

基于 YOLOv3 的正下无人机视角挖掘机实时检测方法

蔡振宇, 王泽锴, 陈特欢*, 李文来

(宁波大学 机械工程与力学学院, 浙江 宁波 315211)

摘要: 针对无人机巡检的智能化要求, 提出一种应对高空巡检场景下的实时挖掘机检测模型. 该模型以 YOLOv3 为基础, 将骨干网络精简至 43 层, 通过特征融合策略使检测任务在两个尺度上进行. 此外模型还借鉴了 focal loss 的思想设计损失函数. 文中实验对象为正下无人机视角的挖掘机目标. 在完成了数据集的搭建工作后, 根据正下无人机视角的目标特性进行训练, 使模型达到最优解. 最终经实验验证, 在相同输入尺寸的情况下, 本文所提出的检测模型比 YOLOv3 准确率更高、鲁棒性更好, 且帧数可提升 10 帧·s⁻¹.

关键词: 无人机视角; YOLOv3; 特征融合; 目标检测

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-5132 (2021) 02-0042-07

当前工程中的巡检任务多为人工实现, 人工巡检存在效率低下、巡检频次低、成本过高等诸多问题^[1]. 而无人机自动巡检技术为巡检任务提供了一种智能化的解决方案, 能够在很大程度上取代人工巡检. 其中对高空环境下目标检测技术的研究是无人机在精准货物投送、自主充电、目标跟随飞行等无人机自主行为领域的基础, 具有较高的应用价值.

在经典的机器学习方法中, 对于目标特征的提取是工作的核心内容, 通常会采用尺度不变特征变换^[2]、方向梯度直方图^[3]等方法进行特征提取. 再将所得的特征输入诸如支持向量机^[4]、迭代器 (AdaBoost)^[5]等分类器之中. 然而所提取的特征严重依赖于人工设计, 且在实际工作环境下, 满足高指标智能巡检需求需要在高空进行检测工作. 处于无人机视角下的检测目标在获取的图像里仅占据着少量像素. 此外, 巡检环境多为郊外地区, 检测过程中存在背景复杂度高、视频数据计算量大等问题^[6]. 2006 年 Hinton 等^[7]首次引入了深度学习的概念, 通过设计深层次的神经网络结构对目标特征进行提取. 深度学习在很大程度上解决了手工

特征的应用局限性, 大幅度地提升了模型的鲁棒性与灵活性^[8].

当前, 基于深度学习的目标检测方法大体上可分为两类, 即基于回归思想的单阶检测方法和基于区域建议的两阶检测方法. 其中基于回归思想的单阶检测方法使用单个前馈神经网络来实现对区域类别的预测并对区域进行微调. 常见的基于回归思想的检测方法包括 SSD^[9]、YOLO^[10]、YOLOv2^[11]、YOLOv3^[12]、RetinaNet^[13]等等.

单阶检测方法无需第二阶段的区域分类操作, 能极大地提高检测效率. YOLOv3 作为一种单阶检测方法, 在保证一定检测效率的同时, 其检测精度比经典的 SSD 以及前代 YOLO 模型更高, 从而被广泛地应用于许多实际场景中. 张富凯等^[14]在 YOLOv3 的基础上设计 6 个不同尺度的卷积特征图, 实现了交通监控中车辆的检测. 戴伟聪等^[15]针对遥感图像中的飞机目标采用一个 49 层网络结构代替模型的骨干网络并对检测尺度进行了扩展. 舒军等^[16]则在 YOLOv3 的基础上引入了更高分辨率的检测模块来强化航拍小目标的检测能力.

当前对 YOLOv3 的研究重点在于提高模型的

收稿日期: 2020-06-06.

宁波大学学报(理工版)网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家自然科学基金 (61703217); 浙江省教育厅科研项目 (Y201840096, Y201839158, Y201840072); 宁波市科技创新 2025 重大专项 (2019B10100).

第一作者: 蔡振宇 (1994-), 男, 湖南株洲人, 在读硕士研究生, 主要研究方向: 深度学习、目标检测. E-mail: bnbnnvh@163.com

*通信作者: 陈特欢 (1988-), 男, 浙江宁波人, 副教授, 主要研究方向: 微流体控制、计算最优控制、机器人控制. E-mail: chentehuan@nbu.edu.cn

检测准确率, 然而模型的检测速率却很少有显著的提升. 本文所考虑的无人机巡检任务中, 主要通过四旋翼飞行器的机载摄像头采集图像, 并交由移动地面站进行处理. 由于移动地面站的图像处理能力相对 GPU 服务器性能较弱, 运行深度学习模型时难以保证检测速率. 这意味着执行无人机巡检任务对模型的检测速率具有很高的要求.

