Advancing Player Identification and Tracking with Global ID Fusion (GIF)

Karol Wojtulewicz*, Minxing Liu*, and Niklas Carlsson

Linköping University, Sweden

LINKÖPING UNIVERSITY

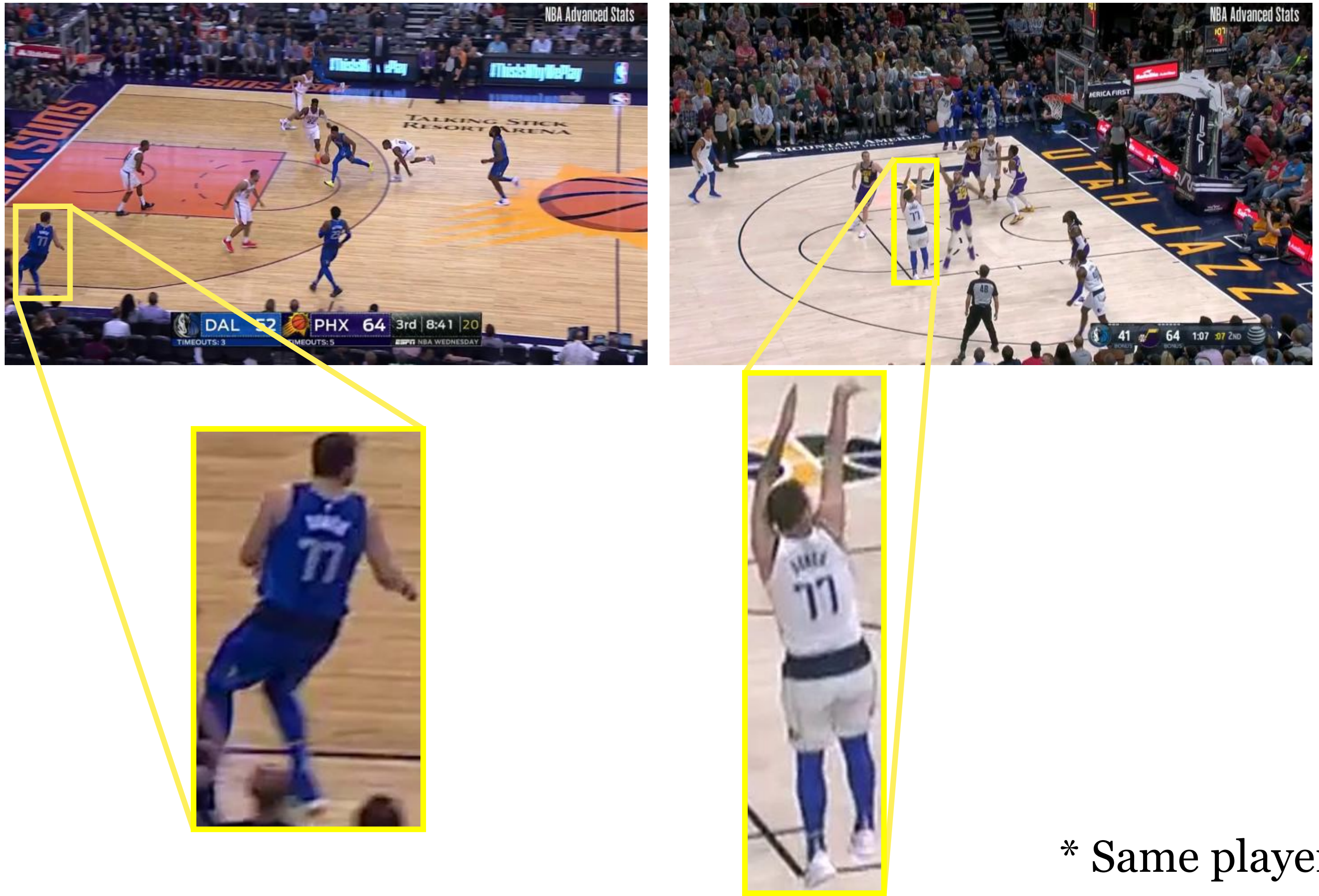
WACV2026, Tucson, Arizona

Multi-perspective Broadcasting Videos

- Multi-perspective
- Clothe-changing
- Zoom-in zoom-out
- Irregular poses
- Uniform

Multi-perspective Broadcasting Videos

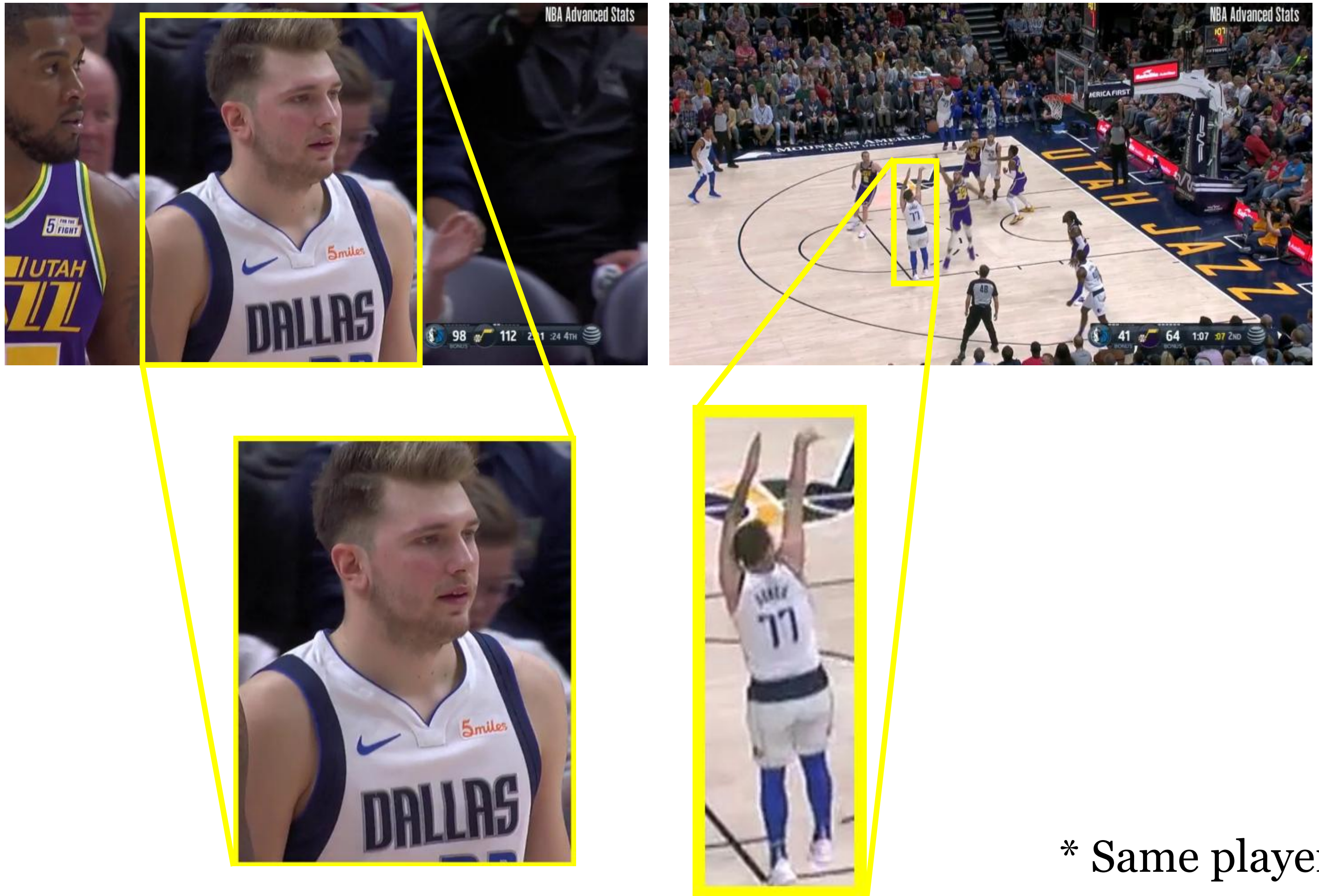
- Multi-perspective
- Clothe-changing
- Zoom-in zoom-out
- Irregular poses
- Uniform



* Same player

Multi-perspective Broadcasting Videos

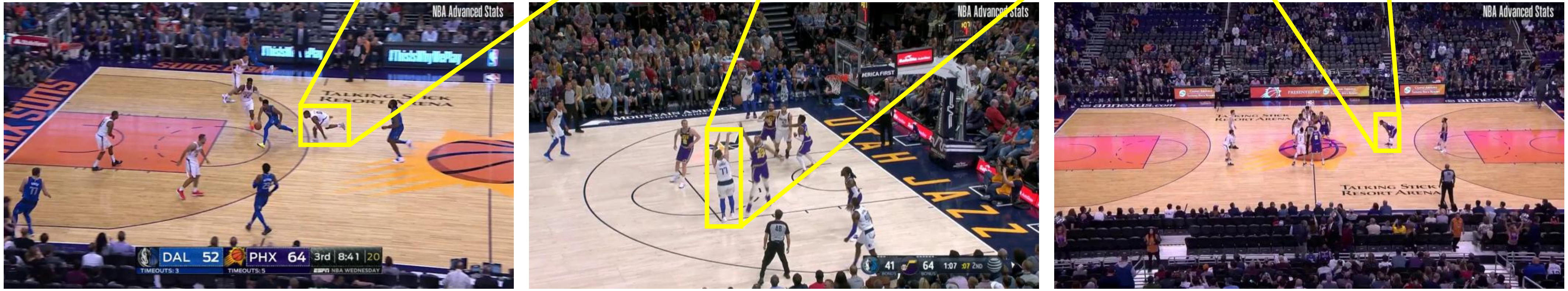
- Multi-perspective
- Clothe-changing
- Zoom-in zoom-out
- Irregular poses
- Uniform



