

DOI: 10.13382/j.jemi.B2104886

基于特征增强 YOLOv4 的无人机检测算法研究*

史雨馨 朱继杰 凌志刚

(湖南大学电气与信息工程学院 长沙 410082)

摘要:现有基于深度学习的目标检测方法在面对空中消费级无人机时,存在鲁棒性差、准确率不足等问题。对此,提出一种基于特征增强的 YOLOv4 目标检测方法—FEM-YOLOv4。首先,针对无人机低、小、慢等特点,改进骨干网络,降低下采样倍数,充分利用包含细粒度信息的浅层特征;其次,加入特征增强模块(feature enhancement module),通过使用不同空洞率的多分支卷积层结构,综合不同深度的语义信息和空间信息,增强小尺度无人机的细节语义特征;另外,利用多尺度融合的特征金字塔结构,突出特征图包含的细节信息和语义信息,提升模型对无人机目标的预测能力;最后,采用 K-means++ 算法对无人机目标候选框的尺寸进行聚类分析。与 6 种目标检算法进行对比,实验结果表明,FEM-YOLOv4 算法的 mAP 和 Recall 分别达到 89.48%、97.4%,优于其他算法,且平均检测速度为 0.042 s。

关键词:卷积神经网络;深度学习;YOLOv4;无人机检测;特征增强模块

中图分类号: TP391.41; TP183

文献标识码: A

国家标准学科分类代码: 510.40

Research on UAV detection method based on feature enhanced YOLOv4 algorithm

Shi Yuxin Zhu Jijie Ling Zhigang

(College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: Consumer-level UAVs have small scale, low fly speed and height, existing deep learning methods hardly achieve high detection accuracy and good robustness on detecting UAVs. In order to address this problem, this paper develops an improved YOLOv4 algorithm with feature enhanced module named as FEM-YOLOv4 for UAVs detection. Firstly, according to the characteristics of UAVs, this paper reduces the subsampling multiple of CSPDarkNet to improve the backbone network and make full use of shallow features containing detailed information. Secondly, this paper introduces the feature enhancement module to replace the SPP module. The feature enhancement module includes multiple branches and dilated convolution, and it obtains different levels of semantic information, which is beneficial to enhance the detailed semantic features and the detection capabilities of the network. Thirdly, delete the PAN module to improve the feature pyramid, and compress the depth of each detection layer to highlight the detailed and semantic information of the feature maps. Finally, the anchor box is initialized by the K-means++ algorithm to make the model more suitable for predicting the UAV targets. Compared with the six target detection algorithms, the experimental results show that the mAP and Recall of FEM-YOLOv4 algorithm reach 89.48% and 97.4% respectively, which are superior to other algorithms, and the average detection speed is 0.042 s.

Keywords: convolutional neural network; deep learning; YOLOv4; UAV detection; feature enhancement module

0 引言

近年来,消费级无人机因其功能强、成本低、易操作等特点,在遥感、交通等诸多领域得到广泛应用^[1-2],有效

地促进了各类行业的发展,同时方便了人们的生活和学习。然而,消费级无人机的普及也引发了诸多问题,如利用无人机犯罪或攻击、无人机黑飞影响机场航班安全等事件屡见不鲜。因此,对无人机进行有效的检测与反制至关重要,其中空中无人机的高效检测成为了反制无人

机的前提,成为各国学者研究的热点^[3]。

目前国内外无人机检测方法主要包括雷达检测^[4]、音频检测^[5]、光视觉检测^[6]等。雷达检测技术通过接收无人机目标反射的雷达信号进行无人机识别与定位,雷达信号虽然能够实现较远距离的检测与定位,但是消费级无人机具有“低、小、慢”等特点,其多普勒效应不明显,导致雷达检测精度受限。音频检测技术通过无人机飞行时的声音特征进行无人机识别,但是音频检测受环境噪声干扰较大,检测距离较短,精度有限。光视觉探测技术主要通过可见光或红外摄像头采集视频信息来识别无人机,与其他探测技术相比,光视觉探测技术具有探测结果直观、精度高、适用场所广等特点,逐步成为了反无人机技术不可或缺的一部分。

