# 人工智能能引论项目报告——第三组

1. 项目计划
   1. 首要目标

本项目的首要目标是，使用Tello无人机，借助一个使用WIFI信号与其进行通信的电脑的算力，使其能够自动导航，从教室的任何一个地方起飞后降落在讲桌上。

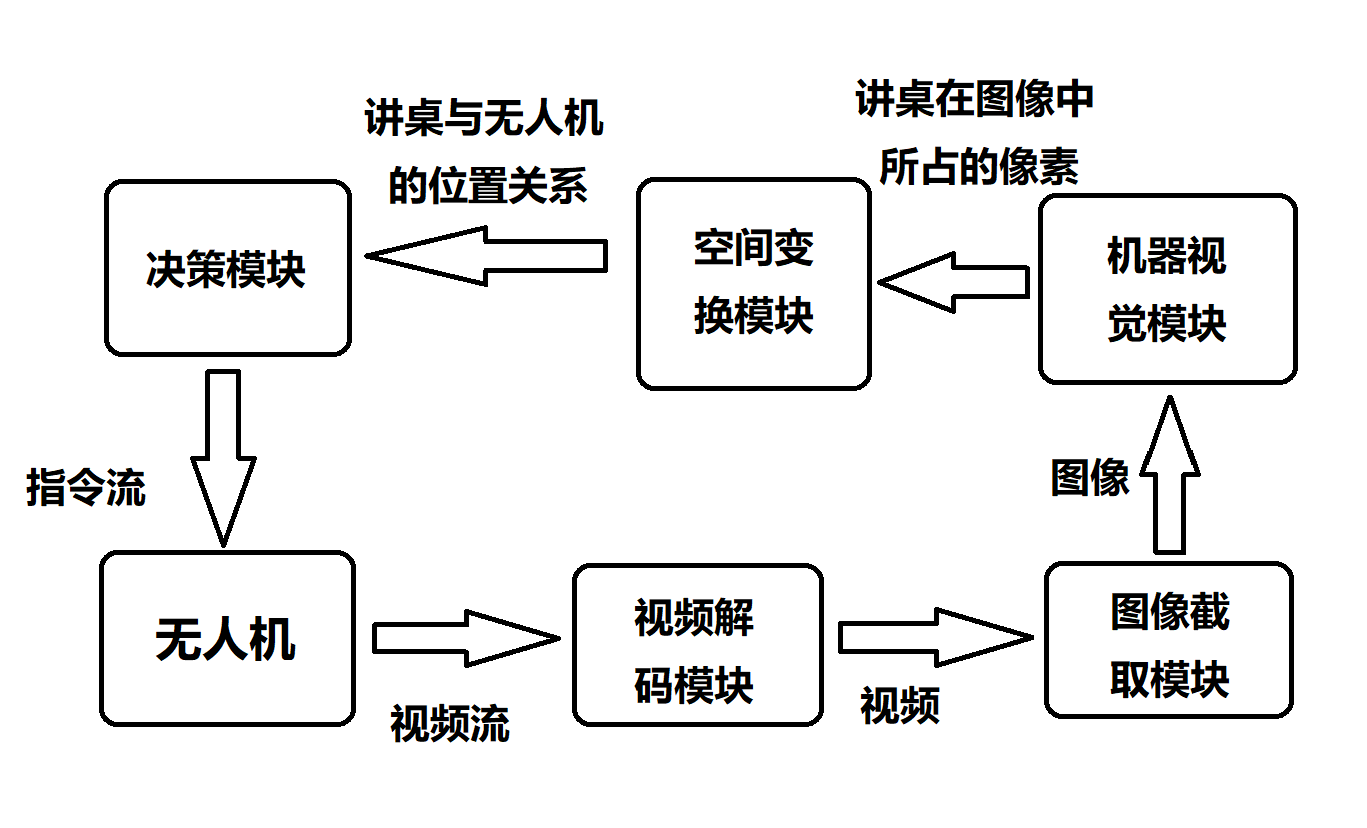
* 1. 初期目标

由于讲台在视频影像中体积比较小、周围环境比较复杂，因此我们首先希望使Tello可以通过自动导航降落在讲台上。

* 1. 探索性目标

如果上述目标顺利完成，我们还有如下探索的方向：

1. 尝试给Tello观看任意一个物品，使其可以在短时间内“认识”该物品，并跟随这个物品移动或降落在这个物品上。
2. 使Tello可以通过目标检测算法识别人体关节的位置，并对人的手势做出反应。
   1. 实施方案



