文章编号:1671-4598(2024)01-0057-07 DOI:10.16526/j.cnki.11-4762/tp.2024.01.009 中图分类号:TP391.41

• 57 •

文献标识码:A

基于改进轻量化 YOLOX 的无人机航拍 目标检测算法

胡 潇,潘申富

(中国电子科技集团公司 第54研究所,石家庄 050081)

摘要:针对小型无人机在巡逻航拍中的应用,提出了一种改进的轻量化目标检测算法,有效解决巡逻过程中空地无线传输信 道和机载端计算能力双重受限的难题;该算法在YOLOX算法的基础上,首先利用 Mobilenetv2 代替 CSPDarknet 骨干网络作为特 征提取网络,降低了模型参数量和计算量,提高目标检测实时性;其次为了弥补轻量化带来的检测精度下降,考虑检测目标框的 长宽比引入 CIOU 定位损失函数,提升目标定位的精度;同时为了平衡训练过程中的正负难易样本,引入 Focal Loss 置信度损失 函数提升模型的检测性能;基于 VisDrone2019-DET 数据集实验表明,改进后算法模型参数量降低了 56.2%,计算量降低了 52.5%,在检测精度没有明显下降情况下单张图片推理时间减少了 41.4%;最后,将改进后的算法部署到 Nvidia Jetson Xavier NX 机载端,测得模型检测帧率可达 22 FPS,改进后算法满足巡逻任务的应用需求。

关键词:无人机;目标检测;轻量化;YOLOX;Focal Loss;CIOU

A Lightweight Algorithm for UAV Aerial Image Objects Detection Based on Improved YOLOX

HU Xiao, PAN Shenfu

(The 54th Research Institute of CETC, Shijiazhuang 050081, China)

Abstract: Aimed at small unmanned aerial vehicle (UAV) patrol applications, a lightweight object detection algorithm is proposed to effectively solve the dual constraints of wireless transmission channel and onboard computing resource during the patrol process. Firstly, the Mobilenetv2 network is used as feature extraction network of YOLOX algorithm to reduce the parameters and computation of the model, and improve the speed of object detection. Secondly, the CIOU loss function is introduced to improve the precision of object positioning. Thirdly, a Focal Loss confidence loss function is introduced to balance the positive and negative or difficult and easy samples in training process, improving the performance of the model. Experimental results based on VisDrone2019-DET dataset show that the improved algorithm reduces the model parameters by 56.2%, the calculation by 52.5%, and the inference time of single image by 41.4%, without a significant decrease in detection accuracy. Finally, the improved algorithm is deployed to the Nvidia Jetson Xavier NX, and the model detection frame rate reaches by 22 FPS, which meets the application requirements of patrol tasks.

Keywords: UAV; object detection; lightweight; YOLOX; Focal Loss; CIOU

0 引言

近年来,多旋翼无人机因其行动灵活、操作门槛低、 设备配置可塑性高,主动控制能力强¹¹等优点在多个领域 中得到了广泛的应用。在边境、城市巡逻中¹²¹,无人机可 以提供不同视角的现场图像,便于工作人员掌握现场情况。 随着人工智能的发展,无人机更加智能自主¹³¹,将目标检 测技术应用于无人机,巡逻人员可以更快速地感知并捕获 现场关键信息,提高巡逻的效率。传统方案将检测算法部署 在远程服务器中需要高清稳定的图像传输,对无线传输信道 的要求高。在无线传输信道受限情况下,将检测算法部署在 机载侧可以实现对连续采集图像进行目标检测,并根据信道 情况向地面工作人员传输图像或文字等类型检测结果,减轻 检测任务对通信条件的依赖,提高巡逻效率。由于小型机载 终端计算资源有限,需要对现有检测算法进行针对性的优化 训练,设计轻量化的无人机航拍目标检测算法。

目前基于深度学习的目标检测算法多基于卷积神经网络,在主流数据集上有很好表现。Sevo和Avramovic^[4]证明卷积神经网络可以有效地结合到航空图像的目标检测算法中。一般来说,基于 CNN 的目标检测算法分为单阶段和双阶段算法。R-CNN^[5]和 Fast R-CNN^[6]是基于区域特征的双阶段算法,双阶段算法在选择候选区域时耗时长、候选帧之间存在大量重叠、提取的特征冗余。单阶段目标检测算

收稿日期:2023-11-02; 修回日期:2023-11-23。

引用格式:胡 潇,潘申富.基于改进轻量化 YOLOX 的无人机航拍目标检测算法[J].计算机测量与控制,2024,32(1):57-63.

