项目计划书

身动翼随——基于人体动作语言识别的无人机控制

项目负责人：李俊逸

所在赛道：本科生创意组

联系电话：17719796315

PROJECT PLANNING PAPER

目录

第一节 基本概要

1.1 项目背景

自 2022 年以来，我国无人机行业发展迅速，市场规模快速增长，预计至 2024 年市场规模达 2000 亿元以上，至 2025 年民用无人机产值达 1800 亿元。

图 1 无人机示意图

无人机技术趋向创新，正逐步向小型化、微型化、长续航、网络集群化方向发展，应用场景不断拓展，在应急抢险、医疗救援、城市管理、物流配送等新兴领域展现出巨大潜力。

我国无人机相关产业链不断完善，从机体、飞控到通信系统，国内都已形成了一定的产业规模，为无人机行业持续发展提供了有力支持。此外，国民对于无人机这一新鲜事物的使用意愿也显著提高，无人机使用正趋向于日常化。

针对以上特点，结合已有知识与创新思路，项目组就无人机在无终端连接情况下的使用做出创新研究。

1.2 项目定位

针对无人机应用过程中，面临的无终端情况下的使用需求，即在如遥控、网络信号传输等远程操控手段不可行的情况下，使用者对于无人机自识别控制的需要。项目组作出研究，认为可以采用人体肢体动作语言来应对这一需求，即通

过无人机对于人体动作的自识别与应答，从而执行动作语言的附加指令，实现在无终端连接情况下的无人机控制。

该技术手段主要应用于远程操控手段缺乏或不可用时，应对紧急情况的补充控制方法，主要应用场景包括军事作战、物流运输、紧急搜救等。同时，该技术手段具有低门槛的特点，使用者只需做出特定动作，由无人机识别后作出自应答，适用人群更广泛。

1.3 创新要点

当前，国内市场上各种无人机控制技术丰富且成熟，但大多为需要使用者通过控制终端下达命令的被动信号传输方式，项目组以无人机的主动识别和自动应答作为主要创新点，用于补充当前市场上无人机控制手段的多样性。其主要区别，便是把控制任务的实现重心从使用者转移到无人机上，从对使用者通过控制终端下达指令的被动接受转变为对控制者动作语言信号的主动检索，并随后做出分析与应答，实现对使用者操作的轻量化。

算法组在程序模块的实现上，做出了很多基于特定任务情景的模型创新与程序设计：

- **数据预处理：**利用 **YOLOv5** 模型进行实时目标检测，能够快速准确地识别视频帧中的人物目标区域，剪裁背景噪音，显著提高模型性能。同时实现了裁剪、尺寸调整和归一化操作，提高了动作分类模型的训练效率和模型的泛化能力；
- **动作检测：**利用 **MediaPipe** 框架，识别和跟踪人体的关键骨骼点。并设计了一套基于大量数据归纳而来的**模态逻辑规则**，能够在绝大多数情况下确保动作检测的准确性，同时引入 **MLP** 模型作为补充，能够处理更复杂的模式识别任务，提供了正确性的额外保证，提高了系统的适应性和泛用性。整个检测流程大幅度降低了动作检测所需的运算量；
- **智能选帧：**利用 **SMART** 算法实现选帧，将采用模型蒸馏技术的**单帧选择器**，与结合关系模型、**LSTM** 网络的**全局选择器**协同使用，减少了计算负担，提高了动作分类的准确性，有效降低后续动作分类的计算成本；

- **动作分类：**采用了新颖的双流网络架构，由基于 **I3D** 模型的 RGB 流和基于 **S2GCN** 的骨架流组成，实现了对视频帧的外观特征和时序动态特征信息的综合利用，提高了动作识别的准确性和稳定性。创新设计 S2GCN 网络来处理人体骨架数据，通过在空间和时间维度上分离图卷积操作，有效地减少了计算复杂性，同时保留了关键的动作表示。

ROS 组在系统模块的实现上，做出了很多基于实际应用需求的工程创新：

- **设备优化创新：**创新采用**树莓派 Raspberry5** 作为无人机搭载的微型计算机，具有低功耗、轻量化、高灵活性的特点，可拓展型强。
- **数据采集处理：**将数据采集，数据处理一体化，集成于记载电脑。利用树莓派 **Picamera2** 框架采集数据，并创新性的在机载电脑进行数据处理，减少了原先在“地-空”数据传输过程中的较大的延迟损耗与可能的数据误差。
- **系统控制：**采用 **ROS2** 框架进行系统整体调控。通过发布，订阅节点，实现数据的采集、处理，飞控。在实现记载电脑控制的同时，利用地面站订阅节点，实时了解无人机的实时状态，确保了安全性与可控性。
- **飞控单元：**利用 **MAVROS** 框架，设计飞控系统体系，与 PIXHAWK 飞控核心进行实时交互。利用机载电脑延迟低的特点，采用较高的控制刷新频率，有效提高了飞控过程中的反应速度，提高无人机的安全性与可靠性。

1.4 团队简介

项目组为本科生队伍，成员均为西安交通大学在读本科生，包括人工智能、自动化、计算机科学与技术、电气等专业学生，团队组成跨学科、跨专业，具有专业交流和学科交叉的团队特色。根据项目需要，主要分为算法、系统、硬件三个模块，团队分工及专业适配如下：

项目负责人	李俊逸	自动化
程序模块	张圣涛	人工智能
	张皓凯	人工智能
	王子诚	人工智能
	赵恒	计算机科学与技术

系统模块	王崇杰	自动化
	钟艺萌	自动化
	蒋梓轩	人工智能
硬件模块	同芃柏	电气
	张淏钦	电气

项目组指导老师共两位：

学院	指导老师	指导方向
人工智能学院	刘龙军	技术指导与项目定位
信息与通信工程学院	毕海霞	项目定位

第二节 项目介绍

2.1 项目总述

基于人体动作语言的无人机控制，便是无人机主动识别目标对象做出的特定动作后，经搭载计算机分析后做出自动应答，实现无人机移动操作的方法。

具体来说，无人机需要搭载视觉识别模块和微型运算模块，通过部署其上的操作系统实现算法的运行和数据的传递，从而实现设计思路。

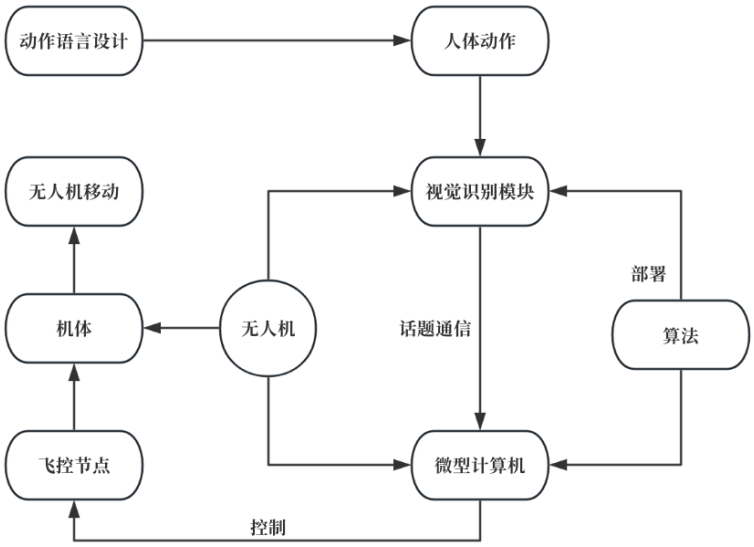

图 2 项目基本思路

项目组结合交警手势与机务手势，设计了一套能够满足三维空间基本移动要求的动作语言。

在无人机上搭载了摄像头提供视觉识别功能，并搭载树莓派这一微型计算机来实现模块控制与信息传输。

利用 ROS 实现算法程序的系统搭载。基于动作语言设计，项目组自行组建了小规模的数据集，用于模型训练，成效显著。

视频数据经多个模型依次处理后，转为为无人机移动指令，经飞控节点实现对无人机的移动操作。

2.2 算法部分

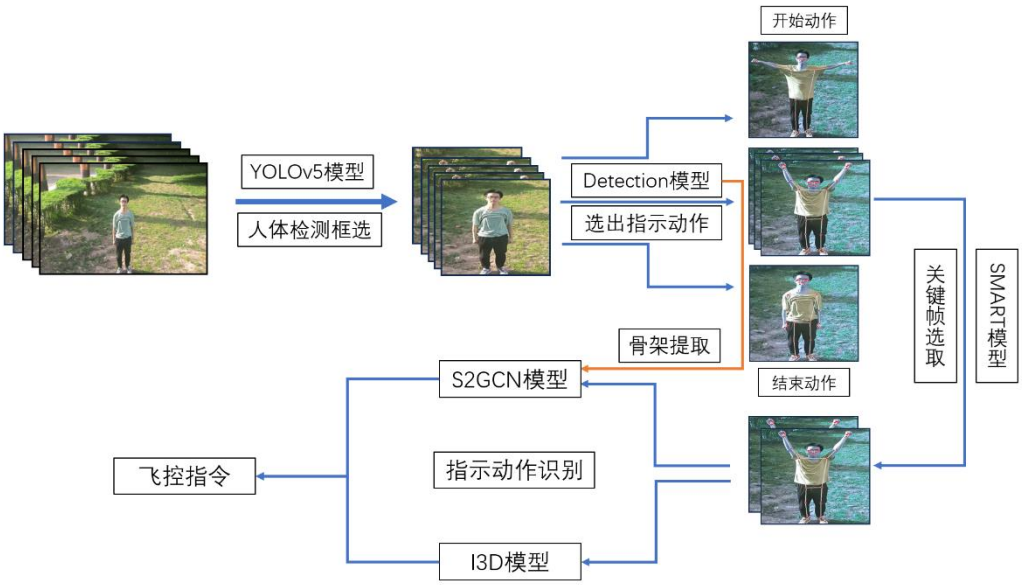

图 3 算法基本思路与流程

通过相机模块获取到视频数据后，利用 **YOLOv5** 模型逐帧进行人体目标检测，根据检测结果选取合适规模的边框，对图像进行裁剪，减小操作者环境背景噪声的影响。

