FOOTBALL INSIGHTS

VIDEO ANALYSIS USING DEEP LEARNING

Takács Tamás and Borsy Máté

Supervisor:
Dos Santos Melício Bruno Carlos

Content:

Project repository and demo videos:

https://github.com/borsym/FootballInsight/tree/main

FOOTBALL INSIGHTS

VIDEO ANALYSIS USING DEEP LEARNING

Introduction:

Football Insights is a project aimed at leveraging cutting-edge computer vision techniques to analyze football matches, generate meaningful statistics, and develop skills in data analysis and visualization.

Motivation:

- Explore recent methods in object detection, tracking, and optical flow.
- Generate meaningful football match statistics using extracted features.
- Provide valuable insights for injury prevention.

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Project Scope:

A. **Detect all football players** on input frames.

B. **Detect the ball** on all input frames.

C. **Track the players** by ID throughout input frames.

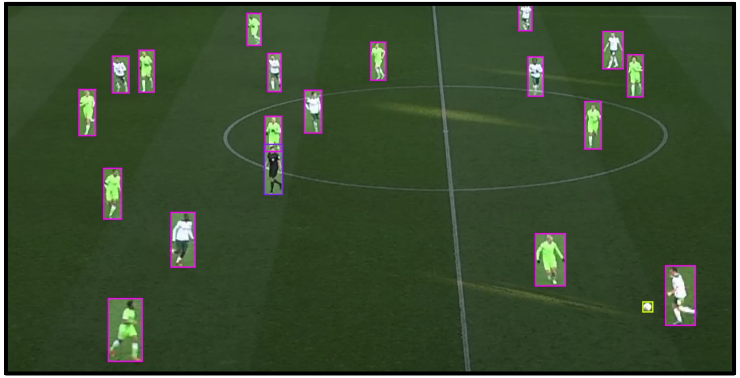
D. **Track the ball** from one input frame

E. **Estimate the poses** of players on input frames.

F. **Calculate actions** for all players on input frames.

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Data:



[Roboflow Dataset: (training)]

- 204 Training Images
- 38 Validation Images
- 13 Test Images
- 4 classes: [ball, goalkeeper, player, referee]
- YOLOv8 Yaml Config
- Used for training a custom YOLOv8 model

[MCI - ARS: (2023)]

- Used for Inference
- 720p Resolution

[FCB - SFC: (2015)]

- Used for Inference
- 1080p Resolution

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Methods:

1. **Detecting and Tracking the Ball**

   SAM + XMem

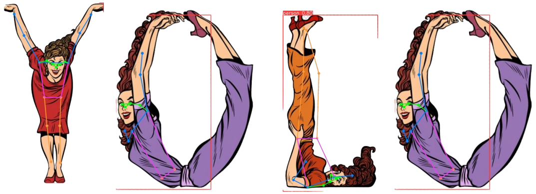
2. **Detecting and Tracking the Players**

   YOLO #1: 0.93

   YOLO

   + DeepSORT

3. **Estimating the Poses and Actions of Players**



+

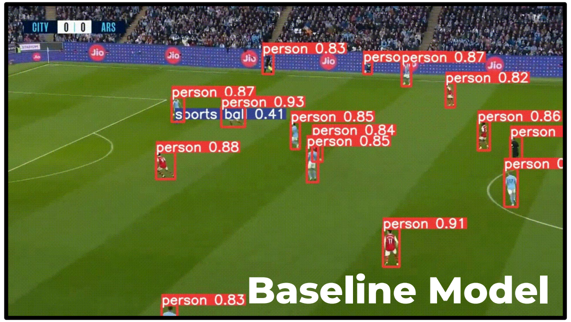
GEOMETRY

FOOTBALL INSIGHTS

VIDEO ANALYSIS USING DEEP LEARNING

A. Detect all football players on input frames:

- YOLOv8 - Ultralytics:
  - You Only Look Once version 8, is an advanced object detection model that builds upon the previous versions of the YOLO.
  - YOLOv8 was launched in early January 2023 so it is still under heavy development.
  - The developers behind version 8 are the original developers behind YOLO.
  - In the project it is used for training a custom object detection model on the **Roboflow** dataset.



