
SCHÜCOARENA
WWW.UNSER-GESUNDHEITSPARTNER.DE
FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING
Takács Tamás and Borsy Máté

Project Scope:

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

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

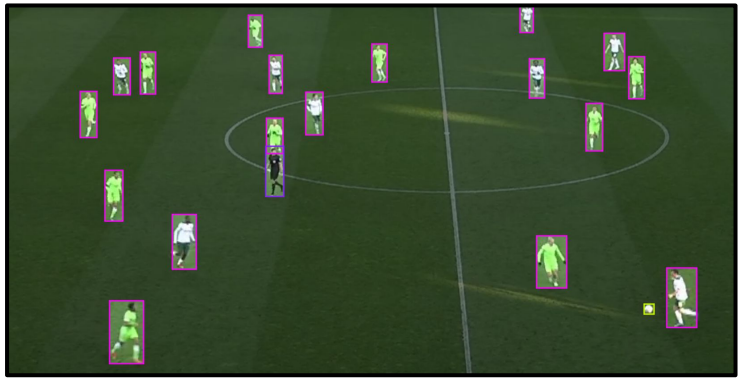
C. **Track the players** by ID throughout input frames.

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

E. **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

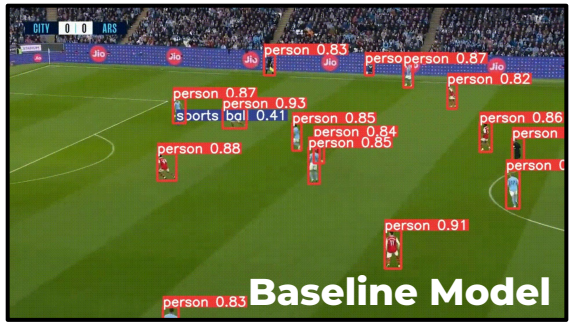
FCB - SFC: (2015)

- Used for Inference
- 1080p Resolution

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

A. Detect all football players on input frames:

- YOLOv7 - Wong Kin Yiu, AlexeyAB:
  - Lacks pose estimation.
  - Larger models are slower.
  - Lower mAP@[.5:.95]. (52.9 for *yolov7-x* vs 53.9 for *yolov8-x*)
- YOLOv8 - Ultralytics:
  - Used for training a custom object detection model on the **Roboflow** dataset.
  - Training Parameters: {epochs: 65, batch_size: 2, image_size: 1920}
  - Results: {prec: 0.927, rec: 0.859, mAP50: 0.902, mAP50-95: 0.678, ball_det: 0.66}



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)

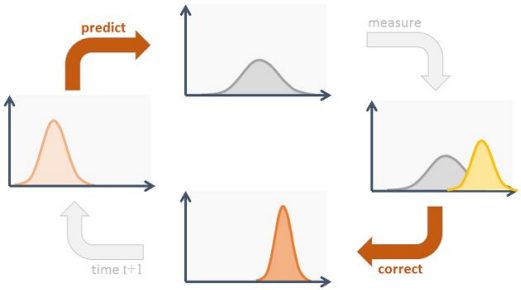
- XMem - Ho Kei Cheng, Alexander Schwing:
  - Tracks segmented objects of an image throughout input frames.
  - Fast Inference (0.05 seconds on an NVIDIA A100 GPU)
  - XMem default model (largest model ~ 264MB)
  - Used for tracking the segmented ball.
  - No labels.

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.



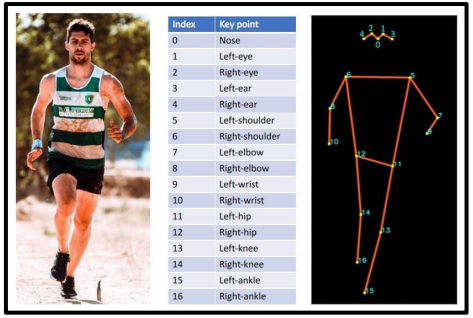
More on:

FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

D. 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:

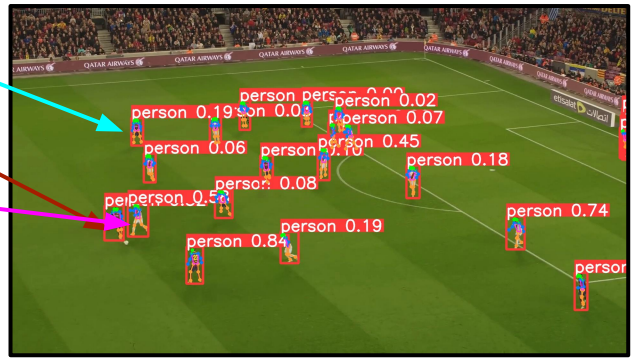


FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

D. 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:
```

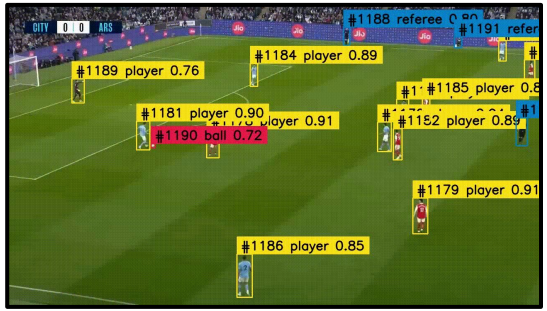


More on:

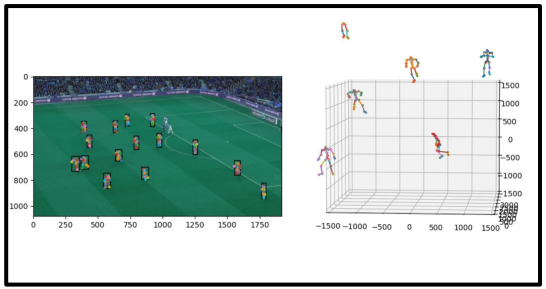
FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

The Road Less Traveled - Failed Attempts:

- 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.

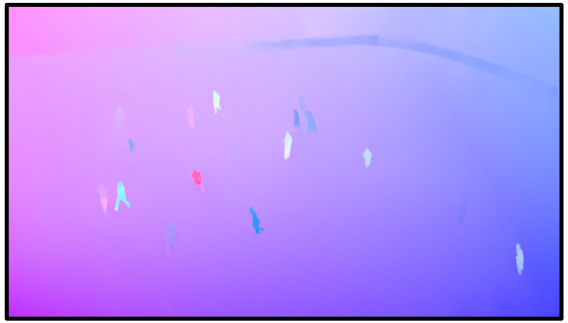


More on:

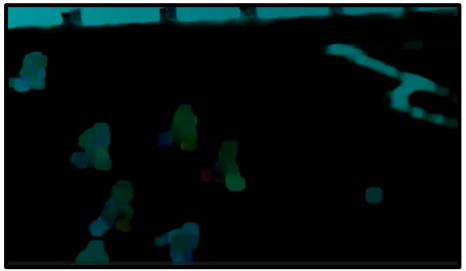
FOOTBALL INSIGHTS
VIDEO ANALYSIS USING DEEP LEARNING

The Road Less Traveled - Failed Attempts:

- 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

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.

Jersey number prediction on a custom model.

Planar Homography.

FOOTBALL INSIGHTS

VIDEO ANALYSIS USING DEEP LEARNING

End Result:

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