对经处理的视频数据，使用 **Mediapipe** 框架进行起始动作与终止动作的动作检测，利用 **SMART（Sampling through Multi-frame Attention and Relations in Time）**模型处理指示动作，选取在空间上和时间上包含信息量较大的视频帧，采用双流网络架构分两路进行动作分类：

其一，利用基于 RGB 的 **I3D（Inflated 3D ConvNet）**模型识别指示动作；

其二，利用基于骨架的 **S2GCN（Spatio-temporal Seperable Graph Convolutional Network）**网络进行分类。将双重分类结果加权计算后得到综合分类结果，通过 ROS 系统向飞控节点传递动作信号，实现无人机的控制。

2.2.1 人体姿态动作检测与切分

YOLOv5 模型具有优越的检测速度和准确性，能够以较低延迟实现实时检测。利用 YOLOv5 模型对视频帧进行目标检测，框选出识别目标所在区域。在检测过程中，取视频数据中所有帧的最大目标范围确定框选尺寸，对检测到的目标对象区域进行裁剪和归一化处理，以确保输入数据的一致性和标准化。

MediaPipe 框架是一种高度优化的深度学习架构，适合于快速处理图像识别任务。

利用 MediaPipe 框架识别起始和终止动作，逐帧提取出视觉识别的 25 个关键点的坐标信息，包括了头部、肩部、肘部、手腕、髋部、膝盖和脚踝等人体的主要关节和部位，构成 50 维度的特征向量，以反映人体的动作姿态，后续步骤中指示动作的识别会利用获取到的骨架信息。

骨架识别如下图示例：

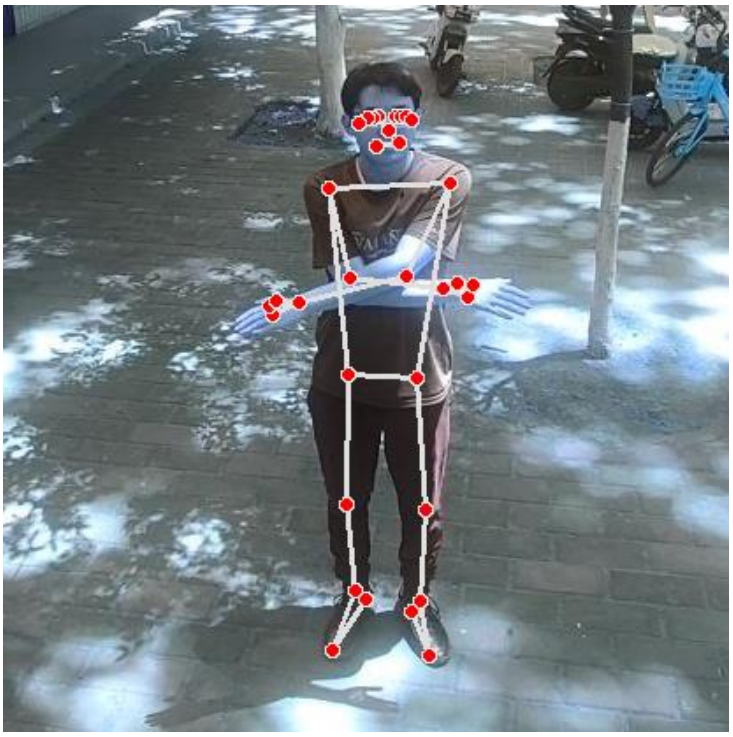

图 4 人体骨架识别示意图

随后，特征向量将由双重路径进行分类处理：

路径一，特征向量被输入**多层感知器（MLP）**神经网络中，对输入的动作特征进行分类。

多层感知器（MLP）神经网络是由多个全连接层组成的深度学习模型，能够学习复杂的非线性关系。

路径二，基于关键点坐标信息归纳出一系列动作特征，编写为模态逻辑规则。

模态逻辑是一种形式化的模糊逻辑系统，能够处理不确定性，并提供一种基于规则的分类方法。

考虑到对 MLP 神经网络和模态逻辑规则的优势的综合利用，项目组采用堆叠泛化技术。将双重分类器的预测结果通过逻辑回归模型融合，产生更可靠的分类结果。

并将逻辑回归模型用于学习基础分类器输出的结合，提高分类结果的准确率。

在训练阶段，对于输入的数据集中样本视频，检测模型能够准确地识别出起始和终止动作，自动截取起始动作至终止动作之间的视频片段，即指示动作，传递给分析节点，由识别模型进行分析。

实际运用中，模型识别到起始动作后开始记录，检测到结束动作后停止，获取完整的指示动作片段，交给识别模型进行分类。这一衔接紧密的数据处理流程确保了训练阶段和实际运用过程的相似性，为系统性能提供保障。

2.2.2 人体指示动作的分析与智能选帧

考虑到视频数据具有高度的时间冗余、较大的噪声以及短暂的不相关帧，算法组利用 SMART 策略实现数据过滤，选择信息量较大的视频帧。该策略具有低计算成本的特点，可以增强识别模型对无人机端侧此类运行资源受限的环境的适应。

SMART 策略的核心是利用两个结构和训练方式不同的选择器对视频帧进行评分，取两个选择器的评分乘积作为综合标准，选取评分前 k 帧以输入识别模型进行动作分类。其中，k 是一个可调节的超参数，用于平衡计算成本和分类准确性。 其中，**单帧选择器（single selector）**负责评估单个视频帧对动作分类的贡献，**全局选择器（global selector）**则专注于捕捉跨帧的时间关系并评分。

单帧选择器的实现基于模型蒸馏技术，能够使计算成本降低的同时减少性能牺牲。训练过程中，采用轻量级的多层感知器（MLP）作为**学生模型**，来模仿高性能但计算密集的**教师模型（EfficientNetV2）**的行为，确保在资源受限的环境下，能够高效地处理单帧动作信息。

全局选择器则采用一种基于关系模型和**长短期记忆（LSTM）**网络的架构，能够从视频序列中提取丰富的时空特征，并通过 LSTM 网络来学习不同帧间的时间依赖关系。 其中，LSTM 网络特别擅长处理序列数据中的长期依赖问题，这对于理解和评分视频中的动作序列至关重要。

选取两个选择器的评分积作为综合标准，有效地结合单帧动作信息和跨帧时间关系的评估，不仅提高了动作分类的准确性，而且通过选帧显著减少了计算负担，使得算法在实时应用中更加可行，能够在有限的计算资源下，实时准确地识别和响应用户的手势命令。

2.2.3 人体指示动作的识别与控制指令的输出

选帧处理完成后，使用双流网络架构对预处理后的视频帧序列进行分类，能够有效地融合来自视频数据的两种不同类型的信息，即外观信息和动态信息。

RGB 流主要关注视频帧的外观特征，捕捉静态图像中的关键信息，而骨架流则专注于时序动态特征，通过分析人体骨架的运动来理解动作的含义。

结合两种信息流，双流网络能够在复杂的背景和恶劣的环境条件下，更准确地识别和分类动作，确保动作指令控制无人机的精确性。

RGB 流 使用 I3D 模型。I3D 模型通过 3D 卷积操作捕捉视频帧之间的时序信息，能够有效识别复杂的动作模式。I3D 模型是将 2D 卷积网络扩展到 3D 卷积网络的一种方法，通过在时间维度上扩展卷积核来处理视频输入。

项目组选择了基于 ResNet50 的 I3D 模型，同时兼顾高性能与轻量化。ResNet50 是一种深层残差网络，通过引入跳跃连接解决了深层网络中的梯度消失问题，具有很强的特征提取能力。I3D 模型通过将在图像上预训练好的 2D 卷积核扩充为 3D 卷积核，使得模型能够同时捕捉空间和时间上的特征，从而更好地理解视频帧之间的时序关系。

骨架流 首先利用动作检测得到的骨架进行动态图的构建，随后利用项目组研发的 S2GCN 网络进行视频的特征提取，最后利用多层感知机作为分类器，得到骨架流的分类结果。

具体介绍如下:

动态图构建：在视频每一帧图像中，提取骨架中目标对象的 25 个关节，对于每一个节点，以在人体骨架中与该节点相连的节点作为邻居，初始输入特征向量（embedding）为前一步骨架检测模型输出的该节点相对人体中心的二维坐标与坐标置信度组成的 3 维向量。