* Same player

Multi-perspective Broadcasting Videos

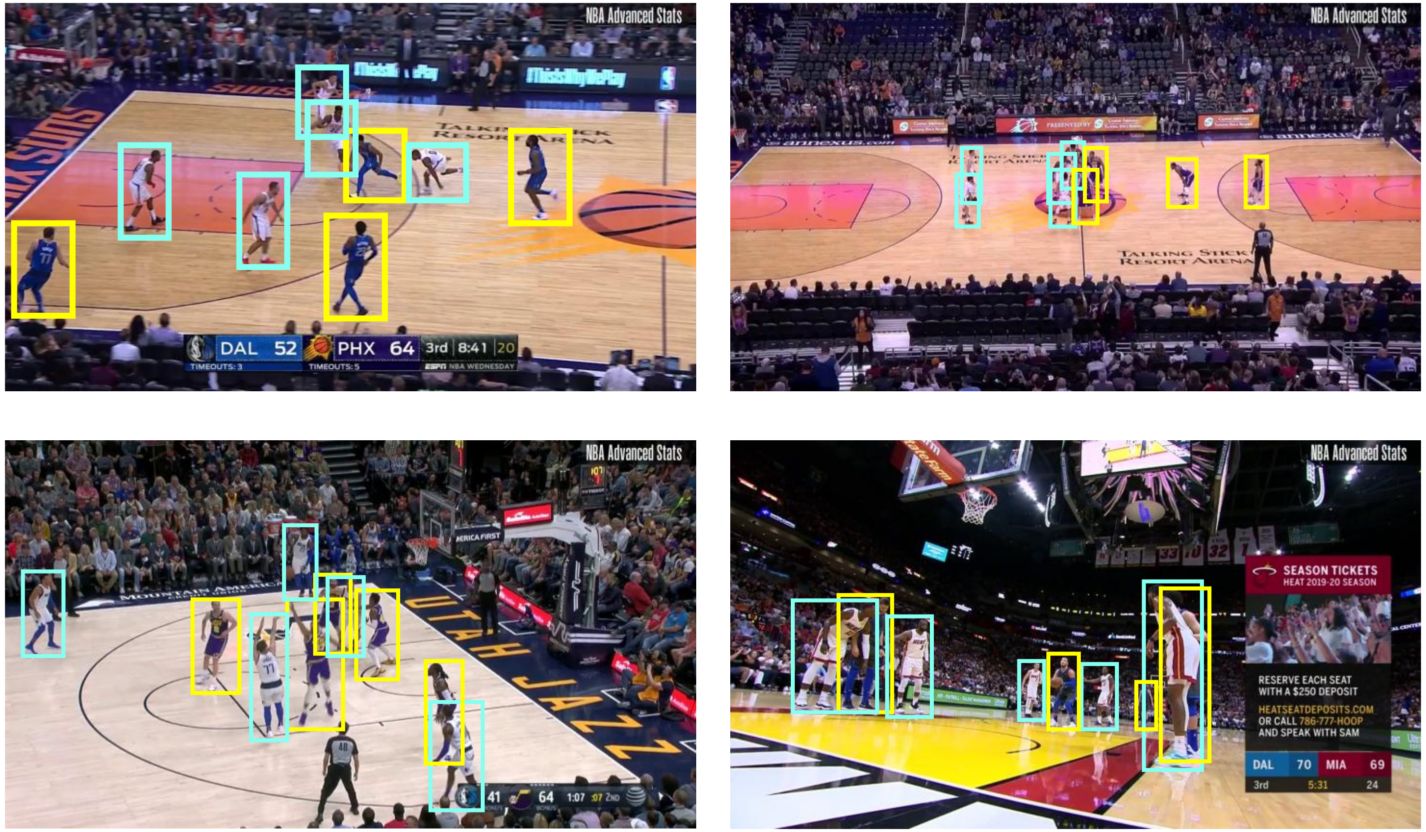
- Multi-perspective
- Clothe-changing
- Zoom-in zoom-out
- Irregular poses
- Uniform



LINKÖPING UNIVERSITY

Multi-perspective Broadcasting Videos

- Multi-perspective
- Clothe-changing
- Zoom-in zoom-out
- Irregular poses
- Uniform



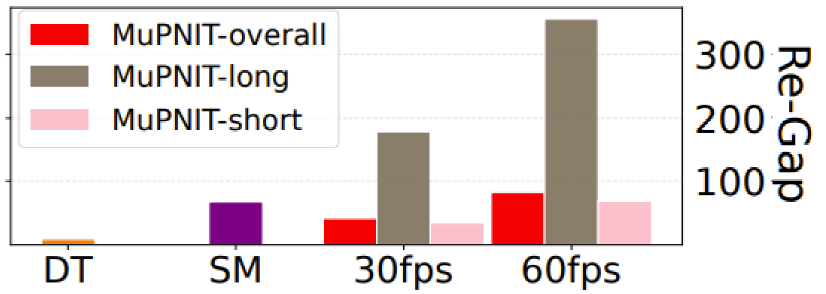
Our Contributions

- **MuPNIT & MuPNIT-ReID Datasets**: A multi-perspective benchmark for sports, featuring 2.19M boxes across 283K frames of NBA games. It addresses critical gaps in tracking across camera cuts, zooms, and cross-season appearance changes (jerseys, lighting, and poses), with a global ID feature bank
- **Global ID Fusion (GIF) Framework**: A high-performance, online fusion approach for real-time tracking and identification
- **Identity Consistency Metrics**: Introduction of five novel Global ID metrics to evaluate tracking stability and provide new insights into identity prediction tradeoffs

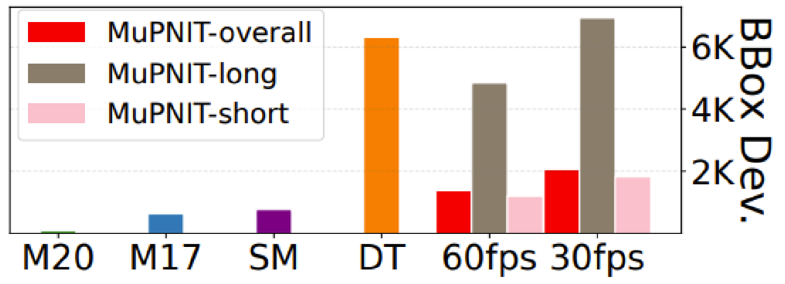
Multi-Perspective NBA Global Identity Tracking (MuPNIT) Datasets - MOT

Table 1. **Comparison of object tracking datasets.** Stats for the long and short videos are shown as long/short/overall. Colors are used to indicate largest and second-largest entry for each feature.

Dataset	MOT17 [51]	MOT20 [14]	DanceTrack [64]	SportsMOT [10]	SMOT [42]	MuPNIT 60fps	MuPNIT 30fps
Videos	14	8	100	240	**3,292**	10 / 200 / 210	
Min. Len. (s)	15.0	**17.2**	9.2	9.0	1.5	**252.4** / 5.5 / 5.5	
Avg. Len. (s)	33.0	**67.1**	52.9	24.7	22.9	**282.7** / 9.4 / 22.4	
Max. Len. (s)	85.3	**132.6**	120.1	73.0	116.0	**339.9** / 15.8 / **339.9**	
Total Len. (s)	462	536	5,292	**6,015**	**75K**	2,827 / 1,882 / 4,709	
#Tracks	796	2,332	990	3,401	**7,792**	1,661 / 2,492 / **4,153**	
#Boxes	340K	**1,699K**	574K	1,629K	335K	**2,187K**	1,094K
#Frames	11K	13K	106K	150K	**151K**	**283K**	141K
fps	**30**	25	20	25	2	**60**	**30**
multi-perspective	-	-	-	-	-	✓	✓
Global IDs	-	-	-	-	-	✓	✓



(a) Average Re-association Gap

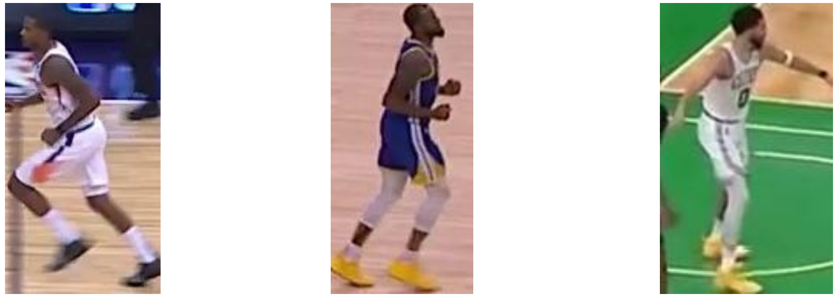


(b) Consec. BBox Size Deviation

Multi-Perspective NBA Global Identity Tracking (MuPNIT) Datasets - ReID

- MuPNIT-ReID: all player crops
- MuPNIT-ReID-light: ~200 images/player

Global ID database



In-game player crops

Table 2. **Comparison of ReID datasets.** Here + denotes image-to-image and * synthetic datasets. Acronyms: uniform (**Uni**), irregular poses (**IrP**), and number of outfits per identity (**#OpI**).

Dataset	#IDs	#Images	#Tracklets	Text	Uni	IrP	#OpI
Market1501+ [85]	751	13K	–	–	–	–	–
CUHK-SYSU+ [74]	8,432	100K	–	–	–	–	–
MARS [86]	1,261	1.19M	20K	–	–	–	1
iLIDSVID [33]	300	42K	600	–	–	–	1
PRCC+ [77]	221	34K	–	–	–	–	–
MTA* [30]	–	37.3M	2,840	–	–	–	1
P-DESTRE [31]	253	14.8M	1,894	–	–	–	–
CCVID [23]	226	348K	2,856	–	–	–	∼2
MEVID [12]	158	1.7M	8,092	–	–	–	∼4
MuPNIT-ReID-light	211	55K	2,260	✓	✓	✓	1–5
MuPNIT-ReID	211	396K	2,415	✓	✓	✓	1–5

Identity Tracking

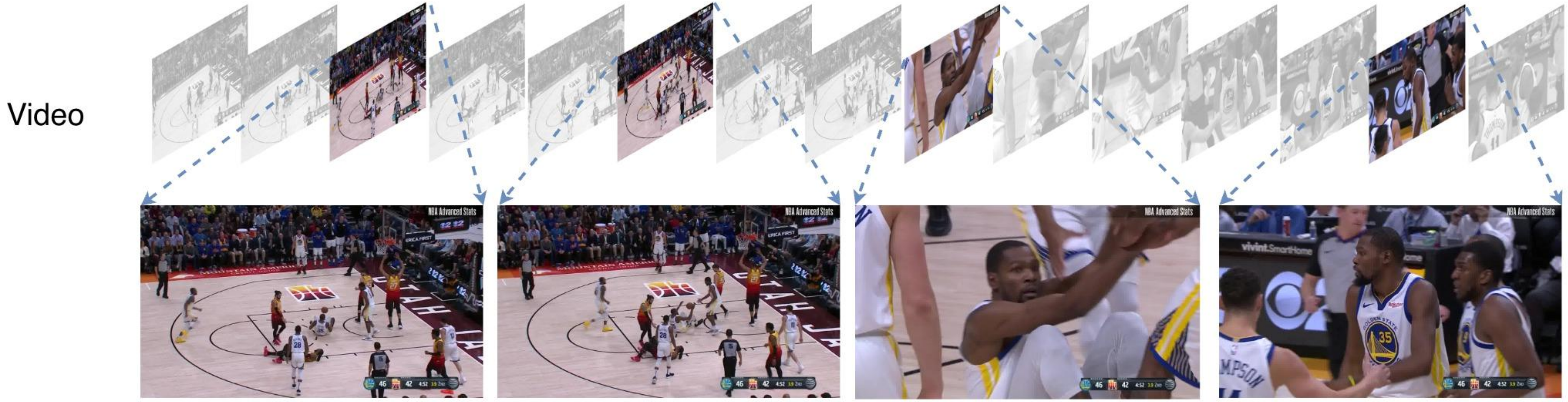


Figure 1. Visualization of GIF. Detection images are mapped against global ID representations (images and/or text) provided as context. The most frequent global ID assigned to a track is selected as its identity, enabling tracking across changes of camera perspectives.