目标检测的对象为无人机电力巡检过程中所出现的违章挖掘机. 考虑到电力系统的布置场景多为郊外环境, 场景复杂性较高, 且文中主要的研究对象为静态图像中的挖掘机, 挖掘机在静态图像中所呈现出的外观形态会随着工作状态发生变化, 这也意味着挖掘机目标的检测比一般静态目标的检测难度更大. 此外, 出于精准定位的需求, 采集的挖掘机图像皆为正下无人机视角. 这也导致图像中的目标特征信息量较低, 在复杂的郊外环境下容易出现目标的误判现象.

本文以 YOLOv3 模型为基础进行改进工作, 构建多尺度特征融合神经网络模型, 并建立以正下无人机视角的挖掘机为样本的数据集, 用于训练和测试所构建的深度学习目标检测模型. 模型的骨干网络部分由 Darknet53^[12]修改而来, 包含 43 个卷积层, 实现了模型参数的精简. 借鉴深度残差网络(ResNet)^[17]中的残差块设计方法加速计算. 此外还将特征金字塔网络(Feature Pyramid Networks, FPN)^[18]的分级检测思想应用于模型之中, 采用将深层与浅层特征融合的策略在大小两个尺寸的特征图上进行物体检测. 为保障模型的检测速率, 先验框的数量设置为 6 个, 先验框的获取则采取 k 均值方法对样本做 6 类聚类处理. 考虑到训练过程中存在正负样本不均衡的现象, 为了保障检测的准确度, 借鉴 RetinaNet 的 focal loss 思想设计损失函数来更新模型. 最后通过实验证明该改进算法实现了对模型的性能提升.

1 检测方法

1.1 YOLO 检测思想

YOLO 算法已发展至第 3 代, 它将目标检测作为回归问题处理. 如图 1 所示, YOLO 将预测图划分为 $s \times s$ 个单元格操作, 当目标的中心坐标位于其中某个单元格时, 该单元格就负责目标的检测工

作. 每个格子预测 B 个边界框位置信息(x, y, w, h)及其置信度(conf), 以及多类别概率(cls_num).

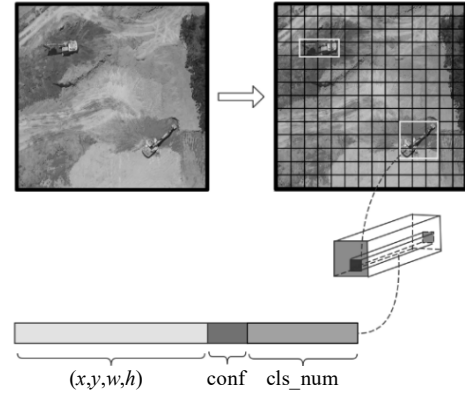


图 1 YOLO 的检测示意图

1.2 网络结构

模型的骨干网络基于 Darknet53 进行修改得到, 为强化网络输入数据的信息量, 将输入尺寸由默认的 416×416 扩展至 608×608 , 通道数为 3. 网络模型的整体结构如图 2 所示.