近几年,随着深度学习的深入研究,基于深度学习的无人机检测方法成为热门的研究方向。Nalamati 等^[7]将 ResNet-101 模型^[8]和 Faster R-CNN^[9]网络进行融合,解决了增加网络深度带来的退化问题,实现无人机目标检测;程怀俊等^[10]在两阶段目标检测框架 Faster R-CNN 的基础上,提出了融合感受野模块和 Darknet 网络的特征提取器,增大感受野(receptive field)和加强网络的特征提取能力,从而有效地提升无人机检测的性能;Lin 等^[11]提出特征金字塔网络(feature pyramid networks, FPN)与 RetinaNet^[12]相融合的无人机检测算法,删除特征金字塔较高的两层,同时添加一个较低层以适应较小目标,并且对数据集重新聚类、选取锚框;王若霄等^[13]基于 SSD 算法提出了 TSSD 无人机检测,该算法削减了骨干层 VGG16^[14]的卷积数目以及通道数目,并且在 Conv4-3 后加入一个特征增强模块,使网络更加轻量化和实时的同时,增强浅层检测小目标的能力。

尽管无人机目标检测研究取得较好的进展,但消费级无人机具有“低、小、慢”等特点,现有光电检测方法存在面对消费级无人机误检率较高、同时存在实时性较差、不利于工程应用等挑战。

为此本文提出了一种面向空中消费级无人机目标检测的特征增强 YOLOv4 (feature enhancement YOLOv4, FEM-YOLOv4)算法。首先对原始 YOLOv4 中特征提取网络和多尺度融合结构进行改进,建立更利于无人机目标特征提取的骨干层以及结构简单、更利于细节特征表达的多尺度融合结构;同时,加入特征增强模块,通过使用不同空洞率的多通道卷积层结构,综合不同深度的语义信息和空间信息,增强特征图的细节语义信息和网络的特征感受能力,提升网络对弱小无人机的检测能力。在训练过程中,使用 K-means++ 算法对无人机数据集的目标候选框尺寸进行聚类分析,采用分析得到的先验框尺寸进行模型训练,以提高模型对消费级无人机目标尺度的适应性。

1 YOLOv4 简介

基于深度学习的目标检测算法一般可以分为两类^[15]:一类是基于候选区域的两阶段算法,如 R-CNN^[16]、Fast R-CNN^[17]等,这类算法准确性更高但实时性较差;另一类是基于端到端学习的单阶段算法,如 SSD^[18]、YOLO^[19-21]系列等,这类算法以速度著称。两阶段目标检测算法以 Faster-R-CNN 为代表,特征图通过特征金字塔进行多尺度特征融合,然后经由候选区域生成网络、分类网络和回归网络,完成目标分类和预测框回归,最后通过非极大值抑制算法完成目标检测。这类算法的准确性相对较高,但速度较慢。单阶段目标检测算法以 YOLO 系列为代表,不使用候选区域生成网络,直接由特征图出发,采用 regression(回归)的方法实现物体位置与分类的预测。这类算法结构更加简洁,检测速度更快,实时性更好。目前,YOLOv4 算法在公共数据集上达到精度与速度最优平衡^[22]。

YOLOv4 网络结构如图 1 所示,其在 YOLOv3 框架的基础上,融合了数据增强、感受野模块、特征金字塔等优化方法。YOLOv4 网络骨干层采用 CSPDarknet53,该结构将跨阶段对等网络(cross stage partial network, CSPNet)^[23]和 Darknet53 网络进行结合,有效解决了网络在推理过程中因梯度信息重复导致计算量过高的问题,能使网络轻量化的同时保证准确率。在 Neck 部分,YOLOv4 添加了空间金字塔池化模块(spatial pyramid pooling, SPP)^[24]和路径聚合模块(path aggregation network, PANet)^[25]。SPP 结构通过不同尺度的最大池化操作,能够增大特征图的感受野,在不影响网络速度的同时,有效地分离数据中的显著特征。PANet 与 FPN 结合,使网络在自上向下传递强语义特征的同时,也能够自下向上传递强定位特征,有助于准确地保存空间信息、正确地定位像素点,进一步提升网络的检测性能。

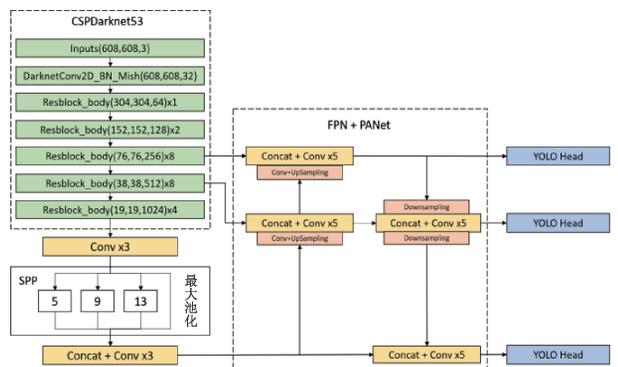


图 1 YOLOv4 结构

Fig. 1 Structure of the YOLOv4

2 面向无人机的 FEM-YOLOv4 网络结构

为了提升深度网络在背景复杂、目标尺度小等情况下的检测性能,在 YOLOv4 网络的基础上,提出了基于特征增强的空中无人机目标检测网络 FEM-YOLOv4,整体网络结构如图 2 所示。针对视频图像中消费级无人机目标较小等特点,首先对特征提取网络进行改进,增强目标的细节特征;并添加特征增强模块,提高最小尺度特征图的特征表达能力;然后对多尺度特征融合结构进行改进,充分利用细粒度特征、增强目标的语义特征;最后采用 K-means++ 算法对先验框重新聚类,从而实现对空中无人机的检测。

