OELP MID SEM PROJECT REPORT

VIDEO ANALYSIS

by:
Jasir K 111901025
Deon Saji 111901022

Faculty Mentor:
Dr. Vivek Chaturvedi

INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

Computer Science and Engineering

Indian Institute of Technology,Palakkad

1.Introduction

Computer vision is one of the areas that is advancing rapidly thanks to deep learning. Today the machines are able to easily solve the tasks that were once considered very difficult. Image recognition, object detection and localisation are three major computer vision tasks and this helps machines to look at each image and make accurate conclusions/ recommendations about them. It uses convolutional neural networks to process visual data at the pixel level and deep learning recurrent neural networks to understand how one pixel relates to another. Because of the ability to figure out the objects around and to perform the same tasks as humans, today self-driving cars are made possible.

Video content analysis is one new technology which uses computer vision and in general different deep learning models.It is used for automatically identifying different events in the videos and also monitor video streams in real time. This is mainly used in the field of sports,transport,security etc.

1.1. Aim and Motivation of the Project

In this project, we will devise a fast learning algorithm which will be able to learn from analyzing video frames to recognize the player's behavior and techniques and then use that information to give suggestions to improve his/her skills and give suggestions for the areas that the player needs to work upon. These results will aid athletes in identifying and correcting their frequent errors as well as modifying their technique if necessary to achieve better results. To begin with, we looked at the existing work in this field in order to determine how we could improve it and what other factors may influence the game through some research paper.

2. Research Paper

To become an export in any sport, one must have good concentration and should be able to make accurate and quick decisions in less time. Because the reaction time is very less it is important to have the knack to make the best possible moves/decisions in a fraction of a second. We read different papers on what factors affect and how to improve alertness, concentration and hand-eye coordination can be improved and visual reaction. In the tests increasing the point of the person seemed to help to improve his attention level even being an repetitive test. In competition by staying on the winning side increases the motivational and other factors thus improving the attention and target. Imagery practices reduce tension and anxiety and thus deal with tactical problems in competitions.

We went through some coach manuals to understand what the rules are and different kinds of shots and its execution and different performance factors of technique, tactics, physical, psychological and lifestyle.

3. Learning process overview

At first we needed to go through some tutorials to understand computer vision and video analysis using python. We started with understanding hand tracking and pose estimation, and face detection, which was used to complete the mini projects. For hands-on experience we made some mini projects for volume gesture control, finger counter, AI trainer and AI mouse control. Working on these mini projects helped me to understand the concepts of the working of pose detections and mediapipe using python.

The frameworks used in the mini projects were mediapipe. In which there was a pretrained model which detects faces at around 50 fps. Which would detect 6 landmarks in the face along with it we can track the pose also. There were many other interesting modules in the mediapipe framework for hair segmentation, box tracking and motion detection.

For hand tracking and pose detection we used opencv along with mediapipe. Which has a pretrained model for locating 21 different landmarks. Thus using these models we were able to count the fingers in real time using python. The module for pose detection would be further used in the project.

Various *Algorithms* were analyzed in order to find the best suited algorithm to continue within the project. For our project we needed to label multiple objects and find the bounding box of that object in the frame. If we use simple CNN then it would be difficult to identify different objects. RCNN and YOLO (You only look once) object detection method was apt for the project. There were different versions of YOLOv5 like YOLOv5s,YOLOv5m,YOLOv5x and YOLOv5l. As our dataset size was small and we needed real time analysis so we selected the smallest model YOLOv5s. Greatest advantage of this model is the speed and it allows us to train custom data from its pre-trained model using transfer learning. The accuracy of these algorithms is pretty much accurate.
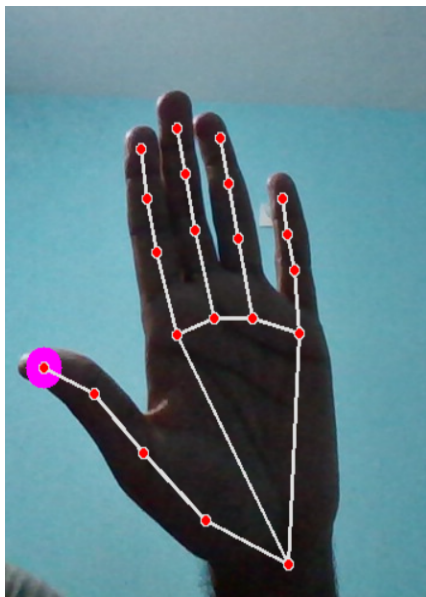