作者简介:胡 潇(2000-),女,硕士研究生。

通讯作者:潘申富(1975-),男,博士,研究员。

第 32 卷

法包括 SSD^[7]、YOLO^[8]和 RetinaNet^[9]等,单阶段算法速度快,更能得到航拍任务的青睐。Zhu^[10]等人在 YOLOv5 算法基础上增加了一个用于小微物体检测的预测头,提出 TPH-YOLOXv5 算法,提高了检测精度。许多学者通过增 加多种模块提升网络的检测性能,但复杂的模型算法需要 很大的内存和计算开销,无法应用于机载终端。

随着嵌入式目标检测成为研究热点,部分学者通过模型剪枝量化、替换轻量化模块等方法进行轻量化网络的设计,解决模型复杂度高无法运行在边缘端的问题。崔令飞^[11]等人将 SSD 算法中的特征提取网络替换为轻量化的 MobileNet 网络进行模型优化,构建了轻量化目标检测算 法。黄海生^[12]等人简化了 YOLOv5 网络,提出了一种轻量 化检测网络 YOLOv5-tiny,提升了检测速度。但小型机载 终端计算资源更加有限,需要针对巡逻应用进一步设计轻 量化算法。

针对巡逻应用,论文首先选用单阶段检测算法中速度更快的 YOLOX 算法,在该算法基础上使用轻量化 Mobilenetv2^[13]网络替换原有 CSPDarknet 骨干网络,降低参数量; 其次训练过程中为了更好的平衡正负、难易样本,使用 Focal Loss 损失函数替代二元交叉熵置信度损失函数;然后将定位 损失改为 CIOU 损失函数,改善网络轻量化后检测精度的下 降问题,得到改进后的轻量化目标检测算法。最后将算法移 植到 Jetson Xavier NX 机载终端,实现目标检测应用。

1 YOLOX 目标检测算法

YOLOX^[14]是 2020 年旷视基于 YOLOv3 提出的高性能 单阶段检测器,目前已经成为工业界实用算法之一。依照模 型缩放的规则 YOLOX 有 YOLOX-s、YOLOX-m、YOLOX-l 和 YOLOX-x 四种不同参数量的模型。其中 YOLOX-s 模型 参数量较少,且不像 YOLOX-Nano 等模型因过小的尺寸损 失了检测精度,达到了检测速度和检测精度的平衡,因此 采用 YOLOX-s 作为改进对象。其网络主要包括输入端、特 征提取网络 Backbone、Neck 特征融合层和 Head 预测头, 网络结构如图 1 所示。

 图像输入端。训练前对输入的训练数据集进行数据 增强,丰富检测数据集。实际检测时输入端向网络输入一 张采集的航拍图像。

2) 骨干网络 Backbone。YOLOX-s 沿用 YOLOv5 采用 的 CSPDarknet 网络进行特征提取,首先使用 Fcous 结构对 输入图像切片堆叠,展开特征信息。然后通过 conv (卷积) +BN (归一化) +SiLu 激活函数的 CBL 模块和融合残差 结构的 CSP 模块进行特征提取,在最后一层经过 SPP (Spatial Pyramid Pooling) 结构改善感受野捕捉特征,最终输出 三个不同尺寸的有效特征图。

3)特征融合网络 Neck。对骨干网络输出不同尺度的三 个特征图实现特征融合。采用 FPN(feature pyramid networks)与 PAN(pyramid attention network)级联结构, FPN 对特征图进行上采样向浅层传递深层特征信息,PAN 下采样实现深层信息和浅层信息的特征融合。