分类过程：考虑到无人机搭载的微型计算机性能有限，分类模型满足轻量化的特点。受 RGB 路线中 R(2+1)D 模型的启发，项目组使用先空间-后时间的特征提取方式，大幅降低了计算量，实现了轻量化。

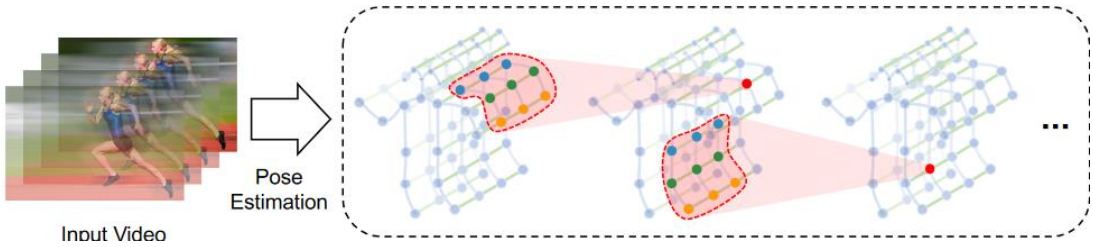

图 5 骨架动态图构建示意

具体来说，在动态图的每一个时刻，使用 **RGCN(Relational Graph Convolution Network)**得到每一个节点的空间表示，对每个节点固定一个时间窗口大小，在时间窗口内使用带有指数衰减的平均池化（距离中心节点越远的节点在平均池化中的权重越低），得到节点带有时间信息的节点表示，利用 Readout 函数得到图的全局表示，并输入 MLP 进行分类。

利用可学习的权重计算将 I3D 模型与 S2GCN 模型预测得到的结果进行加权平均之后得到综合分类结果，将模型输出的分类结果转换为如上升、下降、前进、后退、左转、右转等无人机的飞行动作命令，控制无人机执行相应的操作。

2.3 系统部分

采用 ROS（机器人操作系统）搭载算法、程序，实现跨平台复杂任务的简化。ROS 对于操作系统泛用性高，能够在树莓派平台上运行。树莓派作为微型计算机，具有低成、高性能的特点，与 ROS 的适配性高。

2.3.1 ROS

ROS 是一个为机器人软件开发者提供的灵活框架。它包含了大量库函数、工具、约定和消息传递接口，旨在简化跨机器人平台复杂任务的创建和管理。

ROS 核心特性：

分布式架构：ROS 采用了一种基于发布/订阅模型的分布式系统架构，允许各个模块（节点）之间独立运行并通过消息进行通信。

丰富的工具集：ROS 提供了大量用于开发、调试、测试和可视化的工具，如 rviz、rosbag 等。

广泛的软件生态：ROS 拥有庞大的开源社区和丰富的软件库，开发者可以方便地获取和复用已有的机器人软件模块。

跨平台兼容性：ROS 支持多种操作系统和硬件平台，包括树莓派，使得开发者能够灵活选择最适合其项目的硬件和软件环境。

2.3.2 树莓派

树莓派 5 作为项目的核心硬件平台，具备出色的性能和丰富的接口，搭载了四核 64 位 ARM Cortex-A72 处理器，主频高达 1.8GHz，为实时动作识别提供了强大的计算能力。

同时，树莓派 5 配备了最高 4GB LPDDR4 SDRAM，确保了数据处理的高效性和稳定性。在存储方面，树莓派 5 支持 microSD 卡插槽，可以轻松扩展存储容量，满足项目对存储空间的需求。

此外，树莓派 5 还提供了千兆以太网、Wi-Fi 5 (802.11ac) 无线网络、USB 3.0、USB 2.0、HDMI、GPIO 等多种接口，便于连接各种外设和传感器，实现项目的多样化需求。

树莓派 5 支持多种操作系统，包括 Raspberry Pi OS（基于 Debian 的 Linux 发行版）以及其他主流 Linux 发行版。这些操作系统提供了丰富的软件资源和开发工具，可以根据项目需求进行软件开发和调试。

在编程语言方面，树莓派 5 支持 Python、C/C++、Java 等多种编程语言，提供了灵活的选择。项目组可根据项目需求选择 python 编程语言进行开发，实现无人机动作的实时识别和处理。

2.3.3 相机模块

Camera Module 3 采用了索尼 IMX708 传感器，具有高达 1200 万像素的分辨率，支持拍摄 1080P、50 帧的视频，为动作识别提供了清晰、流畅的视频源。

在实际应用上，Camera Module 3 具有实时图像捕获、精准识别和优秀的扩展性的特点。

通过 Camera Module 3，无人机能够实时捕获高清晰度的图像和视频，为动作识别算法提供输入。

结合树莓派 5 的强大计算能力，Camera Module 3 捕获的图像和视频可以被用于精准的无人机动作识别，包括手势识别、姿态识别等。Camera Module 3 的丰富功能和灵活配置使得无人机动作识别系统具有更好的扩展性，可以适应更多复杂的应用场景。

2.3.4 工作流程

整体系统模块包括微型电脑主板和相机模块，具体运行方式如图。通过摄像头进行外部信息获取，识别人体动作，通过话题通信传递给分析节点，通过部署算法的实时分析，得出动作识别的分析结果，传递给飞控节点，通过发布飞控节点，订阅节点，控制无人机应答。

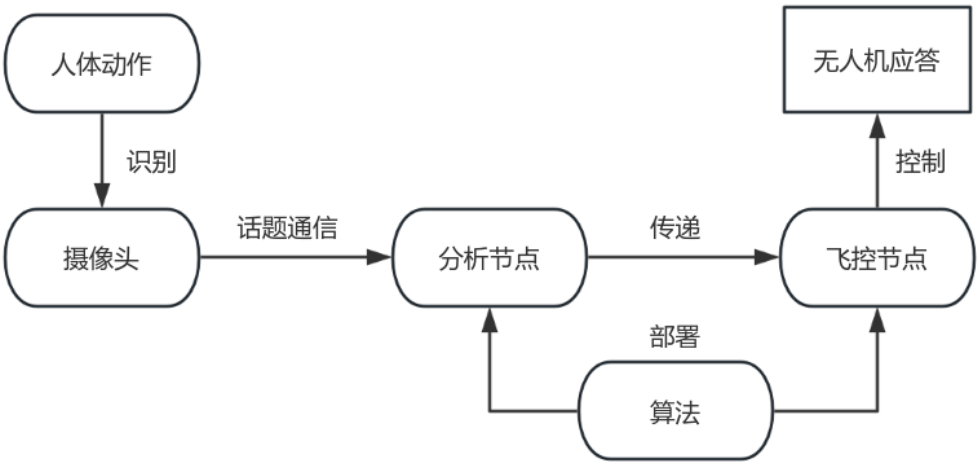

图 6 系统运行基本思路

系统模块的通信连接方式采用**话题通信**，话题通信是基于发布订阅模式的，即一个节点发布消息，另一个节点订阅该消息。大致示意见下图。话题更新模型主要由管理者、发布者、订阅者三部分组成。管理者负责保管发布者和订阅者注册的信息，并匹配话题相同的发布者与订阅者，帮助发布者与订阅者建立连接，连接建立后，发布者可以发布消息，且发布的消息会被订阅者订阅。

摄像头数据采集主要利用 **Picamera2** 框架。在检测到开始动作前，系统以固定频率采集帧数据，并传输给检测模型。在检测到开始动作后，采集视频数据，并处理成张量形式传入检测模型，直到检测到结束动作。在此过程中，系统以固定频率接受检测模型传出的结果，并发布飞控指令，控制飞行器运动。

飞控指令的发布主要采用 **MAVROS** 框架。**MAVLink** 是 PIX4 框架所使用的飞控协议。MAVROS 是一个开源的的 ROS 包，用于将 ROS 与 **MAVLink** 进行串联。飞控指令发布模型基于 MAVROS 系统进行开发，在接受到检测模型的结果后，以固定频率发布节点，控制无人机运动，并在同时发布数据类节点。地面站可通过订阅数据类节点实时获取无人机的相关参数，如高度，速度等。这一过程既可实现机载电脑对无人机的控制，又可实现本地地面站对无人机的操控，确保系统的安全性与可靠性。

