

基于Maixcam视觉模块与大语言模型联动的六足机器人设计

邓庭辰, 王启文, 陈露露, 赵益康

盐城工学院电气工程学院, 江苏 盐城

收稿日期: 2025年4月25日; 录用日期: 2025年6月27日; 发布日期: 2025年7月8日

摘 要

本研究设计并实现了一种蜘蛛仿生六足机器人系统。硬件架构采用ESP32 S3 N16R8微控制器作为主控单元, Maixcam作为摄像头模块, 通过启灵者ZL-IS2 24路舵机控制器控制mg90s舵机, 舵机控制板驱动18个关节舵机, 结合WiFi无线通信模块与高效电源管理电路, 构建完整的运动控制系统。通过API接口与大语言模型集成, 结合Maixcam视觉模块的实时环境感知数据, 实现了智能问答功能与动态自主决策能力。

关键词

六足机器人, Maixcam视觉模块, 大语言模型, 人脸识别, 自主决策

Design of Hexapod Robot Based on Maixcam Visual Module and Large Language Model Linkage

Tingchen Deng, Qiwen Wang, Lulu Chen, Yikang Zhao

School of Electrical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

Received: Apr. 25th, 2025; accepted: Jun. 27th, 2025; published: Jul. 8th, 2025

Abstract

This study designed and implemented a spider-inspired hexapod robot system. The hardware architecture employs an ESP32 S3 N16R8 microcontroller as the main control unit, a Maixcam as the camera module, and utilizes the Qiling ZL-IS2 24-channel servo controller to manage mg90s servos. The servo control board drives 18 joint servos, integrating a WiFi wireless communication module

文章引用: 邓庭辰, 王启文, 陈露露, 赵益康. 基于 Maixcam 视觉模块与大语言模型联动的六足机器人设计[J]. 人工智能与机器人研究, 2025, 14(4): 811-819. DOI: 10.12677/airr.2025.144077

and an efficient power management circuit to construct a complete motion control system. The robot is capable of visual recognition and achieves intelligent interaction and autonomous decision-making by connecting to a large language model via API calls.

Keywords

Hexapod Robot, Maixcam Vision Module, Large Language Model, Face Recognition, Autonomous Decision-Making

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在当今科技飞速发展的时代，机器人技术正朝着智能化、多功能化方向不断迈进。六足机器人因其独特的多足结构，相较于传统的轮式或履带式机器人，在复杂地形适应性和稳定性方面展现出显著优势，广泛应用于灾害救援、探险勘探、工业巡检等诸多领域[1][2]。与此同时，视觉识别技术与大语言模型的蓬勃发展为机器人赋予了更敏锐的感知能力和更高级的智能决策能力。基于此，本研究致力于设计一款集先进视觉功能与智能语言交互决策于一体的六足机器人，旨在拓展机器人在安防、导览、智能家居等多样化场景下的应用范围，提升其服务人类生活与生产的能力。

2. 系统总体架构

2.1. 系统总统框图

系统框图如图1所示，选用18650锂电池7.4 V供电，在经过降压模块的降压将7.4 V降至5 V，给Maixcam摄像头，ESP32控制板以及ZL-IS2 24路舵机控制器供电，同时小智语音助手连接ESP32控制板来供电。摄像头模块主要用来获取图像信息，搭载模型可以具有一定的处理信息的能力，将图片信息传递给ESP32控制板，同时接入小智语音助手搭载大语言模型，使其具有对话能力，同时将目标指令传递给ESP32控制板，控制板控制舵机控制器来完成目标指令。