Identity Tracking

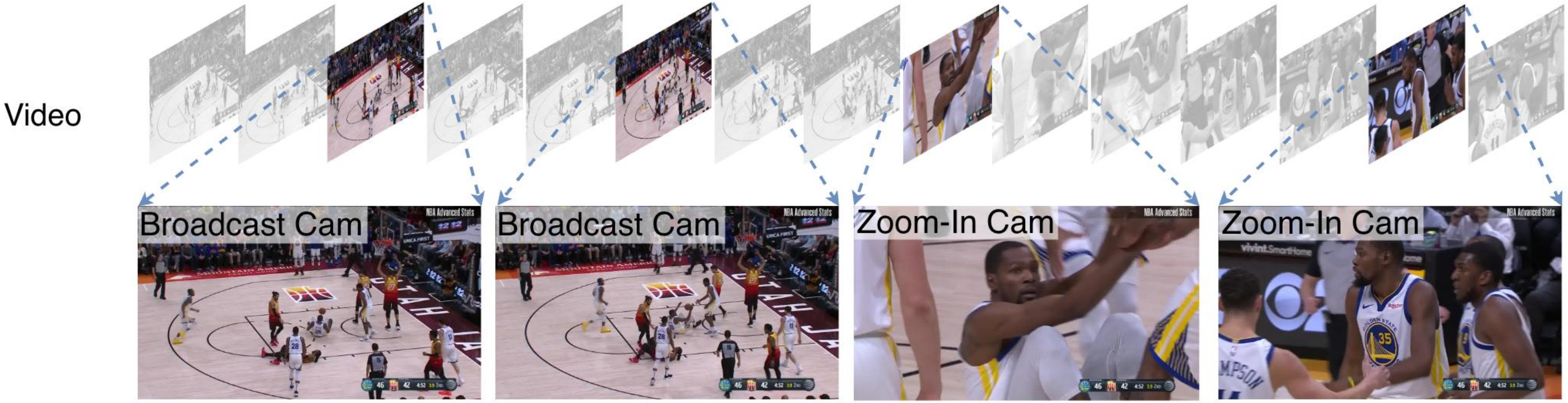


Figure 1. Visualization of GIF. Detection images are mapped against global ID representations (images and/or text) provided as context. The most frequent global ID assigned to a track is selected as its identity, enabling tracking across changes of camera perspectives.

Identity Tracking

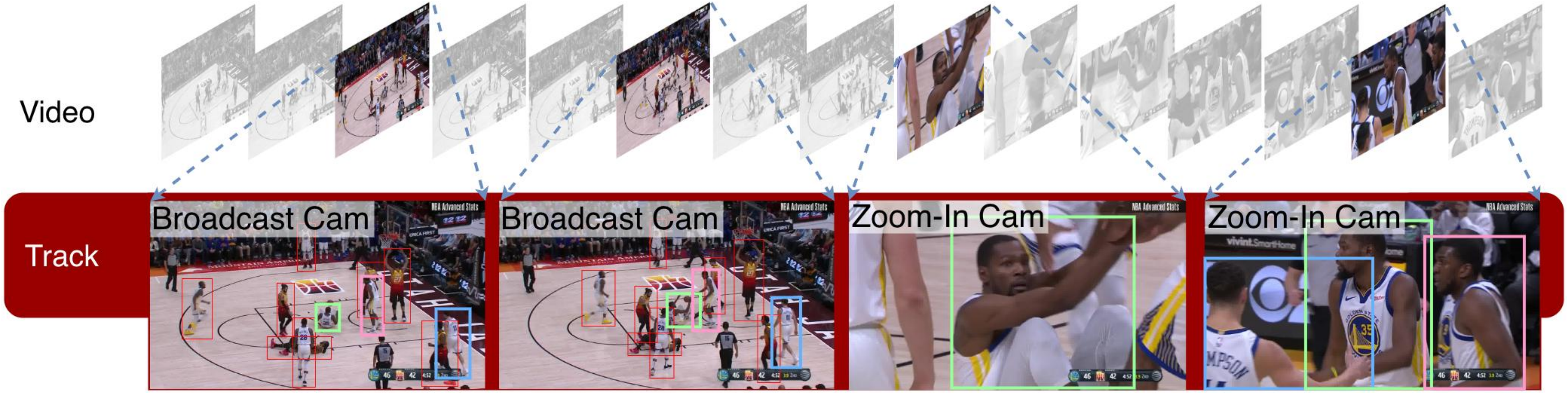


Figure 1. Visualization of GIF. Detection images are mapped against global ID representations (images and/or text) provided as context. The most frequent global ID assigned to a track is selected as its identity, enabling tracking across changes of camera perspectives.

Identity Tracking

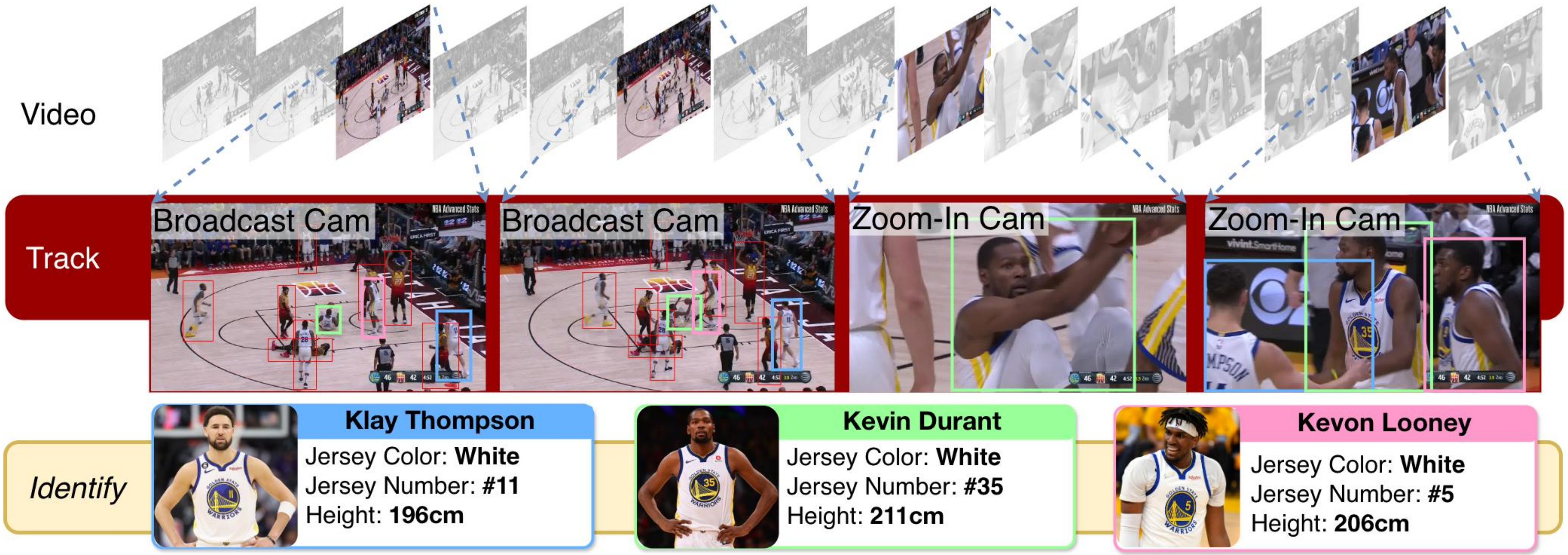


Figure 1. Visualization of GIF. Detection images are mapped against global ID representations (images and/or text) provided as context. The most frequent global ID assigned to a track is selected as its identity, enabling tracking across changes of camera perspectives.

