MARS Lab

VIP3D
Visual Trajectory Prediction 3D

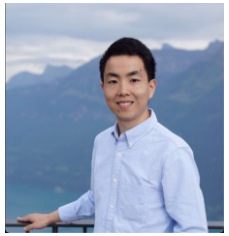
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

ViP3D: End-to-end Visual Trajectory Prediction via 3D Agent Queries




Junru Gu*, Chenxu Hu*, Tianyuan Zhang, Xuanyao Chen,





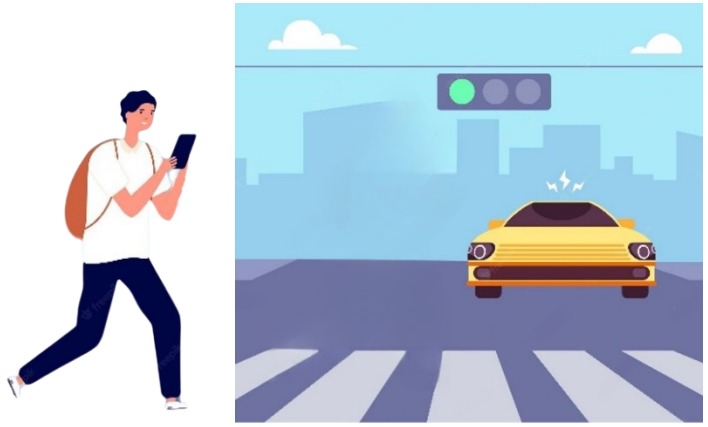
Yilun Wang, Yue Wang, Hang Zhao

TUE-PM-130

https://tsinghua-mars-lab.github.io/ViP3D

Motivation: Why Visual?

- Take advantage of **rich visual information** for motion prediction.

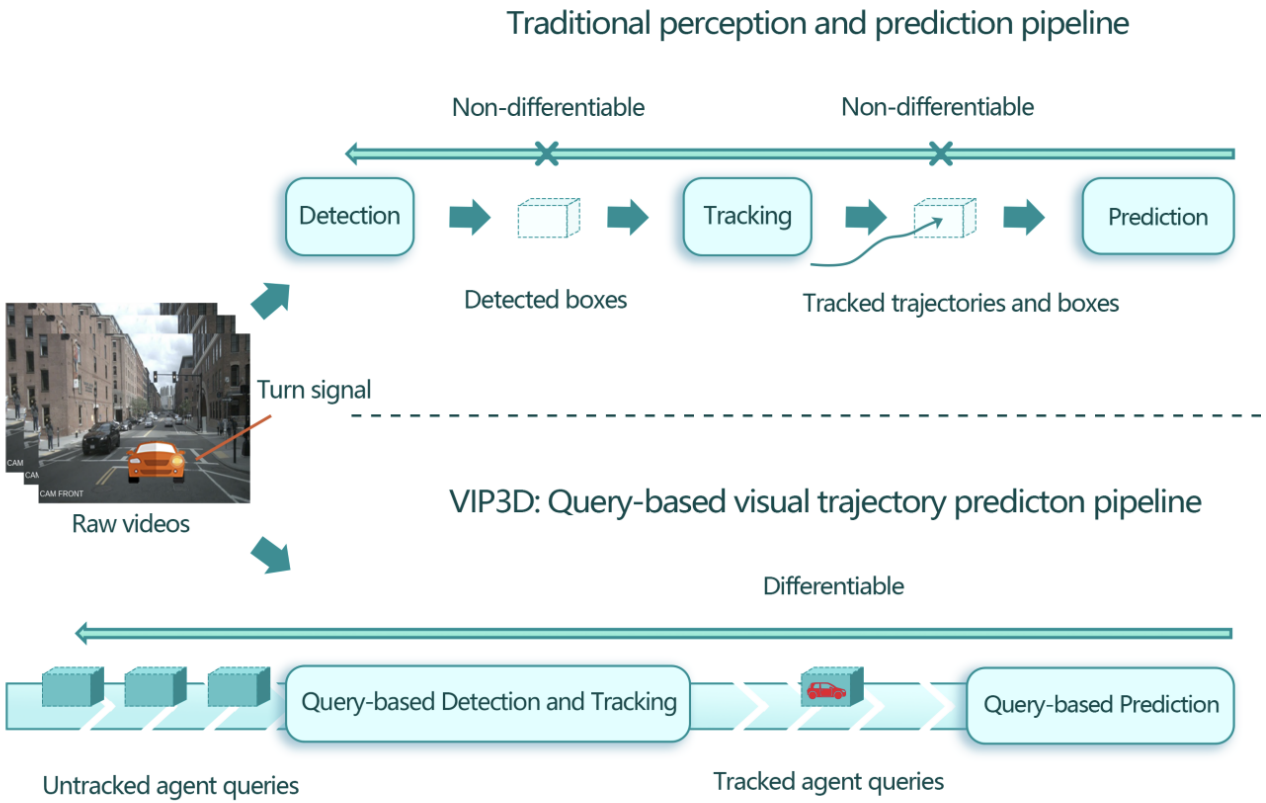


Motivation: Why Visual?

- Take advantage of **rich visual information** for motion prediction.

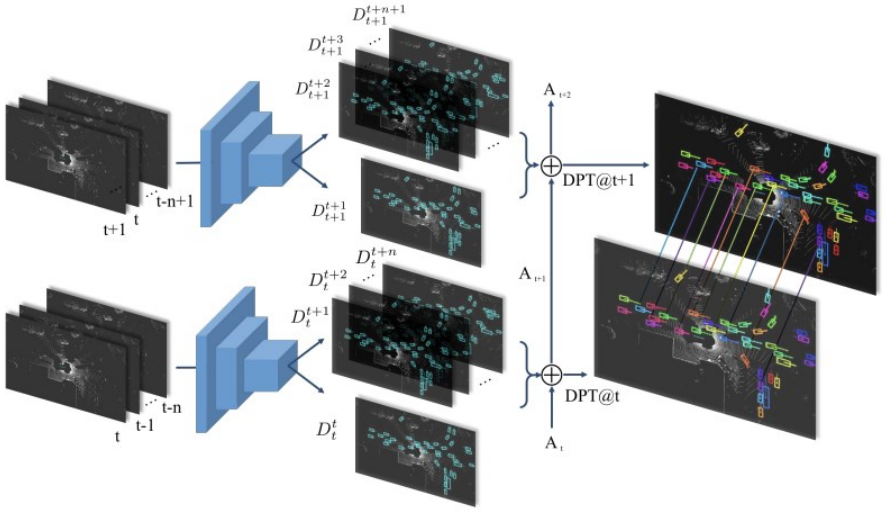
Motivation: Why End-to-End?

