# 无人机项目第二次进度报告

王玺 李卓潼 金泊翰

## 项目构想

本项目以无人机的图像识别和自主决策为核心，在电脑端处理无人机的视频数据、开展决策，以控制无人机完成某种目标，并展现出智能化的特性。目前的探索方向主要是如下两个：

### 人脸跟踪

通过无人机拍摄得到的人脸照片，识别出人脸的范围和人脸的中心。当无人机认识到自己的镜头没有对准中心，就会想相反方向调节，直到正对人脸为止。

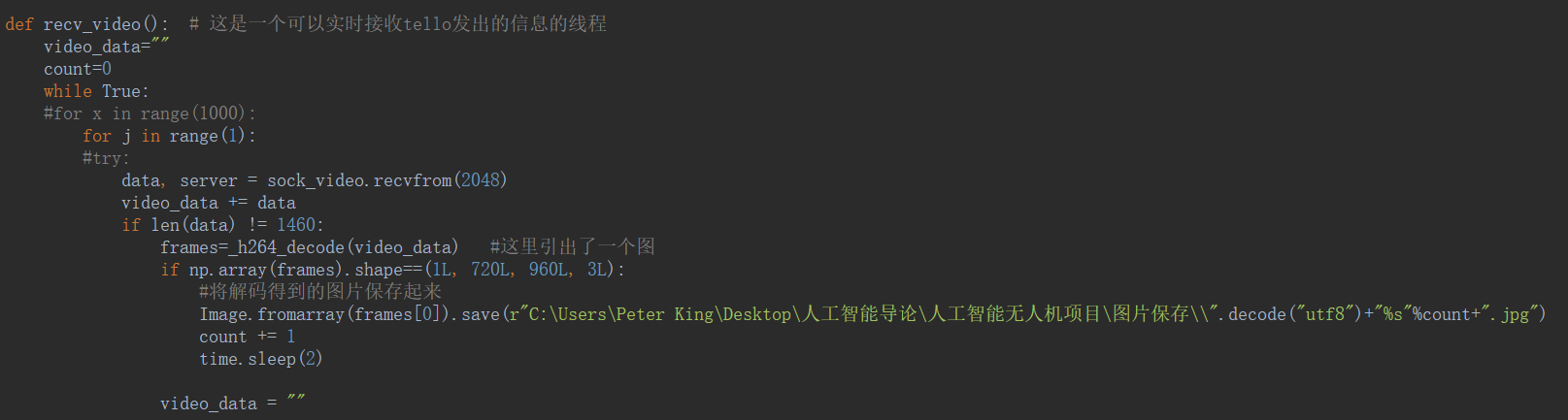
### 讲桌降落

我们希望无人机在教室的任何一个地点起飞后，可以自动地朝讲桌的方向飞行，并落在讲桌上。这就需要无人机对于图像中的讲桌可以进行辨识，根据讲桌和自己的位置关系调整飞行决策，最终实现降落。

## 二、实现情况

### 视频解码模块

在上一个阶段的工作中，我们已经完成了视频流的接收、解码工作。在之前我们充分研究了socket套接字库、threading多线程库的基础上，目前已经可以实现通过调用官方给出的h264解码器对视频进行解码，每隔2秒取一张图片，并保存为np.array的格式，进而将其传入模型中实现判别，最终指导无人机进行决策。实现视频解码的代码核心部分如下图所示，无人机拍摄的视频效果请见附件。



