Object-centric Perception for Autonomous Driving

Foundation Model Group
Tiancai Wang

MEGVII 旷视

MEGVII 旷视

1. 背景介绍
2. PETR系列
3. MOTR系列
4. 总结回顾

背景介绍

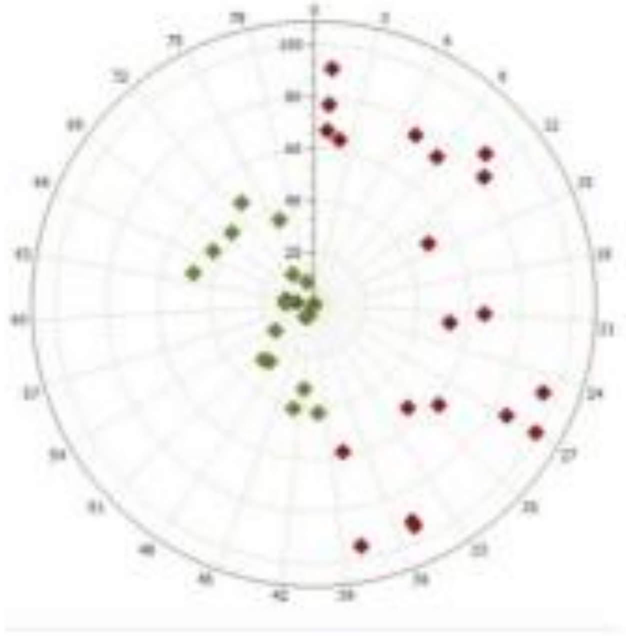
BEV Perception

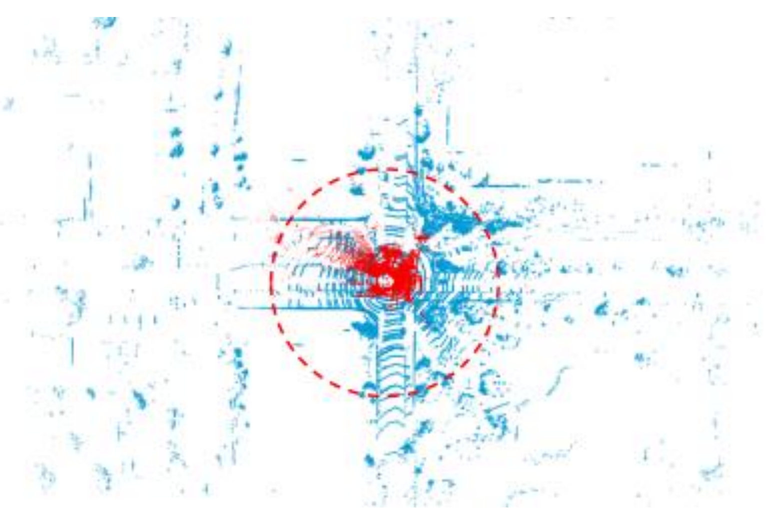
↑

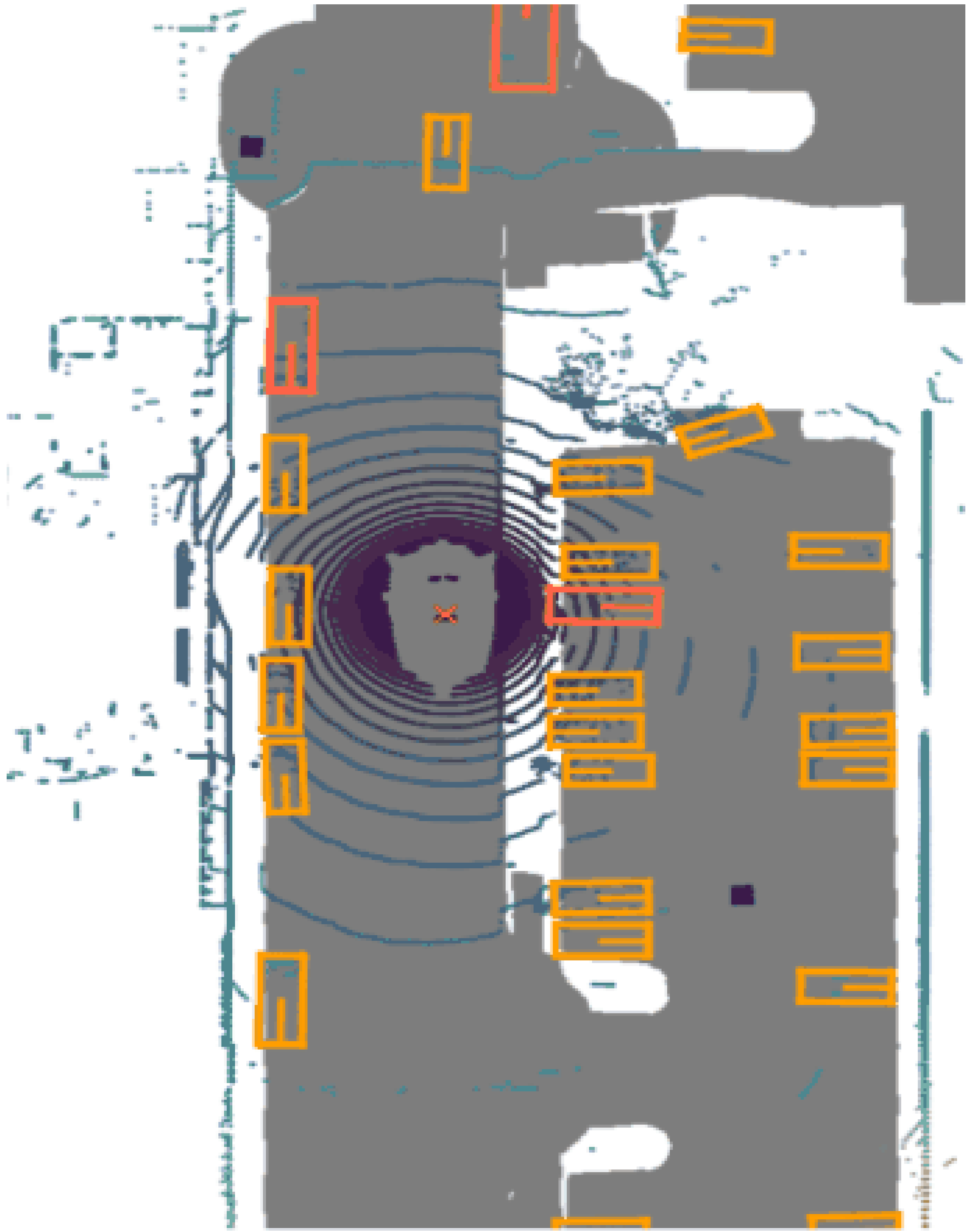
BEV Radar | BEV LiDAR | BEV Cam

↑

Radar | LiDAR | Vision/Camera

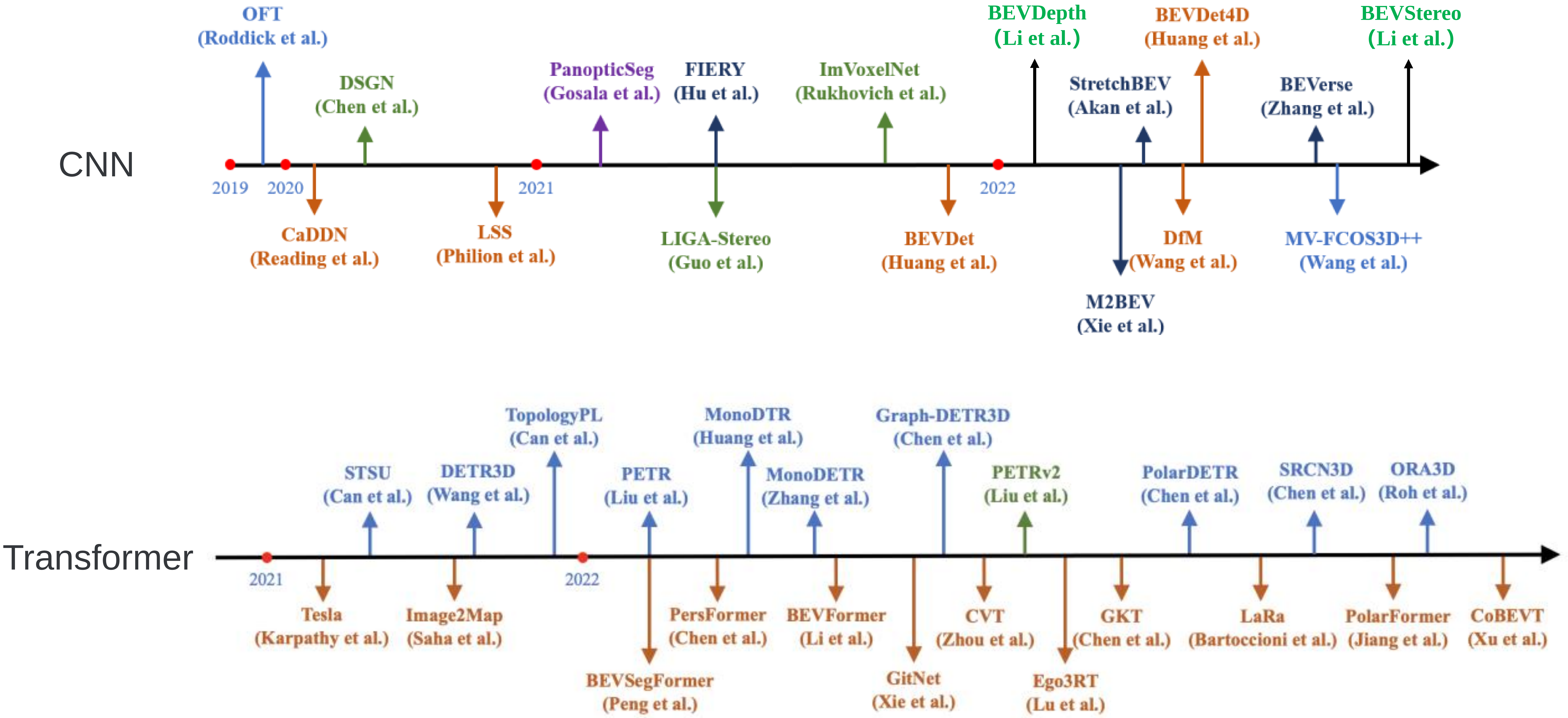






背景介绍

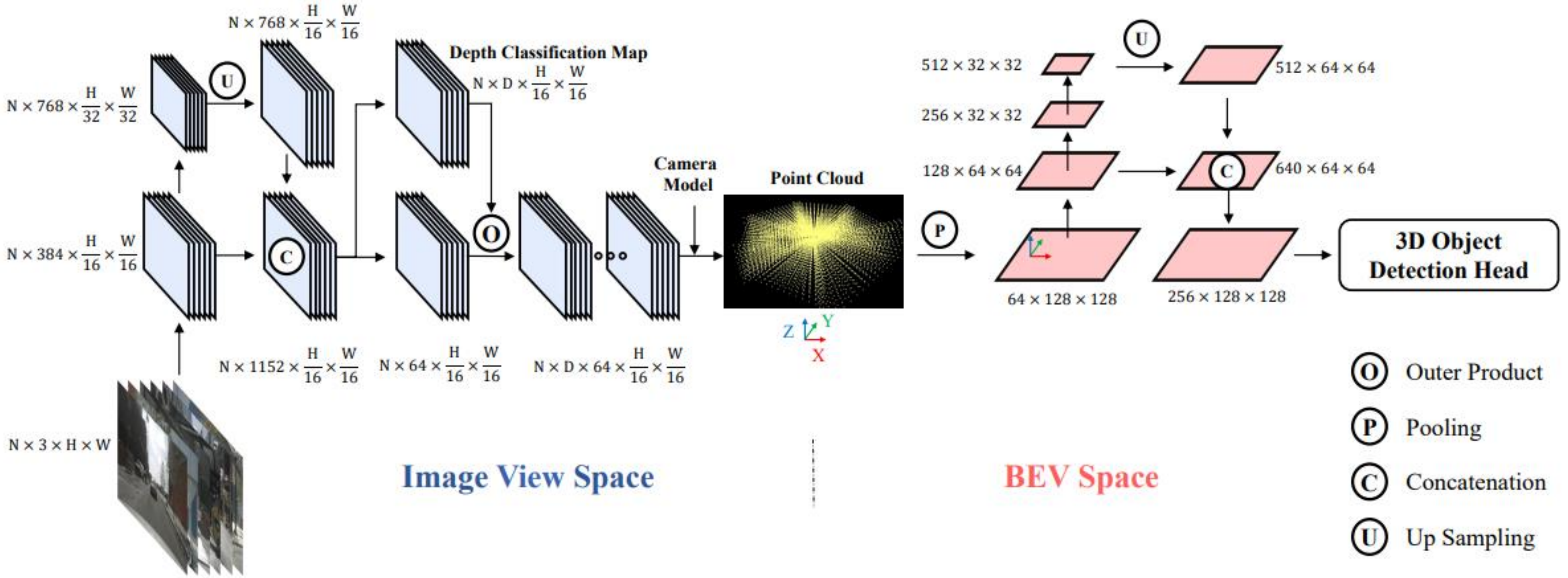
