

低光照下的无人机异物检测与定位^①

傅 强, 蒋雪薇, 成 鹏

(中国民用航空飞行学院 计算机学院, 德阳 618307)

通信作者: 蒋雪薇, E-mail: larotonda@qq.com



摘 要: 为解决无人机在低光照环境下的巡检过程中, 不能对场景中的异物进行识别与定位, 导致后续智能算法无法获得环境语义信息的问题. 本文提出一种将 ORB-SLAM2 算法与适用于低光照目标检测改进的 YOLOv5 模型进行信息融合的方法. 首先, 通过 RGB-D 相机自采集低光照数据集进行深度学习训练及融合算法验证. 然后, 结合关键帧信息、目标检测模块的输出结果以及相机的固有信息完成目标像素坐标提取. 最后, 通过关键帧信息和像素坐标完成目标物体相对世界坐标系的位置解算. 本文实现了低光照环境下目标物体较为准确的识别和目标物体在世界坐标系中分米级的定位, 为低光照环境下无人机智能巡检提供了一种有效的解决方案.

关键词: 视觉 SLAM; 低光照图像; 目标检测; 深度学习

引用格式: 傅强, 蒋雪薇, 成鹏. 低光照下的无人机异物检测与定位. 计算机系统应用, 2024, 33(2): 151-158. <http://www.c-s-a.org.cn/1003-3254/9400.html>

Abnormal Object Detection and Localization of Unmanned Aerial Vehicles in Low-light Environment

FU Qiang, JIANG Xue-Wei, CHENG Peng

(College of Computer Science, Civil Aviation Flight University of China, Deyang 618307, China)

Abstract: Unmanned aerial vehicles (UAVs) cannot identify and locate foreign objects in the scene during the inspection in low-light environments, resulting in the subsequent intelligent algorithms failing to obtain the environmental semantic information. To this end, this study proposes a method to fuse information from the ORB-SLAM2 algorithm with the YOLOv5 model, which is applicable to the improvement of low-light object detection. First, deep learning training and fusion algorithm validation are performed by self-collecting low-light datasets from RGB-D cameras. Then, the target pixel coordinates are extracted by combining the keyframe information, the output of the object detection module, and the inherent information of the camera. Finally, the position of the target object is solved relative to the world coordinate system by keyframe information and pixel coordinates. The study achieves more accurate recognition of target objects in low-light environments and localization of target objects in the world coordinate system at the sub-meter level, which provides an effective solution for intelligent inspection of UAVs in low-light environments.

Key words: visual SLAM; low-light image; object detection; deep learning

相较于大型工业用无人机, 小型无人机在隧道、管道等狭窄、昏暗场景更为灵活, 且具有更高的容错率, 因此被广泛应用于电力、农林、管道、厂区、隧

道等具体场景下的巡检任务^[1]. 早在 2012 年, 电力巡检就已经出现了无人机的身影, 随着相关研究的不断发展与完善, 无人机逐渐将巡检人员从高强度, 高风险,

^① 基金项目: 四川省重点研发项目 (2022YFG0027)

收稿时间: 2023-06-26; 修改时间: 2023-08-08; 采用时间: 2023-09-28; csa 在线出版时间: 2023-12-25

CNKI 网络首发时间: 2023-12-27

高压力的工作中解放出来。无人机自主执行任务时,需要具备感知陌生环境的能力。因此,在未知的环境中实现定位和建图的 SLAM 技术^[2]是无人机的必备能力之一。SLAM 技术按照传感器的不同,可分为视觉 SLAM 和激光 SLAM,在基于视觉的方法中,仅需使用直观的视觉信息,即可及时准确地完成相机运动过程中的定位和建图。近年来随着计算机视觉等技术的飞速发展,相机相对于激光而言更丰富的纹理信息以及更高的性价比等优势得到了进一步扩大。同时视觉 SLAM 与深度学习相结合的研究方向也成为了行业的研究热点之一,并在学术和企业中得到了积极的应用与发展^[3,4]。

在隧道、地下室等无人机巡检的应用场景中,都或多或少存在着信号和光线不足、光照不充分等弊端^[5]。传统视觉 SLAM 算法虽然可以在性能受挫的情况下完成定位和建图,但这些方法大多缺乏对场景中物体识别和检测的功能,不能及时对场景中出现的物体进行属性判断。通过在 SLAM 系统中加入目标识别与检测获取具备语义信息的环境信息,可为无人机在复杂环境下的自主导航和巡检任务过程中的智能决断提供信息支持。但在低光照或复杂光照环境下,图像质量下降会对目标检测的性能造成影响,导致无法准确检测环境中的目标^[6]。因此研究对低光照环境下具有鲁棒性的目标检测与目标三维定位方法,对实现无人机在低光照场景下的智能化作业具有重要意义。