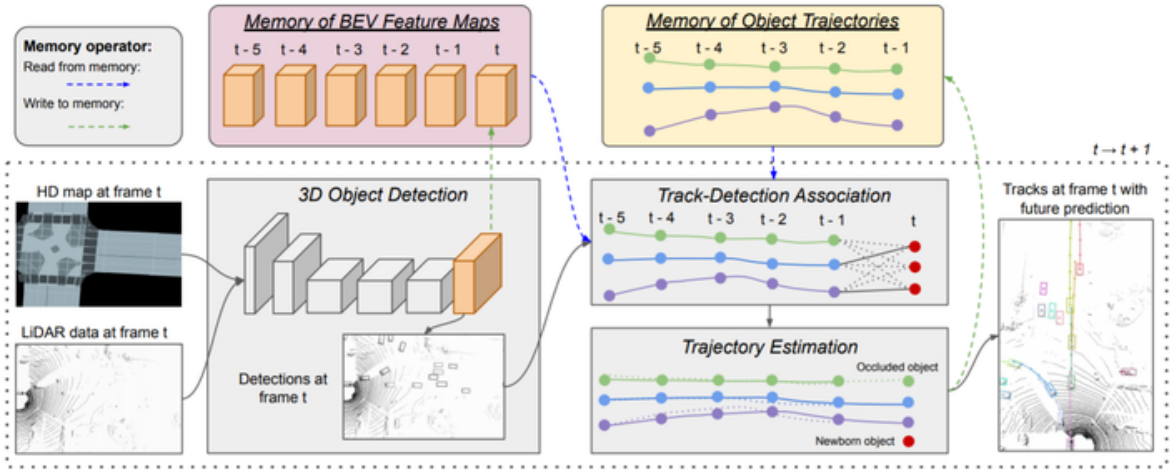
- Avoid **error accumulation** and **data distribution shift.**



Previous End-to-End Prediction Methods

- FaF, CVPR 2018
  - The first CNN model for **joint detection and prediction**, from **LiDAR** inputs.
- PnPNet, CVPR 2020
  - A **tracking-in-the-loop** model for **LiDAR**-based trajectory prediction.





Previous End-to-End Prediction Methods

😔 • Unable able to leverage the abundant **visual information** from cameras

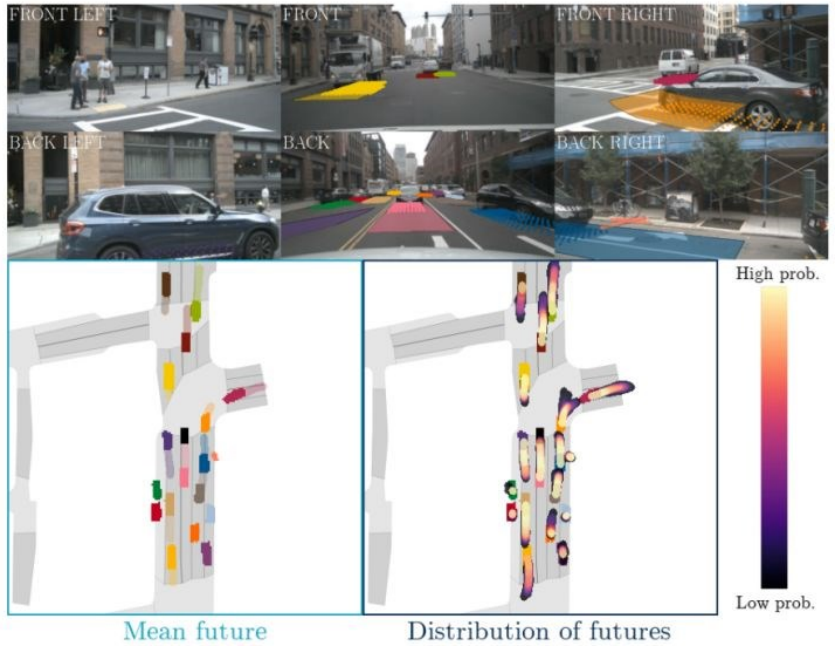
😔 • Use **convolutional feature maps** as their intermediate representations

Suffering from **non-differentiable operations**

- E.g. Non-maximum suppression in object detection
- E.g. Object association in tracking

Previous End-to-End Prediction Methods

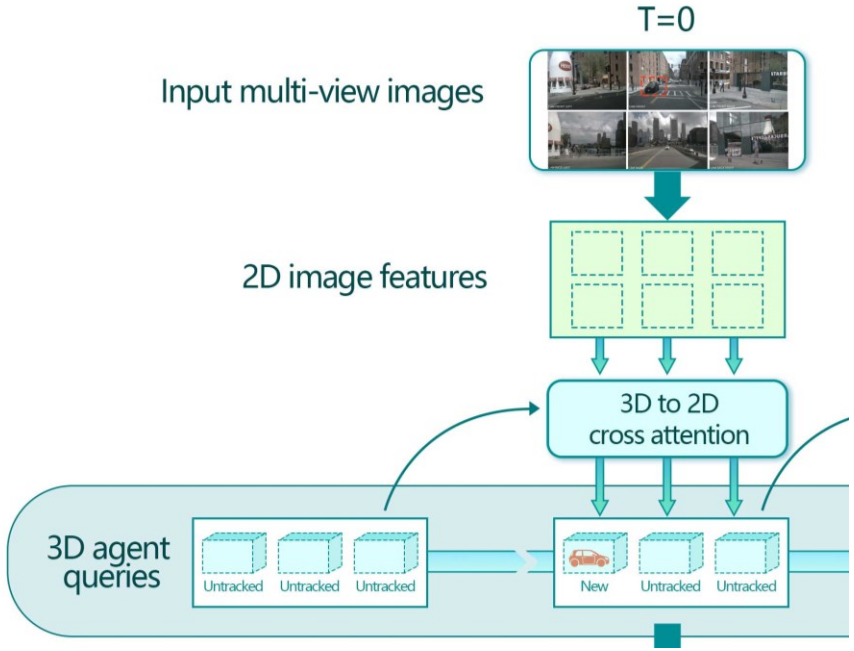
- FIERY, ICCV 2021
  - Predicts future **BEV occupancy heatmaps** from **visual** data.