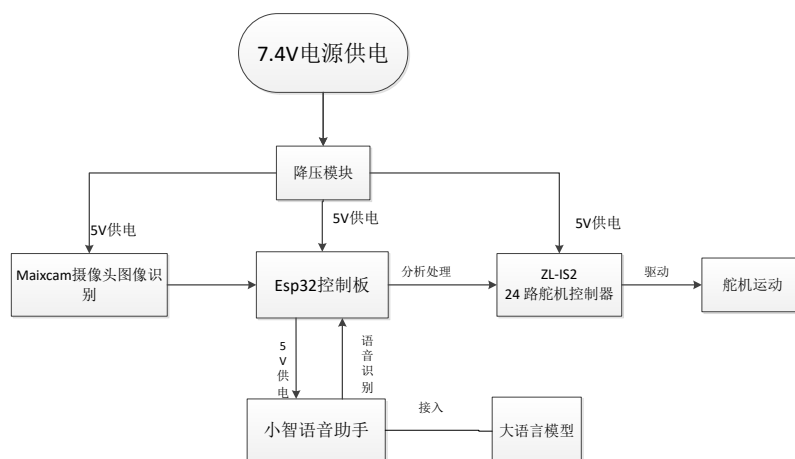


Figure 1. Overall system block diagram

图1. 系统总体框图

2.2. 机械结构设计

在进行机械设计之初，首先明确了项目的总体需求——结构需具备良好的扩展性、可维护性以及一定的模块化水平，以便适应后续功能的迭代与优化。为此，我们参考了现有机器人平台的模块化设计理念，结合 3D 打印制造的优势，决定采用灵活性强、可快速迭代的设计策略。特别是在结构布局上，优先考虑各模块间的协调性与兼容性，使其在保证整体稳定性的基础上，具备良好的拓展能力与可调性。机身框架选用 3D 打印，具体图纸如图 2 所示，具备良好的灵活性，能够快速根据需要修改。框架四周设计有多个安装接口，方便后续腿部机构、传感器以及其他附属部件的固定安装。

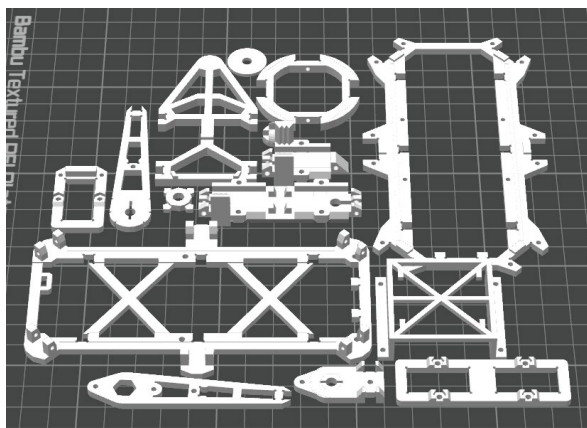


Figure 2. 3D printing drawing of robot mechanical structure
图 2. 机器人机械结构 3D 打印图纸

腿部机构：每条腿部由多个关节和连杆组成，模拟昆虫腿部的仿生结构，设有三个自由度，分别由三个高性能舵机驱动，实现关节的转动，完成腿部的前进、摆动和抬腿等动作。关节处采用高韧度材料连接，有效降低摩擦损耗，提升运动的灵活性和准确性。腿部的长度和粗细经过优化设计，确保机器人具备合适的步幅和支撑能力，适应不同地形的行走需求。设腿部长度为 L ，水平投影长度为 L_x ，垂直投影长度为 L_y ，则步幅 S 与腿部结构参数的关系可近似表示为 $S = 2L \times \cos 30^\circ$ 。

连接关节：腿部与机身框架通过特制连接关节相连，具备旋转和摆动自由度，能够在行走过程中自适应调整腿部角度，增强机器人在复杂地形上的稳定性。

3. 硬件系统设计

3.1. 控制核心

本设计采用高性能的 ESP32 作为核心控制芯片，具备丰富的外设接口和强大的数据处理能力，能够高效执行复杂的控制任务。所选用的 ESP32 3S N16R8 芯片支持 WiFi/蓝牙双模通信，集成 16MB Flash 存储与 8MB PSRAM 大容量内存，特别适用于本地语音处理、AI 边缘计算等对性能要求较高的物联网应用场景。在保持低功耗特性的同时，该模组还提供了优异的系统扩展能力[3]。

