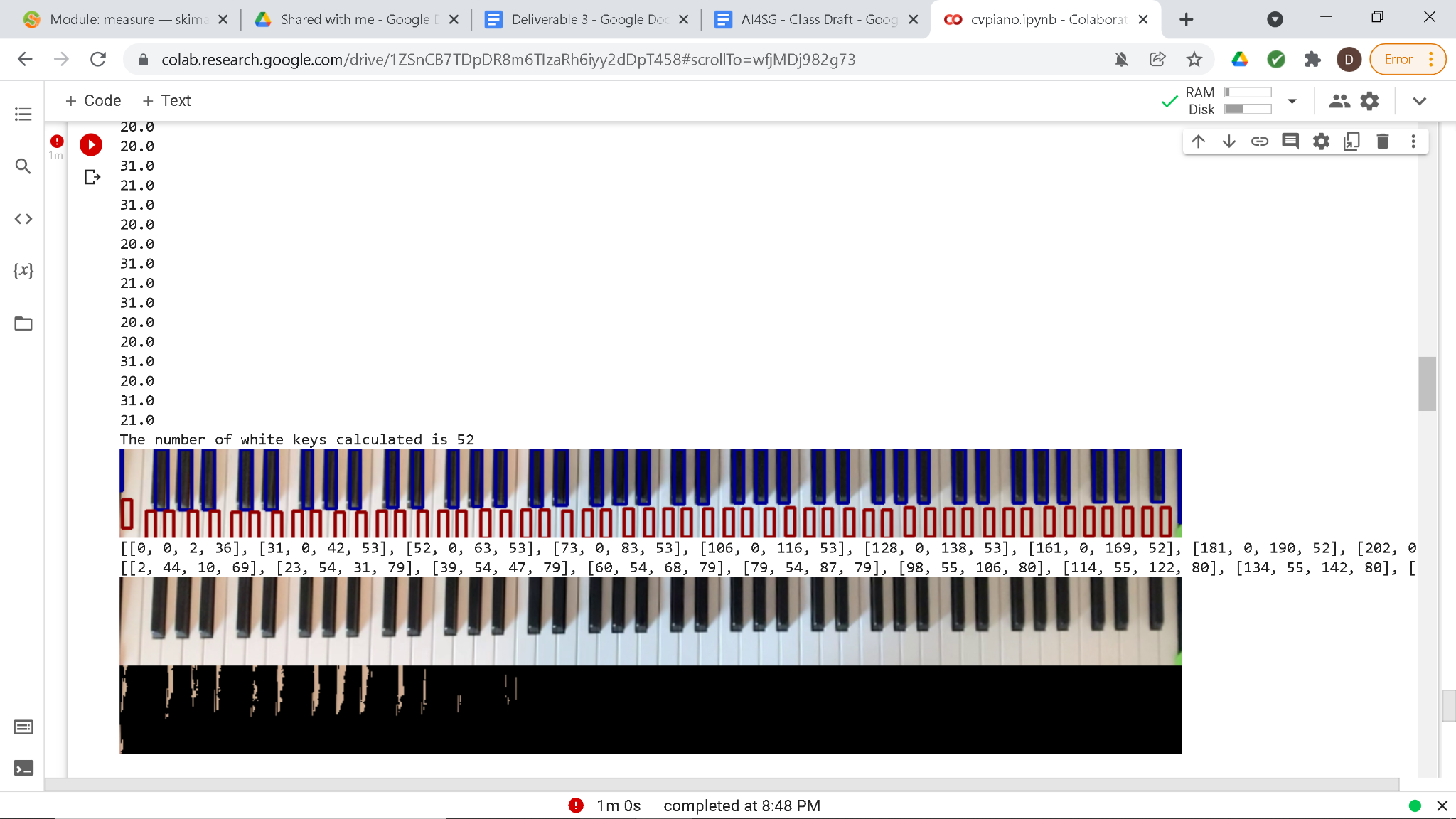
Computer Vision

Semester Project Part 3 Deliverable

Daniel McCarthy and Chris Hunt

Preliminary classification results on validation data and ideas for improvements

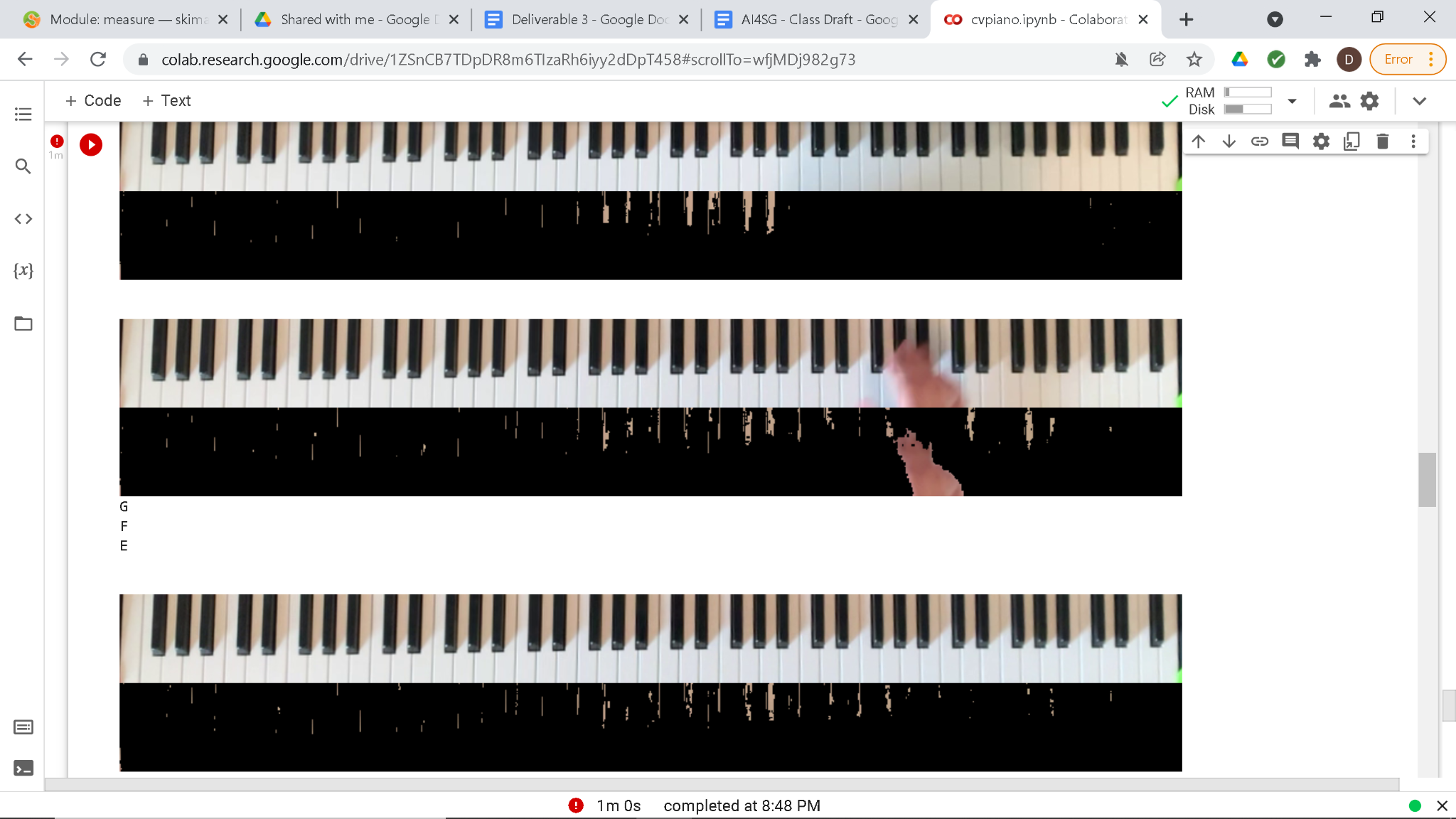
1. In this stage, there are two types of objects that require classification. One object is the actual keys and the other object is the hands. 88-key pianos will always maintain the same order and distances for the keys and the original segmentation will leave the positions in the same place. Black keys and white keys were done separately for further feature extraction. Once we could classify all the keys based on a horizontal pattern, we then used 12 classes of notes to cycle through and classify correctly, from left to right. A formal classifier would be too difficult in this scenario because many keys look exactly the same even though they are classified differently in terms of music notes. It would be too hard to do visually for a formal classifier, but fortunately because of the aforementioned piano patterns, we were still able to use a position-based algorithm for classification. The image below shows the final bounding boxes for each key.