MEGVII 旷视



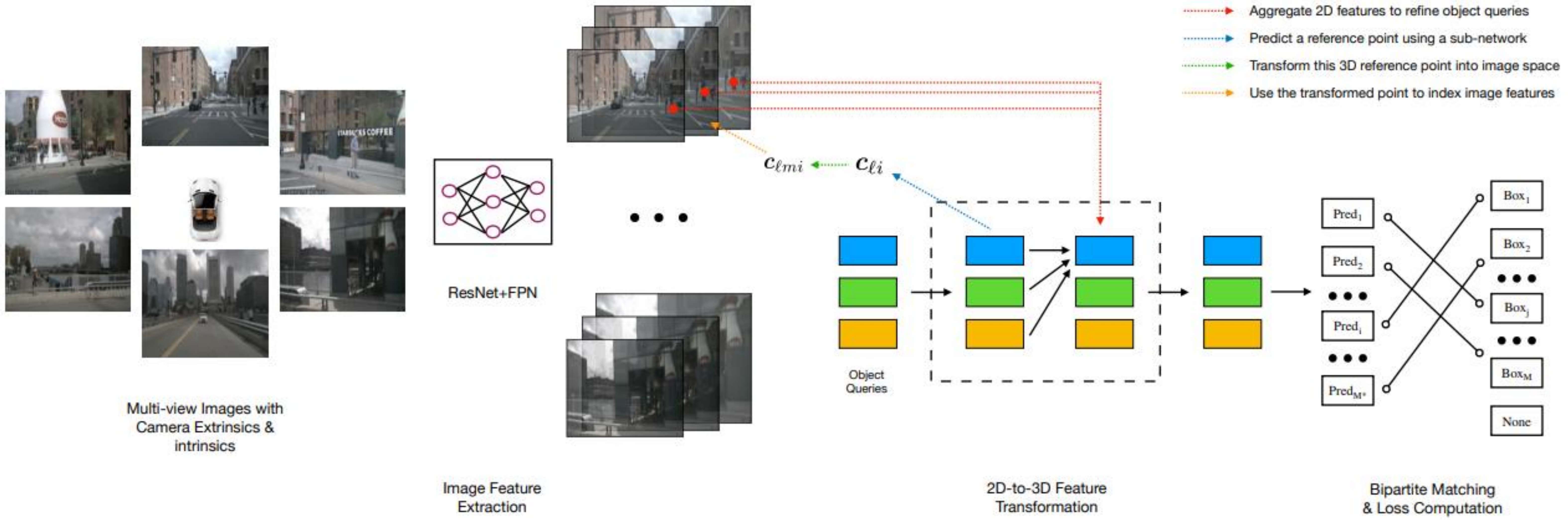
[1] Ma, Yuexin, Wang, Tai et al. “Vision-Centric BEV Perception: A Survey." In arxiv, 2022.

背景介绍

- BEV based框架
  - BEVDet
  - BEVDepth
  - BEVFormer



- Object centric框架
  - DETR3D
  - PETR
  - Sparse4D



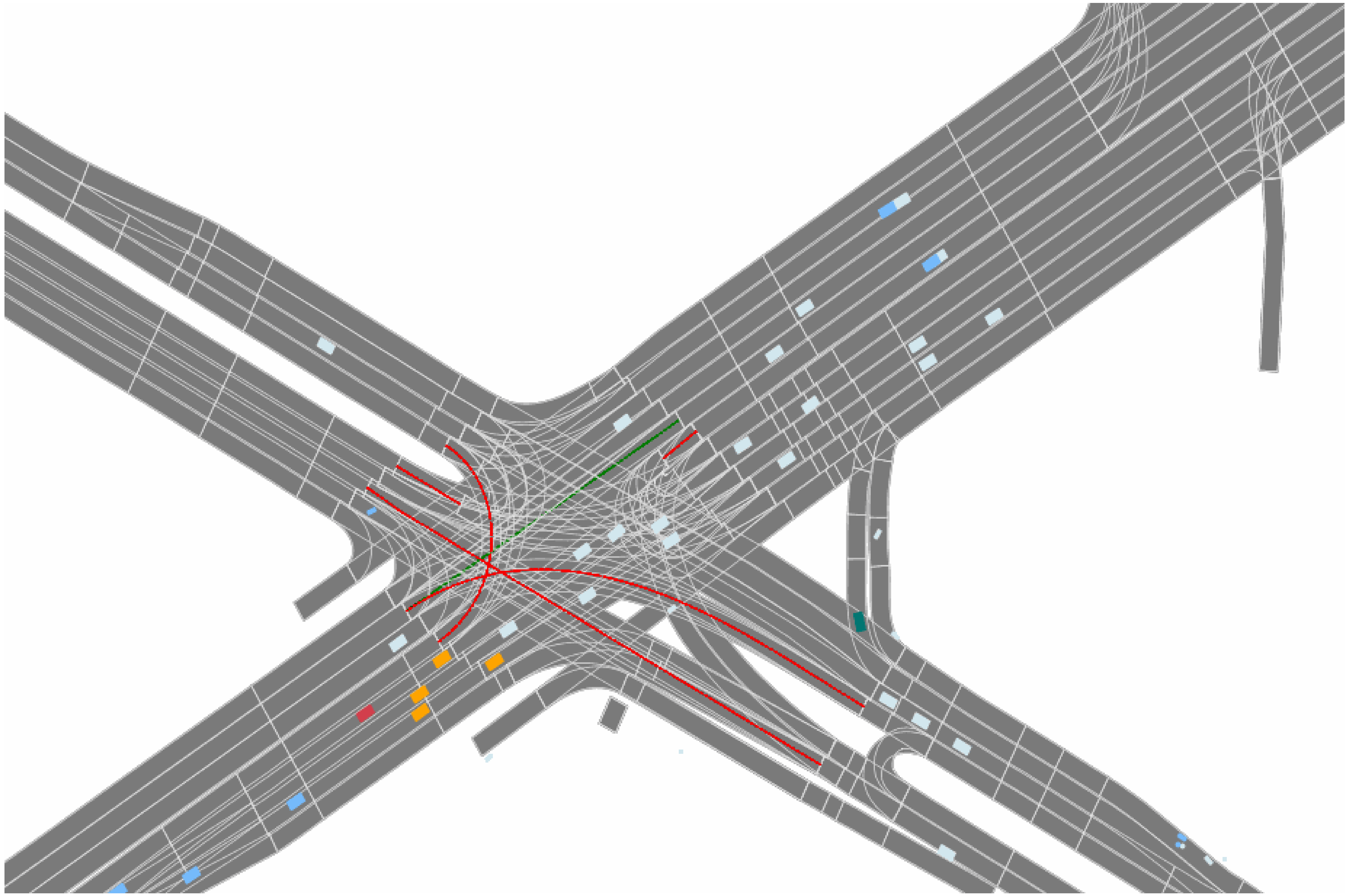
背景介绍

❖ Why Object-centric?

❖ 自驾场景中目标分布稀疏

❖ 动静目标建模的必要性

❖ 考虑交通元素的拓扑关系



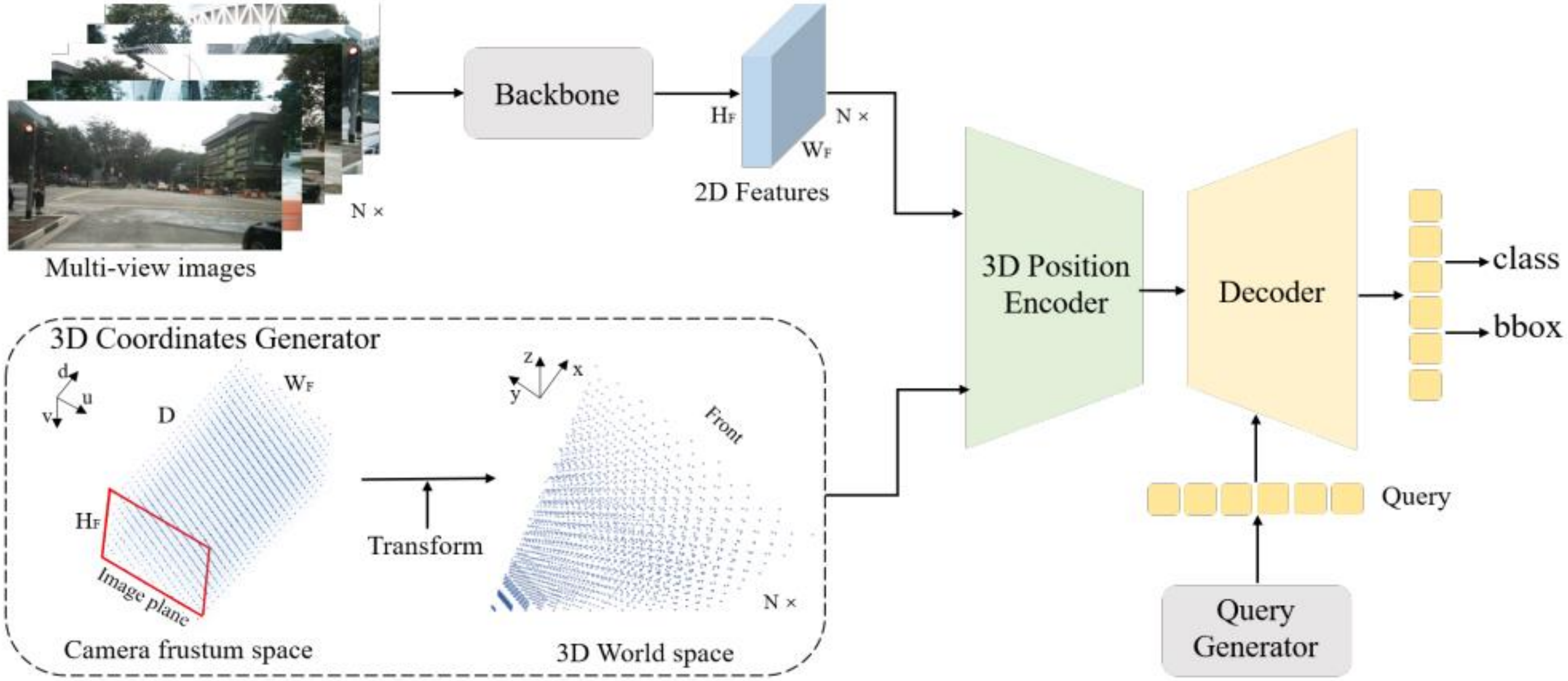
[1] https://thudair.baai.ac.cn/cooplocus