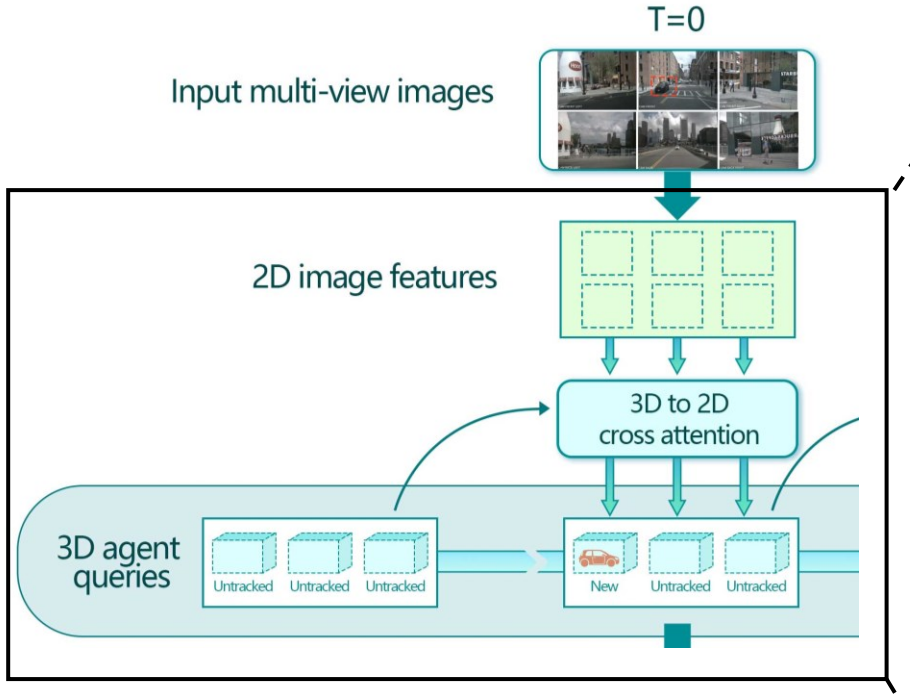
😔
- **Not modularized, non-interpretable**, thus not engineering-wise friendly

😔
- Output representation is **not compatible** with downstream modules like motion planning

