

# 基于优化 YOLOv3 的低空无人机检测识别方法

马旗<sup>1,2\*</sup>, 朱斌<sup>1,2\*\*</sup>, 张宏伟<sup>1,2\*\*\*</sup>, 张杨<sup>1,2</sup>, 姜雨辰<sup>1,2</sup>

<sup>1</sup>国防科学技术大学电子对抗学院脉冲功率激光技术国家重点实验室, 安徽 合肥 230037;

<sup>2</sup>国防科学技术大学, 安徽 合肥 230037

**摘要** 无人机的快速发展与应用在给社会带来便利的同时,也对公共安全、个人隐私、军事安全等构成了严重威胁。快速准确地发现未知无人机变得越来越重要。在无人机检测技术中,基于机器视觉的方法具有成本低廉和配置简便的优点。针对低空快速移动的无人机,提出一种基于优化 YOLOv3 的无人机检测识别方法。利用残差网络及多尺度融合的方式对原始的 YOLO 网络结构进行优化,提出 O-YOLOv3 网络,利用真实拍摄的无人机数据集进行训练和测试。实验结果表明,所提方法的平均准确度优于原始方法,检测速度满足实时性要求。

**关键词** 图像处理; 低空无人机; 目标检测; 残差网络; 多尺度融合

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.201006

## Low-Altitude UAV Detection and Recognition Method Based on Optimized YOLOv3

Ma Qi<sup>1,2\*</sup>, Zhu Bin<sup>1,2\*\*</sup>, Zhang Hongwei<sup>1,2\*\*\*</sup>, Zhang Yang<sup>1,2</sup>, Jiang Yuchen<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Pulsed Power Laser Technology, College of Electronic Engineering,  
National University of Defense Technology, Hefei, Anhui 230037, China;

<sup>2</sup>National University of Defense Technology, Hefei, Anhui 230037, China

**Abstract** The rapid development and application of unmanned aerial vehicles (UAVs) not only bring convenience to the society, but also pose serious threats to public security, personal privacy, and military security. Therefore, rapid and accurate detection of unknown UAV becomes increasingly important. In addition, in UAV detection technology, the method based on machine vision has the advantages of low cost and simple configuration. This paper proposes an optimized YOLOv3 (You Only Look Once version3) based detection and recognition method for low altitude and fast moving UAV. The residual network and multi-scale fusion are used to optimize the network structure of the original YOLO, and the O-YOLOv3 network is proposed. The training and testing are carried out using the real filmed UAV dataset. The experimental results show that the average precision of the optimized method is better than that of the original method, and the detection speed meets the real-time requirement.

**Key words** image processing; low-altitude unmanned aerial vehicles; target detection; residual network; multi-scale fusion

**OCIS codes** 100.4996; 100.3008; 100.2000

## 1 引言

随着现代电子通信技术及航空技术的飞速发展,民用无人机已迅速应用于社会生活的不同方面,在提供很大便利的同时,也对公共安全、个人隐私、军事安全等构成了严重威胁。例如,媒体经常报道一些不法分子通过对低空小型无人机进行改装,即

加装光电摄像设备进行非法拍摄活动,或者挂载危险物品(爆炸物等)来实施恐怖袭击<sup>[1]</sup>。有效的无人机检测技术对于应对无人机威胁问题非常重要。目前,常用的检测技术通常包括基于声波探测<sup>[2-3]</sup>、无线电探测<sup>[4]</sup>和雷达探测<sup>[5-6]</sup>等手段,但是这些技术手段通常需要昂贵的设备和严格的配置。基于机器视觉的技术具有成本低廉、易于配置的优点,因此,基

收稿日期: 2019-04-12; 修回日期: 2019-04-25; 录用日期: 2019-05-20

\* E-mail: 905303927@qq.com; \*\* E-mail: zhubinee@163.com; \*\*\* E-mail: zhw25055@163.com

于机器视觉的低空无人机检测技术具有良好的发展前景。

关于基于视觉的低空无人机检测识别任务：Rozantsev 等<sup>[7]</sup>提出了一种基于回归的视频无人机检测方法，首先利用两个增强树回归器稳定运动，之后使用具有固定尺寸的立方体来检测无人机，当在飞行过程中无人机大小逐渐变化时，这种方法可以获得良好的检测性能，但无人机的实际飞行过程速度快、大小不一，导致该方法的实际表现较差；Martínez 等<sup>[8]</sup>通过重建无人机的三维轨迹来检测装有固定摄像机的无人机，但由于摄像机的观测点固定，因此该方法对无人机的监视范围有限，对背景变化的无人机检测效果也较差。总的来看，基于视觉的无人机检测识别技术仍存在许多亟待解决的问题。

