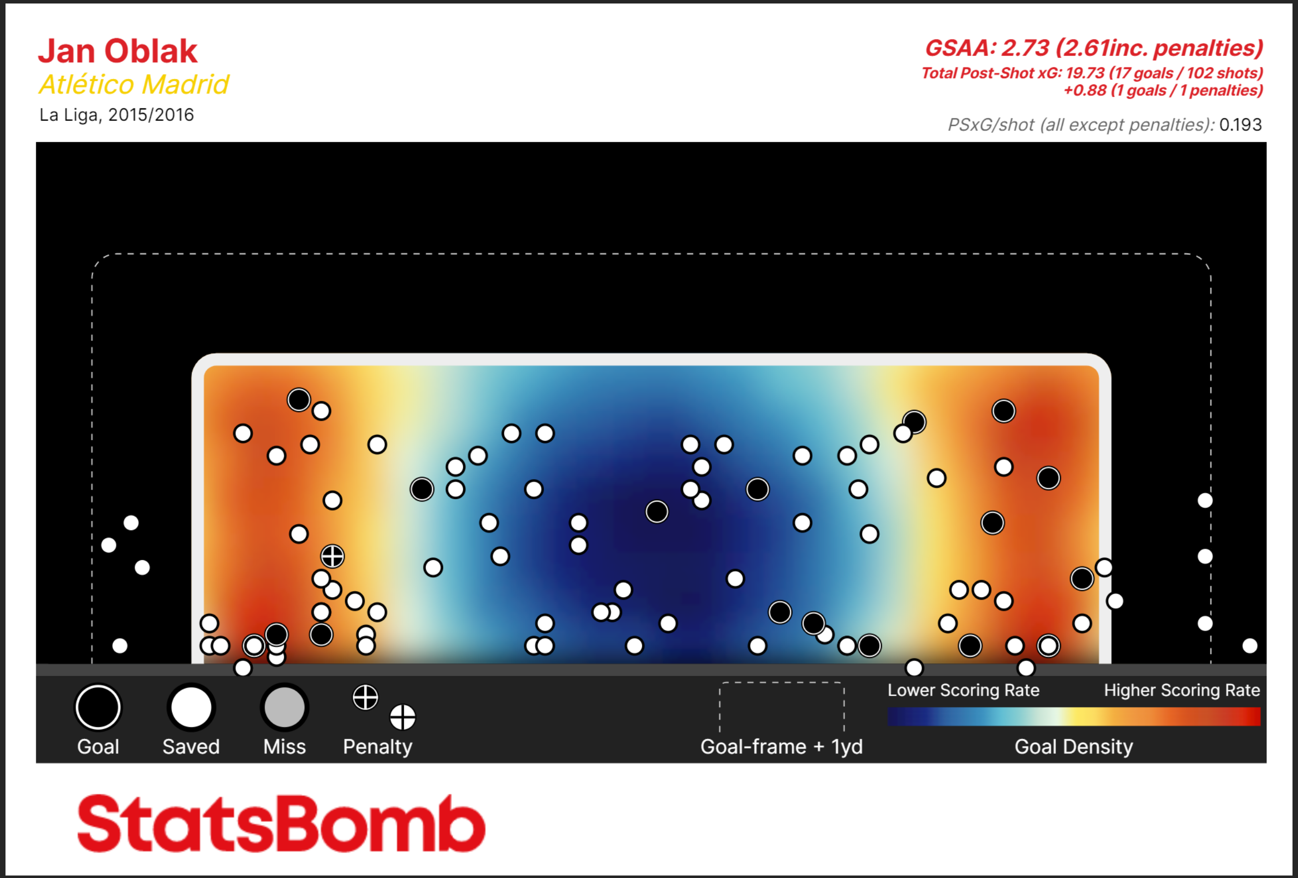
LEVELING UP COLLEGE SOCCER WITH MACHINE LEARNING

SENIOR PROJECT

Alex Fok

BACKGROUND

DATA IS THE GAME

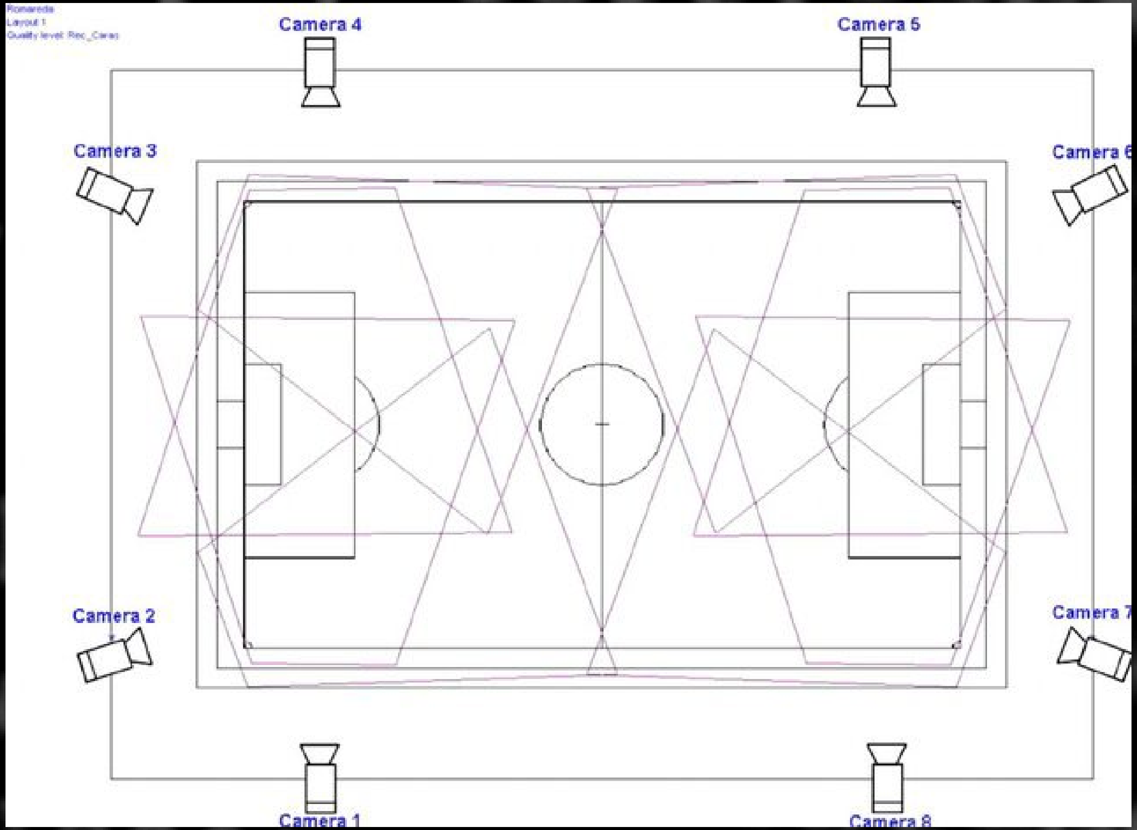


Source: STATSBOMB

Where does data come from?

GPS TRACKER

CAMERA FOOTAGE

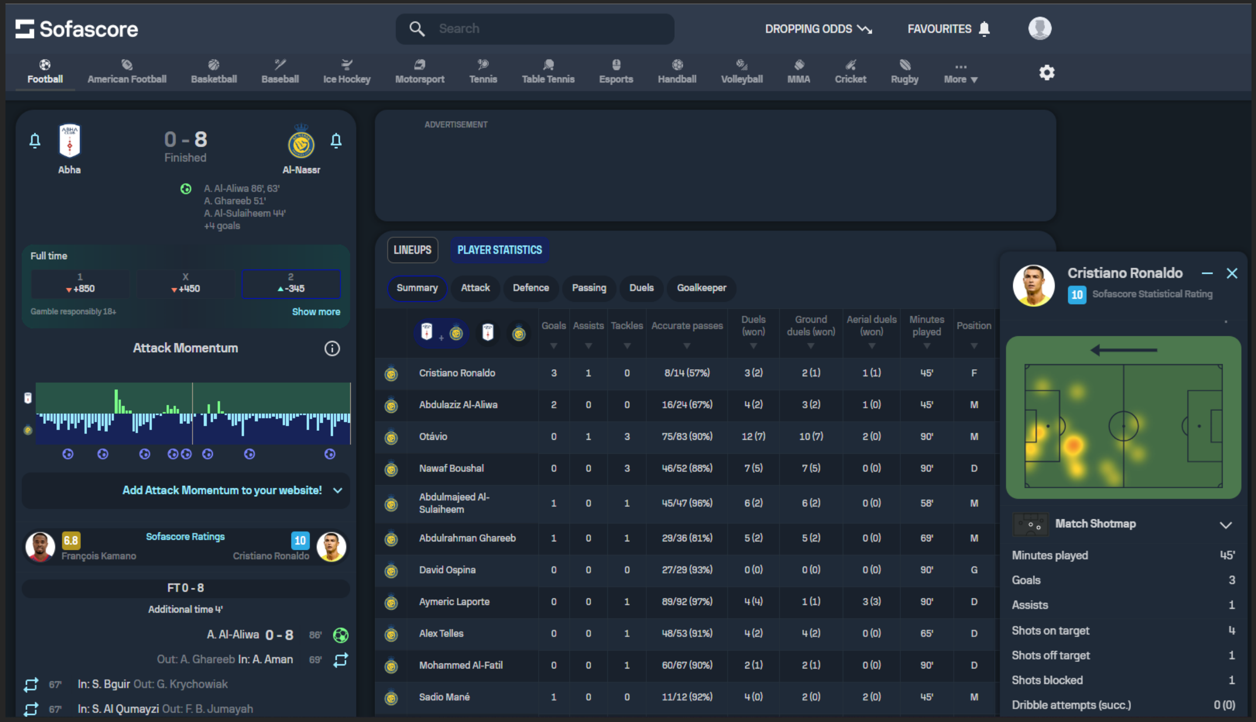


Source: STATSports, Martinez-del-Rincon, Jesus & Herrero Jaraba, Elías & Gómez, J. & Orrite, Carlos & Medrano, Carlos & Montañés-Laborda, Miguel. (2009). Multicamera sport player tracking with Bayesian estimation of measurements. Optical Engineering - OPT ENG. 48. 10.1117/1.3114605.