Identity Tracking

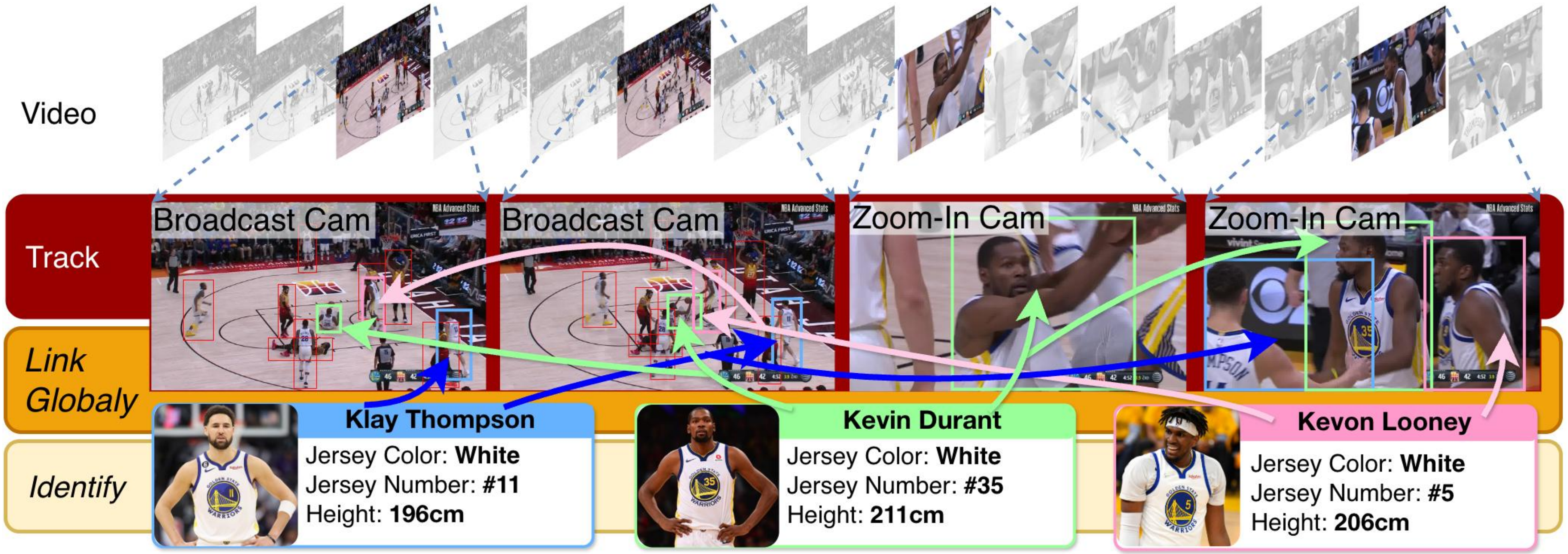
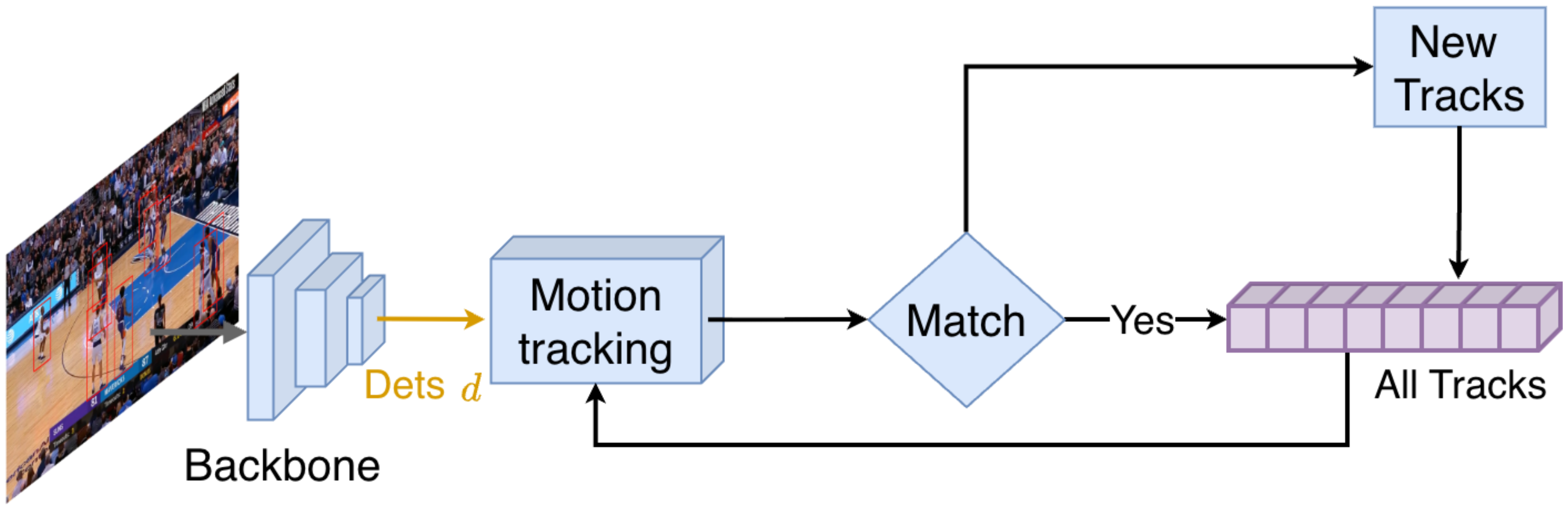
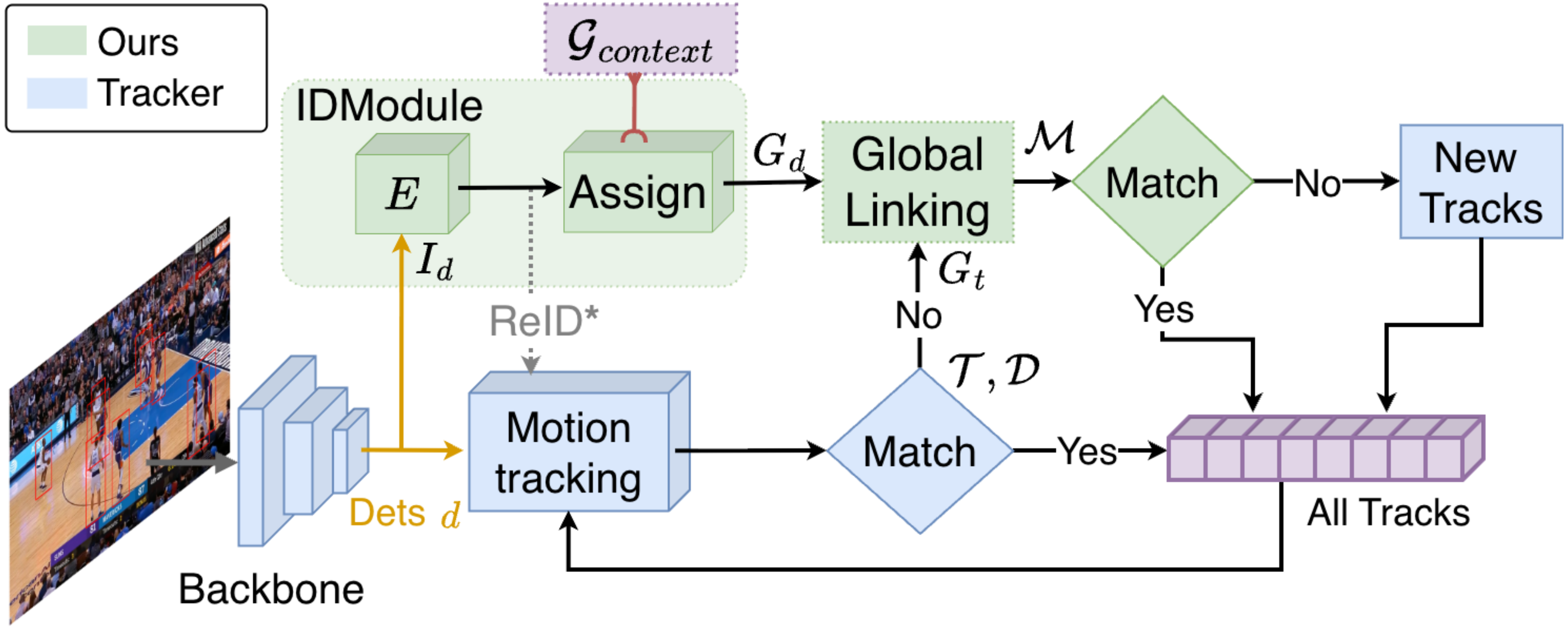


Figure 1. Visualization of GIF. Detection images are mapped against global ID representations (images and/or text) provided as context. The most frequent global ID assigned to a track is selected as its identity, enabling tracking across changes of camera perspectives.

Our Solution



Our Solution



Our Solution (Example)

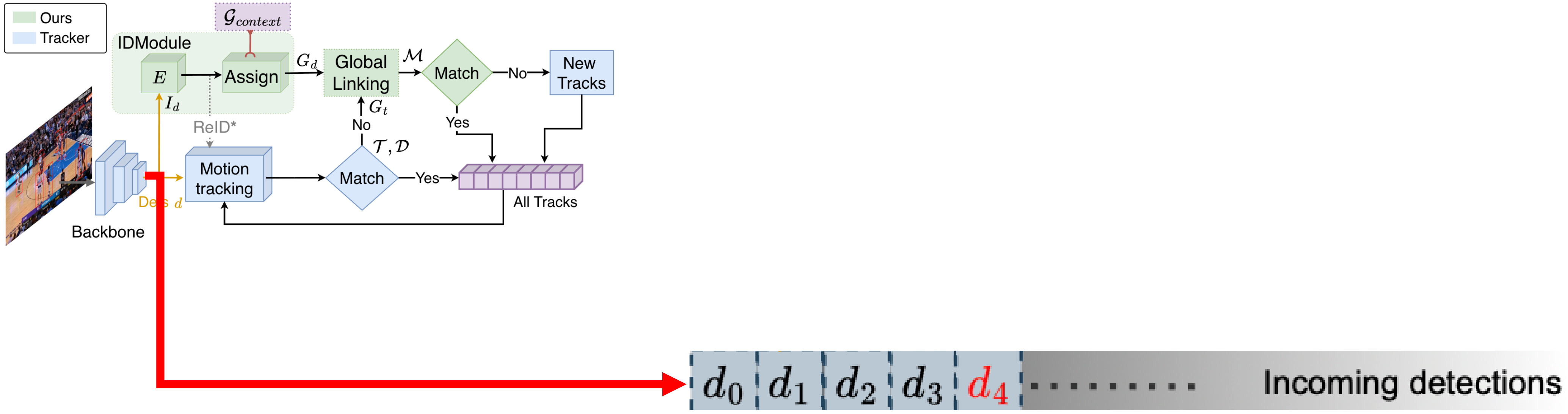


Figure 4. Visualization of successful Global Linking when the BBox assignment fails. G_{id} are global IDs assigned to the tracklet T_{id} which represents track ID, and IDM is our ID module.

Our Solution (Example)

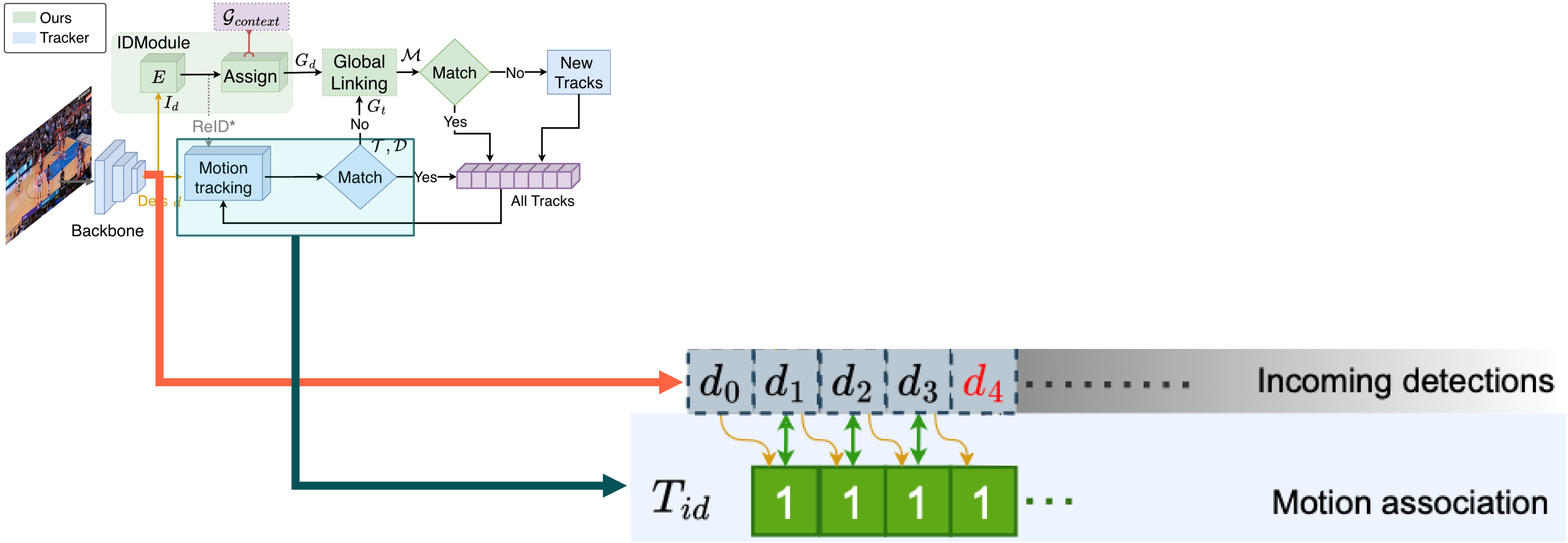


Figure 4. Visualization of successful Global Linking when the BBox assignment fails. G_{id} are global IDs assigned to the tracklet T_{id} which represents track ID, and IDM is our ID module.