More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

B. Detect the ball on all input frames:

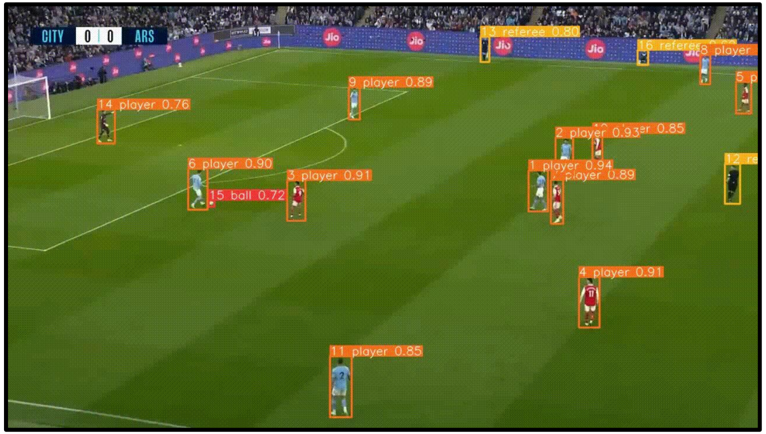
- Segment Anything - Meta AI (SAM):
  - Zero-shot generalization.
  - Segments a large variety of objects extremely well.
  - Fast Inference (0.15 seconds on an NVIDIA A100 GPU)
  - ViT-H SAM model (636M params, largest model ~ 2GB)
  - Segments the playing field and football player extremely well.

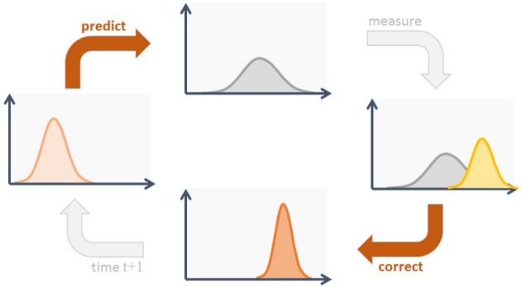
More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

C. Track the players by ID:

- DeepSort - Nicolai Wojke, Alex Bewley, Dietrich Paulus:
  - Simple Online and Realtime Tracking with a Deep Association Metric.
  - Uses Object Detectors such as YOLO to get bounding boxes of objects.
  - CNN to encode deep features of objects (ResNet-152).
  - Kalman Filter to predict future location of objects.
  - Custom re-identification of players in occluded areas.





More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

D. Track the ball from one input frame

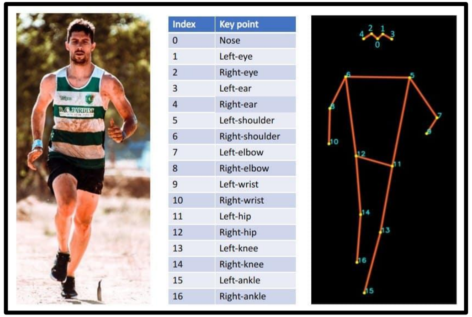
- XMem - [Ho Kei Cheng, Alexander Schwing]:
  - YOLO serves as the starting point for Xmem, an advanced tracking system, which requires an initial detection to begin its tracking operations.
  - SAM Segmentation provides a mask for XMem, indicating the ball's location, allowing for precise tracking.

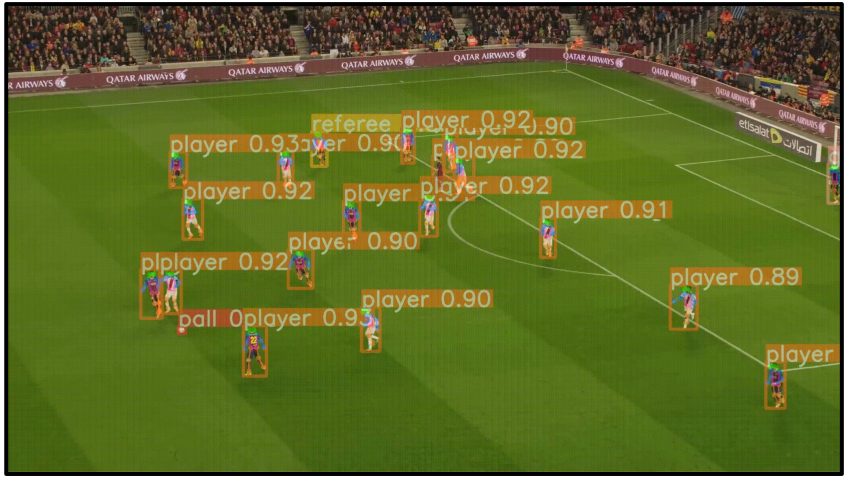
More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