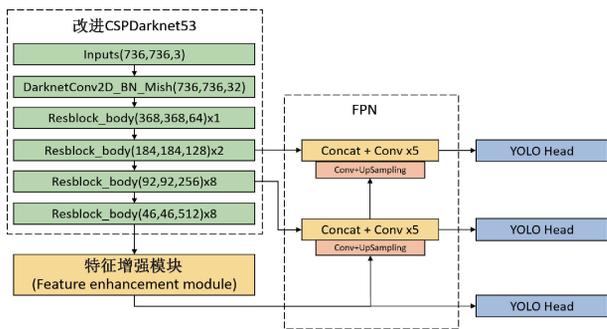


图 2 FEM-YOLOv4 结构

Fig. 2 Structure of the FEM-YOLOv4

2.1 融合浅层特征的骨干网络

根据卷积神经网络的特点,网络层数越深,模型提取到的目标语义信息越丰富。然而随着网络的加深,会导致计算量的增加和细节信息消失,不利于模型的推理速度和对小目标的检测。因此,对 CSPDarknet53 骨干网络进行改进,简化 YOLOv4 网络的特征提取过程。

为适应消费级无人机目标的小尺度特性,将网络输入扩大为 736×736 ,以保留无人机目标更多的信息。为避免高倍下采样导致无人机信息丢失,去掉 CSPDarkNet53 中过深的卷积层,从而将原有骨干网络的 32 倍下采样调整为 16 倍下采样,同时卷积层数降低至 85 个。另外,前移用于预测的特征层,充分利用包含更多细粒度信息的浅层特征。改进后网络结构如图 2 所示,修改后的 CSPDarkNet53 包含 4 个大残差模块、下采样倍数更低、特征输出层更靠前,使得小目标的细粒度信息得以保全。对分辨率为 736×736 pixel 的输入图像,骨干层输出特征图尺寸为 46×46 。

2.2 多分支卷积特征增强模块

YOLOv4 网络中 SPP^[24] 模块接收骨干层的输出特征图,通过不同尺度的最大池化和级联操作,增大了特征图的感受野,同时实现特征图的局部特征和全局特征相融

合,有利于待检测图像中目标尺度差异较大的情况。然而对于消费级弱小无人机目标,最大池化操作可能会导致特征图部分信息丢失;并且单纯地增大感受野,会导致特征图中包含更多的背景、冗余信息。因此,为了保留消费级无人机目标的细节信息,同时减少冗余信息,如图 2 所示,将 SPP 模块删除。目标检测网络 RFBNet^[26] 提出,卷积神经与人类视觉相似,小尺度目标越靠近感受野中心,被检测到的概率越大。受该思想启发,本文针对无人机目标特点,对 RFBNet 中的感受野模块 (receptive fields block, RFB) 进行改进,提出了特征增强模块 (feature enhancement module)。

特征增强模块结构如图 3 所示,包含 3 路分支和 1 个旁路剪枝,每个分支由不同大小的卷积核和空洞卷积组成,最终各通道级联,与旁路剪枝构成输出。这种结构能够在不增加网络参数的情况下,扩大卷积层的感受野,同时使处于网络不同深度的特征图相融合,达到丰富骨干网络输出的特征信息,提升网络的信息表达能力的目的。如图 3 所示,该模块通过核大小为 1×1 的卷积将第 85 层的通道数进行分配,使 3 个分支的通道数均为 128;每个分支由不同数量的卷积层构成,特征图经过不同深度的卷积操作后,能够获得不同层次的语义信息;各分支的输出特征图级联后,能够使网络深层的语义信息和空间信息更加丰富,有利于增强语义特征和网络对无人机目标的检测能力。除此之外,模块采用了空洞卷积,空洞卷积能够在不降低特征图分辨率的情况下,增大特征图的感受野;为了避免感受野过大,特征图包含过多的冗余和背景信息,本文采用空洞率分别为 1、2、3 的空洞卷积;随着通道卷积层的加深,空洞卷积的空洞率也越大,这种结构能使同一特征图产生不同的感受野,有利于网络捕获不同的上下文信息,增强网络的特征感受和特征提取能力,使网络能够响应不同尺度的无人机目标。

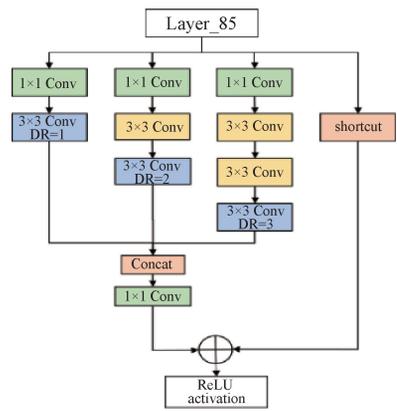


图 3 特征增强模块结构

Fig. 3 Feature enhancement module architecture

2.3 融合多尺度特征的金字塔网络

YOLOv4 网络利用特征金字塔网络 FPN 融合不同层次的特征信息,使浅层网络的细节信息和深层网络的语义信息得到充分利用,在此基础上,对网络结构进行进一步修改。

首先将特征增强模块的输出作为金字塔的顶层特征和网络的预测层 1,该预测层具有较丰富的语义信息。其次将金字塔的顶层特征通过通道降维、上采样操作,与第 54 层的特征图级联,并将输出特征作为预测层 2,增强网络对不同尺度无人机目标的适应能力。此外,将第 23 层的输出作为金字塔的底层特征图,与第 93 层特征图级联,并将输出结果作为预测层 3,提升网络包含的细粒度信息。改进后特征金字塔结构如图 2 所示,在 736×736 的输入下,金字塔输出分别为 46×46 、 92×92 、 184×184 ;金字塔的特征图均前移一个大残差模块,使浅层特征的细粒度信息得到利用。除此之外,逐步压缩每个检测层的深度,去除路径聚合模块,突出特征图包含的细节信息和语义信息。

2.4 基于聚类的锚框选择

作为基于锚框的目标检测网络,YOLOv4 延续了先验框思想,采用 K-means 算法对 COCO 数据集进行聚类分析,使得每个特征图具有 3 种不同尺度的先验框。先验框的尺度会影响模型的收敛速度以及检测的准确性,因此对先验框的分析与选择至关重要。本文使用的无人机数据集中无人机目标尺寸小,而 COCO 数据集包含的目标尺寸多样,所以 YOLOv4 网络初始先验框的维度不适合消费级无人机目标检测任务,需要分析得到新的先验框尺寸。

因此,针对空中消费级无人机检测场景,为使 YOLOv4 网络适应数据集的目标大小,达到更好的训练与检测效果,需要重新对自制无人机数据集进行聚类分析。K-means++ 算法^[27]通过改进初始点的选取方法,提升分类精度,故采用 K-means++ 算法对数据集进行聚类分析。VOC 数据集和本次无人机数据集的先验框尺度对比如表 1 所示,其中大尺度的先验框对应具有较大感受野的小特征图,小尺度的先验框对应大特征图。

表 1 目标候选框尺度对比

Table 1 Scale comparison of target candidate frame

数据集	特征图大小	先验框
VOC	19×19	(142, 110) (192, 243) (459, 401)
	38×38	(36, 75) (76, 55) (72, 146)
	76×76	(12, 16) (19, 36) (40, 28)
无人机	46×46	(34, 47) (44, 54) (67, 97)
	92×92	(23, 29) (20, 35) (29, 38)
	184×184	(1, 15) (14, 22) (18, 26)

3 实验结果与对比分析

3.1 数据集介绍

为了验证所提方法在实际场景下对消费级无人机检测的有效性,本文采用自制无人机数据集 UAV-dataset 和 CVPR 2020 Anti-UAV Workshop & Challenge 中的开源数据集 Anti-UAV dataset,分别进行实验和验证。

本文利用某公司无人机检测终端获取的图像构建数据集 UAV-dataset,其中包括训练集 7 920 张,测试集 2 100 张,无人机数据集图像分辨率为 $1 920 \times 1 080$,无人机目标尺度分布在 26×20 至 140×110 区间内。数据集部分样本如图 4 所示,主要包含室外空旷环境(图 4(a)),和室外复杂环境(图 4(b))下拍摄的无人机图片。采用相关软件对数据集中的图像进行标注,生成与图片相对应的注释文件。



(a) 室外空旷环境 (b) 室外复杂环境
(a) Outdoor open environment (b) Outdoor complex environment

图 4 UAV-dataset 部分样本图像

Fig. 4 Partial sample image of UAV-dataset

开源数据集 Anti-UAV dataset 包括 160 段全高清双模态(可见光 RGB+近红外 NIR)视频(100 段测试视频,60 段训练视频),涵盖多种场景下多尺度无人机数据,为实时追踪算法模型提供训练和测试信息,标注信息中包括 bounding box 以及表示当前帧是否存在目标的标志位。为验证本文改进 YOLOv4 无人机检测模型性能,本文选取 30 段分辨率为 $1 920 \times 1 080$ 的可见光 RGB 视频,无人机目标尺度分布在 $35 \times 20 \sim 110 \times 80$ 区间内,将视频每一帧的标注信息进行转换,获得适合 YOLOv4 模型训练的数据格式。训练集与验证集数据比例为 5 : 1,部分样本如图 5 所示,主要包括小尺度无人机目标(图 5(a))、复杂背景中等尺度无人机目标(图 5(b))。

3.2 实验运行环境与参数设置

本文采用 Darknet 深度学习框架,操作系统为 Windows 10,GPU 硬件采用 NVIDIA GeForce RTX 2080 TI 12 GB,CUDA Toolkit 10.1 版本,cuDNN 神经网络加速库 7.6.4 版本。使用 YOLOv4 网络和改进后的网络分别进行训练,在训练阶段,网络输入图片的尺寸被调整为 736×736 ,批量大小为 32,衰减系数 0.000 5;总迭代次数为 20 000 次,初始阶段选择 0.001 的学习率,当迭代次数



(a) 小尺度无人机目标 (b) 复杂背景中等尺度无人机目标
(a) Small scale drone targets (b) Medium scale drone target with complex background

图 5 Anti-UAV dataset 部分样本图像

Fig. 5 Partial sample image of Anti-UAV dataset

分别为 16 000 和 18 000 时,学习率衰减至 0.000 1 和 0.000 01。

3.3 评价指标

为准确且有效地对算法的性能进行评估,目标检测网络通常使用被检目标的召回率 (Recall) 和准确率 (Precision) 作为评价指标,其定义如式(1)、(2)所示。

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

其中,TP 和 FP 分别表示模型将正样本检测正确与错误的数量, FN 表示误检的数量。召回率表示正确识别无人机的概率,召回率越高表示漏检越少; mAP 表示已识别无人机的平均精度均值, mAP 越高表示虚警越少; 检测速度表示模型检测一张图像的平均时间。因此,本文采用平均检测精度 (mAP)、召回率、检测速度作为评价指标。

3.4 实验结果对比分析

1) 算法模块可行性分析

为了评估所提改进策略在无人机目标检测过程中的有效性,在相同实验条件下,使用不同改进策略,进行多组对比实验,结果如表 2 所示。其中,√表示在该组实验中使用了对应策略, backbone 表示改进特征提取网络, FEM 表示特征增强模块, FPN 表示改进特征金字塔网络, Anchor 表示 K-means++ 算法重新聚类。

表 2 使用不同改进策略检测结果

Table 2 Test results using different improvement strategies

实验	Backbone	FEM	FPN	Anchor	mAP/%	Recall/%
1					87.33	96.1
2	√				88.63	96.8
3	√			√	89.06	97.1
4	√		√	√	89.41	97.0
5	√	√	√	√	89.48	97.4

使用本文的改进策略后,平均精度均值和召回率获得了不同程度的提高。使用改进的 CSPDarknet53 骨干网络后, mAP 提升了 1.3%, Recall 提升了 0.7%, 可见对于小尺度无人机目标检测来说,特征提取网络的下采样

倍数和特征图的细粒度信息十分重要。

使用本文改进策略后, mAP 值为 89.48%, 较原网络提升了 2.15%; Recall 值为 97.4%, 较原网络提升了 1.3%。实验证明,改进策略对本文中的无人机目标检测场景有效。

2) UAV-dataset 实验结果对比

分别选取 UAV-dataset 数据集中空旷环境下小尺度无人机目标、复杂环境下小尺度无人机和复杂环境下多无人机目标 3 种情况进行对比分析, YOLOv4 算法与本文改进 YOLOv4 算法检测结果分别如图 6、7 所示。

