



# SearchTrack: Multiple Object Tracking with Object-Customized Search and Motion-Aware Features

Zhong-Min Tsai<sup>1</sup> Yu-Ju Tsai<sup>1</sup> Chien-Yao Wang<sup>2</sup> Hong-Yuan Liao<sup>2</sup> Youn-Long Lin<sup>3</sup> Yung-Yu Chuang<sup>1</sup>  
<sup>1</sup>National Taiwan University <sup>2</sup>Academia Sinica <sup>3</sup>National Tsing Hua University

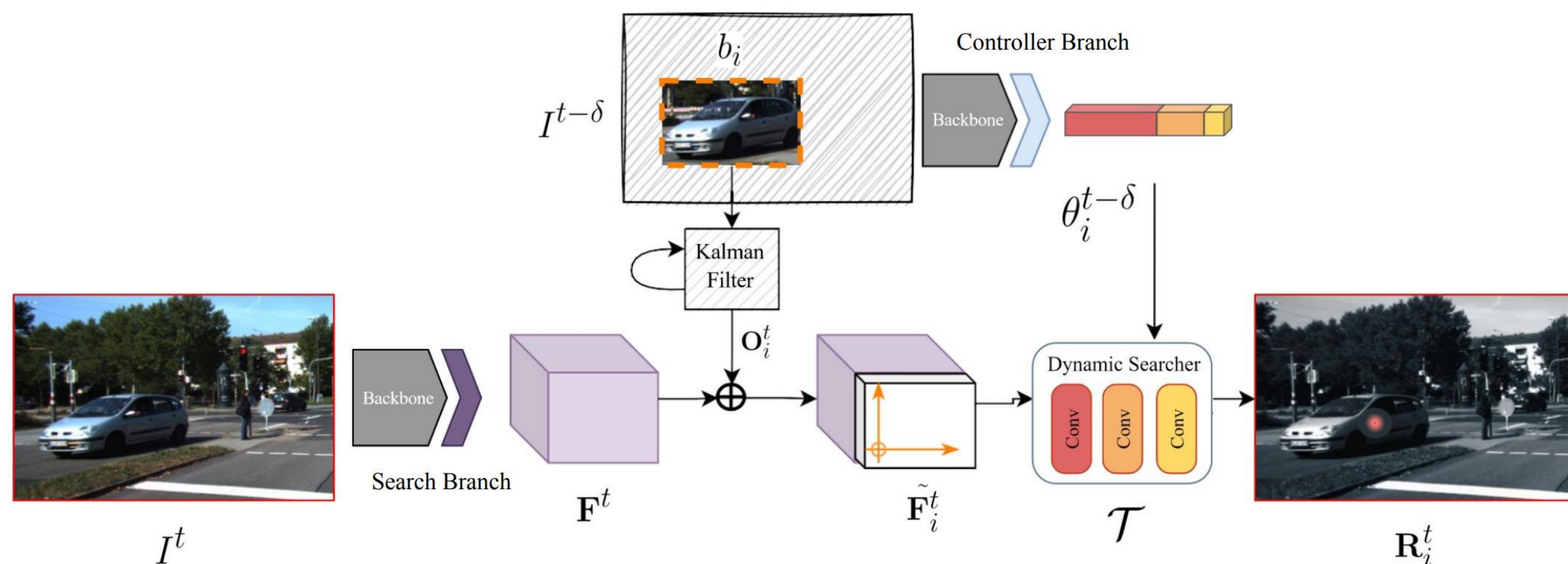


## Contributions:

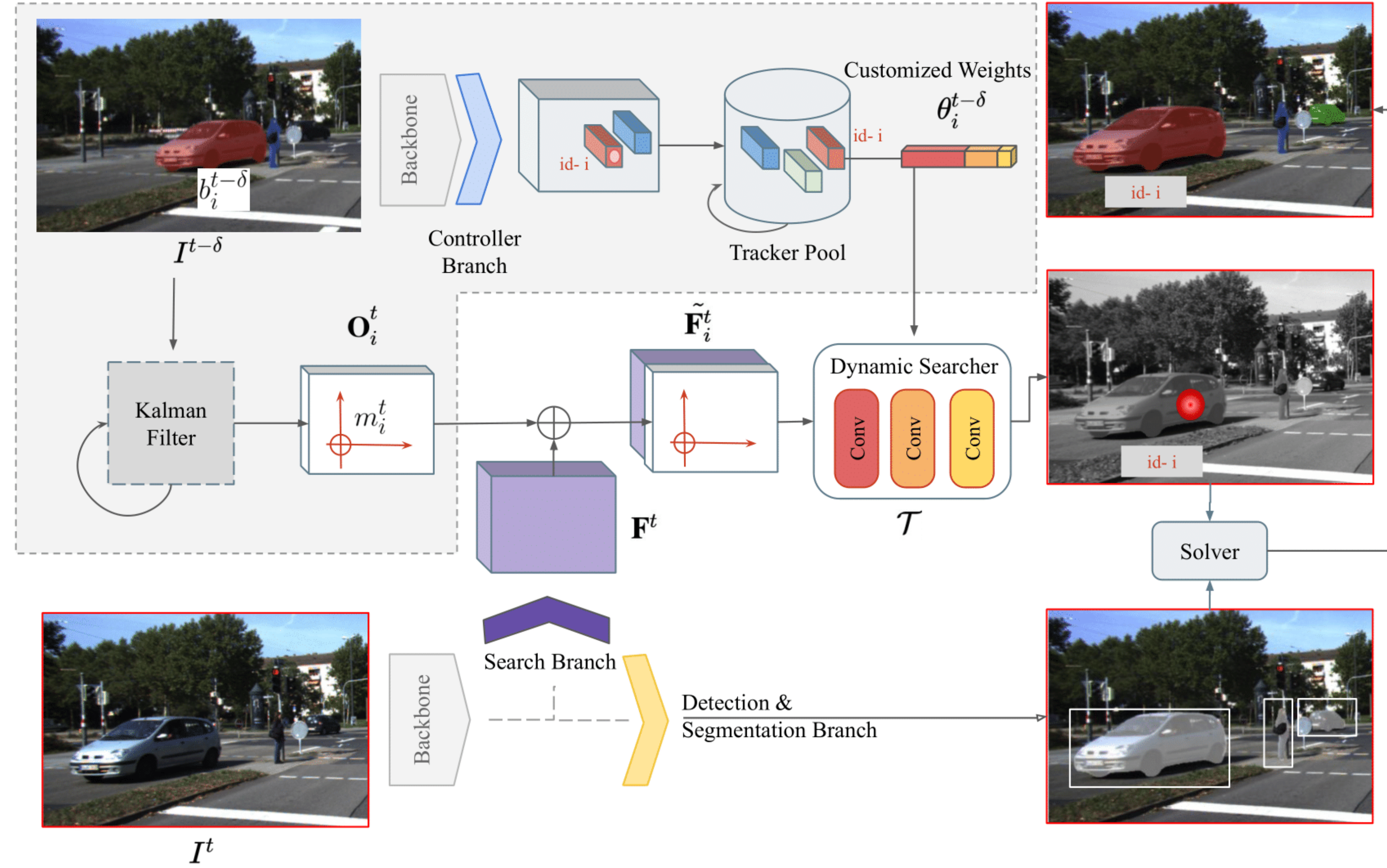
- Propose an algorithm to combine **object appearance** and **motion cues** to solve the **association** problem on the MOT(S).
  - An **online** one-stage point-based tracker
  - Outperform the state-of-the-art **2D online** methods
- Investigate the synergy between **tracking** and **segmentation**
  - Tracking can benefit from the **dense pixel-wise annotations** in our experiments.
  - Networks can learn better separation of **foreground** and **background** in the scene.
- Evaluation on two real-world datasets
  - KITTI MOTS
  - MOT17