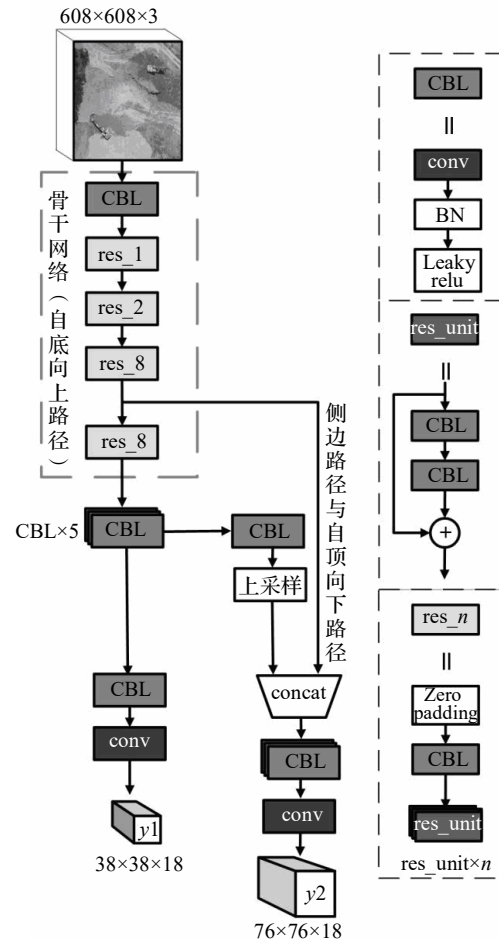


图 2 模型的网络结构

图 2 中, CBL 模块代表一个完整的卷积过程, 该结构在卷积操作后加上批标准化(Batch Normalization, BN)^[19]层进行处理, 减少过拟合现象, 并串联一个激活层, 激活函数选用 Leaky RELU.

网络模型在卷积计算过程中共采取 2 种计算策略, 分别为“SAME”和“VALID”. 对于无法整除的情况则采用向上取整的方式获得输出结果.

$$w_{\text{out}} = \begin{cases} \frac{w_{\text{in}} - F}{x} + 1, \text{ VALID}, \\ \frac{w_{\text{in}} + 2\rho - F}{x} + 1, \text{ SAME}, \end{cases} \quad (1)$$

式中: w_{out} 为卷积输出的特征图尺寸; w_{in} 为进入卷积前的特征图尺寸; F 为卷积核尺寸; x 为卷积核的移动步长; ρ 为特征图的外部填充圈数. 在卷积过程中, 当步长 x 为 1 时按“SAME”运算, x 为 2 时按“VALID”运算.

考虑小目标检测需求, 提取的图像特征将更着重于细节特征, 这意味着高层级的语义信息存在大量信息冗余. 针对这种情况, 对 YOLOv3 模型骨干网络部分的 52 层卷积进行层向剪枝处理以减少模型的参数, 经精简后的模型其特征提取部分构建为一 43 层的全卷积网络结构.

总体来说, 修改后的网络结构包含自底向上路径、自顶向下路径、侧边路径等部分. 在输入尺寸同为 416×416 的条件下, 经修改后的模型参数量与原版模型对比见表 1, 其中参数量以 10 亿次浮点运算量为单位.

表 1 模型参数对比 ×10 亿次

模型	骨干网络参数量	整体参数量
原版 YOLOv3	49.03	65.87
改进 YOLOv3	40.35	51.64

自底向上的过程为模型的骨干网络部分. 在自底向上的过程中网络不断加深, 以此实现多级特征的提取. 自底向上的网络由 res_n 模块组合搭建, 该模块借鉴了深度残差网络的残差结构, 通过对残差结构 res_unit 的使用解决深层网络训练时梯度消失的问题. 自底向上的过程共历经 4 次下采样操作. 为了保留特征中的细节信息, 下采样的整个过程中无池化层参与, 均由步伐大小为 2 的卷积操作实现. 最终采用 38×38 与 76×76 这两个特征图尺度作为该过程的输出.

在自顶向下的路径中, 主要通过上采样的方式扩展深层特征图的尺寸, 再通过侧边路径将扩展尺寸后的特征图和浅层同尺寸特征图进行串联操作, 从而输出信息量更加丰富的特征图. 与 RetinaNet 类似, 对于 76×76 尺度的输出, 采取最邻近上采样的策略将高一层的特征图在分辨率上进行两倍的上采样工作, 并与同尺寸的自底向上输出的最高层特征图进行串联叠加.

最后的输出预测图张量形式分别为 38×38×18 与 76×76×18. 在此基础上进行位置回归和分类.

