Sport Analytics Report

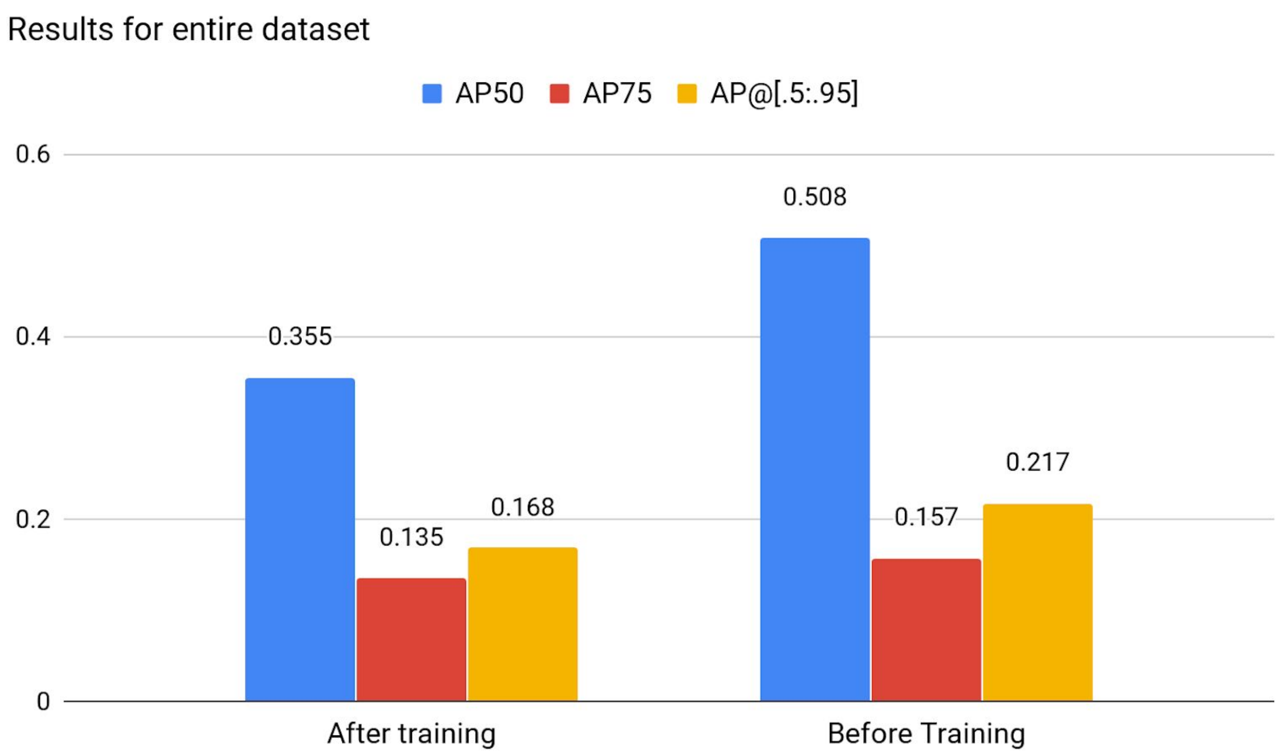
Work done:

Trained detection model on soccer data.	**Model**: YOLOv3 **Data**: 1500 images, top-zoomed out view, players annotated. Test set of 400 images, different views, jersey colors, teams. **Training**: Augmented images, trained for 300 epochs **Result**: Drop in overall accuracy for test set, accuracy for top-zoomed out view almost doubled. More..
Started implementation of tracking.	**Need**: Tracking needed to uniquely identify players positions to help estimate missing detections/registration when not in frame. **Methods**: Explored Centroid based matching, Kalman Filters, Deep SORT tracking. **Implemented**: Finished implementation of Deep SORT. More...
Player Registration	**Finished**: Matching methods to match edge maps. **Unfinished**: Optimisation methods to select the best homography based on the match.
Google Football Environment.	**Need:** Lack of football data for tracking, use it to generate synthetic tracking and detection data. **Finished**: Environment setup to generate data.
Formulated Ideas for using Sequence modelling to improve detection.	**Need**: Can be used to estimate missing detections. Can also be used for registration when player not in frame. **Proposal**: Proposed a three phase pipeline for training and deployment of this model. More...