借助 ESP32 强大的无线通信功能，系统可与智能手机、计算机等终端设备实现稳定连接，为与大语言模型的数据交互及远程指令接收提供了坚实可靠的通信保障。

3.2. 驱动模块

采用启灵者 ZL-IS2 24 路舵机控制板是一款图形化编程，用来控制动作组，和搭配其他单片机控制的

控制板，见图 3，IS2 解决了多舵机控制和图形化编辑动作组的相关问题，能够通过 ZIde 简单的对机器人动作组进行编程，可视化的动作组编程[4]。为腿部微型舵机提供精准控制信号，满足多关节控制需求。其输出电流足以驱动舵机在不同负载下平稳运行，确保腿部动作流畅，见图 4。

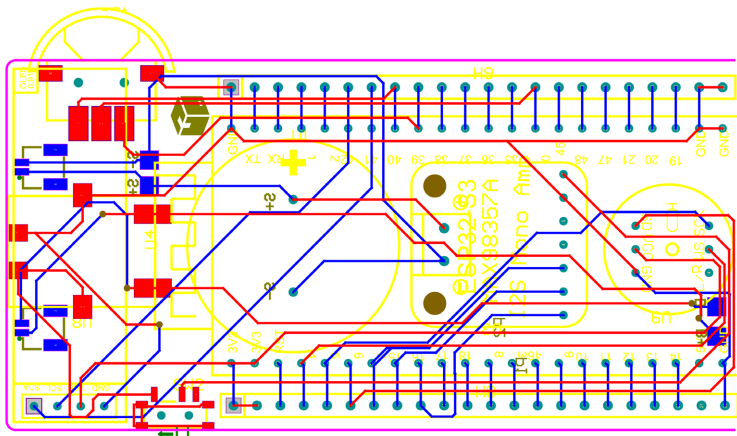


Figure 3. Control circuit PCB diagram
图 3. 控制电路 PCB 图

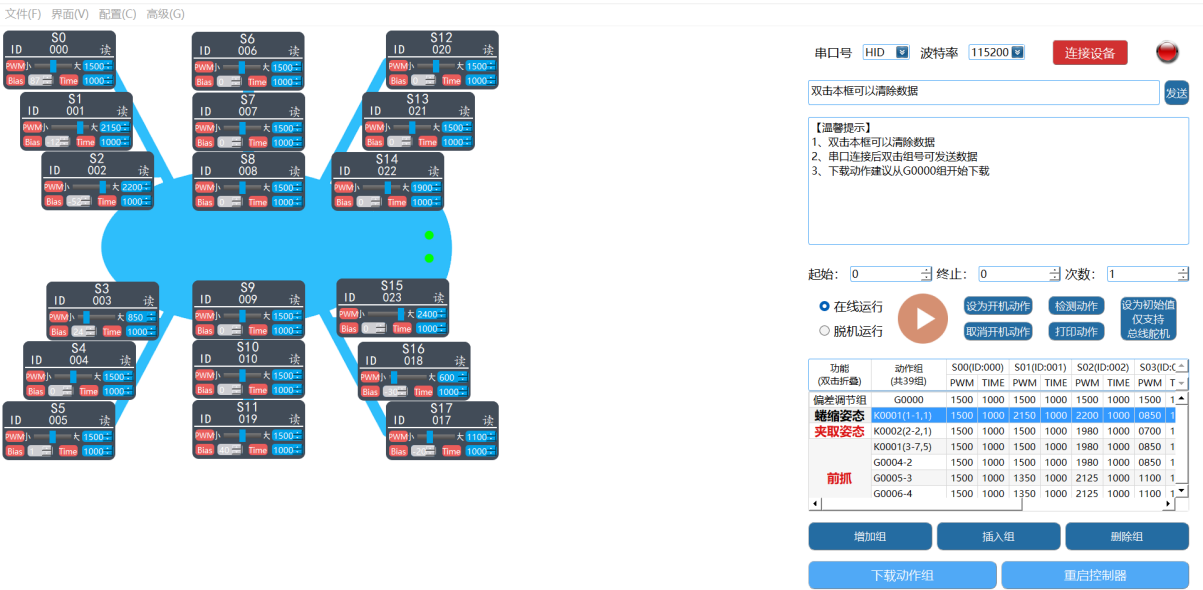


Figure 4. Upper computer diagram of servo drive board
图 4. 舵机驱动板上位机图

3.3. 电源模块

机器人采用航模专用锂电池作为电源，在保证高容量的同时实现了轻量化设计。电池输出电压为 7.4 V，经由高效降压模块稳压至 5 V，为控制板及其他电子元件提供稳定可靠的电力支持[5]。

3.4. 摄像头模块

本系统选用 Maixcam 摄像头模块用于图像采集，并具备基础的图像分析与处理能力。该模块不仅能