4)预测层。针对 Neck 层获得的三个加强特征图预测 目标位置、类别和置信度。传统的耦合检测头使用相同的 特征图通过 1×1 的卷积对被检测对象进行分类定位,无 法获得良好的检测效果^[15]。YOLOX 的预测层 Head 采用 解耦结构,先使用一个 1×1 的卷积对特征融合层的输出 进行降维,然后在分类和回归两个分支中各使用两个 3×3 的卷积实现位置类别和置信度的分别预测,结果更有说 服力。

同时,YOLOX采用了无锚框的方法并引入 SimOTA 方法动态分配正负样本,减少了聚类超参数对于预测结果 的影响,在减少计算量基础上提高模型准确率,与部署在 巡逻无人机机载端中需要考虑网络推理速度快的要求相符 合,因此以 YOLOX-s 为基础算法进行轻量化改进。

然而,YOLOX 算法虽能在实验室的高性能 GPU 中可 以得到优秀的检测效果,但在计算资源有限的无人机机载 侧的边缘终端部署的目标检测算法,需要进一步轻量化, 降低参数规模。

2 改进的 YOLOX 轻量化目标检测算法

2.1 改进后整体网络结构

针对巡逻航拍场景,为降低模型参数保持检测精度, 在YOLOX算法基础上,分别对骨干网络和损失函数进行 了改进,对训练过程进行微调,设计了轻量化改进算法, 网络结构如图2所示。

在改进的网络结构中,为获取更多特征,将网络输入 的图像大小均缩放为 960×960 大小。改进后将 YOLOX 的 特征提取网络替换为 Mobilenetv2 轻量化网络,由于 Mobilenetv2 中最后的池化层和卷积层与 YOLOX 的特征融合 层 PAFPN 功能冗余,为更符合轻量化设计目标将其去除。 利用新的骨干网络对航拍图像特征进行提取,输出 120× 120×32,60×60×96,30×30×320 三种特征图送入 PAF-PN 模块,经过 Neck 层特征传递融合输出三个通道数均为 96 的特征图,分别送入各自不共享权重的三个预测头 Head 中。另外,为提升轻量化后的检测精度,在模型训练过程



图 2 改进后的网络结构

中将置信度损失替换为 Focal Loss,定位损失设计为 CIOU Loss,得到预测框的目标类别、目标边框位置和置信度信息。最后将三种大小特征图的预测框信息进行合并,经过 NMS 非极大值抑制得到改进后网络的最终预测结果。下文详细介绍了具体的改进设计。

2.2 骨干网络轻量化

与 VGG-16 和 Mobilenetv1 骨干网络相比, Mobilenetv2 网络具有更小的体积和计算量,但具有相当的准确率和更快的速度。因此,在对 YOLOX 的轻量化改进中将骨干网络替换为 Mobilenetv2 进行图像特征提取。

将输入图像放缩到 960×960 大小。表1为具体的网络 结构, *t* 为 Bottleneck 结构中的通道扩展因子, *s* 代表步长, *c* 是输出通道数, *n* 是操作重复次数。

输入	操作	t c		п	5
$960^2 imes 3$	Conv2D	—	32	1	2
$480^2 imes 32$	Bottleneck	1	16	1	1
$480^2 imes 16$	Bottleneck 6		24	2	2
$240^2 imes 24$	Bottleneck	6	32	3	2
$120^2 imes 32$	Bottleneck	6	64	4	2
$60^2 imes 64$	Bottleneck	6	96	3	1
$60^2 imes 96$	Bottleneck	6	160	3	2
$30^2 imes 160$	Bottleneck	6	320	1	1
$30^2 imes 320$	Conv2D 1×1	—	1 280	1	1
$30^2 \times 1\ 280$	Avgpool	—	—	1	—
$1^2 \times 1$ 280	Conv2D 1×1	_	k	_	—

表 1 Mobilenetv2 网络结构

Mobilenetv2 网络中最主要的结构是轻量级的 Bottleneck,与 YOLOX-s 网络的 CSP 模块中的 Bottleneck 结构不 同,由于 Mobilenetv2 网络中使用了降低参数量和计算量的 深度可分离卷积,为了不丢失特征,在 Bottleneck 中设计了 倒残差线性瓶颈结构,该结构是网络的核心。