Extended Notes:

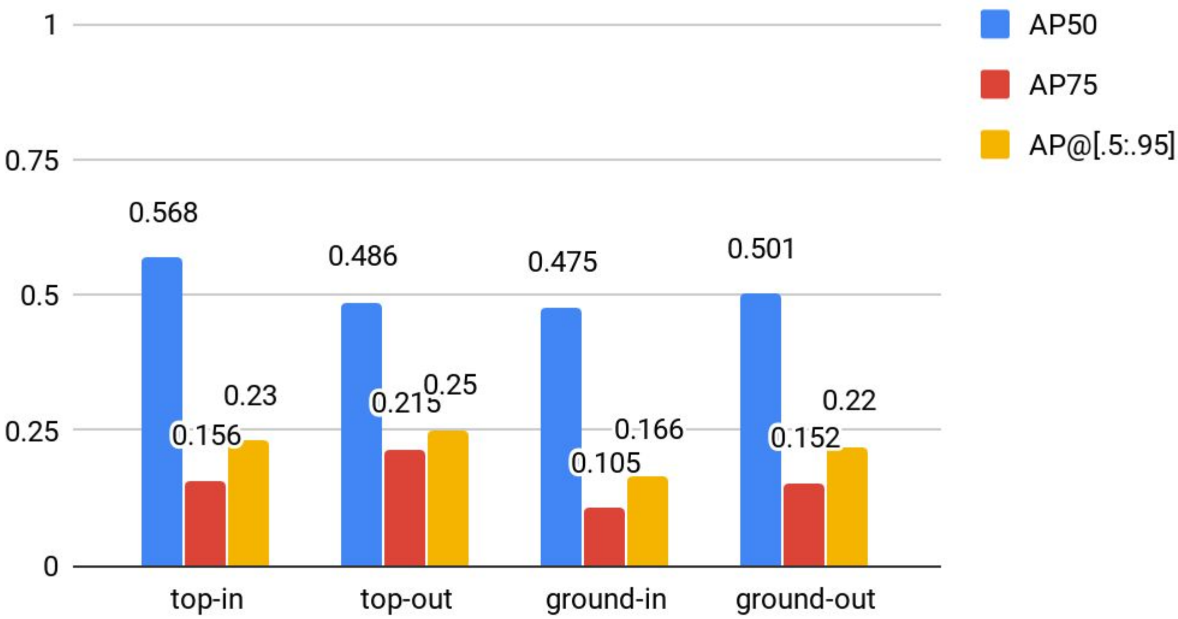
Detection model results:

- **Metrics used**: AP50, AP75 and AP@[.5:.95]
- **Results for the entire dataset:** As compared to the previous results obtained before training on top-view, we observe a loss in general accuracy, this is because the model loses generality but improves its results for a specific view, top-out view in this case.