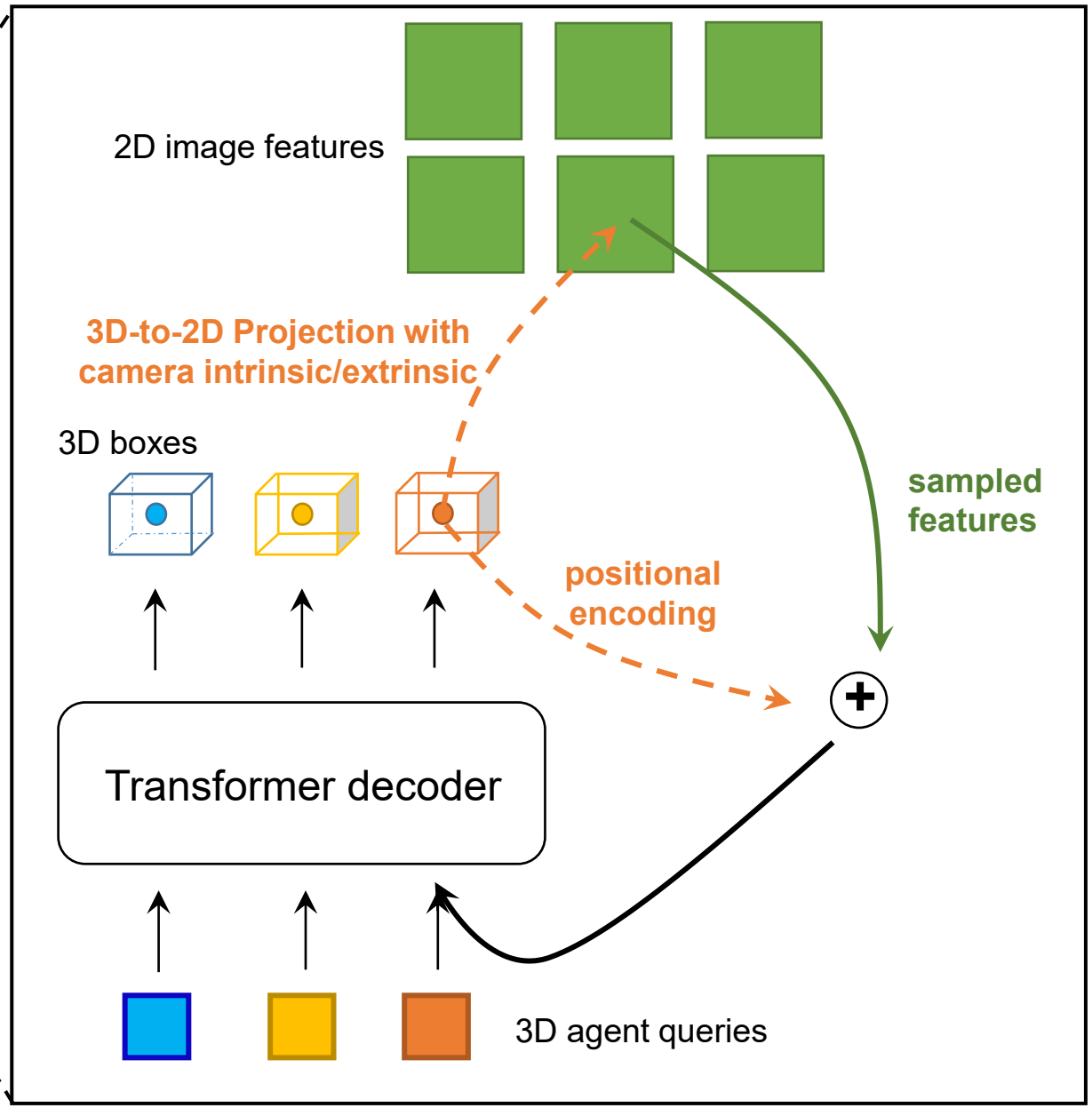
Model Architecture



Model Architecture

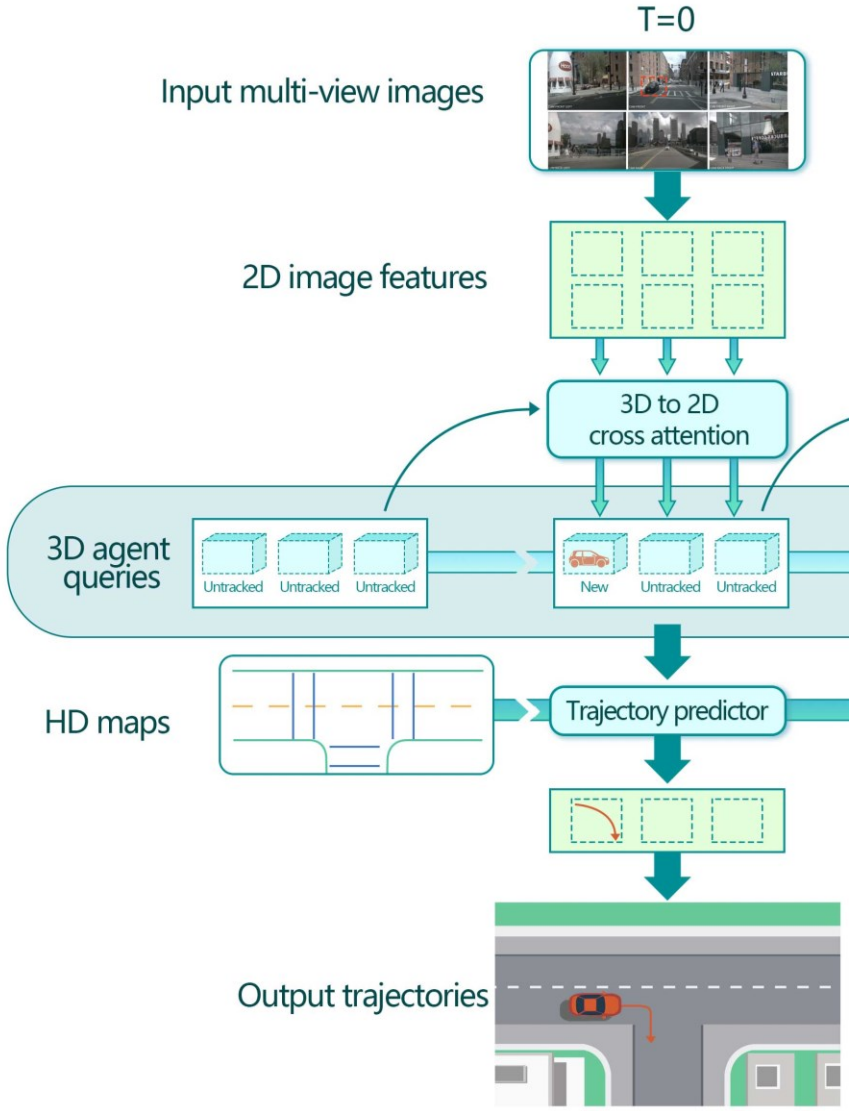


DETR3D

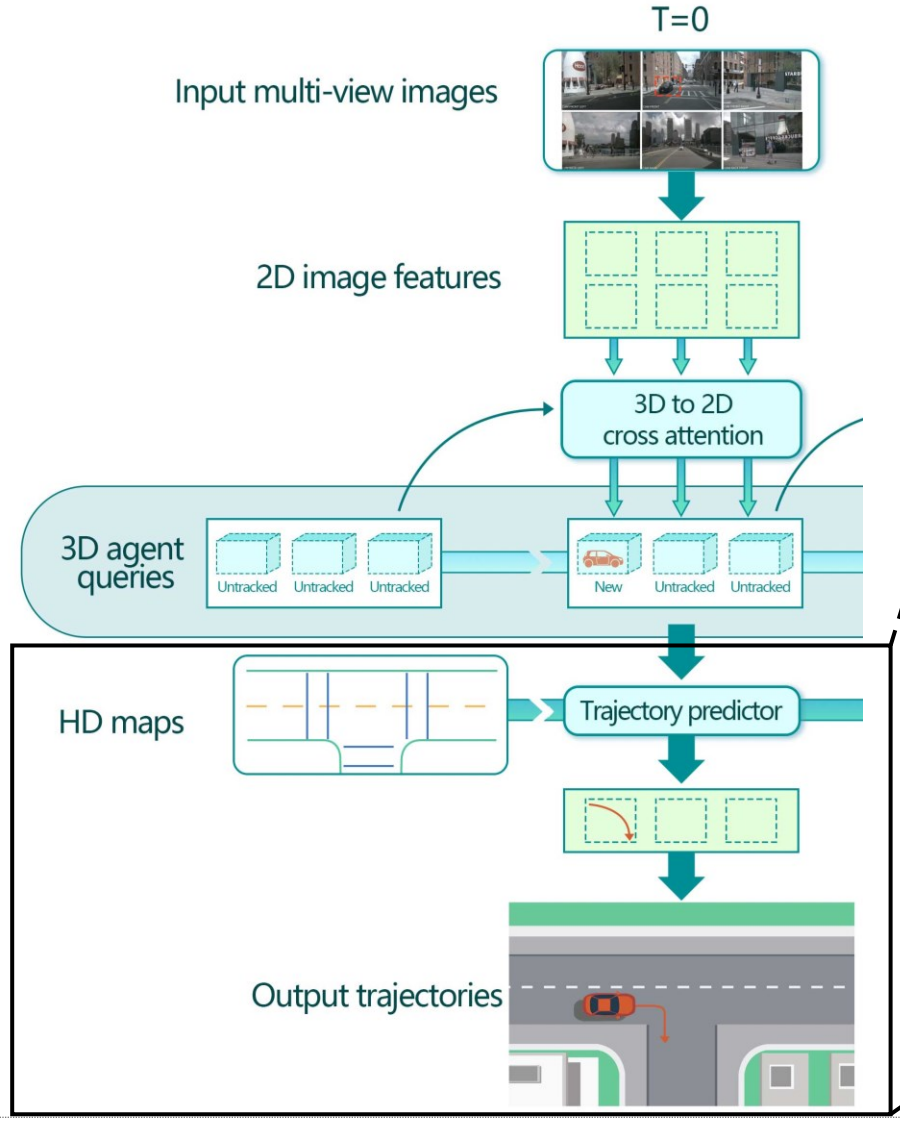


DETR3D: 3D Object Detection from Multi-view Images via 3D-to-2D Queries, Yue Wang, Vitor Guizilini, Tianyuan Zhang, Yilun Wang, Hang Zhao, Justin Solomon, CoRL 2021