E. Estimate the poses of players:

- YOLOv8-Pose - Ultralytics:
  - Recently added to YOLOv8. (as of May 2023)
  - Kinematic Joint-based pose estimation. (17 keypoint model)
  - *YOLOv8x-pose-p6* model. (mAP@[.5:.95]: 71.6, 99.1M params, 0.01s inference)
  - Used for predicting the moves of the players on the pitch.





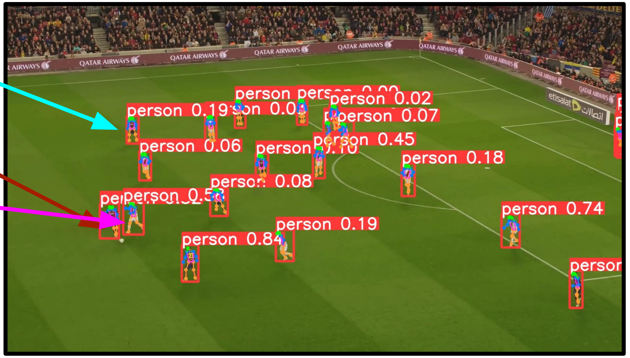
More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

F. Calculate actions for all players:

- Hard Rule-Based Method Without Training:
  - Requires no excess data.
  - Ultrafast, utilizing normalized distance and vector angle metrics.
  - Requires creativity to fine-tune it to human kinematics.
  - Highly flawed due to human bias and inflexibility.
  - Pose Estimation is inaccurate -> shifted pixel values -> incorrect predictions.
  - Requires close-up shots of players to work efficiently.

```
Player on x:542.8930053710938 y:735.842041015625 is running:
Player on x:652.3729248046875 y:560.2882080078125 is standing or doing nothing:
Player on x:375.04571533203125 y:335.5146484375 is standing or doing nothing:
Player on x:313.92767333984375 y:603.463623046875 is kicking:
Player on x:403.02960205078125 y:443.36376953125 is standing or doing nothing:
Player on x:1742.3193359375 y:816.229736328125 is running:
Player on x:949.35107421875 y:432.5810546875 is running:
Player on x:772.974853515625 y:448.326904296875 is standing or doing nothing:
Player on x:907.6070556640625 y:279.8189697265625 is running:
Player on x:1027.827392578125 y:348.4036865234375 is standing or doing nothing:
Player on x:351.14398193359375 y:607.9345703125 is running:
Player on x:700.5921630859375 y:287.96282958984375 is running:
Player on x:625.1990966796875 y:333.45172119140625 is standing or doing nothing:
Player on x:1004.2813720703125 y:296.603759765625 is standing or doing nothing:
Player on x:1226.171630859375 y:488.6516418457031 is running:
Player on x:810.2550048828125 y:694.925537109375 is running:
Player on x:974.1193237304688 y:344.4617919921875 is standing or doing nothing:
Player on x:1523.814453125 y:641.9794311523438 is standing or doing nothing:
Player on x:402.7860107421875 y:742.7982788085938 is running:
Player on x:1886.0013427734375 y:361.64404296875 is running:
Player on x:1886.074462890625 y:362.1553039550781 is running:
```



More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

F. Calculate actions for all players:

Hard Rules might include the following actions:

- Kicking: By analyzing the data of feet keypoints, as well as the keypoints of a player's arms and hands, along with the ball's position, we can make a simple estimation of a kicking action. Specifically, if the angle between the hip and feet keypoints is approximately a right angle, while one of the player's arms is bent and the ball remains close to the feet without being lifted, there is a strong likelihood of a kicking action taking place.
- Running: Running actions often result in minimal instances of double support, where both feet touch the ground simultaneously. By calculating the distance between knee keypoints and evaluating if they surpass a specific normalized threshold, we can effectively estimate a running action. Similar rules as those for kicking actions generally apply in this case. Additionally, if the ball is located outside the player's bounding box, it further reinforces the likelihood of a running action.
- Tackling: Tackling maneuvers encompass a wide range of techniques, including slide tackles, block tackles, and standing tackles, making them inherently complex to identify accurately through pose estimation alone. To mitigate this challenge, a more reliable approach involves utilizing ball possession and player detection techniques.
- Standing/Doing Nothing: If none of the above apply.