够实现水杯、鼠标等多类常见物体的识别，还支持人脸识别、表情识别及人体姿态识别等功能，为机器人的环境感知与智能交互提供有力支持。

将 Maixcam 做为串口模块使用，通过 Type-C 一转二小板，将设备通过串口连接到主控上了。其通讯发送的串口波特率是 115200，数据格式是 8N1，协议遵循 Maix 串口通信协议标准。

简单的信息分析和处理主要依靠 Maixhub 平台的模型实现。Maixpy 可以非常简单地 AI 视觉模型，可以直接调用底层封装好的 API，或者如果需要训练自己的模型，可以通过 MaixHub 在线训练。

3.5. 语音对话模块

小智语音助手以 ESP32 为核心控制板的嵌入式语音交互系统，通过集成轻量化人工智能模型，实现了端侧智能语音交互功能。系统以 ESP32 作为核心处理单元，搭配屏幕，微控开关，麦克风，扬声器模块等构成硬件基础，采用 WebSocket 以及 3D Speaker 等量化模型实现本地语音识别，结合 DeepSeek 模型进行语义理解，支持“你好小智”唤醒词检测和基础对话交互。系统创新性地采用分层处理架构：本地端处理唤醒词识别和简单指令执行(如设备控制)，复杂请求则通过 WiFi 连接云端服务扩展功能，见图 5。通过 FastSpeech2 轻量级语音合成引擎实现自然语音反馈，并利用 ESP32 的低功耗特性，配合语音活动检测(VAD)技术，同时小智的反馈也可以通过 OLED 屏幕进行反馈[6]。

