End-to-End High-Risk Tackle Detection System for Rugby

Naoki Nonaka[1] Ryo Fujihira[1] Monami Nishio[1] Hidetaka Murakami[2]
Takuya Tajima[3] Mutsuo Yamada[4] Akira Maeda[5,6] Jun Seita[1]
[1]Advanced Data Science Project, RIKEN Information R&D and Strategy Headquarters
[2] Murakami Surgical Hospital [3] Faculty of Medicine, University of Miyazaki
[4] Faculty of Health and Sport Sciences, Ryutsu Keizai University
[5] Hakata Knee & Sports Clinic [6] Faculty of Human Health, Kurume University

Presented by Yulu Pan and Aniruddh Doki

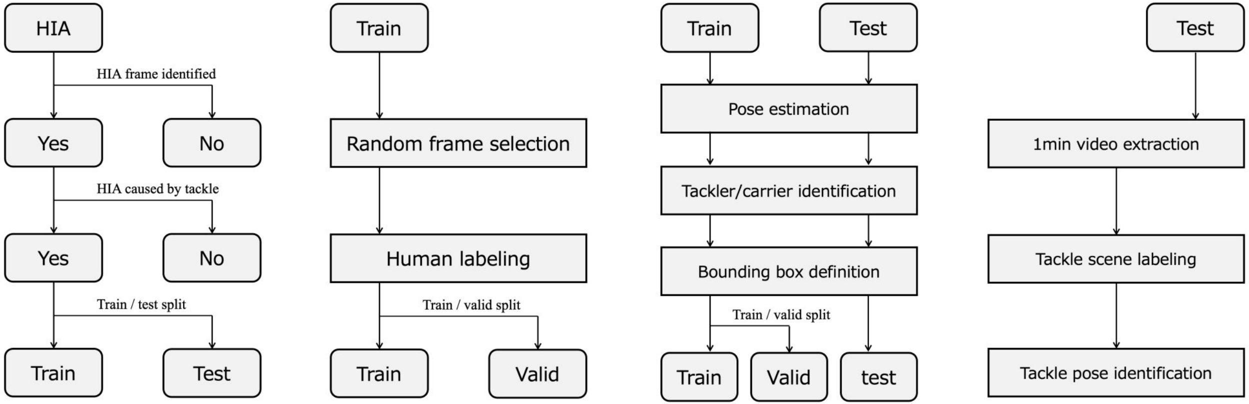
What is Rugby?

- Rugby is a fast-paced collision sport, with a high incidence rate of various injuries, especially concussions
- Concussion is the most common injury in Rugby World Cup
  - 13.9% of all injuries in RWC 2015
  - 15.4% of all injuries in RWC 2019
  - 4.73 per 1,000 player match hours
  - 76% of concussion is caused by tackle
- Head Injury Assessment (HIA)
  - Official match day doctors
  - Only elite level and international games

Gif credits: https://tenor.com/view/rugby-tackle-gif-25042170

Dataset

- Japanese elite league and corresponding official match records from the 2016 to 2018 seasons
  - 360 videos broadcasted on TV, 87 of which contained at least one high-risk tackle frame
  - 226 frames contain event resulted in HIA
  - 87 videos are splitted into training and test set with 9 : 1 ratio



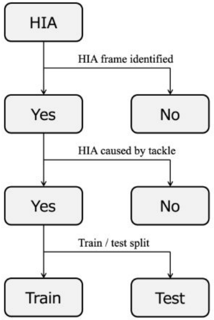
(b) Selection of videos with high-risk tackles. (c) Dataset preparation procedure for tackle frame selection model. (d) Dataset preparation procedure for tackle detection model. (e) Dataset preparation procedure for overall system evaluation.