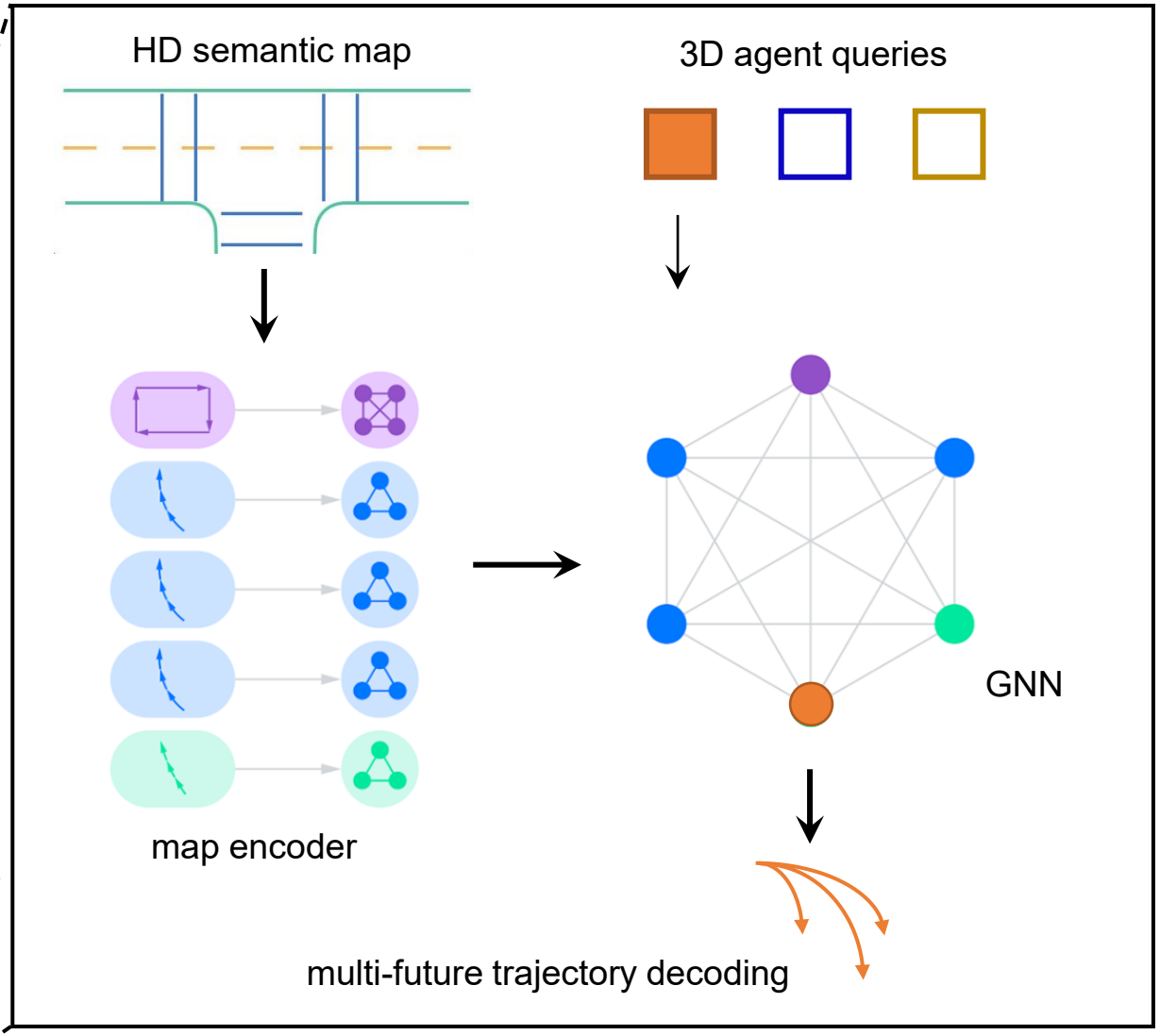
Model Architecture



Model Architecture

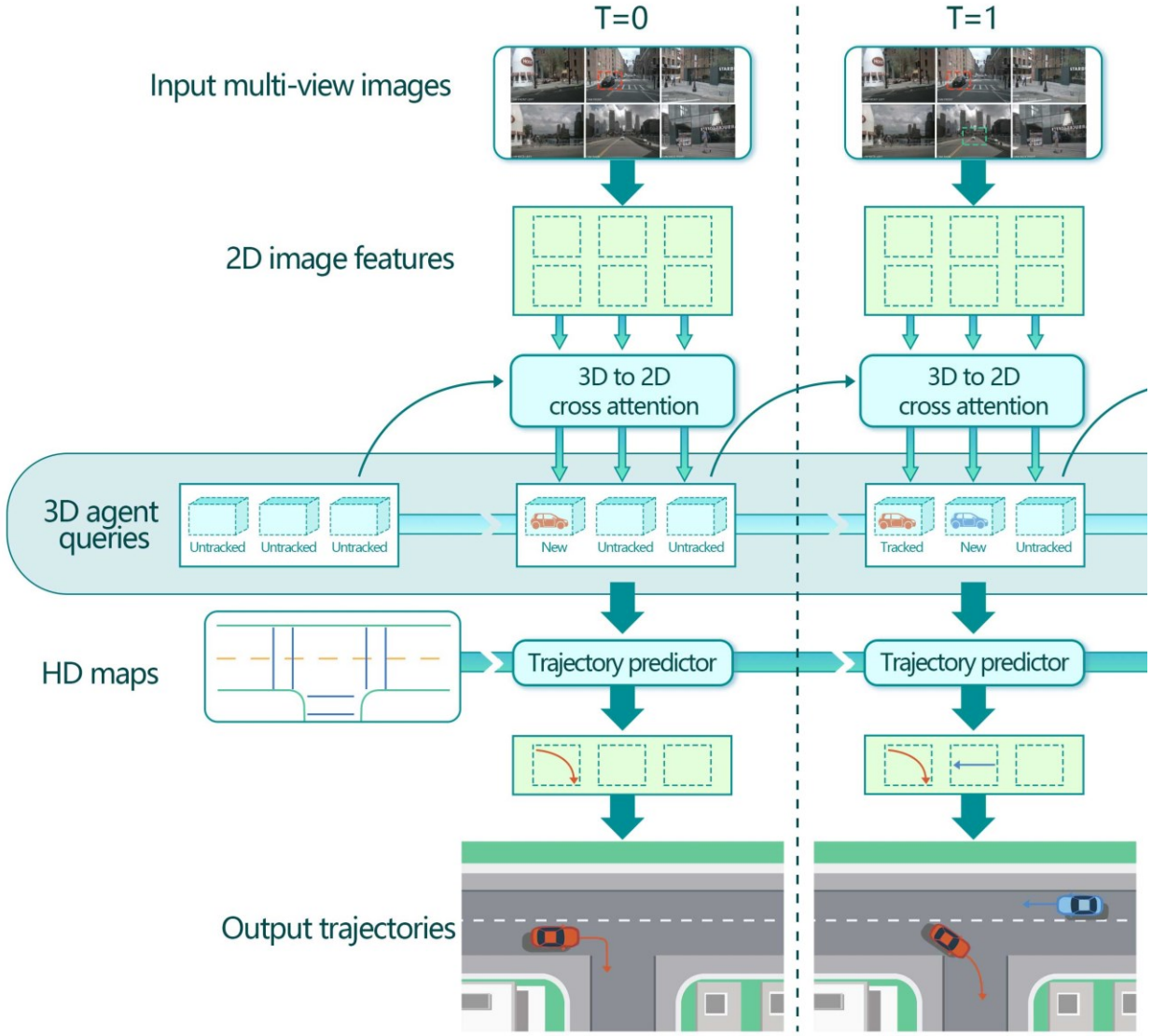


VectorNet

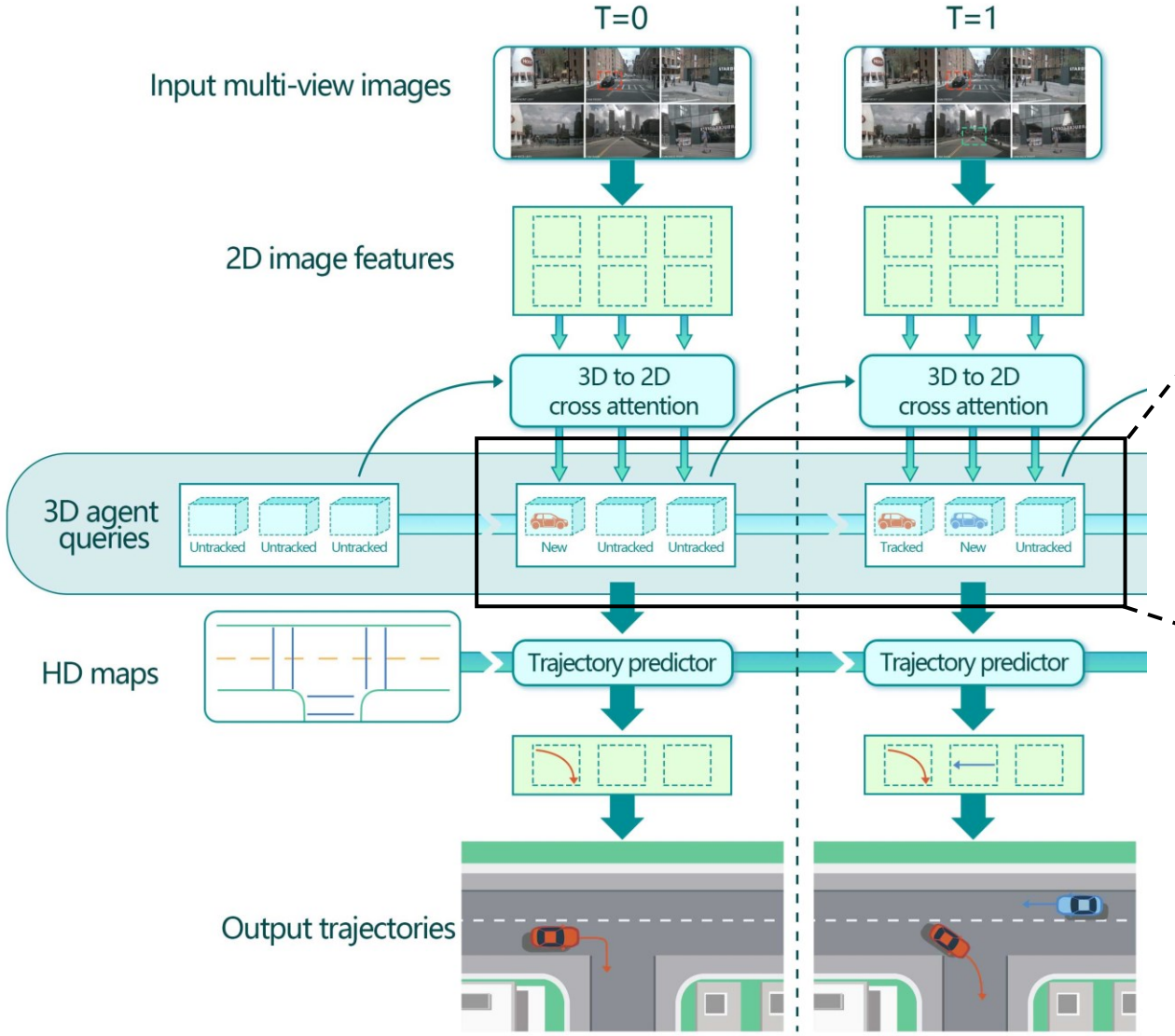


