Soccer Player Re-Identification — Cross-Camera Mapping (Option 1)

1. Approach & Methodology:

The goal of this project was to match players appearing in two separate video feeds (broadcast.mp4 and tacticam.mp4) and assign them consistent IDs across both views. The task was addressed using the following pipeline:

a) **Object Detection**:
- Utilized a pre-trained YOLOv11 (best.pt) model, fine-tuned specifically for player and ball detection.
- Extracted player bounding boxes for each frame in both videos.

b) **Appearance Feature Extraction**:
- Cropped player images based on YOLO detections.
- Passed each crop through a pre-trained **ResNet50** model (ImageNet weights) to extract 2048-dimensional embeddings.

c) **Player Matching**:
- Computed pairwise **cosine similarity** between each player in tacticam and all players in broadcast.
- Assigned each player in tacticam the ID of the most visually similar player in broadcast.

2. Techniques Tried and Outcomes:

Technique	Description	Outcome
YOLOv11	Used to detect players and ball	Accurate detections; fast inference
ResNet50 (Feature Extractor)	Converts player crops into 2048-d feature vectors	Captured visual appearance well
Cosine Similarity Matching	Measures similarity between feature vectors	Worked well for players with clear views
Per-frame Matching	Currently only matches using one frame/sample per player	Works for short clips, could be improved for long-term tracking

Output:

- JSON file (match1.json) mapping tacticam_id to broadcast_id with similarity scores.

Example:

```
[
  {"tacticam_id": 0, "matched_broadcast_id": 3, "similarity": 0.91},
  {"tacticam_id": 1, "matched_broadcast_id": 0, "similarity": 0.87}
]
```

3. **Challenges Encountered:**
    - **Player Occlusion**: Partial visibility in one view reduced feature quality.
    - **Pose Variations**: Players facing different directions affected matching reliability.
    - **Temporal Consistency**: Single-frame matching may not capture player motion or context over time.
    - **Lack of Ground Truth**: No reference IDs to validate the accuracy of cross-camera matching.
4. **If Incomplete: What's Missing & Future Plans:**
    - **Track players over time** using DeepSORT before matching, to use temporal consistency.
    - **Use multi-frame feature aggregation** (averaging embeddings over frames).
    - **Improve re-ID model**: Replace ResNet50 with a network trained specifically for person re-identification.
    - **Visualize results** with side-by-side video frames showing matched IDs.
    - **Evaluate accuracy** using annotated ground truth if available.