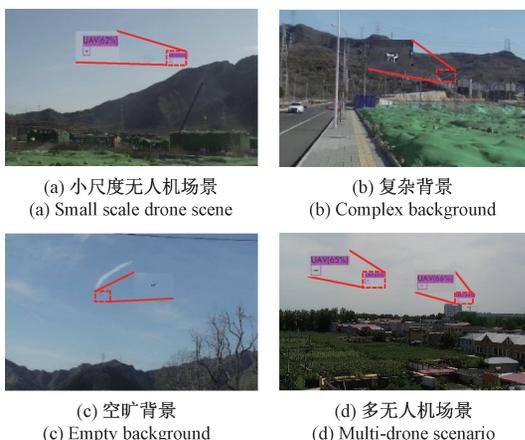


图 6 YOLOv4 算法检测结果

Fig. 6 YOLOv4 algorithm detection results

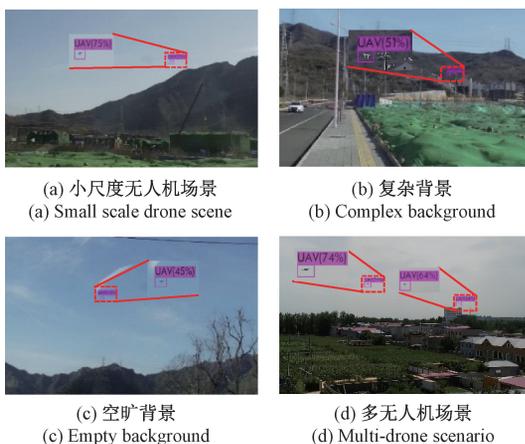


图 7 FEM-YOLOv4 算法检测结果

Fig. 7 FEM-YOLOv4 algorithm detection results

由检测结果可以看出,针对自制无人机数据集 UAV-dataset, FEM-YOLOv4 算法比原 YOLOv4 算法效果更好。YOLOv4 算法在复杂背景(图 6(b))和空旷背景场景(图 6(c))中均未检出无人机目标,而 YOLOv4 算法均能完成无人机目标检测,且置信度分别为 51%、45% (图 7(b)、(c)); 在小尺度无人机场景(图 6(a)、图 7(a))中, FEM-YOLOv4 算法检出的无人机置信度为 75%, 较

YOLOv4 网络高 13%; 在多无人场景中 (图 6(d)、图 7(d)), FEM-YOLOv4 算法检出的无人机置信度分别为 74%、74%。整体高于 YOLOv4 网络的 65%、66%。

为了更全面地评估改进后算法的性能, 采用改进 YOLOv4 网络与 Faster R-CNN、FCOS^[28]、Free Anchor^[29]、ATSS^[30]、Libra R-CNN^[31] 和原始 YOLOv4 网络进行对比实验, 网络输入图像大小均为 736×736。以 mAP、Recall、检测速度作为算法的评价指标, 实验结果对比如表 3 所示。

表 3 UAV-dataset 检测结果

Table 3 The results of detection for UAV-dataset

算法	mAP/%	Recall/%	时间/s
Faster R-CNN	87.48	91.6	0.078
FCOS	85.89	92.5	0.049
Free Anchor	86.59	95.8	0.037
ATSS	89.28	96.2	0.052
Libra R-CNN	88.20	96.5	0.067
YOLOv4	87.33	96.1	0.040
YOLOv5	89.40	92.7	0.015
Ours	89.48	97.4	0.042

注: 加粗字体为每行最优值。

可以看出, 两阶段网络 Faster R-CNN 由于结构加深, 导致检测时间更长。ATSS 网络采用自适应样本训练, mAP 能够达到 89.28%, 但检测速度相对较慢。Free Anchor 网络通过改进损失函数, 能够使物体灵活地匹配锚框, 优化了目标检测的过程, 使得检测速度更快。YOLOv5 网络采用 Mosaic 数据增强机制和 Focus 结构, 使得网络平均检测精度和速度得到提升。

改进网络在未添加特征增强模块时, 针对无人机数据集的目标尺度特点, 基于 YOLOv4 算法进行网络结构调整, 通过采用更加浅层的特征图、简化网络结构、保留了更多细粒度信息, 从而使改进网络 mAP 达到 89.41%, 较原网络提高 2.08%, 且速度稍快。添加特征增强模块后, 通过多通道卷积层, 增强了网络的细节语义信息, 使得本文改进算法在牺牲较少速度的情况下, mAP 与 Recall 均达到最优, 较原网络分别提升 2.15% 和 1.3%。

3) Anti-UAV dataset 实验对比

为验证改进 YOLOv4 模型的适应性和可迁移性, 本文同时在开源数据集 Anti-UAV dataset 上进行验证实验, 针对空旷环境下小尺度无人机目标 (图 8(a)、图 9(a)) 和复杂环境中中等尺度无人机 (图 8(b)、图 9(b)) 两种场景进行对比分析, YOLOv4 网络与本文 FEM-YOLOv4 算法检测结果如图 7 所示, 对比结果如表 4 所示。

由图 8、9 可以看出, 本文改进算法和原 YOLOv4 网络均能实现无人机目标检测, 但对于相同目标的检测, 本文算法比原 YOLOv4 算法的置信度更高。在小目标场景下, FEM-YOLOv4 算法对无人机目标预测的置信度为