Hand Tracking
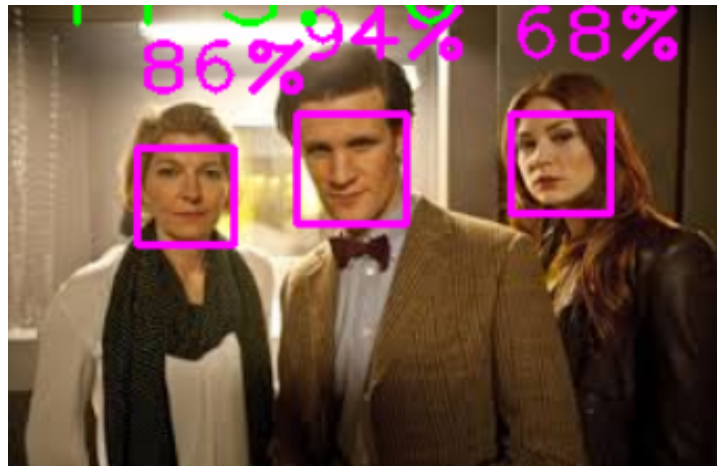
Face detection

3.1 Implementation of Object detection

One of the most important and basic steps in analyzing a video is to detect the objects in each frame of a video. In the field of sports, the main objects to be detected include the players, different sports equipment involved in the game, different types of shots played etc. Therefore we had to train different datasets containing each of the above information so that we could come up with a model to test it on any video. The model is built on the YOLOv5s object detection algorithm. It is a family of object detection architectures and models that were pretrained on large datasets. By using the concept of transfer learning and changing the last layer of YOLOv5s, the open source code of this was used for training the given datasets.

The training dataset consisted of different video frames that already classified each object to different classes and coordinates of their corresponding bounding boxes. YOLOv5 models are trained on labeled data in order to learn classes of objects and the dataset provided by the company already had correct labels and annotations in the required format. There is also an online software named Roboflow which could be used to label and prepare the data suitable for YOLO.

The model was training for 250 epochs and batch size of 32. Colab GPU was used for faster training. Below are the few results we got while detecting objects in the image (player,referee,equipment) for train and validation sets
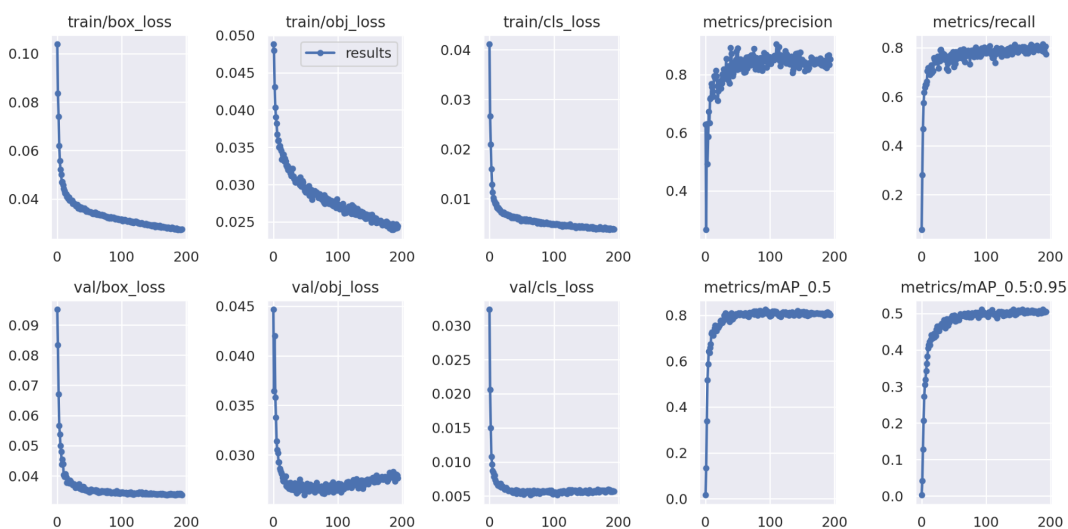
3.2 Analysis of count of shots played by each player

Counting the different shots played by each player at each time frame is an important thing to be found out because it is important to play the right shot at the right time. Depending on the move of the opponent, the best possible move/action must be done by the player.

Since the bounding boxes are already found during the training of the dataset which gives some information about where the player is standing with respect to some frame of reference. For example, in case of table tennis or badminton, by using the coordinates of the bounding boxes of the net and that of the player we could find all the shots/actions made by either player. What we thought of is to take the right bottom y coordinate of the net and compare it with the right bottom y coordinate of each player to find which player is standing on which side of the net. After doing so, we could separately find the shots played by Player 1 and Player 2 during the game.

4. Future Works

As mentioned above, the main objective is to bring out some good and efficient algorithms to improve the game. We have so far used only object detection algorithms and the next step is to use pos detection or similar algorithms to detect backhand forehand shots and to detect the foot movement of the player. This would help in understanding whether the player moves forward or backward and to know whether he/she gets into the right position when it is time for return and the different techniques used by the player. And we would be combining the different trained models and make conclusive analysis from it.

5. Bibliography

[1]
https://www.yumpu.com/en/document/read/53240814/13-a-comparative-study-of-visual-reaction-time-in-badminton-players-

[2]
https://www.researchgate.net/publication/266868606_A_Research_on_Visual_Analysis_of_Badminton_for_Skill_Learning

[3]
https://www.researchgate.net/publication/320168613_A_study_of_attention_and_imagery_capacities_in_badminton_players

[4] Mediapipe Documentation, hand tracking and other modules.
https://mediapipe.dev/

[5] Mini Project CV
https://www.youtube.com/watch?v=01sAkU_NvOY

[6] Plotting results,
https://wandb.ai/site

[7] RoboFlow for Annotations and labeling,
https://roboflow.com/

[8] Yolov5 - Custom Data
https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data