fig.7), making the detection closer to the shape and size of the object.

Results and Discussions The dataset selected in this article is a combination of the self-made dataset (Fig.1) and the Dalian University of Technology drone dataset (Fig.2). The mainly used evaluation indicators are mAP (mean accuracy) and FPS (detection speed), Params (parameter quantity) and GFLOPS (computational quantity) as secondary indicators. Each module was compared with the original algorithm, including attention comparison experiment (Tab.1), RFB module comparison experiment (Tab.2), small target detection layer comparison experiment (Tab.3), Loss function comparison experiment (Tab.4), and deformable convolution comparison experiment (Tab.5). And ablation experiments were conducted (Tab.6), which confirmed the effectiveness and

feasibility of the proposed algorithm through mAP comparison, improving accuracy by 6.1%. On this basis, the detection performance of different algorithms was compared (Tab.7), and the generalization of the algorithm was verified on the VOC public dataset (Tab.8).

Conclusions This article proposes an improved object detection algorithm for anti-drone systems. Through the multi-scale channel attention module, the attention of small targets is enhanced, the fusion RFB increases the Receptive field, adds a small target detection layer to improve the detection ability, and improves the Loss function to improve the training speed and reasoning accuracy. Finally, deformable convolution is introduced to better fit the target size. The improved algorithm has achieved good detection results on different datasets.

Key words: target detection; complex background; attention mechanism; small target detection

Funding projects: Key R & D projects of Jilin Provincial Science and Technology Department (20210203055SF); Science and Technology Research Project of Jilin Provincial Department of Education