

基于嵌入式视觉的无人机目标定位系统

张君轶¹ 陈仁文¹ 刘 飞¹ 萧志鹏²

(1. 南京航空航天大学机械结构力学及控制国家重点实验室 南京 210016; 2. 重庆悠弥德科技有限公司 重庆 400000)

摘 要: 基于传统算法的无人机目标定位系统精确度较低且容易受到光照条件干扰, 针对该问题提出了一种基于嵌入式视觉的无人机目标定位系统。使用深度可分离卷积替换 YOLOv2 原始模型中的 Darknet-19 骨干网络, 大幅减小模型体积; 引入 RepVGG 模块提取复杂特征, 以提升检测精度。将模型部署到嵌入式端并进行了性能测试, 结果表明改进 YOLOv2 算法的网络模型检测精度达到了 96.7%, 检测速度达到 25 fps, 解决了传统算法难以处理光照变化的问题并且有明显的性能提升。设计完成了上机试验, 试验结果验证了无人机目标定位系统的有效性和可靠性。

关键词: 嵌入式视觉; 深度可分离卷积; RepVGG; 目标定位

中图分类号: TP181 **文献标识码:** A **国家标准学科分类代码:** 520.20

UAV target positioning system based on embedded vision

Zhang Junyi¹ Chen Renwen¹ Liu Fei¹ Xiao Zhipeng²

(1. State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. Chongqing Youmisi Technology Co., Ltd., Chongqing 400000, China)

Abstract: The UAV target positioning system based on traditional algorithms has low accuracy and is easily interfered by light conditions. Aiming at this problem, a UAV target positioning system based on embedded vision is proposed. The Darknet-19 backbone in the original YOLOv2 is replaced with depthwise separable convolutions, which greatly reduces the model size. The RepVGG block is introduced to extract complex features to improve detection accuracy. The model is deployed to the embedded system and the performance test is carried out. The results show that the model detection accuracy of the improved YOLOv2 algorithm reaches 96.7%, and the detection speed reaches 25 fps, which solves the problem that the traditional algorithm is difficult to deal with illumination changes and has obvious performance improvement. Test on UAV is designed and completed, the test results verified the effectiveness and reliability of the UAV target positioning system.

Keywords: embedded vision; depthwise separable convolution; RepVGG; target positioning

0 引 言

近年来,随着科学技术的发展进步,无人机(unmanned aerial vehicle, UAV)技术快速发展成熟,应用范围也不断延伸,尤其是在国防军事领域的应用非常广泛,经常代替人来执行战术侦察、窃取情报、通信中继、战术轰炸攻击等危险性极高的任务。随着无人机在现代战争中的广泛应用,无人机目标定位这一关键技术成为了获取战场目标信息和把握战场态势的重要手段^[1]。近些年机器视觉和嵌入式硬件快速发展进步,基于嵌入式机器视觉的

无人机目标定位方法成为热门研究方向。

余昌叶^[2]结合嵌入式系统和机器视觉设计了一个小飞机控制平台,其中在 PC 端完成图像处理与通信进程,并在嵌入式模块中进行飞行控制。洪富祥等^[3]使用搭载 Raspbian 操作系统的树莓派,通过机器学习算法实现了无人机对同心圆靶的目标定位和精准降落。陶孟卫等^[4]设计了一种由多组不同半径比的同心圆组成的视觉标识,通过圆心求解方法提取标识信息,实现了无人机自主降落。

由于图像包含的信息量大,类似上述的传统的机器视觉无人机目标定位方法都需要依靠 PC 端或机载计算机,

成本高且使用过程复杂。张义桢等^[5]提出了一种基于二维码图像识别的无人机目标定位方法,吴文秀等^[6]使用计算量更小的 AprilTag^[7],提出了一种基于 OpenMV 视觉模块的无人机目标定位方法。姜晓栋^[8]设计了四旋翼无人机单元和采用 AprilTag 算法的机器视觉单元,并将两个单元有机结合,在嵌入式平台上实现了基于机器视觉的目标定位,赋予了无人机使用视觉定位追踪目标的能力。然而,上述方法非常依赖于机器视觉的实现效果,当无人机处于夜晚或光线较暗的环境时,将大大降低机器视觉的识别效果,进而导致基于机器视觉的无人机自主降落控制系统的抗干扰能力较差^[9]。