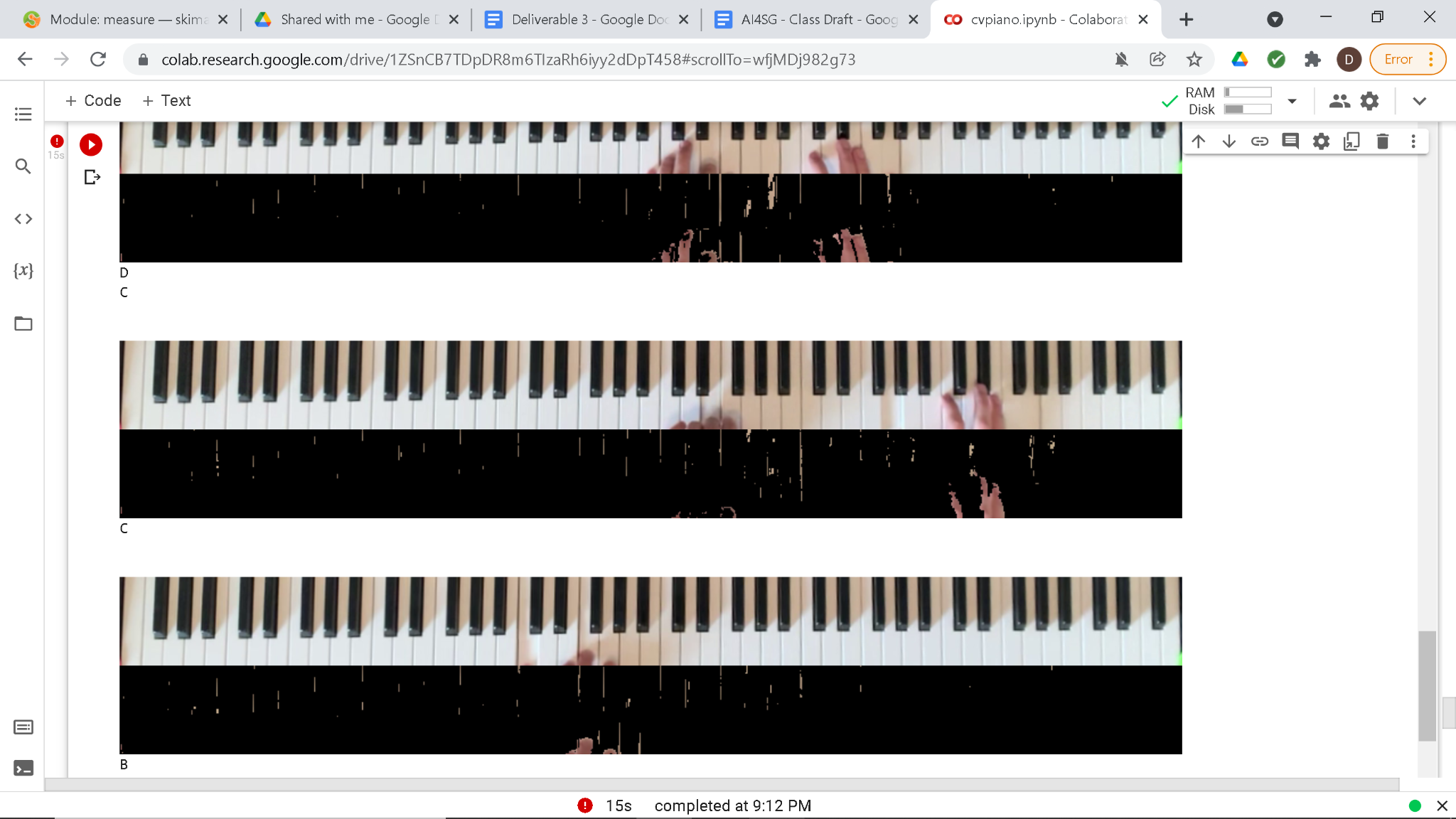


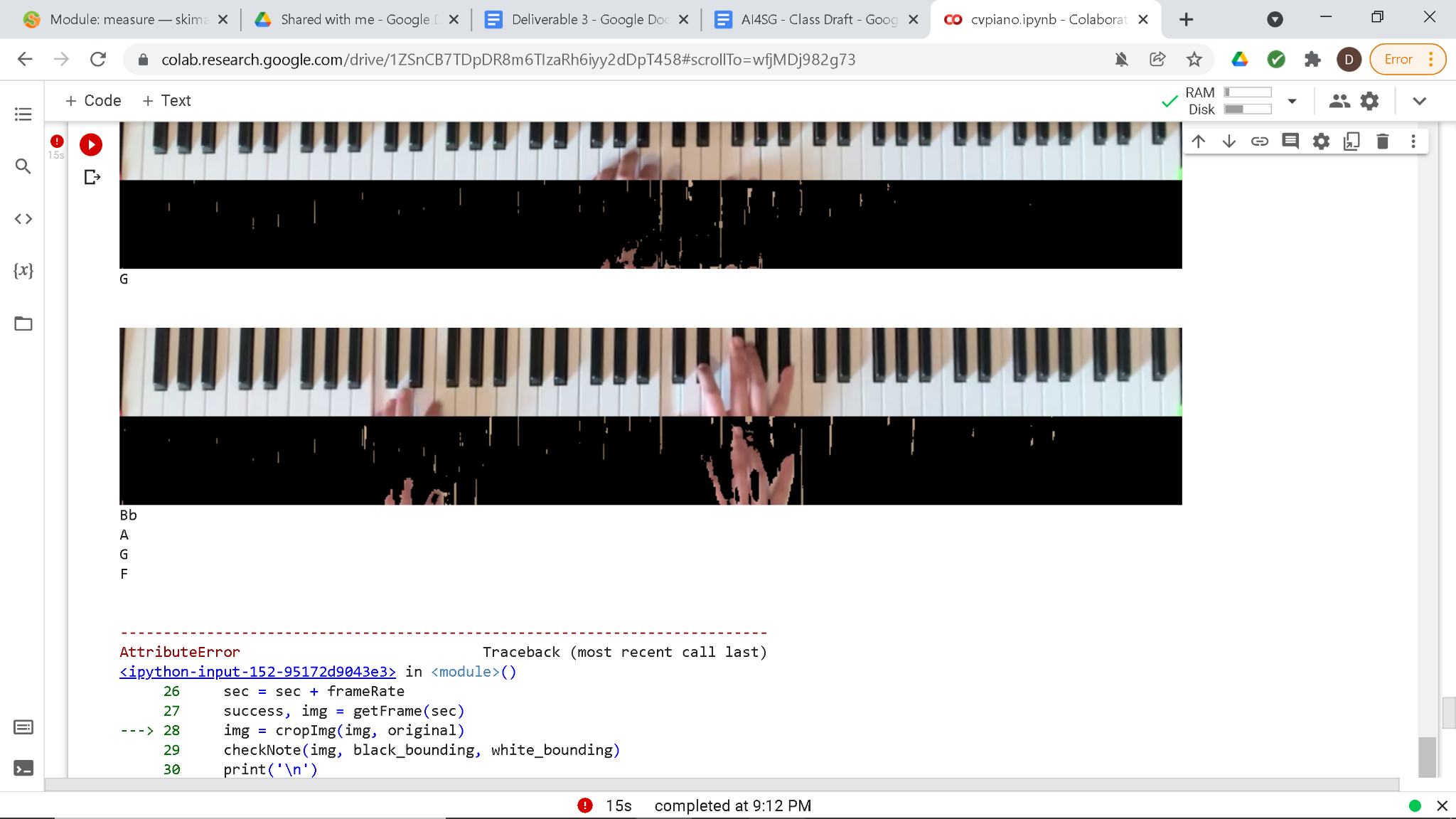
Note that the white keys are only detected in a small area below the black keys, this is a room for improvement in the future. Classification of the hands made use of a simple skin detection algorithm that could create a mask to be used against the piano image. It should be noted that this skin detection was tuned to our piano player’s skin color and would need to be changed for a different colored piano player.

