Player Re-Identification in Sports Video

1. Approach and Methodology

The objective of this project was to assign consistent and unique IDs to players across the frames of a 15-second sports video. This is crucial for sports analytics, performance tracking, and event recognition. The methodology relies on combining object detection and multi-object tracking (MOT) using a two-stage modular pipeline:

a. Object Detection using YOLOv11:

YOLO (You Only Look Once) is a real-time object detection algorithm. In this project, YOLOv11, a custom-trained version of YOLO, was utilized to detect players and the ball in each frame of the video. YOLOv11 outputs bounding boxes, class labels (player, ball), and confidence scores for each detection. The detections were saved per frame and later processed for tracking. Despite its accuracy in detecting the ball, player detection was inconsistent, likely due to:

- Limited and non-diverse training data.
- Variability in player appearance (jersey color, pose, motion blur).

b. Multi-Object Tracking with Deep SORT:

Deep SORT (Simple Online and Realtime Tracking) builds on the original SORT algorithm by incorporating appearance-based features and a Kalman Filter for more reliable tracking. The system uses:

- Kalman filtering for state estimation (position, velocity).
- Cosine distance between deep appearance embeddings for matching detections across frames.

The output of YOLOv11 (bounding boxes with class scores > 0.3) is used as input to Deep SORT. Deep SORT attempts to maintain consistent IDs across frames even with partial occlusion or movement. In practice:

- The ball was successfully tracked.
- Player tracking was less consistent due to detection failures.

2. Techniques Tried and Their Outcome

Player Re-Identification in Sports Video

a. YOLOv11: Custom Object Detector:

YOLOv11 was trained to recognize two object classes: player and ball. While ball detection was almost flawless across all frames, player detection was less accurate. Inconsistent player detection was mainly due to:

- Insufficient and imbalanced training data for diverse player appearances.
- Dynamic scenes such as occlusions, fast movements, or similar background colors.

b. Deep SORT: Tracking and Re-Identification:

Deep SORT was applied only to detections with a confidence score > 0.3. It could maintain identity (ID) of a few players when detection was consistent. Problems occurred when:

- Players re-entered the scene.
- Players occluded each other.
- Detection was missed in a few key frames.

As a result, IDs were not preserved across the full clip for all players.

c. Output Visualization:

For visual inspection, bounding boxes with IDs were drawn on the video. A CSV file was generated containing frame-by-frame tracking information:

- Frame number
- Object ID
- Bounding box coordinates
- Class label

This data serves as a foundation for future analytics, such as player trajectory analysis or performance metrics.

3. Remaining Work and Future Steps

Although the current system achieves basic tracking and ball detection, it requires improvements to be production-ready for real-world applications such as live sports analytics, player heatmaps, and tactical analysis.

a. Improving Detection Performance:

- Retrain YOLOv11 with a larger, labeled dataset that includes diverse player appearances (skin tone, jersey design, body shape).
- Data augmentation techniques (rotation, occlusion, scaling) to improve generalization.

b. Class-Specific Filtering:

- Refine tracking by focusing only on relevant object classes, particularly 'person', to reduce false positives from other detected objects.

c. Appearance-Based Enhancements:

- Incorporate additional cues for re-identification:
  - Jersey color patterns (using color histograms).
  - Pose estimation for distinguishing players by movement style.
  - Player numbers, if visible, could offer a definitive identity feature.

d. Handling Occlusion and Re-Entry:

- Apply ReID networks trained on player appearance datasets to better maintain identity when players exit and re-enter the frame.
- Leverage temporal context using LSTM or Transformer-based trackers to understand motion over time.

e. Extend to Multi-Camera Setup:

- In future iterations, the system can be adapted to work across multiple camera angles (cross-view Re-ID), which is commonly used in professional sports broadcasting.

Conclusion

The project demonstrates the feasibility of basic player and ball tracking in sports videos using deep learning techniques. While the current pipeline works under controlled conditions, further development in data collection, model training, and feature engineering is required for robust and real-time deployment. These improvements can unlock numerous applications in sports analytics, automated highlights, and coaching tools.