2.2.1 深度可分离卷积

Mobilenetv2 的 Bottleneck 主要应用了深度可分离卷积, 卷积是卷积神经网络中算法的核心,深度可分离卷积是 CNN 的一种,与标准卷积相比需要调整的参数更少,计算 量更小,因此采用该操作可以实现网络的轻量化。深度可 分离卷积 2016 年首次在谷歌的 Xception^[16] 网络结构中提 出,Mobilenetv2 的 Bottleneck 模块继承并使用了该结构。 它将传统标准卷积拆分为一层深度卷积和一层逐点卷积, 利用深度卷积分别对特征图各个通道做卷积操作,一个卷 积核对应负责一个通道,深度卷积输出的特征图各通道再 利用逐点卷积进行关联。最终得到的输出特征图效果与标 准卷积相同,但该操作极大的降低了计算量。

下文给出了两种卷积操作的计算量对比,具体方法和 卷积操作对比如图 3 和图 4 所示。



图 3 深度可分离卷积

设 *S_k*×*S_k*×*M* 为卷积核的大小,*N* 为输出特征图的个数,*S_F* 为输出特征图的通道数,分别计算两种卷积的参数 量和计算量得到:

标准卷积的参数量:

$$S_k \times S_k \times M \times N \tag{1}$$



图 4 标准卷积与深度可分离卷积对比

标准卷积的计算量:

 $S_k \times S_k \times M \times N \times S_F \times S_F$ (2)

深度可分离卷积的参数量:

$$S_k \times S_k \times M + 1 \times 1 \times M \times N \tag{3}$$

深度可分离卷积的计算量:

 $S_k \times S_k \times M \times S_F \times S_F + M \times N \times S_F \times S_F$ (4) 通过计算得到,深度可分离卷积的计算量和参数量是 标准卷积的:

$$\frac{1}{N} + \frac{1}{S_k} \tag{5}$$

通过详细计算可以证明该设计极大的降低了卷积的参数量和计算量,由该结构组成的网络更加轻量化,对硬件 计算资源的要求更低。

2.2.2 线性瓶颈残差块

Mobilenetv2 中 Bottleneck 是线性瓶颈残差结构,基于 深度可分离卷积设计,分为线性瓶颈结构和残差结构。残 差结构早已在 ResNet^[17]中证明了对网络检测精度有提高作用,在 YOLOX-s 网络中的 CSP 模块中也有应用。

而 Mobilenetv2 中对原始残差结构在卷积操作和维度 操作上进行了改进,设计了倒残差结构,提升轻量化后网 络的检测精度。传统残差结构先通过 1 * 1 卷积降维、3 * 3 卷积提取特征后再通过 1 * 1 卷积升维。Mobilenetv2 将 残差结构中的卷积操作替换为深度可分离卷积,降低残差 网络的计算量。其次由于深度可分离卷积提取的特征被限 制在输入特征的维度中,若在此基础上使用 ResNet 的残 差块则会在输入特征图的基础上先降维,先压缩再进行特 征提取会丢失到大量的特征。因此,网络设计为先通过 1 * 1 逐点卷积扩张特征图通道数,在更高维的特征图中通 过 3 * 3 卷积提取特征,然后通过 1 * 1 卷积降维,同时将 输入特征与经过逆向残差结构输出的特征图相加,形成最 终的输出特征。

Bottleneck 在倒残差结构中加入线性操作。由于传统残 差结构中采用统一的 Relu 激活函数,在高维空间中 Relu 有 效增加非线性能力,但在深度可分离卷积的低维空间中若 输入特征为负数则该特征被清零破坏造成信息丢失,因此 在逐点卷积后将 Relu 改为 Linear 激活函数减少信息丢失对 网络进行优化。 实际网络设计中表 1 的所有 Bottleneck 中,若步长为 1,输入输出特征图相同则使用残差结构。若步长设计为 2 时输入输出特征图不同则不使用该残差结构。该设计很大 程度上丰富了特征数量,同时保留输入特征的信息,避免 出现梯度消息和梯度爆炸的问题。具体 Bottleneck 模块的结 构如图 5 所示。



图 5 MobilenetV2 中 Bottleneck 模块

在骨干网络替换时,将 Mobilenetv2 中最后的池化层和 卷积层去除用以保证不与 YOLOX 的 Neck 层功能重复,将 其作为 Backbone 加入检测算法中,使模型更加轻量化。