Anchor 框是一系列预设图像上的框, 本文采用 k 均值方法对收集样本进行聚类, 使聚类得到的 Anchor 框可以达到更高的交并比 (Intersection-over-Union, IOU). IOU 用来描述 A_1 、 A_2 交集区域和并集区域面积的比例, 公式为

$$R_{\text{IOU}}(A_1, A_2) = \frac{S(A_1 \cap A_2)}{S(A_1 \cup A_2)}. \quad (2)$$

k 均值聚类的距离公式为

$$d(\text{box}, \text{centroid}) = 1 - R_{\text{IOU}}(\text{box}, \text{centroid}), \quad (3)$$

式中, $R_{\text{IOU}}(\text{box}, \text{centroid})$ 为样本边界框 box 和聚类簇中心 centroid 的 IOU. centroid 的数量设置为 6, 即通过 k 均值方法最终可得到 6 种大小不同的 Anchor 框. 文中在每个预测尺度上放置 3 种不同的框, 其中在大尺度下放置 3 种小尺寸先验框, 负责较小目标的检测, 剩下的框放置于另一较小尺度中以负责视野中较大目标的检测. 每个 Anchor 的预测维度为 (4+1+类别数), 即 4 个位置信息、1 个置信度信息和类别数.

1.3 损失函数

损失函数用来描述真实值和预测值之间的差值, 模型的训练通过最小化损失函数实现. 本文损失函数由 4 部分组成, 分别为坐标误差 l_{xy} 、 l_{wh} 和置信度误差 l_{iou} 以及类别误差 l_{cls} .

$$l = \sum_{i=0}^{s^2} \sum_{j=0}^B l_{xy} + l_{wh} + l_{\text{iou}} + l_{\text{cls}}, \quad (4)$$

$$\begin{cases} l_{xy} = \lambda \times I_{\text{obj}} \times (T_{xy} - P_{xy})^2, \\ l_{wh} = 0.5\lambda \times I_{\text{obj}} \times (T_{wh} - P_{wh})^2, \\ l_{\text{iou}} = I_{\text{obj}} \times E_{\text{BC}}(T_c, P_c) + \\ \quad I_{\text{noobj}} \times (-\alpha \times \sigma(P_c)^\gamma \times \log(1 - \sigma(P_c))), \\ l_{\text{cls}} = I_{\text{obj}} \times E_{\text{BC}}(p(T_c), p(P_c)), \end{cases}$$

式中: T 为真实值; P 为网络输出的预测值; E_{BC} 为预测值和真实值的二值交叉熵; l_{iou} 和 l_{cls} 采用 E_{BC} 进行计算, l_{xy} 和 l_{wh} 则由均方误差表示. 其中预测值需要经过 Sigmoid 函数 $\sigma(x)$ 激活. 对于坐标损失, 考虑对于不同尺寸边界框的预测, 小的边界框预测偏差在相同的尺寸下对 IOU 影响更大. 设定尺寸因子为

$$\lambda = \lambda_{coord} \times (2 - T_w \times T_h), \quad (5)$$

式中: $T_w \times T_h$ 为真实边界框的面积; λ_{coord} 为坐标损失的权重, 以平衡各部分损失之间的关系. 文中 $\lambda_{coord} = 0.5$. λ_{coord} 的设定增加了小目标坐标偏移的损失比重, 平衡了目标尺寸对坐标损失带来的影响. l_{iou} 由 2 部分组成, 其中 I_{obj} 和 I_{noobj} 皆为布尔值. 在 YOLOv3 算法中, 图像被划分为 $s \times s$ 个网格, 并在划分的网格基础上进行目标预测. 当网格负责正样本目标时 I_{obj} 为 1, I_{noobj} 为 0; 当网格负责负样本背景检测时 I_{obj} 为 0, I_{noobj} 为 1. 考虑数据中严重的正负样本不平衡现象, 引入 focal loss 在针对负样本置信度损失计算的交叉熵基础上进行改进:

$$l_f = \begin{cases} -\alpha \times \sigma(P_\psi)^\gamma \times \log(1 - \sigma(P_\psi)), & T_\psi = 0, \\ -\alpha \times (1 - \sigma(P_\psi))^\gamma \times \log(\sigma(P_\psi)), & T_\psi = 1, \end{cases} \quad (6)$$

式中: P_ψ 为输出预测值; T_ψ 为真实值; γ 的作用为降低易分类样本的损失, 使置信度损失能更关注于预测值更低的难区分样本; α 则用以平衡正负样本本身的比例不平衡现象. 当网格负责负样本置信度预测时, $I_{noobj} = 1$, 真实置信度 $T_c = 0$, 从而负样本置信度预测部分表示为

$$l_{f,noobj} = I_{noobj} \times (-\alpha \times \sigma(P_c)^\gamma \times \log(1 - \sigma(P_c))). \quad (7)$$

采用文献[13]中的最优参数 ($\alpha = 0.25$, $\gamma = 2$). 最终模型损失函数为 2 个预测尺度(共 $36 \times 36 + 72 \times 72$)的损失之和.

