Automation of Data Collection from video source using Machine Learning/Deep Learning

Jash Rathod
Denison Mathematics & Computer Science (Data Analytics) Department
Dr. Matthew Neal

SCOUTING
Introduction

I currently work as the data analyst for the varsity women's field hockey team and the varsity baseball team. At work, one major issue I have been facing is that I am not getting quality data. Many questions of interest, such as "Why can't the Denison field hockey score goals more often?", cannot be studied because the only data I have is the number of shots taken, number of assists, number of goals, number of shots on target. Therefore, the question I wanted to ask was, "How can the current data collection process be made efficient, accurate, and more detailed?"

The answer to the question is to use machine learning for object detection.

THE PRACTICE SESSION
Methodology

The methodology is divided into two parts :-

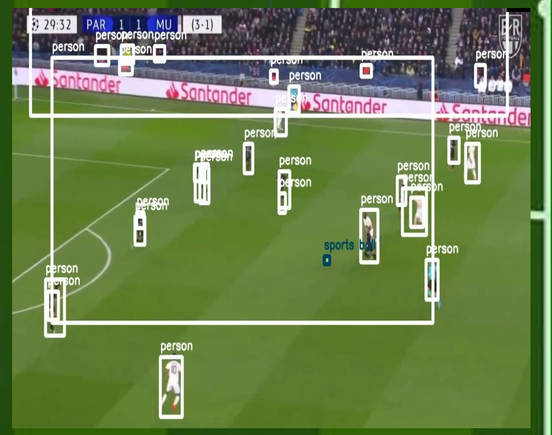
- Getting Quality Video
  To get quality video I decided to use a DJI Tello drone which can be bought for around $99.
  The reason I decided to choose the drone above given its compatibility to run python scripts directly from laptop. Moreover, a drone can get close range videos much better than a camera in the press box.
- Data Collection
  All the coding to get data out of the video was done in python. Some of the libraries used were DJITelloPy for controlling the drone using the laptop and getting the video. After which the video which was getting streamed and converted to a series of matrices frame by frame using the opencv (cv2) library simultaneously. And, once in matrix form I used tensorflow.gpu for object detection api to use a pre trained object detection model. The pretrained object detection model I used is called YOLO v3, which stands for You Only Look Once and COCO dataset to detect common objects like 'person', 'sports ball'. And the final step was getting the coordinates of the bounding boxes at particular time using the library time and using the library pandas just getting the csv file of all the players in the game over a time period before using tableau to plot it.

THE GAME PLAN
Visual Explanation of Methodology

Getting Quality Video

The picture below is the tello drone used for the getting video and it is small enough to fit in ones palm.

The picture below is from the testing phases of the project where I used video from youtube. It's the second leg match between Manchester United and PSG from champions league

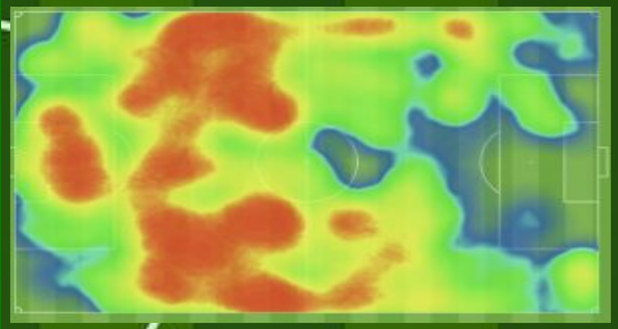


Data Collection

After receiving the video, this is the object detection that was done looked like the picture above where we can see players being detected as person and the football as the sports ball. However, some obstacle faced were that some people from the crowd being detected as person and not being able to differentiate between the two teams. The first obstacle was overcome by increasing the prediction percentage threshold while the second obstacle was overcome by hardcoding filters to detect particular colors only.

GAME DAY
Results

This project was made keeping coaches in mind and as the title suggests, it is a process of live data collection and it would not make sense handing the coach a bunch of csv files during half time or quarter, therefore using tableau and the csv file that we get from applying machine learning on videos received by the drone we could generate heatmaps that would look as the following if generated.



FULL TEAM

INDIVIDUAL PLAYER

Other in game stats that can be achieved using the data would be how much distance players are covering or the average pace of the players, which can be used by coaches to evaluate the fatigue level of players.
Other live stats possible with the data collected would be the possession percentage in sports like soccer, basketball or rugby. For sports like tennis the player movement on the court can also be achieved using the data algorithm.

POST MATCH CONFERENCE
Future

As we can see that all of this was made possible using just a $99 drone. With better drone, some stats like which foot was used to take the shot, or whether the player posture while shooting the ball or maybe what is happening in every pitch in baseball game. The possibilities are limitless and automated .
Other feature I tried working on after the summer research was over was how can I make the drone fly itself by processing the view of the field and its position but was unsuccessful given I am still a beginner in the field of machine learning.

THE COACHES
Acknowledgments

Paula J. Soteriades
(Head Field Hockey Coach - Denison University)
Giving me my first break as a data analyst in the field of sports and being there for me. Be it discussing project ideas or any obstacle I might have (even life).

Dr. Matthew Neal
George R. Distinguished Professor of Mathematics
Trusting in me and giving me the opportunity with an idea that definitely sounded not so feasible when I first discussed it. Moreover, having faith in and being there to bounce off ideasthroughout the summer.

Funding
This project was funded by The William G. Bowen and Mary Ellen Bowen Research Scholars Endowment at Denison University.

PRESS RELEASE
Work Cited

- Redmon, et al. "YOLOv3: An Incremental Improvement." *ArXiv.org*, 8 Apr. 2018, arxiv.org/abs/1804.02767.
- Poon, Jason. "Chalkboard: No Width? No Problem!" *Big D Soccer*, Big D Soccer, 9 June 2015, www.bigdsoccer.com/2015/6/9/8752933/scratching-the-chalkboard-uswnt-austrailia-2015-world-cup.
- "USL League One, Week Two Recap: A League Meant For TV Drama - BGN - Beautiful Game Network." *BGN*, 9 Apr. 2019, www.bgn.fm/usl-league-one-week-two-recap-a-league-meant-for-tv-drama/.
- "Ryze Tech Tello Quadcopter." *CP.PT.00000252.01 B&H Photo Video*, www.bhphotovideo.com/c/product/1383649-REG/ryze_tech_cp_pt_00000252_01_tello_quadcopter.html.