WHAT ABOUT NON-PROFESSIONALS?

Source: Veo

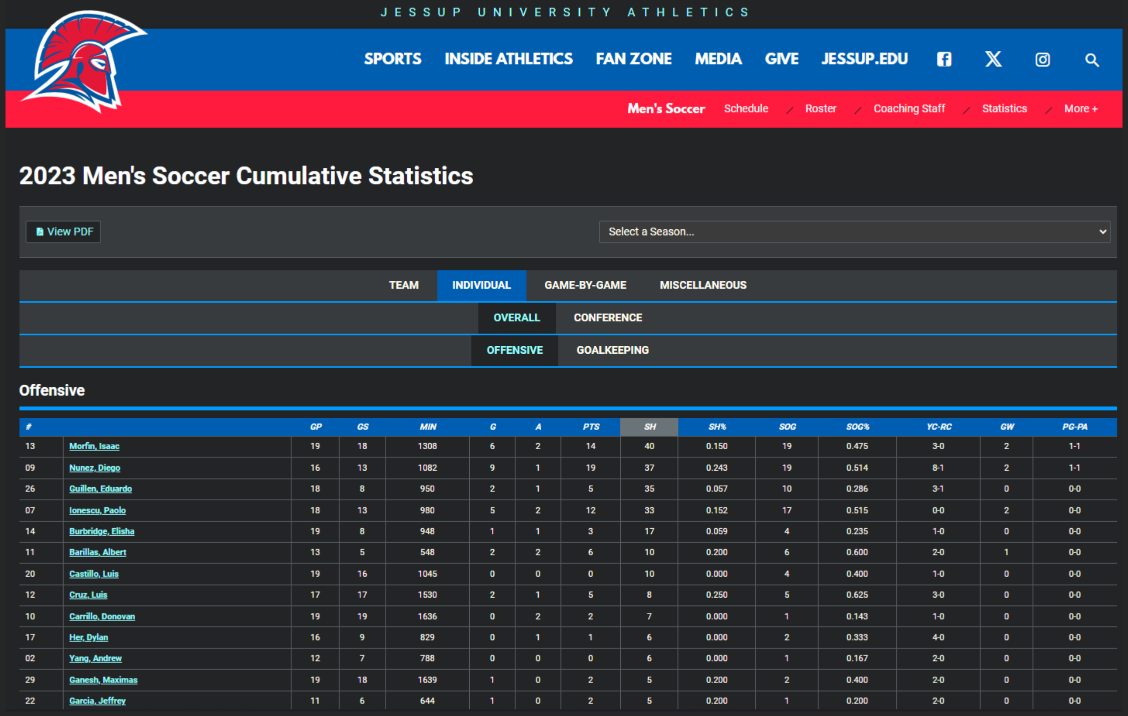
STATISTICS AVAILABLE



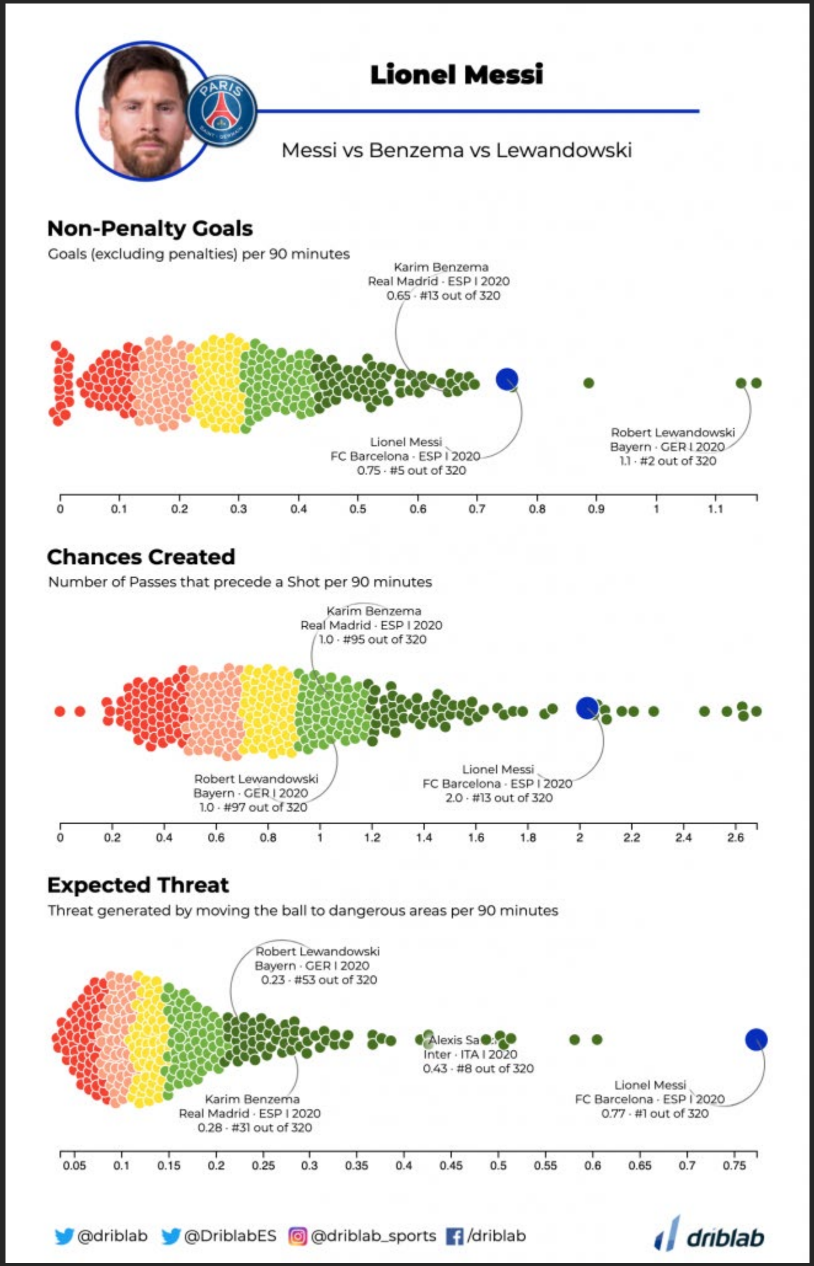
	Goals	Assists	Tackles	Accurate passes	Duels (won)	Ground duels (won)	Aerial duels (won)	Minutes played	Position
Cristiano Ronaldo	3	1	0	8/14 (57%)	3 (2)	2 (1)	1 (1)	45'	F
Abdulaziz Al-Aliwa	2	0	0	16/24 (67%)	4 (2)	3 (2)	1 (0)	45'	M
Otávio	0	1	3	75/83 (90%)	12 (7)	10 (7)	2 (0)	90'	M
Nawaf Boushal	0	0	3	46/52 (88%)	7 (5)	7 (5)	0 (0)	90'	D
Abdulmajeed Al-Sulaiheem	1	0	1	45/47 (96%)	6 (2)	6 (2)	0 (0)	58'	M
Abdulrahman Ghareeb	1	0	1	29/36 (81%)	5 (2)	5 (2)	0 (0)	69'	M
David Ospina	0	0	0	27/29 (93%)	0 (0)	0 (0)	0 (0)	90'	G
Aymeric Laporte	0	0	1	89/92 (97%)	4 (4)	1 (1)	3 (3)	90'	D
Alex Telles	0	0	1	48/53 (91%)	4 (2)	4 (2)	0 (0)	65'	D
Mohammed Al-Fatil	0	0	1	60/67 (90%)	2 (1)	2 (1)	0 (0)	90'	D
Sadio Mané	1	0	0	11/12 (92%)	4 (0)	2 (0)	2 (0)	45'	M

	Touches	Accurate passes	Key passes	Crosses (acc.)	Long balls (acc.)	Notes	Position	Sofascore rating
Cristiano Ronaldo	22	8/14 (57%)	1	0 (0)	0 (0)	Big chances created: 1	F	10
Abdulaziz Al-Aliwa	31	16/24 (67%)	1	3 (0)	1 (0)	Big chances created: 1	M	8.9

	Duels (won)	Ground duels (won)	Aerial duels (won)	Possession lost	Fouls	Was fouled	Offsides	Position	Sofascore rating
Cristiano Ronaldo	3 (2)	2 (1)	1 (1)	7	0	1	1	F	10
Abdulaziz Al-Aliwa	4 (2)	3 (2)	1 (0)	11	1	1	0	M	8.9
Otávio	12 (7)	10 (7)	2 (0)	11	0	4	0	M	8.6



#		GP	GS	MIN	G	A	PTS	SH	SH%	SOG	SOG%	YC-RC	GW	PG-PA
13	Morfin, Isaac	19	18	1308	6	2	14	40	0.150	19	0.475	3-0	2	1-1
09	Nunez, Diego	16	13	1082	9	1	19	37	0.243	19	0.514	8-1	2	1-1
26	Guillen, Eduardo	18	8	950	2	1	5	35	0.057	10	0.286	3-1	0	0-0
07	Ionescu, Paolo	18	13	980	5	2	12	33	0.152	17	0.515	0-0	2	0-0
14	Burbridge, Elisha	19	8	948	1	1	3	17	0.059	4	0.235	1-0	0	0-0
11	Barillas, Albert	13	5	548	2	2	6	10	0.200	6	0.600	2-0	1	0-0
20	Castillo, Luis	19	16	1045	0	0	0	10	0.000	4	0.400	1-0	0	0-0
12	Cruz, Luis	17	17	1530	2	1	5	8	0.250	5	0.625	3-0	0	0-0
10	Carrillo, Donovan	19	19	1636	0	2	2	7	0.000	1	0.143	1-0	0	0-0
17	Her, Dylan	16	9	829	0	1	1	6	0.000	2	0.333	4-0	0	0-0
02	Yang, Andrew	12	7	788	0	0	0	6	0.000	1	0.167	2-0	0	0-0
29	Ganesh, Maximas	19	18	1639	1	0	2	5	0.200	2	0.400	2-0	0	0-0
22	Garcia, Jeffrey	11	6	644	1	0	2	5	0.200	1	0.200	2-0	0	0-0



Stats of top 3 Ballon d'Or 2021 contenders until end of voting

Players	Messi	Lewandowski	Benzema
Games	49	45	50
Goals	40	54	35
Assists	16	8	14
Shots on target	103	98	87
Big chances missed	12	35	24
Accurate passes	2525	715	1589
Dribbles completed	224	48	60
Chances created	118	55	75
Big chances created	26	16	14
Tackles	31	12	23
Recoveries	107	98	101
Points won by goals	22	20	19
Points won by G/A	39	25	30
Match winning goals	9	8	6
Match winning G/A	15	10	10
Man of the match	27	13	11
Team trophies	Copa del Rey Copa America	Bundesliga Club World Cup	UEFA Nations League
Individual awards	LaLiga top scorer Copa America top scorer Copa America top assister Copa America POTT	Bundesliga top scorer European Golden Shoe CWC Golden Ball	
Records broken	Most NT goals in CONMEBOL Most goals in a single club	Most goals in a single BuLi season	

Stats from January 1, 2021 until October 24, 2021 (Ballon d'Or 2021 voting deadline day)

BEST PLAYER DEBATE

Image: Tribuna

Position		Club	Played	Won	Drawn	Lost	GF	GA	GD	Points
1	▲	Arsenal	33	23	5	5	77	26	51	74
2	▲	Liverpool	33	22	8	3	75	32	43	74
3	▼	Manchester City	32	22	7	3	76	32	44	73
4	•	Aston Villa	34	20	6	8	71	50	21	66
5	•	Tottenham Hotspur	32	18	6	8	65	49	16	60
6	•	Newcastle United	32	15	5	12	69	52	17	50
7	•	Manchester United	32	15	5	12	47	48	-1	50
8	•	West Ham United	34	13	9	12	54	63	-9	48
9	•	Chelsea	31	13	8	10	61	52	9	47
10	•	Brighton And Hove Albion	32	11	11	10	52	50	2	44

DATA ACCURATE AS OF 4/21/2024

"Data is important in football, but at the end of the day I prefer my own eyes." – Ten Hag

"I use data a lot to understand certain concepts of the game, and I try to amplify performance." – Arteta

SOURCE: X, GOAL.COM, PREMIER LEAGUE, GETTYIMAGES

SOURCE: LIVERPOOL FC

YOU'LL NEVER WALK ALONE
LIVERPOOL
FOOTBALL CLUB
EST·1892

Google DeepMind

TACTIC AI

4 YEARS OF DEVELOPMENT

Only about 10 corner kicks are played in each match in the Premier League every season.

Sealed a spot to the Final and won the club’s 6th European Cup in 2019

GETTING DATA

BE ONE STEP AHEAD

Track player locations from single camera footage

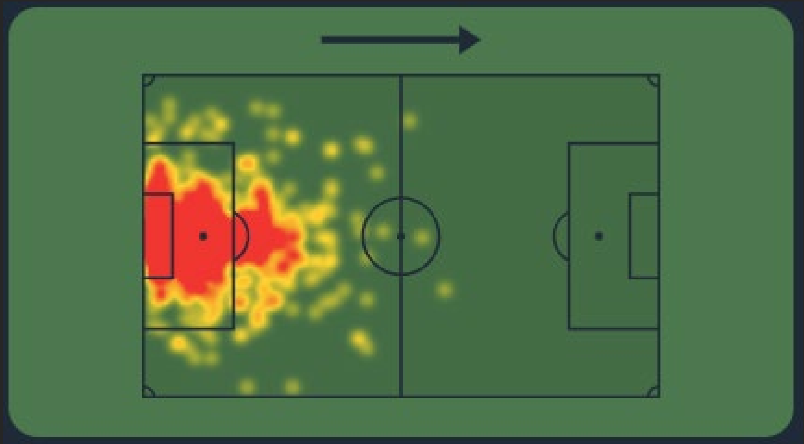
Generate Heat Maps

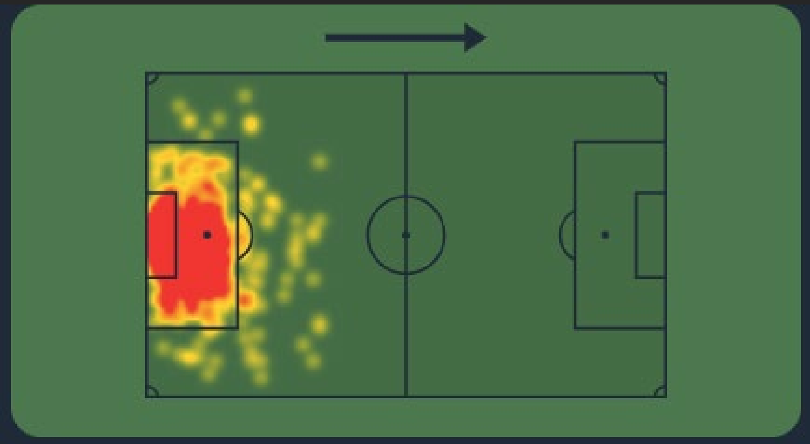
Perform Data Analysis

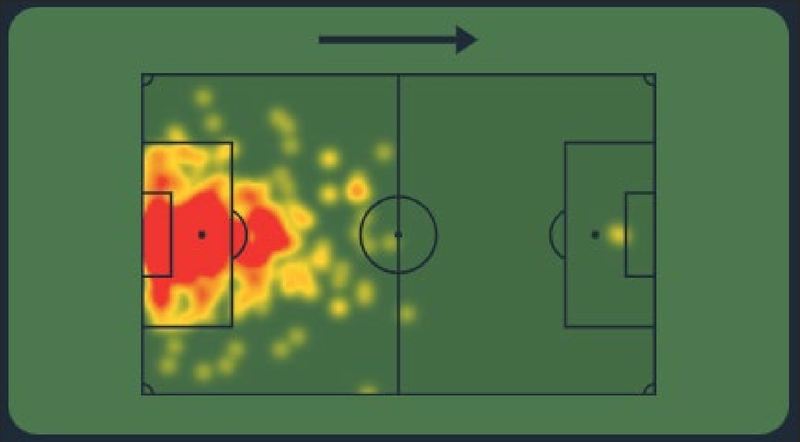
Plot Average Position

PROJECT GOALS

HEAT MAPS







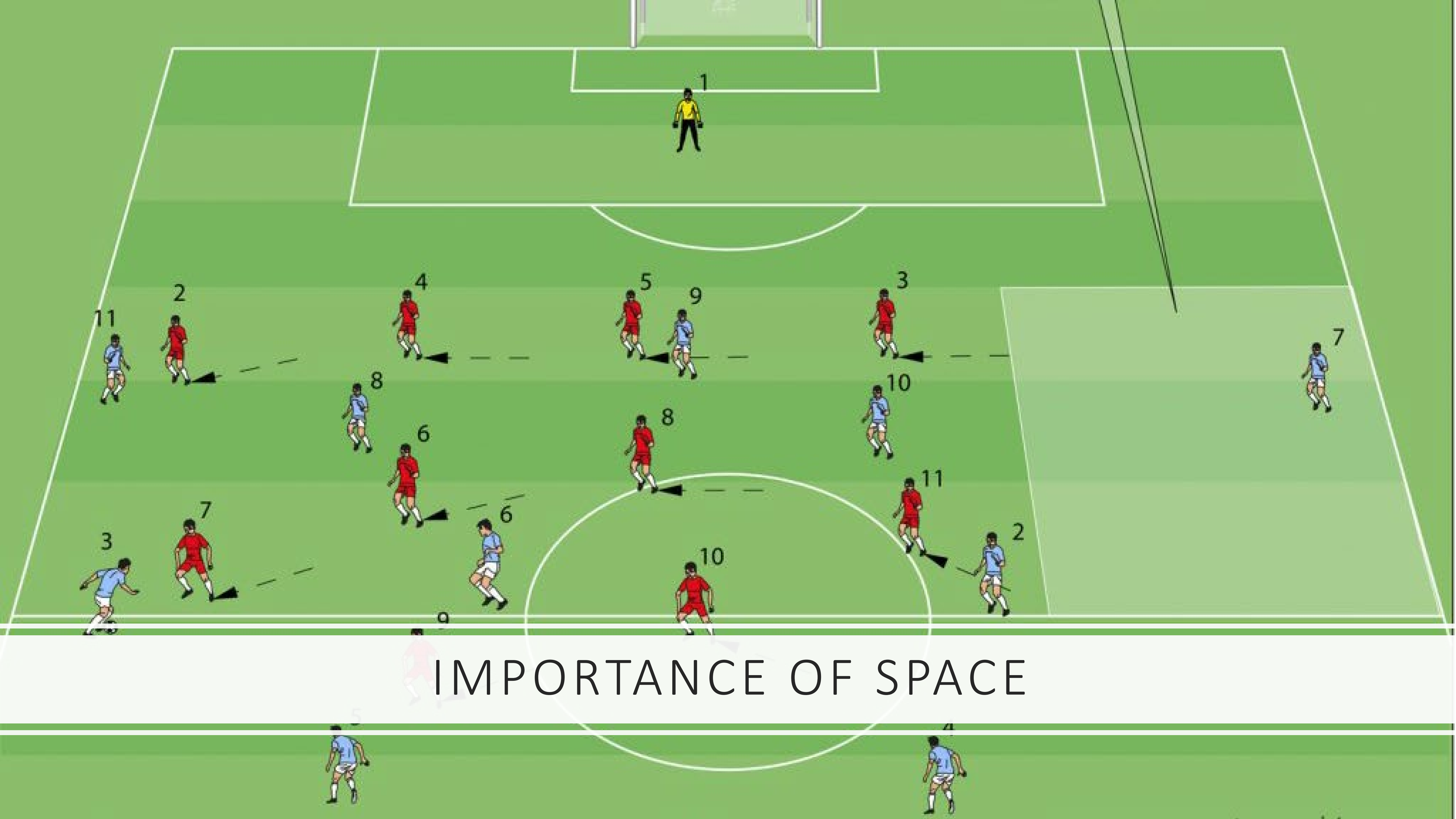
Andre Onana

Jan Oblak

Manuel Neuer

Source: SofaScore Data as of 4/1/2024

Images: sortitoutsi.net



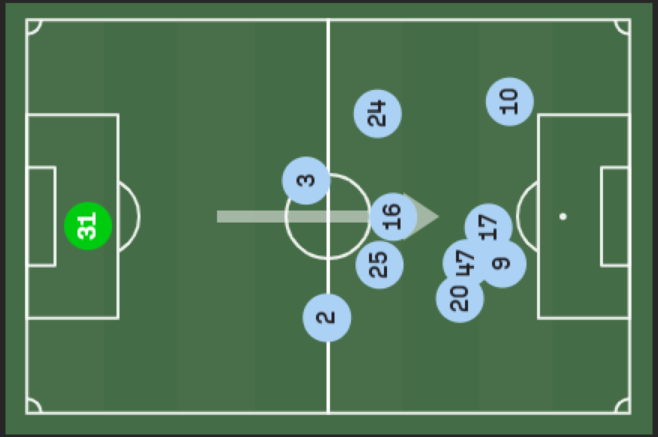
IMPORTANCE OF SPACE

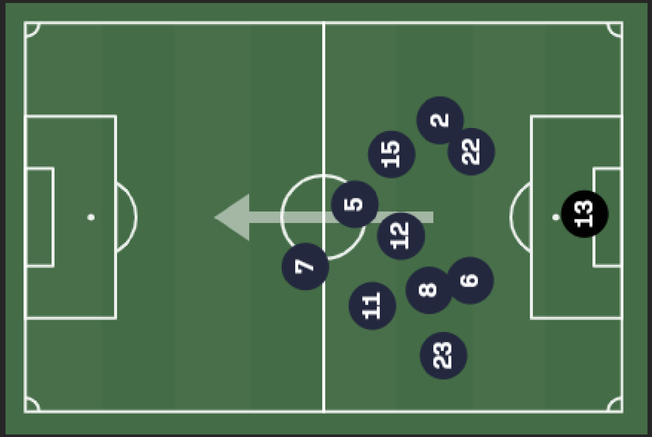
TACTICS - POSITION SHIFTING

UEFA Champions League,
Knockout stage,
Quarterfinal Second Leg

4-17-2024

Manchester City vs Real Madrid





Source: Sofascore

PROJECT SETUP

Programming Language: Python

Object Detection and Classification Model: Custom YOLOv8 model

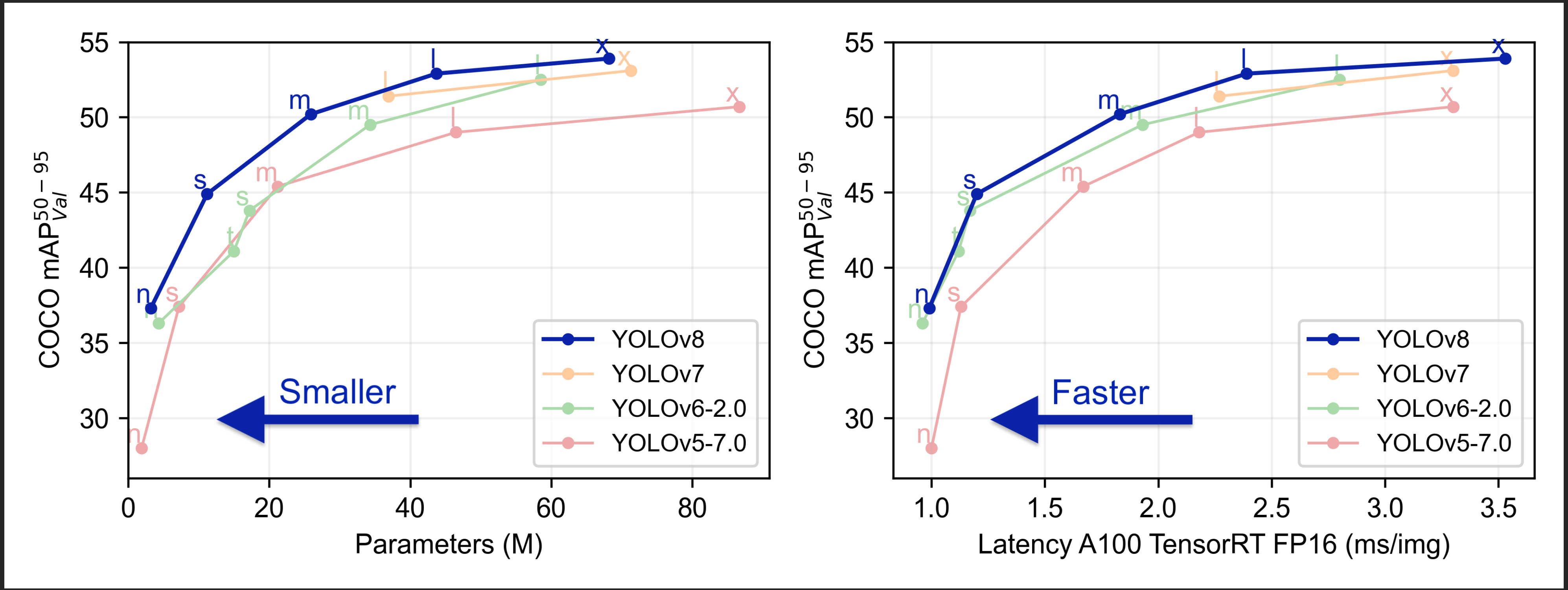
Machine Learning Framework: TensorFlow

Computer Vision Library: OpenCV

Graphing Library: Matplotlib, PyGWalker

Front-end: Streamlit

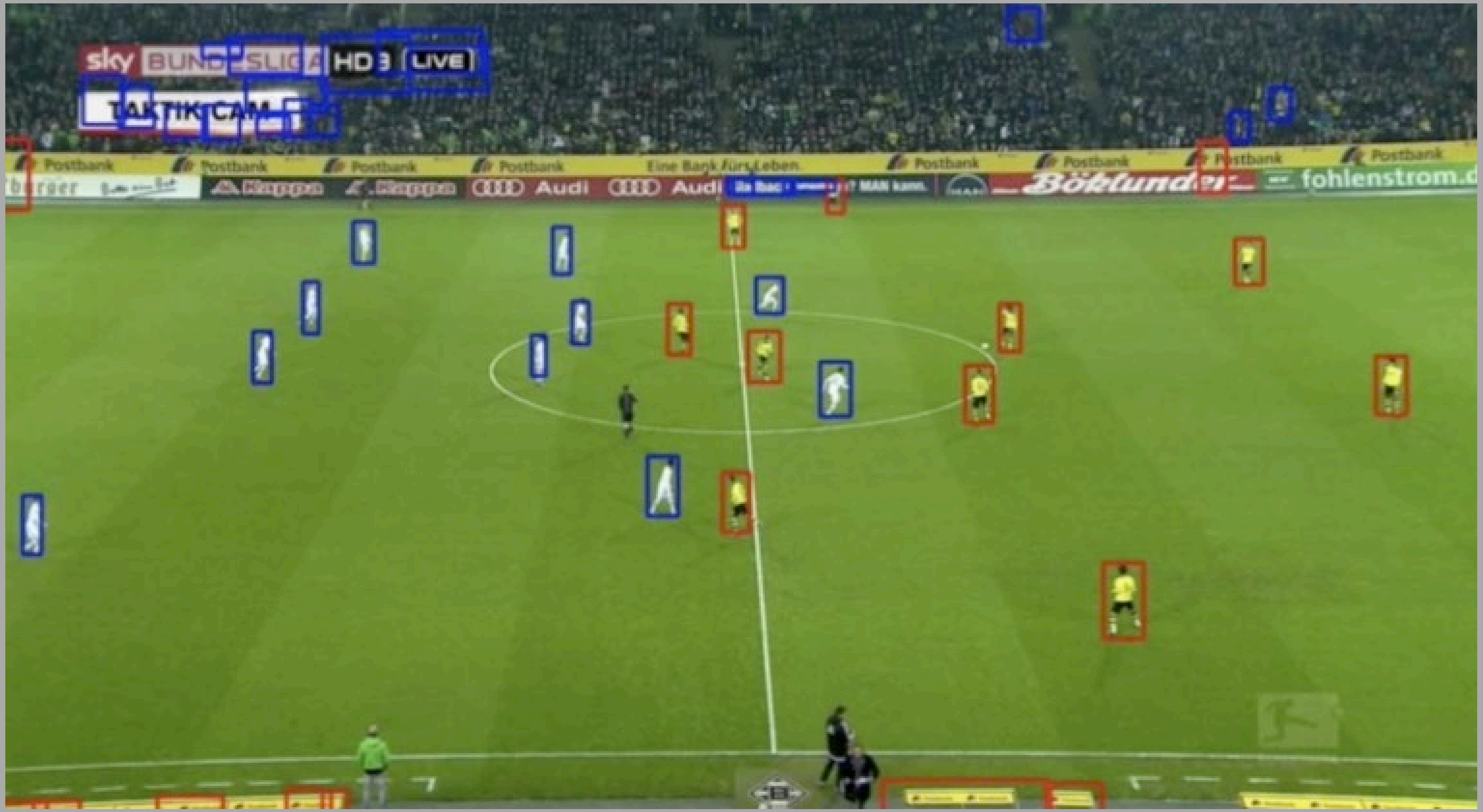
DETECTING PLAYERS



Open-Source Computer Vision Model

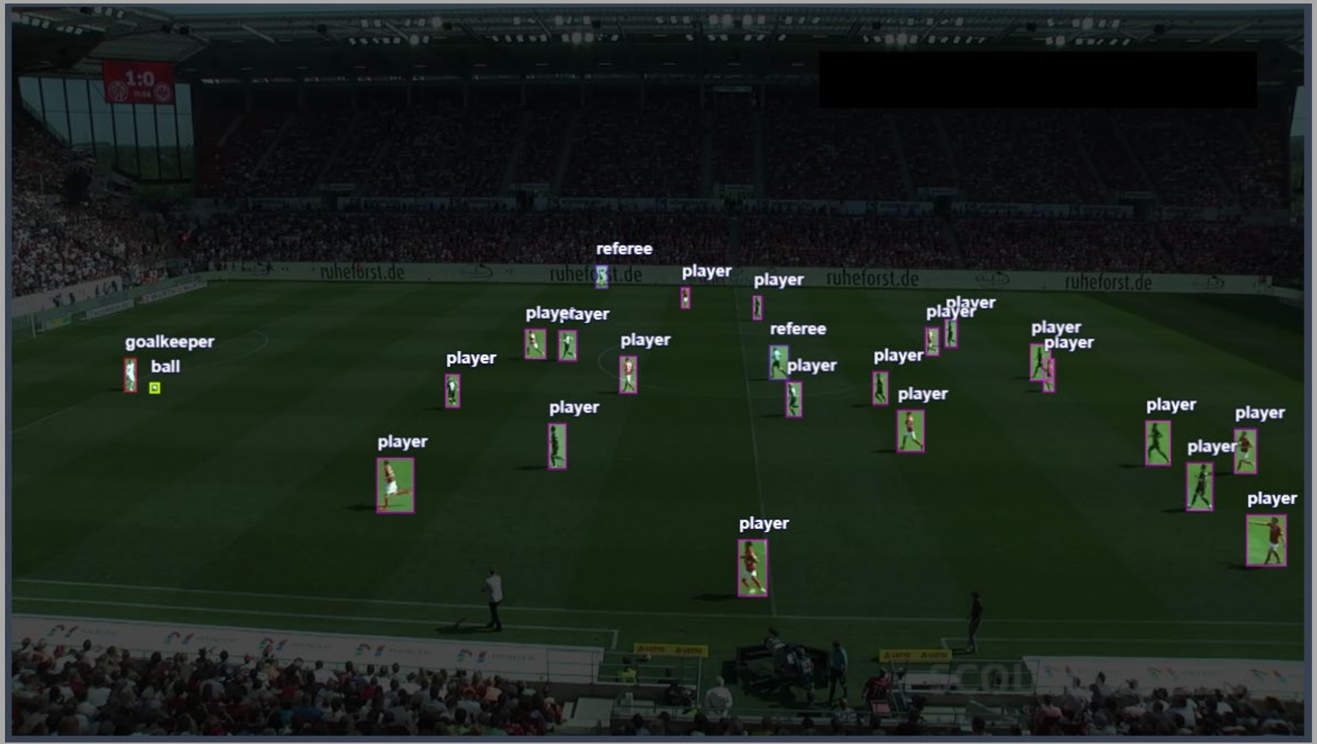
Convolutional Neural Network (CNN)

OBJECT DETECTION



MESSY

Image Source: Manuel Stein at FC Barcelona Sports Tomorrow



TRAINING A MODEL WITH CUSTOM DATA

- Limiting detections to 4 classes
- Free Dataset from Roboflow
- Trained on ~200 images

TRAINING A MODEL WITH CUSTOM DATA