为了解决上述问题,本文提出了一种基于嵌入式视觉的无人机目标定位系统。考虑到嵌入式设备计算成本高和延迟大的问题,以 YOLOv2 算法为基础引入了深度可分离卷积和 RepVGG 模块,在保证模型精确度的同时大幅减小模型参数量;设计制作了目标定位模块,将模型部署到目标定位模块并进行多组实验测试模型性能;实现了完整的无人机目标定位系统,并完成上机试验。

1 无人机目标定位系统设计

本文设计的无人机目标定位系统的典型应用场景如图 1 所示,先通过远程基站控制无人机进入目标活动区域,然后无人机自主搜索该区域并精确定位目标。



图 1 无人机目标定位系统典型应用场景

本文以高精度、实时性、低能耗为目标,设计响应速度快、稳定性强的无人机目标定位系统。根据任务所需的性能指标,要求目标检测精度达到 95% 以上,串口通信频率达到 10 次/s 且至少有一次是有效的目标位置信息。系统整体框图如图 2 所示。

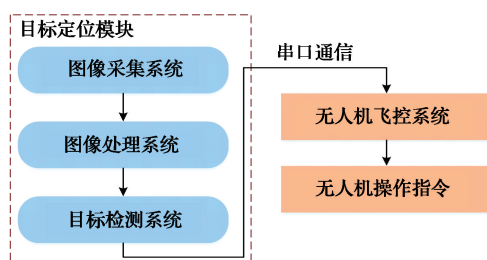


图 2 系统整体设计

1.1 硬件系统设计

本文使用 RISC-V 架构的 Kendryte K210 芯片作为嵌入式视觉的 MCU, K210 具有良好的功耗性能、稳定性和可靠性^[10], 其内置的神经网络处理器 (KPU) 算力可达到 0.8 TFLOPS, 能够通过硬件加速实现嵌入式端的神经网络模型推理。K210 芯片的架构如图 3 所示。

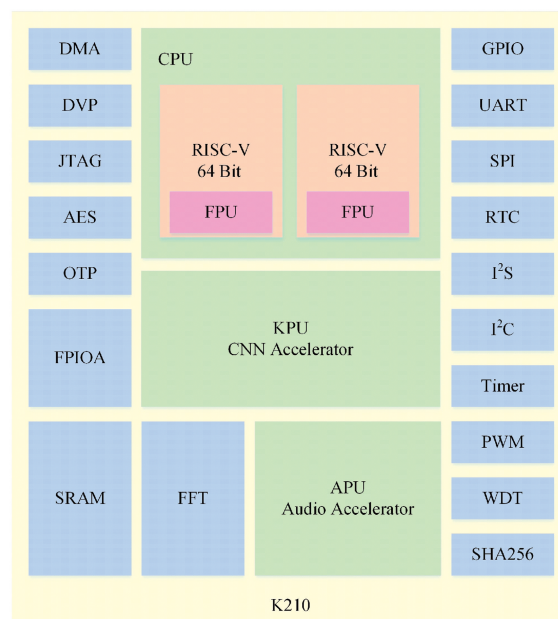


图 3 K210 芯片架构

目标定位模块的硬件系统以搭载 K210 芯片的 Maix Bit 开发板为核心, 同时与实现各项功能所必需的硬件模块相连接, 以实现硬件系统的模块化。

目标定位模块实物如图 4 所示, 重量仅 57.6 g。针对夜间目标定位的需求, 本文使用红外夜视灯和红外摄像头组成图像采集系统。USB 转 TTL 模块用于开发板供电和串口通信, LCD 显示屏用于调试设备时显示检测结果。