定位精度在低光照环境下显著降低,这个缺点是视觉 SLAM 算法应用到户外移动的机器人中的主要障碍之一。近年来,处理在低光照条件下拍摄的图像的问题在多种不同的方法中得到了解决,其中大多数方法都集中在输入图像的预处理上,其预处理算法也从简单的数学变换^[7,8]到更复杂和更先进的方法^[9,10]。这些方法已被证明在昏暗条件下略微增加了视觉算法的灵活性,然而,它们仍然无法在极暗的环境中达到理想的结果。另一项工作^[11]利用具有自适应伽马校正和对比度自适应直方图均衡的图像预处理来克服低光条件,并通过插入关键帧和三角测量不断补充相应的映射点来保持跟踪。但是,这种方法增加了系统的运算复杂度,且该研究缺乏在黑暗条件下的实验结果,仅在光线较为充足的环境中收集的数据集上进行了测试。此外,研究者们还提出了处理夜间条件下的图像的另一种方法^[12,13]。这些方法要么使用多光谱成像,要么使用额外的红外基础设施,这也导致成本显著增加。

为解决上述问题,本文提出一种信息融合方案,将改进的 YOLOv5 目标检测模型与 ORB-SLAM2 系统有效结合,实现低光照环境下的自动目标检测与目标三维空间定位。该方法通过 YOLOv5 检测图像上目标的像素坐标,再与 SLAM 系统建立的三维地图进行匹配,获得目标精确的三维坐标。与现有单一 SLAM 或检测方法相比,该方案实现了低光照环境下异物的检测与三维空间定位功能,为无人机的智能巡检工作提供了技术支持。

1 相关技术

1.1 视觉 SLAM

传统的视觉 SLAM 按照是否需要特征点提取,分为基于特征点的间接法视觉 SLAM 和基于像素亮度值的直接法视觉 SLAM^[14]。基于特征点法的视觉 SLAM^[15]指的是对输入的图像进行特征点检测及提取,根据提取特征点的信息对相邻时间所采集的图片进行数据关联,从而对运动位姿进行估计。基于直接法的视觉 SLAM^[16,17]不用计算关键点和描述子,而是根据像素灰度信息来最小化光度误差估计出相机运动以及点的投影,把该问题当作能量最小化问题,位姿估计则通过能量函数进行迭代优化。

两种方法的优点各不相同,大多数基于特征点的 SLAM 方法如 ORB-SLAM2、DXSLAM 等^[18-21]都具有基于关键点特征提取器的前端,其中 ORB-SLAM2 采用 ORB 作为特征点描述子,对如地下室这种光照不均匀的场景具有一定程度的适应性。基于直接法的视觉 SLAM 计算量小,适应视角快速变换,但是由于直接法基于灰度不变假设,直接法对光照变化十分敏感。总体而言,特征点法虽然会消耗更多的计算资源,但整体表现更为鲁棒。

在特征点法的中 ORB-SLAM2^[22]是首个支持单目、双目和 RGB-D 相机的完整 SLAM 算法,能够实现地图重用和重定位的功能,它使用 ORB 特征点,并利用词袋模型实现特征点的加速匹配和闭环检测。该算法框架完整逻辑清晰,十分便于二次开发,本文的改进工作也在其基础上完成。

该算法具有 3 个主线程。其中跟踪线程通过查找当前图像与局部地图之间的特征匹配来跟踪每帧的位置,并且通过光束平差法 (bundle adjustment, BA) 来最小化重投影误差。局部建图线程则是通过执行局部

BA 管理局部地图并进行优化. 闭环检测线程主要是检测大的闭环, 并通过执行位姿图优化来校正漂移误差, 该线程将触发全局 BA, 即在位姿图优化之后计算整个系统最优结构和运动结果.

1.2 YOLO 目标检测

YOLO^[23]是近年来受到广泛关注和使用的是一种目标检测模型. YOLO 的主要特点在于其简单高效的检测框架, 这带来了速度快、检测精度高与泛化能力强等优点. YOLO 系列目标检测仍然存在着检测精度与速度之间的平衡问题, 通过精度和速度的对比, YOLOv5 显示出较强的特征表达能力, 在低光照场景下有更好的表现. 所以本文针对无人机低光照环境中的巡检任务, 最终采用精度较高的 YOLOv5 模型进行改进.