### 人脸识别模块

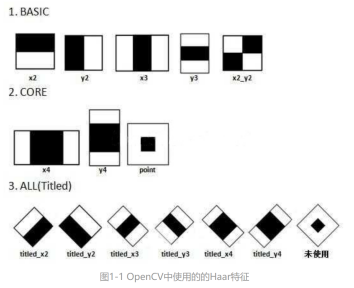
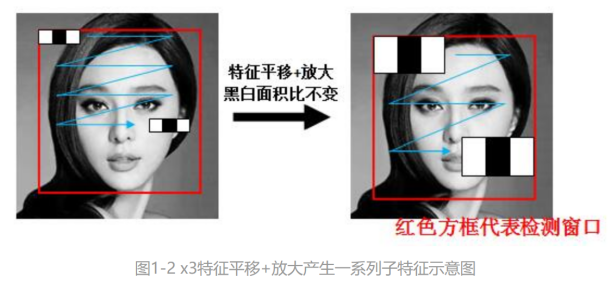
#### 1.Opencv cascade级联分类器原理：

**（1）特征的获取**

①在分类器的实现中采用由Paul Viola等人提出的Haar特征。其获取步骤概述如下：

***Step1：***

采用不同的特征形式检测窗口在图片中放大，平移，得到不同大小，不同位置的Haar系列特征。不同检测窗口获取不同类型的图像信息，如横向信息，纵向信息。



***Step2：***

计算Haar特征值。其公式为：



即Haar特征值=整个Haar区域内像素和×权重 + 黑色区域内像素和×权重：不同特征检测窗口的对应的各区域权重不同。图1-2中的x3特征的weight\_black=-3，相当于白色区域像素-2\*黑色区域像素。

***Step3：***

标准化特征值。由于计算特征量时用到区域内像素和，导致仅仅12\*18大小的窗口计算出特征值变化范围为-2000-6000.不利于量化评定特征值。因此采用计算平均值，计算标准化因子等多种方式对同一位置检测窗口特征值进行标准化。

②特征获取结果：

获取了一系列具有区分图像信息功能的一维特征值，在保存图像信息的同时实现降维。

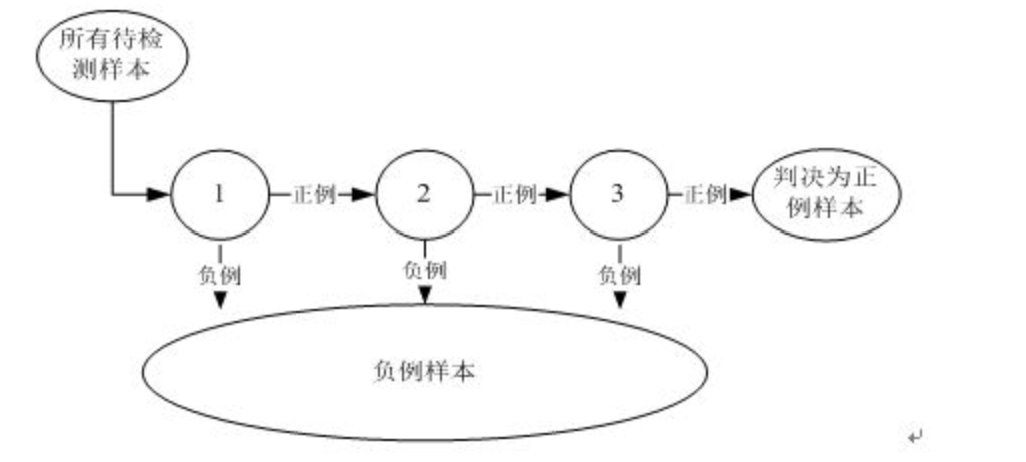
特征值的个数与窗口大小和图片大小有关。

③为什么Haar特征是有效的：

Haar特征的获取过程类似于CNN，特征窗口类似于CNN内核。只不过在CNN中，内核的值由训练确定，而Haar特征是手动确定的。从检测窗口的不同形式来看，这些特征用于检测边缘，线，矩形，斜线。

**（2）Adaboost级联分类器**

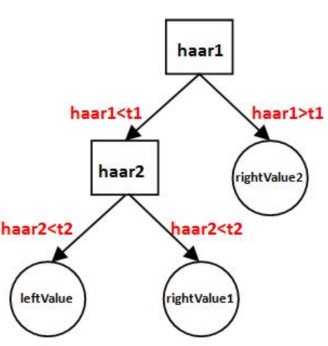
①总体结构：树状结构，每个stage代表一级强分类器。某个检测窗口通过所有强分类器时被认为是目标，否则拒绝。