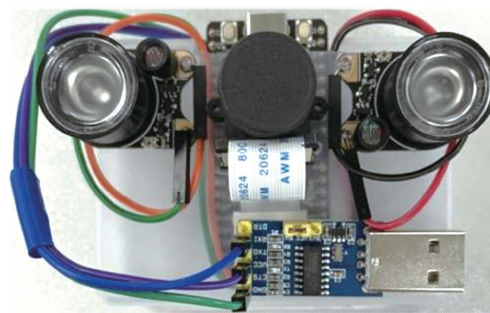


图 4 目标定位模块

1.2 软件系统设计

在嵌入式 MCU 中是无法训练神经网络的, 而在实际的应用场景中也只需要 MCU 能进行模型的推理即可。近年来随着深度学习模型轻量化研究的进展, 通过

使用 PC 端的 GPU 训练神经网络模型,然后通过一系列的量化和压缩操作,将模型数据体积大大减小后,部署于嵌入式端的技术已经逐渐趋于成熟。软件方案流程如图 5 所示。

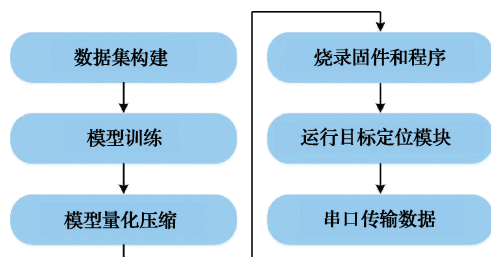


图 5 软件方案流程

针对部署于嵌入式端的参数要求,本文以 YOLOv2 算法为基础,设计了轻量化的神经网络,用于训练自主构

建的数据集。将模型文件量化压缩后部署到 MCU 芯片,运行目标检测程序并通过串口传输数据。

2 改进的 YOLOv2 算法

无人机机载平台硬件资源条件有限,结构简约速度更快的一阶段方法更适合应用于无人机平台^[11]。YOLO 是一种单阶段式(one-stage)目标检测算法,将目标边界定位问题转换成回归问题,具有速度快、可移植性好的特点。由于无人机目标定位任务对实时性和轻量化性能要求较高,本文综合考虑了 YOLO 系列算法的准确性、效率及模型大小,基于 YOLOv2 算法^[12]对神经网络进行了改进。

改进的 YOLOv2 模型结构如图 6 所示。Backbone 部分使用深度可分离卷积替换 YOLOv2 原始模型中的 Darknet-19 骨干网络,以减少网络参数;Neck 部分引入两个 RepVGG 模块,特征相加后传递到预测层;YOLOv2 Head 部分根据输入特征做出预测。