目前，深度学习技术已能模拟人脑建立学习分析的神经网络，利用人脑的机制实现数据分级特征的理解和表达，具有强大的视觉信息处理能力，在处理速度和处理性能上均能较好地满足当前图像视频大数据的要求。利用卷积神经网络构建的模型在目标检测识别领域已得到了十分广泛的应用，如 Faster R-CNN<sup>[9]</sup>、R-FCN<sup>[10]</sup>、SSD<sup>[11]</sup>、YOLO<sup>[12]</sup>等方法。

无人机的检测识别需要一定的实时性。SSD 与 YOLO 这 2 种方法在目标检测过程中都能达到实时检测。同时，文献<sup>[13]</sup>表明，YOLOv3 在标准数据集上的检测精度和速度都优于 SSD 方法。但原始的 YOLOv3 方法是针对常规目标数据集的检测方法，并不适用于对低空无人机的检测识别。针对无人机的实际检测情况，本文提出一种基于优化 YOLOv3 的低空无人机检测识别方法。该方法利用残差网络和多尺度预测相结合的优点，基于  $K$ -means 算法得出最优锚点，对原始的 YOLO 网络结构进行优化，并利用实际拍摄的数据集进行训练和测试。结果表明，优化后的方法与原始方法相比，能够较为准确地检测出低空无人机目标，且具有实时

检测的性能。

## 2 YOLO 方法的原理及优化

### 2.1 YOLO 检测过程

因为低空无人机的类别不止一种，图像中包含多种类别的目标，所以需要每个预测框的不同类别做出判断，具体的检测过程如下。

1) 将整个图片划分为  $S \times S$  的格子，每个格子包含 5 个预测边框，每个预测边框包含 5 个参数，即  $(x, y, w, h, I_{\text{object\_conf}})$ ，其中  $(x, y)$  为预测边框中心与对应格子边界的偏移值， $(w, h)$  为边框的宽高与整幅图像宽高之比， $I_{\text{object\_conf}}$  为目标的置信度分数，即边框中包含目标的置信度值，可表示为

$$I_{\text{object\_conf}} = P_r(o_{\text{bject}}) \cdot I_{\text{pred}}^{\text{truth}}, \quad (1)$$

式中： $P_r(o_{\text{bject}})$  表示预测边框对应的单元格中是否包含目标，包含时为 1，不包含时为 0； $o_{\text{bject}}$  为图像中的目标； $I_{\text{pred}}^{\text{truth}}$  表示预测边框与真实边框的交并比。

2) 利用卷积神经网络提取特征并进行预测。当网络中类别的个数为  $C$  且格子中有物体存在的条件概率为  $P_r(c_{\text{lass}} | o_{\text{bject}})$  时 ( $c_{\text{lass}}$  为目标类别)，得到某一类别的概率为  $P_r(c_{\text{lass}, k})$  ( $c_{\text{lass}, k}$  为第  $k$  个类别)。设定某一类别的概率与其对应交并比的乘积为该类别的置信值：

$$I_{\text{object\_conf}, i} = P_r(c_{\text{lass}} | o_{\text{bject}}) \cdot P_r(o_{\text{bject}}) \cdot I_{\text{pred}}^{\text{truth}} = P_r(c_{\text{lass}, k}) \cdot I_{\text{pred}}^{\text{truth}}. \quad (2)$$

3) 通过非极大值抑制 (NMS) 筛选预测的边框，去掉重复的边框，输出最终的预测结果。其中，损失函数的设计使用了  $S \times S \times (B \times 5 + C)$  维向量与图像真值的均方值和误差值，以达到优化模型结构的目的。但因为很多格子内并无目标物体存在，所以在损失函数中分别给有目标存在与无目标存在的预测框设置不同比例的系数来平衡差异。此外，还加入边界框的损失系数和判断类别的损失系数，以使边界框内包含目标的损失系数有较高的权重。损失函数可表示为

$$l_{\text{oss}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 + \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{C \in C_{\text{lasses}}} [p_i(C) - \hat{p}_i(C)]^2, \quad (3)$$