②弱分类器结构及工作原理

弱分类器也是树状结构。以一个深度为2的弱分类器为例，通过将两个特征值Haar1和Haar2与阈值t1,t2对比，选择输出leftvalue, rightvalue中的哪一个。

训练弱分类器的目的即是确定特征值的最优阈值t，使得这个弱分类器对所有训练样本的分类误差最低。

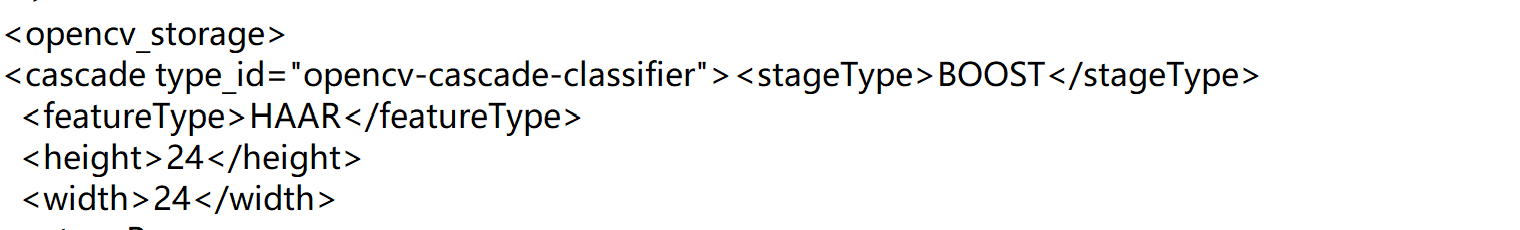


③强分类器结构及工作原理

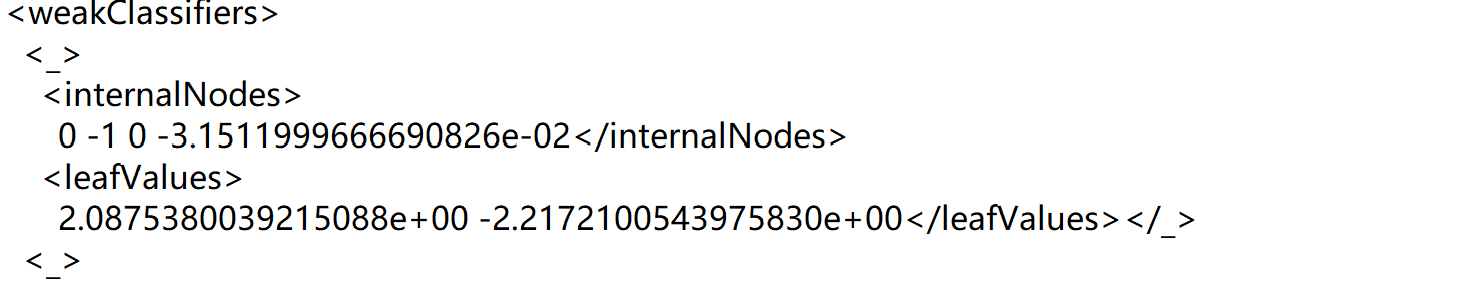
在OpenCV中，强分类器是由多个弱分类器“并列”构成，即强分类器中的弱分类器是两两相互独立的。在检测目标时，每个弱分类器独立运行并输出cascadeLeaves[leafOfs - idx]值，然后把当前强分类器中每一个弱分类器的输出值相加，之后与本级强分类器的stageThreshold阈值对比，当且仅当结果sum>stageThreshold时，认为当前检测窗口通过了该级强分类器。当前检测窗口通过所有强分类器时，才被认为是一个检测目标。