## Observation:

- The tracked object in MOT(S) is highly correlated across **consecutive frames**, tracing its trajectory is helpful for tracking.
  - We encode the **object motion** predicted by the **Kalman filter** with the **object appearance** feature to produce a **motion-aware feature**.
  - We propose the **object-customized dynamic searcher** to identify the association for a given object.



## Network Architecture:



## Ablation study on motion-aware feature:

	Pedestrian			Car		
	HOTA↑	DetA↑	AssA↑	HOTA↑	DetA↑	AssA↑
feature without motion	43.9 %	<b>56.4 %</b>	34.4 %	77.1 %	79.3 %	75.5 %
motion-aware feature	<b>59.4 %</b>	56.0 %	<b>63.5 %</b>	<b>78.7 %</b>	<b>79.4 %</b>	<b>78.5 %</b>

## Ablation study on segmentation branch:

	Pedestrian				Car			
	HOTA↑	DetA↑	AssA↑	LocA↑	HOTA↑	DetA↑	AssA↑	LocA↑
w/o segmentation	55.9 %	55.5 %	56.4 %	82.2 %	73.9 %	<b>73.7 %</b>	74.3 %	87.3 %
w/ segmentation	<b>58.2 %</b>	<b>56.8 %</b>	<b>59.8 %</b>	<b>83.8 %</b>	<b>74.5 %</b>	73.3 %	<b>76.0 %</b>	<b>87.6 %</b>

Source code: <https://github.com/qa276390/SearchTrack>

## KITTI MOTS Dataset:

“\*” denotes offline method.