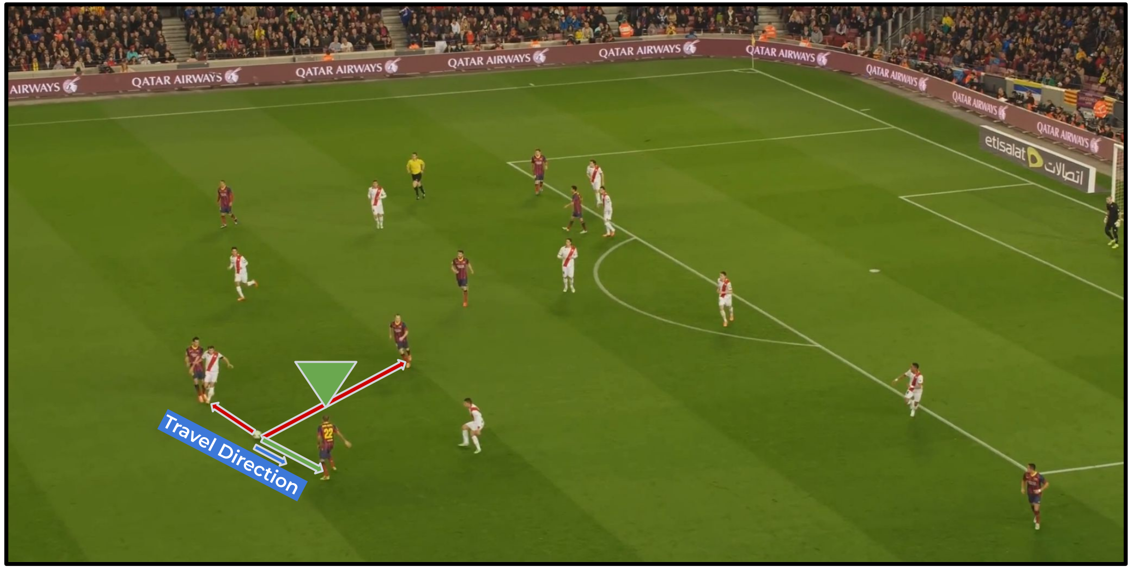
Event Predictions:

What are the different detectable events within our methods?

- Number of times **ball was touched** by a player.
- The number of **passes** a player has completed.
- **Risky maneuvers** which could lead to **injuries** such as ankle sprains or knee strains.
- **Ball possession percentage** of every player.
- The **number of tackles** completed without foul.
- The **number of ball catches** by goalkeepers.

Ball Control:

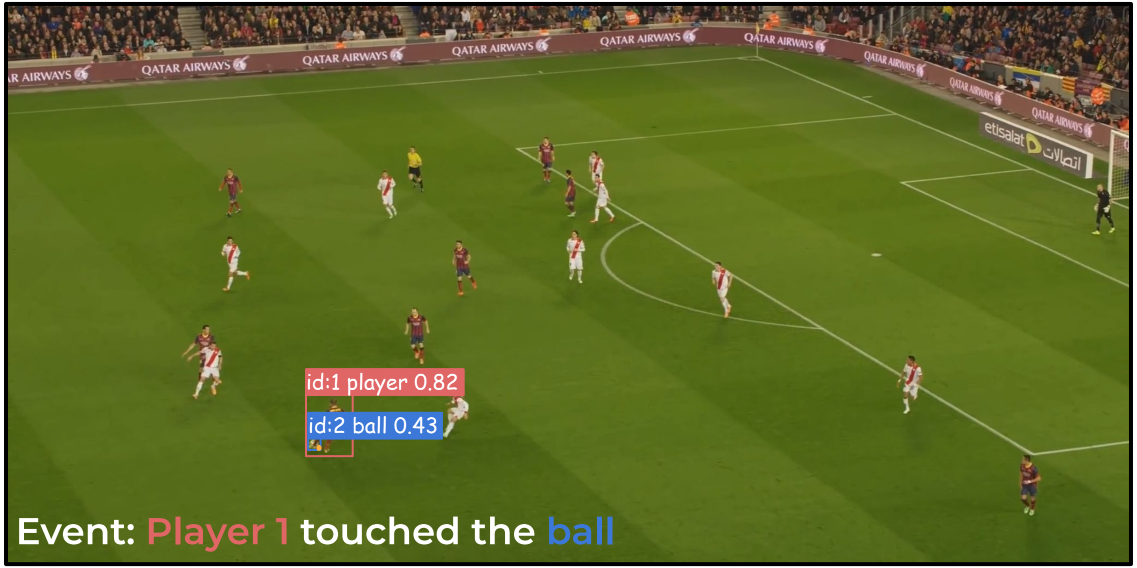
- By leveraging the distance between the ball and foot keypoints, along with the traveling direction of the moving ball, it becomes possible to determine which player has control over the ball.



FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Ball Touches:

- **Ball touches** serve as a significant summarization metric, and they can be obtained by assessing the **overlap between a player's bounding box and the ball**. To enhance this method, **foot keypoints** are incorporated, ensuring that they overlap with the ball as well.



FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Passes:

The process to calculate passes between two players requires four steps:

- Detect the ball within the bounding box of person with ID X. Pass is in "**started**" phase.
- Observe the ball moving away from person with ID X, control is lost from X. If the ball left the bounding box of the player the pass switches to an "**ongoing**" state.
- Observe the ball getting closer to person with ID Y, control is on Y. The pass is still in the "**ongoing**" state.
- Detect the ball within the bounding box of person with ID Y. The pass switches to a "**completed**" state.
- Log the event as: "**Player X completed a pass to Player Y**"

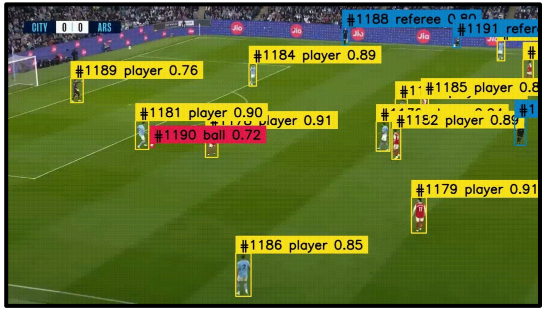
FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Knee Injury:

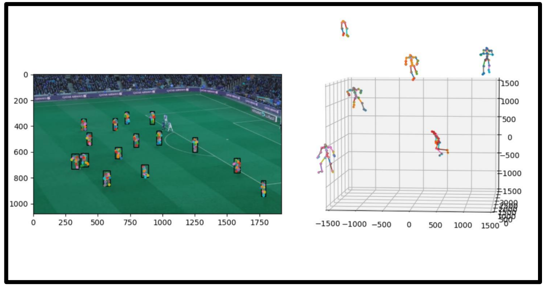
- When the foot keypoint is in close proximity to the **knee keypoint, the probability of injury is higher, ranging between 0.9-1.** This indicates a situation where the foot is positioned in a **potentially risky manner** relative to the **knee joint**.
- Conversely, if the **foot keypoint** is lower in relation to the knee, the probability falls within the range of 0.5-0.8. This suggests a **less critical positioning** but still poses some risk.
- When the **foot keypoint** is located above the knee, the probability decreases further, ranging between 0.1-0.4. This indicates a **lower risk scenario** where the foot is positioned in a manner that is less likely to result in an injury.

Experiments:

- ByteTrack - [Yifu Zhang, Peize Sun]:
  - Faster Inference. (0.05s on an A100 GPU)
  - Less Accurate in occluded areas.
  - Offset between object centers and bounding boxes.
  - *bytetrack_x_mot20* model. (MOTA: 93.4 , FPS: 17.5)



- MeTRAbs - [István Sárándi, Timm Linder]:
  - 3D body joint estimation.
  - Slower Inference. (0.05 seconds)
  - Better suited for VR and HC interactions.
  - Does not work well for a lot of objects.
  - 17 keypoint model.