YOLOv5 的工作原理主要依靠一个端到端的检测网络实现, 通过特征提取、特征融合与多尺度预测实现对图像中多个类别目标的检测. 它将目标检测转化为一个回归问题来求解, 最终输出检测框的坐标, 达到检测效果. YOLOv5 使用更强的特征提取网络, 实现特征表达能力的提高, 从而对小目标与场景变化具有更强的适应性, 这是 YOLOv5 有较高精度的原因之一.

YOLOv5 的特征提取部分使用 CSPDarknet53 作为主干网络. 该网络包含多个卷积层与残差块, 可以有效提取图像丰富的语义特征. 将不同层次的特征进行串联与堆叠, 进行特征融合, 从而得到更加抽象和语义化的特征表示, 这有助于检测多个类别与尺度的目标.

2 系统设计

本文设计的针对低光照环境的目标检测与定位系统建立在 ORB-SLAM2 原有的 3 个线程的基础上, 加入了目标检测与目标三维空间定位两个新的处理过程.

程序的整体实现流程如图 1 所示, 可以分为 3 个阶段. 首先采用深度相机进行图像信息采集, 并将采集到的 RGB 图和深度图提供给 ORB-SLAM2 进行处理. 然后将 ORB-SLAM2 挑选出的关键帧所对应的 RGB

图片传递给目标检测模块, 目标检测处理后得到目标物体的像素坐标. 最后结合 ORB-SLAM2 得到的该帧位姿信息和深度图信息, 解算出目标物体的世界坐标位置.

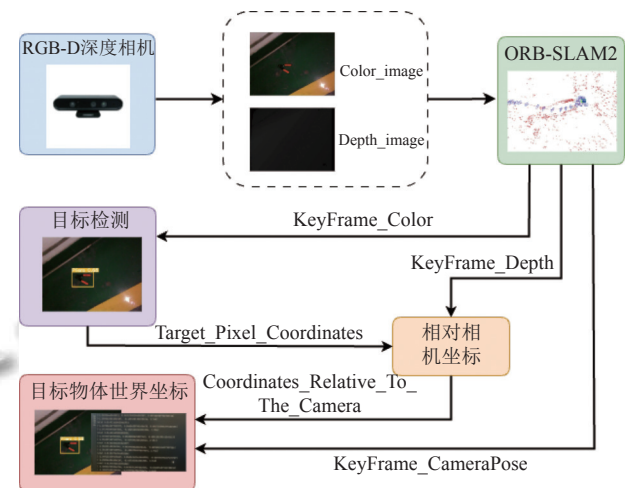


图 1 系统流程图

2.1 关键信息挑选

ORB-SLAM2 通过 Astra Pro 深度相机采集到的信息进行定位与建图, 在这一过程中 ORB-SLAM 会挑选出在跟踪过程中具有代表性的关键帧. 本文利用这一特性, 通过 ROS 系统将所挑选关键帧所对应的彩色图、深度图以及相对世界的位姿发布到参数服务器中, 供后续算法使用. 通过该方法一方面避免了所有帧都进行物体检测的无意义运算, 另一方面也记录了每一帧相机相对于世界的坐标转换关系.

2.2 目标检测模块

目标检测模块主要实现目标识别与检测. YOLOv5 是一种单阶段目标检测算法, 检测速度快, 适合实时应用. 同时, YOLOv5 有着较高的准确性, 支持多类别检测任务.

2.2.1 Backbone

YOLOv5 的 Backbone 主要分为 Conv 模块、CSPDarknet53 和 SPPF 模块, 如图 2 所示.

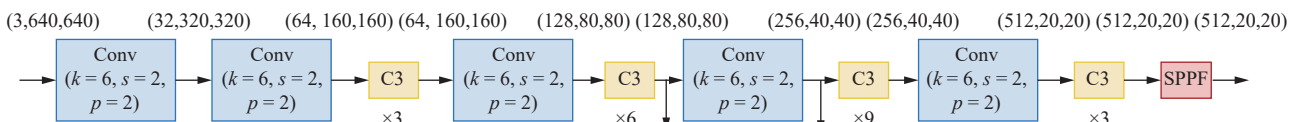


图 2 Backbone 结构图

在 Conv 模块中封装了 3 个功能: 包括卷积 (Conv2d)、batch normalization 和激活函数, 同时使用 autopad(k, p)

实现了填充效果. 其中使用 SiLU 作为激活函数, 代替了旧版本中的 LeakyReLU. SiLU 有着 ReLU 相同的激

活性能,但有着更好的稳定特性, *SiLU* 其实本质上来说是 *Sigmoid* 激活函数的一种变体. *Sigmoid*、*SiLU* 激活函数的公式如下:

$$Sigmoid(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

$$SiLU = x \times Sigmoid(x) \quad (2)$$

YOLOv4 和 YOLOv5 均借鉴了 CSPNet 的思想, 将其运用于 DarkNet53 骨干网络. YOLOv5 使用了 C3 模块, 替代了早期的 BottleneckCSP 模块.

如图 3 所示, 这两者结构作用基本相同, 均为 CSP 架构, 只是在修正单元的选择上有所不同, C3 模块包含了 3 个标准卷积层以及多个 Bottleneck 模块. C3 模块相对于 BottleneckCSP 模块所不同的是经过 Bottleneck 模块输出后的 Conv 模块被去掉了.

YOLOv5 使用了 SPPF 模块来代替空间金字塔池化 (spatial pyramid pooling, SPP) 模块^[24], YOLOv5 借鉴了 SPPNet 的思想.

SPPF 模块采用多个小尺寸池化核级联代替 SPP 模块中单个大尺寸池化核, 从而在保留原有功能, 即融合不同感受野的特征图, 丰富特征图的表达能力的情

况下, 进一步提高了运行速度^[25].

2.2.2 添加注意力机制

现有的 YOLOv5 模型在常规数据集目标检测方面有更好的效果, 但在低光照情况下表现不佳. 经实验后发现, 常规卷积针对本数据集识别效果可以通过添加 CBAM 注意力机制^[26]进行优化. CBAM 更关注于识别目标物体, 其包括两个独立模块: 通道注意力模块 (channel attention module, CAM) 和空间注意力模块 (spatial attention module, SAM). 在节约参数和算力的基础上, 保证其可以集成到现有的模块中. 本系统中将 Backbone 中的 C3 模块改进为 CBAMC3, 在进行特征提取时会增加一项注意力机制的计算, 其作用是对特征图中的不同目标进行不同维度上的注意力加权计算, 提高算法对特征图中主要特征的提取, 以此来提高目标检测算法的准确性改进后的算法框架如图 4 所示.

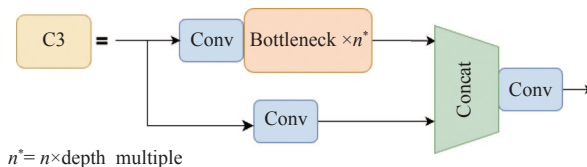


图 3 C3 结构图

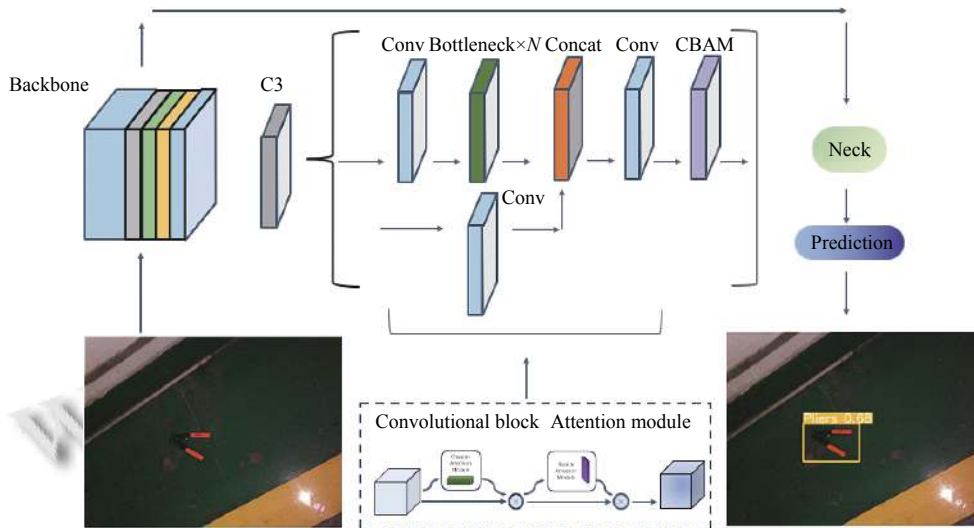


图 4 改进 YOLOv5 框架图

2.3 目标三维空间定位

目标三维空间定位模块主要分为两个步骤, 如图 5 所示.

首先收集关键帧信息和目标检测模块的输出结果, 结合相机的固有信息完成目标相对相机位姿的估计. 之后通过关键帧信息挑选所得到的相机位姿数据, 完

成目标物体相对世界坐标系的位置解算.

根据关键帧的彩色图所检测的目标点像素坐标、关键帧的深度图信息和深度因子结合相机内参, 计算出目标物体相对于相机的空间位置. 为降低深度值波动带来的误差影响, 本文采用的深度值为目标点为中心 5*5 方格的平均值. 计算目标物体相对于相机的位

姿的具体计算公式如下所示:

$$z = \frac{d}{depth} \tag{3}$$

$$x = \frac{(u - c_x)z}{f_x} \tag{4}$$

$$y = \frac{(v - c_y)z}{f_y} \tag{5}$$

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_0 q_2 - q_1 q_3) \\ 2(q_0 q_3 - q_1 q_2) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_0 q_1 - q_2 q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \tag{6}$$

其中, (x_0, y_0, z_0) 为目标物体相对于相机的空间位置, (q_0, q_1, q_2, q_3) 为相机位姿的四元数表示, (x_1, y_1, z_1) 为检测的目标相对于世界坐标系的定位。

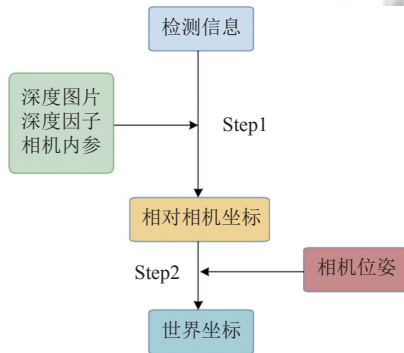


图5 三维空间定位流程图

3 实验

3.1 数据集与实验环境

3.1.1 数据集

本文数据集分为目标检测数据集和算法验证数据

其中, (u, v) 为像素坐标, d 为深度值, $depth$ 为相机的深度因子, f_x, f_y, c_x, c_y 为相机内参信息, (x, y, z) 为目标相对于相机的坐标。

最后, 结合关键帧的相机位姿和得到的目标物体相对于相机的空间位置, 通过四元数坐标旋转公式计算出目标物体相对于世界坐标. 转换公式如下所示:

集, 均使用奥比中光的 Astra Pro 深度相机 (RGB-D) 作为采集设备. 目标检测数据集通过 RGB-D 相机在地下车库采集真实数据进行制作. 算法验证数据集, 则通过在 ROS 系统中录制 RGB-D 相机的运行过程中的 rosbag 来制作。

应用于目标检测模块的数据集则需要根据 VOC 数据集的格式进行调整. 从 RGB-D 相机中提取到的数据是未完成标注的, 需要进行手动标注. 标注的文件记录了目标的类别、标记框等信息. YOLOv5 模型的输入格式要求 txt 文件, 为满足训练要求, 需要对标注好的数据集进行格式转换. 通过人工筛选后, 共保留分辨率为 640×480 的 2 005 张原始图像数据. 共计 5 个类别, 分别为 Pliers、Screwdriver、Utility knife、Hammer、Pop can.

本实验中的数据将拆分为训练集 (80%)、验证集 (20%) 两个部分, 并采用完全相同的预处理操作, 以保证训练、验证的数据一致性. 最终数据传入模型得出的预测目标并进行标记的部分图像如图 6 所示。



图6 测试集训练及预测标签部分展示

算法验证的数据集, 提取并保存的 rosbag 包中包含相机运动过程中采集的 RGB 图像和深度图像信息,

使其可以在计算机上复现该运动过程. 提取的数据集包含 Depth、RGB 两个话题的信息, rosbag 包在保存

图像信息的同时,也保存了二者之间时间上的对应关系.

3.1.2 实验环境

本文算法实现采用的软件环境为:编程语言 Python 3.8.16,深度学习框架 PyTorch 1.8.2+cu111.实验使用的计算机硬件配置为:处理器 Intel(R) Core(TM) i7-12700 CPU@2.10 GHz,显卡 NVIDIA GeForce RTX 3070Ti 16 GB.

3.2 评价指标

常见目标检测问题的评价指标有:平均准确度均值 (mAP)、速度指标 (FPS)、准确率 (Precision)、召回率 (Recall) 等. 本文将使用 mAP、Precision、Recall 作为最终算法评估标准. 其计算公式分别如下所示:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$mAP = \frac{\sum_{i=1}^n (Recall_{x_2} - Recall_{x_1}) \times MaxPrecision}{n} \quad (9)$$

其中, TP 表示被正确分类的正样本数量, FP 表示被错误分类的正样本数量, FN 表示被错误分类的负样本数量, x_2 表示当前样本 x_1 表示上个样本, $MaxPrecision$ 表示当前点以及后边所有 Precision 值最大的点.

