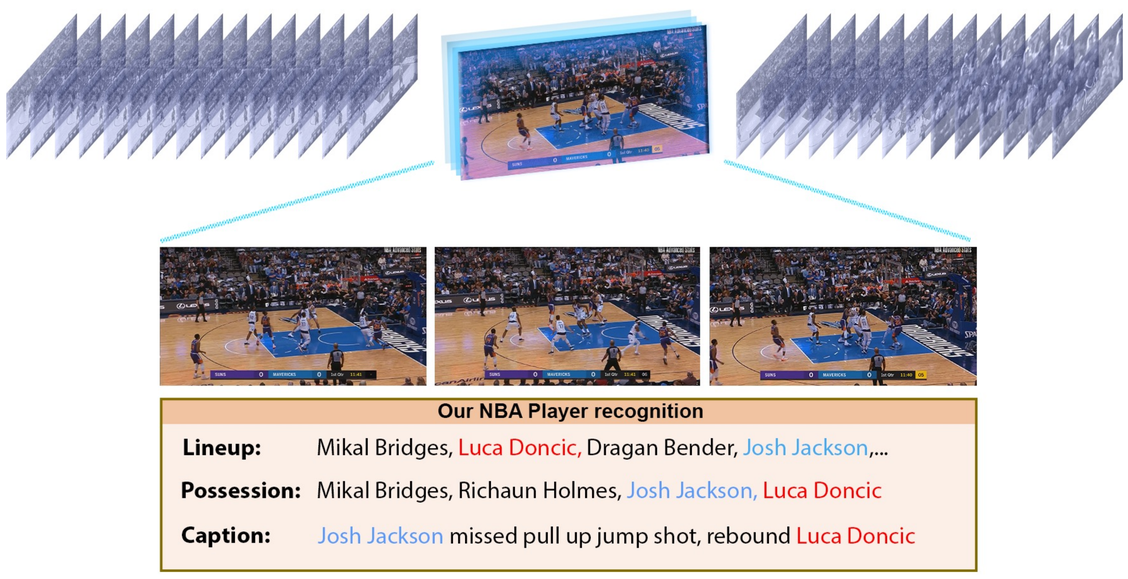
Video Models for Sports Analytics

Karol Wojtulewicz, Linköping University
Department of Computer and Information Science (IDA)

Problem Definition

Current player tracking systems in sports, particularly basketball, face significant challenges due to the sport's dynamic nature, including non-linear player movements and frequent occlusions. These limitations often result in the loss of accurate tracking. This research aims to develop a robust tracking model that integrates player tracking with an ID verification network, ensuring continuous and precise player identification even in zero-shot scenarios. Additionally, the study explores the fusion of video streams with game data, such as rosters, to enhance the accuracy and detail of sports video analysis. The proposed solutions seek to improve both the accuracy and computational efficiency of sports analytics models, facilitating real-time inference and decision-making, ultimately enhancing the overall quality of sports analysis.

Related Work



Multimodal Learning: The integration of multiple data sources, such as video and audio, has been shown to improve prediction accuracy in various domains. For instance, in emotion recognition, combining facial expressions (video) with voice tonality (audio) provides a more holistic understanding of the subject's emotional state [1].

Video Analytics: Video models, like convolutional neural networks (CNNs), have been pivotal in activity recognition. The temporal nature of videos is captured using 3D CNNs and then can be further analyzed with a transformer network [2].

Sports Analytics: The sports industry has been revolutionized with the advent of video analytics. Coaches and analysts use video replays not just for post-match analysis but also for real-time strategy planning. Predicting game strategies and player movements with video has become an active research area [3].

Under Investigation

Zero-Shot Player Tracking: Creating systems for accurate player tracking in zero-shot scenario based on physical features and limited options space. Current tracking systems have limited understanding of objects they are tracking, and it is easy to lose a track (on things...). Basketball videos specifically are a difficult challenge for tracking models because of the fast dynamic of the sport like non-linear player movements and frequent occlusions.

Idea Expansion: Develop a reliable tracking model that can be fused together with an ID verification network, that keeps track on the players that are visible out of the ones in the player dictionary.

Sports Video Annotation: Accurate and detailed sports video analysis involves fusing a video stream together with additional information that are available during games, like rosters.

Idea Expansion: Explore the area of information fusion and investigate the potential of injecting game data for video analysis.

Real-time Inference: For applications like sports analysis, real-time decision-making is in high demand. This requires models that are not only accurate but also computationally efficient. Edge computing, where processing occurs closer to the data source, can reduce latency and enable real-time insights.

Idea Expansion: Explore fast detection networks that can make it possible to perform end-to-end tracking and prediction frame-by-frame.

References

[1] Mustaqeem, Soonil Kwon, (2020). MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach. Expert Systems with Applications

[2] Luo H., Ji ,L. (2020). UniVL: A Unified Video and Language. Pre-Training Model for Multimodal Understanding and Generation. Microsoft STCA

[3] Williams, S., & Taylor, L. (2018). SoccerNet: A scalable dataset for action spotting in soccer videos. CVPR Workshops