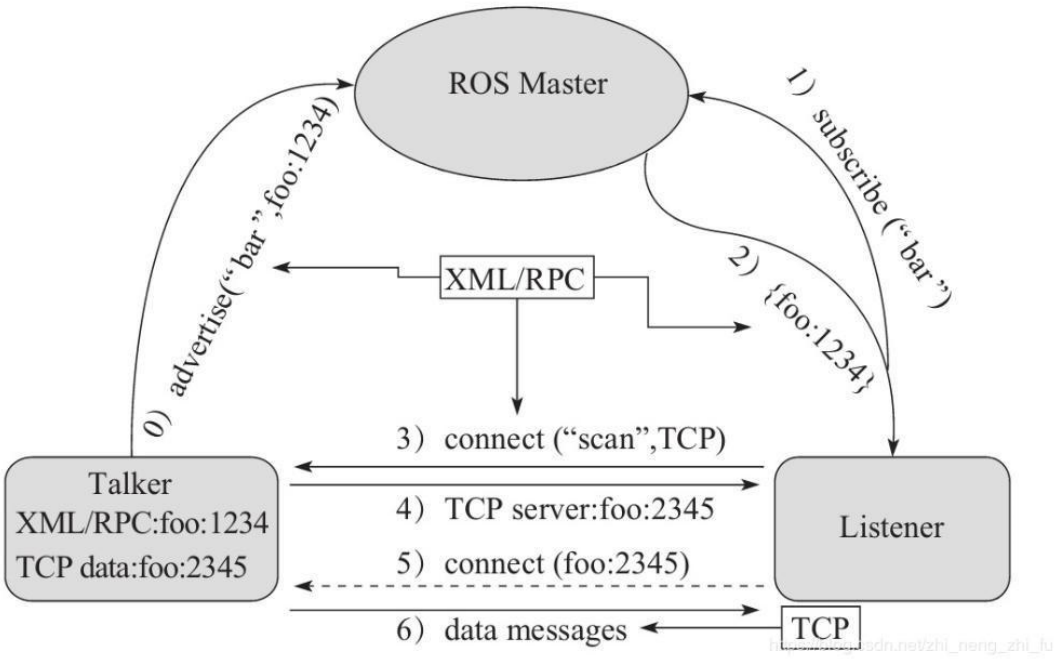

图 7 话题通信内容示意

2.4 硬件部分

在构建无人机飞行平台时，选择了图腾 Q250 四轴穿越机机架作为核心结构。这款机架将电池供电与电调集成在底板上，不仅简化了无人机的整体结构，还显著减小了体积，为后续的仪器搭载和飞行提供了更大的便利。

飞控系统方面，采用了 Pixhawk 2.4.8，它搭载了高性能的 32 位 ARM Cortex-M4 处理器，并运行 Nuttx RTOS 实时操作系统。这款飞控不仅功能强大，而且支持多种扩展模块，如气压计、电子罗盘和 GPS 等。这些模块为无人机提供了精确的定位、导航和避障能力，使得无人机能够执行更复杂的任务，如精确坐标定位、高度测量、自动导航和超声波探障等，极大地扩展了无人机的应用场景和实用性。

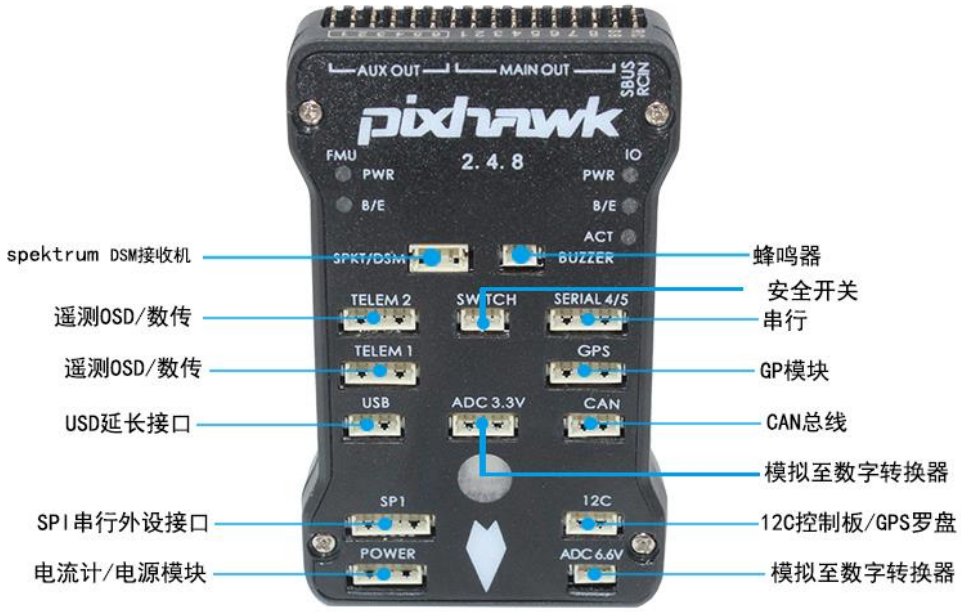

图 8 无人机端口示意

GPS 所使用的为飞控内置罗盘，虽然精度存在不足，但本项目核心不需要长距离的飞行，因此可以满足要求。

为了满足飞行需求，选择了 TMOTOR 的 V2306-2400Kv 型号电机，其提供的升力足以将未来可能搭载在无人机上的各种仪器顺利带入空中。

同时，选用了 14.8V 4S 2300MAH 45C XT60 型号的电池，该电池能够提供稳定的电压输出和巨大的电池容量，满足长时间飞行和测试的需求。电池大小与

机架设计相匹配，为整体设计提供了极大的便利。

设计图如下：

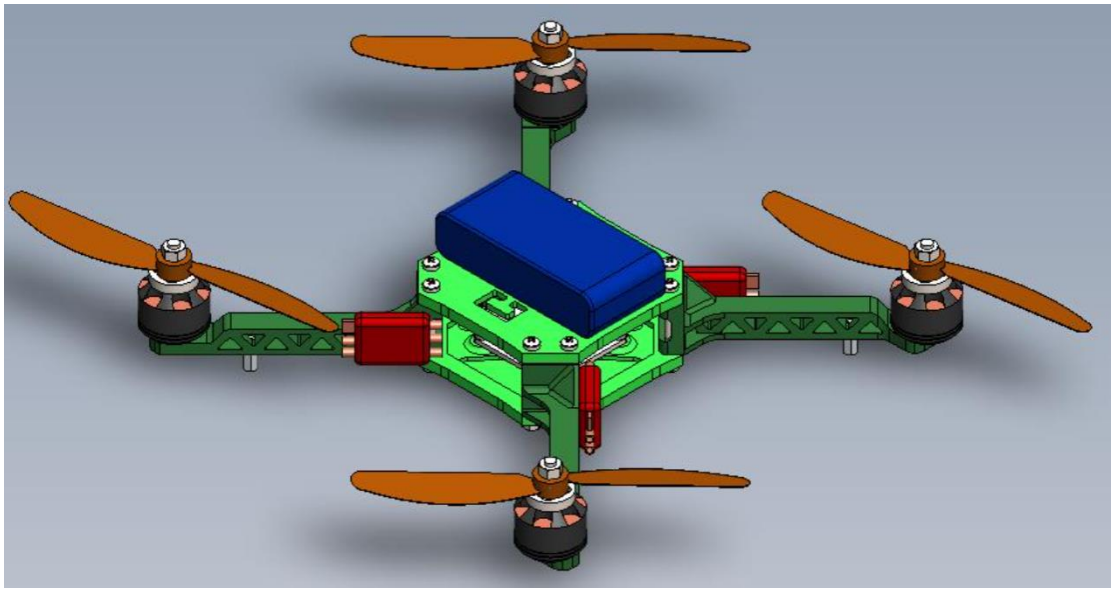

图 9 无人机设计图

2.5 数据集部分

为了实现对无人机的精确控制，项目组精心设计了一套动作语言，并为其创建了一个定制化的数据集。在设计这套动作语言时，遵循了简洁性和实用性的原则，确保每个动作都能直观地传达特定的控制指令。

数据集的制作过程中，致力于提高其多样性，通过选择不同性别、年龄和体型的表演者，在各种不同的室内外场景下，以及采用不同的拍摄角度和距离来捕捉动作。这些动作包括明确的起始动作、待识别动作和结束动作。

待识别动作是控制无人机行为的核心，包含了无人机的控制信息，而起始动作和结束动作则用于明确地标识动作的开始和结束，以便在实际应用中能够准确地截取有效信息。

针对无人机的六种基本控制动作，即向上、向下、向左、向右、向前和向后运动，设计了六种对应的手势动作。在实际的数据集制作中，邀请了 10 位表演者，他们在 10 个不同的场景中，以大约 45 度的俯视角拍摄了这六组动作，共计产生了 600 个视频样本。每个视频样本的长度约为 10 秒，其中前 3 秒为起始动作，中间 4 秒为待识别动作，后 3 秒为结束动作。

尽可能地缩短了动作之间的转换时间，以减少对数据集质量的潜在影响。最后，将视频的帧率设置为 30 帧/秒，分辨率设置为 1280*720，并将视频转换为一组图像帧，以便进行后续的处理和分析。

以下为项目组设计的基本动作语言：

图 10-1 起始动作

图 10-2 终止动作

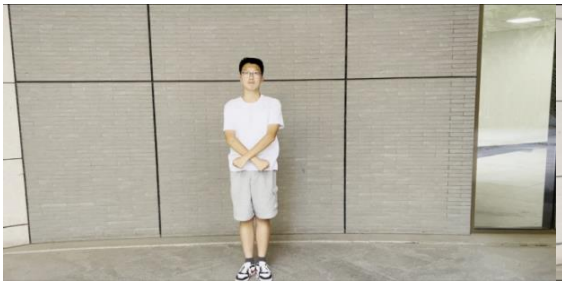

图 10-3 向前

图 10-4 向后

图 10-5 向上

图 10-6 向下

图 10-7 向左

图 10-8 向右

在数据集的设计与收集中，项目组工作满足以下要求：

动作设计的封闭性：动作设计采用了封闭性原则，确保每个动作都以起始动作开始，以结束动作结束。这一原则不仅有助于在实际应用中最大限度地保留有效信息，还能够在较低帧率的条件下，使制作的数据集与实际采集的数据保持较高的相似度，从而避免了因录入的起始和结尾噪声而影响动作识别的准确性。