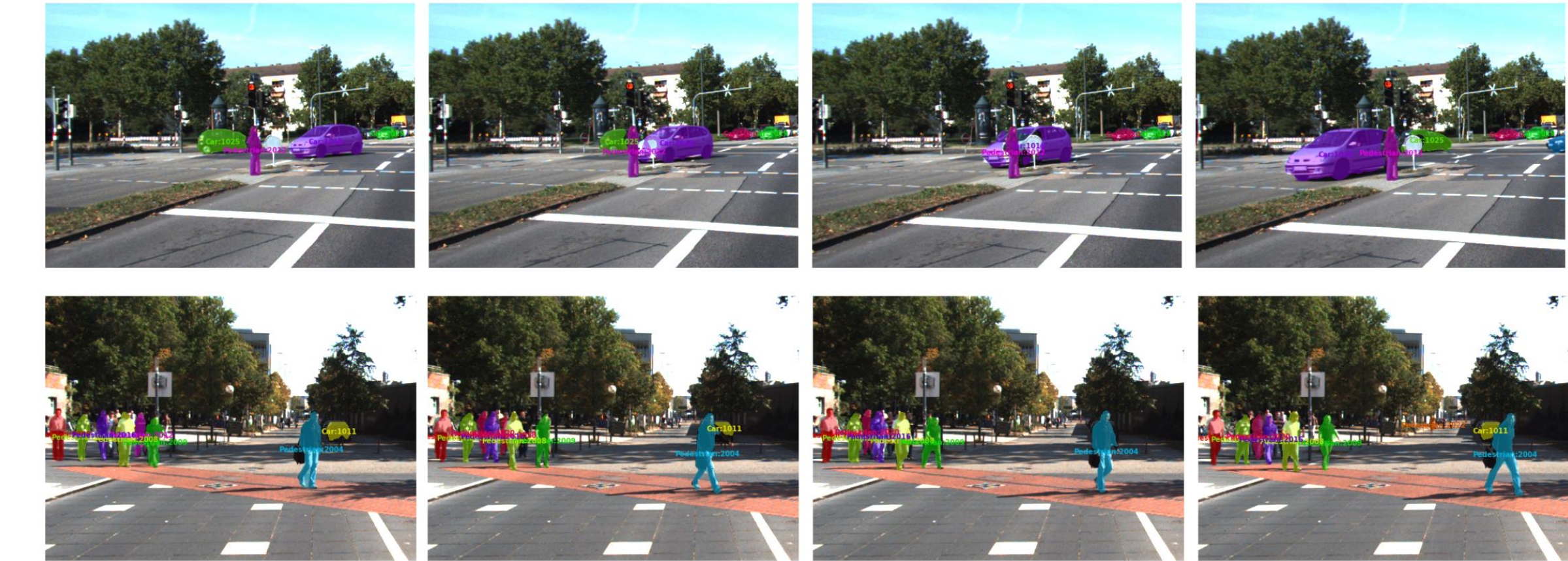
	time	Pedestrian					Car				
		HOTA↑	DetA↑	AssA↑	LocA↑	sMOTSA↑	HOTA↑	DetA↑	AssA↑	LocA↑	sMOTSA↑
TrackRCNN [15]	0.5	41.9	53.8	33.8	78.0	47.3	56.6	69.9	46.5	86.6	67.0
GMPhD_SAF [16]	<b>0.08</b>	49.3	<b>65.5</b>	38.3	<b>83.8</b>	<b>62.9</b>	55.1	77.0	39.8	<b>88.7</b>	75.4
PointTrack [17]	<b>0.05</b>	54.4	62.3	48.1	83.3	61.5	62.0	<b>79.4</b>	48.8	88.5	<b>78.5</b>
ReMOTS* [18]	3	<b>58.8</b>	<b>68.0</b>	<b>52.4</b>	<b>84.2</b>	<b>66.0</b>	<b>71.6</b>	<b>78.3</b>	<b>66.0</b>	<b>89.3</b>	<b>75.9</b>
Ours	0.19	<b>57.6</b>	63.7	<b>53.1</b>	80.9	60.6	<b>71.5</b>	76.8	<b>67.1</b>	88.0	74.9

## MOT17 Dataset:

Method	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	MT↑	ML↓	FP↓	FN↓	IDSW↓
Tracktor++ [19]	44.8	44.9	45.1	56.3	55.1	21.1%	35.3 %	<b>8866</b>	235449	1987
Visual-Spatial [20]	46.4	45.3	47.9	56.8	58.3	22.8%	37.4 %	11567	230645	<b>1320</b>
CenterTrack [21]	48.2	49.0	47.8	61.5	59.6	26.4%	31.9 %	14076	200672	2583
TMOH [22]	50.4	49.6	50.9	62.1	62.8	26.9%	31.4 %	<b>10951</b>	201195	<b>1897</b>
SiamMOT [23]	-	-	-	65.9	63.3	34.6%	23.9%	18098	170955	3040
PermaTrack [24]	<b>54.2</b>	<b>58.0</b>	<b>51.2</b>	<b>73.1</b>	<b>67.2</b>	<b>42.3%</b>	<b>19.1%</b>	24557	<b>123508</b>	3571
Ours	<b>53.4</b>	<b>55.6</b>	<b>51.6</b>	<b>68.0</b>	<b>65.7</b>	<b>39.1%</b>	<b>21.1%</b>	25651	<b>150786</b>	4254

## Visualization:

### ➤ KITTI MOTS



### ➤ MOT 17