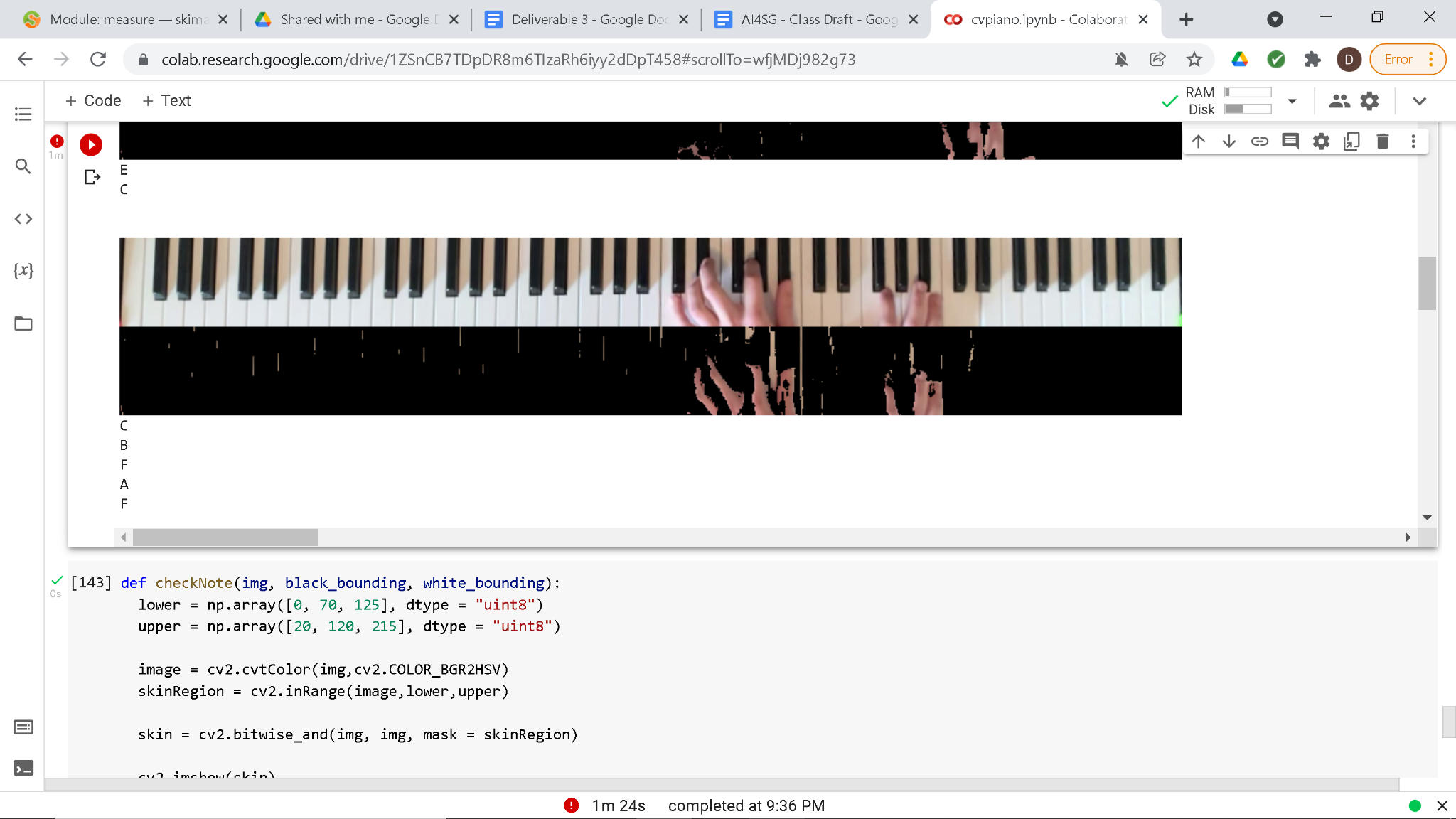
1. For object classification, A video feed of Chris playing the piano was sampled at a rate of 1 frame per second. Each frame was checked for skin detection and then checked to see if skin was located on the center of a key’s bounding box. We understand that this is a rudimentary way for detecting the press of a key and are actively looking to improve our method. Below are images displaying the accuracy/inaccuracy of our method.

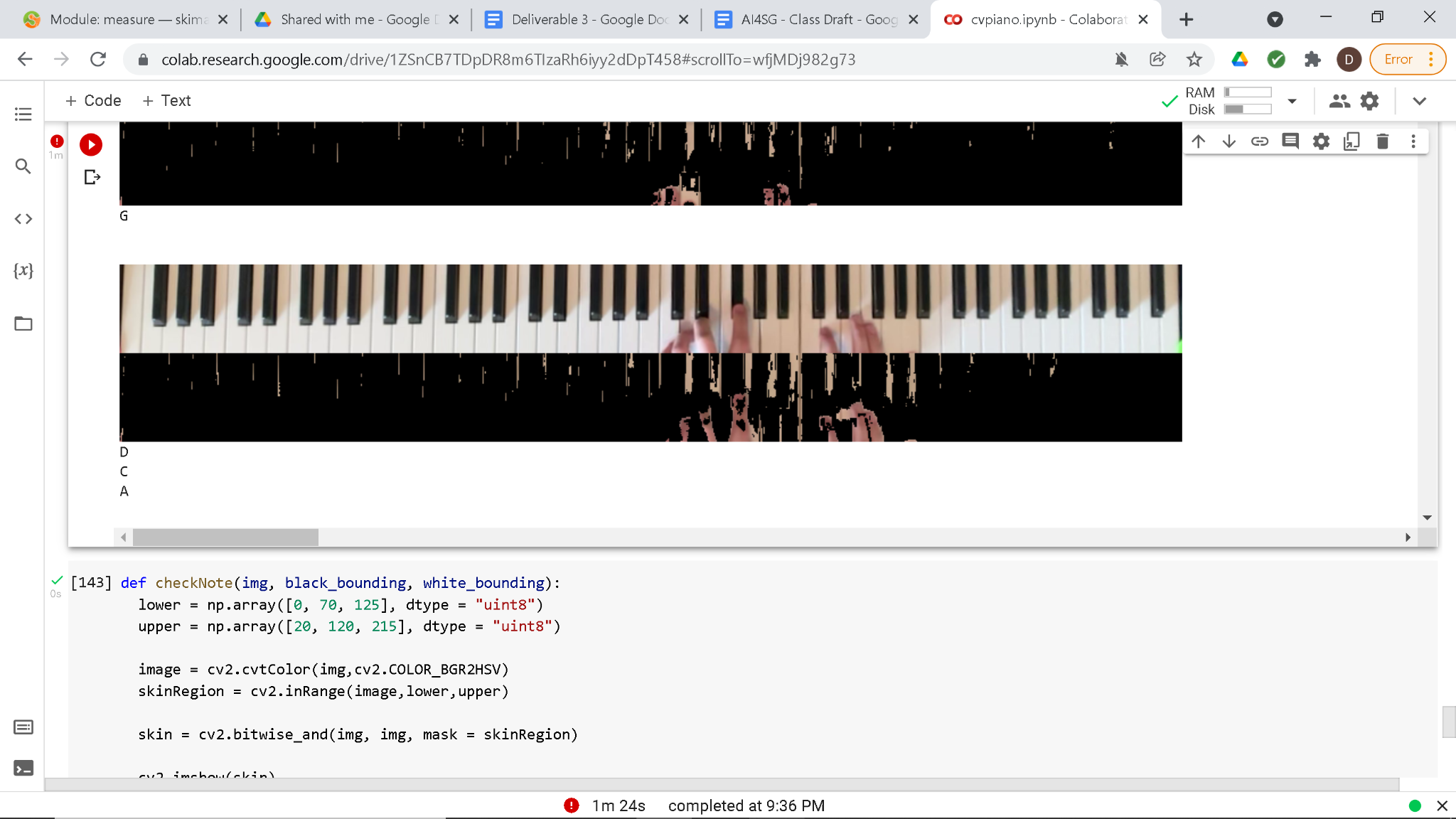
*Note: Our final piano image is flipped and therefore the far right piano key is the key A and subsequent keys proceed alphabetically from right to left*



The above image demonstrates the greatest weakness of our classification. While setting up the keyboard, Chris reached over to turn on the power. This was caught in one frame where Chris’s hand is detected as playing the E, F, and G keys. While in fact he is not playing these keys, his hand is covering the center of their bounding boxes and therefore is detected as being played. 

In this image, we see another limitation of our program. When the player's hands are cast in shadow or otherwise do not look the same color that they normally do in regular lighting