多样性的保证：通过在多个角色和多个场景中录制数据，有效减少了背景对动作识别的干扰，提高了模型对不同环境和条件的适应性，从而在一定程度上保证了模型的鲁棒性。

真实情境的还原：通过引入多样化的背景和采用俯角拍摄的方式，的数据集尽可能地模拟了无人机在实际航拍时的真实情境。这种设计不仅增强了数据集的有效性和可靠性，还有助于训练出的模型更好地适应实际操作中的各种复杂环境。

光照条件的考虑：在不同的时间段和光照条件下进行了视频拍摄，包括早晨的柔和光线、中午的强烈日光和傍晚的昏暗光线。这样做的目的是为了使模型能够适应光照变化，提高其在不同光照条件下的识别能力。

质量控制的重视：在视频录制的过程中，严格控制了动作的执行质量，确保每个动作都被清晰、准确地执行，避免了模糊或不完整的动作对数据集的整体质量产生负面影响。同时，还对视频进行了后期处理，以消除可能的噪声和干扰，确保每个视频帧都能够清晰地反映出手势的细节。

综上所述，数据集设计和制作过程在多样性、真实性、质量控制和精细标注等方面都下了极大的功夫，这些因素共同为的手势控制无人机项目提供了坚实的基础，并为神经网络模型的训练和优化创造了有利条件。通过这样的数据集，能够训练出更加精确和可靠的动作识别模型，从而实现对无人机的有效控制。

第三节 团队组成

3.1 成员组成

项目组学生成员共 10 人，皆为西安交通大学在读本科生，团队组成跨学院、跨专业、跨学科，包括人工智能、自动化、计算机科学与技术、电气等专业。

具体介绍如下：

李俊逸，自动化专业，具有丰富的学生任职和社会任职经历，参与“星航计划”、“正心计划”等，曾获社会活动先进个人等荣誉，具有卓越的团队组织和项目运营能力，曾组织参加过腾飞杯等科研竞赛，大创项目负责人，参与“新蕾计划”，曾获校级奖学金、美赛 M 奖等奖项。

张皓凯，人工智能专业，有良好的 C++和 Python 编程基础，熟悉 Pytorch 框架的使用与机器学习、深度学习常见模型，现在在计算机学院罗敏楠老师本科生实验室（LUD）进行图神经网络（GNN）相关科研，有一篇剧透检测领域论文在投，曾获校级奖学金、优秀学生、美赛 H 奖等荣誉。

张圣涛，人工智能专业，认真学习并熟练掌握专业课程相关知识，熟悉 C++语言、Python 语言程序设计。已学习机器学习、深度学习，并熟悉 Pytorch 深度学习框架，有相关实践经验。目前在计算机学院罗敏楠老师本科生实验室（LUD）进行动态图神经网络（DGNN）相关领域的学习研究。曾获校一等奖学金、校优秀学生、美赛 H 奖等荣誉。

王子诚，就读于人工智能学院人工智能实验班，担任班级学习委员。曾获校级三等奖学金、数模美赛 H 奖。参与开发 Matterhorn——基于 PyTorch 的脉冲神经网络框架，被收录入 PyPI；参与大创项目“基于脉冲神经网络的动态吸引子时序特征提取研究”。

赵恒，计算机科学与技术专业，曾带队参加美国大学生数学建模竞赛并荣获 M 奖，获得校级二等奖学金，现致力于人工智能领域的学习和研究。

张溟钦，电气工程及其自动化专业，曾获电气王汝文奖学金一等奖，专业能力优秀，并对跨学科融合的本课题有浓厚兴趣，与同为电气专业的同芃柏同学负责无人机硬件设计与调控。同时也曾任团组织宣传部成员，多次参与美育志愿活动，德智体美劳全面发展。

同芃柏，电气工程及其自动化专业，曾获电气王汝文奖学金三等奖，进行过单片机以及嵌入式开发方面的学习与研究，能够胜任本项目硬件部分的实现，曾担任班级组织委员，多次参加美育志愿活动。

王崇杰，自动化专业，熟悉 Python、C 等代码语言，学习了 Pytorch、ROS 等框架，具有丰富的学生组织任职经历，曾获优秀学生、仲英榜样等荣誉，曾参加过腾飞杯等科研竞赛，参与过新蕾计划科研训练，获得校级二等奖学金等奖项。

钟艺萌，自动化专业，熟悉 Python，C 语言，ROS 框架，有丰富的学生工作和社会任职经验，积极参加腾飞杯，全国大学生数学建模竞赛，以及各类社会实践等等，荣获多项奖项。是“信息新蕾”ITP 科研训练计划成员之一。

蒋梓轩，人工智能专业。曾获校级三等奖学金、数模美赛 M 奖。有良好 C++和 python 语言基础。

根据项目需要，主要分为算法、系统、硬件三个模块：

团队分工及专业适配如下：

项目负责人	李俊逸	自动化
程序模块	张圣涛	人工智能
	张皓凯	人工智能
	王子诚	人工智能
	赵恒	计算机科学与技术
系统模块	王崇杰	自动化
	钟艺萌	自动化
	蒋梓轩	人工智能
硬件模块	同芃柏	电气
	张溟钦	电气

3.2 指导老师

指导老师共有两位，分别是人工智能学院的刘龙军老师和信息与通信工程学院的毕海霞老师。

刘龙军，副教授，博导。2015 年毕业于西安交通大学模式识别与智能系统专业，获工学博士学位，博士期间赴美国佛罗里达大学（University of Florida，UF）国家公派联合培养两年。在人工智能、集成电路设计、智能系统体系架构等领域期刊会议上如 TPDS，TCS-I，TCSVT，TMM，ISCA，CVPR，AAAI，ACM MM 等发表多篇论文。获得中国自动化学会“CAA 自然科学奖”一等奖（“从芯片到系统的高效智能计算架构关键技术研究”，第二完成人）；高等教育（研究生）国家级教学成果一等奖;获得 IEEE Computer Architecture Letter（IEEE 计算机体系结构快报，第一作者）期刊年度最佳论文奖“Best of CAL”；以第一作者

投稿的学术论文获得计算机体系结构领域国际顶级学术会议 IEEE/ACM International Symposium on Computer Architecture（ISCA，IEEE/ACM 国际计算机体系结构学术年会，第一作者）收录并在大会上做论文口述报告。获得 IEEE International Conference on ASIC（IEEE 国际专用集成电路设计学术会议）“优秀学生论文奖”（第一作者）。中国自动化学会科普奖“CAA 科普奖”（“AI 科普行动”）等等，承担国家自然科学基金、国家重点研发计划、国家科技重大专项“核高基”等多个项目。目前项目合作包括与航天九院共同开发无人机视觉检测算法研究及无人机导航规划等研究。

毕海霞，研究员，博士生导师，陕西省高层次人才青年项目、西安交通“青年拔尖人才计划”入选者。2003 年和 2006 年于中国海洋大学获得学士（计算机科学与技术）和硕士学位（地理信息系统）。2006 至 2013 年分别于华为通信技术有限公司和爱立信通信技术有限公司从事移动通信产品软件研发工作。2018 年于西安交通大学获得博士学位（计算机科学与技术）。2018 年至 2021 年分别于英国布里斯托大学和德比大学任博士后研究员、Research Associate 职位。2021 年 9 月入职西安交通大学，入选西安交通大学“青年拔尖人才计划”。2022 年入选陕西省高层次人才青年项目。研究方向为人工智能基础算法研究及其应用研究，尤其是不同监督程度的机器学习算法、多源数据融合算法及其在遥感图像处理及健康医疗数据中的应用。在机器学习及图像处理领域知名期刊（IEEE TIP, IEEE TGRS, IEEE-JBHI 等）和国际会议发表学术论文 30 余篇，其中 ESI 高被引论文 2 篇。长期担任 IEEE TIP, TGRS, TSMCS, IOT, GRSL 等多个知名期刊和 ECML-PKDD 等多个国际会议的论文评审人。主持国家自然科学基金、陕西省高层次人才项目、陕西省科技厅创新创业人才项目等多项国家级和省部级项目；作为项目骨干参与科技部重点研发计划等国家级重大项目。

第四节 项目进展

4.1 进展总述

目前，已能够基本实现实际应用中无人机对于人体动作的识别与应答，包括各层次模型的建立、数据集的收集、模型的深度学习训练、无人机的组装与树莓

派、摄像头的搭载，ROS 系统的嵌入等。

（此处待插入实机演示画面）

4.2 算法进展

算法实现大致分为三部分内容，即人体姿态动作检测与切分、姿态视频分析与智能选帧、人体指示动作的识别与控制指令的输出。

4.2.1 人体姿态动作检测与切分

项目实施过程中，利用 YOLOv5 模型实现了对视频帧中人物区域的实时检测。YOLOv5 的高效性使得系统能够以较快的速度运行，满足了无人机控制对实时性的要求。通过对视频帧的检测，能够准确地框选出人物区域，并根据视频的所有帧确定了统一的框选尺寸，确保了检测结果的稳定性和一致性。