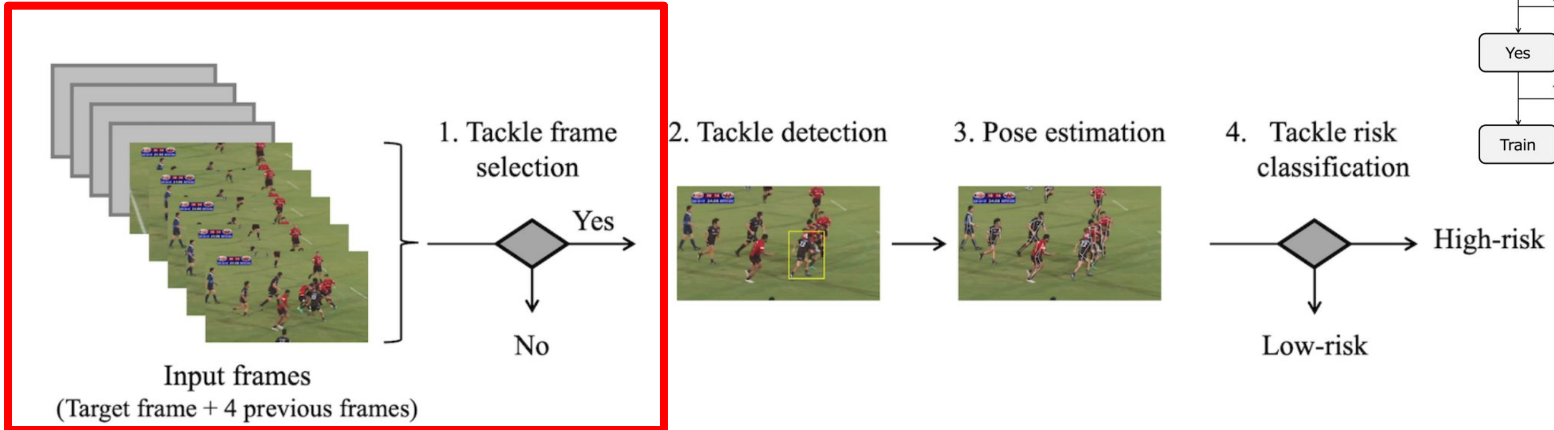
High-risk Tackle Detection System

- High-risk tackle: tackles that lead to a Head Injury Assessment in the official record
  - Potential weakness- model quality is dependent on the quality of the official record
- Four models
  - Tackle frame selection model
  - Tackle detection model
  - Pose estimation model
  - Tackle risk classification model

Tackle Frame Selection Model

- Determine whether a video clip contains a tackle or not (binary classification)
- Take 100 video clips of 2 seconds from each 78 training video dataset (7800)
- Manually checked each video clip and labeled whether the final frame of video clip contains tackle or not
  - 199 video clips with and 7601 video clips without tackle



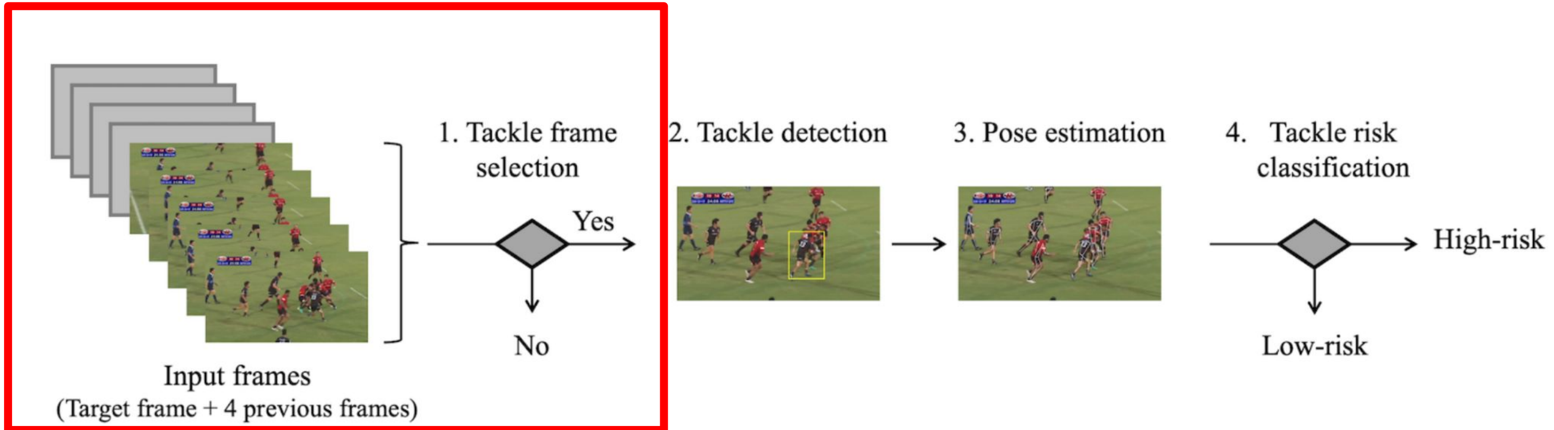


Tackle Frame Selection Model

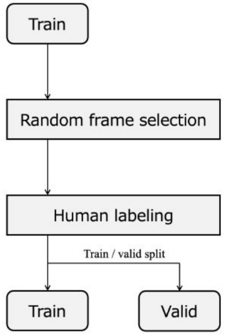
- Pre-trained with Kinetics-400, fine-tuned with previous data