1. 背景介绍
2. PETR系列
3. MOTR系列
4. 总结回顾

PETR系列

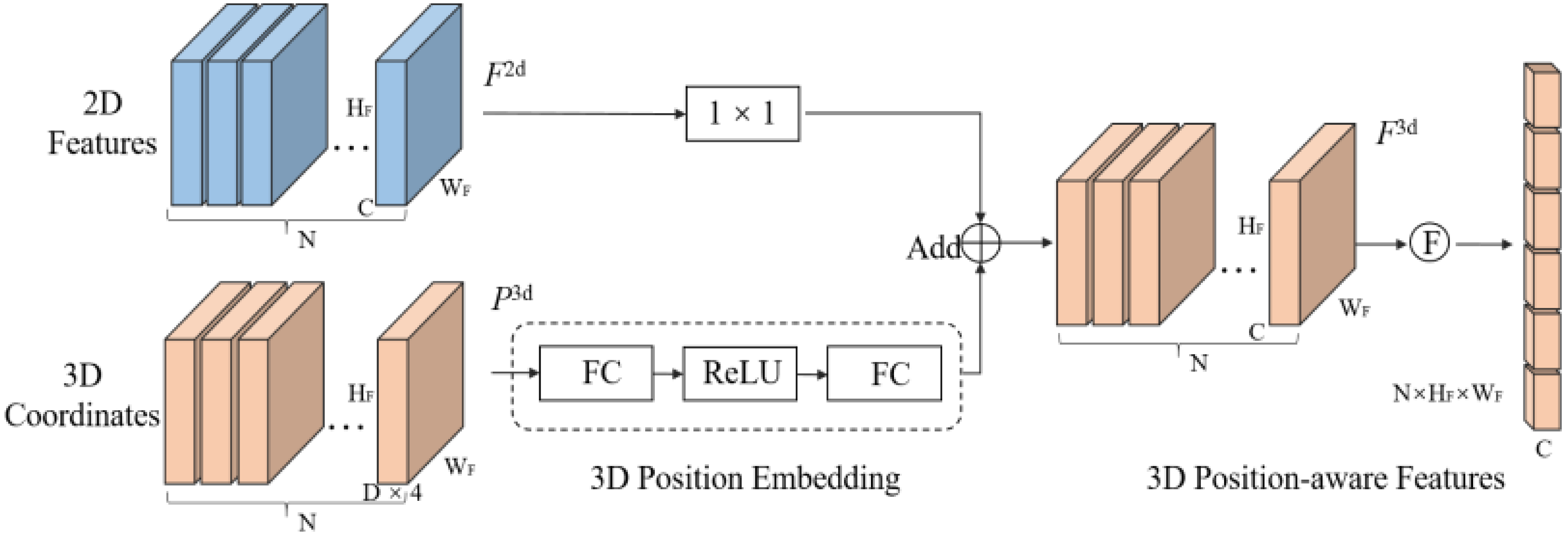
MEGVII 旷视

❖ PETR提出**3D位置编码**，轻松实现端到端3D感知

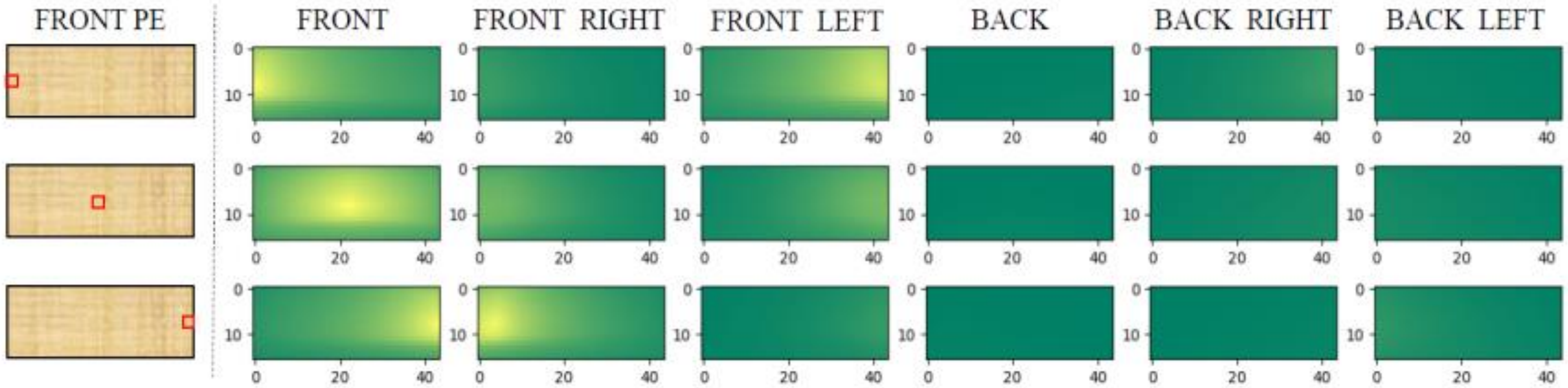


[1] Liu, Yingfei, Wang, Tiancai, Zhang, Xiangyu, and Sun, Jian. “PETR: Position Embedding Transformation for Multi-View 3D Object Detection." In ECCV, 2022.

PETR系列



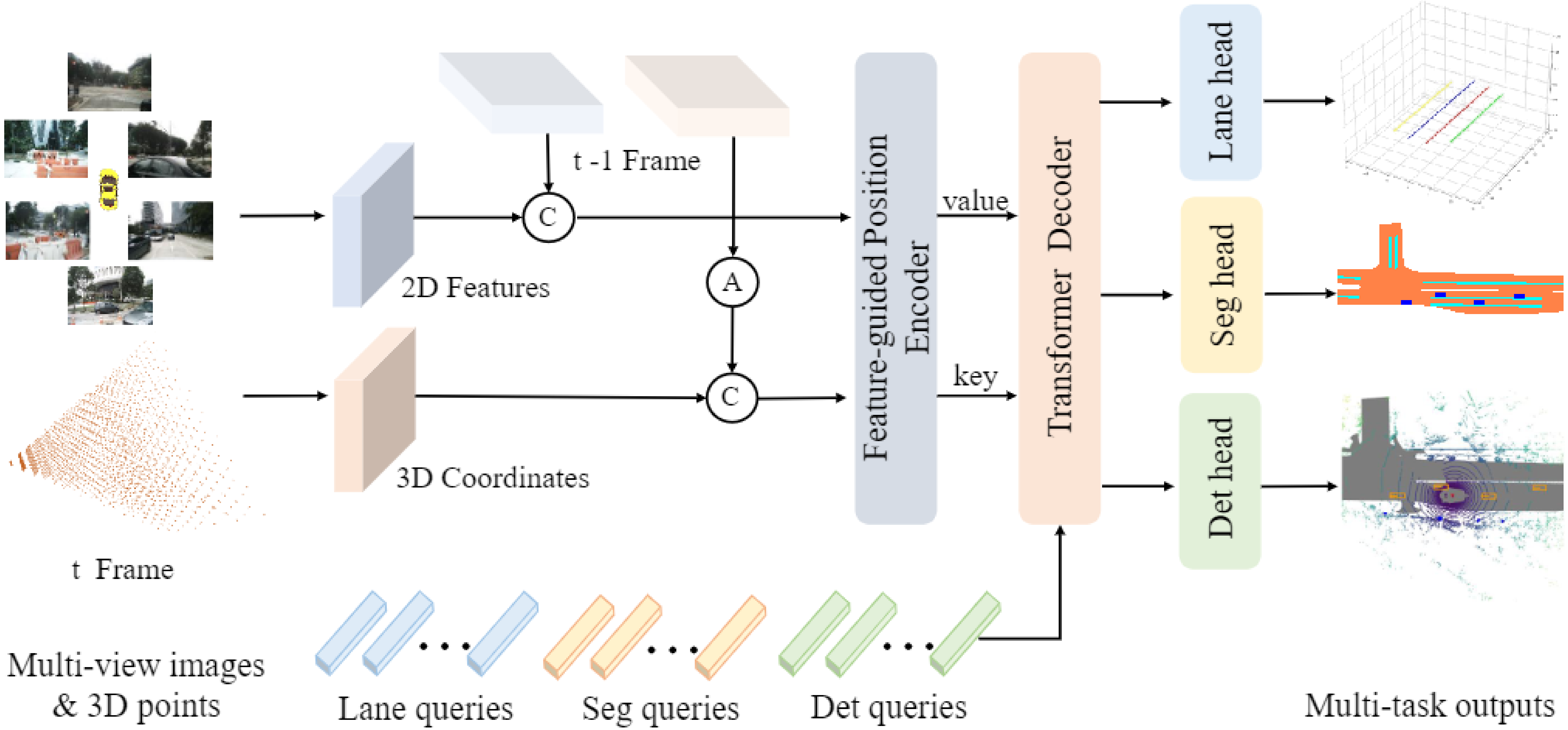
2D
Features
F^{2d}
1 × 1
Add
F^{3d}
3D
Coordinates
P^{3d}
FC
ReLU
FC
3D Position Embedding
3D Position-aware Features
N × HF × WF
C



FRONT PE
FRONT
FRONT RIGHT
FRONT LEFT
BACK
BACK RIGHT
BACK LEFT

PETR系列

❖ PETRv2进一步引入**时序建模**、扩展至**多个感知任务**

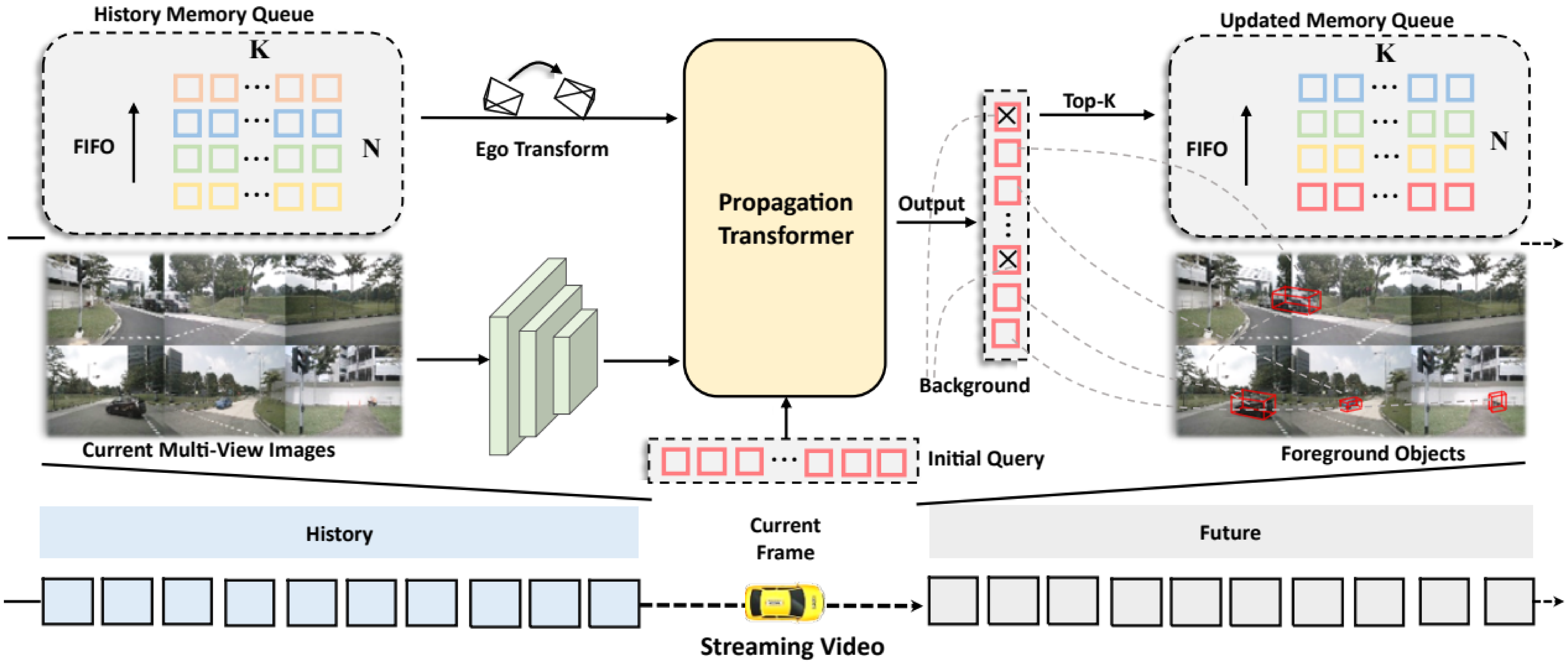


[1] Liu, Yingfei, Yan, Junjie, Jia, Fan, Shuailin Li et al. “PETRv2: A Unified Framework for 3D Perception from Multi-Camera Images." In arxiv, 2022.

PETR系列

MEGVII 旷视

❖ StreamPETR引入**RNN**的概念，扩展至**无限长时序**建模



[1] Wang, Shihao, Liu, Yingfei, Wang, Tiancai, and Zhang, Xiangyu. “Exploring Object-Centric Temporal Modeling for Efficient Multi-View 3D Object Detection." In arxiv, 2023.