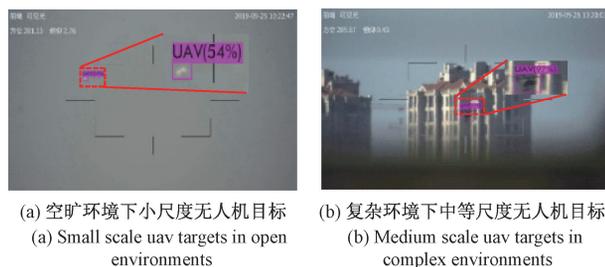


图 8 YOLOv4 算法检测结果

Fig. 8 YOLOv4 algorithm detection results

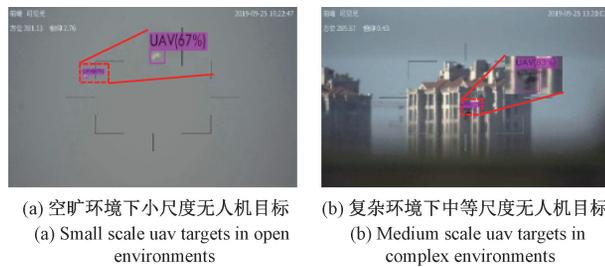


图 9 FEM-YOLOv4 算法检测结果

Fig. 9 FEM-YOLOv4 algorithm detection results

67% (图 9(a)), 比 YOLOv4 算法高 13% (图 8(a)); 在背景复杂、中等尺度目标场景下, 改进 YOLOv4 算法对无人机目标预测的置信度为 97% (图 9(b)), 比 YOLOv4 算法高 14% (图 8(b))。

表 4 Anti-UAV dataset 检测结果对比

Table 4 The results of detection for Anti-UAV dataset

算法	mAP/%	Recall/%	时间/s	模型大小/MB
YOLOv4	93.94	95.2	0.041	244
Free Anchor	92.87	94.1	0.036	290.1
Faster R-CNN	94.03	92.3	0.080	330.2
FCOS	89.86	91.4	0.047	256.1
ATSS	94.17	94.8	0.055	256.3
Libra R-CNN	93.88	93.7	0.064	292.4
YOLOv5	94.73	91.6	0.018	167.8
Ours	94.59	95.5	0.042	77.3

由表 4 可以看出, 在对开源数据集 Anti-UAV dataset 的实验中, 本文 FEM-YOLOv4 算法平均准确率均值为 94.59%, 比原 YOLOv4 网络高 0.65%, 比 YOLOv5 网络低 0.14%; 在召回率方面, FEM-YOLOv4 算法比原 YOLOv4 网络高 0.4%, 比 YOLOv5 网络高 3.9%; 在检测速度方面, YOLOv5 网络达到 0.018 s, FEM-YOLOv4 算法和 YOLOv4 算法效果无明显差距; FEM-YOLOv4 算法在经过简化网络结构后, 模型大小仅为 77.3 MB, 更有利于模型部署。

结合对自制 UAV-dataset 数据集和开源 Anti-UAV dataset 数据集进行的实验结果可以看出, 本文的 FEM-YOLOv4 算法在维持原有检测速度的基础上, 能够一定程度地提高平均准确率均值和召回率; 主要是因为本文

方法融合了网络的浅层特征,在保留细粒度信息的同时利用特征增强模块,丰富细节语义特征信息,提升网络的无人机检测精度;除此之外,本文进行网络结构简化,使得网络在处理大特征图的同时降低检测速度,满足应用要求。由此可见,基于特征增强的 YOLOv4 改进网络模型对无人机目标检测效果更好,能够较好地完成无人机目标检测任务。

4 结 论

本文针对消费级无人机目标特点,提出了一种面向空中消费级无人机目标检测的特征增强 YOLOv4 (FEM-YOLOv4) 算法。首先针对图像中消费级无人机目标尺度小,特征不明显等特点,改进、压缩特征提取网络,避免高倍下采样导致的信息丢失,保留更多细节特征。然后加入特征增强模块,通过利用不同深度和不同感受野的卷积层,提高特征提取能力,增强网络的细节语义信息,以此提升预测层对较小无人机目标的检测能力。其次压缩并改进网络的特征金字塔层,突出目标的细节信息和语义信息。最后通过采用 K-means++ 方法调整先验框维度,使网络更加适应无人机目标的尺度特点。在自制无人机数据集和 Anti-UAV 数据集上进行实验,结果表明,本文提出的 FEM-YOLOv4 无人机检测算法在两组真实无人机图像数据集上均表现出良好的检测性能,并在检测精度上有一定的优势。在后续的研究中,可以对如何轻量化网络进行进一步探索,促进深度学习方法在消费级无人机目标检测中的应用。