Our Solution (Example)

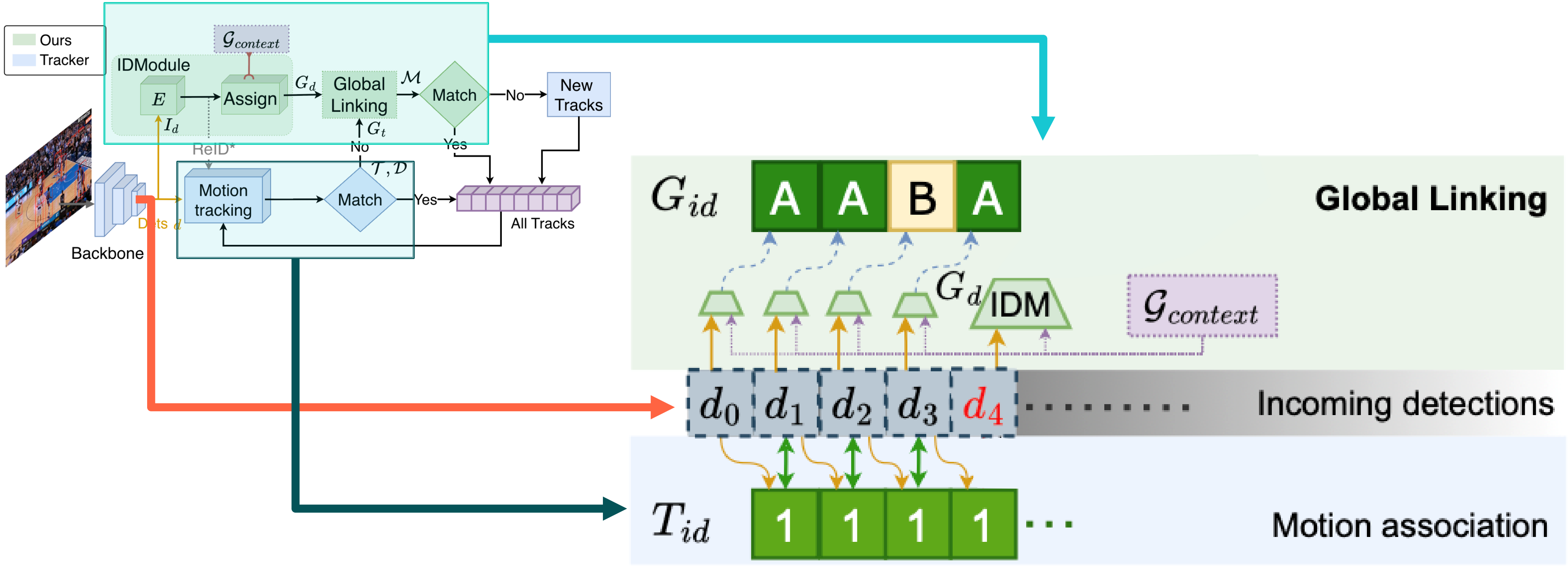


Figure 4. Visualization of successful Global Linking when the BBox assignment fails. G_{id} are global IDs assigned to the tracklet T_{id} which represents track ID, and IDM is our ID module.

Our Solution (Example)

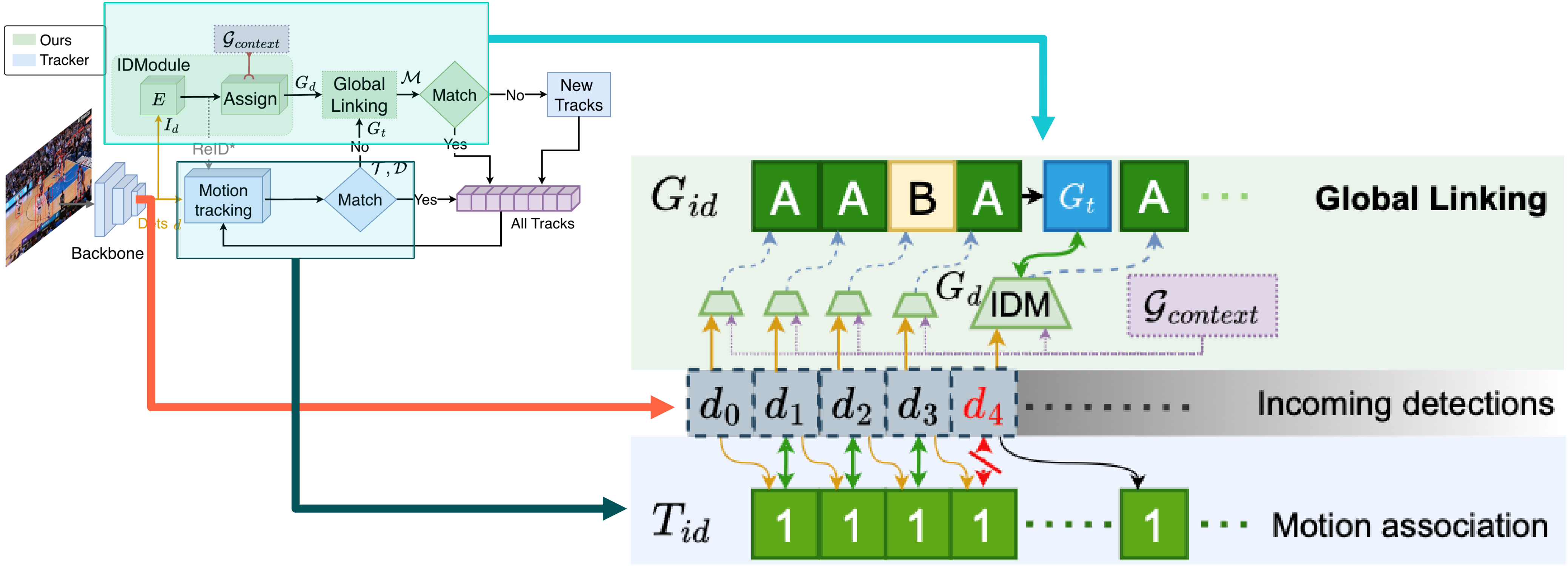
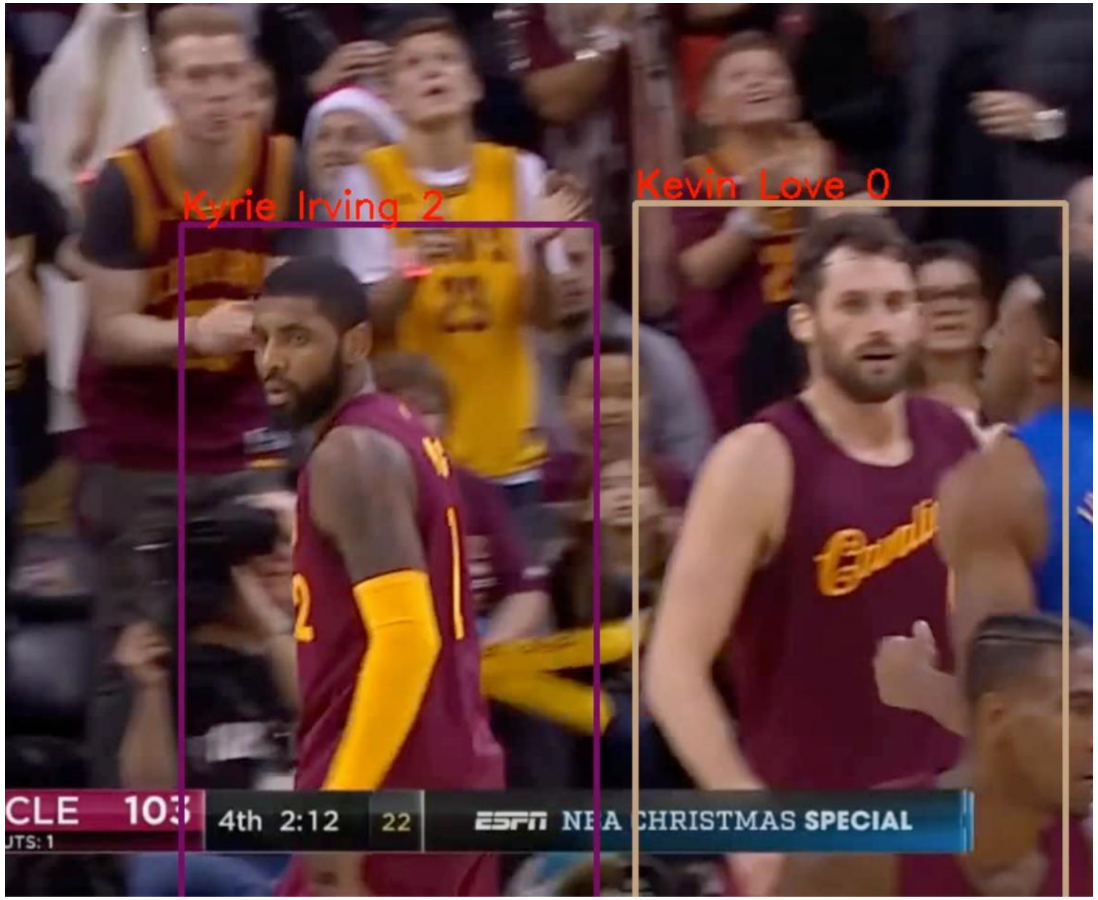


Figure 4. Visualization of successful Global Linking when the BBox assignment fails. G_{id} are global IDs assigned to the tracklet T_{id} which represents track ID, and IDM is our ID module.

Global ID Metrics

- **Frames to First ID (FTFI):**
- The average time to first classify all tracklets
- **Frames to First Correct ID (FTFIc):**
- The average time to correctly classify IDs
- **Average Class Switches (IDsw):**
- The average number of tracklets' ID switches
- **Mean Correct Ratio (MCorr):**
- The ratio of correct ID-tracklet matches over time
- **Mean Average Precision (mAP):**
- The precision of ID classification for a tracklet



Results: ReID

- CLIP family and AP3D stand out, motivating our MOT+ReID selection

Table 4. ReID models comparison on MuPNIT-ReID-light. *proj* and *txt* represent extra projection layer and additional textual features respectively. *Facial features only.