式中： $\lambda_{\text{coord}}$  用来增强边界框在损失计算中的重要性；

$\lambda_{\text{noobj}}$  用来减弱非目标区域对目标区域置信度计算的

影响;  $C_i$  为对应目标框的可信度分数;  $p_i(c)$  为目标条件类别概率,即在网络含有目标中心的条件下的目标类别概率;  $c$  为目标类别;  $c_{lasses}$  为所有类别的集合;  $I_{ij}^{obj}$  表示目标是否出现在第  $i$  个网格中第  $j$  个框中,若是则取值为 1,否则为 0;  $(\hat{x}_i, \hat{y}_i)$  为真实边框中心与对应格子边界的偏移值;  $(\hat{w}_i, \hat{h}_i)$  为真实框的宽高与整幅图像宽高之比;  $\hat{C}_i$  为预测框与真实框的交并比;  $\hat{p}_i(c)$  为目标实际类别对应的概率。

### 2.2 O-YOLOv3 的网络结构

提出基于 Optimized-YOLOv3 (简称 O-

YOLOv3)的低空无人机检测识别方法,其实现框架如图 1 所示。首先,O-YOLOv3 使用深度残差网络<sup>[14]</sup>(网络结构如表 1 所示)提取样本中低空无人机的特征;然后,在残差网络后添加 4 个卷积层,形成金字塔型的多尺度预测网络;接着,将预测网络中得到的特征图与之前深度残差网络中对应尺寸的特征图进行融合,以获得更有价值的语义信息;最后,利用 NMS 的方法去除重复的预测框,得到最终的低空无人机检测识别结果。与原始的 YOLOv3 相比,O-YOLOv3 易于训练且效果良好。