PETR系列

MEGVII旷视

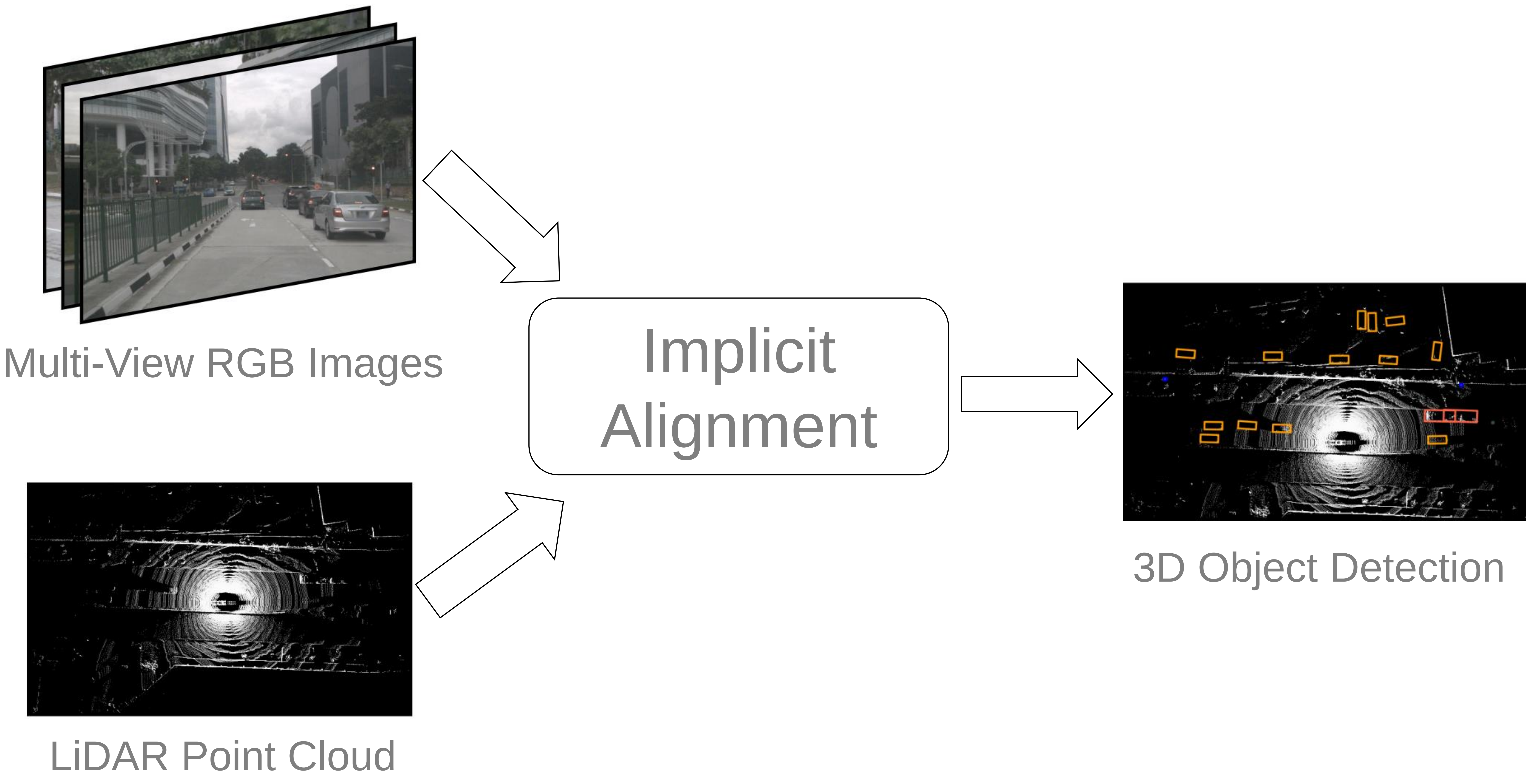
❖ 坚持不用未来帧，纯视觉榜单3D检测第二、多目标追踪第一

	Method					Metrics						
	Date	Name	Modalities	Map data	External data	mAP	mATE (m)	mASE (1-IOU)	mAOE (rad)	mAVE (m/s)	mAAE (1-acc)	NDS
			Camera	All	All							
>	2023-04-05	HoP	Camera	no	no	0.624	0.367	0.249	0.353	0.171	0.131	0.685
>	2023-05-03	StreamPETR-Large	Camera	no	no	0.620	0.470	0.241	0.258	0.236	0.134	0.676
>	2023-02-07	VideoBEV	Camera	no	no	0.592	0.385	0.246	0.323	0.174	0.137	0.670
>	2022-12-21	BEVDet-Gamma	Camera	no	no	0.586	0.375	0.243	0.377	0.174	0.123	0.664
>	2022-12-08	BEVFormer v2 Opt	Camera	no	yes	0.580	0.448	0.262	0.342	0.238	0.128	0.648

	Method											
	Date	Name	Modalities	Map data	External data	AMOTA	AMOTP (m)	MOTAR	MOTA	MOTP (m)	RECALL	GT
			Camera	All	All							
>	2023-05-03	StreamPETR-Large	Camera	no	no	0.653	0.876	0.762	0.553	0.564	0.733	17081
>	2023-05-17	E2E-Tracker-Base	Camera	no	no	0.582	0.919	0.793	0.536	0.381	0.675	17081
>	2023-04-11	DORT	Camera	no	no	0.576	0.951	0.771	0.484	0.536	0.634	17081
>	2023-03-04	QTrack-StreamPETI	Camera	no	no	0.566	0.975	0.711	0.460	0.576	0.650	17081
>	2022-10-24	MV-ByteTrack	Camera	no	no	0.564	1.005	0.748	0.471	0.616	0.635	17081

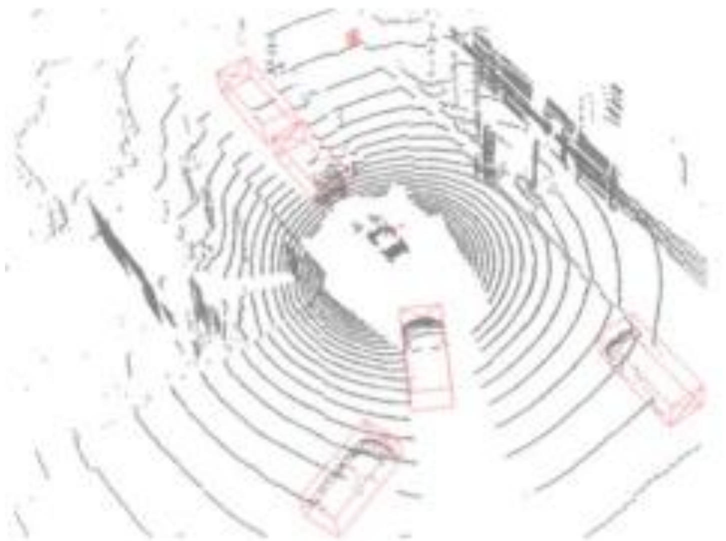
PETR系列->CMT

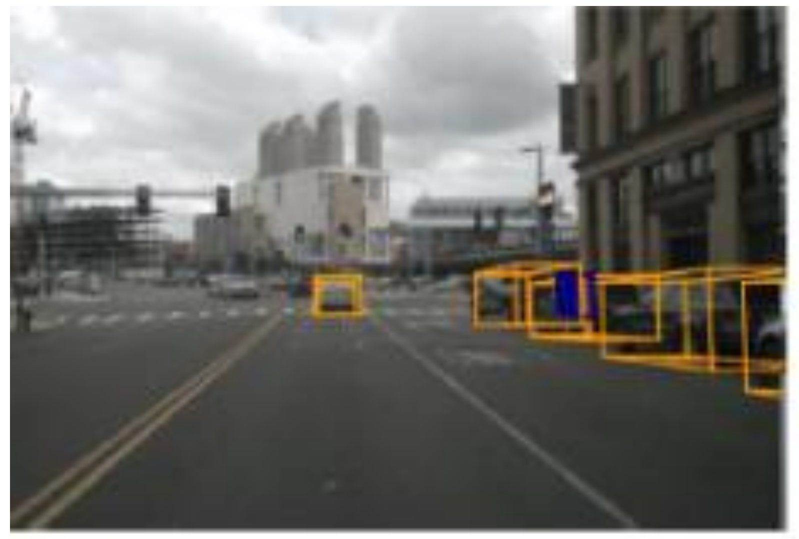
❖PETR的设计思想能够扩展到多模态场景?

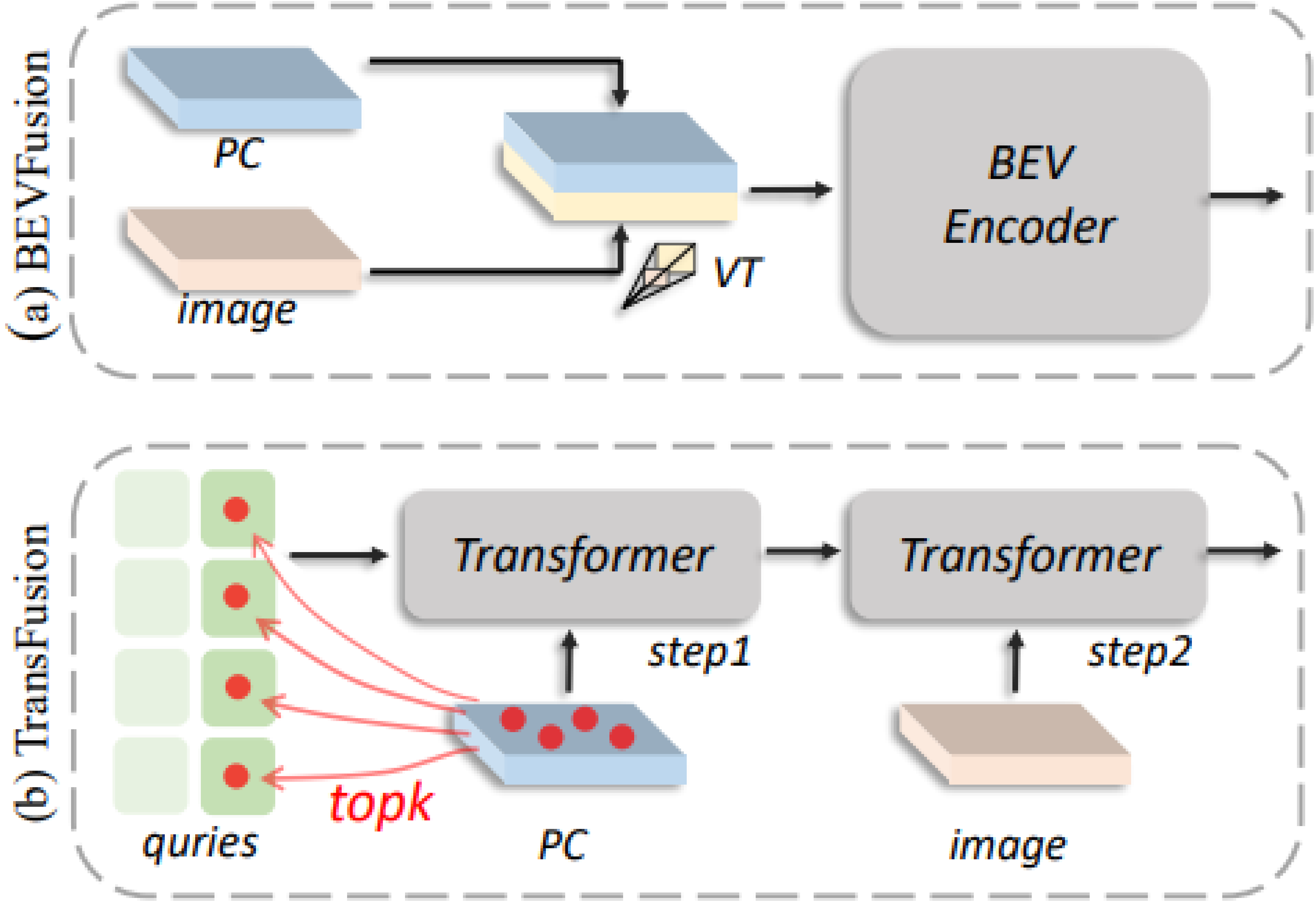


PETR系列->CMT

- ❖ 多模融合框架存在的问题
  - ❖点云、图像不易对齐
  - ❖多阶段，前后依赖







[1] Liu, Zhijian, et al. " BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation." In ICRA, 2023.
[2] Bai, Xuyang, et al. " TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers." In CVPR, 2022.

