Football Player Re-identification Report

1. Introduction

Football player re-identification is the task of consistently identifying and tracking individual players as they move in and out of a camera's field of view. This is challenging due to frequent occlusions, uniform similarity, and rapid motion. In this project, we process a 15-second video clip using a YOLO fine-tuned detection model, a temporal tracking algorithm (ByteTrack), and a re-identification network (OSNet) to maintain consistent identity assignments across frames.

2. Data and Preprocessing

- **Video Input**: A 15-second soccer clip captured by a single camera.
- **Detection Model**: YOLOv11 fine-tuned on a football dataset, detecting only players (no IDs).
- **Output Exploration**: Initial YOLO output provided bounding boxes and class labels (players), but inconsistent identity assignment across frames.

3. Object Tracking

3.1 Alternative Trackers Explored

- **DeepSORT**: Combines Kalman filtering and appearance features. Rejected because uniform similarity in team sports yields poor discriminative embeddings.
- **SORT**: Uses only motion (Kalman filter + Hungarian). Dropped due to ID switches under occlusion and crowding.
- **ByteTrack**: A temporal tracking algorithm that leverages both high- and low-confidence detections to form robust tracklets. It first associates tracks with high-confidence detections, then recovers missed matches by incorporating low-confidence ones, reducing ID fragmentation.

3.2 ByteTrack Implementation

- **Parameters Tuned**
  - `conf_thres` (detection confidence): 0.5
  - `iou_thres` (association IoU): 0.45
  - `frame_buffer`: 50 frames (temporal memory)

- **Results**: After tuning, ByteTrack maintained consistent IDs during normal motion. However, under occlusions (when a player is temporarily hidden behind another object or leaves the frame), IDs were often reassigned.

4. YOLO Confidence Tuning

YOLO occasionally produced high-confidence false detections (e.g., player legs) and failed to detect players in non-standard poses (proned). I adjusted its confidence threshold to balance precision and recall, improving detection quality for tracking.

5. Player Re-identification

5.1 Color Histograms

- **Concept**: Represent each detected player's RGB distribution as a feature vector. Map these histograms to tracking IDs to recover identity after occlusion.
- **Limitations**: Uniform kits produce near-identical histograms for same-team players. Insufficient for fine-grained discrimination in soccer.

5.2 OSNet Re-identification

- **Model**: OSNet (Omni-Scale Network), pre-trained on pedestrian datasets, captures multi-scale features (clothing, accessories, gait). More discriminative than simple color features.
- **Dataset Preparation**:
  - Extracted all player crops from the first 5 seconds.
  - Organized into training, query, gallery splits at a 70:15:15 ratio per player.

- **Finetuning Setup**:
  - Framework: `torchreid`
  - Base layers frozen; adapter (classifier) layers trained.
  - Output layer Embedding dimension: 512.
  - Optimizer: Adam (learning rate = 0.0003).
  - LR scheduler: single-step decay at epoch 20.
  - Batch strategy: ensure balanced samples per identity.

6. Results

Computing distance matrix with metric=euclidean ...
Computing CMC and mAP ...
** Results **
mAP   : 79.4%
CMC curve
Rank-1 : 89.5%
Rank-5 : 94.6%
Rank-10 : 96.0%
Rank-20 : 97.8%
Checkpoint saved to og/osnet/model/model.pth.tar-1
Elapsed 0:15:57

- **mAP (Mean Average Precision)**: 79.4% average retrieval precision across all queries.

- **CMC (Cumulative Matching Characteristic)**:
  - *Rank-1*: 89.5% of queries correctly match the true identity at the top-1 position.
  - *Rank-5*: 94.6% within top-5 candidates.

- **Training Time**: ~15 minutes and 57 seconds on a colab CPU.

7. Discussion

While OSNet significantly improved identity consistency, confusion persists among visually similar teammates. Challenges include:

- Limited appearance variation in same team uniforms.
- Short tracking buffer leading to ID switch after long occlusions.

8. Future Improvements

1. **Motion Cues**: Integrate proximity and estimated player speed to link re-entries plausibly.
2. **Entry Constraints**: Only allow new IDs at frame borders, assuming a single static camera without cuts.
3. **Camera Motion Compensation**: Detect camera movement and adjust player trajectories accordingly.
4. **Multi-view Fusion**: Use additional camera angles for robust identity linking.
5. **Cross-domain Techniques**: Explore satellite/space object tracking (e.g., NASA methods) or animal tracking literature for advanced association strategies.
6. **Vision Transformers**: Employ state-of-the-art re-identification models based on Vision Transformers which capture long-range spatial dependencies and may outperform CNN-based approaches like OSNet.