epochs	Total number of training epochs. Each epoch represents a full pass over the entire dataset. Adjusting this value can affect training duration and model performance.
patience	Number of epochs to wait without improvement in validation metrics before early stopping the training. Helps prevent overfitting by stopping training when performance plateaus.

Source: Ultralytics

- Training Parameters:
  - Epochs: 50
  - Patience: 5

TRAINING A MODEL

Cloud Setup

32GB RAM

Tesla V100 GPU w/16GB VRAM

~$1500

My laptop

16GB RAM

MX250 w/ 2GB VRAM

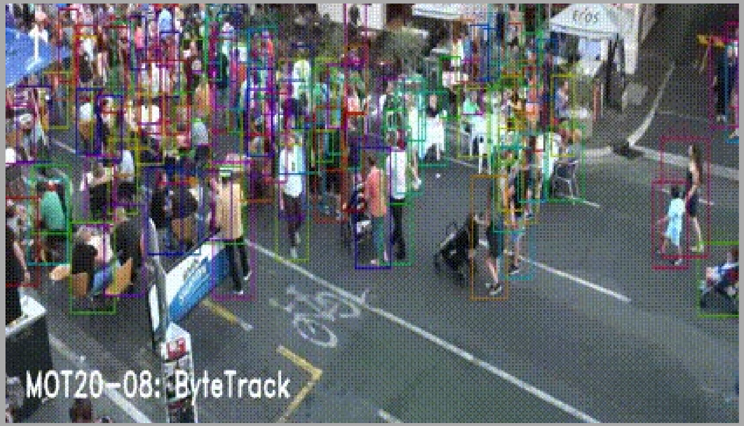
BEFORE AND AFTER

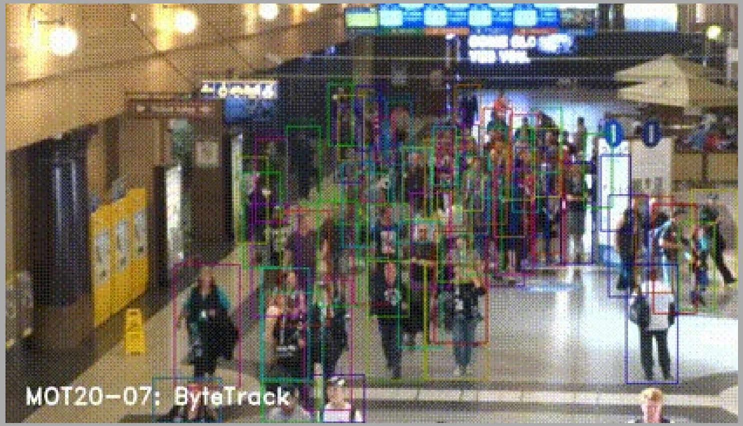
```
Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
 1/50      3.57G      1.901      4.184     0.9688        499        640: 100%|██████████| 13/13 [00:13<00:00,  1.00s/it]
               Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 2/2 [00:00<00:00,  4.00it/s]
```

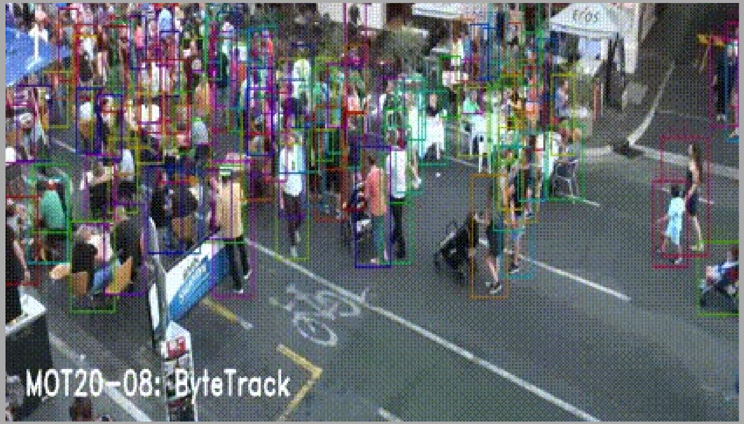
```
      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
      44/50      2.45G      1.119     0.7021     0.8158        284        640: 100%|██████████| 13/13 [00:01<00:00,  9.86it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 2/2 [00:00<00:00,  9.71it/s]
Stopping training early as no improvement observed in last 5 epochs. Best results observed at epoch 39, best model saved as best.pt.
To update EarlyStopping(patience=5) pass a new patience value, i.e. `patience=300` or use `patience=0` to disable EarlyStopping.
```