PETR系列->CMT

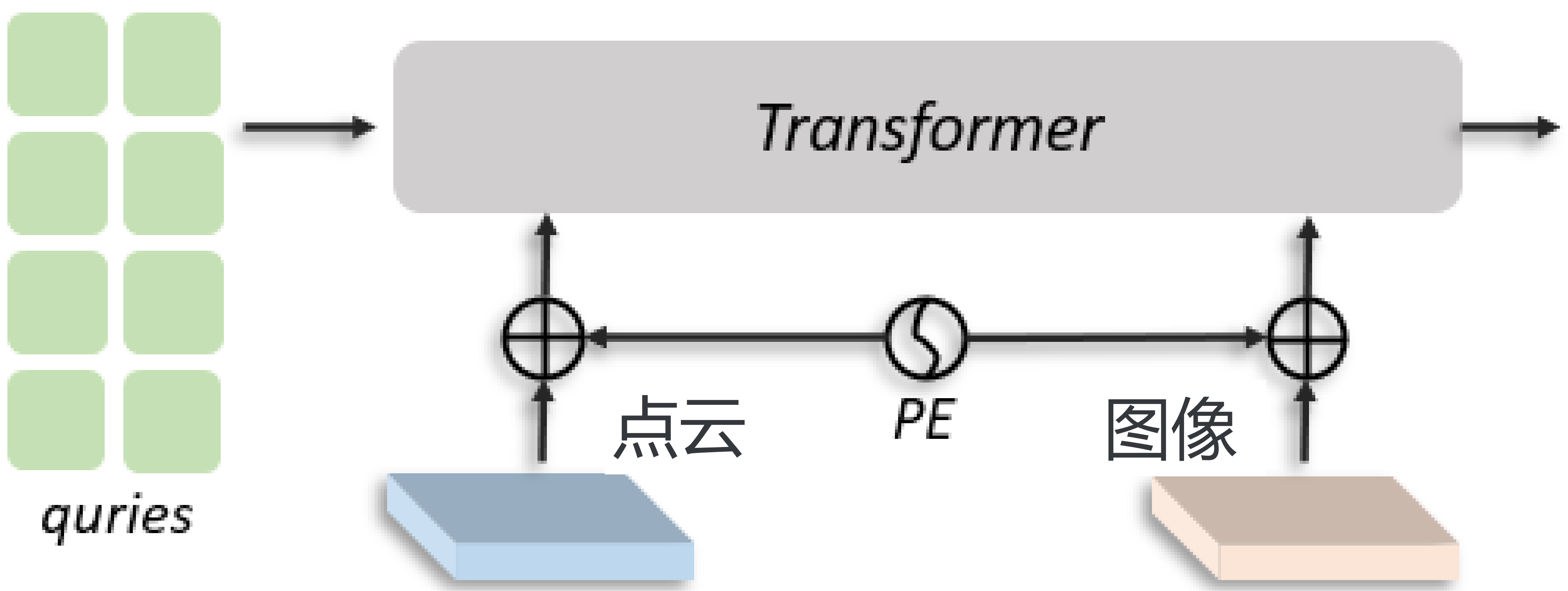
❖ 如何统一多模态输入、应对传感器损坏?

❖对策

❖利用**3D位置编码**，统一多模态

❖**模态掩码训练**

CMT框架示意图

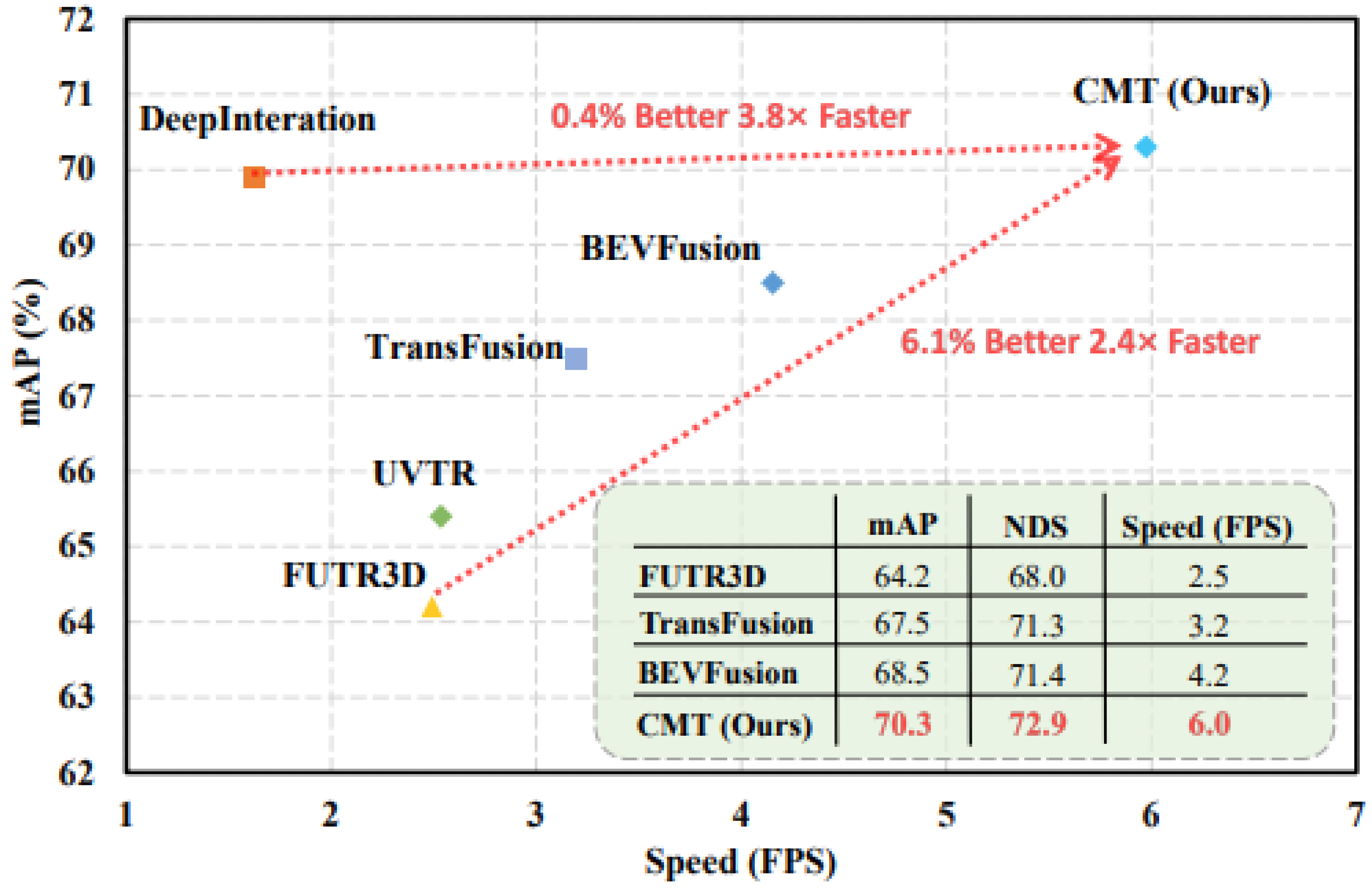


Metric	Vanilla training			Masked-modal training		
	CMT	only LiDAR	only Cams	CMT	only LiDAR	only Cams
NDS ↑	0.716	0.594	0.067	0.719 (↑0.3%)	0.677 (↑**8.3%**)	0.434 (↑**36.7%**)
mAP ↑	0.685	0.472	0.000	0.694 (↑0.9%)	0.613 (↑**14.1%**)	0.386 (↑**38.6%**)

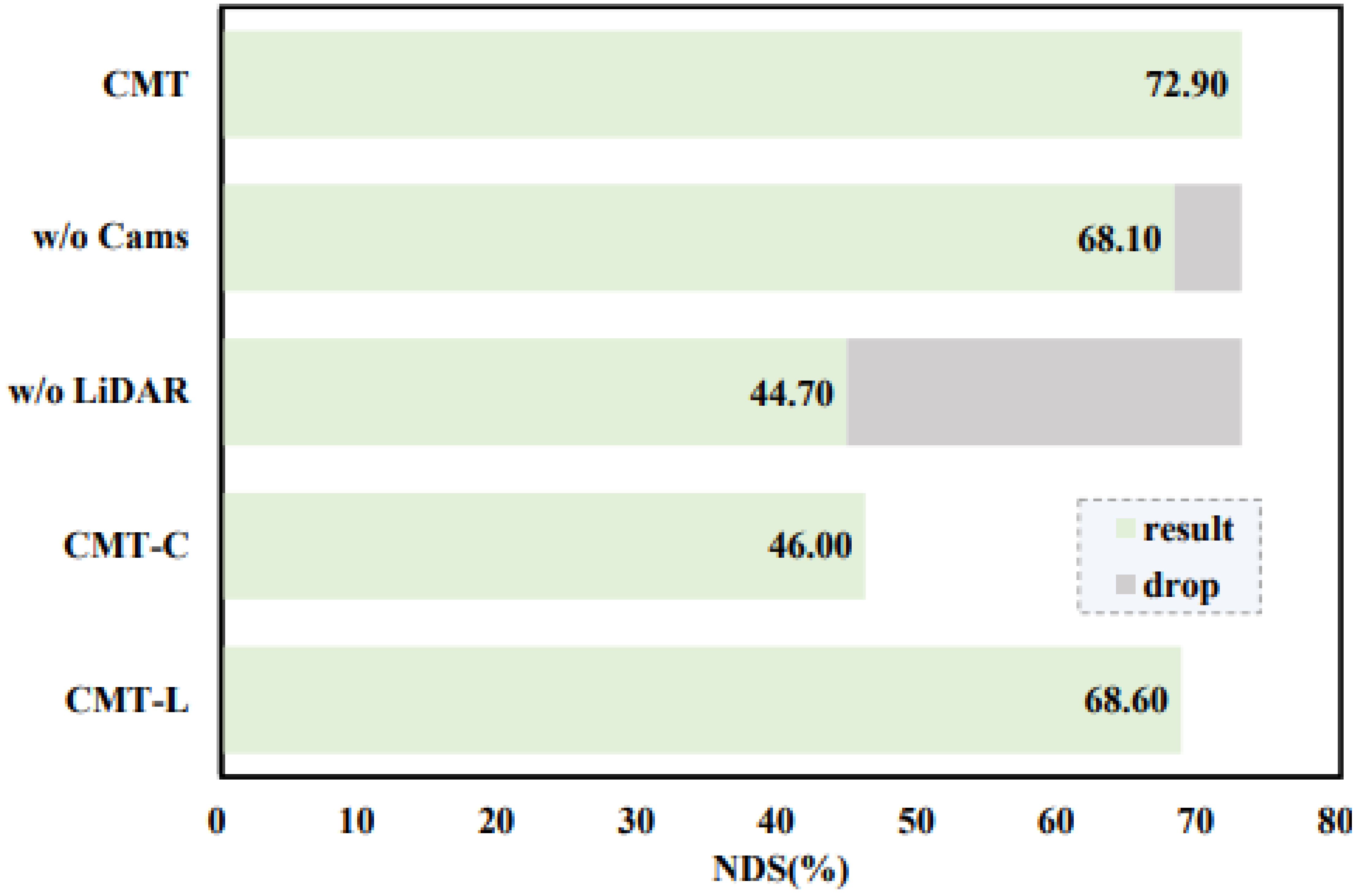
[1] Yan, Junjie, et al. " Cross Modal Transformer via Coordinates Encoding for 3D Object Detection." In arxiv, 2022.