为了实现上述目标，我们设计了如图所示的流程图。无人机首先自动执行起飞程序，开始向计算机传回套接字视频流。通过视频解码模块将其转化为可以用程序读取的格式。图像截取模块从视频流中每隔一段时间截取一帧（具体时间间隔取决于整体的运算速度）。将这一帧图像输入机器视觉模块，通过语义分割等算法，确定讲桌在图像中的范围。再根据这个范围通过灭点法确定无人机和讲桌的位置关系。最后，通过决策模块，利用控制理论的知识引导无人机逐渐缩小其与桌子的距离，最终实现降落。

1. 设备配置

本项目使用大疆Tello无人机，其主要的技术参数如下：

* 照片：500万像素
* 视场角：82.6°
* IP：192.168.10.1
* 命令接收和回复端口：8889
* 状态传输端口：8890
* 视频传输端口：11111
* 视频格式：libh264
* 视频清晰度：HD720P30

1. 应用的方法
   1. 硬件交互方面
      1. 网络通信

在Tello与电脑之间进行通信的过程中使用WIFI局域网。由于视频流的传输不需要弥补损失掉的帧，该过程使用UDP(User Datagram Protocol)协议。Tello和计算机之间以数据报套接字(SOCK\_DGRAM)的形式进行通信，将命令、状态、视频等信息都以二进制字符的形式传输，再通过设备上的软件进行解码。套接字通信可以通过Python中的socket模块实现。

* + 1. 多线程

计算机在于Tello进行通信的过程中需要同时作出发送指令、接收回复、接收视频等操作，因此要使一个运行中的代码同时地、持续地进行多项任务，这就要使用到多线程。多线程可以通过Python中的threading模块来实现。虽然Python中的threading库不能实现多核计算，只是将计算资源在极短的时间内依次分发给不同进程，但是也已经足以满足我们的需求。

* 1. 图像识别方面
     1. 问题归类：

最开始我们将目标“无人机停在讲桌上”定性为目标检测，在查阅了目标检测相关的文献代码，试用了yolo3后发现大多数目标检测算法具有对多目标划分及识别的功能，但没有对对象的筛选功能，教室里桌子和讲桌类别都为dining table。