针对目标定位模型的评估,本文考虑实时性和定位精度两个方面. 定位模型在实时性方面的表现是指其能否在短时间内对目标物体的位置进行有效的估计. 智能无人机在实际应用中,往往需要快速移动,这就要求目标物体的定位具备一定的实时性. 为了评价定位模型的实时性,本文多次测量求平均的方式测试了模型的性能.

定位模型的精度是指目标物体位置估计值与实际位置的误差大小. 精度评价是对定位模型最基本的评价指标之一,而且精度是所有定位应用的要素之一. 为了评价定位模型的精度,本文采用比较精确的实验方法来进行测试. 在实验过程中,将目标物体放置在已知位置上,并通过定位系统来估计该目标的位置,然后与实际位置进行比较来评估模型的精度. 通过欧氏距离完成了对测量精度的评估,评估公式如下所示:

$$error = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \quad (10)$$

3.3 实验结果与分析

实验结果表明,本文提出的方法能有效实现低光照场景下异物检测与定位功能,其检测精度、实时性以及定位精度能满足无人机异物巡检的需求.

本文改进的模型与现有经典目标检测网络方法进行对比,结果如表 1 所示. 从表 1 数据可以看出,本文提出的模型在精度上有较好的表现,能有效适应低光照环境. DDOD 和 Faster-RCNN, 尽管精度较高,但是推理速度较慢. 综上所述,本文提出的改进模型更能满足低光照环境下无人机巡检任务的需求.

表 1 检测精度对比

模型	mAP@0.5	mAP@0.5:0.95
FPN ^[27]	0.796	0.525
Faster-RCNN ^[28]	0.834	0.621
DDOD ^[29]	0.816	0.603
SSD512 ^[30]	0.712	0.531
YOLOv3 ^[31]	0.421	0.213
YOLOv5	0.499	0.301
Our	0.874	0.679

模型在学习过程中的数据变化,最终预测得出的混淆矩阵、检测准确率如图 7 所示.

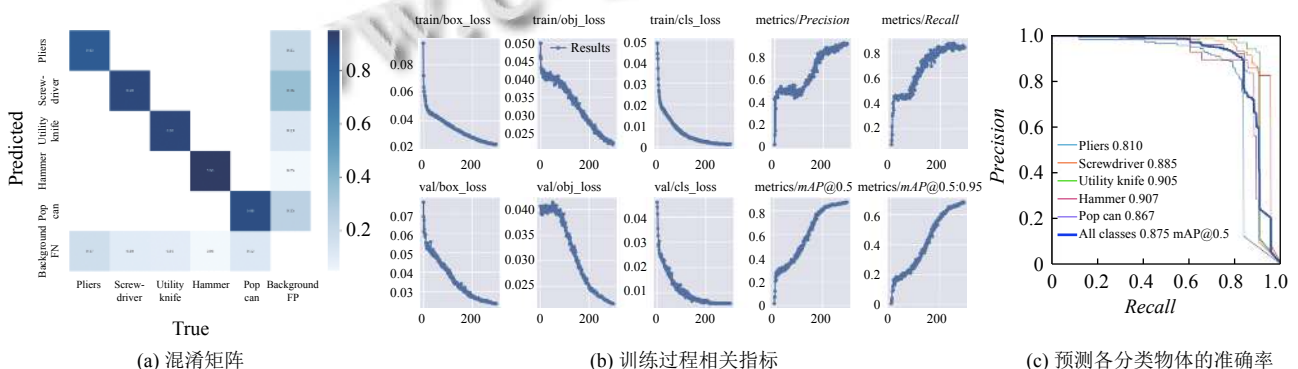


图 7 检测算法学习过程与最终结果展示图

通过混淆矩阵可以看出,目标之间相互影响较小,但环境的影响较大,导致目标检测会出现误检和漏检

的情况. 其中绿色目标最容易出现漏检情况,原因可能在于与场景颜色过于相似. 从结果分析,排除了可见光

弱等影响,目标检测本身还是会收到背景颜色和目标颜色重叠的干扰.同时由于自采数据集筛选后存在类别分布不均匀的情况,导致目标检测的召回率和准确率没有达到最理想的情况.