PETR系列->CMT

❖ CMT框架性能评测

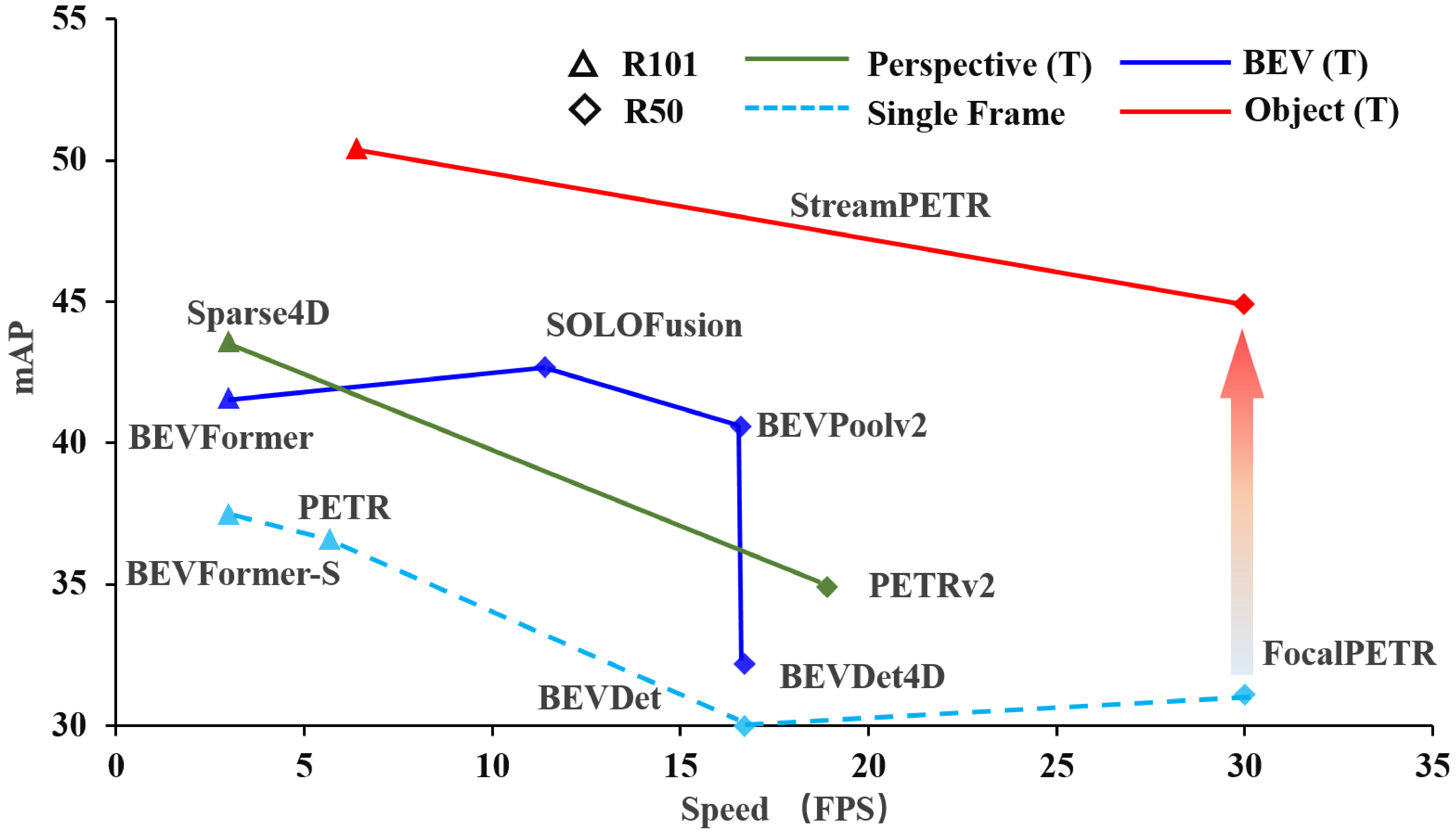


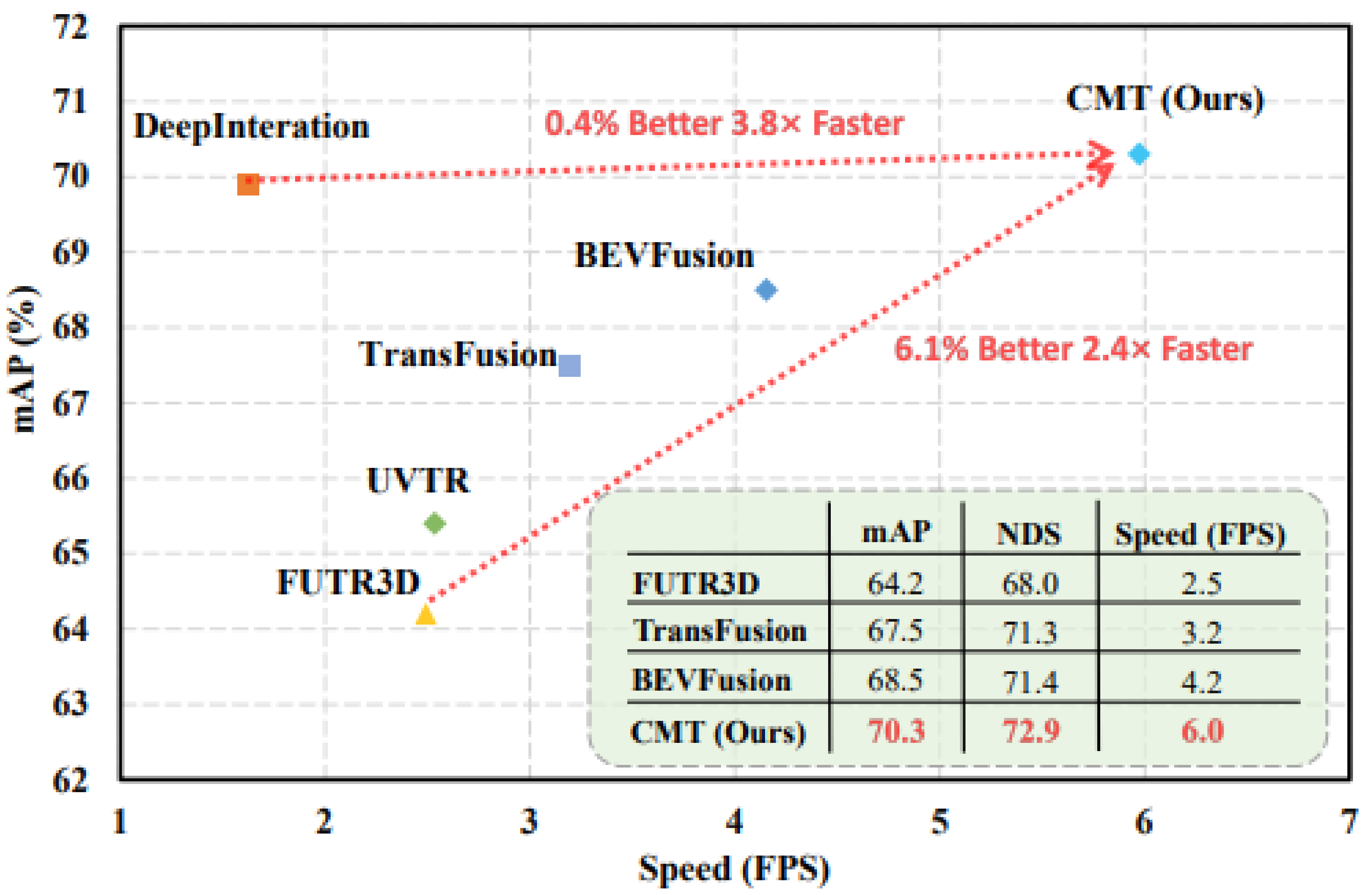
	mAP	NDS	Speed (FPS)
FUTR3D	64.2	68.0	2.5
TransFusion	67.5	71.3	3.2
BEVFusion	68.5	71.4	4.2
CMT (Ours)	70.3	72.9	6.0



PETR系列->速度评测

❖ 问题：Transformer模型没有CNN模型快?

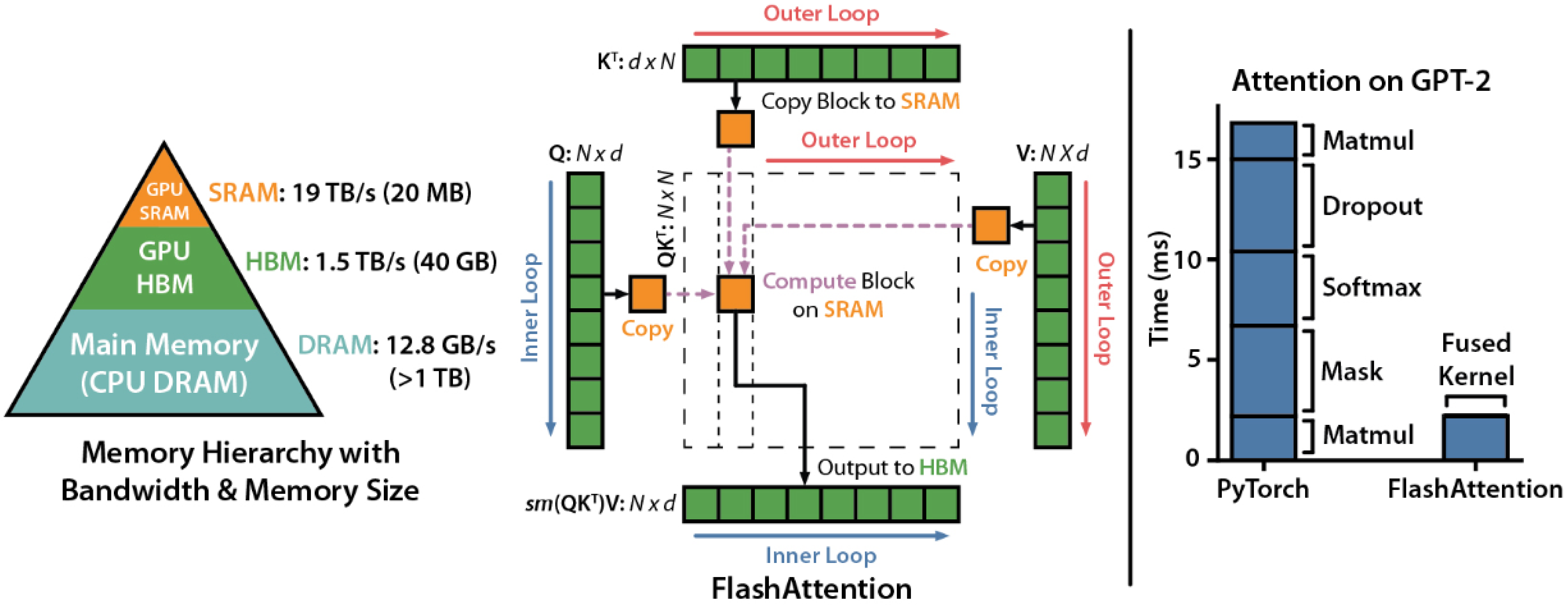




	mAP	NDS	Speed (FPS)
FUTR3D	64.2	68.0	2.5
TransFusion	67.5	71.3	3.2
BEVFusion	68.5	71.4	4.2
CMT (Ours)	70.3	72.9	6.0

PETR系列->速度评测

❖ FlashAttention [1] 高度适用于naïve attention加速



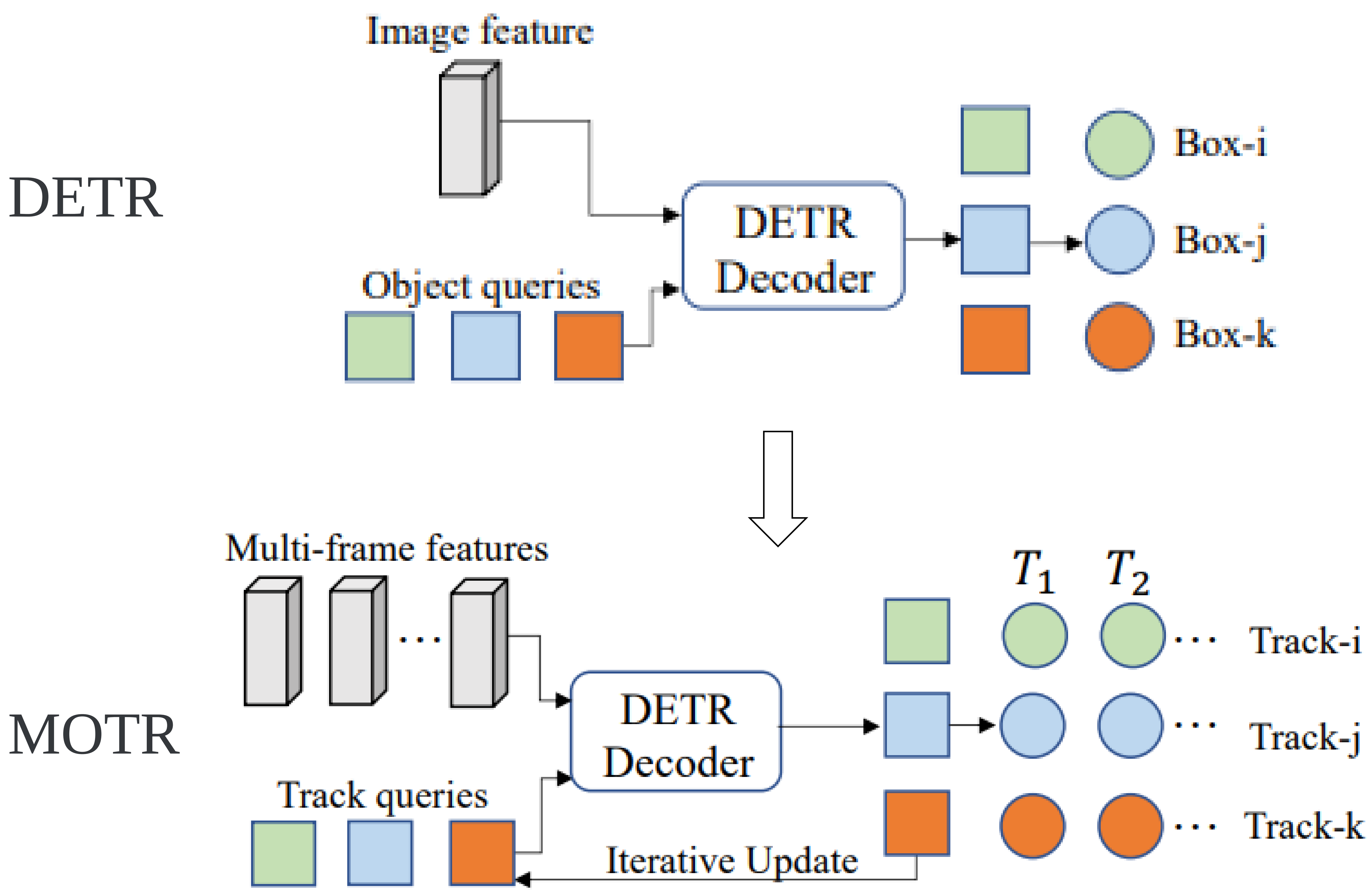
[1] Dao, Tri, et al. "FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness." In arxiv, 2022.