**（3）Xml文件参数部分解释**

①feature类型为HAAR, 检测窗口高度和宽度均为24



②弱分类器包含了<internalNodes>和<leafValues>两个标签。其中<leafValues>标签中的2个浮点数由左向右依次是rightValue、leftValue；而<internalNodes>中有整数和浮点数，其中浮点数是弱分类器阈值，整数用于标示所属本弱分类器Haar特征存储在<features>标签中的位置，以及用于控制弱分类器树的形状

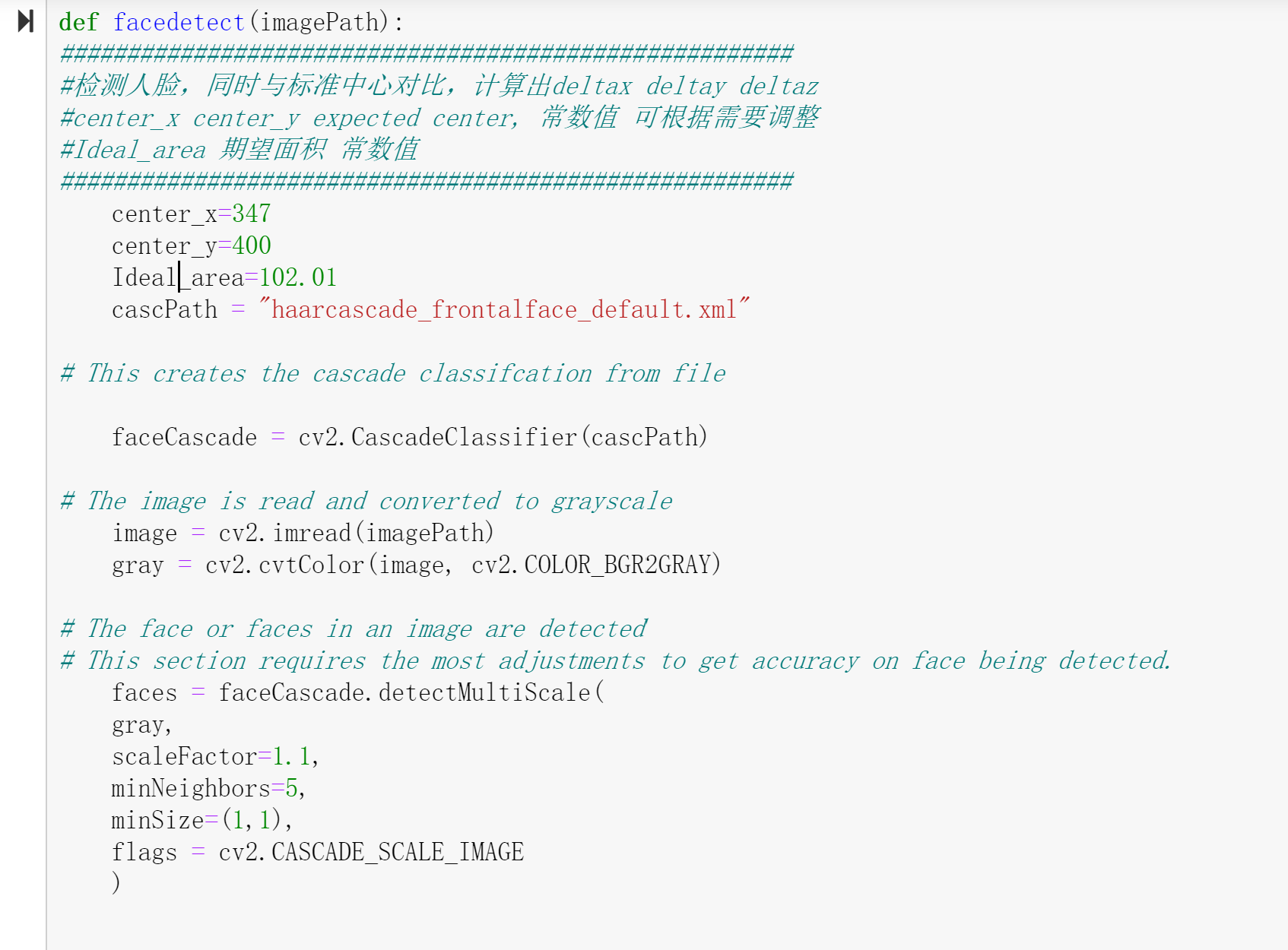


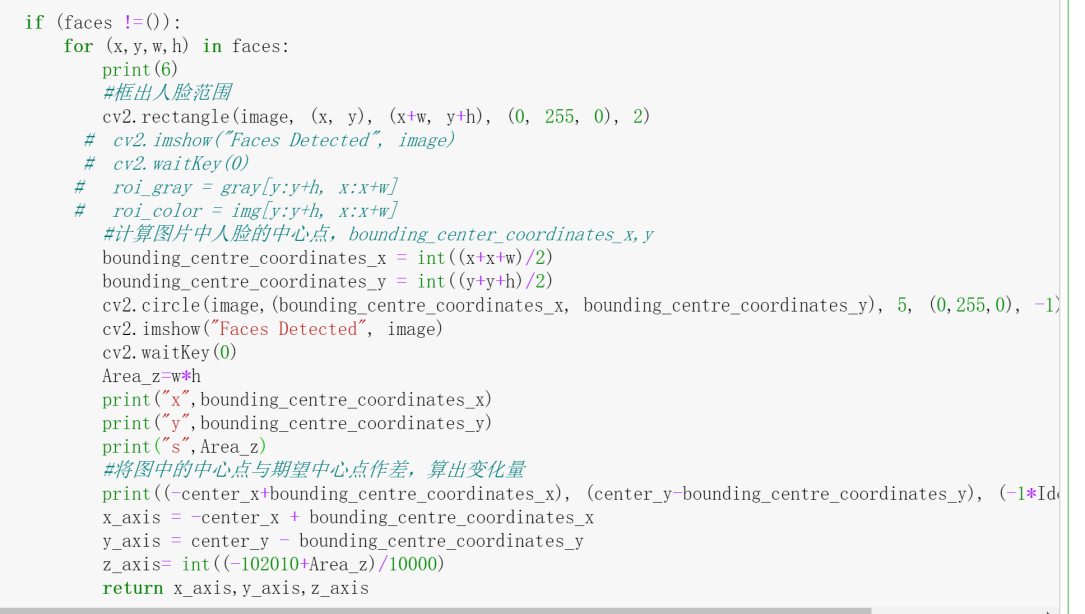
#### 2.人脸识别和控制飞行器移动代码及注释：

Facedetect说明文档