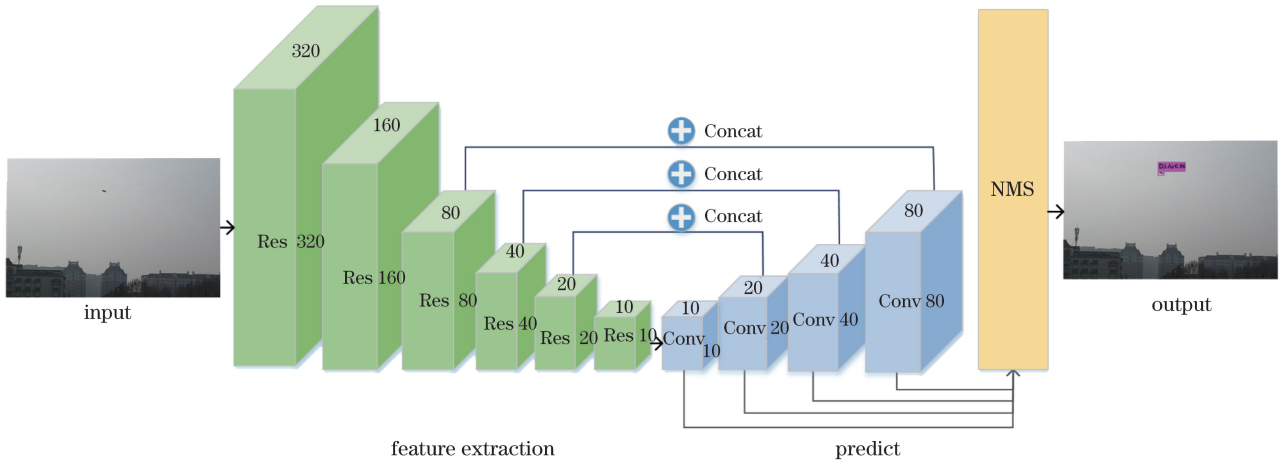


图 1 用于低空无人机的检测识别结构图

Fig. 1 Structural diagram of detection and recognition for low-altitude UAV

1) 利用深度残差网络提取无人机特征

O-YOLOv3 根据残差网络的原理,构建了深度残差网络用于提取无人机的特征。残差网络是由一系列残差块组成,如图 1 中 Res 所示。

残差块的结构如图 2 所示。两个残差块之间按照顺序堆叠,图中带加号的实心圆圈表示跳过连接,其相应的公式为

$$x_{l+1} = x_l + F(x_l, W_l), \quad (4)$$

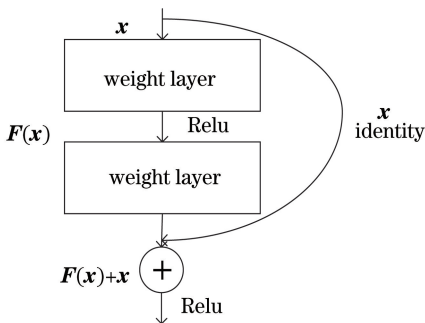


图 2 残差块结构图

Fig. 2 Residual block structural diagram

式中:  $x_l$  和  $x_{l+1}$  分别是第  $l$  个残差块的输入与输出向量;  $F(x_l, W_l)$  表示转换函数,对应于由堆叠层组成的残差分支;  $W_l$  为第  $l$  个残差块中卷积层的权重。这样组成的深度残差网络易于信息流动,并且容易训练。

O-YOLOv3 的网络结构如表 1 所示,表中 Type 表示网络中每一层的类型,Filters 表示卷积核的个数,Size 表示卷积核的大小,Output 表示输出特征图的大小,左侧  $1 \times$ 、 $2 \times$  等表示残差块堆叠的个数。以  $640 \text{ pixel} \times 640 \text{ pixel}$  的输入图像为例:第一层卷积用 16 个尺寸为  $3 \times 3$  的卷积核实现滤波;将第一层的输出作为第二层的输入,用 32 个尺寸为  $3 \times 3$  的卷积核、步长为 2 进行滤波,实现下采样操作,之后添加残差块(由  $1 \times 1$  卷积层和  $3 \times 3$  卷积层组成)来增加网络的深度,此时得到的特征图尺寸为  $320 \text{ pixel} \times 320 \text{ pixel}$ ;执行包含  $2 \times$ 、 $8 \times$ 、 $8 \times$ 、 $4 \times$ 、 $4 \times$  残差块的 5 组网络,分别得到  $160 \text{ pixel} \times 160 \text{ pixel}$ 、 $80 \text{ pixel} \times 80 \text{ pixel}$ 、 $40 \text{ pixel} \times 40 \text{ pixel}$ 、 $20 \text{ pixel} \times 20 \text{ pixel}$ 、 $10 \text{ pixel} \times 10 \text{ pixel}$  分辨率的特

表 1 用于低空无人机检测识别的 O-YOLO 网络结构

Table 1 O-YOLO network structure for detection and recognition of low altitude UAV

Repetition times	Type	Filters	Size	Output
1×	Convolutional	16	3×3	640×640×16
	Convolutional	32	3×3/2	320×320×32
	Convolutional	16	1×1	
	Convolutional	32	3×3	
	Residual			320×320×32
2×	Convolutional	64	3×3/2	160×160×64
	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			160×160×64
1×	Convolutional	128	3×3/2	80×80×128
8×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			80×80×128
8×	Convolutional	256	3×3/2	40×40×256
	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			40×40×256
1×	Convolutional	512	3×3/2	20×20×512
4×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			20×20×512
1×	Convolutional	1024	3×3/2	10×10×1024
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			10×10×1024
1×	Avgpool		Global	
	Connected		3	
	Softmax			

征图。在这 5 组残差块组成的网络中,除了卷积核数与特征图的尺度不同之外,每一个残差块都相似。其中,该网络的所有卷积层上均添加批量归一化层<sup>[15]</sup>以加速网络训练收敛速度。最后,选取分辨率为 80 pixel×80 pixel,40 pixel×40 pixel,20 pixel×20 pixel,10 pixel×10 pixel 的特征图与下一节提出的上采样特征图进行融合,构成特征金字塔<sup>[16]</sup>用于低空无人机的检测识别。

## 2) 多尺度网络预测低空无人机

O-YOLOv3 网络在多个不同尺度的特征图上回归预测无人机的位置和类别。如图 1 右侧部分所示,将不同分辨率的卷积层(图 1 中 Conv)分成 4 个分支(特征图的分辨率分别为 10 pixel×10 pixel,20 pixel×20 pixel,40 pixel×40 pixel,80 pixel×80 pixel),每个分支独立地对无人机进行预测。其中,每个分支都配备 6 层对应大小的卷积层。为增