ORB-SLAM2 可以达到实时跟踪和映射,其帧率可以达到 20–30 Hz.它使用 FAST 关键点检测器和 ORB 描述子,计算量小且易于硬件加速,这使得其可以快速提取并匹配特征点,满足实时应用.本文所提出的系统在基于 ORB-SLAM2 上实现异物定位,这一部分为相对于原 ORB-SLAM2 多出来的运算任务.本实验选用 10 个目标分别进行 20 次测试,其系统检测到物体并反馈定位的平均响应时间如表 2 所示.

表 2 系统检测与定位响应时间

目标序号	Time (s)
1	0.014 7
2	0.014 4
3	0.013 6
4	0.013 7
5	0.019 4
6	0.019 5
7	0.013 4
8	0.013 1
9	0.019 5
10	0.014 4
平均时间	0.015 57

从表 2 可以看出,每完成一个目标的检测与定位,系统反馈定位平均时间仅为 15 ms 左右,不影响整体算法的实时性.

ORB-SLAM2 可以达到厘米级的绝对定位精度,本文使用的深度相机也可完成对深度值的毫米级测量,结合检测和定位算法可实现对目标物体的厘米级定位.本文通过将目标物体放在预先测量好分别为 1.4 m、0.5 m、2.8 m 的位置,通过定位系统进行估计,并对多次测量的结果求平均,得出如表 3 所示的实验结果,从实验结果可以看出本文提出的目标定位算法可以达到分米级别的精度.

4 总结

本文提出一种信息融合方案,将 YOLOv5 目标检测模型与 ORB-SLAM2 系统有效结合,实现低光照环境下的目标检测与其三维空间定位.该方法通过 YOLOv5 检测图像上目标的二维位置,与 SLAM 系统建立的三维地图进行匹配,获得目标的三维空间坐标.

与现有单一 SLAM 或目标检测方法相比,该方案实现了低光照环境下目标物体较为准确的识别和目标物体在世界坐标系中分米级的定位,为低光照环境下无人机智能巡检提供了一种有效的解决方案.

表 3 测试定位坐标的均值与误差

序号	x (m)	y (m)	z (m)	误差 (m)
1	1.351 0	0.401 5	2.765	0.115 448 040
2	1.371 5	0.402 8	2.906	0.146 615 449
3	1.354 3	0.435 3	2.871	0.106 374 715
4	1.350 6	0.447 1	2.765	0.080 397 575
5	1.385 1	0.479 8	2.743	0.062 282 020
6	1.435 4	0.570 5	2.788	0.079 796 053
7	1.383 8	0.535 3	2.797	0.038 955 487
8	1.431 0	0.604 7	2.834	0.114 363 849
9	1.428 7	0.589 4	2.811	0.094 535 972
10	1.345 5	0.572 7	2.7	0.135 113 064
Average	1.383 69	0.503 91	2.798	0.097 388 222

本文还有一些不足的地方,未来的工作还需要对目标物体的检测精度和定位精度上做进一步的改进.在 ORB-SLAM2 对低光照的适应性上需要做更深入的研究,考虑融合图像增强算法以减小低光照图像质量下降带来的影响.同时,可以在深度值选取部分加入滤值来取代平均值,进一步降低由于原本深度值偏差对定位精度的影响.

参考文献

- 1 段海滨,何杭轩,赵彦杰,等. 2021 年无人机热点回眸. 科技导报, 2022, 40(1): 215–227.
- 2 Chen WF, Shang GT, Ji AH, *et al.* An overview on visual slam: From tradition to semantic. *Remote Sensing*, 2022, 14(13): 3010. [doi: 10.3390/rs14133010]
- 3 Ondruš J, Kolla E, Vertaľ P, *et al.* How do autonomous cars work? *Transportation Research Procedia*, 2020, 44: 226–233.
- 4 Jennings D, Figliozzi M. Study of road autonomous delivery robots and their potential effects on freight efficiency and travel. *Transportation Research Record: Journal of the Transportation Research Board*, 2020, 2674(9): 1019–1029. [doi: 10.1177/0361198120933633]
- 5 董康兴,包振山,任智慧,等. 智能井场巡检机器人结构设计. *机械设计*, 2022, 39(S2): 31–34.
- 6 Savinykh A, Kurenkov M, Kruzhkov E, *et al.* DarkSLAM: GAN-assisted visual SLAM for reliable operation in low-light conditions. *Proceedings of the 95th IEEE Vehicular Technology Conference: (VTC2022-Spring)*. Helsinki: IEEE, 2022. 1–6.
- 7 Huang SC, Cheng FC, Chiu YS. Efficient contrast