在目标检测的基础上，对检测到的人物区域进行了裁剪和归一化处理。项目组开发了一套数据预处理流程，包括裁剪人物区域、调整尺寸和归一化操作，确保了输入到动作分类模型的数据具有高度的一致性和标准化，减少了模型训练过程中的不稳定性。经过预处理的数据集为后续的动作分类任务提供了坚实的基础。

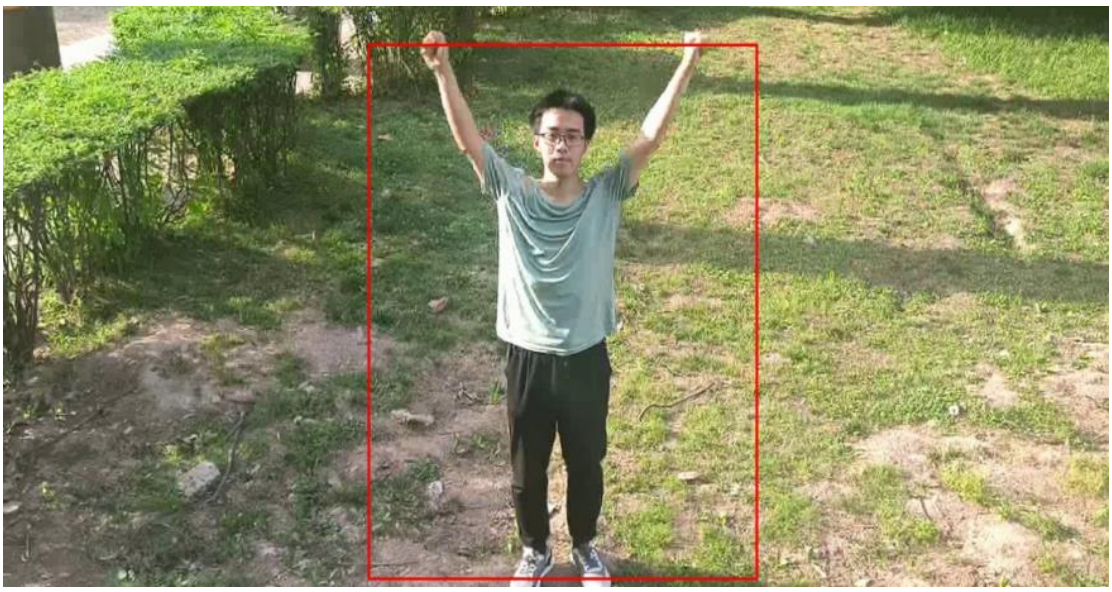

图 11 图像框选示意

之后，使用谷歌的 MediaPipe 框架作为基础技术。经初步研究，成功实现了从视频帧中提取人体关键点坐标的功能，由此构建了一个 50 维度的特征向量。这一步骤为后续的算法开发奠定了基础。

后续迭代中，设计并实现了一个多层感知器（MLP）神经网络，用于对提取的特征向量进行分类。同时，基于动作检测的相关识知和先验经验，归纳出了一系列动作特征，并将其编写为模态逻辑规则。这两个分类器的初步版本在内部测试集上进行了评估，结果显示 MLP 在处理复杂动作时表现出色，而模态逻辑规则在处理简单和重复动作时更为准确。

为进一步提升系统的准确性，引入了堆叠泛化技术。通过逻辑回归模型，初期尝试融合 MLP 和模态逻辑的预测结果。在多次实验和参数调整后，发现这种融合策略能够显著提高分类的准确率。

训练过程中，采用随机梯度下降法（SGD）来优化逻辑回归模型的参数。经过数轮的训练和验证，模型在测试集上达到了 **96.33%**的精确率。这一成果超出项目组的预期，并证明算法在动作检测任务上的有效性和准确性。

4.2.2 人体指示动作的分析与智能选帧

算法实现的过程中，项目组使用 SMART 策略，通过两个不同的选择器对视频帧进行评分，并最终选择得分最高的 k 帧进行动作分类。这一过程的关键在于两个选择器的训练和集成。

首先，单帧选择器的设计和训练是基于模型蒸馏的方法，使用一个轻量级的多层感知器（MLP）来学习一个高性能但计算成本较高的 oracle 模型（即 EfficientNetV2 模型）的行为。通过这种方式，可以有效地减少计算成本，同时保持较高的动作识别准确率。单帧选择器的训练包括两个阶段：第一阶段是训练 oracle 模型以获得高质量的动作分类分数；第二阶段是使用这些分数作为目标，训练 MLP 以模仿 oracle 模型的行为。

其次，全局选择器的实现则侧重于捕捉和理解视频序列中的时间关系。采用了一种基于关系的模型，该模型能够从连续帧中提取特征，并通过长短期记忆（LSTM）网络来处理这些特征。LSTM 网络能够捕捉时间序列数据中的长期依赖关系，这对于理解和评分视频中的动作至关重要。全球选择器的训练涉及到端到端的优化过程，其中 LSTM 网络与关系模型一起被训练来预测每一帧对动作分类的重要性。

在集成两个选择器的过程中，项目组采用了一种简单的乘法策略，即将单帧选择器和全局选择器的评分相乘，以得到每一帧的最终评分。这种方法能够有效地结合两个选择器的优势，即单帧选择器对单帧动作信息的敏感性和全局选择器对时间序列关系的理解。

为进一步优化算法，项目组还探索了不同超参数 k 的影响，即选择不同数量的 Top-k 帧进行动作分类，以确保算法的有效性和鲁棒性。

算法进展包括了单帧选择器和全局选择器的实现、训练和集成，以及对超参数的调优。通过这些努力，能够实现一个既轻量级又高效的视频动作分类系统，这对于手势控制无人机项目来说是至关重要的。

4.2.3 人体指示动作的识别与控制指令的输出

数据预处理完成后，项目组引入了一种创新的算法流程，即双流网络架构，结合了基于 ResNet50 的 I3D 模型与项目组自研的 S2GCN（Spatio-temporal Seperable Graph Convolutional Network）网络。不仅能够利用 I3D 模型强大的时序信息捕捉能力，还能够通过 S2GCN 高效地分析人体骨架的运动特征，双流结构大幅提升了动作分类的准确性和稳定性。

I3D 模型采用 mmaction2 库实现，该库提供了高效的视频处理和动作识别工具。同时，S2GCN 网络的加入使得算法能够更加专注于人体骨架的动态变化，进一步提高了动作识别的精确度。通过在动态图构建中使用 RGCN 和带有指数衰减的平均池化，S2GCN 网络能够捕捉到关键的空间和时间特征，同时保持轻量级的特点。

经过一系列的训练和优化，模型在自定义的人体动作数据集上达到了 **94%**的 Top-1 准确率，这一优异表现证明了模型在解析视频帧时序信息和识别复杂动作模式方面的卓越能力。

在动作分类的基础上，项目组开发了一套命令生成系统，能够根据双流网络架构输出的分类结果，自动转换为无人机的飞行动作命令。例如，当模型识别出“上升”动作时，系统会生成相应的上升命令，利用 ROS 控制无人机执行上升操作。通过这种方式，实现了从人体动作到无人机飞行动作的紧密衔接，使得无人机能够根据操作者的动作实时调整飞行状态。

4.3 系统进展

在树莓派上成功搭建了 ROS（Robot Operating System）环境，整个系统已经具备了强大的机器人开发和操作能力。我们精心挑选并连接了与树莓派兼容的摄像模块，确保高质量的图像捕捉和实时性。完成摄像模块的连接后，我们对相机进行了细致的调试，确保它能够稳定地捕获外界信息，并清晰、流畅地输出视频数据。

在 ROS 的架构下，我们构建了一个专门的摄像节点，负责从摄像模块读取视频流，并将其作为 ROS 消息发布出去。这个节点在系统中起到了桥梁的作用，将物理世界中的视觉信息转化为 ROS 能够理解的数字信号。通过这个摄像节点，我们能够实时地获取到周围环境的信息，为后续的分析和决策提供了重要的数据支持。

在获取到视频数据后，我们将其传递给一个专门的分析节点。这个分析节点是我们整个系统的核心部分，它负责处理来自摄像节点的图像信息，并通过预先构建的通信算法与其他节点进行信息交换。

分析节点通过复杂的算法对图像信息进行分析，提取出有用的特征和数据。这些特征和数据不仅可以帮助我们了解人体的运动状态，还可以为我们提供关于环境、物体等更多方面的信息。在得到分析结果后，分析节点会将结果转化为运动信号，并将其作为 ROS 消息发布出去。

运动信号是我们在系统中定义的一种特殊消息类型，它包含了关于人体运动的详细信息。其他节点可以通过订阅这些运动信号来获取到这些信息，并根据这些信息执行相应的任务或做出决策。通过这种方式，我们实现了一个高效、灵活的信息传递和处理机制，为整个系统的运行提供了有力的支持。

综上所述，我们的 ROS 系统已经具备了从摄像模块获取视频数据、通过摄像节点将视频数据转化为 ROS 消息、通过分析节点对图像信息进行分析并得到分析结果以及运动信号的能力。这些功能的实现为我们后续的开发和应用奠定了坚实的基础。