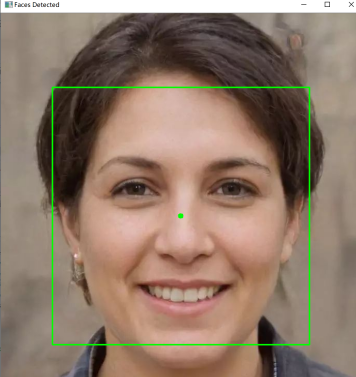
Facedetect函数：

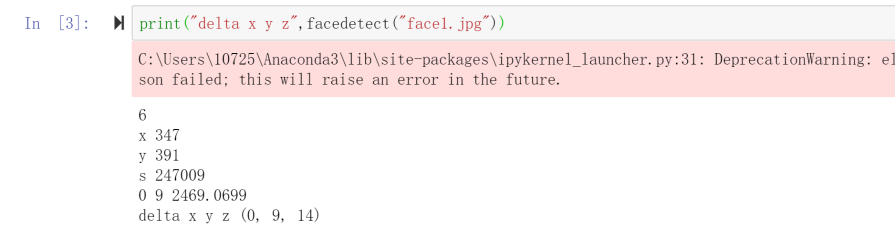
输入图片路径，输出识别图像以及与标准框相比需要移动的x,y,z轴变化量。





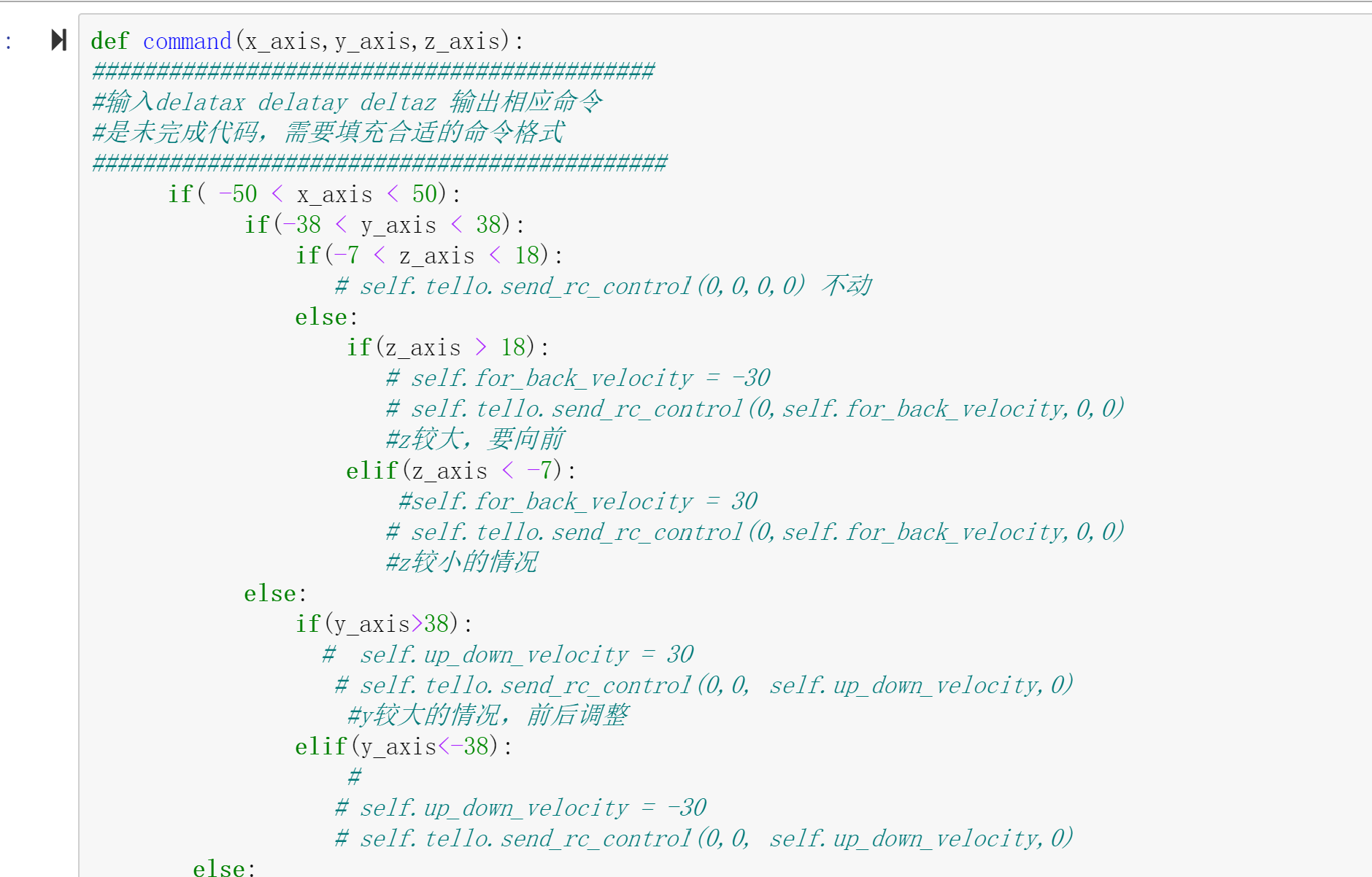
结果展示：





在控制飞行器方面，模型主要使用如下两个函数：

* Command函数：思路是x,y,z在某一范围内不变化，x控制左右移动，y控制上下移动，z控制前后移动。返回命令如“forward 20”
* Run函数：主要思路是保存某时刻截图在某一文件名下，facedetect识别，command函数输出命令，命令执行，间隔几秒，循环。





