New York Yankees Summer Associate Report

Quantitative Analysis: Computer Vison

Duration of Employment: 5/16/2024 - 8/30/2024

Author: Ruben Alias

Manager: David Grabiner

Mentor: Jorge Ortiz PhD.

Why Use Computer Vision in Sports?

During my time with the New York Yankees, I was exposed to an interesting and unique application of machine learning: sports. Computer vision primarily is utilized for understanding and analyzing player performance, strategy, and coordination.

Given that professional sports teams hold an abundance of video footage, radar data, as well results-based player stats from their games, computer vision and other deep learning techniques find plenty of use in developing sophisticated analysis tools for player performance. We also find that player performance is not solely understood from results-based stats such as runs or strikeout rate, but also from more meaningful features isolating the player's own athleticism and talent. Extracting these features, or a close first-order approximation of them, particularly in the absence of radar data is a key role of computer vision in performance science. State-of-the-art models such as YOLOv8 are employed for not just identifying players but also mapping out their bodies and understanding how they pose themselves, how quickly they may react to a certain stimulus, and so on.

An additional application of computer vision in player performance analysis is assisting in filtering and processing of large datasets. When working at the professional level of a sport, the amount of video feed you are exposed to increases exponentially. Further, it is not guaranteed that the type of video you receive is always of the same quality or content, and manually filtering this kind of video would not be as beneficial as automating it via existing tools.

Project 1: Video Feed Classifier

- **Context**: When exchanging video feed form other stadiums, it is often the case that the labels for various angles of video feed are incorrect or mislabeled as different angles. Being able to sift through large quantities of video for a particular angle is very useful in developing more advanced computer vision tools.
- **Purpose**: To take large quantities of unlabeled video feed and automate labeling.
- **Methodology**: The method was to use existing tools, namely Ultralytics' YOLOv8 Classification Model. It was trained on manually labeled frames extracted from video feed at various levels of play. To smoothen out the effects of per-frame error arising from miscellaneous movements in certain videos, we decided on taking a majority vote on a fixed number of frames sampled at a regular interval throughout the full video.
- **Challenges**: We found that the original model, although understanding larger concepts like what constitutes a Pitcher Angle versus a Batter Angle, struggled with picking apart whether a pitcher or batter was facing the camera.
- **Solutions:** The solution we came up with was to train extra classifiers designed for those subclasses of batter and pitcher angles to determine whether a player's body was facing the camera throughout the movement.
- **Results**: We achieved around a 93% accuracy per video, at a rate of approximately 1 second per video.
- **Things I would do better**: I would consider potentially moving towards extracting features such as the existence of home plate, the body, and the types of players in the video feed, and then train a classifier on those features and their relative positions in space. My reason for doing so is that this would have 'killed multiple birds with one stone' per se, allowing for those smaller models to be used in other tasks for detecting and tracking certain facets of video.

Project 2: Bat/Home Plate key-point detector

- **Context:** When dealing with large quantities of video in baseball, we may want to extract features from the footage based on the location of various facets such as the bat, particularly the bat head, midpoint, and knob.
- **Purpose**: Design a feature extractor to retrieve the key-points of the bat as well as the vertices of home plate.
- **Methodology:** The first step in our approach was to first train Ultralytics YOLOv8 key point model to detect the bat and home plate in each frame of video, as well as the coordinates of various parts of the bat in pixels (head, barrel, knob). This required taking the previously constructed angle classifier, run it over a large quantity of video from minor leagues as well as high speed footage, and select the videos where the angle was Hitter Open (Batter angle with the body facing the camera).
- **Challenges:** One challenge faced was in searching for useful training data. Given on standard footage the bat may move too fast to be visible because of a high exposure time, it may cause issues for the model during detection.
- **Solution:** We worked with a mix of high-speed data as well as low-exposure video data retrieved from the database.
- **Results:** We achieved a mAP50 of .84 and .96 on detections for Bats and Home Plate respectively.
- **Things I would do better:** I would consider including more data on additional classes potentially like the batter box key points(for zone depth) or the location of the ball(for retrieving impact frames)