强金字塔对特征的代表能力,对于分辨率为 20 pixel×20 pixel,40 pixel×40 pixel,80 pixel×80 pixel 的分支,执行 2 倍大小的上采样。考虑到计算速度,上采样方式选择常用的最近邻上采样。将上采样后的特征与深度残差网络中对应尺寸的特征图进行级联,使每个分支能充分利用上下文的语义信息来对目标进行检测,并且也能够共享从残差网络中提取到的特征,从而达到提升检测效果的目的。

在预测阶段,对于输入的低空无人机图像,将会预测一个三维张量,包括无人机的边界框、置信度和类别。其中,O-YOLOv3 网络将特征图划分成  $S \times S$  的格子(对于不同尺度的特征图, $S$  大小不同,分别为 10,20,40,80),每个格子预测 3 个不同的边界框,张量可以表示为  $S \times S \times [3 \times (4+1+3)]$ ,即 4 个边界框的偏移量、1 个目标置信度、3 个目标类型。

## 3 实验及结果与评价

### 3.1 数据集的准备

由于基于深度学习的低空无人机检测识别方法较少,而且也没有公开的数据集或是标准的数据集。为此,首先进行数据集的采集及构建。

1) 数据采集。利用可见光探测器在不同时间、不同背景对3种型号民用无人机进行拍摄。无人机型号分别为大疆-精灵3(DJ-3)、大疆-御 Pro(DJ-Pro)和大疆-御 Air(DJ-Air)。为确保数据的多样性,拍摄过程中充分考虑了低空无人机的各种飞行姿态,包括盘旋、急速升降、平稳飞行等。最终共得到3524张可见光图像。

2) 数据标注。采取人工标注方式。为确保数据的有效性,对于目标被遮挡面积大于等于50%的样本进行标注。最终,对3012张可见光图像中的低空无人机目标进行了边界框的标注,得到了具有标注信息的数据。按照4:1的比例将其分成训练集和测试集。

3) 图像增强。为提升方法的检测识别精度,采取目标检测领域中通用的图像增强方式对训练集进行增强处理,包括亮度、对比度的操作处理。

4) 数据扩充。考虑到无人机飞行时的姿态不全是保持水平,还有倾斜等,故对训练集进行了翻转及 $\pm 10^\circ$ 、 $\pm 20^\circ$ 的旋转操作。如果旋转处理后图像中靠近边缘的目标出现残缺或完全丢失的情况,则放弃该样本数据。

5) 数据集构建。通过对训练集的图像进行增强和扩充处理,分别获得3类数据集。如表2所示,UAV\_A的训练集是由原始图像的低空无人机目标组成;UAV\_B的训练集是由原始图像(UAV\_A)与经过图像增强处理后的图像组成;UAV\_C的训练

集是由UAV\_A和UAV\_B及其扩充后的样本组成(包含数据集UAV\_A和UAV\_B)。其中,测试集选用相同的样本用于方法验证。

表2 不同训练集的处理方式及样本数

Table 2 Processing methods and sample numbers for different train datasets

Train dataset	Processing method	Number of samples
UAV_A	Original	2400
UAV_B	Original + Image enhancement	4800
UAV_C	Original + Image enhancement + Data augmentation	46000

### 3.2 YOLO 网络训练

针对原始YOLOv3网络及优化后的O-YOLOv3网络进行对比实验。实验环境如下:网络的训练为Darknet框架;服务器平台配置为Ubuntu 16.04操作系统,Intel(R) Xeon(R) CPU E5-2673 v3 @ 2.40 GHz,搭载 GeForce GTX 1080Ti/PCIe/SSE2显卡,内存32 GB。

针对低空无人机的训练集,通过K-means聚类方法<sup>[17]</sup>自动生成锚点,并与原始方法进行比较。原始的YOLOv3采用3种尺度的9个锚点,分别为:(10×13),(16×30),(33×23),(30×61),(62×45),(59×119),(116×90),(156×198),(373×326),平均重叠率为79.62%;而O-YOLOv3采用4种尺度12个锚点,如图3所示,分别为:(16×15),(19×16),(21×20),(23×17),(24×20),(28×22),(41×24),(45×29),(61×32),(72×41),(159×62),(179×76),平均重叠率为86.97%。

在整个训练过程中,2个网络的训练参数相同,最大迭代次数为50000,动量和权重衰减分别为0.9和0.0005,批量大小设置为64,初始学习率为 $10^{-3}$ ,并在后两个阶段降低为 $10^{-4}$ 和 $10^{-5}$ 。相对应每个阶段