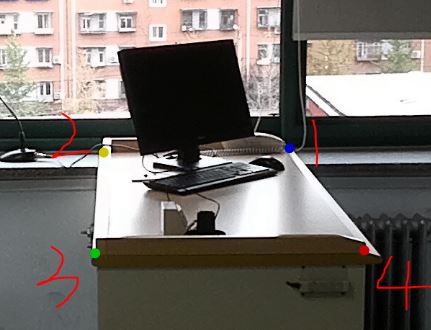
### 训练集构建

#### 1.拍摄

本研究运用大疆的tello无人机，选择了教二和教三的两间教室进行拍摄。首先将无人机按照说明书连接手机官方tello软件。在手机上操纵无人机飞行和图片拍摄，获得的图片直接保存在手机上。照片的主体为教二教三讲桌以及投影仪等障碍物，改变无人机和讲桌的位置关系，包括距离、角度和高度，得到两个数据集：tello和tello2，分别对应教三、教二教室。

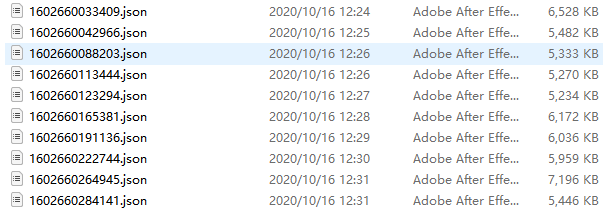
#### 2.定点

将手机上的图片导入电脑打包，并用数据标注软件labelme标注讲桌的四个点。这四个对应的数字如下图：



