1. **Image Dataset Preparation**

It was necessary to prepare a large enough dataset of images of the individual capuchin monkeys under study, labelled with rectangular bounding boxes around the individual capuchins’ faces and their identifying names in order to train an appropriate object detection model that can then detect these individual capuchin faces in real-time and also correctly identify the detected capuchins by their names. I was provided several recorded videos of the capuchin monkeys along with timestamped intervals logged on BORIS as well as the names of the individual capuchins present during these intervals in the video. The naïve but time-inefficient way to extract images of capuchins from the videos is to simply use BORIS to navigate through and take a screenshot of every frame of the video within every labelled interval and save the image in the corresponding folder with the capuchin’s name.

Table

Description automatically generated with medium confidenceHowever, I wanted to automate this process for every observation by implementing a Python program with the help of OpenCV, a Python library of programming functions aimed at computer vision and video / image processing tasks. BORIS has a feature that gives the user the option to convert the labelled intervals format to a binary table (every entry in the table is either 0 or 1) format for every capuchin under study for every video observation where rows represent successive frame timestamps throughout the observation, separated at a fixed, user-defined interval, the 5 columns represent the 5 possible capuchin behaviors – Apparato, Banana, Mecate, Observar, Platforma – and every entry in the table is either 1 or 0 indicating whether the specific capuchin was present in the video performing the specific behavior (column) at the frame at the corresponding timestamp (row). These tables would make the process of extracting images of capuchins from the videos with OpenCV easier and more straightforward to implement. I did not want to extract candidate images from every frame in all the videos, because this could result in creating an extremely large image dataset, much larger than necessary, and with a lot of very similar images. This would also mean a lot more images to be labelled, and with a lot of very similar images in the dataset, this introduces bias into the dataset, giving the object detection model more opportunities to learn patterns specific to the similar images, making it more prone to overfitting i.e. having a low ability to generalize to new images outside of the images that the model is trained on. Therefore, I generated the tables for frames at 0.3 second intervals for all observations, to be able to extract a maximum of 3.33 candidate frame images of capuchins per second. For example, for the observation S17 P3, for the capuchin “Pio”, shown alongside is an extract of the binary table generated by BORIS for frames at 0.3 second intervals.