参考文献

- [1] KE S, DING X, YANG D, et al. Saliency-based artificial object detection for satellite images [C]. Visual Information Processing and Communication VI. International Society for Optics and Photonics, 2015, 9410: 94100J.
- [2] LI W, LI H, WU Q, et al. Simultaneously detecting and counting dense vehicles from drone images [J]. IEEE Transactions on Industrial Electronics, 2019, 66(12): 9651-9662.
- [3] SHI X, YANG C, XIE W, et al. Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges [J]. IEEE Communications Magazine, 2018, 56(4): 68-74.
- [4] LIU Y, WAN X, TANG H, et al. Digital television based passive bistatic radar system for drone detection [C]. 2017 IEEE Radar Conference (RadarConf). IEEE, 2017: 1493-1497.
- [5] MEZEI J, FIASKA V, MOLNÁR A. Drone sound

- detection [C]. 2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI). IEEE, 2015: 333-338.
- [6] LEE D, LA W G, KIM H. Drone detection and identification system using artificial intelligence [C]. 2018 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2018: 1131-1133.
- [7] 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.
- [8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [10] 程怀俊,陈鹏,陈杰,等. 基于视觉感受野的无人机目标检测 [J]. 现代计算机, 2020(15): 108-111, 116.
CHENG H J, CHEN P, CHEN J, et al. UAV object detection based on visual perception field [J]. Modern Computer, 2020(15): 108-111, 116.
- [11] 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.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [13] 王若霄,徐智勇,张建林. 基于 SSD 的实时轻量级无人机检测算法 [J]. 半导体光电, 2020, 41(2): 296-300.
WANG R X, XU ZH Y, ZHANG J L. Real-time lightweight UAV detection method based on SSD algorithm [J]. Semiconductor Optoelectronics, 2020, 41(2): 296-300.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [15] 刘颖,刘红燕,范九伦,等. 基于深度学习的小目标检测研究与应用综述 [J]. 电子学报, 2020, 48(3): 590-601.
LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep

- learning[J]. Acta Electronica Sinica, 2019, 48 (3): 590-601.
- [16] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [17] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [20] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [23] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.
- [24] HUANG Z, WANG J, FU X, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection [J]. Information Sciences, 2020, 522: 241-258.
- [25] WANG K, LIEW J H, ZOU Y, et al. Panet: Few-shot image semantic segmentation with prototype alignment[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9197-9206.
- [26] LIU S, HUANG D. Receptive field block net for accurate and fast object detection [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385-400.
- [27] ARTHUR D, VASSILVITSKII S. K-means ++: The advantages of careful seeding[R]. Stanford, 2006.
- [28] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9627-9636.
- [29] ZHANG X, WAN F, LIU C, et al. Freeanchor: Learning to match anchors for visual object detection[J]. arXiv preprint arXiv:1909.02466, 2019.
- [30] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759-9768.
- [31] PANG J, CHEN K, SHI J, et al. Libra R-CNN: Towards balanced learning for object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 821-830.

作者简介



史雨馨, 2019 年于湖南大学获得学士学位, 现为湖南大学硕士研究生, 主要研究方向为机器视觉与图像处理。

E-mail: syx_hnu@163.com

Shi Yuxin received her B. Sc. degree from Hunan University in 2019. She is now a M. Sc. candidate at Hunan University. Her main research interests include image processing and computer vision.



朱继杰, 2017 年于洛阳理工学院获得学士学位, 2021 年于湖南大学获得硕士学位, 主要研究方向为目标检测与深度学习。

E-mail: 1398614801@qq.com

Zhu Jijie received his B. Sc. degree from Luoyang Institute of Science and Technology in 2017, M. Sc. degree from Hunan University in 2021. His main research interests include object detection and deep learning.



凌志刚 (通信作者), 2000 年于长安大学获得学士学位, 2003 年和 2010 年于西北工业大学分别获得硕士学位和博士学位, 现为湖南大学电气与信息工程学院副教授, 国家机器人视觉感知与控制技术工程研究中心研究员, 主要研究方向为计算机视觉和模式识别。

E-mail: zgling@hnu.edu.cn

Ling Zhigang (Corresponding author) received his B. Sc. degree from Changan University in 2000, M. Sc. and Ph. D. degrees from the Northwestern Polytechnical University in 2003 and 2010, respectively. He is currently an associate professor with the College of Electrical and Information Engineering and a research fellow with the National Engineering Research Center for Robot Visual Perception and Control Technology, Hunan University, Changsha, China. His main research interests include computer vision and pattern recognition.