2.3 损失函数改进

训练过程中,模型损失的计算是将网络预测结果与真 实结果进行对比,计算梯度回传从而不断调整模型的网络 参数进行优化,模型训练的效果与损失函数的设计强相关。 在训练过程中,所有特征点生成的预测框被 SimOTA 策略 动态分配为正样本和负样本,网络通过正样本学习目标特 征,通过负样本学习背景特征。

在 YOLOX 的解耦头中,总共有三个小的预测分支, 分别为得到预测框类别的分支,判断预测框是目标或背景 的分支以及对预测框的坐标信息(*x*,*y*,*w*,*h*)进行回归 的分支。因此网络的损失计算分为三部分,分别是分类损 失、置信度损失和定位损失。定位损失和分类损失仅针对 正样本进行计算,置信度损失包含正负所有样本。计算公 式如下:

$$L = L_{\rm cls} + \lambda L_{\rm reg} + L_{\rm obj} \tag{6}$$

定位部分 L_{reg} 采用 $IoU^{[18]}$ 交并比作为边界框预测的损失 函数,分类 L_{cls} 和置信度 L_{obj} 均采用二元交叉熵损失函数。 2.3.1 CIOU Loss

在训练过程中,已知所有目标的真实框和其对应的特征点,网络可以针对该特征点得到一个预测框,计算定位损失回传使预测框向真实框方向移动。YOLOX 预测头的定位损失采用交并比 IOU,计算预测框与真实框之间的重叠面积。IOU 损失函数计算公式如下:

$$IOU = \frac{(A \cap B)}{(A \cup B)} \tag{7}$$

$$L_{\rm IOU} = 1 - IOU \tag{8}$$

但是当两个框不相交或者互相包含时,IOU并不能正确反映预测框与真实框之间的距离和位置关系,IOU值相同时两个框的重合效果也不同,无法快速准确的确认定位框的优化方向。而CIOU损失函数在惩罚项中加入了预测与真值框两个框之间的重叠率、框中心点间距离和目标框的长宽比,在不断训练中使预测框更接近真实框,并且加快网络的收敛速度和回归精度,因此将定位损失替换为CI-OU。CIOU计算定义如下:

$$L_{\text{CIOU}} = L_{\text{IOU}} + \frac{\rho^2(b, b^{\text{gr}})}{c^2} + \alpha v \qquad (9)$$

$$\alpha = \frac{v}{L_{\rm IOU} + v} \tag{10}$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gr}}{h^{gr}} - \arctan \frac{w}{h})^2$$
(11)

其中: α 为权重函数, v 为真实框和预测框长宽比一致 性函数。b 和 bst为预测框与真实框的中心点, wst、w、hst、 h 分别为真实框与预测框的宽和高。ρ 为两中心点之间的欧 氏距离。

2.3.2 Focal Loss

YOLOX 解耦头中判断预测框中是否存在目标的置信度 损失为二元交叉熵损失,其计算定义如下:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=0}^{N} y_i \log(p(y_i))$$
(12)

式中, y_i 为预测框的标签值, p (y_i) 为边界框属于 y 标签的概率。

在 YOLOX 中每个真实目标对应可变 K 值个正样本预测 框,其余特征点生成的预测框均为负样本,而图像中的目标 数量相对所有待预测特征点是少数,因此负样本远多于正样 本,导致置信度损失计算时出现正负不平衡问题。同时在训 练中还存在对一些样本预测置信度值较低的难分样本,这些 样本对模型训练精度的贡献更大,YOLOX 中仅使用二元交 叉熵损失无法对难易样本进行平衡。因此加入 Focal Loss 损 失函数,同时平衡正负和难易样本,计算公式如下:

$$L_{\rm FL} = -\frac{1}{N} \sum_{i=0}^{N} \alpha_i (1-p_i)^{\gamma} y_i \log (p(y_i))$$
(13)

在该损失函数中,加入了参数 α_i,通过设置 α_i的值控 制正负样本对总损失的权重。引入了调制参数 γ 解决难易 分类样本不平衡的问题,通过减少损失中易分类样本权重, 促使模型对难分类的样本在训练中更关注。当模型预测样 本概率较大时,(1-*p_i*)^γ趋近于 0,降低其损失值,当预测 概率较小样本难度较大时,(1-*p_i*)^γ趋近于 1,提高该样本 对回传梯度的贡献。