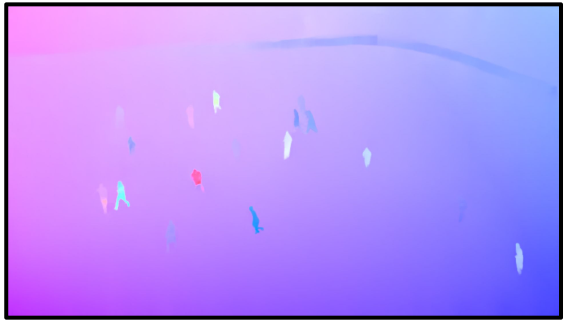


More on:

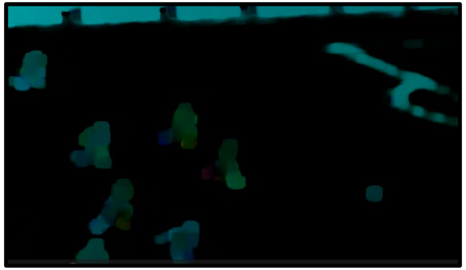
FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Experiments:

- UniMatch - Haofei Xu, Jing Zhang:
  - Unifies Optical Flow, Rectified Stereo matching and Stereo Depth Estimation.
  - Transformer-based model.
  - Really slow inference speeds. (0.1 sec on A100 GPU)
  - Using largest models. (~22.1 M params)
  - Hard to segment out the ball from results.



- Dense Optical Flow - OpenCV:
  - Gunnar Farneback's algorithm. (2003)
  - Tracking the movement of pixels.
  - Fast Inference. (0.005 sec on A100 GPU)
  - Not suited for many occlusions.
  - Hard to segment out the ball from results.

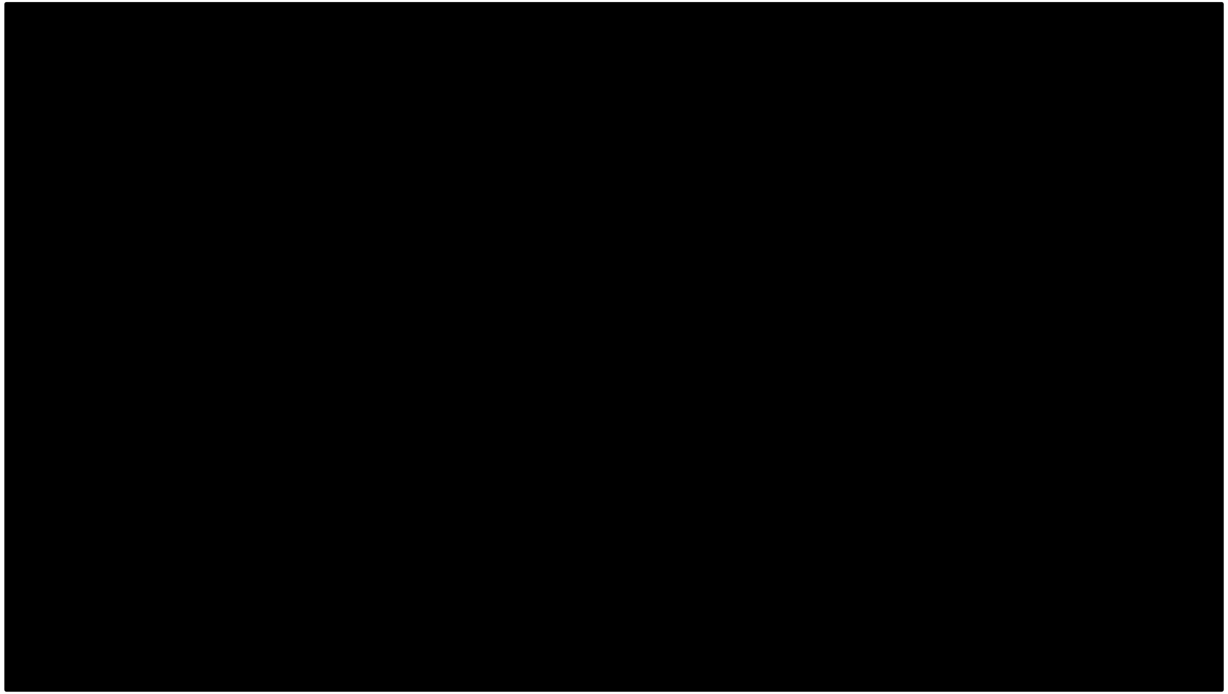


More on:

FOOTBALL INSIGHTS

VIDEO ANALYSIS USING DEEP LEARNING

Results:



PARAMS:

YOLOv8: {model: yolov8x, imgsz: 1920, epoch: 65, batch_size: 2, conf: 0.55}

DeepSort: {conf: 0.51, inertia: 0.39, iou_thresh: 0.22, max_age: 50, min_hits: 1, delta_t: 1}

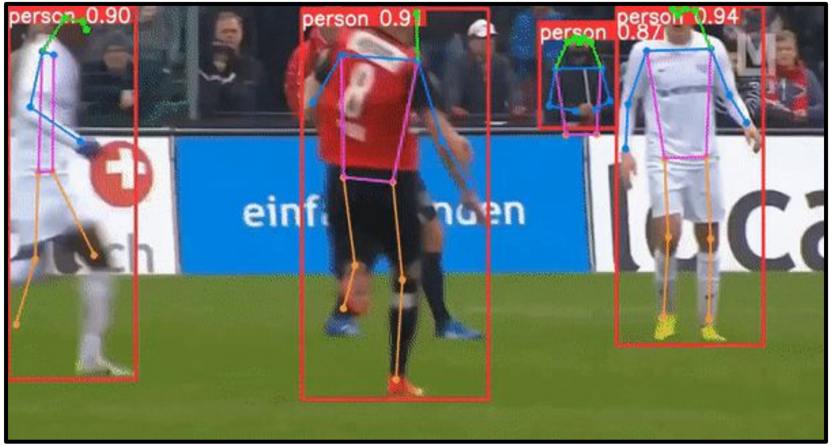
YOLOv8-Pose: {model: yolov8x-pose-p6, conf: 0.01, imgsz: 1920

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Results:

Pass Detection

Knee Injury Detection



More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

Improvement Ideas:

TRACKING

Reserve ID's according to number of players.

Reidentify players by current missing ID's for improved tracking.

Rewrite code to be compatible with newest YOLOv8 models.

The ball usually remains the same, so no re-identification needed.

POSE + ACTION PREDICTION

Train a custom model.

Use XGBoost or Random Forest for predictions.

Use zoom cameras for accurate pose estimation.

Train a custom CNN for pose ROIs.

EXTRAS

Crowd Elimination.

Planar Homography.

Jersey number prediction on a custom model.

Future Work:

- **Integration with Player Tracking Systems:** Integrating the injury prediction model with player tracking systems, such as GPS or wearable sensors, would enhance the accuracy of injury probability assessments by considering additional factors like player speed, direction, and biomechanics.

- **Collaboration with Sports Science Experts:** Collaborating with sports science experts, coaches, and medical professionals could provide valuable insights and domain-specific knowledge to further refine and validate the injury prediction model and statistical analysis methods.

- **Collaboration with a Camera Crew and Technical Team:** Engaging in collaboration with camera experts to establish a comprehensive system tailored specifically for the vision model would exponentially expand our capabilities in summarizing football matches. By constructing a composite vision system, we would entirely change the way the game is perceived.

Conclusion:

Despite the current state-of-the-art tools, analyzing football matches remains computationally intensive and intricate. However, by harnessing the capabilities of task-specific models in tandem, we have demonstrated the ability to accomplish a remarkably comprehensive analysis.

With the increasing popularity of football, the number of ongoing matches has surpassed manual analysis capabilities. To address this, leveraging tools like Football Insights can serve as a valuable resource, enabling the automation of analysis processes. This not only expedites the analysis but also provides interactive platforms for fans to engage with their favorite teams, players, or matches.

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING