INTRODUCTION

The game of football Player re-identification (ReID) is an essential subsidiary activity of broader area of sports video recognition and intelligent surveillance. With the growing need to monitor player behavior, it is emerging that there is a growing desire to be able to track players consistently across a variety of camera angles like a broadcast and a tactical angle. As opposed to controlled conditions, football contests have some special obstacles to overcome, players are in the same jerseys, have moments of fast movements and occlusions, and come into view on different intensities and size depending on the camera angle. It is because of these complications that traditional tracking is inadequate when it comes to cross-view identity association.

This project tackles these issues by creating a multi-stage pipeline that combines state-of-the-art methods of computer vision; detection, tracking and re-identification based on deep learning. Identification of the players on each frame is developed through custom-trained YOLO model that is made available by Liat.ai. MARS embeddings are appended to the DeepSORT algorithm and enables it to assign unique track IDs deterministically over time in the same camera stream. However, as due to the limitations of DeepSORT MARS embeddings according to which the training samples and the tested ones have to belong to the same camera domain, we simply enhance the pipeline: we add OSNet (Omni-Scale Network) models, which are characterized by high-quality discrimination of appearance features that can support ReID tasks.The goal is to sample embeddings of DeepSORT-tracked players with OSNet, and compare them in the two views with respect to cosine similarity to assign identity results. This approach enables the system re-recognize users between the videos, even without available explicit ground truth. To aid this procedure, average embeddings are calculated per track as a function of time and visual comparison is done as a frame-to-frame comparison, view in a side-by-side construct with IDs added.

Overall, this project proposes a working system to unify spatial, visual and temporal information to address the problem of the re-identification of the player under a cross-view scenario in football. It outlines the advantages of uniting modular solutions: object detection, tracking, and deep learning ReID, as well as limitations that can be faced because of the lack of data and the ground truth. The solution will serve as a good basis on future developments of sport video understanding and analytics automatically.

METHODOLOGY

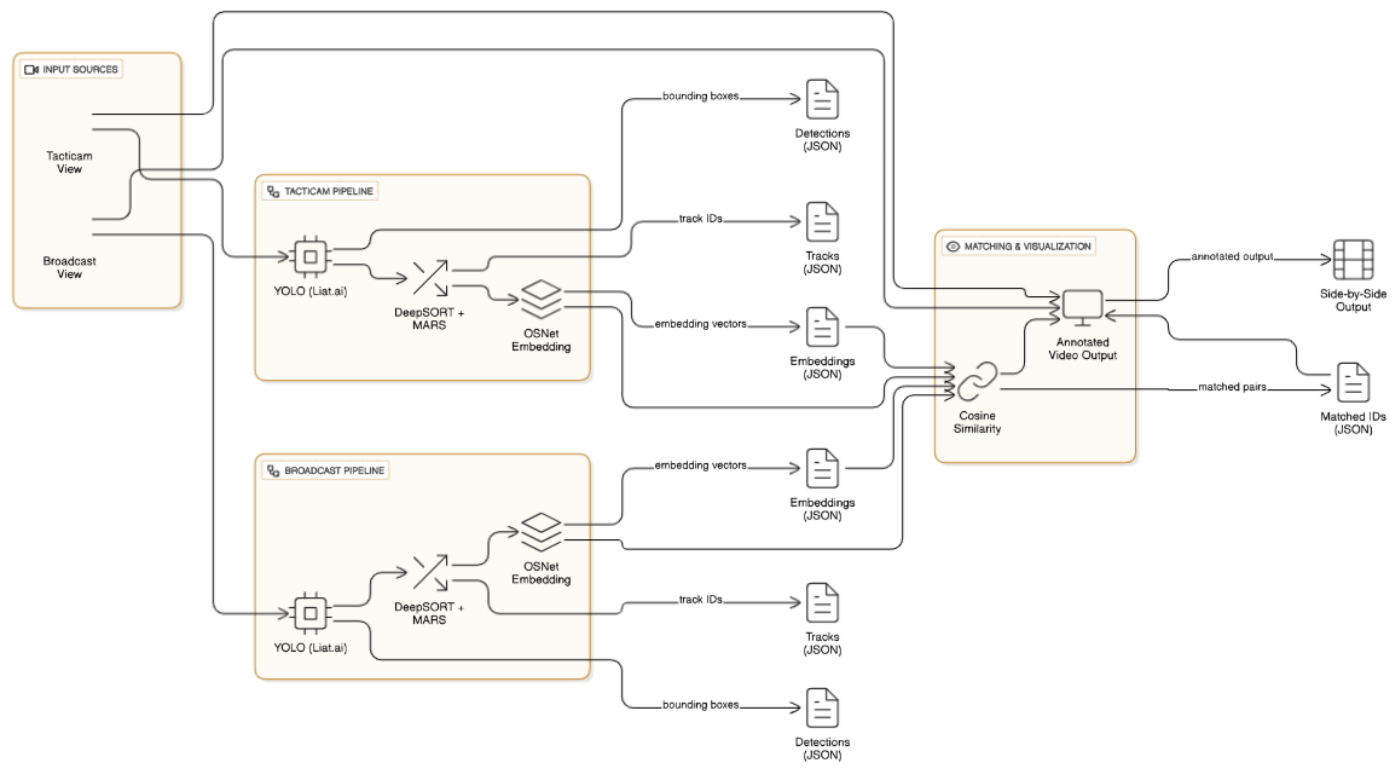
The modular pipeline followed during this project includes detection, tracking, extraction of the embedding and matching. First, human detection was carried out on both videos broadcast and tacticam with the model based on YOLO offered by Liat.ai. In every frame, this model produces one bounding box as well as confidence per detected person. Output of the detecting is saved in ordered JSON files.Such detections are fed into DeepSORT, an online multi-object tracker which combines motion (Kalman filtering) and appearance (deep metric learning). It has the MARS embedding model to depict the look of every individual. This model helps DeepSORT to be persistent on overlaps and quick actions. Consequently, every participant in the video is allotted an individual and consistent track ID in the video.