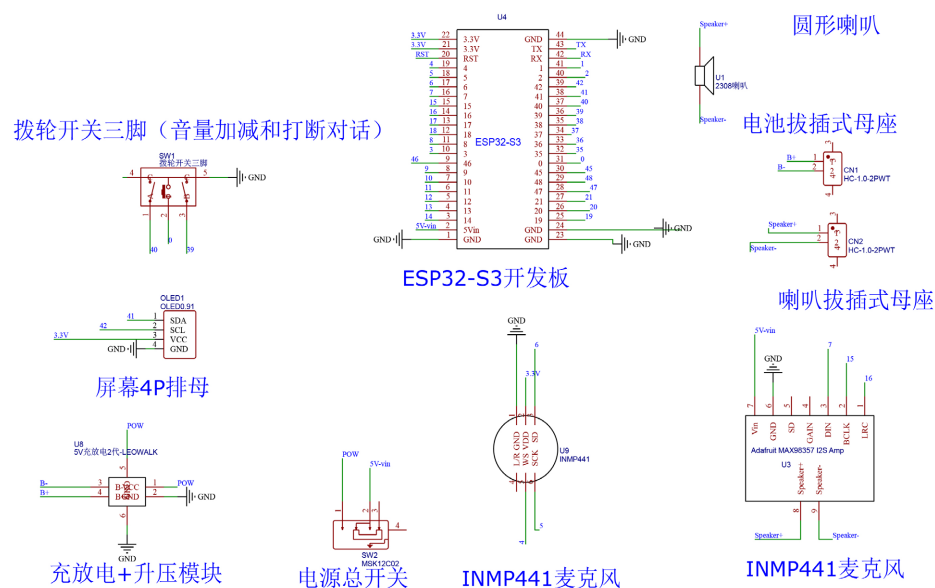


Figure 5. Schematic diagram of Xiaozhi voice assistant circuit
图 5. 小智语音助手电路原理图

4. 软件系统设计

4.1. 摄像头模块程序设计

摄像头模块程序设计流程如图 6 所示，摄像头模块首先通过 Maixcam 模块完成图像采集，获取当前环境中的图像信息。接着，系统利用人脸识别算法检测图像中的人脸区域，并通过仿射变换对图像中的人脸进行几何对齐，以标准化输入，增强识别精度。随后，提取对齐后人脸的特征向量，并将其与人脸特征库中的向量进行余弦相似度对比。最后，根据相似度判断识别结果，并输出对应的身份标签，实现对人脸的有效识别与分类[7]。



Figure 6. Camera module program design process
图 6. 摄像头模块程序设计流程

4.2. 人脸识别

在六足机器人中，主要用到简单的人脸识别使用 AI 模型检测人脸，获得坐标和五官的坐标。利用五官的坐标仿射变换将图中的脸拉正对其到标准脸的样子，方便模型提取脸的特征。使用特征提取模型提取脸的特征值。与库中记录的人脸特征值进行对比(计算保存的和当前画面中的脸的特征值的余弦距离，得出最小的距离的库中的人脸，小于设定的阈值就认为当前画面中就是这个库中的人，见图 7)。

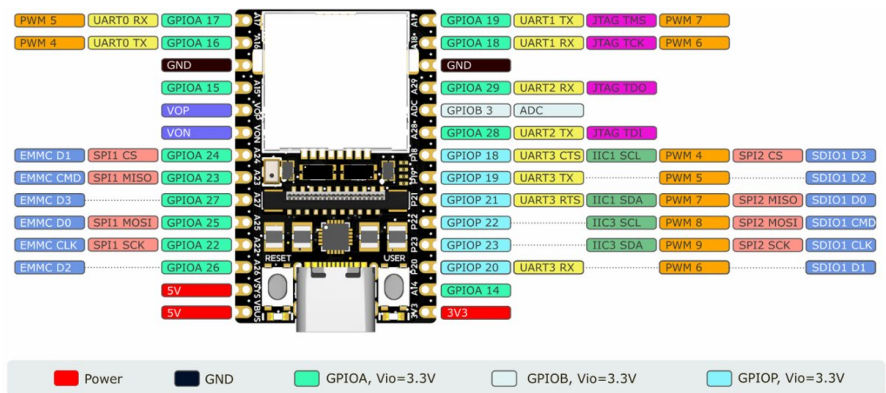


Figure 7. Functional pin diagram of camera module
图 7. 摄像头模块功能引脚图

4.3. 物品识别

系统采用 YOLOv5 等目标检测模型，在 MaixHub 平台上进行数据集标注、模型训练与部署[8] [9]。训练过程中，输入图像中的目标物体(如水杯、鼠标、书本、手机等)通过标注框进行识别标记，模型不断迭代学习其特征信息。训练完成后，模型被部署至 Maixcam 模块，实现对实际环境中物体的实时检测与识别。通过这一流程，系统具备了高效的物品识别能力，为后续任务执行与环境交互奠定基础。