Conclusion

My time with Quantitative Analysis at the New York Yankees was very educational. I learned about a practical application for CV, as well as learning about the application of statistics in understanding player performance. Seeing the intersection between theory and practice was an enlightening experience, as you can see the scientific approach taken in sports analytics. Additionally, I was given the chance to learn how to train and test models using YOLOv8, develop CV datasets with Roboflow, and utilize CV techniques such as Homography.

Diagrams/Graphics

Figure 1: Architecture of video feed classifier**.**

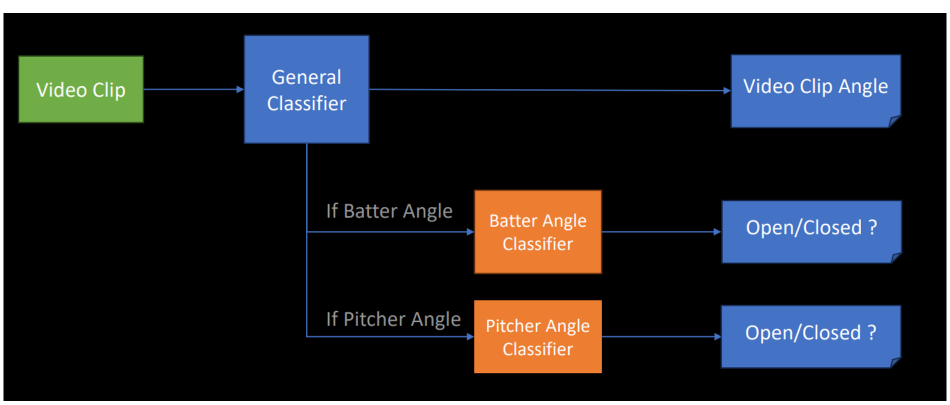


Figure 2: The kinds of frames the video feed classifier model would classify.