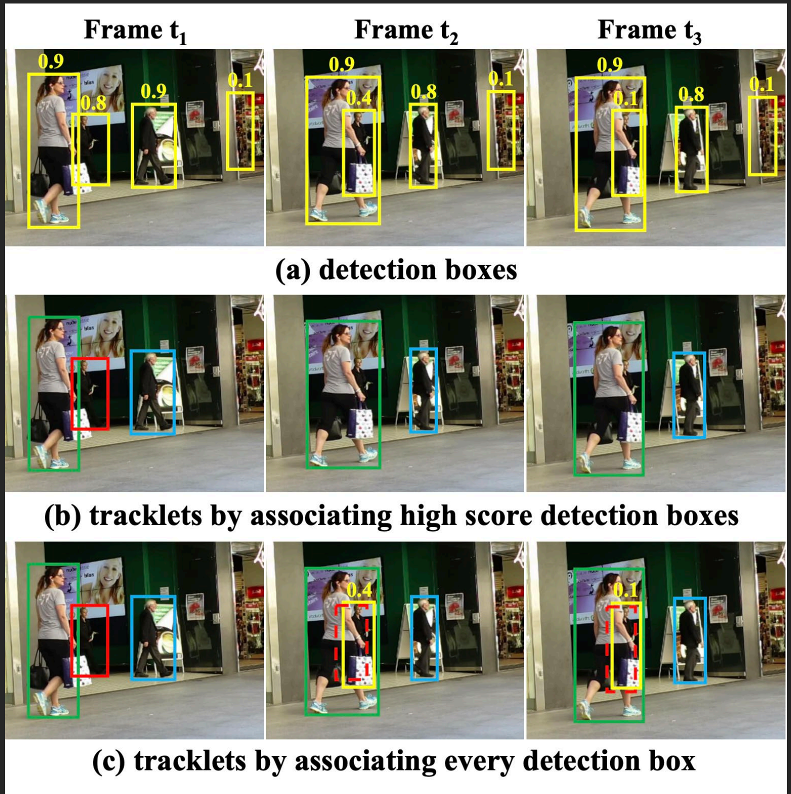
~2 minutes in training time

PROBLEMS WITH MULTIPLE OBJECT DETECTION





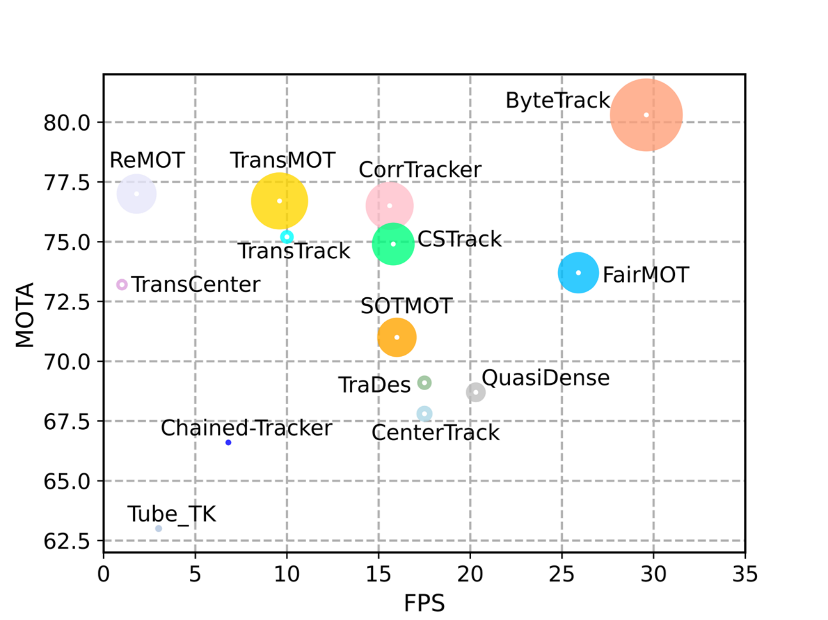




OBJECT REIDENTIFICATION

ByteTrack

Multi-object tracking (MOT)



144 tracking IDs → 97 tracking IDs

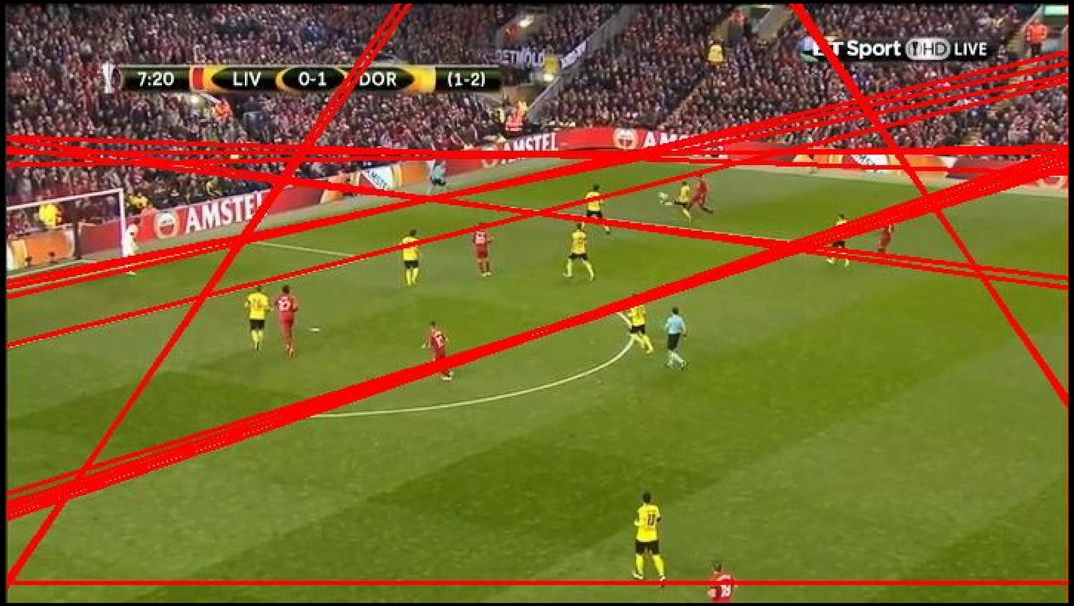
HOMOGRAPHY TRANSFORMATION

3D TO 2D

IMAGE WRAPPING

$$
\begin{array}{cccc}
 & \textbf{2N x 8} & \textbf{8 x 1} & & \textbf{2N x 1} \\
\begin{array}{c}\textbf{Point 1}\\ \\ \textbf{Point 2}\\ \\ \textbf{Point 3}\\ \\ \textbf{Point 4}\\ \\ \end{array} &
\begin{bmatrix}
x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x'_1 & -y_1x'_1 \\
0 & 0 & 0 & x_1 & y_1 & 1 & -x_1y'_1 & -y_1y'_1 \\
x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2x'_2 & -y_2x'_2 \\
0 & 0 & 0 & x_2 & y_2 & 1 & -x_2y'_2 & -y_2y'_2 \\
x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3x'_3 & -y_3x'_3 \\
0 & 0 & 0 & x_3 & y_3 & 1 & -x_3y'_3 & -y_3y'_3 \\
x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4x'_4 & -y_4x'_4 \\
0 & 0 & 0 & x_4 & y_4 & 1 & -x_4y'_4 & -y_4y'_4
\end{bmatrix} &
\begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} & = &
\begin{bmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{bmatrix} \\
\textbf{additional points} & \vdots & & & \vdots
\end{array}
$$

```
pred_homo = get_perspective_transform(src,dst)
temppoints= np.hstack(track).reshape(-1,1,2).astype(np.float32)[0][0]
track_transformed = track_transformed_history[track_id]
transformed_points = utils.scale_points(temppoints,(st_env.size),(st_env.kp_model_shape))
transformed_points = utils.warp_points([[transformed_points]],pred_homo)
track_transformed.append(transformed_points)
```