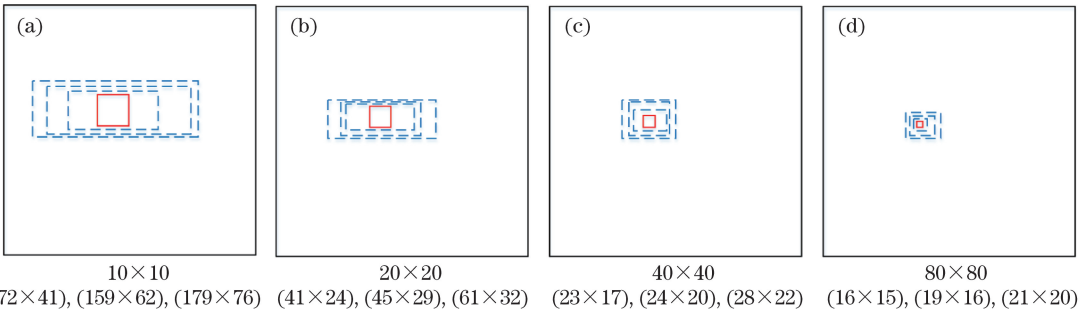


图3 4种尺度的12个锚点示意图。在每一种尺度下,实线框为一个对应的方格,虚线框为对应的锚点框。

(a)  $10 \times 10$ ; (b)  $20 \times 20$ ; (c)  $40 \times 40$ ; (d)  $80 \times 80$

Fig. 3 Diagrams of 12 anchors with 4 scales. At each scale, the solid box is a corresponding square, and the dotted boxes are the corresponding anchor boxes. (a)  $10 \times 10$ ; (b)  $20 \times 20$ ; (c)  $40 \times 40$ ; (d)  $80 \times 80$

的学习率,网络分别迭代 25000、15000 和 10000 次。

### 3.3 测试结果及评价

分别用 UAV\_A、UAV\_B 和 UAV\_C 的训练集来训练 2 种检测识别网络,并与经原始及优化后的网络训练得到的模型进行比较,重点比对其对不同类型低空无人机的检测识别性能。

在对测试集进行检测时,将低空无人机目标测

试集图像输入训练好的网络,进行目标的位置回归:当模型预测的目标边界框与手工标注的边界框交集比( $I_{OU}$ )不小于 0.7 且无人机的类型识别准确时,则认为检测识别正确;否则为错误。同时,选取平均准确率(AP)、平均 AP 值(mAP)和检测速率作为评价准则,以兼顾模型的准确率、召回率以及实时性。实验结果如表 3、图 4 所示。

表 3 不同方法在不同测试集上的评价结果

Table 3 Evaluation results of different methods on different test datasets

Train set	Method	Detection speed /s <sup>-1</sup>	mAP /%	AP /%		
				DJ-Air	DJ-Pro	DJ-3
UAV_A	YOLOv3	30.28	70.11	64.14	60.62	85.57
	O-YOLOv3	26.13	77.87	73.86	69.12	90.63
UAV_B	YOLOv3	29.87	71.24	64.56	62.13	87.02
	O-YOLOv3	25.75	79.43	75.01	71.24	92.04
UAV_C	YOLOv3	30.05	73.86	67.23	65.02	89.34
	O-YOLOv3	25.85	82.15	77.86	73.25	95.33

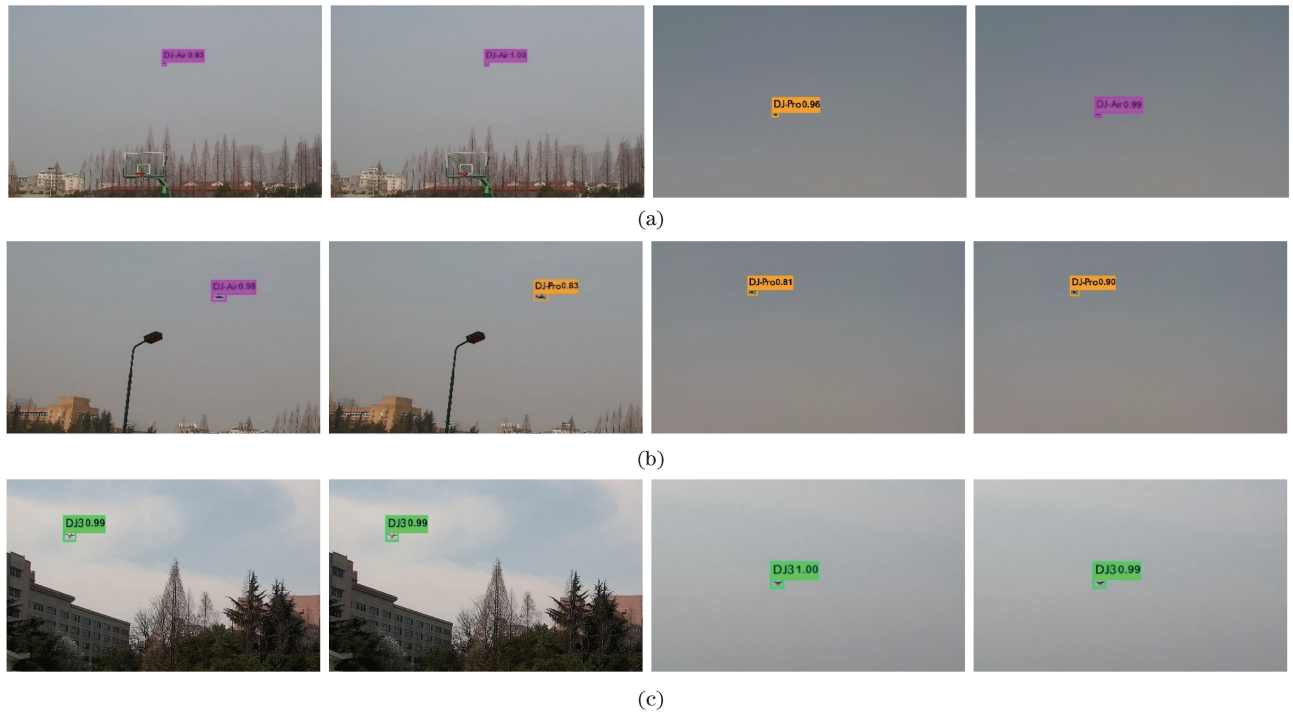


图 4 测试集中部分检测结果对比图(第 1 列和第 3 列为 YOLOv3 方法,第 2 列和第 4 列为 O-YOLOv3 方法)。

(a) DJ-Air;(b) DJ-Pro;(c) DJ-3

Fig. 4 Comparison of partial detection resultson test dataset (The first and third columns are the results of the YOLOv3 method, and the second and fourth columns are the results of the O-YOLOv3 method). (a) DJ-Air; (b) DJ-Pro; (c) DJ-3

1) 相同网络结构对应不同训练集的结果比较。由 YOLOv3 网络结构对应的结果可以看出,经过图像增强的训练集(UAV\_B)对应的 mAP 值为 71.24%,相较于原始训练集(UAV)提高了 1.13%,而经过图像增强和扩充的训练集(UAV\_C)对应的 mAP 值为 73.86%,相较于原始训练集(UAV\_A)

提高了 3.75%。对于 O-YOLOv3 网络结构,经过图像增强以及图像增强和扩充对应的 mAP 值分别提高了 1.56%和 4.32%。由此可见,图像增强与数据扩充对方法的检测性能有一定的提升。

2) 相同训练集对应不同网络结构的结果比较。从表 3 可知,UAV\_C 对应的 mAP 值都是最优的。

在此条件下比较不同的网络结构,可以看出, O-YOLOv3的 mAP 值为 82.15%, 相比于 YOLOv3 的 73.86% 提高了 8.29%, 检测性能有较大幅度的提升, 说明网络结构的优化对提升检测效果是非常重要的。

3) 检测速度比较。YOLOv3 的检测速度保持在  $30\text{ s}^{-1}$  左右, 而 O-YOLOv3 的检测速度约为  $26\text{ s}^{-1}$ , 说明在进行检测的同时两种方法都能够保持一定的实时性能, 这在实际应用中有很大的作用。

由图 4 的对比可以看出, 对于 DJ-Air 和 DJ-Pro 两种无人机, 原始的方法存在部分误检, 优化方法较好地减少了误检; 对于 DJ-3 型无人机, 两种方法都能较地进行检测识别。误检情况的减少说明优化方法的检测精度有一定提升。综上所述, 优化后的方法在检测精度上高于原始方法, 虽然检测速度稍低于原始方法, 但仍具备实时检测的能力, 并且具有一定的稳健性。