4.4 硬件进展

在无人机平台的搭建和测试过程中，取得了显著成果。在 Mission Planner 地面站中，成功识别并配置了无人机飞控系统，并在无桨状态下对飞控的 GPS 识别、俯仰角识别以及不正常姿态下的电机自动调节转速等功能进行了测试，所有测试均成功完成。

此外，项目组还在地面站中设置了自动返回、悬停等飞行模式，这些模式不仅提升了无人机的飞行安全性，还增强了其可靠性和稳定性。在已经完成的实验中，充分验证了无人机机身结构的牢固性和稳定性，证明了其能够胜任各种复杂环境和任务需求。

在无人机平台的设计中，始终将安全性放在首位，采取了多项措施来防止无人机在飞行过程中发生碰撞或失控，确保其在各种环境下都能保持安全稳定的状态。同时，也将继续优化和升级无人机的性能，以更好地满足未来项目的需求。

组装无人机实物图如下：

图 12 搭建无人机实物图

第五节 项目拓展

5.1 雨雾天气增强

5.1.1 多模态信息融合

在雨雾天气下，传统的 RGB 相机受环境影响较大，图像质量会显著下降，导致人体动作识别的准确性和稳定性不足。为了增强系统在恶劣天气条件下的鲁棒性和可靠性，我们可以引入多模态信息融合算法。多模态信息融合是指利用来自不同传感器或信息源的数据进行融合，以获得更全面和准确的环境感知能力。具体来说，我们可以结合 RGB 相机与其他类型的传感器，例如红外相机、激光雷达（LiDAR）、毫米波雷达等。

5.1.2 多模态信息融合算法设计

1.数据预处理：

同步与校准：在多模态数据融合之前，首先需要对来自不同传感器的数据进行时间同步和空间校准，以确保数据在同一时刻和空间参考系下进行融合。

数据对齐：将不同传感器的数据对齐到共同的参考坐标系中，例如将 LiDAR 点云数据投影到 RGB 图像平面上，或者将红外图像与 RGB 图像进行配准。

2.特征提取：

单模态特征提取：从每种传感器的数据中提取有用的特征，例如从 RGB 图像中提取颜色和纹理特征，从红外图像中提取热辐射特征，从 LiDAR 点云中提取几何形状特征，从毫米波雷达中提取运动特征。

特征融合：将从不同传感器提取的特征进行融合，可以采用特征拼接、特征加权等方法。例如，可以将 RGB 图像的视觉特征与红外图像的热特征进行拼接，形成多模态特征向量。

3.融合策略：

早期融合：在特征提取阶段进行融合，将来自不同传感器的原始数据进行融合，再输入到统一的特征提取模块中。

中期融合：在特征提取完成后进行融合，将各自提取的特征进行结合，再输入到统一的决策模块中。

晚期融合：在单模态决策完成后进行融合，将各自的决策结果进行综合，形成最终的决策结果。

4.决策层融合：

加权平均：对不同传感器的决策结果进行加权平均，根据各自的置信度分配权重，形成综合决策。

贝叶斯融合：利用贝叶斯理论对不同传感器的决策结果进行融合，计算后验概率，形成最优决策。

深度学习融合：利用深度学习模型（如卷积神经网络、递归神经网络等）对多模态特征进行融合和决策，通过端到端学习提升融合效果。

5.1.3 预期效果

去噪：多模态信息融合可以显著减少噪声干扰。例如，在雨雾天气下，RGB图像可能受到光学噪声的影响，而红外相机和雷达传感器则不受此影响。通过融合不同传感器的数据，可以有效去除噪声，提高识别的准确性。

抗干扰：不同类型的传感器对环境干扰的敏感度不同。通过融合多模态数据，可以增强系统对环境干扰的抵抗能力。例如，雷达传感器可以穿透雨雾，而 RGB 相机则可以提供详细的视觉信息，两者结合可以互为补充。

增强鲁棒性：多模态信息融合可以提高系统的鲁棒性，使其在各种复杂环境下都能稳定运行。无论是强光、低光、雨雾还是其他恶劣天气条件，多模态融合技术都能提供可靠的数据支持，确保无人机的远程控制和动作识别的稳定性。

5.2 端侧多模态大模型

5.2.1 结合语音和视觉对无人机动作作决策

在远程控制无人机的过程中，单一的视觉信息往往不足以应对复杂多变的环境和任务需求。为了提高系统的智能化水平和决策能力，可以结合语音和视觉信息，利用端侧多模态大模型对无人机动作进行综合决策。

5.2.2 多模态大模型算法设计

1.语音识别与指令解析：

语音预处理：对语音信号进行预处理，包括降噪、归一化等操作，提取语音特征。

语音识别模型：利用深度学习模型（如卷积神经网络、长短期记忆网络等）进行语音识别，将语音信号转化为文字指令。

指令解析：对识别出的文字指令进行解析，提取出具体的控制命令。

2.视觉感知与环境理解：

视觉预处理：对 RGB 图像和其他传感器数据进行预处理，包括调整图像尺寸、归一化等操作。

视觉特征提取：利用卷积神经网络等深度学习模型从图像中提取特征，例如物体检测、人体动作识别等。

环境理解：结合视觉特征和先验知识，对环境进行理解和建模，识别出障碍物、目标物体和人体动作等信息。

3.多模态大模型融合：

特征融合：将语音特征和视觉特征进行融合，可以采用特征拼接、特征加权等方法。例如，可以将语音指令的特征向量与视觉感知的特征向量进行拼接，形成多模态特征向量。

决策模型：构建多模态大模型（如多模态 Transformer、融合神经网络等），将融合后的特征输入模型中，通过深度学习算法进行综合决策。

4.动作决策与控制：

动作规划：根据多模态大模型的决策结果，规划无人机的动作和路径。

控制执行：将规划好的动作转化为具体的控制命令，通过控制算法（如 PID 控制、模型预测控制等）执行无人机的飞行和操作。

5.2.3 预期效果

自然交互：通过结合语音和视觉，用户可以以更加自然和直观的方式与无人机进行交互。语音指令可以简化操作流程，提高用户体验和操作效率。

智能决策：多模态大模型可以综合分析和处理多种信息源，做出更加智能化的决策。例如，在复杂环境中，系统可以结合语音指令和视觉数据，自动调整飞行路径，避免障碍物，确保任务顺利完成。

增强鲁棒性：多模态大模型可以提高系统在复杂环境下的鲁棒性和适应性。例如，当语音信号受到干扰时，系统可以依靠视觉信息进行补偿，确保无人机的稳定控制和操作。

5.1.4&5.2.4 实现路径

数据采集与标注：收集大量的语音和视觉数据，并进行精确标注，构建多模态数据集。数据集应覆盖各种飞行环境和任务场景，确保模型的泛化能力和适应性。

模型训练与优化：利用先进的深度学习算法和多模态融合技术，训练端侧多模态大模型。在训练过程中，可以采用迁移学习、数据增强等技术，提高模型的性能和鲁棒性。

系统集成与测试：将训练好的多模态大模型集成到无人机控制系统中，进行全面的功能测试和性能优化。通过实际飞行测试和用户反馈，不断改进和优化系统，确保其在各种复杂环境下的稳定性和可靠性。

5.1.5&5.2.5 未来展望

实时性与高效性：随着计算机硬件性能的不断提升，未来的端侧多模态大模型将能够在无人机上实现实时、高效的语音和视觉处理与决策，提高无人机的自主性和智能化水平。

多任务协同：未来的多模态大模型将能够支持多任务协同，实现更加复杂和多样化的任务。例如，无人机可以同时执行跟随、避障、拍照等多项任务，并在任务之间进行智能切换和协同工作。

人机协同与智能交互：未来的无人机系统将更加注重人机协同和智能交互，通过多模态信息融合和大模型决策，提高人机协同工作的效率和效果，实现更加智能化和便捷化的无人机操作体验。

5.1.6&5.2.6 应用拓展

未来的多模态大模型无人机系统将在更多领域得到应用和推广。例如，在农业、物流、安防、救援等领域，无人机可以通过多模态信息融合技术，提升工作效率和精度，带来更多社会和经济效益。

综上所述，通过雨雾天增强和端侧多模态大模型的技术拓展，我们的大学生创新创业项目将能够在复杂环境下实现更加稳定和智能的无人机远程控制和人体动作识别。未来，随着技术的不断进步和应用的拓展，我们的项目将具备更广阔的发展前景和应用价值。

5.3 摄像设备拓展

5.3.1 红外摄像头

后续将对摄像设备进行优化，固然 Camera Module 3 已经可以满足基本使用需求，但仍存在一些缺陷。可将其拓展为红外摄像头。

Camera Module 3 缺点:

夜视限制：标准版不专为夜间设计，虽然有 NoIR 版本去除红外截止滤镜，但需要外部红外光源辅助才能在黑暗中成像。

光线依赖：在极低光环境下，如果没有额外的光源辅助，其成像质量会显著下降。

5.3.2 红外摄像机技术原理

红外摄像机是将摄像机、防护罩、红外灯、供电散热单元等综合成为一体的摄像设备。数码摄像机用 CCD 感应所有光线这就造成所拍摄影像和我们肉眼只看到可见光所产生的影像很不同。为了解决这个问题，数码摄像机在镜头和 CCD 之间加装了一个红外滤光镜，其作用就是阻挡红外线进入 CCD,让 CCD 只能感应到可见光，这样就使数码摄像机拍摄到的影像和我们肉眼看到的影像相一致了。目前大多数的红外摄像机都采用 LED 红外发光二级管作为红外摄像机的主要材料。