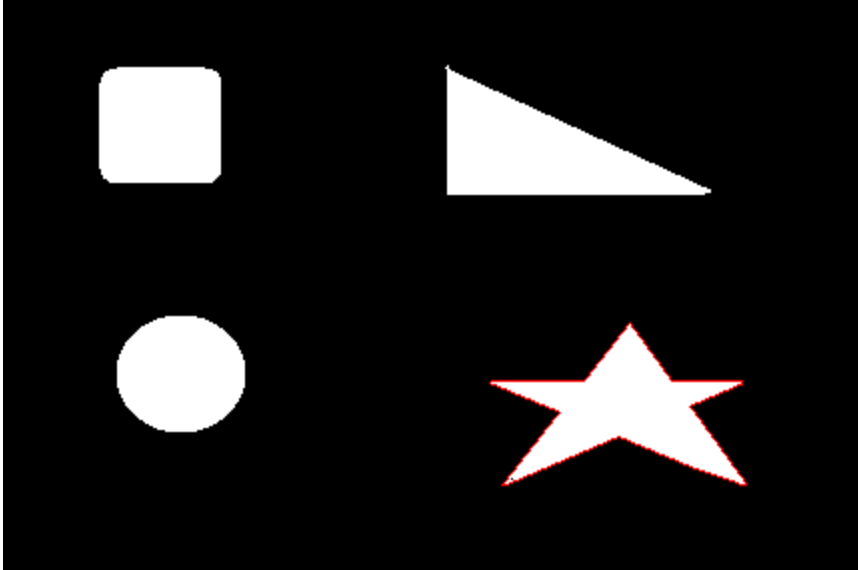
yolo3检测出的图片结果

之后我们调研了单目标检测，其主要方法是通过轮廓分割出图像中不同的轮廓后（相当于将图片分块）选出某个特定的部分，此方法和我们的选题目标很接近

* + 1. 单目标检测流程
       1. 将图片分割

方法1：轮廓检测，使用opencv findcontours函数

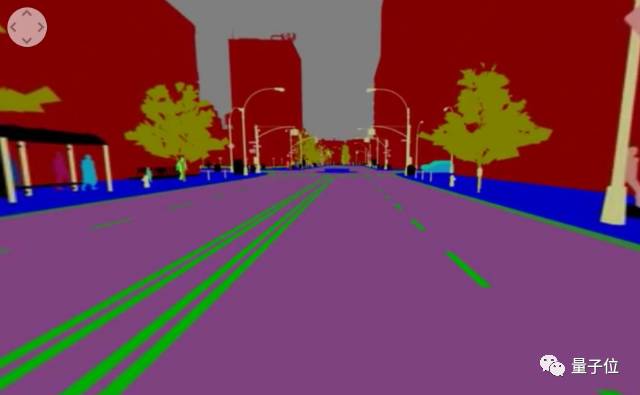
检测效果：如下是检测效果，可以看出轮廓检测在划分大体几何图形上有较好的表现，可以应用于我们检测教室中讲桌的目的：



方法2.语义分割

方法：使用神经网络训练

效果：如下是语义分割的效果，可以看出相比较轮廓划分其效果更佳细节。



* + - 1. 将讲桌从多个部分中识别出来

方法1：特定情况下的识别

原理：此方法较为讨巧，是我们考虑到进行神经网络训练没有足够的数据集且效果不一定好（讲桌和桌子之间的识别即使通过神经网络训练也较为困难）的情况下想出的替代措施。即是通过计算轮廓区域的面积算出最大的面积作为目标。

局限：局限在于只能识别距离足够近的物体。

方法2：CNN神经网络识别

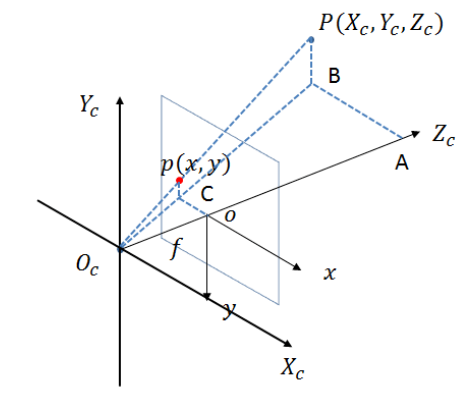
原理：此方法如果采取语义分割可以衔接，识别出标记为讲桌的目标

局限：没有足够的数据集，训练效果可能不好

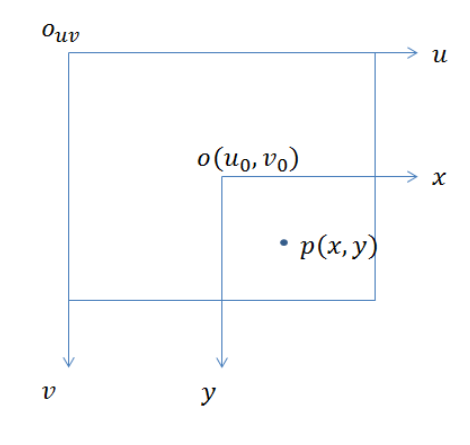
* 1. 空间变换方面

当我们通过图像识别获得了讲桌的像素坐标时，下一步是通过坐标变化，把二维像素点转换到三维的世界坐标系中，确定讲桌中点在世界坐标系的三维坐标。反过来获得无人机在教室中的位置信息，进而无人机自动判与讲桌中心的距离和相对位置，来确定飞行方向和距离。