Method	mAP↑	Top-1↑	Top-5↑	Top-10↑	Top-20↑
PCB [66]	11.0%	16.3%	39.6%	52.5%	72.4%
InsightFace* [2]	16.0%	16.1%	27.3%	36.7%	50.0%
MGN [69]	17.1%	18.6%	41.0%	53.2%	73.8%
IANet [27]	17.5%	26.3%	54.5%	66.3%	79.2%
PSTA [70]	17.5%	26.6%	51.1%	64.7%	79.5%
RGA-SC [83]	17.6%	25.9%	54.5%	69.2%	81.6%
ReIDCaps [28]	18.0%	29.8%	51.6%	64.7%	80.0%
CAL [23]	18.4%	26.7%	50.4%	63.7%	78.4%
TransReID [26]	30.4%	46.7%	71.4%	82.7%	93.1%
CLIP [56]	13.5%	18.6%	38.0%	50.0%	58.8%
CLIP_{proj} [56]	25.3%	32.2%	59.6%	69.8%	81.8%
$\text{CLIP}_{proj,txt}$ [56]	33.6%	**46.8%**	70.2%	80.9%	89.6%
AP3D [22]	**48.1%**	30.3%	**71.6%**	**88.1%**	**98.2%**

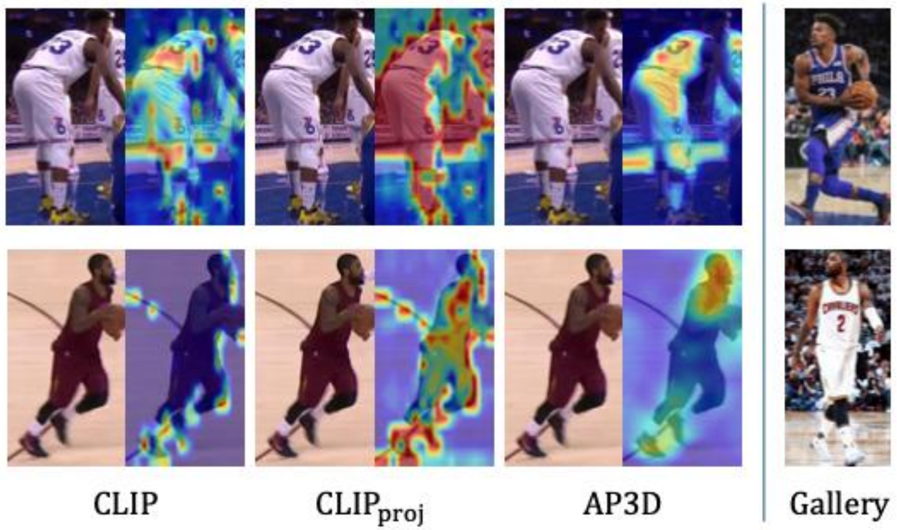


Figure 6. Activation maps on query images (left), when matching with gallery images containing different jersey colors (right).

Results: ReID

- Low face visibility makes face-based ReID fail

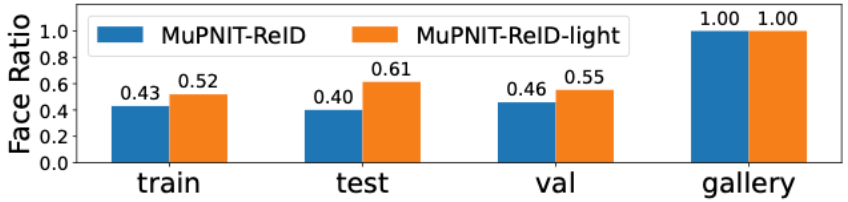


Table 4. ReID models comparison on MuPNIT-ReID-light. *proj* and *txt* represent extra projection layer and additional textual features respectively. *Facial features only.

Method	mAP↑	Top-1↑	Top-5↑	Top-10↑	Top-20↑
PCB [66]	11.0%	16.3%	39.6%	52.5%	72.4%
InsightFace* [2]	16.0%	16.1%	27.3%	36.7%	50.0%
MGN [69]	17.1%	18.6%	41.0%	53.2%	73.8%
IANet [27]	17.5%	26.3%	54.5%	66.3%	79.2%
PSTA [70]	17.5%	26.6%	51.1%	64.7%	79.5%
RGA-SC [83]	17.6%	25.9%	54.5%	69.2%	81.6%
ReIDCaps [28]	18.0%	29.8%	51.6%	64.7%	80.0%
CAL [23]	18.4%	26.7%	50.4%	63.7%	78.4%
TransReID [26]	30.4%	46.7%	71.4%	82.7%	93.1%
CLIP [56]	13.5%	18.6%	38.0%	50.0%	58.8%
CLIP_{proj} [56]	25.3%	32.2%	59.6%	69.8%	81.8%
$\text{CLIP}_{proj,txt}$ [56]	33.6%	**46.8%**	70.2%	80.9%	89.6%
AP3D [22]	**48.1%**	30.3%	**71.6%**	**88.1%**	**98.2%**

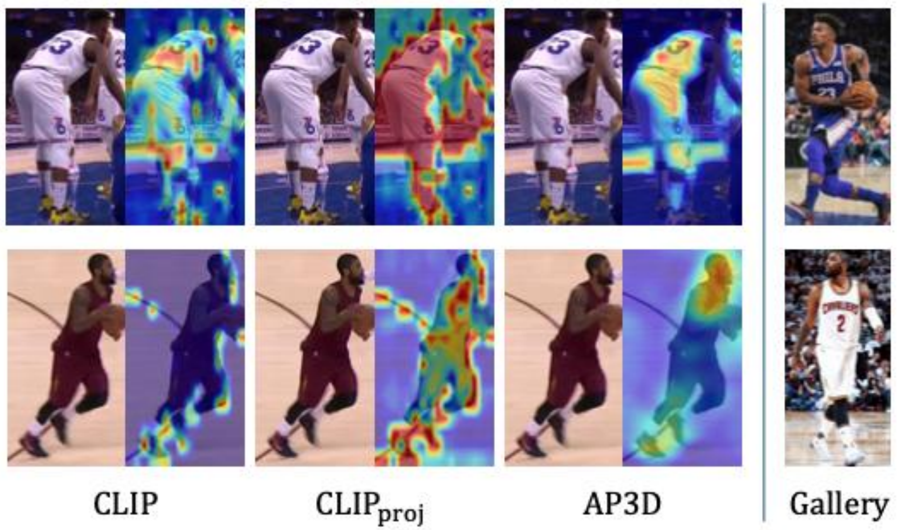


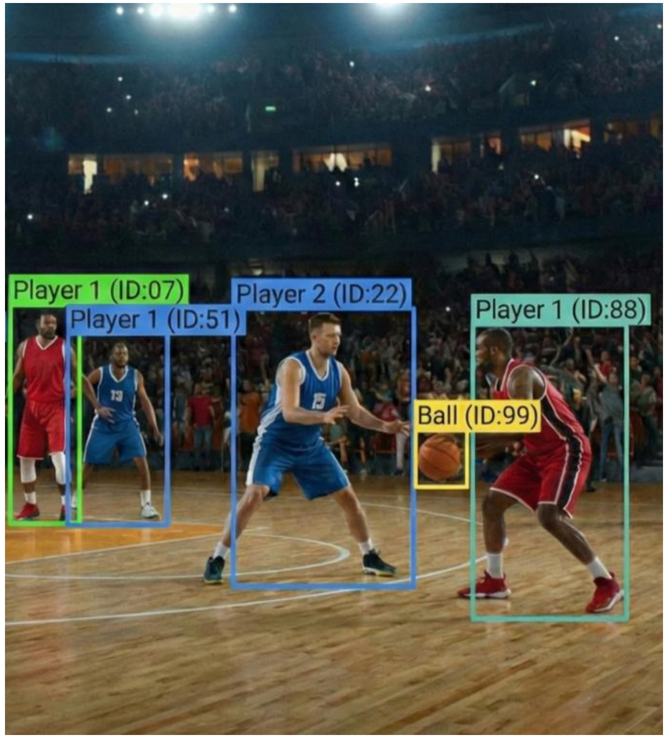
Figure 6. Activation maps on query images (left), when matching with gallery images containing different jersey colors (right).

Results: Object Tracking

Table 3. Results of tracking models on MuPNIT. SORT models use YoloX-X weights trained on *train* (*t*) and *trainval* (*tv*) splits.

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	IDs↓
MASA [39] (RN50)	27.39	48.764	15.825	40.78	20.467	1946
MASA (GDINO)	27.638	48.765	16.11	40.758	20.594	2045
MOTRv1 [80]	30.338	72.676	12.88	87.728	20.46	1579
GTR [88]	31.614	73.618	13.723	88.739	22.36	1088
ByteTrack$_t$ [81]	31.539	73.809	13.574	90.677	22.916	1122
SORT$_t$ [5]	32.118	75.191	13.84	90.729	22.856	1221
MOTRv2 [82]	32.32	73.245	14.393	86.783	23.704	**850**
HybridSORT$_t$ [76]	32.518	75.284	14.208	90.711	23.033	1138
ByteTrack$_{tv}$	32.859	77.208	14.046	**93.925**	24.138	958
SORT$_{tv}$	32.994	78.261	13.975	93.236	23.375	1057
OC-SORT$_t$ [6]	33.349	78.725	14.231	93.248	23.621	1001
HybridSORT$_{tv}$	33.451	78.522	14.363	93.218	23.533	1009
OC-SORT$_{tv}$ (OC)	33.496	78.761	14.359	93.263	23.669	1002
OC$_{tv}$+CLIP	33.536	78.742	14.395	93.286	23.712	998
OC$_{tv}$+AP3D	**33.943**	**78.769**	**14.741**	93.286	**24.140**	997

- Both CLIP family and AP3D help the strongest baseline, OC-SORT



Results: Object Tracking

- GL significantly improves ID tracking

Table 5. Identity-aware tracking results with various ID modules, with/without global linking, across temporal windows (1, 60, 240).

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	FTFI↓	FTFIc↓	IDsw↓	MCorr↑	mAP↑
Baseline [6]	33.496	78.761	14.359	93.263	23.669	-	-	-	-	-
Global **context (All players in the test set)**										
GL$_{CLIP}$	35.642	78.403	16.309	93.369	31.241	**42.097**	**92.783**	123.180	34.7	19.0
GL$_{AP3D}$	**38.384**	**78.404**	**18.900**	93.395	**35.512**	72.147	105.145	142.688	41.0	23.8
CLIP+GL$_{60}$	36.616	78.284	17.235	93.396	32.973	101.174	182.076	14.500	56.2	23.5
AP3D+GL$_{60}$	38.035	78.394	18.559	93.403	34.088	101.887	152.601	11.883	61.7	30.6
AP3D+GL$_{240}$	38.310	78.314	18.845	**93.418**	34.890	116.199	180.034	**7.151**	**72.8**	**31.0**
Game **context (Only players in the game)**										
GL$_{CLIP}$	34.419	78.672	15.166	93.295	26.36	**16.264**	29.808	118.825	40.6	30.2
GL$_{AP3D}$	34.577	**78.719**	15.298	93.311	26.598	17.152	**28.935**	88.740	52.4	39.5
CLIP+GL$_{60}$	38.729	78.134	19.298	93.414	37.106	117.046	177.122	19.101	60.1	38.0
AP3D+GL$_{240}$	40.703	78.205	21.282	**93.463**	41.019	101.472	159.487	**9.994**	**77.7**	46.0
AP3D+GL$_{60}$	**41.959**	78.032	**22.655**	93.452	**42.489**	85.004	133.528	15.416	68.8	**46.1**

Results: Object Tracking

Table 5. Identity-aware tracking results with various ID modules, with/without global linking, across temporal windows (1, 60, 240).