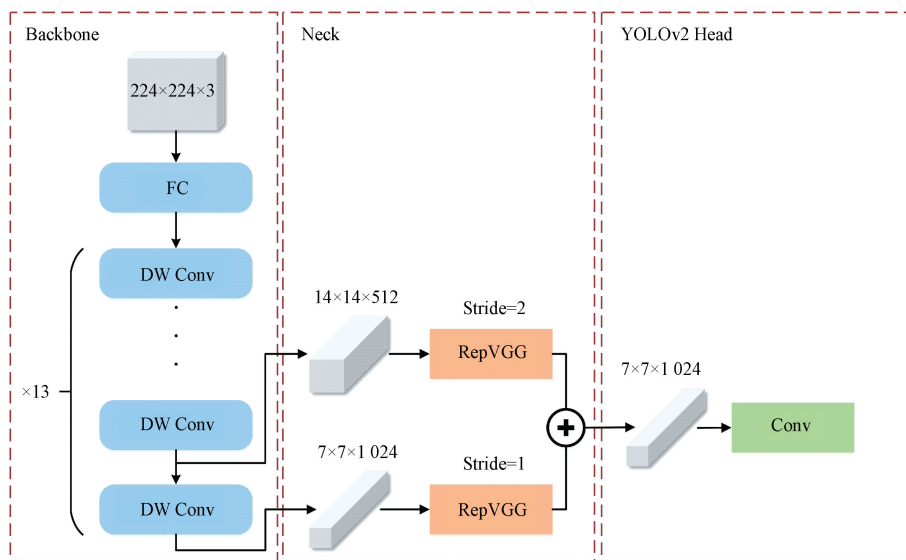


图 6 改进的 YOLOv2 模型结构

2.1 深度可分离卷积

可分离卷积是一种算法层压缩加速提高学习网络运行速度的方法,运用可分离卷积的实现深度学习网络的轻量化,提高学习网络的运行速度^[13]。深度可分离卷积是轻量化网络 MobileNet^[14]的核心,在特征提取效果上与传统卷积相同。将骨干网络中的传统卷积用深度可分离卷积代替,可以大幅减少模型的参数量(Params)和计算量(floating point operations, FLOPs),结构如图 7 所示。深度可分离卷积由一个负责空间滤波的深度卷积和一个负责线性组合的点卷积组合而成,将标准卷积分成了两步。

根据如下推导和对比可以得出,相比于标准卷积,深度可分离卷积使网络参数和模型体积减少,从而能够显著提高网络的检测速度。

假定每个标准的卷积层输入 $x_0 \times y_0 \times p_0$ 的特征图,

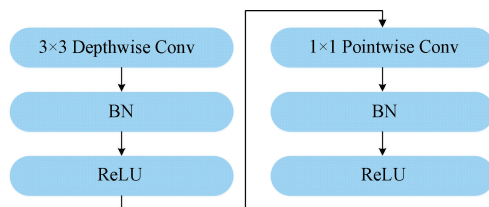


图 7 深度可分离卷积

并得到一个 $x_1 \times y_1 \times p_1$ 的输出特征图。其中 x_0, y_0 表示输入特征图的宽和高, x_1, y_1 表示输出特征图的宽和高, p_0, p_1 表示输入和输出的通道数。标准卷积层的卷积核 K 大小为 $k \times k \times p_0 \times p_1$, 其中 k 为卷积核空间尺度。假设步长和填充为 1, 则标准卷积的输出计算公式如下:

$$Out_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot In_{k+i-1,l+j-1,m} \quad (1)$$

标准卷积和深度可分离卷积的参数数量及计算量公式如下:

$$Params_{standard} = k^2 p_0 p_1 \quad (2)$$

$$Params_{dw} = k^2 p_0 + p_0 p_1 \quad (3)$$

$$FLOPs_{standard} = k^2 \cdot p_0 p_1 x_1 y_1 \quad (4)$$

$$FLOPs_{dw} = k^2 \cdot p_0 x_1 y_1 + p_0 p_1 x_1 y_1 \quad (5)$$

深度可分离卷积相比于传统卷积的计算量缩减如下:

$$\frac{k^2 \cdot p_0 x_1 y_1 + p_0 p_1 x_1 y_1}{k^2 \cdot p_0 p_1 x_1 y_1} = \frac{1}{p_1} + \frac{1}{k^2} \quad (6)$$

本文使用 3×3 的深度可分离卷积,相较于标准卷积可以减少 8~9 倍的计算量。

2.2 RepVGG 模块

2021 年, Ding 等^[15] 提出了 RepVGG 网络结构。RepVGG 吸收了残差网络的优点,在深层网络中可以有效解决训练时梯度消失的问题,使网络收敛得更快。本文引入 RepVGG 模块作为神经网络的 Neck 部分,对 Backbone 提取的特征进行再处理,通过提取更复杂的特征来提升检测模型的精确度。

RepVGG 模块结构如图 8 所示, RepVGG 模块由 3×3 和 1×1 的卷积操作以及 ReLU 激活函数组成。其中有两种残差结构:一种仅包含 1×1 卷积分支,另一种包含 1×1 卷积分支和 Identity 恒等映射分支。多分支的残差结构为模型增加了两条梯度流通的路径,可有效解决模型的退化问题和梯度消失问题,使模型在训练阶段更易于收敛且精度更高。