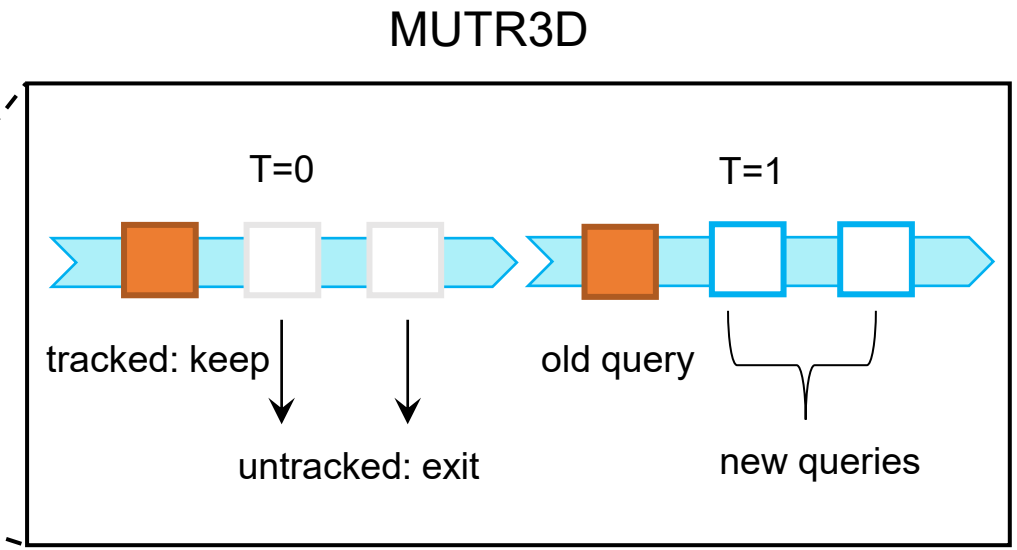
Vectornet: Encoding HD maps and agent dynamics from vectorized representation, Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, Cordelia Schmid, CVPR 2020