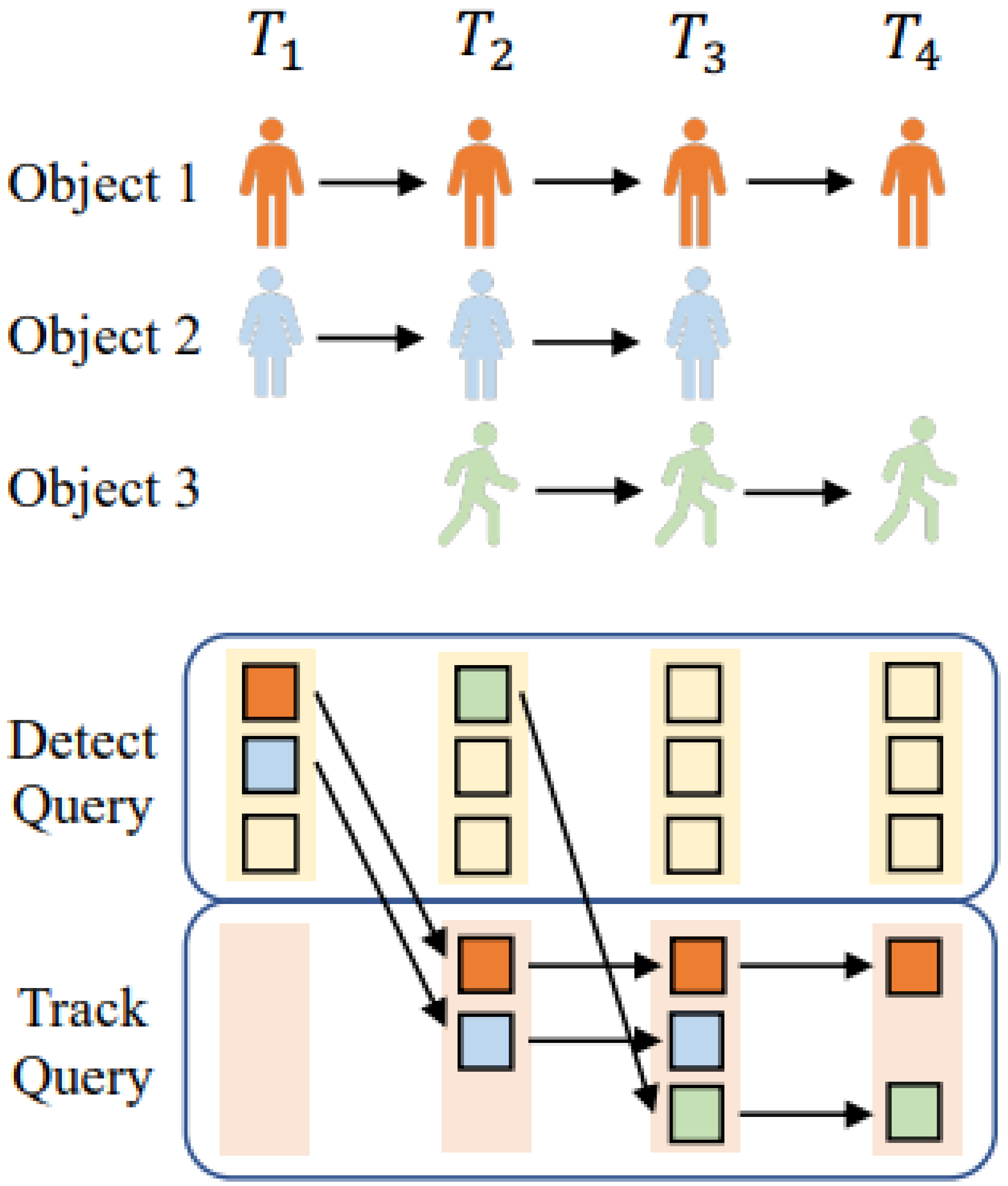
MEGVII 旷视

MOTR系列

MEGVII 旷视

❖ MOTR**重启RNN范式**，实现**在线端到端**多目标追踪





[1] Zeng, Fangao, Dong, Bin, Zhang, Yuang, and Wang, Tiancai. “MOTR: End-to-End Multiple-Object Tracking with Transformer." In ECCV, 2022.

MOTR系列

❖ MOTR的检测性能**垃圾**，对复杂运动估计有明显**优势**

Methods	HOTA↑	AssA↑	DetA↑	IDF1↑	MOTA↑	IDS↓
CNN-based:						
Tracktor++[2]	44.8	45.1	44.9	52.3	53.5	2072
CenterTrack[44]	52.2	51.0	53.8	64.7	67.8	3039
TraDeS [40]	52.7	50.8	55.2	63.9	69.1	3555
QDTrack [20]	53.9	52.7	55.6	66.3	68.7	3378
GSDT [35]	55.5	54.8	56.4	68.7	66.2	3318
FairMOT[43]	59.3	58.0	60.9	72.3	73.7	3303
CorrTracker [32]	60.7	58.9	62.9	73.6	76.5	3369
GRTU [33]	62.0	62.1	62.1	75.0	74.9	1812
MAATrack [27]	62.0	60.2	64.2	75.9	79.4	1452
ByteTrack [42]	63.1	62.0	64.5	77.3	80.3	2196
Transformer-based:						
TrackFormer [18]	/	/	/	63.9	65.0	3528
TransTrack[29]	54.1	47.9	**61.6**	63.9	**74.5**	3663
MOTR (ours)	**57.8**	**55.7**	60.3	**68.6**	73.4	**2439**

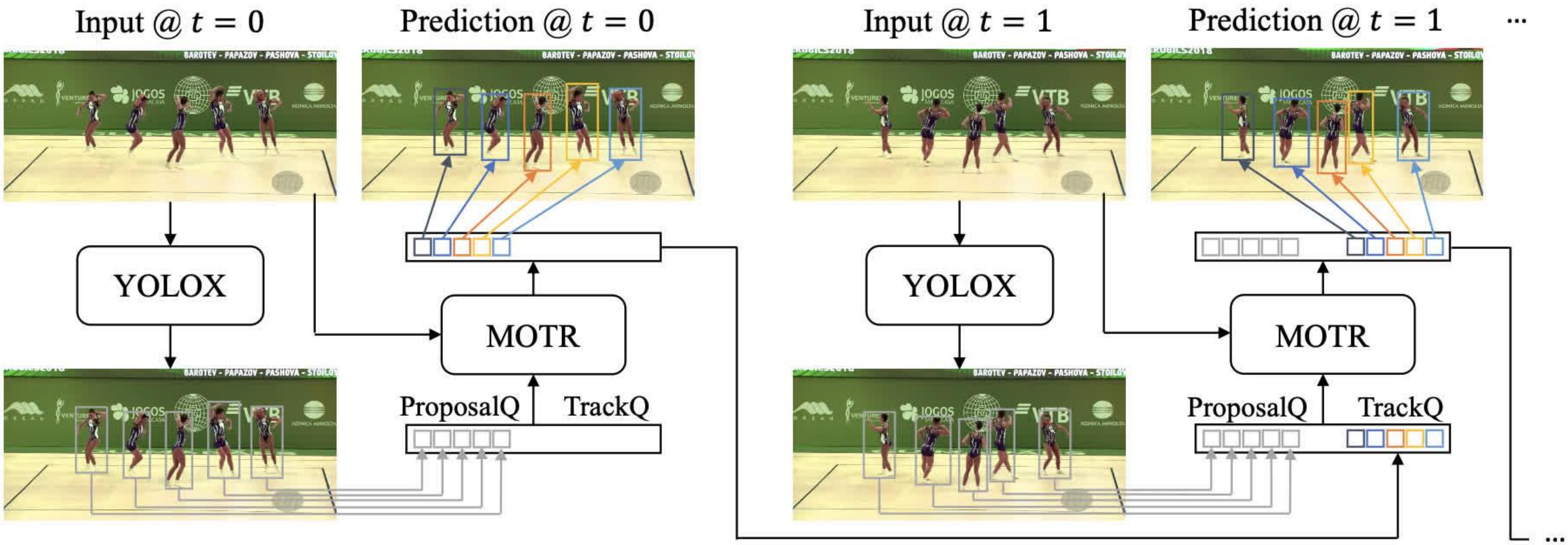
Methods	HOTA	AssA	DetA	MOTA	IDF1
CenterTrack [44]	41.8	22.6	**78.1**	86.8	35.7
FairMOT [43]	39.7	23.8	66.7	82.2	40.8
QDTrack [20]	45.7	29.2	72.1	83.0	44.8
TransTrack [29]	45.5	27.5	75.9	88.4	45.2
TraDes [40]	43.3	25.4	74.5	86.2	41.2
ByteTrack [42]	47.7	32.1	71.0	**89.6**	**53.9**
MOTR (ours)	**54.2**	**40.2**	73.5	79.7	51.5

Method	IoU match	NMS	ReID
TransTrack [29]	✓		
TrackFormer [18]		✓	✓
MOTR (ours)			

MOTR系列

MEGVII 旷视

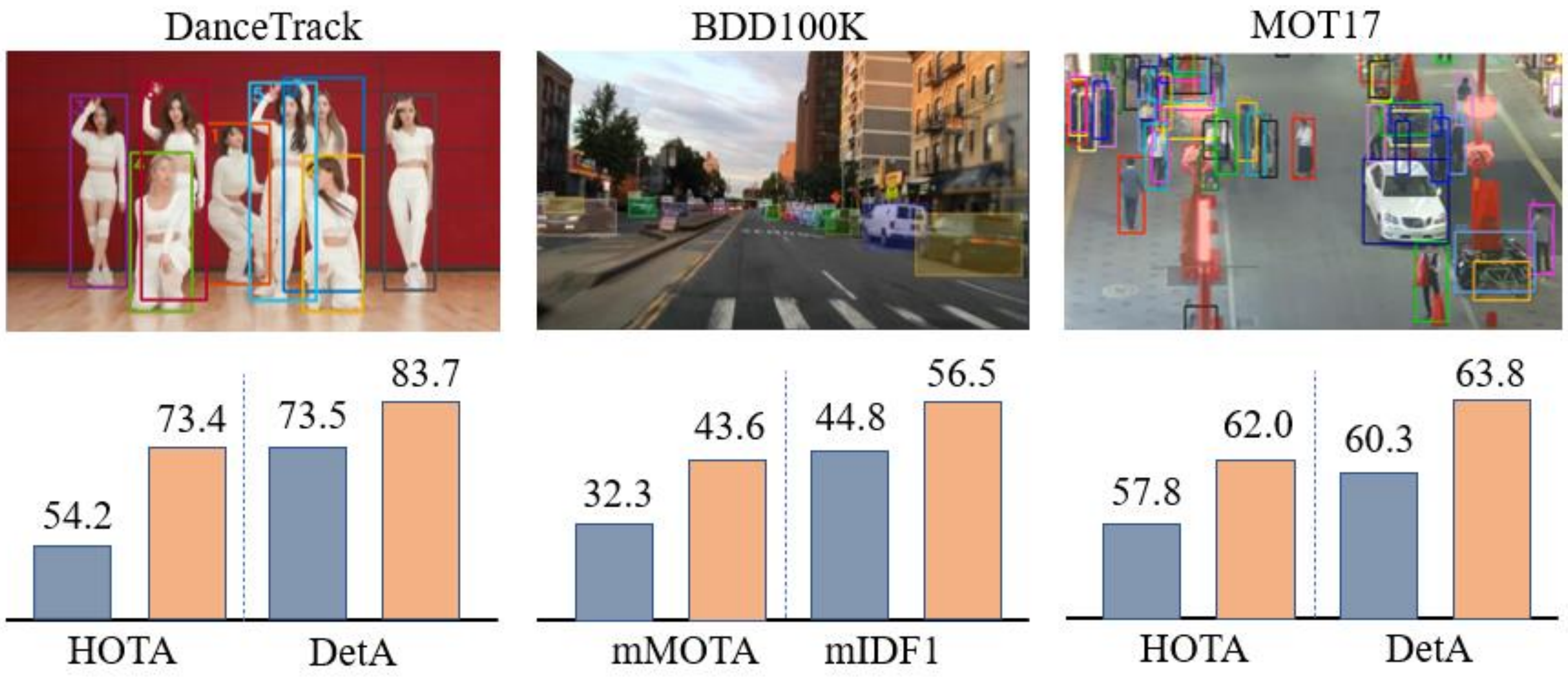
❖ MOTRv2将**MOTR和YOLOX有机结合，**缓解动静态优化问题



[1] Zhang, Yuang, Wang, Tiancai and Zhang, Xiangyu. “MOTRv2: Bootstrapping End-to-End Multi-Object Tracking by Pretrained Object Detectors." In CVPR, 2023.