通过对损失函数的改进增加了模型检测的性能,虽然 一定程度会上导致计算速度下降,增加模型的训练时间, 但损失函数计算量的小幅度增加并不会对检测的推理应用 速度产生影响,符合轻量化设计目标。

3 基于 Jetson Xavier NX 的模型部署

巡逻航拍目标检测任务的完成包括训练阶段和移植部 署应用阶段。训练阶段在 PC 端将大量标注好的数据投入到 改进的目标检测模型中进行训练,通过计算网络预测输出 和真值之间的损失值进行梯度回传更新网络参数完成训练, 得到迭代训练后的检测网络模型。部署阶段是将上述改进 后的模型训练后加速优化移植到机载端,最终在机载端调 用机载摄像头完成目标检测。

由于机载端推理并不需要网络的反向传播,为解决机 载端的算力受限问题,部署阶段利用 TensorRT 对已训练好 的改进后 YOLOX 模型优化加速,部署在 Jetson Xavier NX 上。TensorRT^[19]是一个基于深度学习前向传播架构的高性 能推理优化器,通过去除无用层、将卷积层、BN 层和激活 层合并为 CBR 层等操作对网络结构进行重构,加快模型在 边缘端的推理速度。

详细的部署过程如图 6。首先在机载端配置好航拍检测 所依赖的环境,将 PC 端利用 Pytorch 框架训练生成的 pth 格式模型转换为 onnx 中间格式模型,然后利用 TensorRT 加载解析 onnx 模型优化为 Engine 格式模型,最后通过加载 Engine 模型完成对机载摄像头获取图像的检测推理。



图 6 模型部署步骤

4 实验与结果分析

4.1 数据集情况

VisDrone2019-DET^[20]是一个在中国不同城市不同场景 拍摄的无人机航拍图像数据集,共有不同地点、不同高度 图像10209张,其中训练集、验证集和测试集分别为6471 张、548张和3190张。图像大小囊括现有无人机航拍器的 拍摄分辨率,场景涉及了雨天、雾天、夜晚等不同的天气 状况和光照条件,还包含目标遮挡和视角变化等多种挑战 性的情况,可使训练出的模型泛化性能更强。数据集共标 注了人、汽车、行人、面包车、公共汽车、卡车、摩托车、 自行车、遮阳篷三轮车和三轮车十类目标,满足巡逻任务 对车辆和行人分布情况的感知应用。

实验选取该数据集作为算法训练和测试的数据集。在 该数据集基础上为了提高训练出模型的鲁棒性,训练过程 中采用 Mosaic^[21]和 MixUp^[22]数据增强策略进行随机抽取缩 放旋转拼接,丰富现有数据集增加训练样本。

4.2 实验环境与步骤

实验算法的训练和测试均采用显存为 12 GB 的 GPU 型 号 NVIDIA GeForce RTX 3060 的环境完成。实验系统版本 为 Ubuntu20.04,内核为 Linux,运行环境基于开源深度学 习框架 PyTorch1.13,采用 CUDA11.6 加速计算,实验代

码由 python3.8 编写, YOLOX 模型的训练基于官方提出的 YOLOX-s 算法,改进后的算法基于 MMdetection 框架 完成。

由于 VisDrone2019-DET 数据集以小目标为主,因此在 训练硬件性能允许的情况下,设置图像的输入分辨率大小 为 960×960。设置网络最多训练轮次为 300 次。在训练过 程中,设置初始学习率为 0.001 25,使用随机梯度下降 (SGD)方法,采用带预热策略的余弦学习率调度,权重衰 减为 0.000 5, SGD 动量设为 0.9。在训练的最后 15 轮关闭 数据增强操作,避免训练出的模型脱离实际的巡逻目标检 测应用。将训练好的网络模型作为各算法最终模型进行检 测测试。