conditions, they may not be detected by our skin detection algorithm. This issue is occurring on the right side of the image with the hands not being recognized in the skin detection that follows it. Also, in an effort to eliminate the noise of regular white keys cast in shadow being detected as skin, we made our skin detection very specific to the point where it sometimes doesn’t fully recognize skin that is in good lighting which causes fingers that are covering notes to not be detected as playing that note. You can see on the right side of the image that while multiple notes are semi-covered, only the C note is covered in the center of its bounding box and printed out as detected. A third and final issue presented by this image is that we can clearly see there is a note being pressed between the two hands, likely by a thumb, but that thumb is “off screen” as the manipulation of the original image cuts off the very lower part of the keys. We may incorporate actual sheet music as a means of measuring accuracy quantitatively rather than qualitatively for our final deliverable. For now a series of more images displaying more of the former mentioned pros/cons of our method are provided below.





1. If we had sheet music and played the song at a rate that we can easily sample. We could have a measurable way of knowing which notes should be played at what time and be able to compare these notes to our own to determine a statistic for accuracy. For accuracy improvement, we would like to find a way to detect only fingertips rather than entire hands and also adjust the sizes of our bounding boxes to be more inclusive of the entire areas of keys. Another note on bounding boxes is that we would like to change our method of detecting whether there is skin in the center of the bounding box to detecting whether there is skin anywhere in the bounding box. Or, we may decide to pursue a different method of press detection that does not involve skin such as whether the color of the key itself changes from frame to frame as this could indicate a change of light reflection from being played. Finally, a last ditch effort to improve accuracy would be to change how we are physically playing the piano to be more suited towards our detection algorithm. This would entail limiting the amount that our hands cover notes that are not being played and ensuring that our keypresses are located in the center of where our key’s bounding box would be.
2. <https://github.com/chunt4/cv-piano>