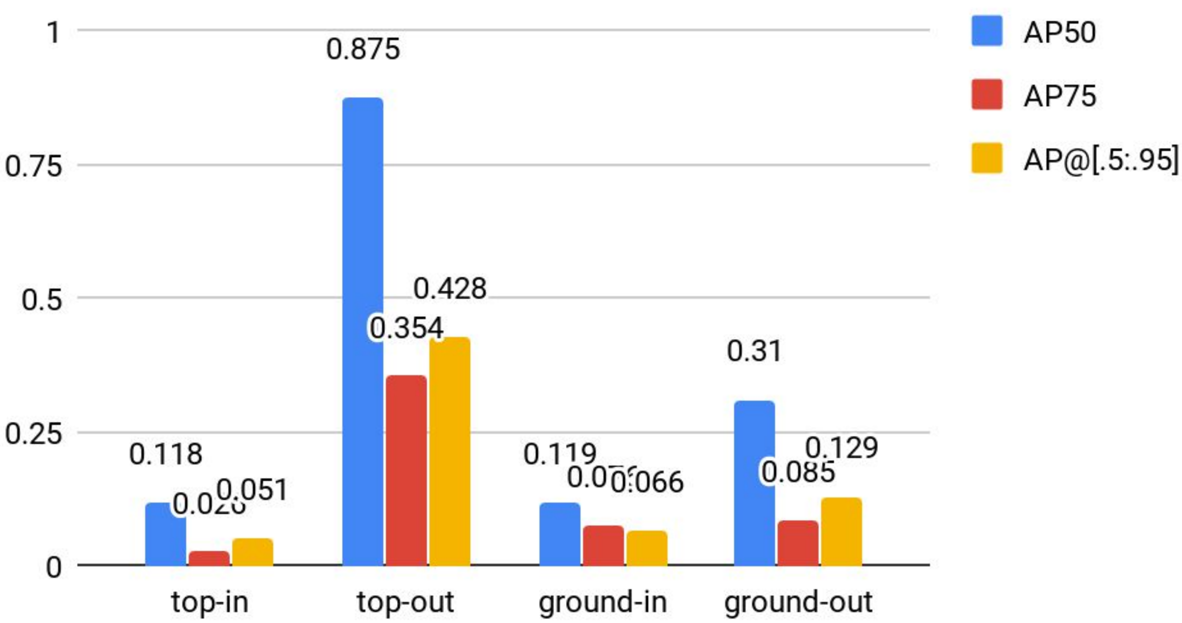


- **Results for different camera views**: We see a large jump in top out view accuracy. This is a desirable result for us, since registration can use only top-out views to map players on the field map. We see accuracy almost double for AP50 and AP@[.5:.95], with a significant increase even in AP75.

Camera based view before training



Camera based view after training



Tracking:

- **Need**: We want to be able to estimate player positions when detection is unable to identify bounding boxes for a player even if the player is in the frame. We use information from the players previous position and velocity to be able to estimate his current position(bounding box). For this we need to be able to differentiate the bounding boxes for each player, which can be done only with tracking. For tracking I researched various methods that are commonly used for tracking:
    - Centroid based matching
    - Kalman filter based tracking
    - Deep SORT tracking algorithm(State of the art)

- **Implementation**: Have implemented Deep SORT into the implementation of YOLO I am using to be able to track players. However, it still needs to be evaluated on a sports dataset. There are also no standard metrics to measure the accuracy of tracking methods, so I also need to decide upon the metric that needs to be used.

Sequence Modelling:

- Need: Sequence to Sequence modelling can be used to improve both detection as well as top-view registration. The Sequence model can be used to estimate missing detections. We can individually model the movements of each player by fine-tuning a generic player movement model.

- **Implementation details:**
    - A generic player movement model: This model learns to estimate missing detections given a series of detections with gaps in between. This will be a standard Sequence to Sequence model with an LSTM/GRU encoder-decode pair. We train this model to learn the relationship between the players past position and velocity to predict future values. This is modeled by sequence of bounding box coordinates(xyxy) given as input with missing detections in between given as a special token. The decoder output would be the correct sequence where the decoder replaces the tokens with the correct bounding box coordinates.

- Fine tuning generic model: Once we have a generic model that has modeled player movements, we finetune this model for each player in the field in realtime. We use shots that are clearer to be train the model, with fewer misses in detection/tracking. Then use this model to estimate the tracking/detection results for shots that are not as clear.

- Using model for registration: Once a player is out of the camera frame, we do not really know the exact location of the player on the field. Therefore it becomes impossible for us to estimate the player position just from camera views. Movement models that have been finetuned on each players movements can then be used to estimate the movement of each player, when he/she is not present in the frame.