Frame selection model	Macro F1	Recall	Precision
No classifier	0.114	1.	0.136
ResNet Mixed Convolution	0.564	0.199	0.312
ResNet (2+1)D	0.565	0.21	0.301
ResNet 3D	0.534	0.127	0.275

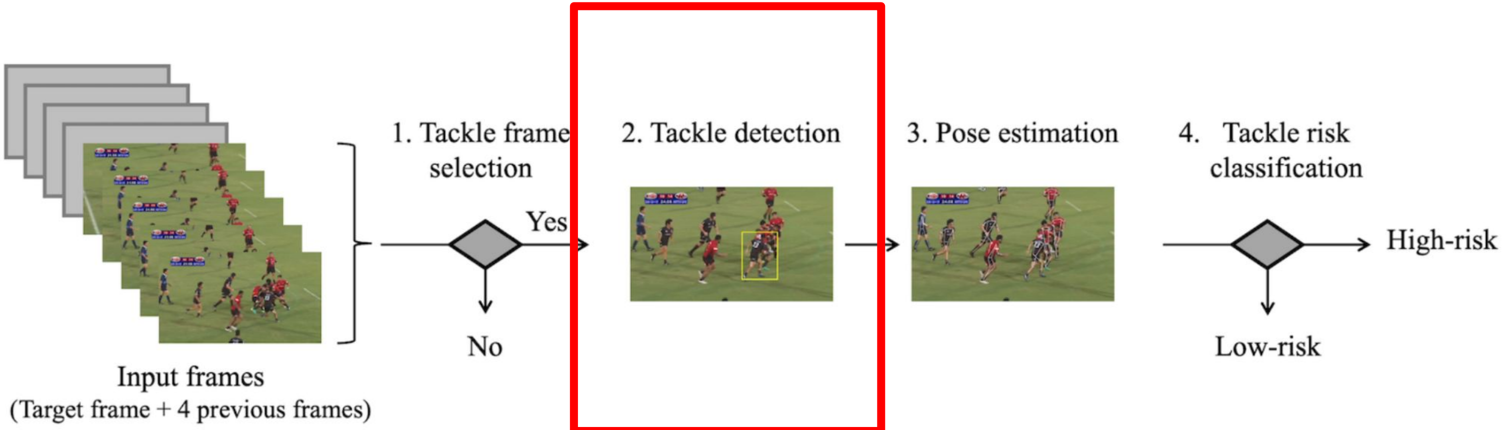
No Classifier: assuming all frames as tackle frame
Recall = TP / (TP + FN)
Precision = TP / (TP + FP)
F1 = 2* (Precision * Recall) / (Precision + Recall)
Macro: 1/N * Σ F1



Tackle Detection Model



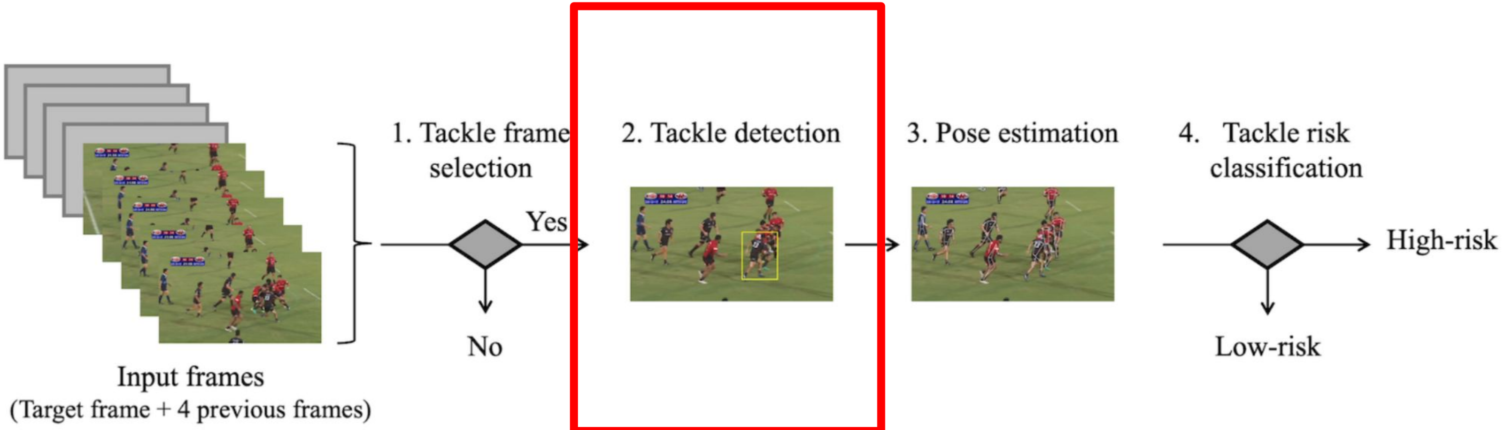
- Select low risk tackles and high risk tackles 4:1 ratio
- CenterTrack
  - Identified tackler and ball carrier
- Selected frames with 5 or more key-points detected for both players
- Tackle area: rectangular area covering both posture from coordinates of the players



Tackle Detection Model

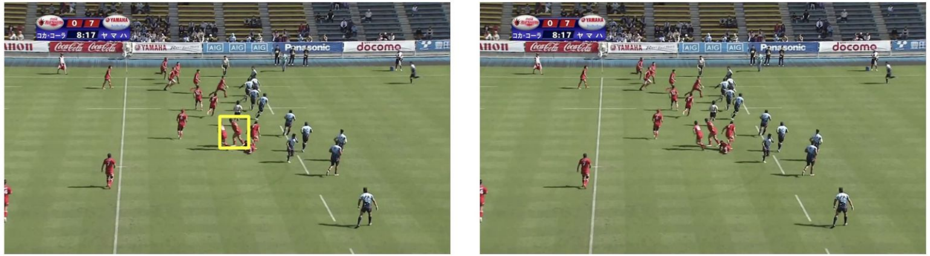
- Pre-trained with COCO, fine-tuned previous data

	Top confidence bbox IoU	Average bbox IoU	Best bbox IoU	ratio of detection
DETR	0.647	0.646	0.679	0.939 (31/33)
RetinaNet	0.655	0.577	0.655	0.939 (31/33)
YOLOv3	0.277	0.277	0.277	0.364 (12/33)



Tackle Detection Model

(a) Example of an image in which both DETR (left) and RetinaNet (right) were successful in detecting a tackle.

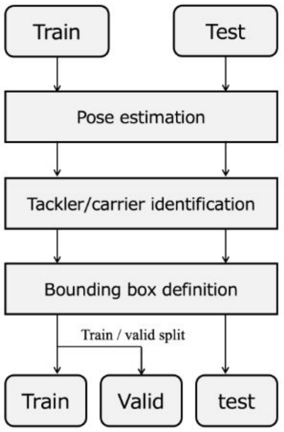


(b) Example of an image in which a false positive occurs in DETR (left) but not in RetinaNet (right).

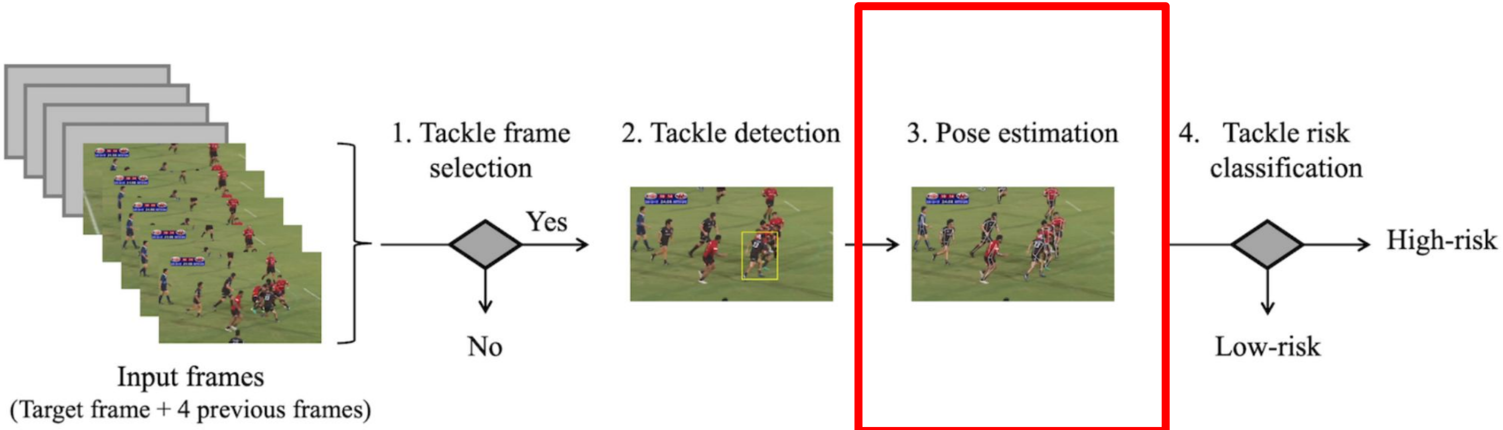
(c) Example of an image in which correctly detected in DETR (left) but not in RetinaNet (right).

(d) Example of an image in which both DETR (left) and RetinaNet (right) failed in detecting a tackle.

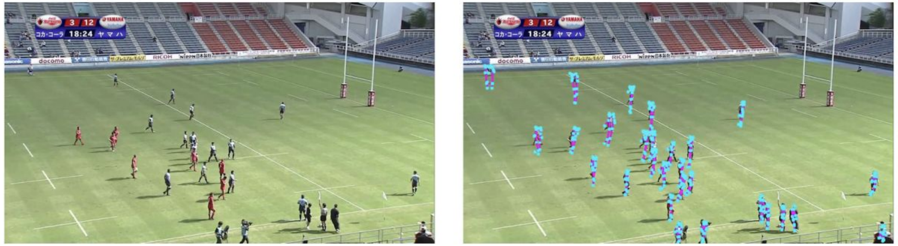
Pose Estimation Model



- HRNet and CenterTrack
- Pre-trained with COCO dataset and no additional training
- Apply pose estimation model to extract posture of all players
- Extract tackle related players
  - player's part of torso is located inside tackle region given by tackle detection model

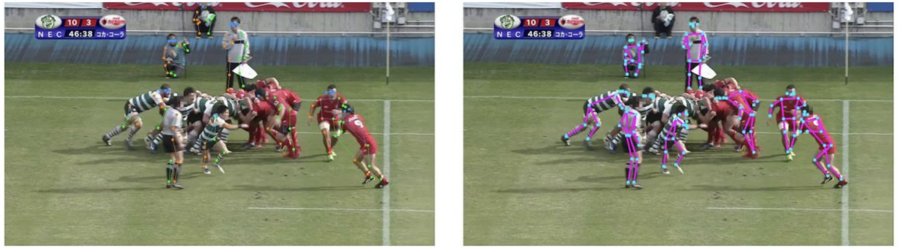


Pose Estimation Model



(a) Example of pose estimation with HRNet (left column) and CenterTrack (right column)

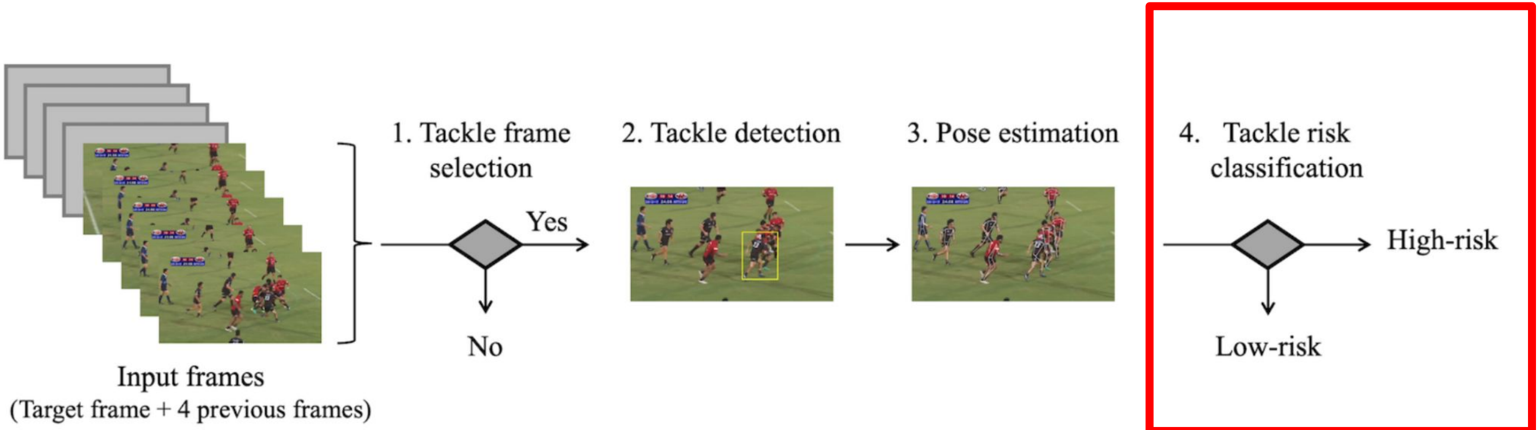
(b) Example of an zoom in image, both HRNet (left) and CenterTrack (right) succeeded.