Figure 3: Classifier methodology

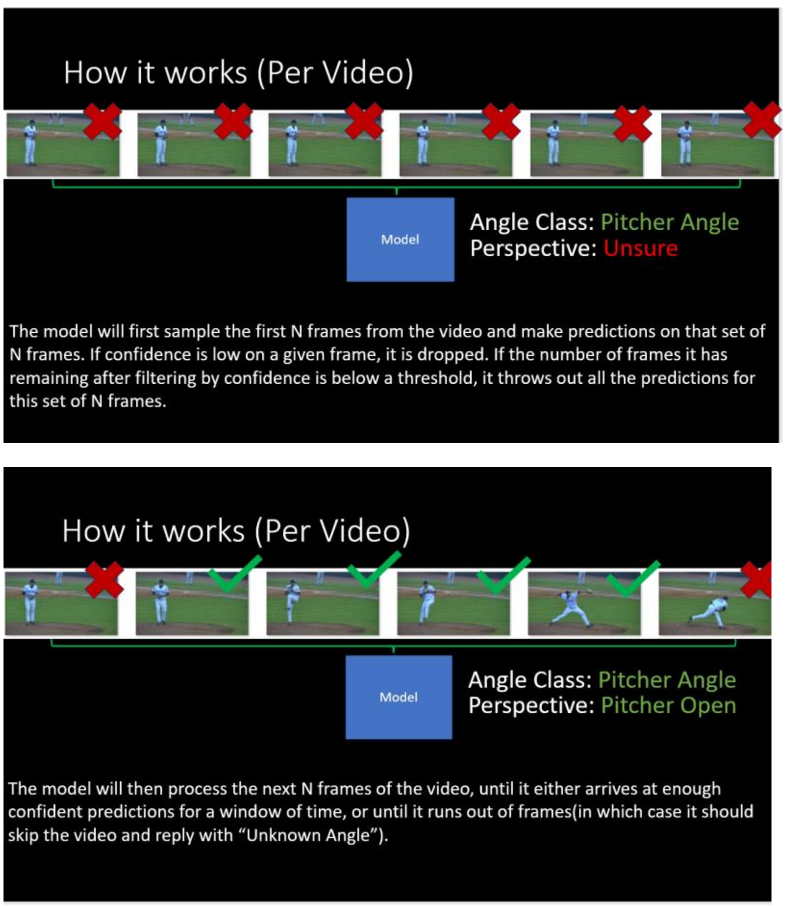
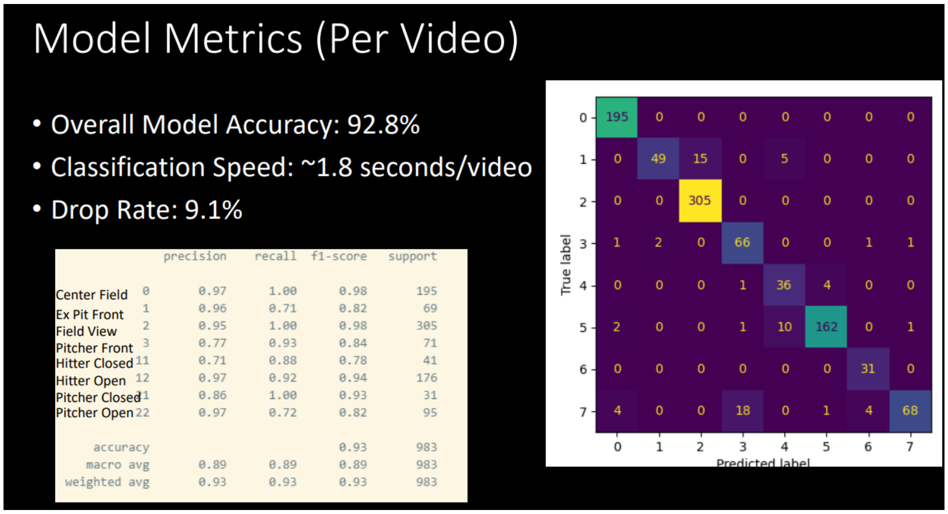


Figure 4: Classifier stats per class (left) as well as confusion matrix(right)



		precision	recall	f1-score	support
Center Field	0	0.97	1.00	0.98	195
Ex Pit Front	1	0.96	0.71	0.82	69
Field View	2	0.95	1.00	0.98	305
Pitcher Front	3	0.77	0.93	0.84	71
Hitter Closed	11	0.71	0.88	0.78	41
Hitter Open	12	0.97	0.92	0.94	176
Pitcher Closed	1	0.86	1.00	0.93	31
Pitcher Open	22	0.97	0.72	0.82	95
accuracy				0.93	983
macro avg		0.89	0.89	0.89	983
weighted avg		0.93	0.93	0.93	983

	0	1	2	3	4	5	6	7
0	195	0	0	0	0	0	0	0
1	0	49	15	0	5	0	0	0
2	0	0	305	0	0	0	0	0
3	1	2	0	66	0	0	1	1
4	0	0	0	1	36	4	0	0
5	2	0	0	1	10	162	0	1
6	0	0	0	0	0	0	31	0
7	4	0	0	18	0	1	4	68

Figure 5: Sample of 3 classes being detected (Bat, Plate, and Ball)

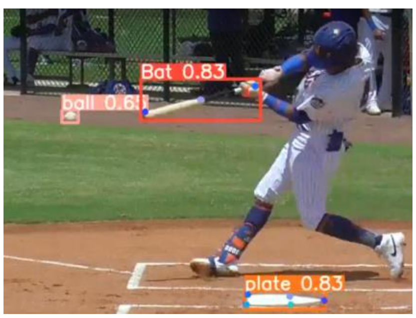


Figure 6: [Video]() of detections