2 实施方法

2.1 数据集

出于对无人机定位准确度的考虑, 数据集的图像采集工作由机载正下视镜头完成, 通过飞行器对挖掘机进行航拍采样. 无人机采用大疆的消费级飞行器精灵 4 Pro, 相机型号为 FC6310. 最终获得一共包含 3 599 张图片的数据样本, 共 4 932 个目标, 平均每张图片包含 1.37 个目标. 采用样本标注软件 LabelImg 对图像数据中的目标进行人工

标注. 带标注的样本数据按 8:1:1 的比例分为 3 部分: 训练集 2 879 张、验证集 360 张、测试集 360 张. 数据集的样本尺寸分布情况如图 3 所示.

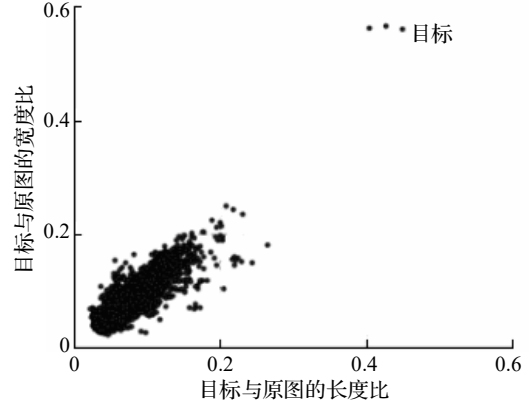


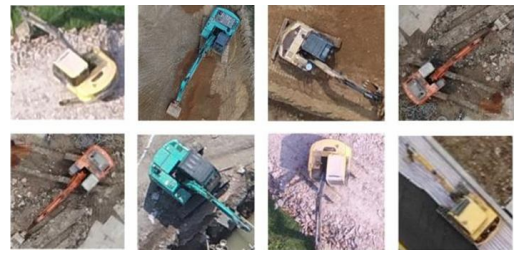
图3 数据集的样本尺寸分布

具体数据集平均目标信息统计情况见表 2.

表2 数据集信息统计

目标平均宽度/像素	目标平均长度/像素	目标平均面积/像素
58.04	51.44	3 603.29
目标平均宽度与 原图宽度比例	目标平均长度与 原图长度比例	目标平均面积占比
0.055 3	0.057 8	0.003 8

文中检测对象为施工中的挖掘机个体. 对于输入的图像样本, YOLOv3 模型将样本划分为 $s \times s$ 个网格, 文中设置每个网格的 Anchor 数为 3. 其中目标中心坐标所在的 Anchor 为正样本, 其余负责背景的 Anchor 为负样本, 样本示例如图 4 所示.



(a) 正样本



(b) 负样本

图4 数据样本示例

2.2 模型训练

为了训练检测模型,文中设置每个网格预测 3 个边界框,并通过对损失函数的数值衰减来更新模型。此外,本文还使用了数据增广与迁移学习方法来改善模型的性能。采取多尺度训练的策略,在训练过程中以每 100 批次量为一个周期,使模型的输入尺寸在 320 至 608 的范围内以 32 的倍率随机变动,从而实现模型性能的提升。

数据增广:数据增广是应对样本数据不足时提高模型性能的一种手段。综合考虑巡检过程中光照以及天气等因素的影响,本文增广策略包括调整输入样本的 HSV 值,即色相(Hue)、饱和度(Saturation)、明度(Value),以 50%的波动范围对三者进行调节。此外,由于下视镜头的中心旋转特性,文中增加数据的角度旋转增广。

训练:以公开数据集 PASCAL VOC 2012 为基础进行预训练操作,然后在 GPU 配置为 NVIDIA GTX Titan XP 的本地主机上对收集的挖掘机数据进行训练。在本文的训练策略中,批量大小设为 64,历经 600 轮次训练,初始学习率为 0.001。分别在轮次为 480、540 时下降至前一阶段的 10%。