Model Architecture



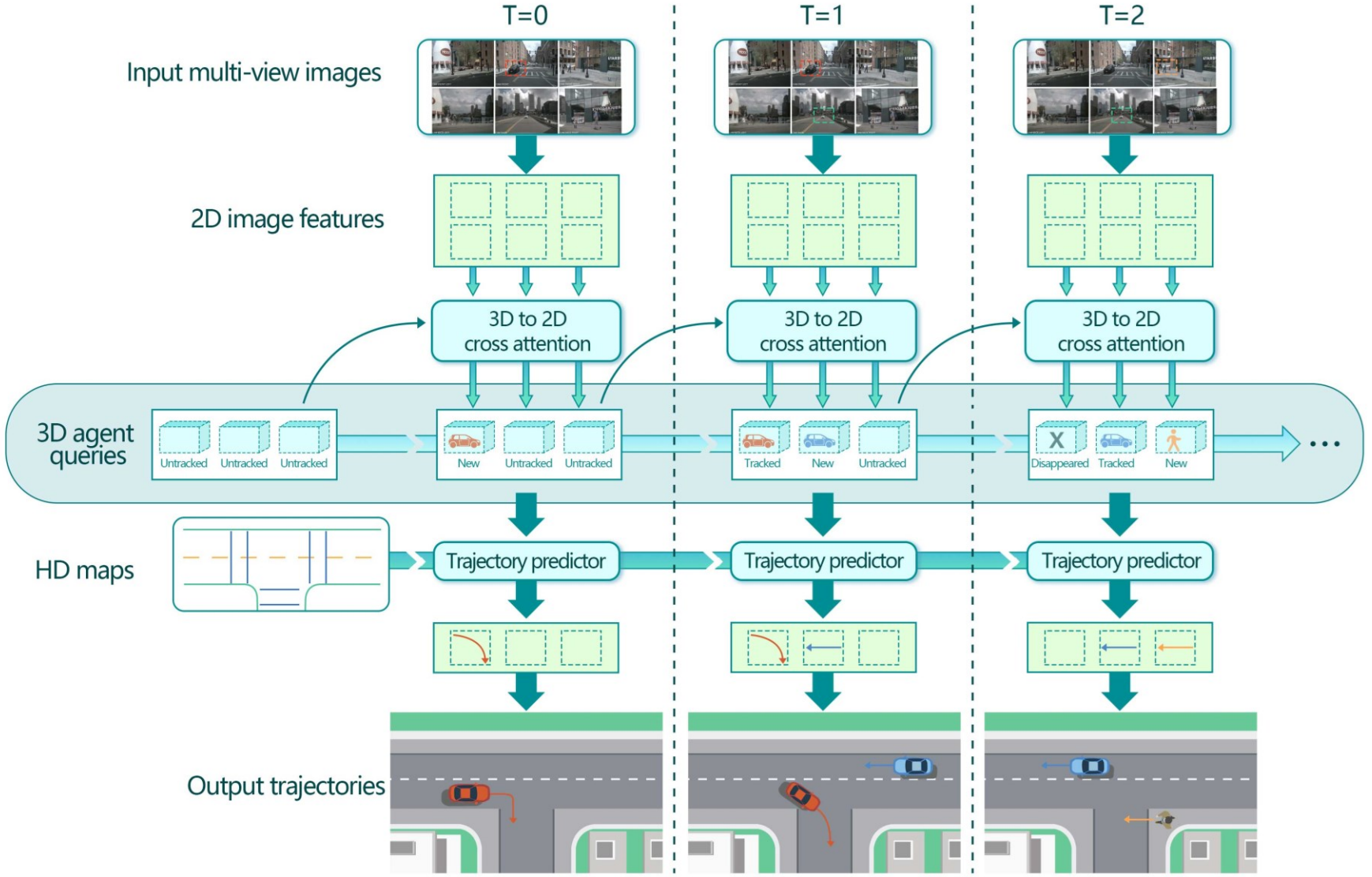
Model Architecture





MUTR3D: A Multi-camera Tracking Framework via 3D-to-2D Queries, Tianyuan Zhang, Xuanyao Chen, Yue Wang, Yilun Wang, Hang Zhao, CVPRW 2022

Model Architecture



Streaming

Queries initialized, updated and discarded over time…

Experiments

- Comparison with traditional pipeline and the SOTA model PnPNet.

		Traditional		PnPNet-vision [30]		ViP3D (Ours)
Architechture	detector	DETR3D		DETR3D		DETR3D
	detector-tracker interface	boxes		boxes		queries
	tracker	Kalman Filter	CenterPoint	Kalman Filter	CenterPoint	query-based
	tracker-predictor interface	trajectories		cropped features		queries
	predictor	regression-based		regression-based		regression-based
Metrics	minADE↓	2.07	2.06	2.04	2.04	**2.03**
	minFDE↓	3.10	3.02	3.08	3.03	**2.90**
	MR↓	0.289	0.277	0.277	0.271	**0.239**
	EPA↑	0.191	0.209	0.198	0.213	**0.236**

Ablation Study

- Differentiability is important
- Historical trajectories are no longer necessary, since agent queries captured these information implicitly