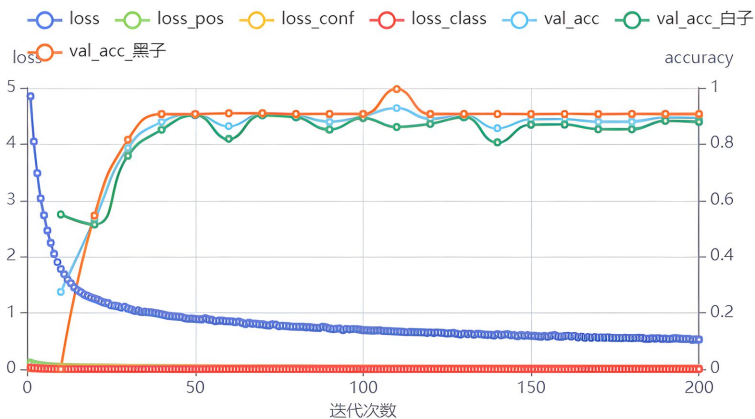


Figure 8. Schematic diagram of item recognition training
图 8. 物品识别训练示意图

图 8 是物品识别训练示意图, 图中展示了物品识别模型在训练过程中的各项损失值(loss)与验证集准确率(val_acc)的变化趋势[10] [11]。其中:

(1) loss (蓝线)表示总损失值, 在训练初期较高, 随着训练轮数的增加迅速下降, 说明模型在不断优化参数、提高对训练数据的拟合能力;

(2) loss_pos、loss_conf、loss_class 分别代表不同子损失项(正样本定位损失、置信度损失、类别损失), 它们在训练初期波动较大, 逐渐趋于稳定, 表明模型各部分正在逐步收敛。

(3) val_acc_黑子与 val_acc_白子分别为验证集中识别“黑子”和“白子”的准确率(橙色与绿色曲线), 准确率从最初较低逐渐上升并趋于平稳, 接近 0.9, 说明模型已经具备良好的识别能力。

综上可见, 模型训练过程稳定, 损失有效下降, 验证准确率较高, 说明该模型已经具备了较好的泛化能力, 可部署于实际物品检测任务中。

5. 系统调试

在完成整机结构搭建与硬件连接后, 系统进入调试阶段。通过多次试运行与模块级联调, 机器人各模块协同工作情况良好, 核心控制单元 ESP32 与 Maixcam 摄像头之间的通信稳定, 舵机控制响应及时, 语音识别与人机交互逻辑顺畅。在系统调试阶段, 通过多轮闭环校准实验, 精确标定舵机角度参数与脉冲宽度, 确保各关节运动轨迹一致, 并基于逆运动学算法优化步态时序, 最终实现六足机器人的稳定协调运动。机器人实物图如图 9 所示。

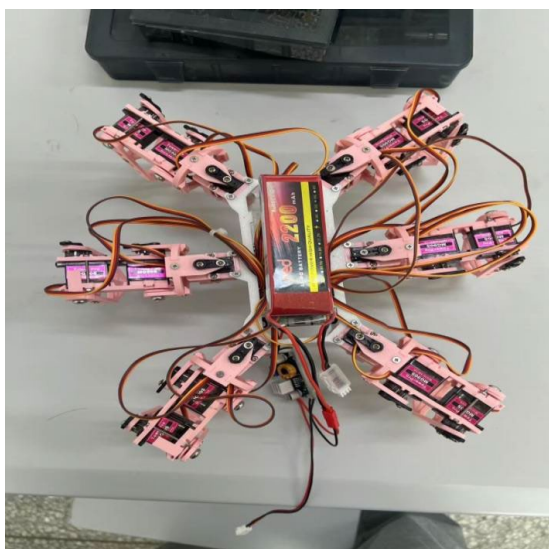


Figure 9. Physical picture of the robot
图 9. 机器人实物图

机器人可以实现如下功能:

- (1) 基础动作控制: 支持机器人前进、后退、左右平移与原地转向;
- (2) 复杂姿态动作: 实现了挥手、快走、扭腰等动作, 展现了良好的动态平衡与协调性;
- (3) 图像识别与反馈: Maixcam 模块通过 YOLOv5 模型实现了物品识别(如鼠标、水杯、棋子等), 并基于人脸识别算法完成了人脸检测、特征提取与身份匹配;
- (4) 人机交互功能: 语音模块支持自定义唤醒词检测和基础对话交互, 结合大语言模型实现了智能问答与指令响应, 验证了分层处理架构(本地 + 云端)的可行性, 见图 10;

(5) 无线通信控制：系统支持 Wi-Fi 无线通信，可远程发送控制指令，实现远程调试与控制，如图 11。

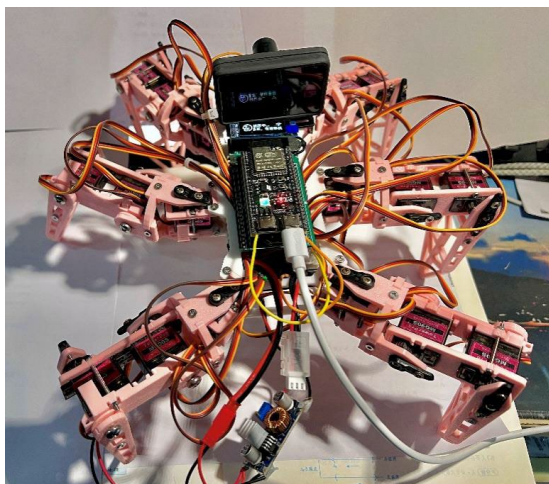


Figure 10. Function display of the voice module
图 10. 语音模块功能展示图



Figure 11. The screen of the control terminal of the mobile phone
图 11. 手机控制端画面

6. 结论

本系统采用 ESP32 控制器作为主控单元，Maixcam 作为摄像头模块，通过启灵者 24 路舵机控制器控制 mg90s 舵机，控制板驱动 18 个关节舵机，结合 WiFi 无线通信模块与高效电源管理电路，能够完成前进，后退，平移，转向，人脸识别，人机对话等功能，以及包括挥手，快走，扭腰等高灵活度动作，为未来六足机器人发展提高了良好的范例。

参考文献

- [1] <https://github.com/sipeed/MaixPy>
- [2] <https://github.com/78/xiaozhi-esp32>
- [3] 董正. 六足机器人自主指令生成与人机协同操控方法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2024.
- [4] 王鑫鹏. 基于深度强化学习的六足机器人运动规划研究[D]: [硕士学位论文]. 绵阳: 西南科技大学, 2023.
- [5] 冯鲁文, 董伟. 基于多传感器的 6 足救援机器人设计[J]. 技术与市场, 2020, 27(2): 37-38.
- [6] 张宏江, 鄂维南, 张亚勤. 大语言模型[J]. 中文信息学报, 2025, 39(1): 175.
- [7] 陈潇磊, 尤波, 李佳钰, 丁亮, 董正. 基于驾驶员模型的六足机器人自主/协同决策[J]. 仪器仪表学报, 2023, 44(4): 91-100.
- [8] 潘海鸿, 钱广坤, 陈希良, 等. 基于改进 YOLOv5 的柑橘采摘机器人识别定位方法[J]. 传感器与微系统, 2025, 44(5): 57-61.
- [9] 李新科, 毛含莹, 孙玉瑾, 等. 基于 DCNN 的水稻叶瘟病检测与取样装置设计实现[J]. 电子制作, 2025, 33(6): 36-41.
- [10] 钟坤华, 陈芋文, 秦小林, 等. 基于机器学习的脓毒症预测与干预决策研究综述[J]. 计算机应用, 2024, 44(S2): 357-363.
- [11] 王凯, 黄陈蓉, 顾杰, 等. 基于 MFBN-YOLOv5 输电线路绝缘子缺陷检测研究[J]. 计算技术与自动化, 2025, 44(1): 80-87.