讲桌右上角标记为1，标号2、3、4分别对应讲桌左上角、左下角和右下角（图示方向）。并且这四个点在实际空间内位于同一平面，即讲桌的桌面，并没有在两侧的桌沿或者右侧向下倾斜的台面上。

标记后得到了一系列格式为json的数据集，文件内容包括标记点的像素位置信息。这部分将作为机器训练的主要数据来源。



桌角的位置和图像的信息已经保存在了json文件中，后续我们将用python提取桌角位置到Excel中，方便训练。

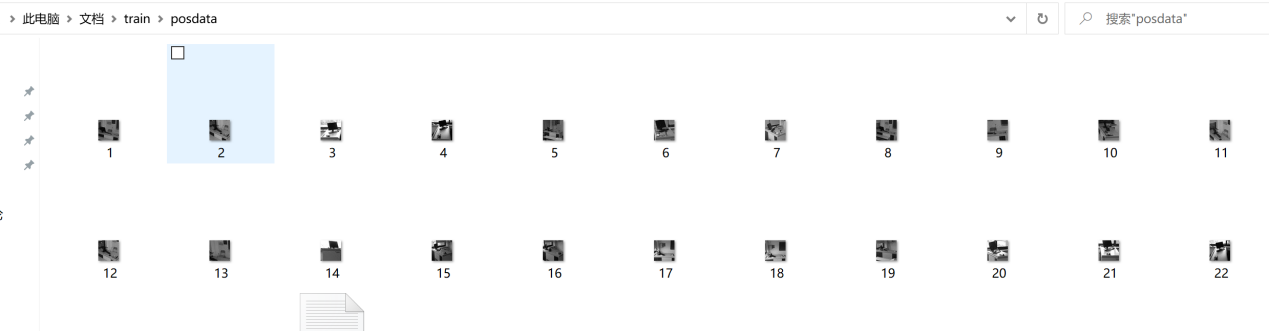
## 三、遇到的问题

### 关于讲桌的尝试

在实现了人脸检测后，尝试用联级分类器来识别讲桌。

**Step1:**

截取了与讲桌相关的图片，设定大小40\*40，灰度处理作为positive sample:



**Step2:** 准备negative data.

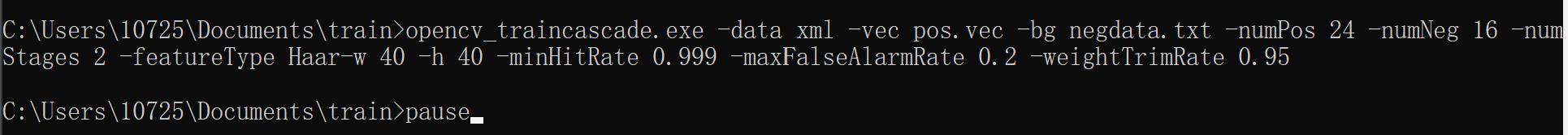
**Step3:**准备vec文档，用trainscade运行相关命令：

opencv\_traincascade.exe -data xml -vec pos.vec -bg negdata.txt -numPos 24 -numNeg 16 -numStages 10 -featureType LBP-w 40 -h 40 -minHitRate 0.999 -maxFalseAlarmRate 0.2 -weightTrimRate 0.95

pause



结果：可能由于所用数据集包含照片太少，训练器没有反应



1. **关于无人机照片**

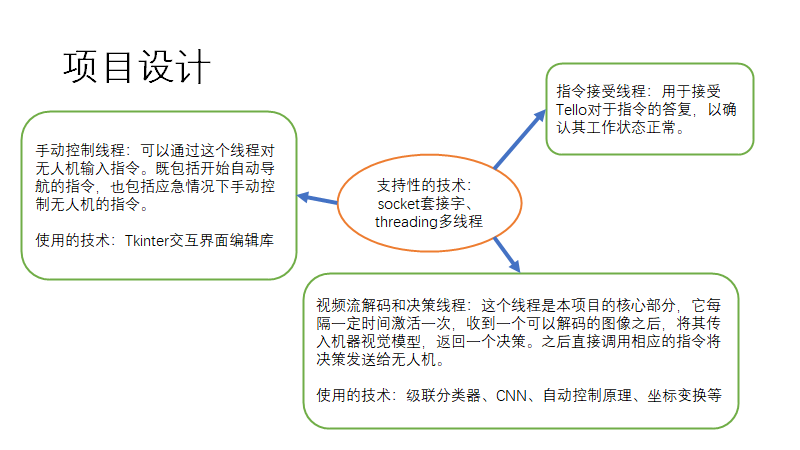


在接收和解码视频的时候，我们发现，无人机在悬停时拍摄的效果较好，在移动过程中拍摄的效果较差，有的时候甚至完全不能辨认。但是其实人也会在移动的过程中难以注意周边的情况，如果能从模糊的照片中获取有价值的信息，其实也是一种有价值的工作。

## 四、未来规划

目前，我们各方面的技术探索基本成功，下面主要的工作是将各种独立的技术整合起来，形成一个相互配合、协同运转的工程项目。目前我们设计了如下的代码框架：

本项目以“线程”为核心，主要分为手动控制线程、视频流解码和决策线程、指令答复接收线程三个部分。具体内容如下图所示：



对于人脸识别、讲桌降落两个任务，只需要修改其中的机器视觉模型和决策模块即可，不需要对整个项目的大框架进行修改。

下一步我们将分别编写这三个模块的程序，并进行整合测试，以达到实验目的。