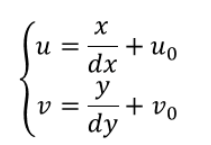
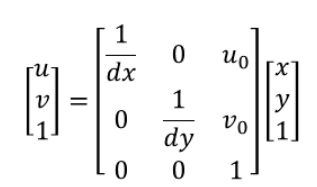
完成此任务的主要思路为坐标转换，主要涉及的坐标系有世界坐标系、相机坐标系、图像坐标系、像素坐标系。世界坐标系是为了描述相机的位置，原点设计在靠近讲台的墙角处（需要除去讲台的高度），以教室的三条墙角缝为坐标轴；相机坐标系是以无人机相机的光心为原点，透光的直线方向为坐标轴，假设为Z轴；在图像坐标系中，原点位于Z轴上，与相机坐标系原点相距f（焦距），它是图片的中心；像素坐标系的原点为图像左上角，单位为pixel。在下图中，P是世界坐标系中的一点，即生活中真实的一点，p是图像中的成像点，在图像坐标系中的坐标为（x，y），在像素坐标系中的坐标为（u，v）。



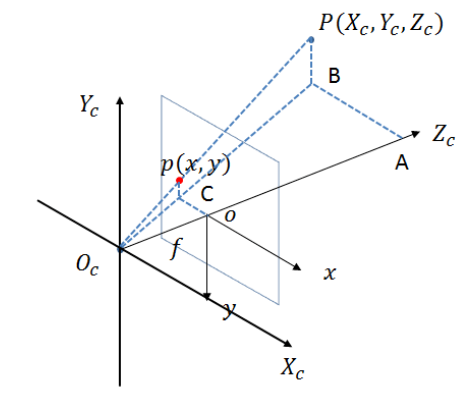
（1）首先构建物理尺寸桥梁，用图像坐标系替换像素坐标系。



已知p在像素坐标系中的坐标为（u，v），它的单位是pixel，并且每一像素的行和列的宽度分别为dx，dy。那么（x，y）和（u，v）的对应关系为：

（2）再根据透视投影关系进行刚性变化，得到相机坐标。



如上图所示，△ABP和△oCOc为相似三角形，那么存在几组比例关系：



带入数据得：



进一步整理成矩阵形式：



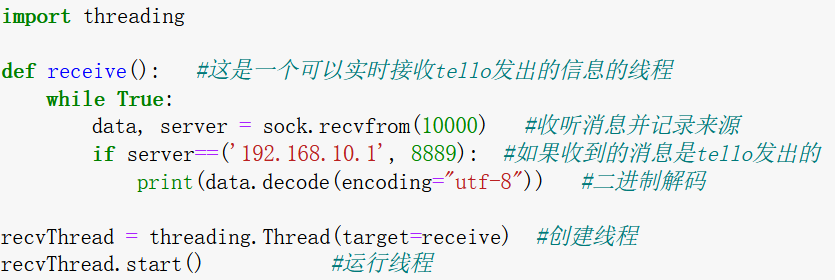
（3）最后比较讲桌的相机坐标系坐标和世界坐标系坐标，确定通过怎样的变化将世界坐标系变化为相机坐标系。

1. 现有的探索
   1. 硬件方面
      1. 网络通信



使用计算机与Tello进行通信的基本代码实现原理如上图所示。运行后，我们成功将socket绑定在恰当的接口。再运行代码：sock.sendto(b“command”.encoding=”utf-8”, tello\_address)，可以看到Tello的指示灯从黄色闪烁状态变为了绿色，表示Tello已经收到了指令并进入了二次开发模式。

* + 1. 多线程



使用多线程的方法如上图所示。将一个while True的永久循环打包成一个线程运行起来之后，就可以实时收听某个端口发出的消息。使用同样的方式，我们同样可以持续地对无人机传回的视频流进行接收。但是目前仍然存在视频流解码的问题，下一步我将着力解决。

1. 下一步的工作
   1. 硬件交互方面：进行视频流的解码，使计算机可以实时将无人机的影响传入机器视觉模块进行飞行决策
   2. 数据集构建方面：使用labelme对教室内的影像进行标注，构建语义分割数据集
   3. 机器视觉方面：尝试使用轮廓检测、语义分割等方法，对讲桌进行提取，并使其能和教室内其余的桌子区分开来
   4. 无人机控制方面：尝试借助控制理论进行调节，尽可能摆脱对三维几何的依赖