

Soccer Player Tracking Report -

Arnavv Agnihotri (arnavvagnihotri06@gmail.com)

1. Approach and Methodology

The objective of this assignment was to track soccer players and the ball in a single video feed using a pre-trained YOLOv8 object detection model in combination with the DeepSORT tracking algorithm. The methodology followed includes:

- Utilizing a fine-tuned YOLOv8n model to detect players and the ball.
- Leveraging DeepSORT for object re-identification and consistent ID assignment over time.
- Optimizing detection thresholds and DeepSORT parameters to reduce false positives and overlapping boxes.
- Enhancing the visualization by labeling players with consistent and readable Player IDs.

2. Techniques Tried and Outcomes

Technique 1: YOLOv11 (custom-trained)

- Outcome: Failed due to model architecture incompatibility and version conflicts with Ultralytics packages. Required modules like `ultralytics.nn.modules.conv` were not found due to architecture mismatch or save format.

Technique 2: Switching to YOLOv8n

- Outcome: Successful. YOLOv8n provided real-time inference and stable bounding box detection of players and the ball.

Technique 3: Default DeepSORT settings

- Outcome: Caused inconsistent player ID assignment and multiple overlapping boxes per player. IDs changed when players left and re-entered the frame.

Technique 4: Tuning DeepSORT parameters (`n_init`, `max_age`, `nn_budget`, `max_cosine_distance`)

- Outcome: Improved temporal ID consistency, reduced drift, and controlled the number of tracks per player.

Technique 5: Feature-based ID hashing

- Outcome: Attempted using track features for consistent ID mapping using a hash dictionary. Ultimately dropped due to the absence of `.feature` attributes in DeepSort v0.3.1 onward.

Technique 6: Filtering redundant boxes

- Outcome: Partial success. Filtered duplicate or overlapping detections using basic heuristics (e.g., class confidence threshold and box overlap tolerance), but redundant boxes persisted due to close-proximity detections by YOLOv8.

3. Challenges Encountered

- **Incompatible YOLOv11 Model:** Could not load due to missing internal architecture dependencies. Switched to YOLOv8n.
- **torchvision errors:** Encountered `AttributeError: module 'torchvision' has no attribute 'extension'` due to a corrupted or mismatched torchvision installation.
- **Changing Player IDs:** A significant issue with ID switching when players re-entered the scene. Attempted multiple strategies to ensure stable ID retention.
- **Redundant Boxes:** Despite filtering, some frames still show multiple bounding boxes for the same player.
- **Audio preservation:** Video output lacked audio since OpenCV cannot handle audio streams. Required `ffmpeg` post-processing to restore audio.

4. Incompleteness and Future Work

While the current pipeline assigns consistent IDs to each player using DeepSORT, redundant bounding boxes still appear for a single player in certain frames. This occurs because:

- YOLOv8 sometimes detects the same player multiple times with overlapping boxes due to pose variation or rapid movement.
- DeepSORT, though tuned, creates separate tracks for close detections if temporal consistency is not met.

Proposed Fixes (If more time/resources were available):

- Implement Non-Maximum Suppression (NMS) on YOLOv8 outputs before passing to DeepSORT to merge overlapping player detections.
- Upgrade to a customized tracking pipeline (e.g., BoT-SORT or FairMOT) that integrates detection and tracking more tightly.
- Integrate a re-identification model that can compare embeddings between frames and reassign old IDs to re-entering players.

- Use a spatial clustering mechanism or box association heuristic to merge redundant tracks.

Conclusion

The current implementation achieves the core goals of tracking soccer players and assigning readable Player IDs that mostly persist across frames. Despite known issues with overlapping boxes, it is well-suited for prototyping and demonstrates a robust multi-object tracking pipeline using modern deep learning tools.

Further enhancements would polish the output and improve the visual clarity and ID consistency for professional-grade re-identification applications.