## 4 结 论

针对低空无人机目标存在的威胁, 采用可见光探测技术, 以 YOLOv3 网络为基础, 通过深度残差网络及多尺度的预测来优化网络结构, 提出 O-YOLOv3 网络。构建了低空无人机的数据集, 并对两种方法进行比较评价。由实验分析可知, 改进方法性能得到提高的原因有: 1) 通过具有针对性的图像增强与数据扩充操作, 数据集的数量和多样性增多, 提高了模型的泛化性能和检测精度; 2) 原始的网络结构与锚点值仅适用于常规目标, 而优化后的网络结构与重新聚类得到的锚点值更适用于低空无人机目标, 最终提高了模型的检测精度并且减少了误检; 3) 优化后的网络结构的网络层数较原始网络有所加深, 这使得网络参数增加, 也导致检测速度有所下降, 但并不影响其实时检测的能力。

虽然优化后的方法相较于原始方法在性能上有了较大的提升, 但仍存在一些不足: 1) 数据集的类别较为有限, 如能获取更多类别的低空无人机数据, 性能将会有更大的提升; 2) 采集的无人机图像均处于光照良好的可见光条件下, 对于夜间或较为昏暗的光线条件下的无人机检测识别仍有待进一步的改进和验证。这将是未来深入研究的内容。

## 参 考 文 献

[1] Zhao Y Y, Chang J D. Detection and handling of drones in low altitude airspace [J]. Electronic

Warfare, 2016(4): 44-49.

赵云雨, 常晋聃. 对低空无人机的探测与管控新技术综述[J]. 电子对抗, 2016(4): 44-49.

[2] Busset J, Perrodin F, Wellig P, *et al.* Detection and tracking of drones using advanced acoustic cameras [J]. Proceedings of SPIE, 2015, 9647: 96470F.

[3] Wang T J, Huang W N, Fan Z X, *et al.* Research on the methods of UAV detection technology [J]. China Radio, 2017(11): 57-58.

汪庭霖, 黄伟宁, 范振雄, 等. 无人机侦测技术方法研究[J]. 中国无线电, 2017(11): 57-58.

[4] Nguyen P, Ravindranatha M, Nguyen A, *et al.* Investigating cost-effective RF-based detection of drones [C] // Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, June 26-26, 2016, Singapore. New York: ACM, 2016: 17-22.

[5] Li Q, Huang K M. Simulation analysis of radar detection range for small low-altitude UAV [J]. Radio Engineering, 2018, 48(4): 303-307.

李琴, 黄卡玛. 低空小型无人机雷达探测距离仿真分析[J]. 无线电工程, 2018, 48(4): 303-307.

[6] Klare J, Biallawons O, Cerutti-Maori D. Detection of UAVs using the MIMO radar MIRA-CLE Ka [C] // Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, June 6-9, 2016, Hamburg, Germany. Germany: VDE, 2016: 731-734.

[7] Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 4128-4136.

[8] Martínez C, Mondragón I F, Olivares-Méndez M A, *et al.* On-board and ground visual pose estimation techniques for UAV control [J]. Journal of Intelligent & Robotic Systems, 2011, 61: 301-320.

[9] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN [J]. Acta Optica Sinica, 2018, 38(6): 0615004.

冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测 [J]. 光学学报, 2018, 38(6): 0615004.

[10] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks [C] // Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain.

- New York: Curran Associates Inc., 2016: 379-387.
- [11] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [12] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [13] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J/OL]. (2018-04-08) [2019-03-10]. <https://arxiv.org/abs/1804.02767>.
- [14] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [15] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J/OL]. [2015-03-02] [2019-03-10]. <https://arxiv.org/abs/1502.03167>.
- [16] Ren Z J, Lin S Z, Li D W, *et al.* Mask R-CNN object detection method based on improved feature pyramid [J]. Laser & Optoelectronics Progress, 2019, 56(4): 041502.  
任之俊, 蔺素珍, 李大威, 等. 基于改进特征金字塔的 Mask R-CNN 目标检测方法[J]. 激光与光电子学进展, 2019, 56(4): 041502.
- [17] Wei Y M, Quan J C, Houyu Q Y. Aerial image location of unmanned aerial vehicle based on YOLO v2[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111002.  
魏湧明, 全吉成, 侯宇青阳. 基于 YOLO v2 的无人机航拍图像定位研究[J]. 激光与光电子学进展, 2017, 54(11): 111002.