Player Re-Identification in Soccer

Problem:

Whenever the player goes out of the frame or is being occluded and then comes back he/she is assigned a new player ID. We need to fix this problem in a way that the same player get's the same ID when he get's back in the frame

Breakdown of the main issue:

Video → divide it into frames → Apply detection algorithms → Apply tracking algorithms (this also includes reidentification)

Object Tracking Algorithms:

1. SORT (2016) - Simple Online Realtime Tracking
2. DeepSORT(2017)
3. StrongSORT(2023)
4. ByteTrack (2021) - Fast, simple but less accurate
5. BoTSORT (2022) - slow, more accurate (handles occlusion and reidentification pretty well)

1. Intuition based on Initial Research (Manual way):

What if we can store the image or features of a player with its assigned id and whenever a new player comes in the frame we check for similarity in the database and if it exists, Assign the player it's old player ID.

A more refined version based on a reddit answer:

For every person the system detected in a video frame, they saved a cropped image (called a *chip*) of that person. They only kept chips from *last X seconds*. Eg: the last 10 seconds — not forever.
If the same person is detected consistently across a few frames (e.g., 5 frames in a row), it's not just a glitch, it's a real person. So now it's worth tracking. Take the new person's image and compare it to all the previously stored chips using a siamese network
(a neural network that tells how visually *similar* two images are). This helps figure out if this new person was seen before.

→ If the new image is similar to any stored chip, it's the **same person**, so assign it the same ID.
If not, it's a new person, so add their image to the catalog as a new ID.

Few Questions:
How to decide X (seconds)?

How to decide an optimal amount of frames to discard the possibility of a glitch..?

2. Check out some of the ready-made tools/libraries out there

TorchReid: https://github.com/KaiyangZhou/deep-person-reid?tab=readme-ov-file

DeepSORT : https://github.com/levan92/deep_sort_realtime

StrongSORT and ByteTrack: https://github.com/FoundationVision/ByteTrack

TransTrack: https://github.com/PeizeSun/TransTrack?tab=readme-ov-file

MOTRv2: https://github.com/megvii-research/MOTRv2?tab=readme-ov-file

Some questions I had regarding the assignment:

1. Do you want us to provide a solution which work for the dataset(2 videos) given in this folder or in general?
2. Does the solution need to have ReID over different gameplay for the same player? Also it that thing even feasible rn?
3. Should I start with the easy way (Using existing libraries or frameworks) or try the harder way (implement my own algorithm)

Two types of methods exists:

1. Traditional CNN based methods
2. Transformer based

Methods I tried:

1. DeepSORT with mobilenet embedder:
   Fast but less accurate, ID switches frequently even during occlusion,
   FPS around (16ps)
2. DeepSORT with osnet embedder:
   slower than mobilenet but more accurate, ID switches but less during occlusion and more when the player comes back after going out of the frame
   FPS around (8ps)
3. BotSORT = false positivies negligible but ReID problem still persists, slow
4. BotSort with reid = with_reid is a parameter introduced in the new version. Very slow, still couldn't reidentify once player goes out of the frame and comes back
5. ByteTrack = Faster but more freq ID switches

What Could Have Been Done

Given more time and resources, the following could enhance the solution:

- **Fine-tuning a custom ReID model** on soccer-specific data using frameworks like TorchReID or CLIP-based vision-language models (as explored in CLIP-ReIDent).
- **Siamese or Triplet networks** trained on soccer player crops to improve matching performance during occlusions or out-of-frame intervals.
- **Multiview fusion** using camera angles (if available) to enhance ReID consistency.

Evaluation Criteria:

1. Manually checking for ID switches and false positives (not very reliable, although Al Pacino once said "The eyes, Chico, they never lie" but here they just might)
2. Annotating the video file to get ground truth values but that seems impractical for this assignment. As we may have many videos and we can't do manual annotation for all. Although I did give it a try (detector is working fine) but it's difficult to keep track of ID switches with each frame.

Resources:

1. **SoccerNet challenge:** https://www.soccer-net.org/challenges/2023
2. Reddit QnA: https://www.reddit.com/r/computervision/comments/1dgpiyo/computer_vision_ai_development_for_sports/
3. https://www.reddit.com/r/computervision/comments/1cxhjnm/reidentification_in_sports/
4. https://www.reddit.com/r/computervision/comments/1dtweye/reid_problem/
5. https://www.reddit.com/r/computervision/comments/1ky9e6j/how_to_maintain_consistent_player_ids_in_football/
6. A nice research paper https://arxiv.org/pdf/2203.02281

Some Great papers to check out:

SoccerNet 2023 winner: https://arxiv.org/pdf/2309.06006, https://arxiv.org/pdf/2303.11855(paper: CLIP- ReIDent)

https://arxiv.org/abs/2001.04193

> 💡 Introduced a new evaluation metric called mean Inverse Negative Penalty (mINP) for person re-ID

https://arxiv.org/abs/2401.06960v2

https://arxiv.org/abs/1903.07071

https://www.nature.com/articles/s41598-024-51767-4

https://link.springer.com/chapter/10.1007/978-981-96-1445-5_2#Sec4