实验采用的机载计算边缘端型号为 Nvidia Jetson Xavier NX,训练好的算法模型最终部署到该平台上用于实际应用 推理。算法运行环境基于边缘端的 Jetpack5.1,由 cuda 11.4、python3.8、pytorch2.0和 tensorRT8.5 搭建的深度 学习框架实现算法目标推理检测。

4.3 实验结果与分析

4.3.1 模型评价指标

在目标检测任务中,模型的评估指标包括平均精度 AP (average precision)、参数量、计算量即浮点运算次数 GFlops (giga floating-point operations)和单张图像的推理 速度等, AP 值与精确率 Precision 和召回率 Recall 有关。 mAP 是模型检测所有类别的平均精度。

$$Precision = \frac{TP}{TP + FP}$$
(16)

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$AP = \int_{0}^{1} P(r) dr$$
 (18)

其中: TP 表示模型正确预测的正样本个数。FP 表示 算法模型将负样本错误预测为正样本个数。FN 表示模型将 正样本预测为负样本的个数。为了证明改进算法的有效性, 进行了消融实验。

4.3.2 消融实验

为了评估改进后算法的检测效果,分别设计了加入不同模块后改进算法的消融实验,在 PC 端 VisDrone2019-DET 测试集上对训练好的模型进行测试推理评估,测试环 境为模型训练时使用的运行环境。选取 IOU 阈值为 0.5 分 别计算各类别和模型的 mAP 值,模型对各类别检测情况如 表 2 所示。

在测试集上对各类别的检测结果表明,将特征提取网络替换为 Mobilenetv2 轻量化网络,替换后模型的 mAP 降低了 0.2%。在加入 Focal Loss 损失函数和 CIOU Loss 损失函数进行训练后,模型 mAP 增长了 0.3% 和 0.4%,与 YOLOX 算法的检测精度相当,改进后的算法提高了人、自行车等小型目标的检测精度。下面分别计算每个训练好模型的参数量、计算量和 PC 端单张图像推理时间,各算法的模型性能如表 3 所示。

表 2 以进制后候望住丨关目协工的 mAr						
类别	YOLOX	+ Mobilenetv2	+Focal Loss	+CIOU Loss		
行人	0.391	0.369	0.37	0.372		
人	0.200	0.231	0.229	0.235		
自行车	0.121	0.124	0.122	0.128		
汽车	0.772	0.772	0.775	0.776		
厢式货车	0.361	0.39	0.4	0.397		
卡车	0.369	0.362	0.365	0.363		
三轮车	0.214	0.172	0.175	0.175		
带篷三轮车	0.160	0.159	0.152	0.162		
大巴车	0.541	0.529	0.539	0.532		
摩托车	0.368	0.375	0.379	0.377		
mAP	0.350	0.348	0.351	0.352		

お洪哉巨揖到左上米日長上的 … A D

表 3 各算法模型性能

模型	参数量	计算量 FLOPS	推理时间
YOLOX	8.94 M	26.78 G	14.73 ms
Mobilenetv2-YOLOX	3.91 M	12.73 G	8.63 ms
Mobilenetv2-YOLOX-Focal Loss	3.91 M	12.73 G	8.63 ms
Mobilenetv2-YOLOX-Focal Loss-CIOU	3.91 M	12.73 G	8.63 ms

由实验结果可知,将 YOLOX 算法的骨干网络换为 Mobilenetv2 后,网络模型参数量由 8.94 M 降低为 3.914 M,降 低了 56.2%。计算量由 26.78 G FLOPS 降低为 12.733 G FLOPS,单张图片的推理时间降低了 41.4%,后续损失函 数的改进对参数量和计算量并无明显影响。与基准算法 YOLOX 相比改进后的算法在保持检测精度的情况下大大提 升了检测速度。最后将改进的 Mobilenetv2-YOLOX-Focal Loss-CIOU 算法移植到 Nvidia Jetson Xavier NX 巡逻无人机 机载边缘端中运行测试,推理速度达到 22 FPS,符合应用 要求。

4.3.3 实验结果

将航拍图像输入改进后的网络上进行检测推理,得到 算法检测到的目标可视化结果。可以看出算法在(a)、 (b)、(c)、(d)四幅图像所示的密集、稀疏等不同场景、傍 晚、黑夜等不同光照条件和模糊、旋转等多种图像中都表 现出不错的效果,检测结果如图7所示。