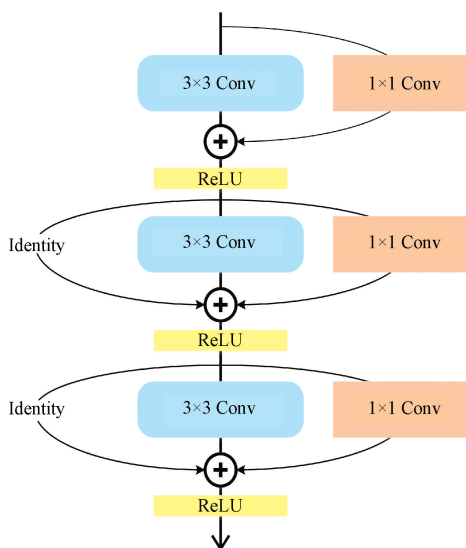


图 8 RepVGG 模块

3 实验与结果分析

3.1 数据集构建和模型训练

本文使用搭配补光灯的红外镜头采集样本,共采集关键帧图像 600 张,其中白天和夜间各 300 张,图像分辨率为 224×224 ,样本中包含目标物和干扰物两个类别。如

图 9 所示,本实验自定义的目标物直径 10 cm,贴有反光条便于夜间目标定位;干扰物与目标物大小形状相似。

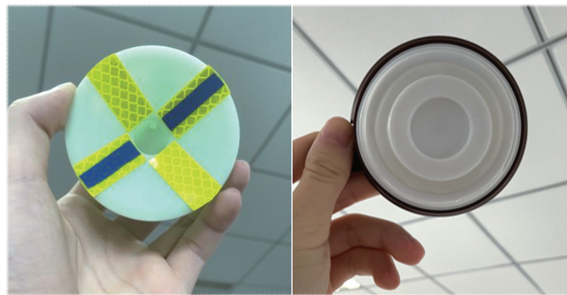


图 9 目标物和干扰物

本实验的训练平台使用的 CPU 为 AMD Ryzen5-3600@3.60 GHz $\times$ 6, GPU 为 NVIDIA GeForce RTX 2060S,操作系统为 Ubuntu20.04,模型训练基于 TensorFlow2.9.1 深度学习框架。

3.2 模型部署与性能测试

模型训练完成后,本文对其进行了量化压缩并部署到了嵌入式系统中,模型体积仅 1.8 MB。

为测试部署后模型的性能,本文进行了相关的实验验证。性能评价指标采用检测速度和检测精度(precision),检测速度即每秒检测的图像帧数,检测精度计算公式如下:

$$precision = \frac{T}{T+F} \times 100\% \quad (7)$$

式中: T 和 F 分别表示被正确检测和错误检测的视频帧数量。在测试中跟踪每一帧图像及其前一帧图像的目标定位结果,规定若两者的位置坐标差超过 20 pixels,则判定该视频帧为错检。

实验分别在白天和夜间对部署后的模型进行了测试,测试距离取 0.2、0.5 和 1.0 m,并且检验了相似目标干扰下的模型性能,图 10 所示为嵌入式设备 LCD 屏幕上截取的测试结果。

可以看出,部署到嵌入式端的模型能够在不同光照强度和不同距离下检出并精确定位目标,在相似目标干扰下可以准确区分出目标物和干扰物。每个视频帧的检测时间约为 40 ms,即检测速度达到了 25 fps,满足实时性要求。

为了对比传统算法和改进前后的 YOLOv2 算法在嵌入式端的目标定位性能,本文在树莓派 4B 平台上使用基于 OpenCV 的 SURF(speeded up robust features)算法实现了相同目标的检测定位。SURF 算法^[16] 是一种稳健的局部特征点检测和描述算法,基本思路分为 3 步:局部特征点的提取、特征点的描述和特征点的匹配。实验中采用相同的性能指标,测试距离同样选取 0.2、0.5 和 1.0 m,测试结果如图 11 所示。

