Voice Activity Detector

1. Instead of using deepspeech which is the best offline speech to text library we use webrtcvad library which tells you whether a sound is present or not in the audio clip. The reason for using it –

* As we are only concerned about someone speaking or not there is no need to convert the speech to text.
* Deepspeech doesn’t provide models to translate hindi.

1. It takes the input audio of length 10, 20, 30 ms and then output 0 or 1 based on whether sound is there or not.
2. It has 3 mode of working. Mode 3 is hard mode while mode 1 is soft mode. Means in mode 3 it only detects sounds which are louder and in mode 1 it detects slightest of sound and sometimes detect the noise as sound.
3. So we have made a queue of 300 ms and if in that 300 ms queue 240 ms are of sound if say that there is a sound and if in 300 ms, 180 ms are of silence we say that the sound has stopped.

Background Matching

1. In this we only match the background of the present student whether it is changing or not.
2. So for that we first detect the faces of the person as they can be moving from frame to frame.
3. Then we make all the pixels which are coming in the face bounding box in either of the image.
4. Then we compare both the images and set a threshold to declare them as same or changed.
5. We looked for different technique like SSIM, Euclidean distance, Mahalanobis distance, feature detectors, CNN based approach.
6. But looking at complexity of our use case we selected SSIM as it is fast and less dependent on the lighting conditions.

Head Pose Estimation

1. In this we will predict the yaw, pitch and roll of the head of the person. These data will be useful in predicting whether the candidate is looking away for cheating.
2. The second and important use case is when we register the candidate. We will take 5 photos from different angles. The yaw, pitch and roll will help to determine the current pose of the candidate.
3. For this there were two methods –

* First one is use opencv and detect the facial landmarks and then by using PnP algorithm determine the yaw, pitch and roll.
* Second was to use pretrained deep learning model.
* We selected second approach as the first approach also depends upon the camera while the second approach is not dependent on camera.

Object Detection

1. For this we will fine tune the pretrained object detection models on our required classes.
2. We will first download the images related to our required classes.
3. Then we will annotate the images.
4. Then we will retrain the model.
5. Deploy the model.