3 实验

实验的软件环境为操作系统 Ubuntu 16.04,深度学习框架采用基于 C 语言的 Darknet。实验环境的主要硬件配置如下:处理器为 Intel Xeon(R) CPU E5-2678 v3 2.3 GHz×48,内存为 64 GB, GPU 为 GTX Titan XP。实验对象为航空视角下的挖掘机。检测器的输入主要有 2 种形式,分别为视频和图像。采用测试集进行评估,通过 $mAP@0.5IOU^{[20]}$ 评估所制作的测试集作为准确性的参考依据。

在评估改进的 YOLOv3 模型性能之前进行了实验 A。选择目前存在的 Faster R-CNN^[21]、SSD 以及原版 YOLOv3 这 3 种主流目标检测模型,在文中所搭建的挖掘机数据集上进行数据训练,模型参数设定参照论文中的数值。模型在当前数据集下的性能见表 3。

本文在表 3 中测试了各模型的检测精度以及检测速率。并通过对网络模型输入尺寸的调节来测试模型在不同分辨率下的性能表现。通过对比 Faster R-CNN、SSD 以及原版 YOLOv3 这 3 种主

流目标检测模型,可以看出针对文中提出的数据集,综合考虑检测准确率以及速率,YOLOv3 具有更加优秀的性能表现。

表 3 检测模型基准实验结果

模型	输入尺寸	mAP@0.5IOU	帧速/(帧·s ⁻¹)
Faster R-CNN	400×400	81.7	17
(Resnet50)	512×512	88.1	20
Faster R-CNN	400×400	78.6	49
(MobileNet)	512×512	82.5	45
SSD	400×400	72.5	77
(MobileNet)	512×512	82.2	92
YOLOv3	480×480	60.1	48
	512×512	88.4	44
	608×608	88.6	32

在实验 B 中,考虑正下无人机视角特性,采用合适的数据增广策略来提高模型的检测精度,加入旋转的方式对数据进行增广,选取输入尺寸为 608×608 的 YOLOv3 模型进行测试,结果见表 4。

表 4 不同旋转角度下的模型性能

角度/(°)	mAP@0.5IOU	角度/(°)	mAP@0.5IOU
0	88.6	60	89.0
15	89.0	75	88.9
30	89.3	90	88.7
45	88.9		

本文在 YOLOv3 的基础上对模型进行了多项修改以提高模型的目标检测性能。实验结果见表 5。

表 5 模型改进对比实验

实验	模型与改进	mAP@0.5IOU	帧速/(帧·s ⁻¹)
C1	YOLOv3	88.6	32
C2	YOLOv3+改进网络结构	88.3	41
C3	YOLOv3+focal loss	89.8	32

在实验 C1 中,测试了原始的基于 DarkNet53 的 YOLOv3 模型,并将其作为其他实验的基准模型,输入尺寸设置为 608×608。

在实验 C2 中,对模型的网络结构采用文中所描述的策略进行修改,以 43 层全卷积结构作为骨干网络,并使检测任务在两个检测尺度上进行。实验结果显示,针对单目标检测的情况,网络结构的精度不会带来检测准确率的大幅度波动,并显著地提升了模型的检测速率。

在实验 C3 中, 将 focal loss 引入模型的损失函数中进行对比实验, 证明损失函数的改进可提升模型的检测精度.

在实验 D 中, 对基于基准的修改进行整合, 测试改进后的 YOLOv3 模型. 并对网络模型输入尺寸进行调节来测试模型在不同分辨率下的性能表现, 实验结果见表 6.

表 6 不同输入尺寸下的模型性能对比实验

输入尺寸	mAP@0.5IOU	帧速/(帧·s ⁻¹)
320×320	86.2	75
416×416	89.1	66
480×480	89.2	58
512×512	89.4	54
608×608	89.6	41

通过对比表 3 中原版 YOLOv3 的实验结果可以看出, 改进后的模型在检测速率上实现了明显的提升, 在相同的输入尺寸下最高可提升 10 帧·s⁻¹. 此外由于网络层级过高导致细节特征在卷积过程中损失严重, 当 YOLOv3 模型的输入尺寸降低至 480×480 及以下, 其对于实验数据集的小目标检测性能呈现为断层式下跌. 文中改进模型采取的特征融合策略能更好地保留细节特征, 针对小目标实验对象具有良好的检测性能, 在模型的输入尺寸由大至小的实验测试环境下始终维持着较高的准确度, 具有极高的鲁棒性. 本文方法的实验效果如图 5 所示.