表 1 为 3 种算法的性能指标对比,测试结果表明,基

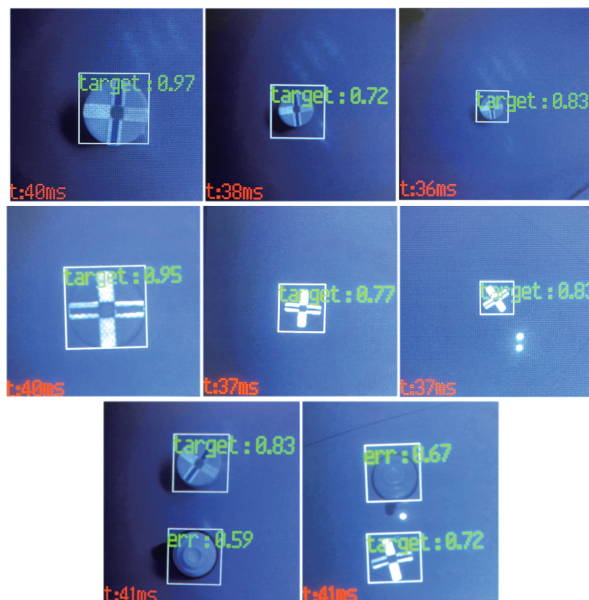


图10 模型性能测试结果

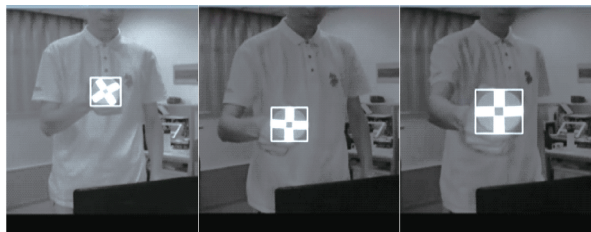


图11 SURF算法测试结果

于 YOLOv2 的改进算法相比于 SURF 算法有显著的性能提升。在检测速度指标上提升了约 10 fps,这是由于深度可分离卷积大幅减小了网络模型的参数量,并且 K210 内置的神经网络处理器对模型推理有加速作用。在检测精度方面,改进的 YOLOv2 算法比 SURF 算法提升了约 5%~7%。

由测试结果可以看出,引入 RepVGG 模块后的 YOLOv2 算法提升了 2.5%检测精度,而在检测速度上有小幅降低,可见 RepVGG 模块能够更好地利用骨干网络

表1 性能指标对比

算法	SURF			DW Conv+YOLOv2			DW Conv+RepVGG+YOLOv2		
测试距离/m	0.2	0.5	1.0	0.2	0.5	1.0	0.2	0.5	1.0
检测速度/fps	15.2	16.6	17.3	25.1	26.0	27.1	24.5	25.2	26.7
检测精度/%	92.3	89.2	88.1	95.5	94.8	92.4	97.6	97.1	95.3
平均检测精度/%		89.9			94.2			96.7	

提取的图像特征。

另外,测试结果表明,改进的 YOLOv2 算法随着测试距离的增加,检测精度有明显的下降。这是因为 YOLOv2 只有一个检测头,对小目标的检测效果不佳。

综合上述实验可以看出,由于传统算法不善于处理视角变化和光照变化,因此容易在目标移动时出现错检。而基于 YOLOv2 的改进算法在各项性能指标上有显著的提升,并且不易受到光照条件的影响,更加适用于无人机目标定位的复杂应用场景。

3.3 上机试验

本文搭建了无人机目标定位系统试验平台,并完成了上机试验。将部署模型后的目标定位模块挂载在四旋翼无人机的下方,组成无人机目标定位系统。上机试验平台如图 12 所示。

目标定位模块与无人机飞控系统之间使用 UART 串口通信,目标定位模块每检测到一次目标,就输出一次目标在相机坐标系下的 x 、 y 坐标并通过串口传输到无人机飞控系统,飞控系统根据输入的目标位置信息改变无人机进给方向,直至目标位置接近相机坐标系原点,即无人机收敛于目标的正上方。