When stable track identifications are made, the system uses OSNet (Omni-Scale Network) of Torchreid library to retrieve discriminative feature embeddings for every tracked individual. OSNet takes the crop of each player and extracts 512-dimensional feature vector as a representation of visual identity. To increase dependability several embeddings of a player (of various frames) are

averaged. Such averaged features are re-identified by calculating the cosine similarity between the two views of the camera.Lastly, matching of players is done based on a nearest-neighbor algorithm in the feature space and matching pairs are then shown in side-by-side annotated video. The result delivers quantitative data (matched IDs) and qualitative visual material on the performance of re-identification.

This project compares the performance and role of multiple models:

Model Name	Purpose	Strengths	Notes
YOLOv5 (Liat.ai)	Detection	Fast, accurate bounding boxes for humans	Custom-trained for football domain
DeepSORT + MARS	Tracking + Appearance	Real-time capable, stable tracking with ID continuity	MARS helps match people across frames
OSNet_x1_0	Visual ReID	Balanced speed and accuracy, captures multi-scale features	Used for embedding extraction



CHALLENGES FACED

A significant challenge was the lack of ground truth data to quantitatively assess accuracy. Without labeled identity correspondences between views, metrics like precision, recall, or F1 score couldn't be computed. As a workaround, reliability was assessed visually via side-by-side frame comparisons.

Another challenge was processing time. OSNet is relatively computationally expensive, and embedding extraction for each player in every frame took considerable time. To optimize this, the embedding code was modified to extract features for a maximum of three frames per track ID, significantly speeding up the process without major loss of performance.

Maintaining ID consistency between detection, tracking, and embedding was also non-trivial. Errors in one stage could propagate to the next. It was important to debug ID mismatches carefully and ensure that bounding boxes were correctly associated across frames and views.

RESULTS AND OBSERVATIONS

The visual results suggest good performance. The side-by-side videos with bounding boxes and identity labels show that most players are matched accurately between the broadcast and tacticam views. The DeepSORT tracking was stable, even through partial occlusions, and the OSNet embeddings proved effective for cross-view re-identification.

Qualitatively, matched players appeared in similar poses, movements, and outfits in both views. This alignment suggests that the feature representations captured by OSNet were robust to camera view changes. However, due to the lack of ground truth, the results are qualitative and subjective.

FUTURE WORK

To move beyond subjective analysis, the next step would be to create or acquire a labeled dataset with cross-view identity matches. This would enable precise computation of metrics such as top-1 accuracy, mean Average Precision (mAP), and cumulative match characteristic (CMC) curves.

Another area of improvement is fine-tuning OSNet on sports-specific datasets, especially football datasets, where player poses, uniforms, and background clutter differ from general-purpose datasets like Market-1501.

Future extensions could include:

- Using temporal sequence models like LSTMs to capture movement patterns
- Incorporating spatial information from player positions on the field
- Using graph-based matching algorithms to resolve one-to-many or ambiguous matches
- Scaling to more than two cameras for real-world broadcast setups