Global

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	FTFI↓	FTFIc↓	IDsw↓	MCorr↑	mAP↑
Baseline [6]	33.496	78.761	14.359	93.263	23.669	-	-	-	-	-
Global **context (All players in the test set)**										
GL_{CLIP}	35.642	78.403	16.309	93.369	31.241	**42.097**	**92.783**	123.180	34.7	19.0
GL_{AP3D}	**38.384**	**78.404**	**18.900**	93.395	**35.512**	72.147	105.145	142.688	41.0	23.8
CLIP+GL_{60}	36.616	78.284	17.235	93.396	32.973	101.174	182.076	14.500	56.2	23.5
AP3D+GL_{60}	38.035	78.394	18.559	93.403	34.088	101.887	152.601	11.883	61.7	30.6
AP3D+GL_{240}	38.310	78.314	18.845	**93.418**	34.890	116.199	180.034	**7.151**	**72.8**	**31.0**
Game **context (Only players in the game)**										
GL_{CLIP}	34.419	78.672	15.166	93.295	26.36	**16.264**	29.808	118.825	40.6	30.2
GL_{AP3D}	34.577	**78.719**	15.298	93.311	26.598	17.152	**28.935**	88.740	52.4	39.5
CLIP+GL_{60}	38.729	78.134	19.298	93.414	37.106	117.046	177.122	19.101	60.1	38.0
AP3D+GL_{240}	40.703	78.205	21.282	**93.463**	41.019	101.472	159.487	**9.994**	**77.7**	46.0
AP3D+GL_{60}	**41.959**	78.032	**22.655**	93.452	**42.489**	85.004	133.528	15.416	68.8	**46.1**

- Performance is relatively insensitive to window size

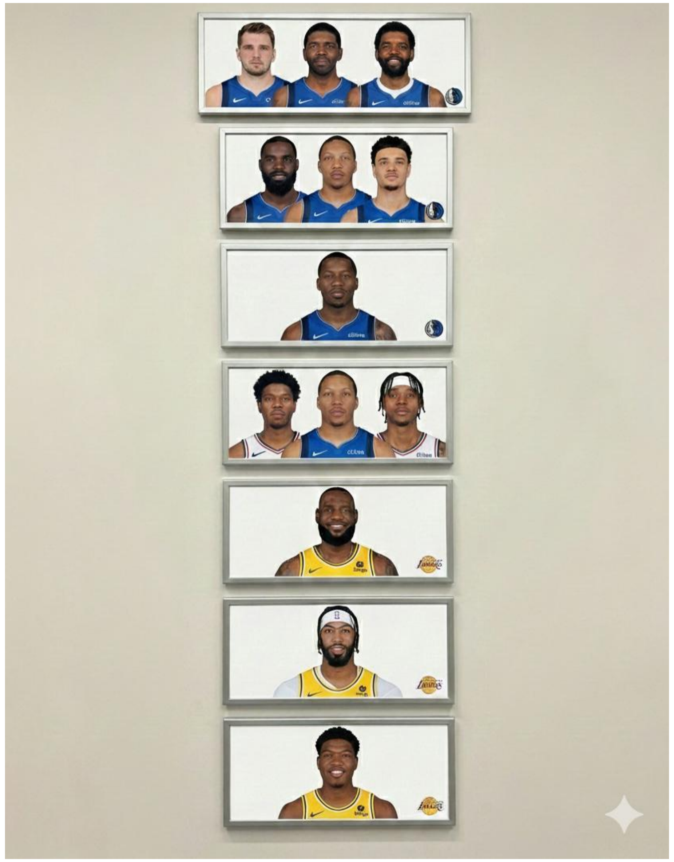
Results: Object Tracking

Table 5. Identity-aware tracking results with various ID modules, with/without global linking, across temporal windows (1, 60, 240).

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	FTFI↓	FTFIc↓	IDsw↓	MCorr↑	mAP↑
Baseline [6]	33.496	78.761	14.359	93.263	23.669	-	-	-	-	-
Global **context (All players in the test set)**										
GL_{CLIP}	35.642	78.403	16.309	93.369	31.241	**42.097**	**92.783**	123.180	34.7	19.0
GL_{AP3D}	**38.384**	**78.404**	**18.900**	93.395	**35.512**	72.147	105.145	142.688	41.0	23.8
CLIP+GL_{60}	36.616	78.284	17.235	93.396	32.973	101.174	182.076	14.500	56.2	23.5
AP3D+GL_{60}	38.035	78.394	18.559	93.403	34.088	101.887	152.601	11.883	61.7	30.6
AP3D+GL_{240}	38.310	78.314	18.845	**93.418**	34.890	116.199	180.034	**7.151**	**72.8**	**31.0**
Game **context (Only players in the game)**										
GL_{CLIP}	34.419	78.672	15.166	93.295	26.36	**16.264**	29.808	118.825	40.6	30.2
GL_{AP3D}	34.577	**78.719**	15.298	93.311	26.598	17.152	**28.935**	88.740	52.4	39.5
CLIP+GL_{60}	38.729	78.134	19.298	93.414	37.106	117.046	177.122	19.101	60.1	38.0
AP3D+GL_{240}	40.703	78.205	21.282	**93.463**	41.019	101.472	159.487	**9.994**	**77.7**	46.0
AP3D+GL_{60}	**41.959**	78.032	**22.655**	93.452	**42.489**	85.004	133.528	15.416	68.8	**46.1**

Local (Game)

- The benefit of increasing window size is non-monotonic



Results: Object Tracking

Table 5. Identity-aware tracking results with various ID modules, with/without global linking, across temporal windows (1, 60, 240).

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	FTFI↓	FTFIc↓	IDsw↓	MCorr↑	mAP↑
Baseline [6]	33.496	78.761	14.359	93.263	23.669	-	-	-	-	-
Global context (All players in the test set)										
GL$_{CLIP}$	35.642	78.403	16.309	93.369	31.241	**42.097**	**92.783**	123.180	34.7	19.0
GL$_{AP3D}$	**38.384**	**78.404**	**18.900**	93.395	**35.512**	72.147	105.145	142.688	41.0	23.8
CLIP+GL$_{60}$	36.616	78.284	17.235	93.396	32.973	101.174	182.076	14.500	56.2	23.5
AP3D+GL$_{60}$	38.035	78.394	18.559	93.403	34.088	101.887	152.601	11.883	61.7	30.6
AP3D+GL$_{240}$	38.310	78.314	18.845	**93.418**	34.890	116.199	180.034	**7.151**	**72.8**	**31.0**
Game context (Only players in the game)										
GL$_{CLIP}$	34.419	78.672	15.166	93.295	26.36	**16.264**	29.808	118.825	40.6	30.2
GL$_{AP3D}$	34.577	**78.719**	15.298	93.311	26.598	17.152	**28.935**	88.740	52.4	39.5
CLIP+GL$_{60}$	38.729	78.134	19.298	93.414	37.106	117.046	177.122	19.101	60.1	38.0
AP3D+GL$_{240}$	40.703	78.205	21.282	**93.463**	41.019	101.472	159.487	**9.994**	**77.7**	46.0
AP3D+GL$_{60}$	**41.959**	78.032	**22.655**	93.452	**42.489**	85.004	133.528	15.416	68.8	**46.1**

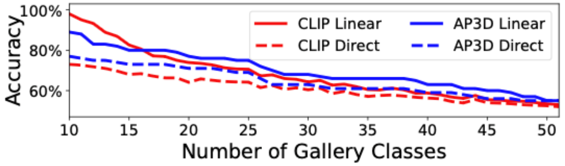
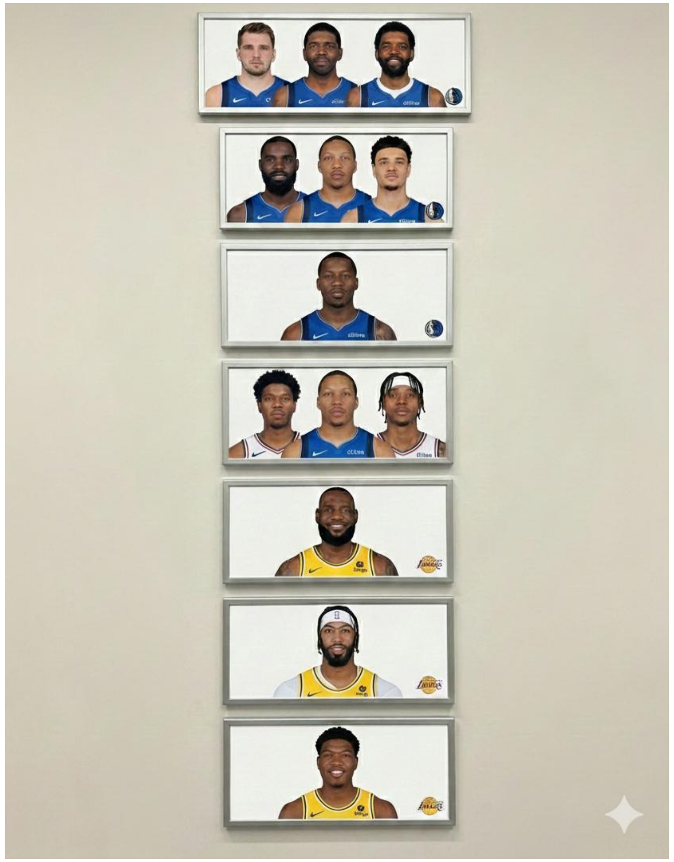


Figure 7. Classification accuracies of 10 fixed query classes vs. number of gallery classes for CLIP$_{proj}$ and AP3D models using direct and linear assignment strategies.

- Linear assignment consistently outperforms direct matching



Results: Generalization

Table 9. Generalization performance on MuPNIT, SportsMOT, and SoccerNet datasets using 500 general text embeddings.