ATTEMPT WITH LINE DETECTION

- Picture by TravelScape

ATTEMPT WITH SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

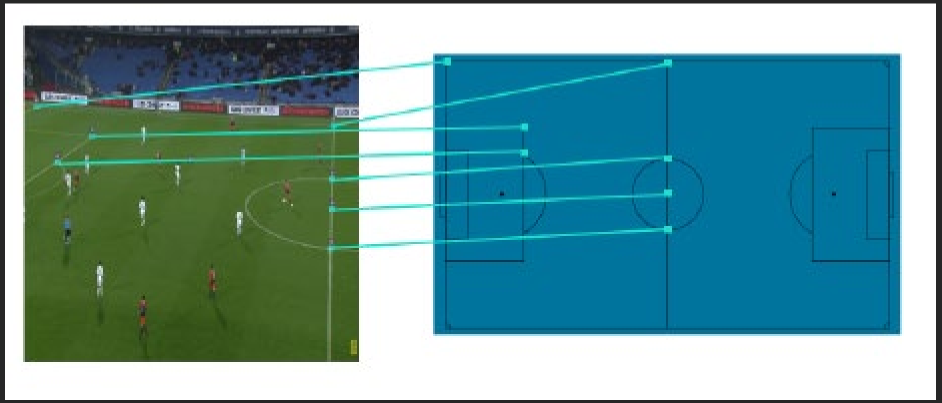
KEYPOINT MODEL

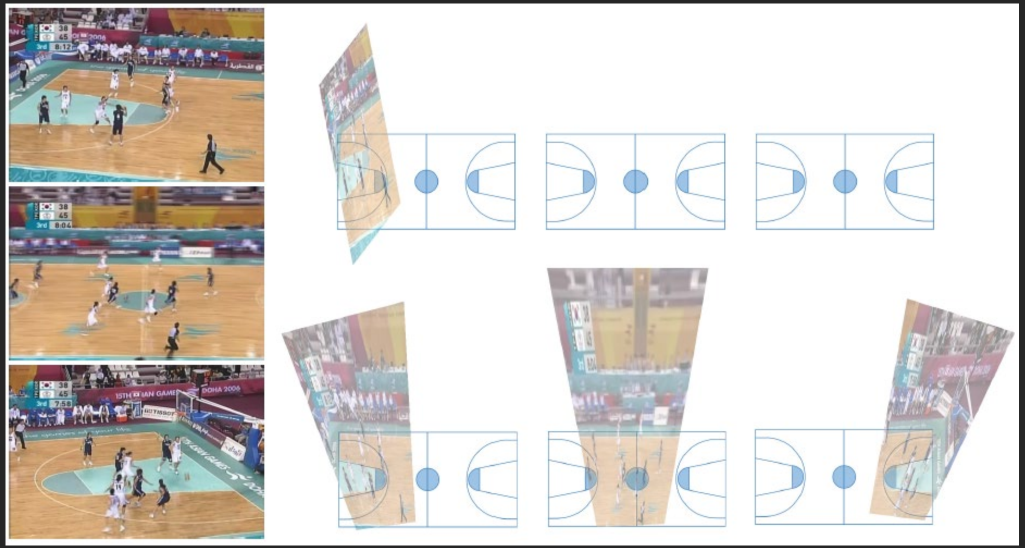
Convolutional Neural Network (CNN)

28 key points

Trained with dataset from Narya paper

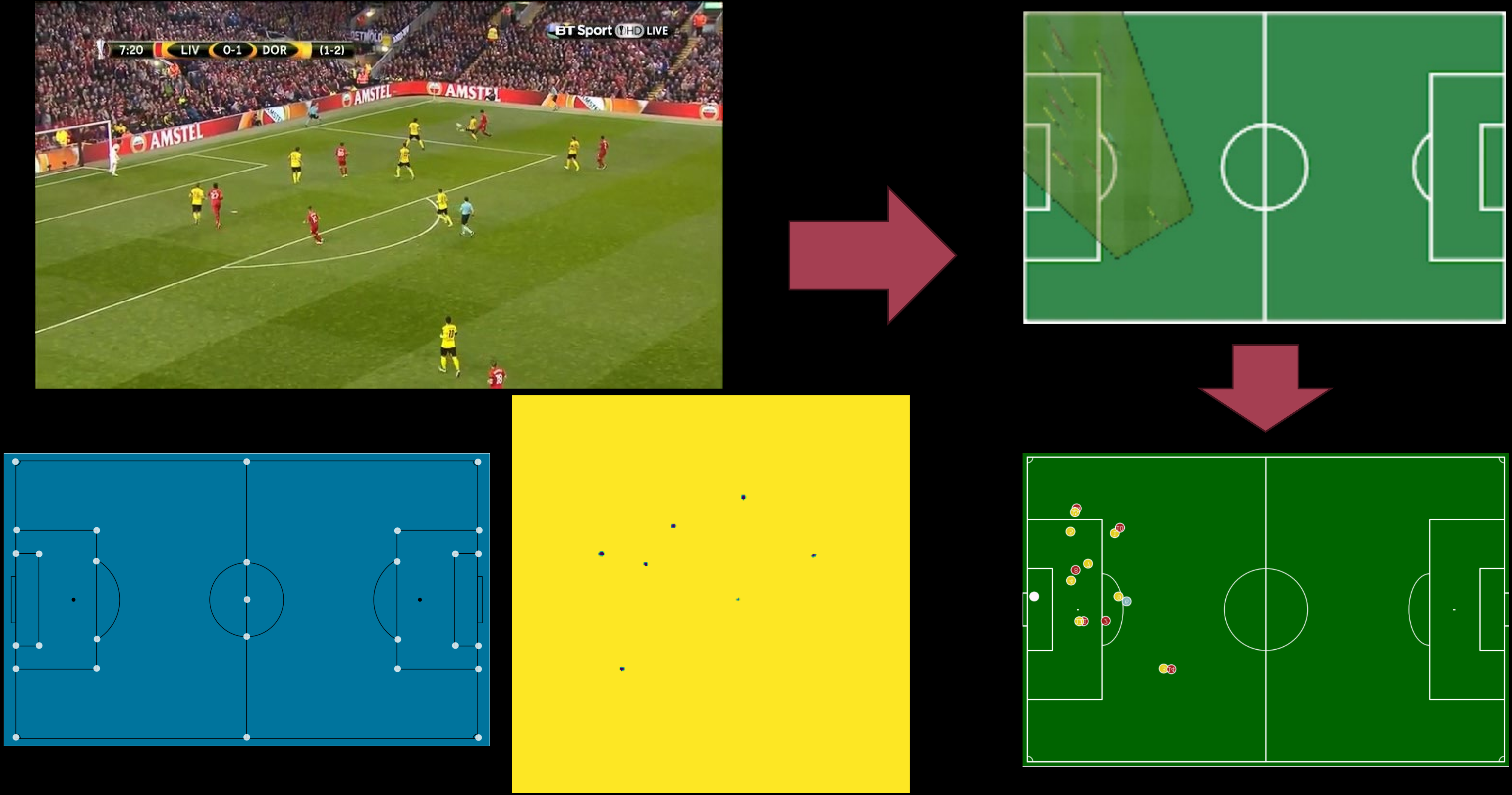
~58 images





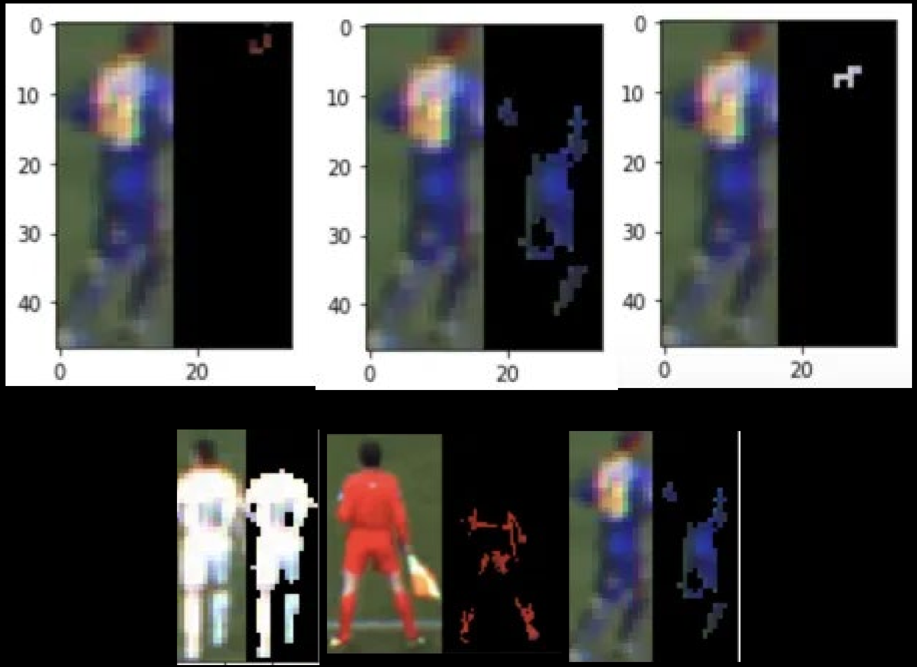
Reference: Garnier, P., & Gregoir, T. (2021). Evaluating Soccer Player: from Live Camera to Deep Reinforcement Learning. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2101.05388

Image: Wen PC, Cheng WC, Wang YS, Chu HK, Tang NC, Liao HM. Court Reconstruction for Camera Calibration in Broadcast Basketball Videos. IEEE Trans Vis Comput Graph. 2016 May;22(5):1517-1526. doi: 10.1109/TVCG.2015.2440236. PMID: 28113142.