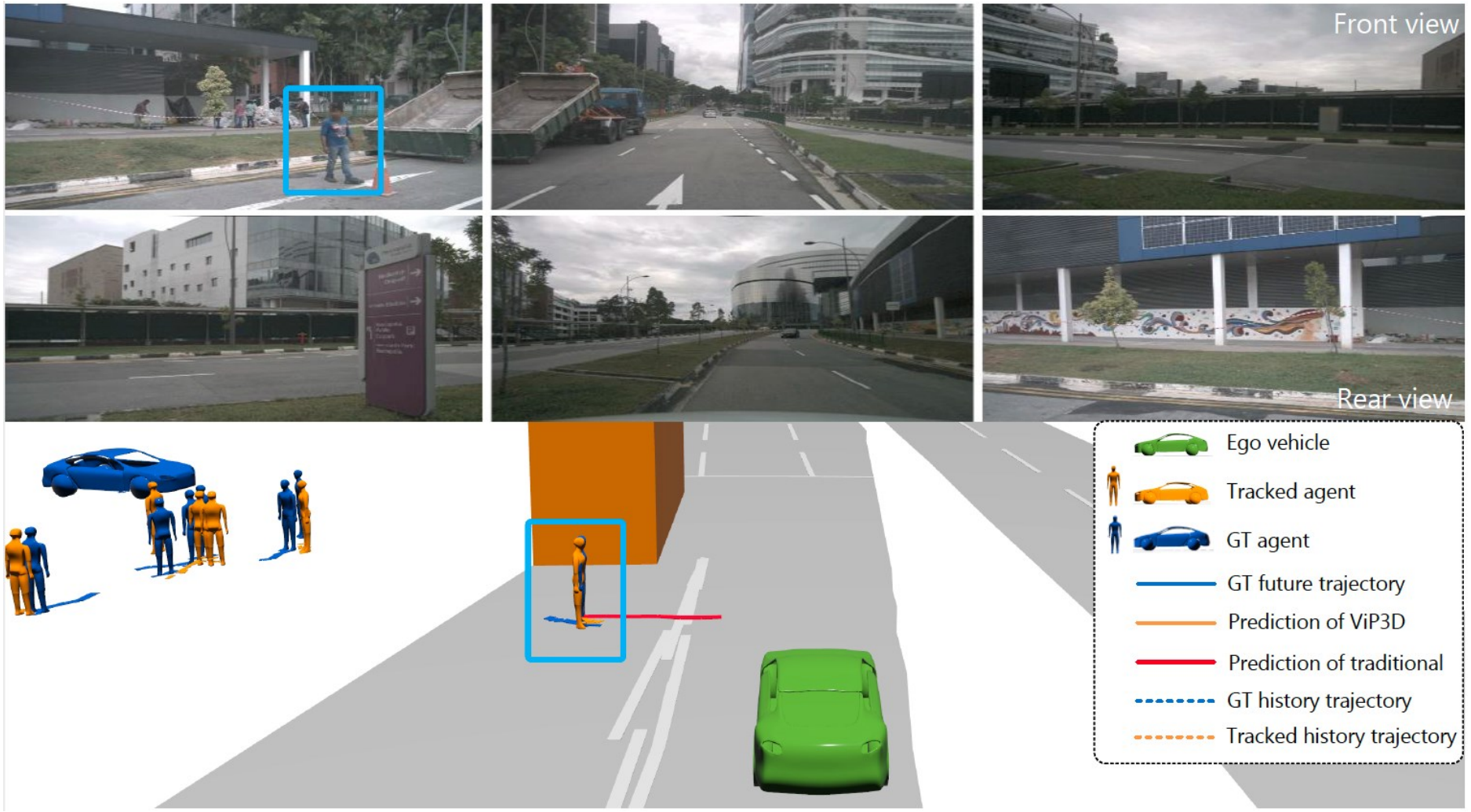
	Prediction inputs	Differentiable	minADE ↓	minFDE ↓	MR ↓	EPA↑
	Agent trajectories	✗	2.30	3.33	0.282	0.186
	Agent trajectories + Agent queries	✗	2.20	3.19	0.274	0.211
ViP3D	Agent queries	✓	**2.03**	**2.90**	**0.239**	**0.236**

Ablation Study

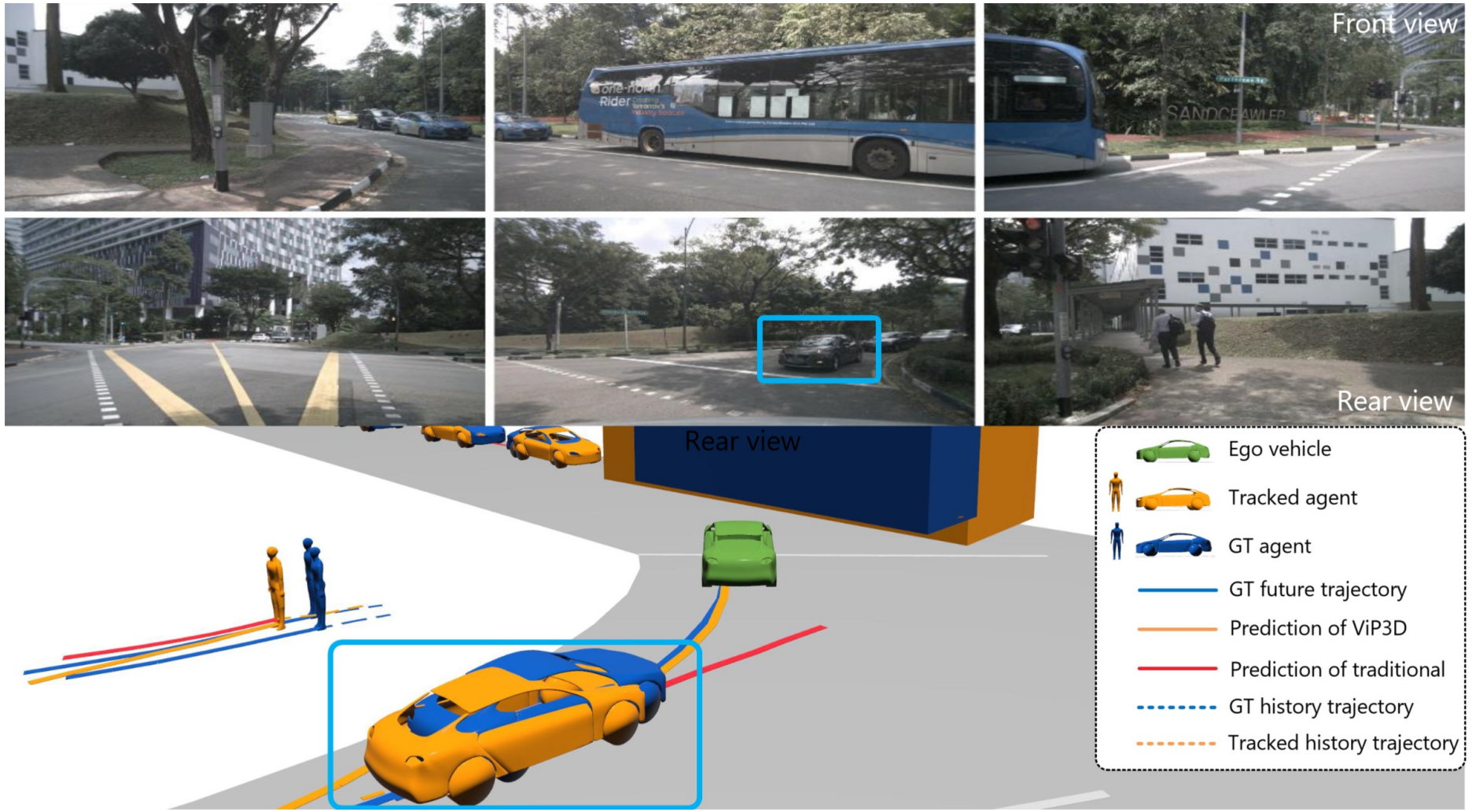
- ViP3D is compatible with multiple trajectory decoders.
- Such as TNT (goal-based) and HOME (heatmap-based).

Decoder	Pipeline	mADE	mFDE	MR	EPA
Goal [60]	Traditional	2.50	3.93	0.266	0.195
	ViP3D	**2.24**	**3.33**	**0.238**	**0.219**
Heatmap [14]	Traditional	2.53	3.81	0.264	0.197
	ViP3D	**2.33**	**3.42**	**0.218**	**0.214**

Qualitative Results



Qualitative Results



Project Page

MARS Lab

THE END THANKS