- enhancement using adaptive Gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 2013, 22(3): 1032–1041. [doi: [10.1109/TIP.2012.2226047](https://doi.org/10.1109/TIP.2012.2226047)]
- 8 Pizer SM, Amburn EP, Austin JD, *et al.* Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 1987, 39(3): 355–368. [doi: [10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X)]
- 9 Dusmanu M, Rocco I, Pajdla T, *et al.* D2-Net: A trainable CNN for joint description and detection of local features. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 8084–8093.
- 10 Revaud J, Weinzaepfel P, De Souza C, *et al.* R2D2: Repeatable and reliable detector and descriptor. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2019. 1113.
- 11 Karpyshev P, Ilin V, Kalinov I, *et al.* Autonomous mobile robot for apple plant disease detection based on CNN and multi-spectral vision system. *Proceedings of the 2021 IEEE/SICE International Symposium on System Integration (SII)*. Iwaki: IEEE, 2021. 157–162.
- 12 Kalinov I, Petrovsky A, Agishev R, *et al.* Impedance-based control for soft UAV landing on a ground robot in heterogeneous robotic system. *Proceedings of the 2021 International Conference on Unmanned Aircraft Systems (ICUAS)*. Athens: IEEE, 2021. 1653–1658.
- 13 Şimşek B, Satir S, Bılge HŞ. Performance comparison of direct and feature based vSLAM algorithms. *Proceedings of the 29th Signal Processing and Communications Applications Conference (SIU)*. Istanbul: IEEE, 2021. 1–4.
- 14 Ferrera M, Eudes A, Moras J, *et al.* OV² SLAM: A fully online and versatile visual SLAM for real-time applications. *IEEE Robotics and Automation Letters*, 2021, 6(2): 1399–1406. [doi: [10.1109/LRA.2021.3058069](https://doi.org/10.1109/LRA.2021.3058069)]
- 15 Min ZX, Dunn E. VOLDOR⁺ SLAM: For the times when feature-based or direct methods are not good enough. *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an: IEEE, 2021. 13813–13819.
- 16 任伟建, 高强, 康朝海, 等. 移动机器人同步定位与建图技术综述. *计算机测量与控制*, 2022, 30(2): 1–10, 37.
- 17 Mur-Artal R, Montiel JMM, Tardos JD. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 2015, 31(5): 1147–1163. [doi: [10.1109/TRO.2015.2463671](https://doi.org/10.1109/TRO.2015.2463671)]
- 18 Li DJ, Shi XS, Long QW, *et al.* DXSLAM: A robust and efficient visual SLAM system with deep features. *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas: IEEE, 2020. 4958–4965.
- 19 Zhao YP, Vela PA. Good feature matching: Toward accurate, robust VO/VSLAM with low latency. *IEEE Transactions on Robotics*, 2020, 36(3): 657–675. [doi: [10.1109/TRO.2020.2964138](https://doi.org/10.1109/TRO.2020.2964138)]
- 20 Campos C, Elvira R, Rodríguez JJG, *et al.* ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 2021, 37(6): 1874–1890.
- 21 Mur-Artal R, Tardós JD. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 2017, 33(5): 1255–1262.
- 22 Jiang PY, Ergu D, Liu FY, *et al.* A review of YOLO algorithm developments. *Procedia Computer Science*, 2022, 199: 1066–1073.
- 23 周彦, 孟江南, 吴佳, 等. 改进 YOLOv5s 的钢板表面缺陷检测算法. *激光与光电子学进展*, 2023, 60(4): 0415009.
- 24 乔健, 陈能达, 伍雁雄, 等. 融合注意力机制的金属锅圆柱表面缺陷检测. *光学精密工程*, 2023, 31(3): 404–416.
- 25 赵振兵, 王帆帆, 刘良帅, 等. 基于注意力特征融合 YOLOv5 模型的无人机输电线路航拍图像金具检测方法. *电测与仪表*, 2023, 60(3): 145–152.
- 26 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018, 3–19.
- 27 Ghiasi G, Lin TY, Le QV. NAS-FPN: Learning scalable feature pyramid architecture for object detection. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 7029–7038.
- 28 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. 2015. 91–99.
- 29 Chen ZH, Yang CHY, Li QF, *et al.* Disentangle your dense object detector. *Proceedings of the 29th ACM International Conference on Multimedia*. ACM, 2021. 4939–4948.
- 30 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21–37.
- 31 Redmon J, Farhadi A. YOLOv3: An incremental improvement. *arXiv:1804.02767*, 2018.

(校对责编: 牛欣悦)