图 5 模型的检测效果示意

4 结束语

本文将无人机的高空巡检工作和深度学习的视觉检测算法相结合, 对现有检测性能优良的 YOLOv3 的整体网络与损失函数进行改进, 以适用于正下无人机视角的挖掘机检测. 文中搭建了一个正下无人机视角的单类挖掘机目标数据集, 修改了模型的骨干网络并以 FPN 的思想设计检测网络部分, 使检测任务在大小两个尺度上进行. 为了解决样本严重不平衡问题, 在损失函数中加入了 focal loss. 文中提出的模型平均精度最高可达 89.6%, 在相同输入尺寸的条件下对比 YOLOv3 模型可实现最高 10 帧·s⁻¹ 的检测速率提升, 且具有更强的鲁棒性.

由于条件的限制, 所收集的数据集有待进一步扩容, 这对最终模型性能的表现有着显著的影响. 此外, 模型的损失函数还有待进一步调优. 为此, 在接下来的工作中, 还需进一步对样本集进行扩容, 优化损失函数, 并在保障性能的前提下继续进行模型的轻量化工作.

参考文献:

- [1] 陈志强. 浅谈变电站机器人巡检系统优化应用[J]. 工程技术(文摘版)·建筑, 2016(2):156.
- [2] Ma X X, Grimson W E L. Edge-based rich representation for vehicle classification[C]//Tenth IEEE International Conference on Computer Vision, 2005:1185-1192.
- [3] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Conference on Computer Vision and Pattern Recognition, 2005:886-893.
- [4] Kazemi F M, Samadi S, Poorreza H R, et al. Vehicle recognition using curvelet transform and SVM[C]//Fourth International Conference on Information Technology, 2007:516-521.

- [5] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.
- [6] 王健, 王晓东, 郭磊. 基于混合高斯模型的运动目标检测[J]. 宁波大学学报(理工版), 2019, 32(3):35-39.
- [7] Hinton G E, Salakhutdinov R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [8] 周晓杰, 蔡元强, 夏克江, 等. 基于火焰图像显著区域特征学习与分类器融合的回转窑烧结工况识别[J]. 控制与决策, 2017, 32(1):187-192.
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision, 2016:21-37.
- [10] 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, 2016:779-788.
- [11] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:6517-6525.
- [12] Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB/OL]. [2019-08-17]. <https://arxiv.org/abs/1804.02767>.
- [13] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//International Conference on Computer Vision, 2017:2999-3007.
- [14] 张富凯, 杨峰, 李策. 基于改进 YOLOv3 的快速车辆检测方法[J]. 计算机工程与应用, 2019, 55(2):12-20.
- [15] 戴伟聪, 金龙旭, 李国宁, 等. 遥感图像中飞机的改进 YOLOv3 实时检测算法[J]. 光电工程, 2018, 45(12):84-92.
- [16] 舒军, 吴柯. 基于改进 YOLOv3 的航拍目标实时检测方法[J]. 湖北工业大学学报, 2020(1):21-24.
- [17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.
- [18] Lin T, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:936-944.
- [19] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB/OL]. [2019-07-12]. <https://arxiv.org/abs/1502.03167>.
- [20] Everingham M, Eslami S M A, van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.
- [21] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.

Real-time excavator detection under direct UAV view based on improved YOLOv3 method

CAI Zhenyu, WANG Zekai, CHEN Tehuan*, LI Wenlai

(Faculty of Mechanical Engineering & Mechanics, Ningbo University, Ningbo 315211, China)

Abstract: A real-time excavator detection method for high-altitude patrol scenarios is proposed to meet the requirements of Unmanned Aerial Vehicle (UAV) intelligent inspection. The model is based on YOLOv3 and adopts a 43-layer backbone network. The object can be detected on two scales through feature fusion strategy. Besides, it also defines the loss function with the idea of focal loss. The experimental object in this article is an excavator under the direct UAV view. Upon the completion of the construction of the dataset, the optimal solution of the model is obtained by training the model with characteristics of the object under direct UAV view. The experimental results show that the proposed model is more accurate and robust than YOLOv3 for the same input size, and has increased by 10 frame per second in terms of frame number.

Key words: UAV view; YOLOv3; feature fusion; object detection

(责任编辑 韩 超)