红外摄像技术分为被动红外摄像技术和主动红外摄像技术。

被动红外摄像技术是利用任何物体在绝对零度(一 273℃)以上都有红外光发射的原理。由于人的身体和热物体发出的红外光较强，其它非发热物体发出的红

光很微弱，因此，利用特殊的红外摄像机就可以实现夜间监控。被动红外摄像技术由于设备造价高且不能反映周围环境状况，因此在夜视系统中很少被采用。主动红外技术利用特制的“红外灯”人为产生红外辐射，这些红外光辐射照明景物和环境。摄像机通过图像传感器感受周围环境反射回来的红外光，获取比较清晰的黑白图像画面，实现夜视监控。在夜间或恶劣天气条件下，能够正常工作，准确检测到目标物体。目前的红外摄像技术多数采用主动红外摄像技术，技术相对成熟，广泛应用在监控领域。

5.3.3 红外摄像机对于本课题的优势性

对于“人体姿态控制无人机运动”这个课题，将 Camera Module 3 拓展为具备红外功能的摄像头具有以下几点必要性和优势：

增强夜间或低光环境下的操作能力：在黄昏、夜晚或昏暗的室内环境中，传统可见光摄像头的性能会大幅下降，难以准确捕捉人体姿态。而红外摄像头即使在几乎没有可见光的条件下也能有效识别目标，使得无人机能够在各种光照环境下稳定执行任务，提高了全天候作业能力。

提高姿态识别精度：人体散发的红外辐射相对均匀且不易受外界光线影响，使用红外摄像头可以帮助更稳定地追踪人体的热量分布，这对于精确识别身体各部分的位置和姿态变化特别有利，从而提升无人机跟随或响应人体动作的准确性。

增加隐蔽性和隐私保护：在某些应用场合下，如安全监控或秘密军事行动，红外摄像头的隐蔽性可以减少被监控对象的察觉，同时保持高效的人体姿态识别，平衡了功能需求与隐私考量。

优化动态追踪和避障：在复杂多变的环境中，红外摄像头可以辅助无人机在低光条件下有效识别障碍物与人体，结合热成像和机器视觉技术，提高无人机的自主避障能力和动态路径规划，确保飞行安全。

提升能源效率：在某些红外应用中，特别是结合主动红外照明时，相较于高功率的可见光照明，红外 LED 的能耗更低，有助于延长无人机的续航时间。

综上所述，将 Camera Module 3 拓展为红外摄像头对于“基于人体动作语言识别的无人机控制”的研究和应用具有重要意义，不仅拓宽了无人机的作业时间窗口，还提升了在复杂环境下的识别和控制精度，是推动该领域技术进步的关键

因素之一。

5.3.4 树莓派可用红外摄像机调研

名称/品牌	型号	价格	特点
Waveshare-H 型 OV5647	RPi Camera (H)	¥143.42	对角视场角（FOV）：160 度
Raspberry Pi	RPi NoIR Camera V2	¥119.18	IMX219 800 万像素，支持红外夜视，兼容树莓派系列、Jetson Nano 和 VisionFive2 等主板

5.4 硬件拓展

5.4.1 无人机平台升级

机架选择与材料革新：虽然图腾 Q250 四轴穿越机机架结构紧凑，但为了应对未来可能搭载的更多设备（如红外摄像头、传感器等），我们建议转向更大的机架，如大疆经纬系列八旋翼无人机机架，其最大起飞重量可达 6 kg，能更好地支持复杂任务。展望未来，随着材料科学的进步，无人机机架将更加轻量化、模块化和智能化。碳纤维复合材料和纳米材料的应用，将造就更轻、更强、更耐用的机架，而模块化设计将简化部件更换和升级流程，智能设计则允许机架根据任务需求自动调整结构，提高适应性。

电机与电池的前沿探索：为满足实际飞行需求，如飞行速度、续航时间，优化电机和电池的选择至关重要。未来视角显示，新型电机如永磁同步电机和直驱技术，以及固态电池等先进电池技术，将极大提升无人机的性能。这些技术不仅提高效率和能量密度，还增强安全性，预示着无人机将拥有更远的飞行距离、更长的续航时间和更快的速度，拓宽了无人机的应用边界。

飞控系统的智慧升级：考虑使用更先进的飞控系统，集成避障、路径规划等功能，提升无人机的自主性和智能化。未来趋势是飞控系统将变得更加智能和自主，集成了多种传感器和人工智能算法，实现精准定位、导航、避障及智能任务执行。通过深度学习和强化学习，无人机将学会适应各种飞行环境和任务，自主完成复杂操作，如编队飞行、搜救任务、环境监测等，同时保证飞行的安全性和可靠性。

5.4.2 传感器与扩展接口的未来融合

传感器：感知世界的进化

增加传感器，如测距、测速和姿态检测传感器，能显著提升无人机的感知能力和飞行稳定性。例如，超声波传感器或激光雷达可用于测距，霍尔传感器或编码器用于测速，而加速度计、陀螺仪或磁力计则用于姿态检测。展望未来，无人机将集成视觉传感器、激光雷达、毫米波雷达等多样化传感器，形成多传感器融合系统，这不仅能增强环境感知，还能利用深度学习和强化学习技术，更准确地理解环境和任务需求，提升复杂环境下的适应性和任务执行能力。

扩展接口：模块化的无限可能

增加扩展接口，如 USB、串口或 GPIO 接口，便于连接相机、传感器等设备，为无人机的功能扩展提供了便利。未来视角中，无人机将展现出更高的模块化和可扩展性。更多的扩展接口将允许用户轻松连接各类设备，从相机、传感器到机械臂，满足多样化的任务需求。借助软件定义无线电技术，无人机将实现无线连接和数据传输，极大地增强了其灵活性和应用场景的多样性，为无人机技术的未来开辟了广阔前景。

5.4.3 硬件设计与集成：前瞻性的创新融合

轻量化设计：材料与结构的革新

在确保无人机性能的基础上，轻量化设计是提升飞行效率的关键。采用碳纤维复合材料或泡沫塑料等轻质材料制造机架和外壳，以及优化电机和电池选型，都是实现轻量化的有效途径。展望未来，无人机设计将更加聚焦于轻量化，利用石墨烯、碳纳米管等前沿材料，以及拓扑优化和仿生设计等创新方法，构建更轻、更强、更耐用的无人机结构，确保在保持高性能的同时，实现重量的最小化。

模块化设计：灵活性与维护性的双重提升

模块化设计不仅便于维护和升级，还提升了无人机的灵活性。将电机、电池、飞控系统等组件划分成独立模块，有助于快速响应不同任务需求。未来视角中，无人机将展现出更高程度的模块化和可维护性。通过热插拔技术，损坏模块的更换将更加便捷，显著降低维护成本，同时增强无人机的可靠性，确保其在复杂环境下的稳定运行。

散热设计：高效冷却技术的引入

面对硬件设备数量增多带来的散热挑战，设计合理的散热方案至关重要。使用散热片、风扇等传统散热措施的同时，未来方向指向更高效的散热解决方案。石墨烯、碳纳米管等新材料的散热片，以及液冷或相变冷却技术的应用，将大幅提升散热效率，确保无人机即使在高温条件下也能维持稳定运行，延长使用寿命，适应更广泛的作业环境。

5.4.4 综合考量：优化中的前瞻性策略

在推进硬件优化的过程中，我们必须全面考虑成本控制、安全性和可维护性，确保无人机系统的整体效能。

成本控制：经济性与效率并重

选择性价比高的硬件设备是成本控制的核心。采用国产电机、电池和飞控系统等高性价比选项，能在保证性能与可靠性的前提下，有效降低开支。展望未来，无人机产业将愈发重视成本效益，通过 3D 打印技术生产零部件，大幅削减制造成本；借助开源软硬件资源，减少研发投入。这些举措将推动无人机技术更加普及，拓展其应用范围至更多领域，如农业、物流和监测等。

安全性：智能技术的守护

确保无人机在各类环境下的安全飞行是不可忽视的责任。通过增强飞行控制系统（如 GPS 定位、姿态稳定系统）和避障系统（如激光雷达、毫米波雷达），无人机的安全性能得以显著提升。未来视角下，无人机将集成更先进的人工智能技术，如深度学习与强化学习，实现智能化的飞行控制与避障；采取多冗余设计（如多电机、多电池、多传感器），显著增强系统的可靠性。这将赋予无人机在复杂场景中安全稳定飞行的能力，拓宽其在搜救、灾害监控等高风险领域的应用。

可维护性：便捷与智能的维护方案

设计时充分考虑可维护性，通过模块化结构和故障诊断系统（如传感器数据监测、故障代码显示），简化日常维护流程，加速故障排查。未来方向中，无人机将展现出更出色的可维护性，采用易于拆装的设计，便于用户自行更换与修复部件；提供远程诊断及维护服务，确保用户能迅速解决遇到的技术难题。这不仅将提升无人机的长期可靠性，还将显著减轻用户的维护负担，促进无人机技术在个人与企业用户间的广泛采用。