Objective

To match player identities across two distinct camera perspectives—**Tacticam (top-down)** and **Broadcast (side-view)**—using detection, tracking, and feature-based identity modeling.

Methodology Overview

1. Detection with YOLOv11

- Used a fine-tuned `best.pt` YOLOv11 model to detect key objects: players, referees, goalkeepers, and the ball.
- Inference performed frame-by-frame on both video streams to generate bounding box detections.

2. Tracking with Deep SORT

- Deep SORT assigned consistent `track_id` values to each player across frames, enabling temporal identity persistence.
- Bounding boxes were reformatted properly to comply with Deep SORT input expectations.

3. Feature Embedding via Deep ReID (OSNet)

- Extracted appearance embeddings for each tracked player crop using a pretrained OSNet model from `torchreid`.
- Embeddings were averaged over time for each player to enhance stability.

4. Matching Across Views

- Used **cosine similarity** between embeddings and applied **mutual nearest neighbor matching**:
    - Only retained matches that were each other's best candidates with similarity above a threshold.

5. Spatial Alignment (Optional)

- Computed homography matrix between the views based on manually selected keypoints (e.g., field lines).
- Warped Tacticam player coordinates and compared them to Broadcast positions to validate spatial plausibility.

6. Visualization

- Created side-by-side comparisons of matched players on synchronized frames using OpenCV.
- Enabled manual inspection to verify matching accuracy visually.

Challenges & Solutions

Challenge 1: Many-to-One Matching

- Early appearance features (color histograms) led to multiple Tacticam players mapping to the same Broadcast ID.
- ✅ **Solution**: Replaced low-discriminative features with Deep ReID embeddings via OSNet.

Challenge 2: Detection Gaps & Frame Skips

- YOLOv11 produced frames with no detections or incomplete coverage (due to occlusion or motion blur).
- **Solution**:
    - Allowed empty detection frames gracefully within the pipeline.
    - Used Deep SORT's age mechanism to retain player identities across minor detection gaps.

Challenge 3: Inconsistent Lighting & Pose

- Broadcast view had strong lighting variation and frontal poses, while Tacticam showed more top-down geometry.
- **Solution**: Leveraged pose-invariant embeddings from the OSNet ReID model, which is trained on diverse view angles.

Challenge 4: False Positive Matches

- Some matched players had different field positions or roles.
- **Solution**:
    - Applied mutual nearest neighbor constraint.
    - Introduced optional homography-based position validation.

Results

- Final mapping produced cleaner one-to-one associations.
- Visual inspection confirmed stronger matching precision.

Example output:
python
Player Mapping: {'3': '16', '7': '53', '10': '19', '11': '21', '13': '15', ...}

-

Conclusion

This pipeline demonstrates a multimodal approach combining **visual features**, **spatial reasoning**, and **temporal consistency** to solve the challenging task of cross-camera player mapping. The evolution from simple appearance descriptors to ReID-enhanced embeddings, combined with homographic projection, significantly improved match quality and robustness.