MOTR系列

❖ MOTRv2性能评测

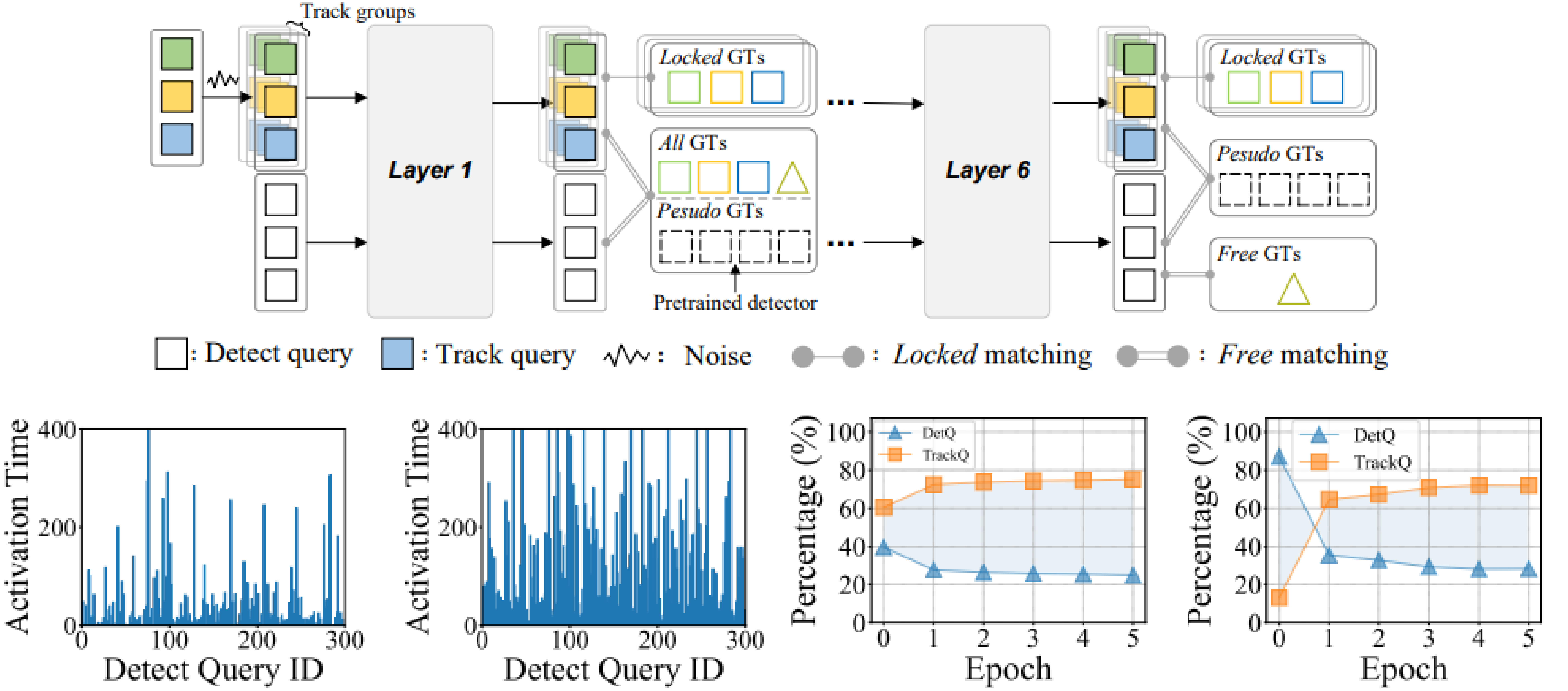


Methods	HOTA	DetA	AssA	MOTA	IDF1
FairMOT [45]	39.7	66.7	23.8	82.2	40.8
CenterTrack [47]	41.8	78.1	22.6	86.8	35.7
TransTrack [28]	45.5	75.9	27.5	88.4	45.2
TraDes [39]	43.3	74.5	25.4	86.2	41.2
ByteTrack [44]	47.7	71.0	32.1	89.6	53.9
GTR [39]	48.0	72.5	31.9	84.7	50.3
QDTrack [22]	54.2	80.1	36.8	87.7	50.4
MOTR [43]	54.2	73.5	40.2	79.7	51.5
OC-SORT [6]	55.1	80.3	38.3	92.0	54.6
MOTRv2 (ours)	69.9	83.0	59.0	91.9	71.7
MOTRv2* (ours)	**73.4**	**83.7**	**64.4**	**92.1**	**76.0**

MOTR系列

MEGVII 旷视

❖ MOTRv3提出**释放-收回匹配规则**，解决动静态优化问题



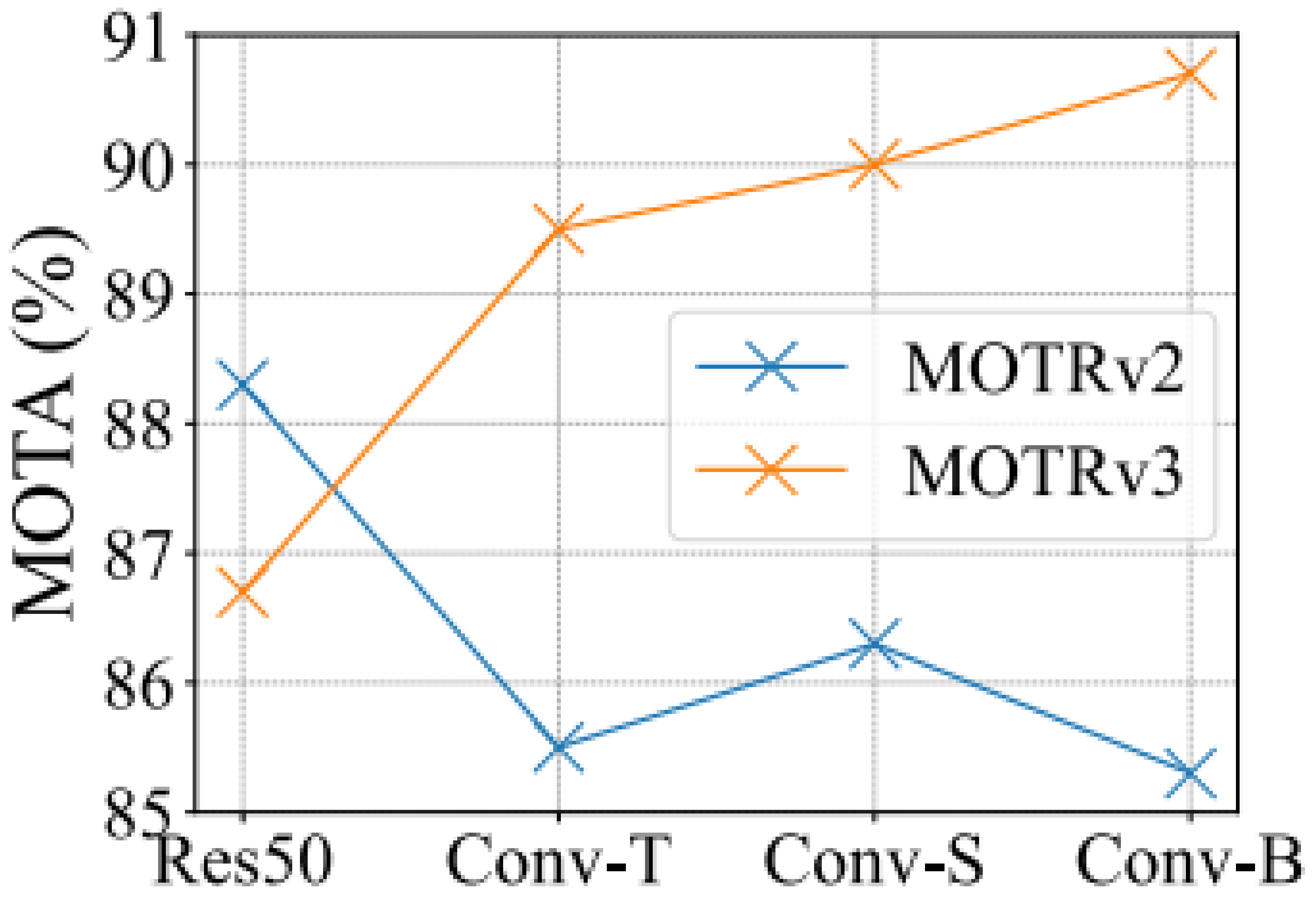
(a) MOTR (b) MOTR with RFS (c) MOTR (d) MOTR with RFS

[1] Yu, En, Wang, Tiancai et al. “MOTRv3: Release-Fetch Supervision for End-to-End Multi-Object Tracking." In arxiv, 2023.

MOTR系列

❖ MOTRv3兼顾**高性能、端到端特性**，观测到了模型**Scaling up**能力

Method	*End to end*	HOTA↑	AssA↑	DetA↑	MOTA↑	IDF1↑
CNN-based						
QDTrack [25]	✗	54.2	36.8	80.1	87.7	50.4
FairMOT [42]	✗	59.3	58.0	60.9	73.7	72.3
CenterTrack [44]	✗	41.8	22.6	78.1	86.8	35.7
ByteTrack [41]	✗	47.7	32.1	71.0	89.6	53.9
OC-SORT [8]	✗	55.1	38.3	80.3	92.0	54.6
Transformer-based						
TransTrack [31]	✗	45.5	27.5	75.9	88.4	45.2
MOTR [40]	✓	54.2	40.2	73.5	79.7	51.5
MOTRv2 [43]	✗	69.9	59.0	83.0	91.9	71.7
MOTRv3	✓	**70.4**	**59.3**	**83.8**	**92.9**	**72.3**



MOTR系列

MEGVII 旷视

- MOTR系列模型的扩展应用
  - 2D Multiple-Object Tracking
    - **MeMOT**
  - 3D Multiple-Object Tracking
    - **MUTR3D、PF-Track**
  - Video Instance Segmentation
    - **GenVIS**
  - Referring Multiple-Object Tracking
    - **RMOT**
  - Trajectory Prediction
    - **ViP3D**
  - Unified Driving Framework
    - **UniAD**

MEGVII 旷视

总结回顾

MEGVII 旷视

- 以物体为中心的表征：
  - 自驾场景中，障碍物、道路元素等均可建模为稀疏的实例化表征
  - 以物体为中心的稀疏表征，便于进行长时序的感知建模
  - 稀疏表征，便于建模运动物体，统一上下游的预测、规划任务
  - 可直接进行拓扑关系的生成，以及适配多模态数据（语音、文本）的接入
  - 对于密集预测任务（如BEV分割、OCC预测），没有明显的优势

总结回顾

MEGVII 旷视

- 关于时序建模：
  - RNN范式的时序建模能够获取较多的时序信息，对运动估计较好
  - 目前的时序探索更多集中在下游任务，上游视频预训练才是关键
  - 在三维空间进行时序建模更为合理

MEGVII 旷视

以非凡科技，为客户和社会持续创造最大价值