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	IDs↓
MuPNIT test split						
Base [6]	33.496	**78.761**	14.359	93.263	23.669	1002
Base+GL	**35.053**	78.618	**15.742**	**93.334**	**27.866**	**696**
Gain	+4.6%	-0.2%	+9.6%	+0.1%	+17.7%	-30.5%
SportsMOT val split (basketball)						
Base [6]	73.361	**89.609**	60.236	97.318	74.790	331
Base+GL	**73.502**	89.564	**60.507**	**97.319**	**75.169**	**314**
Gain	+0.2%	-0.1%	+0.4%	-	+0.5%	-5.1%
SportsMOT val split (football)						
Base [6]	67.425	84.904	54.518	90.531	65.696	1030
Base+GL	**68.243**	**85.177**	**55.666**	**90.845**	**67.816**	**830**
Gain	+1.2%	+0.3%	+2.1%	+0.4%	+3.2%	-19.4%
SportsMOT val split (volleyball)						
Base [6]	69.979	86.104	57.756	92.503	71.034	493
Base+GL	**71.104**	**86.553**	**59.366**	**92.879**	**72.943**	**416**
Gain	+1.6%	+0.5%	+2.8%	+0.4%	+2.7%	-15.6%
SoccerNet test split						
Base [6]	**52.864**	**71.909**	**39.034**	**85.908**	55.432	5932
Base+GL	51.917	71.700	37.746	85.209	**56.171**	**3823**
Gain	-1.8%	-0.3%	-3.3%	-0.8%	+1.3%	-35.6%

- We report the generalizability of GL on multiple benchmarks

A player with yellow jersey and number 30
A player with white jersey and number 07
A player with red jersey and number 77
A player with blue jersey and number 24
A player with green jersey and number 55
A player with yellow jersey and number 01
A player with white jersey and number 68
A player with red jersey and number 19
A player with blue jersey and number 33
A player with green jersey and number 91
A player with yellow jersey and number 15
A player with white jersey and number 42

Results: Generalization

Table 9. Generalization performance on MuPNIT, SportsMOT, and SoccerNet datasets using 500 general text embeddings.

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	IDs↓
MuPNIT test split						
Base [6]	33.496	**78.761**	14.359	93.263	23.669	1002
Base+GL	**35.053**	78.618	**15.742**	**93.334**	**27.866**	**696**
Gain	+4.6%	-0.2%	+9.6%	+0.1%	+17.7%	-30.5%
SportsMOT val split (basketball)						
Base [6]	73.361	**89.609**	60.236	97.318	74.790	331
Base+GL	**73.502**	89.564	**60.507**	**97.319**	**75.169**	**314**
Gain	+0.2%	-0.1%	+0.4%	-	+0.5%	-5.1%
SportsMOT val split (football)						
Base [6]	67.425	84.904	54.518	90.531	65.696	1030
Base+GL	**68.243**	**85.177**	**55.666**	**90.845**	**67.816**	**830**
Gain	+1.2%	+0.3%	+2.1%	+0.4%	+3.2%	-19.4%
SportsMOT val split (volleyball)						
Base [6]	69.979	86.104	57.756	92.503	71.034	493
Base+GL	**71.104**	**86.553**	**59.366**	**92.879**	**72.943**	**416**
Gain	+1.6%	+0.5%	+2.8%	+0.4%	+2.7%	-15.6%
SoccerNet test split						
Base [6]	**52.864**	**71.909**	**39.034**	**85.908**	55.432	5932
Base+GL	51.917	71.700	37.746	85.209	**56.171**	**3823**
Gain	-1.8%	-0.3%	-3.3%	-0.8%	+1.3%	-35.6%

- GL consistently improves ID tracking in MuPNIT and SportsMOT

A player with yellow jersey and number 30
A player with white jersey and number 07
A player with red jersey and number 77
A player with blue jersey and number 24
A player with green jersey and number 55
A player with yellow jersey and number 01
A player with white jersey and number 68
A player with red jersey and number 19
A player with blue jersey and number 33
A player with green jersey and number 91
A player with yellow jersey and number 15
A player with white jersey and number 42

Results: Generalization

Table 9. Generalization performance on MuPNIT, SportsMOT, and SoccerNet datasets using 500 general text embeddings.

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	IDs↓
MuPNIT test split						
Base [6]	33.496	**78.761**	14.359	93.263	23.669	1002
Base+GL	**35.053**	78.618	**15.742**	**93.334**	**27.866**	**696**
Gain	+4.6%	-0.2%	+9.6%	+0.1%	+17.7%	-30.5%
SportsMOT val split (basketball)						
Base [6]	73.361	**89.609**	60.236	97.318	74.790	331
Base+GL	**73.502**	89.564	**60.507**	**97.319**	**75.169**	**314**
Gain	+0.2%	-0.1%	+0.4%	-	+0.5%	-5.1%
SportsMOT val split (football)						
Base [6]	67.425	84.904	54.518	90.531	65.696	1030
Base+GL	**68.243**	**85.177**	**55.666**	**90.845**	**67.816**	**830**
Gain	+1.2%	+0.3%	+2.1%	+0.4%	+3.2%	-19.4%
SportsMOT val split (volleyball)						
Base [6]	69.979	86.104	57.756	92.503	71.034	493
Base+GL	**71.104**	**86.553**	**59.366**	**92.879**	**72.943**	**416**
Gain	+1.6%	+0.5%	+2.8%	+0.4%	+2.7%	-15.6%
SoccerNet test split						
Base [6]	**52.864**	**71.909**	**39.034**	**85.908**	55.432	5932
Base+GL	51.917	71.700	37.746	85.209	**56.171**	**3823**
Gain	-1.8%	-0.3%	-3.3%	-0.8%	+1.3%	-35.6%

- GL consistently improves ID tracking in MuPNIT and SportsMOT

A player with yellow jersey and number 30
A player with white jersey and number 07
A player with red jersey and number 77
A player with blue jersey and number 24
A player with green jersey and number 55
A player with yellow jersey and number 01
A player with white jersey and number 68
A player with red jersey and number 19
A player with blue jersey and number 33
A player with green jersey and number 91
A player with yellow jersey and number 15
A player with white jersey and number 42

Results: Generalization

Table 9. Generalization performance on MuPNIT, SportsMOT, and SoccerNet datasets using 500 general text embeddings.

Model	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	IDs↓
MuPNIT test split						
Base [6]	33.496	**78.761**	14.359	93.263	23.669	1002
Base+GL	**35.053**	78.618	**15.742**	**93.334**	**27.866**	**696**
Gain	+4.6%	-0.2%	+9.6%	+0.1%	+17.7%	-30.5%
SportsMOT val split (basketball)						
Base [6]	73.361	**89.609**	60.236	97.318	74.790	331
Base+GL	**73.502**	89.564	**60.507**	**97.319**	**75.169**	**314**
Gain	+0.2%	-0.1%	+0.4%	-	+0.5%	-5.1%
SportsMOT val split (football)						
Base [6]	67.425	84.904	54.518	90.531	65.696	1030
Base+GL	**68.243**	**85.177**	**55.666**	**90.845**	**67.816**	**830**
Gain	+1.2%	+0.3%	+2.1%	+0.4%	+3.2%	-19.4%
SportsMOT val split (volleyball)						
Base [6]	69.979	86.104	57.756	92.503	71.034	493
Base+GL	**71.104**	**86.553**	**59.366**	**92.879**	**72.943**	**416**
Gain	+1.6%	+0.5%	+2.8%	+0.4%	+2.7%	-15.6%
SoccerNet test split						
Base [6]	**52.864**	**71.909**	**39.034**	**85.908**	55.432	5932
Base+GL	51.917	71.700	37.746	85.209	**56.171**	**3823**
Gain	-1.8%	-0.3%	-3.3%	-0.8%	+1.3%	-35.6%

- GL yields clear gains in IDF1 and IDs for SoccerNet, despite incomplete IDs and lack of a global ID bank here

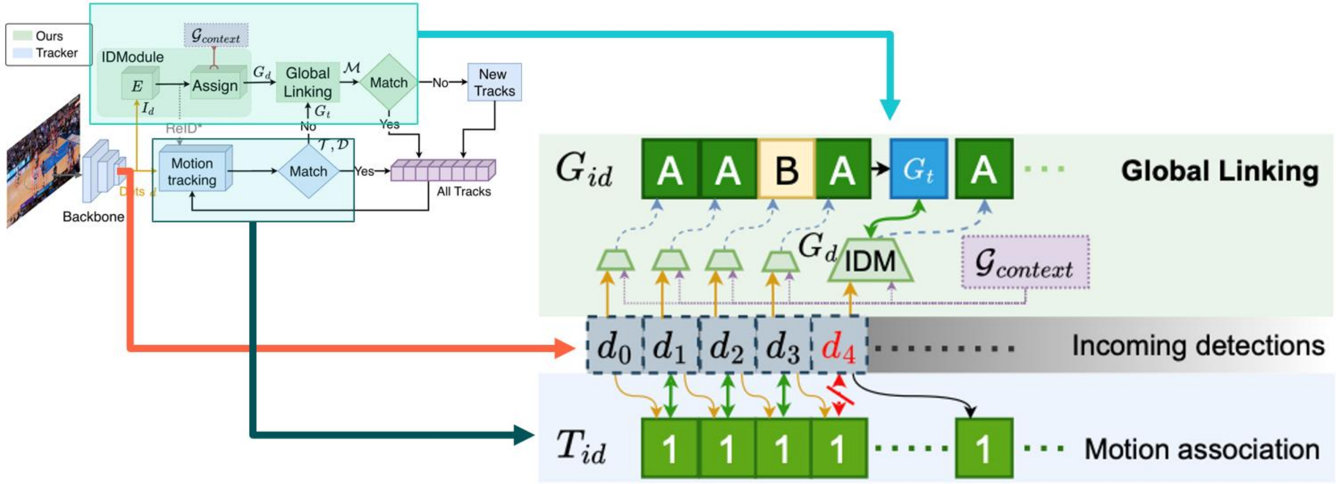
A player with yellow jersey and number 30
A player with white jersey and number 07
A player with red jersey and number 77
A player with blue jersey and number 24
A player with green jersey and number 55
A player with yellow jersey and number 01
A player with white jersey and number 68
A player with red jersey and number 19
A player with blue jersey and number 33
A player with green jersey and number 91
A player with yellow jersey and number 15
A player with white jersey and number 42

Our Contributions

- Two datasets for multi-perspective ID-aware MOT
- GIF, a generalizable, online tracking-with-ID approach
- Five novel Global ID tracking metrics evaluating ID prediction tradeoffs in multi-perspective MOT

Global ID database

In-game player crops



Advancing Player Identification and Tracking with Global ID Fusion (GIF)

Karol Wojtulewicz*, Minxing Liu*, and Niklas Carlsson

Linköping University, Sweden

LINKÖPING UNIVERSITY

Contact: karol.wojtulewicz@liu.se