测试时,控制无人机在 1.0 m 的水平高度上低速接近目标物,观察无人机的收敛情况。白天和夜间的多次试验



图12 上机试验平台

结果表明,当镜头视野覆盖到目标物之后,无人机能够做出相应机动,不断接近目标并最终停留在目标的正上方。

上机试验结果证明了本文设计的无人机目标定位系统具有良好的有效性和可靠性。

4 结论

针对传统的无人机目标定位方法精确度较低且容易受到光照条件干扰的问题,提出了一种基于嵌入式视觉的无人机目标定位系统。通过网络模型部署后的性能测试和上机试验,得出如下结论。

本文使用深度可分离卷积替换 YOLOv2 原始模型中的 Darknet-19 骨干网络,并引入 RepVGG 模块更好地利用骨干网络提取的图像特征。模型体积缩减至 1.8 MB,易于部署到嵌入式端。实验结果表明,改进后的算法模型的平均检测精度达到了 96.7%,比 SURF 算法提升了 6.8%,检测速度达到 25 fps,满足实时性要求,并且解决了传统算法中容易受到光照条件干扰的问题。通过上机试验结果表明,无人机目标定位系统性能良好,为无人机目标定位技术的发展提供了理论依据和技术方案。

参考文献

- [1] 邵慧. 无人机高精度目标定位技术研究[D]. 南京:南京航空航天大学,2014.
- [2] 余昌叶. 基于嵌入式系统与机器视觉的小飞机控制平台设计与实现[D]. 哈尔滨:哈尔滨工业大学,2014.
- [3] 洪富祥,丘仲锋,孙亚坤. 基于机器视觉的无人机同心圆靶精准降落方法: CN110618691A[P]. 2019.
- [4] 陶孟卫,姚宇威,元海文,等. 无人机自主降落视觉标识设计及位姿测量方法[J]. 仪器仪表学报,2022, 43(5):155-164.
- [5] 张义楨,向婕,唐立军. 基于二维码图像识别的无人机定位方法研究[J]. 计算机应用与软件,2019,36(9): 117-121.
- [6] 吴文秀,关志全,汪磊. 基于 openmv 的无人机定点降落视觉识别技术研究[J]. 智能计算机与应用,2021, 11(1):41-45.
- [7] OLSON E. AprilTag: A robust and flexible visual fiducial system [C]. 2011 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2011.
- [8] 姜晓栋. 基于机器视觉的无人机对移动目标追踪系统研究[D]. 苏州:苏州大学,2019.
- [9] 杨岳航,陈武雄,朱明,等. 基于机器视觉的无人机自主着陆技术[J]. 国外电子测量技术,2020,39(4): 57-61.
- [10] 种丹丹. 基于 RISC-V 的开源芯片生态发展现状及未来机遇[J]. 中国集成电路,2021,30(8):25-30.
- [11] 韩玉洁,曹杰,刘琨,等. 基于改进 YOLO 的无人机对地多目标检测[J]. 电子测量技术,2020,43(21): 19-24.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.
- [13] 王云艳,周志刚,罗帅. 基于数据增强的太阳能电池片缺陷检测[J]. 电子测量与仪器学报,2021,35(1): 26-32.
- [14] GHOSH T, ABEDIN M, REZA S, et al. Bangla handwritten character recognition using mobilenetv1 architecture[J]. Bulletin of Electrical Engineering and Informatics, 2020, 9(6): 2547-2554.
- [15] DING X, ZHANG X, MA N, et al. RepVGG: Making VGG-style convnets great again[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021:13728-13737.
- [16] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up robust features[C]. Proceedings of the 9th European Conference on Computer Vision-Volume Part I. Springer-Verlag, 2006.

作者简介

张君轶,博士研究生,主要研究方向为仪器科学与技术、目标检测与目标跟踪等。
E-mail:zhangjunyi@nuaa.edu.cn