(c) Example of an image with occlusion, both model failed with occluded players.

Tackle Risk Classification Model

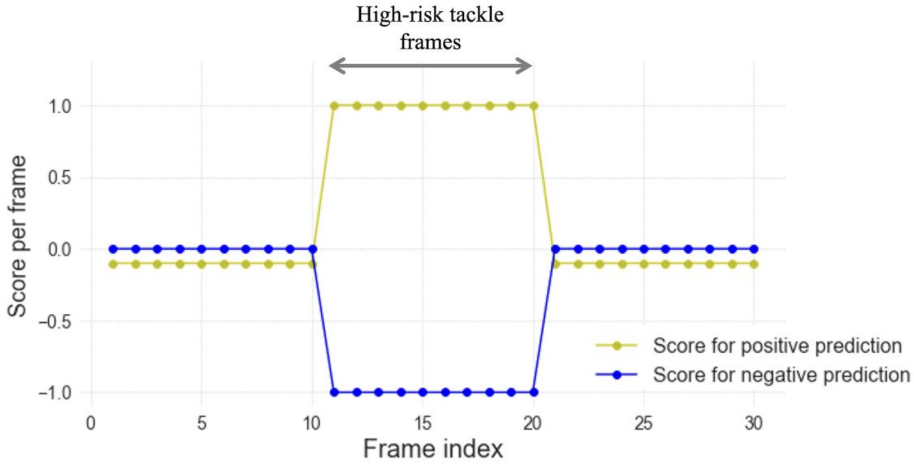
- Classify whether tackle in given frame is high-risk or not
  - Using tackle related players’ posture pair
  - If three or more postures are related to tackle, take all combination of pairs and evaluate each pair by Naive Bayes model



Evaluation

- Positive example: identify frame 1.5 seconds before and after the high-risk tackle
- True positive: +1
- False negative: -1
- False positive: -0.1
- True negative: 0

$$U_{score} = \frac{U_{total} - U_{neg}}{U_{max} - U_{neg}}$$



Results

Frame selection model	Tackle detection model	Pose estimation model	Score	Recall
Human labels	RetinaNet	HRNet	0.3449	0.583
		CenterTrack	0.4905	0.833
	DETR	HRNet	0.2249	0.417
		CenterTrack	0.5397	0.917
No selection	RetinaNet	HRNet	0.2312	0.583
		CenterTrack	0.2759	1.000
	DETR	HRNet	0.2204	0.583
		CenterTrack	0.2224	1.000
ResNet Mixed Convolution	RetinaNet	HRNet	0.1837	0.333
		CenterTrack	0.0793	0.167
	DETR	HRNet	0.1825	0.333
		CenterTrack	0.1680	0.333
ResNet 2+1D	RetinaNet	HRNet	0.0840	0.167
		CenterTrack	0.2807	0.500
	DETR	HRNet	0.000	0.000
		CenterTrack	0.2719	0.500
ResNet 3D	RetinaNet	HRNet	0.0867	0.167
		CenterTrack	0.0400	0.083
	DETR	HRNet	0.0866	0.167
		CenterTrack	0.0820	0.167

Limitations

- Multiple deep neural network models resulting in slow processing speed
- Fail to perform pose estimations when players are occluded

Thank You