图 7 改进后算法的检测效果

为更好的验证部署后的算法在实际巡逻中的应用效果, 采用 DJI M300 RTK 无人机拍摄了国内某城市的巡逻图像, 将其送入部署在 Jetson Xavier NX 机载端的模型中进行检测 推理,各关注目标均被明显标出,得到实际目标检测效果 如图 8 所示。



图 8 部署算法巡逻应用情况

5 结束语

通过对无线传输信道受限情况下城市巡逻中无人机航 拍目标检测的轻量化技术进行研究,提出了改进的 YOLOX 轻量化目标检测算法,并在机载边缘端进行了实验验证。 提出的算法在 YOLOX 基础上采用 Mobilenetv2 特征提取网 络,降低了网络的参数量和计算量;其次将置信度损失函 数换为 Focal Loss 损失函数,更好的平衡正负样本和难易 样本;然后对定位损失进行改进,在 IOU 交并比基础上加 入目标框长宽比惩罚项,将其替换为 CIOU 损失函数,提 升网络的回归精度。通过改进前后模型参数、运行速度和 精度对比实验,证明设计的轻量化的航拍目标检测网络在 同等检测精度下模型更小、检测速度更快,更适合部署在 机载边缘终端。为验证算法的实际应用性能,将轻量化模 型移植到部署在小型无人机上的 Jetson Xavier NX 中进行测 试,检测速度达到 22FPS,满足巡逻任务的应用需求,对 机载目标检测算法的轻量化和工程实现有参考意义。

参考文献:

- [1] 杨浩然,张雨晗. 基于计算机视觉的无人机目标检测算法综述 [J]. 电子测试, 2022, 36 (4): 44-45.
- [2] 江 波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标 检测研究综述 [J]. 航空学报, 2021, 42 (4): 137-151.
- [3] 王荣浩,高星宇,向峥嵘. 有人/无人机协同系统及关键技术 综述[J]. 兵器装备工程学报,2023,44 (8):72-80.
- [4] SEVO I, AVRAMOVIC A. Convolutional Neural Network Based Automatic Object Detection on Aerial Images [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13 (5): 740 -744.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
- [6] GIRSHICK R. Fast R-CNN [C] //IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
- [7] WEI L, DRAGOMIR A, DUMITRU E. SSD: Single Shot MultiBox Detector [C] // European con-ference on computer

vision, 2016: 21-37.

- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detectionp [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
- [9] LIN TY, 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.
- [10] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C] // IEEE/CVF International Conference on Computer Vision Workshops, 2021: 2778 - 2788.
- [11] 崔令飞,郭永红,修全发,等. 基于国产嵌入式智能计算平 台的无人机检测方法 [J]. 兵工学报,2022,43 (s1):146 -154.
- [12] 黄海生,饶雪峰.面向无人机航拍场景的轻量化目标检测.计 算机系统应用,2022,31 (12):159-168.
- [13] SANDLER M, HOWARD A, ZHU M, et al. Inverted Residuals and Linear bottlenecks: mobile networks for classification, detection and segmentation [EB/OL]. [2018 - 01 - 13]. https://doi.org/ 10.48550/arXiv.1801.04381.
- [14] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [EB/OL]. [2021 - 07 - 18]. https://doi.org/ 10.48550/arXiv.2107.08430.
- [15] SONG G, LIU Y, WANG X. Revisiting the sibling head in object detector [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11560 – 11569.
- [16] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1800 - 1807.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [18] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network [C] // Proceedings of the 24th ACM international conference on Multimedia, 2016: 516-520.
- [19] JEONG E, KIM J, TAN S, et al. Deep learning inference parallelization on heterogeneous processors with tensorRT [J]. IEEE Embedded Systems Letters, 2022, 14 (1): 15 -18.
- [20] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results
 [C] // IEEE/CVF International Conference on Computer Vision Workshop, 2019: 213 226.
- [21] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2020 - 04 - 23]. https:// doi.org/10.48550/arXiv.2004.10934.
- [22] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization [EB/OL]. [2017 - 10 - 25]. https://doi.org/ 10.48550/arXiv.1710.09412.