7:20 LIV 0-1 DOR (1-2)
BT Sport HD LIVE
AMSTEL

SAMPLE FRAME

TEAM IDENTIFICATION



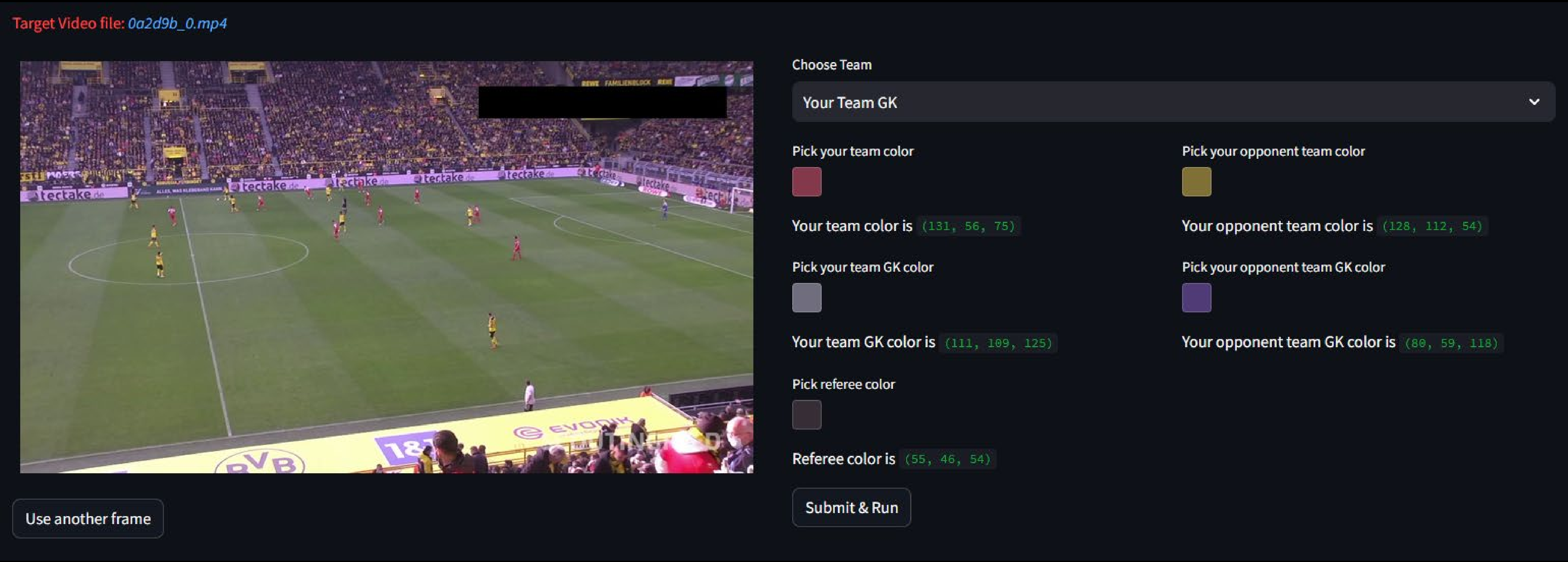
COLOR MASK BASED ON RGB BOUNDARY RANGE

$$Ratio = \frac{Non - black\ pixels}{total\ pixels}$$

Threshold ~ 20-30%

SAMPLE DETECTIONS

```
for team,color in bgrteamcolorlist.items():
    tempframe=frame
    crop_img = tempframe[int(y1):int(y2),int(x1):int(x2)]
    mask = cv2.inRange(crop_img, color[1],color[2])
    # get nonblack vs total pixel ratio
    colorpixelratio=cv2.countNonZero(mask)/(w*h)
    if colorpixelratio > 0.3:
        box_color=color[0]
        label = team + label
        team_name = team_history[track_id]
        team_name.append(team)
        break
```



COLOR PICKER WIDGET

DEMO

APPLICATION

Image Source: freepik

Scouting talents

Identifying players that suits the team’s playstyle

Opponent analysis

Identify opponent weakness and exploit them

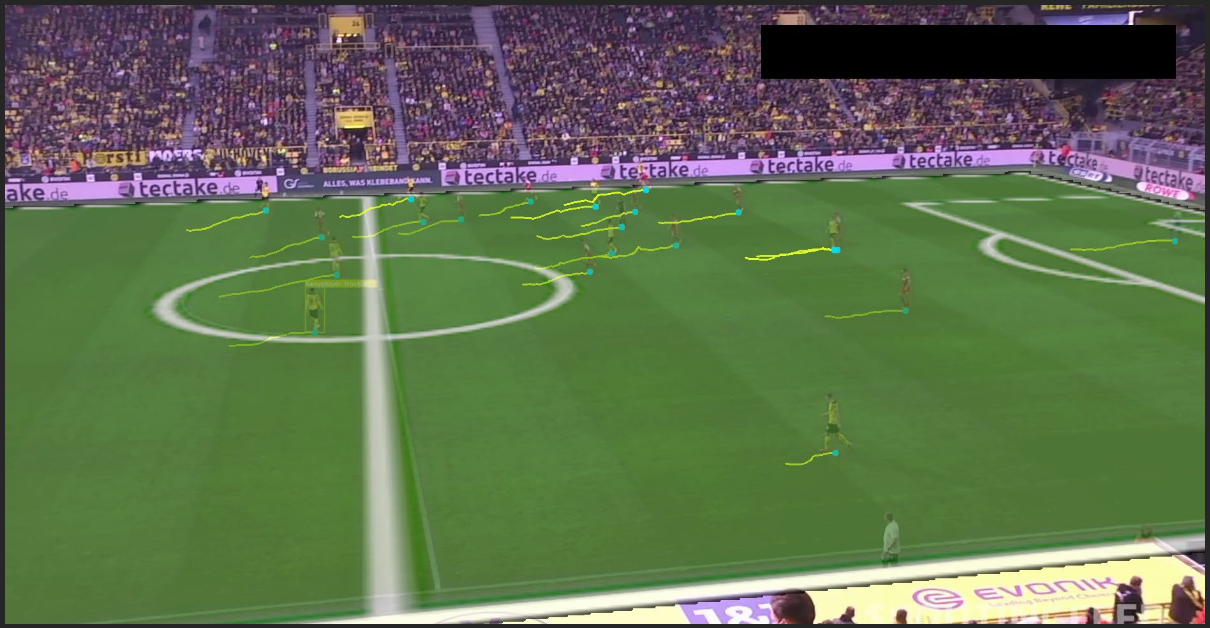
Training analysis

Analyze player strengths and weaknesses

Team dynamics

Analyze players that play well together

NEXT STEPS



Source: Manuel Stein at FC Barcelona Sports Tomorrow

Update Team Identification System

Teams with more than one color on their jersey

Upgrade Classification Model

Have better detections on players as well as on the ball

Upgrade Keypoint Model

Refine homography estimation

Pose Model

Identify body language

Implement AI in analytics

Prompt to visualization with LLM

Metrics

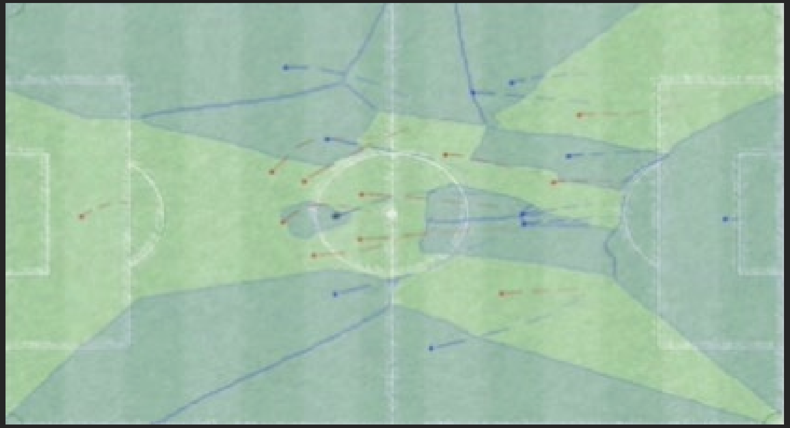
- Identify passes and shots
- Find best passing option and best time to shoot

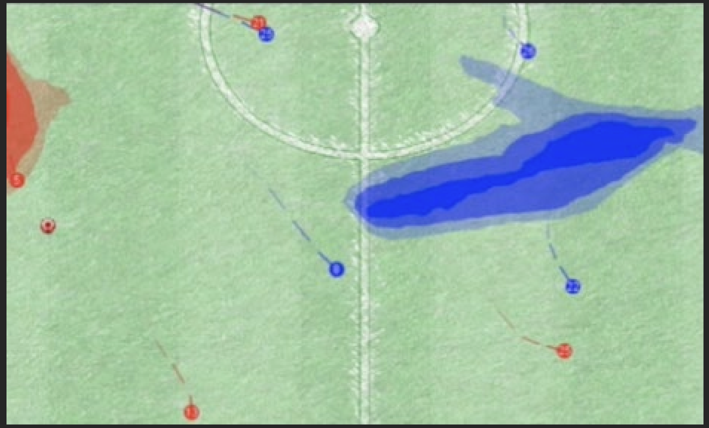
xGs and xAs

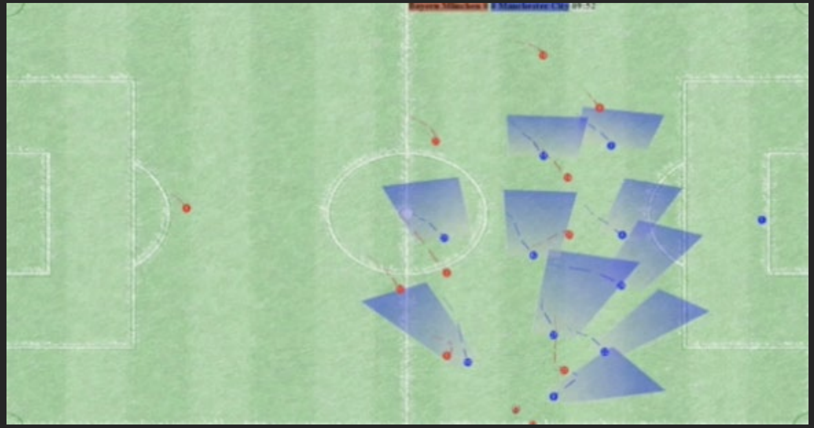
- Predict game results

Identify spaces

- Free space on the pitch
- Dominant zones and cover shadows







Source: Manuel Stein at FC Barcelona Sports Tomorrow

THANK YOU

thealexfok

thisisalexfok@gmail.com

SOURCES

Images

https://www.sofascore.com/player/andre-onana/787498

https://www.sofascore.com/player/jan-oblak/69768

https://www.sofascore.com/player/manuel-neuer/8959

https://www.amazon.com/STATSports-Activity-Football-Performance-Technology/dp/B07NYYP1RC?th=1

https://yolov8.com/

https://soccer-coaches.com/importance-space-pep-guardiola/

https://www.sportperformanceanalysis.com/article/automating-data-collection-and-match-analysis-from-video

SOURCES

Datasource:

https://universe.roboflow.com/roboflow-jvuqo/football-players-detection-3zvbc

Paper references:

Garnier, P., & Gregoir, T. (2021). Evaluating Soccer Player: from Live Camera to Deep Reinforcement Learning. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2101.05388