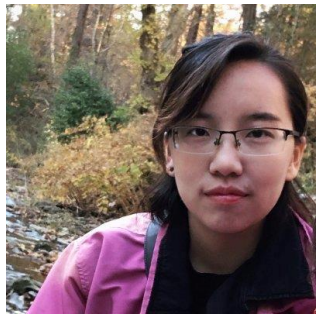
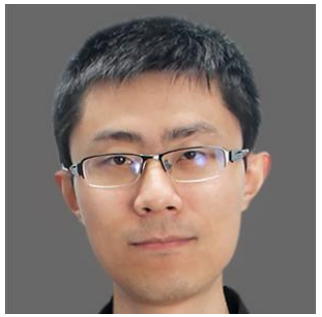
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Standing Between Past and Future: Spatio-temporal Modeling for Multi-Camera 3D Multi-Object Tracking

Ziqi Pang[1], Jie Li[2], Pavel Tokmakov[2], Dian Chen[2], Sergey Zagoruyko[3], Yu-Xiong Wang[1]





UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN
TOYOTA RESEARCH INSTITUTE
woven planet

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Standing Between Past and Future: Spatio-temporal Modeling for Multi-Camera **3D Multi-Object Tracking**

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

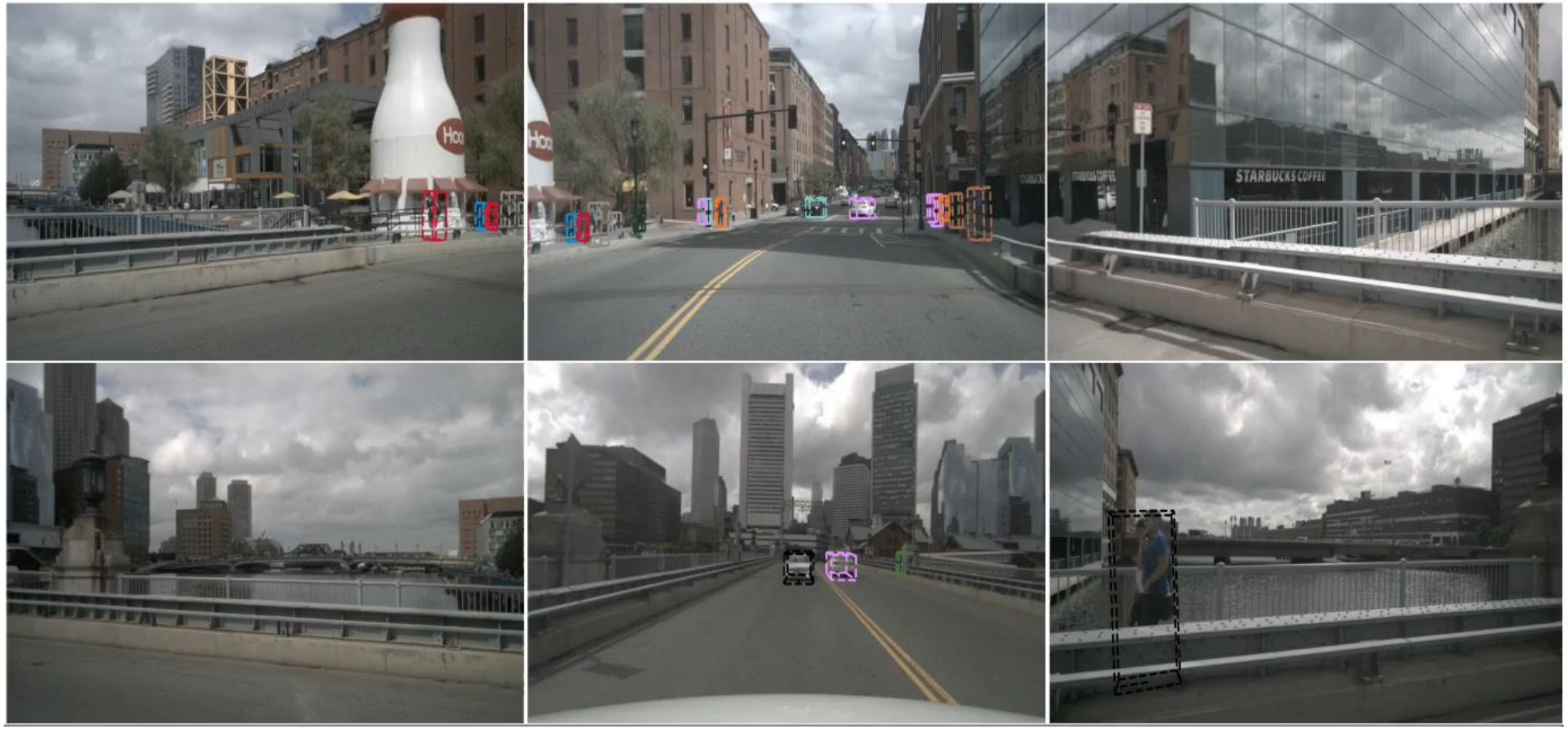
Standing Between Past and Future: Spatio-temporal Modeling for Multi-Camera 3D Multi-Object Tracking

"PF-Track"

Overview of
PF-Track

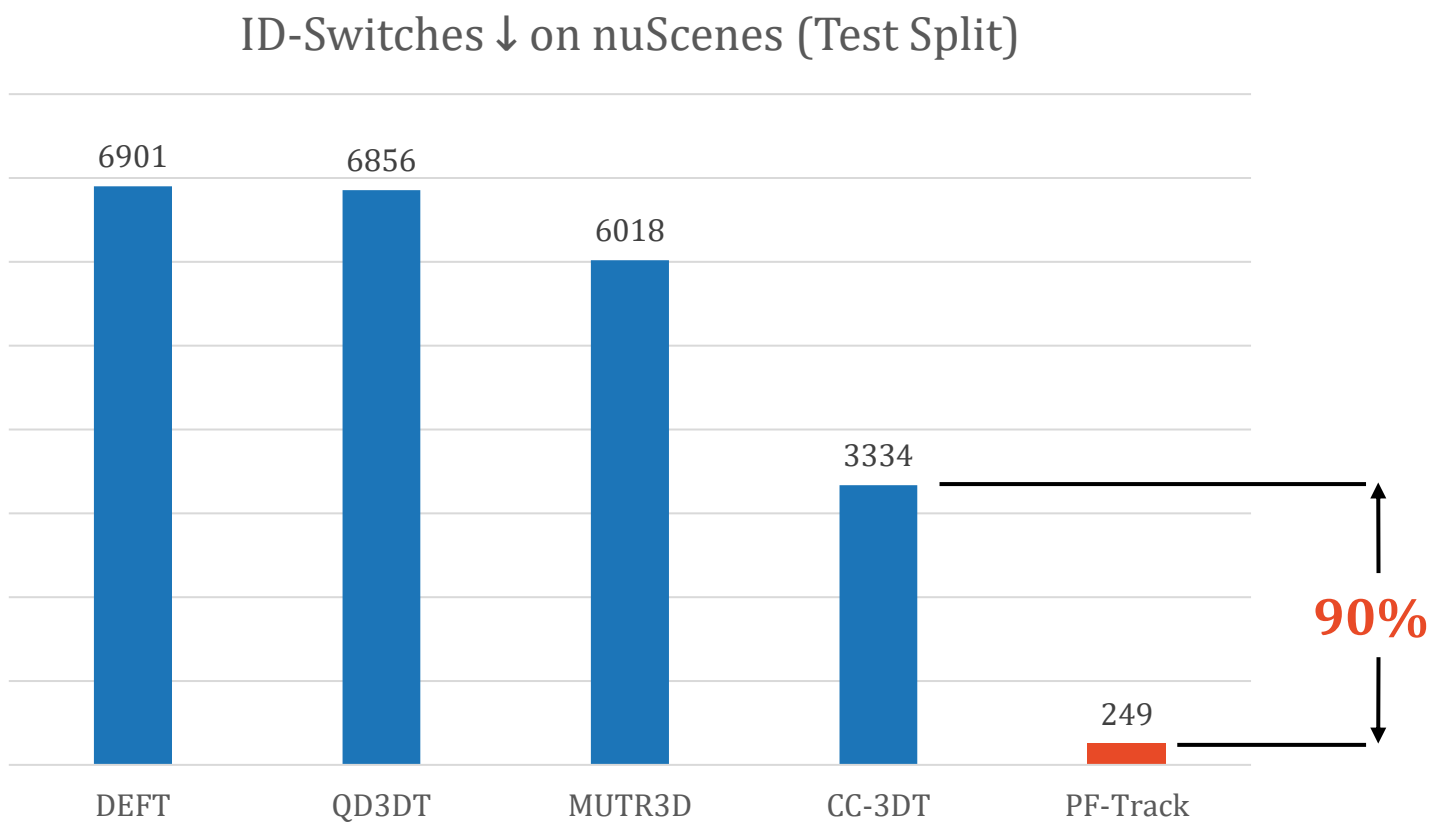
Overview: 3D MOT

- Coherently track objects overtime.



Overview: PF-Track's Effectiveness

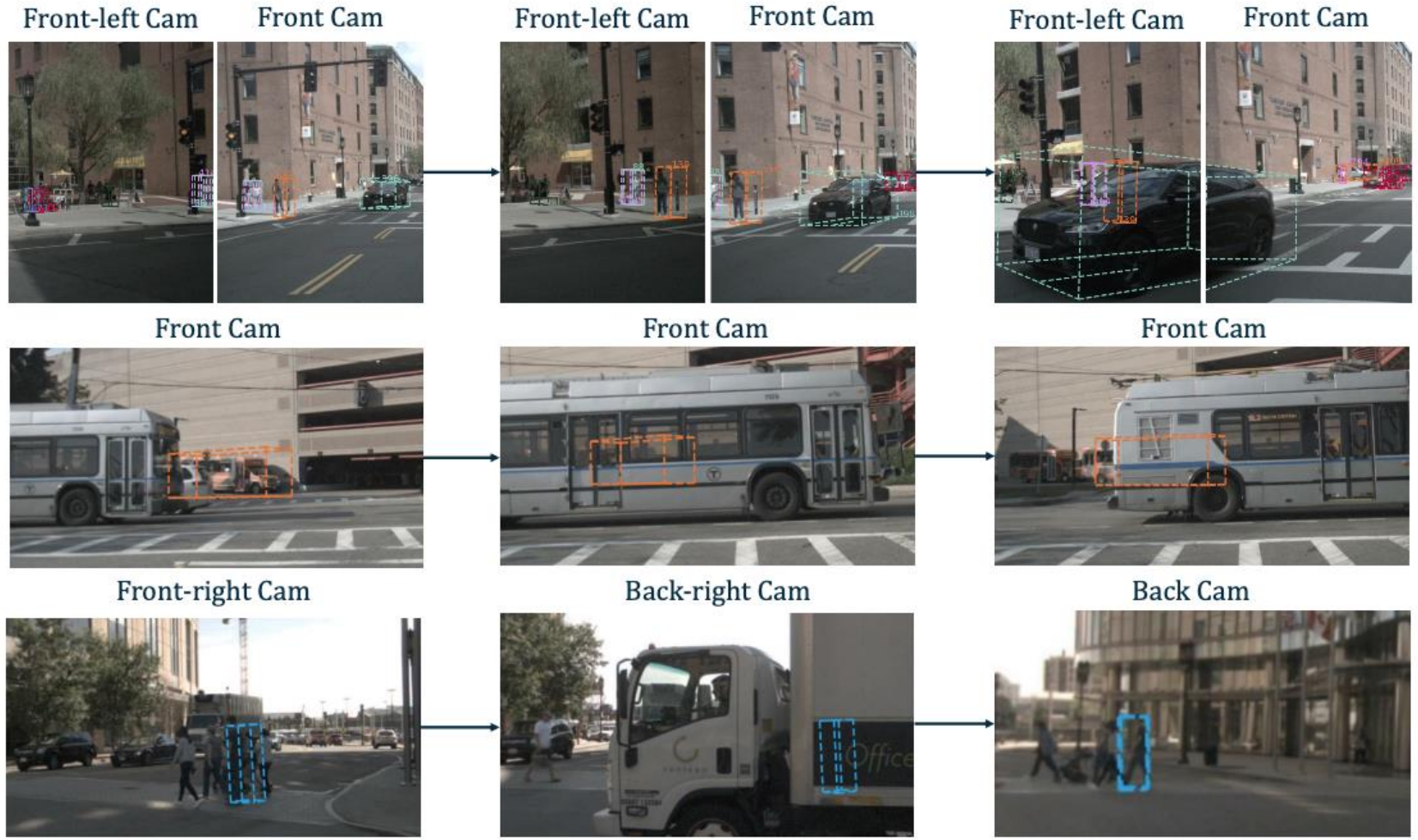
- **Accurate association**: **90% fewer** ID-Switches.



Compared on nuScenes test split (Caesar et al.). Caesar, Holger, et al. "nuScenes: A multimodal dataset for autonomous driving." *CVPR*. 2020.
Chaabane, Mohamed, et al. "Deft: Detection embeddings for tracking." arXiv preprint arXiv:2102.02267.
Hu, Hou-Ning, et al. "Monocular quasi-dense 3d object tracking." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.2 (2022): 1992-2008.
Fischer, Tobias, et al. "CC-3DT: Panoramic 3D Object Tracking via Cross-Camera Fusion." *CoRL* 2022.
Zhang, Tianyuan, et al. "MUTR3D: A multi-camera tracking framework via 3D-to-2D queries." *CVPR Workshop*. 2022.

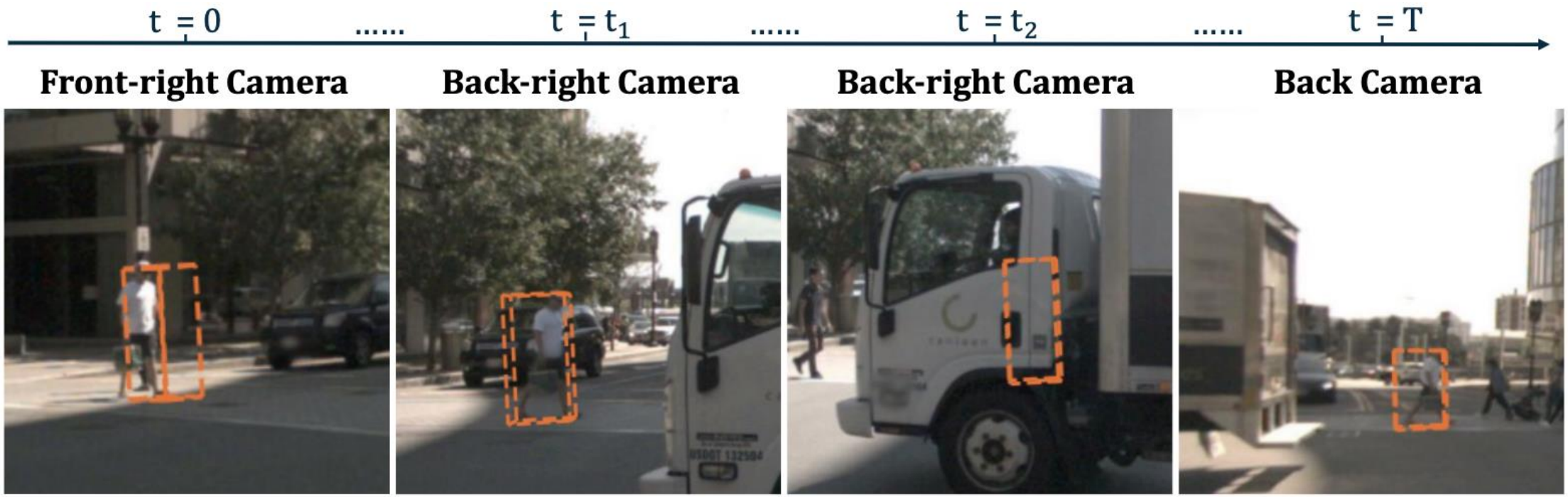
Overview: PF-Track's Effectiveness

- **Accurate association**: **90% fewer** ID-Switches.
- **Robustness**: **Camera hand-overs** and **Occlusions**.



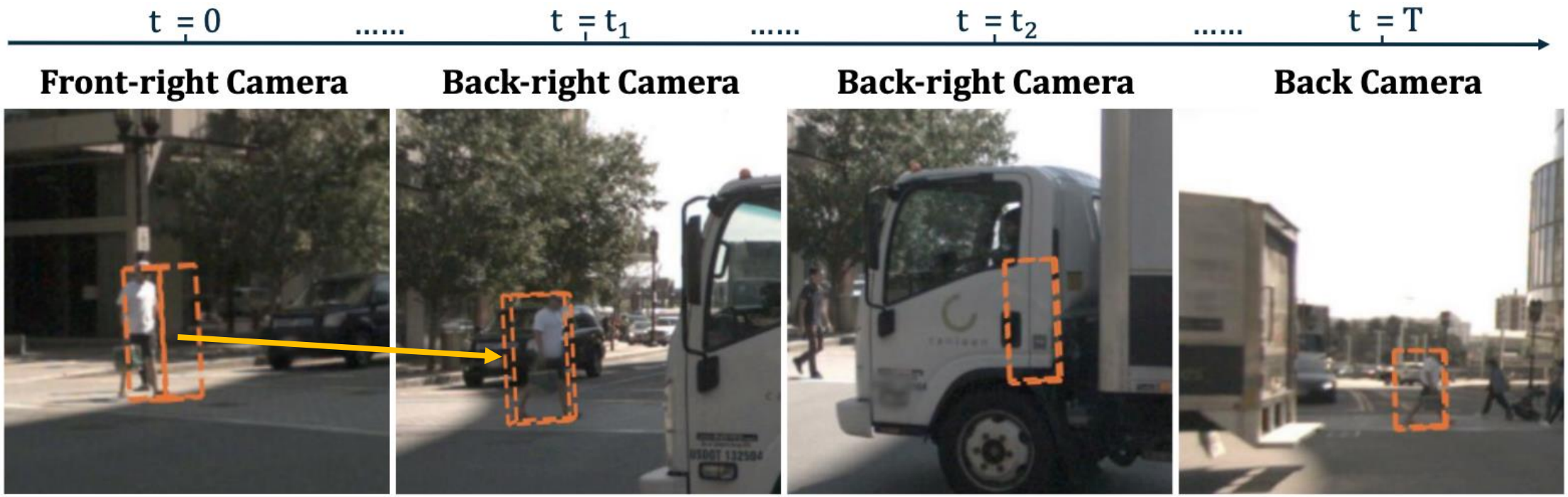
Overview: Recipe of PF-Track

- **Spatio-temporal modeling**: **Past** and **Future** reasoning.



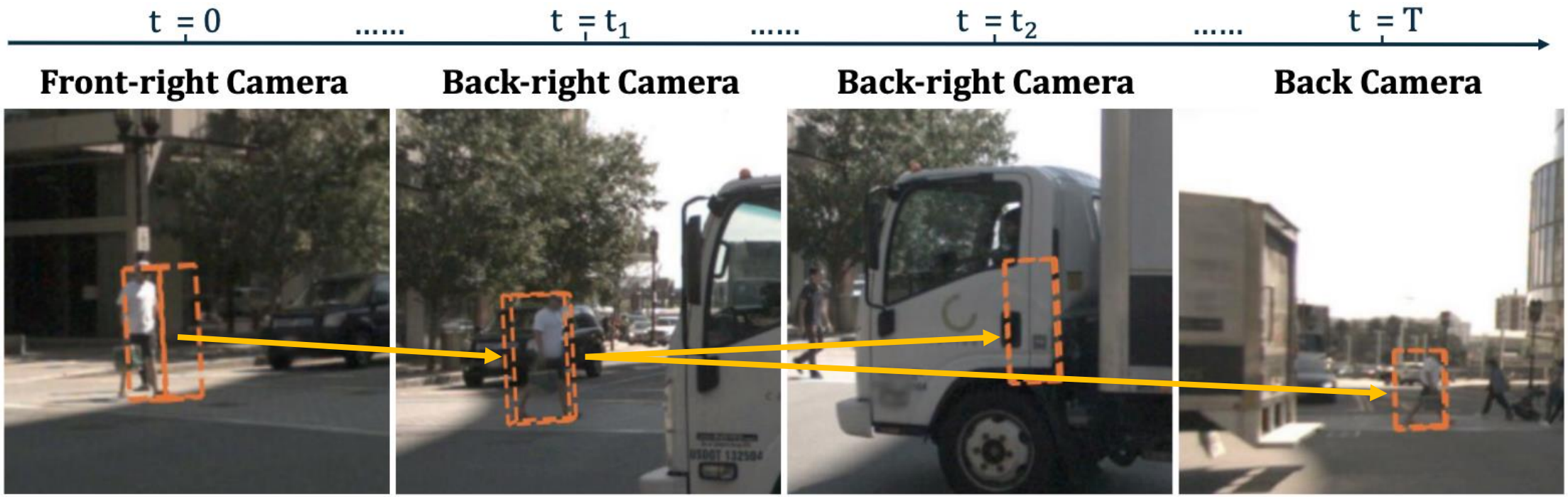
Overview: Recipe of PF-Track

- **Spatio-temporal modeling**: **Past** and **Future** reasoning.
- **Past reasoning**: Aggregate historical information for better bounding boxes.



Overview: Recipe of PF-Track

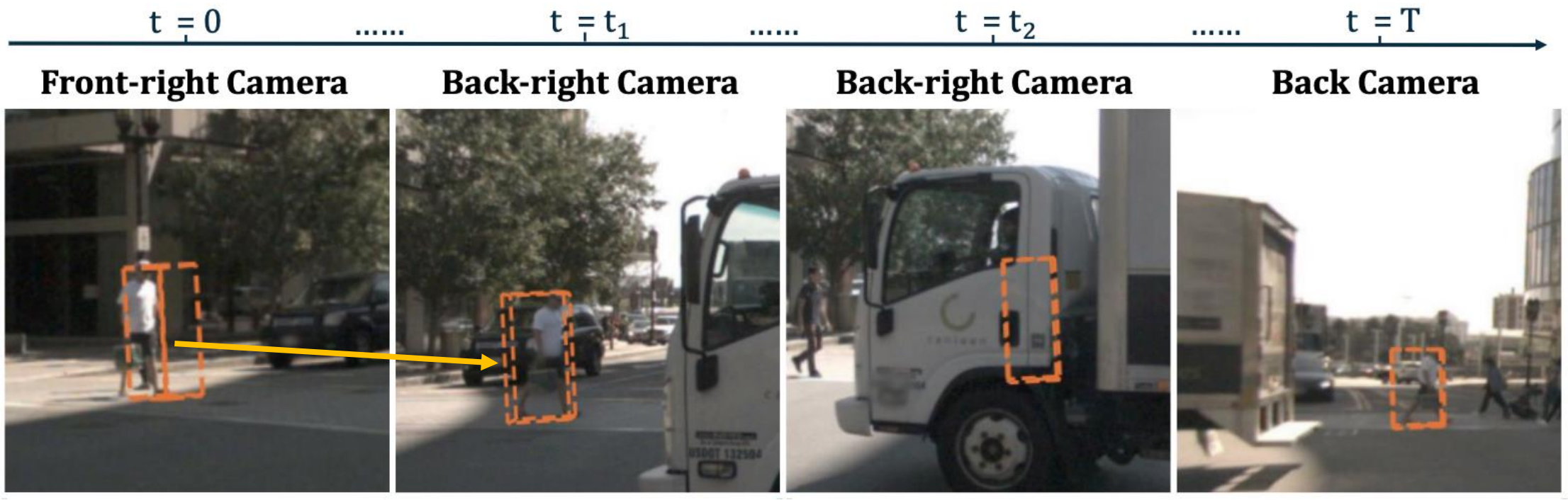
- **Spatio-temporal modeling**: **Past** and **Future** reasoning.
- **Past reasoning**: Aggregate historical information for better bounding boxes.
- **Future reasoning**: Motion prediction for tracking.



Framework of
PF-Track

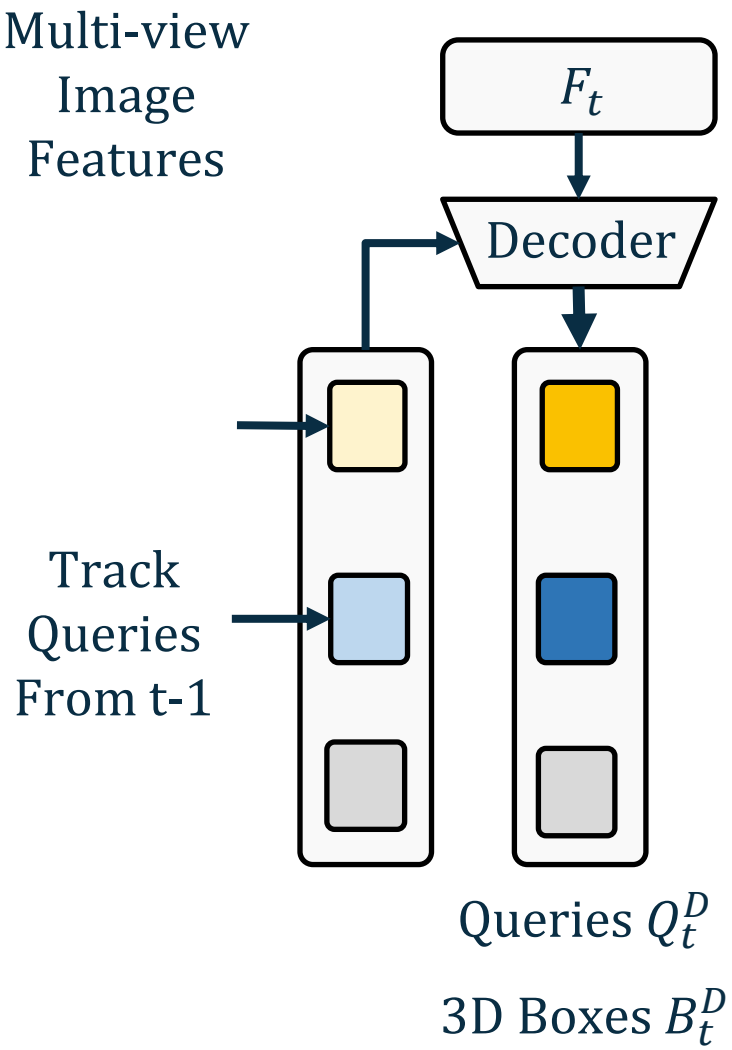
Framework

- **Query-centric framework**: Represent objects as queries.



Represent Objects as Queries

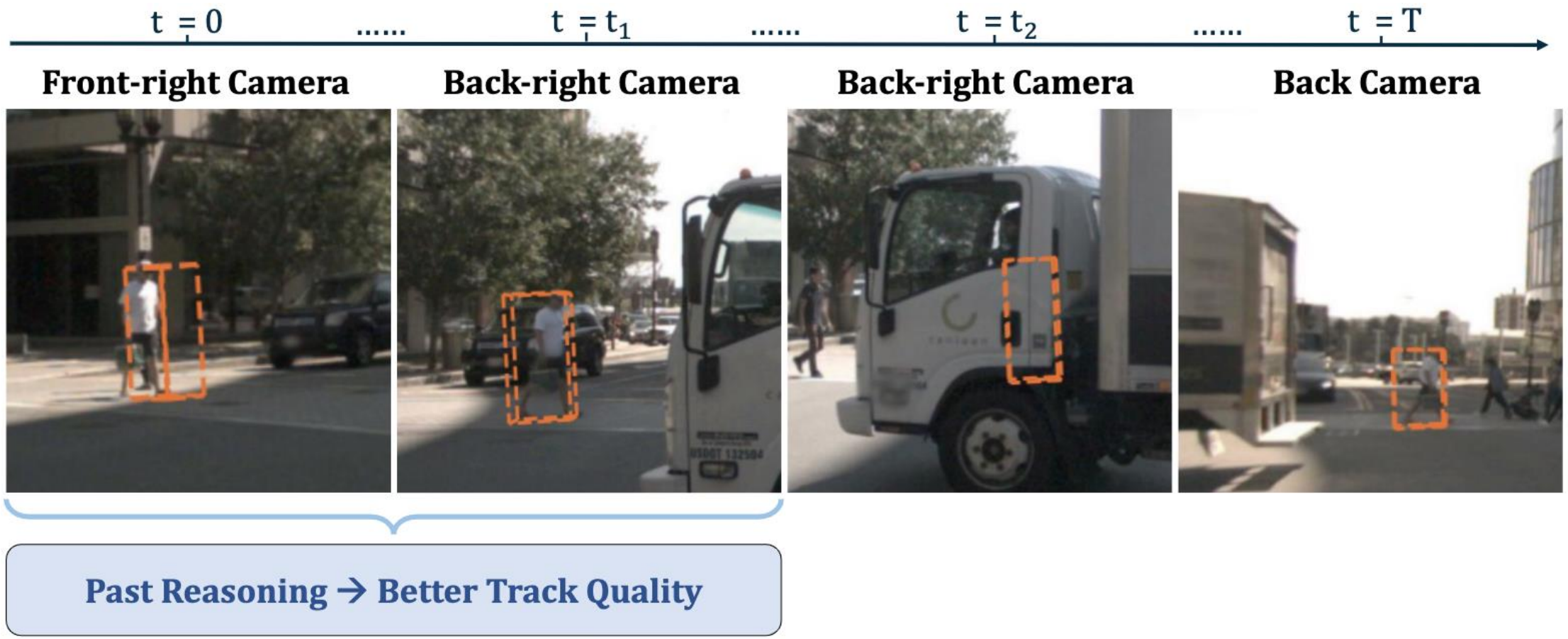
- Query-based detector: DETR3D, PETR, BEVFormer, etc.



Wang, Yue, et al. "DETR3D: 3D object detection from multi-view images via 3D-to-2D queries." *CoRL*, 2021.
Liu, Yingfei, et al. "PETR: Position embedding transformation for multi-view 3D object detection." *ECCV*, 2022.
Li, Zhiqi, et al. "BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers." *ECCV*, 2022.

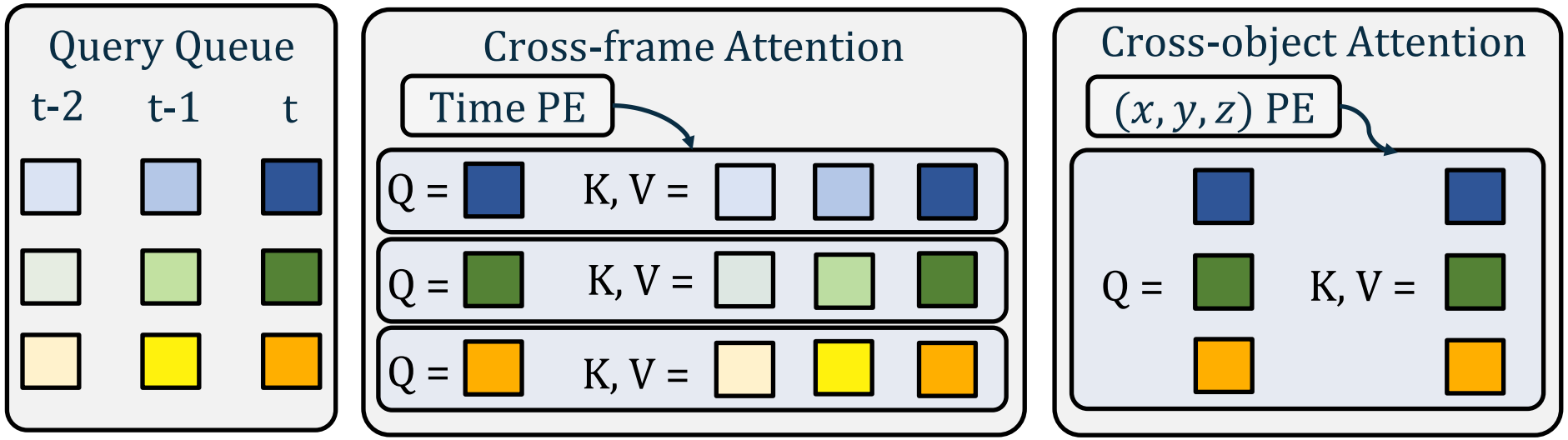
Framework: Past Reasoning

- **Query-centric framework**: Represent objects as queries.
- **Past reasoning**: Aggregate historical information.

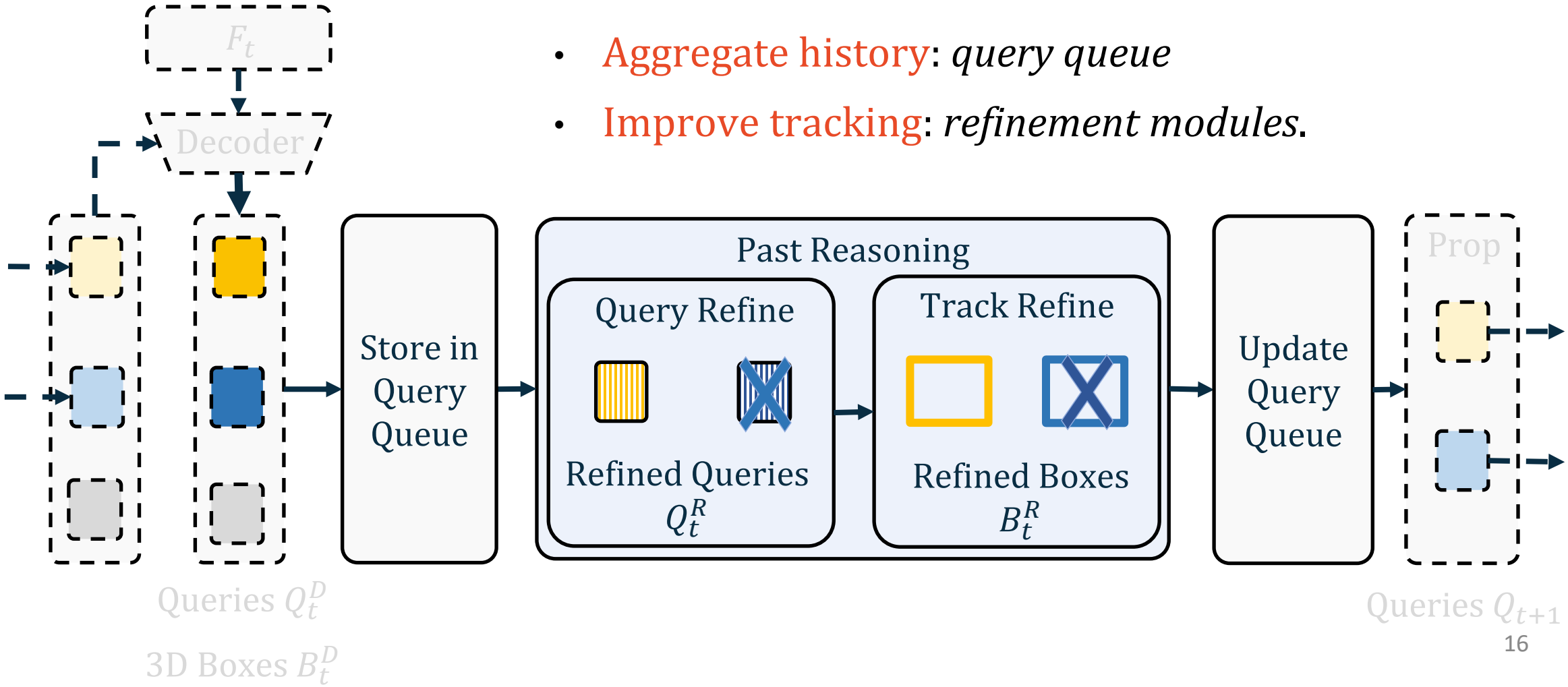


Past Reasoning

- How to leverage multi-frame information?
- Cross-frame attention → Temporal relationship.
- Cross-object attention → Spatial relationship.



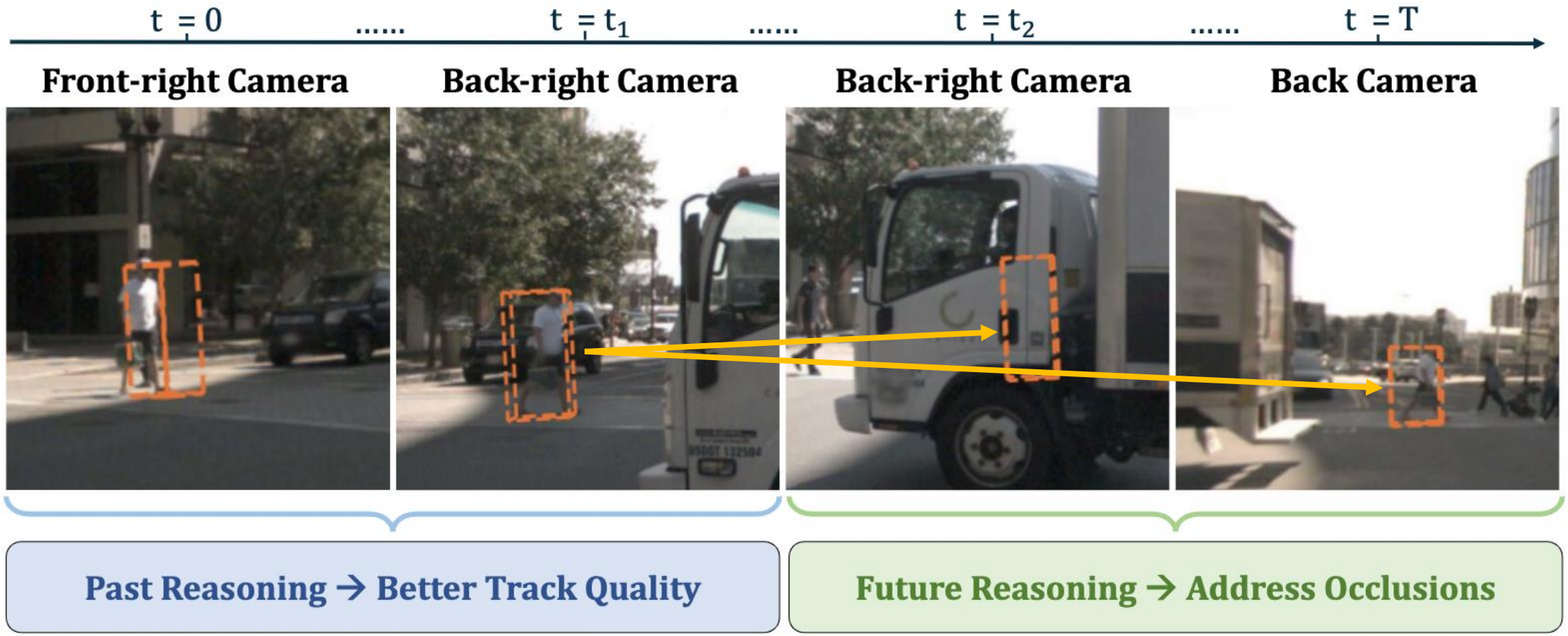
Past Reasoning



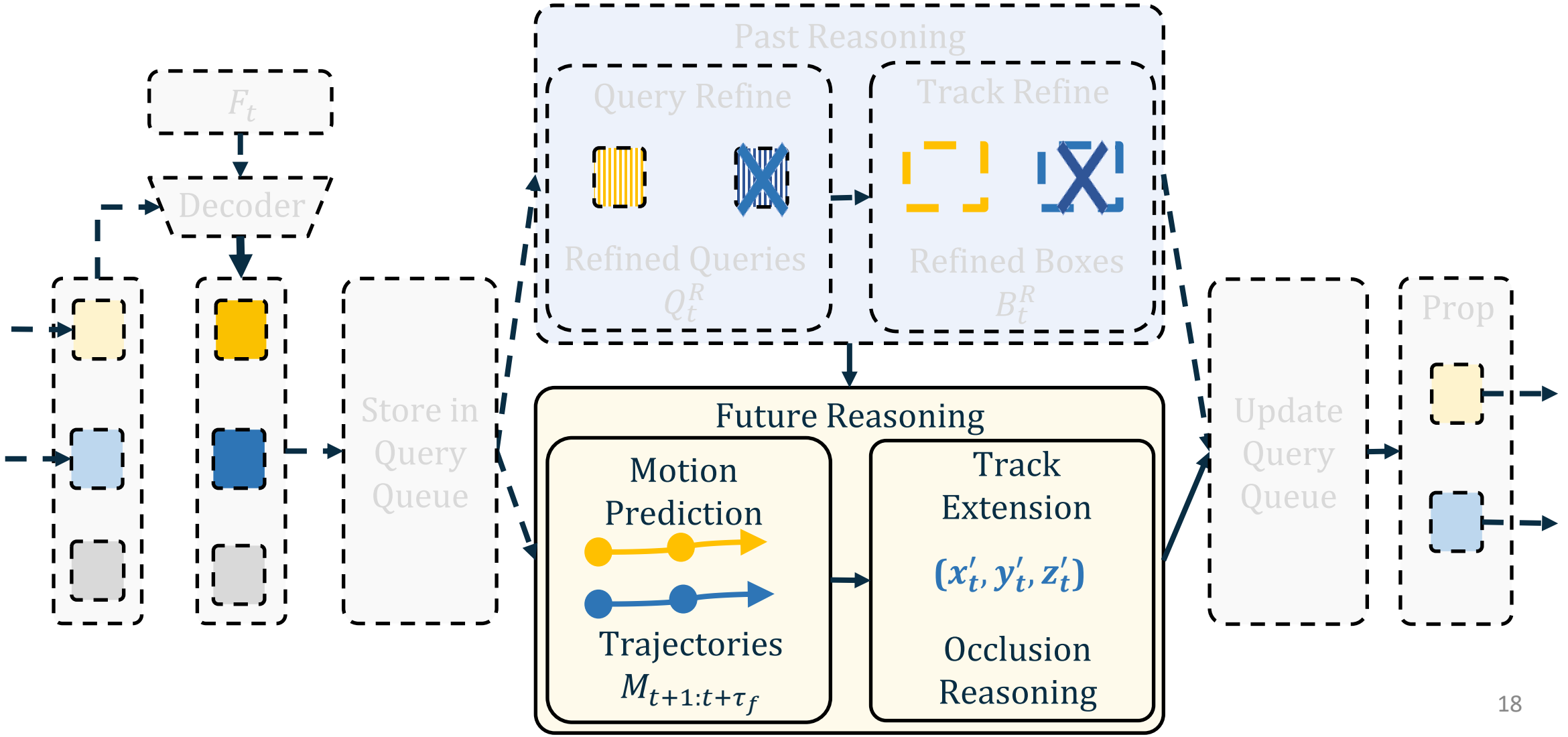
- Aggregate history: *query queue*
- Improve tracking: *refinement modules*.

Framework: Future Reasoning

- **Query-centric framework**: Represent objects as queries.
- **Past reasoning**: Aggregate historical information.
- **Future reasoning**: Forecast future movements for robust occlusion reasoning.

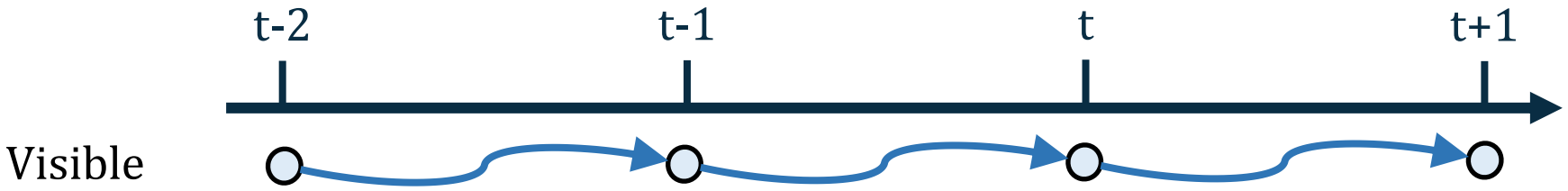


Future Reasoning



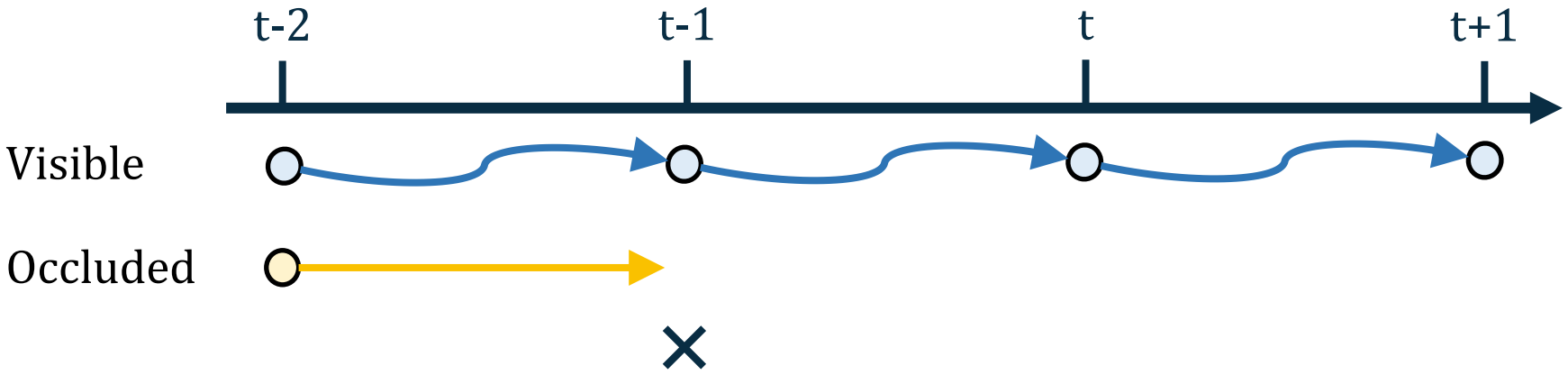
Method: Track Extension for Occlusion Reasoning

- Predictions on occluded frames are inaccurate.
- Solution: use trajectories from visible frames to propagate objects.



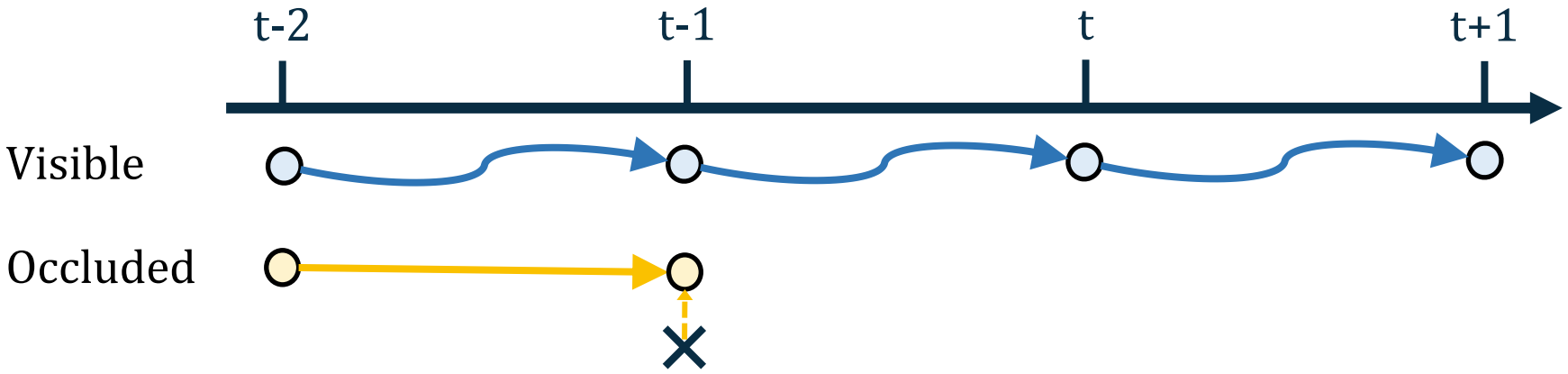
Method: Track Extension for Occlusion Reasoning

- Predictions on occluded frames are inaccurate.
- Solution: use trajectories from visible frames to propagate objects.



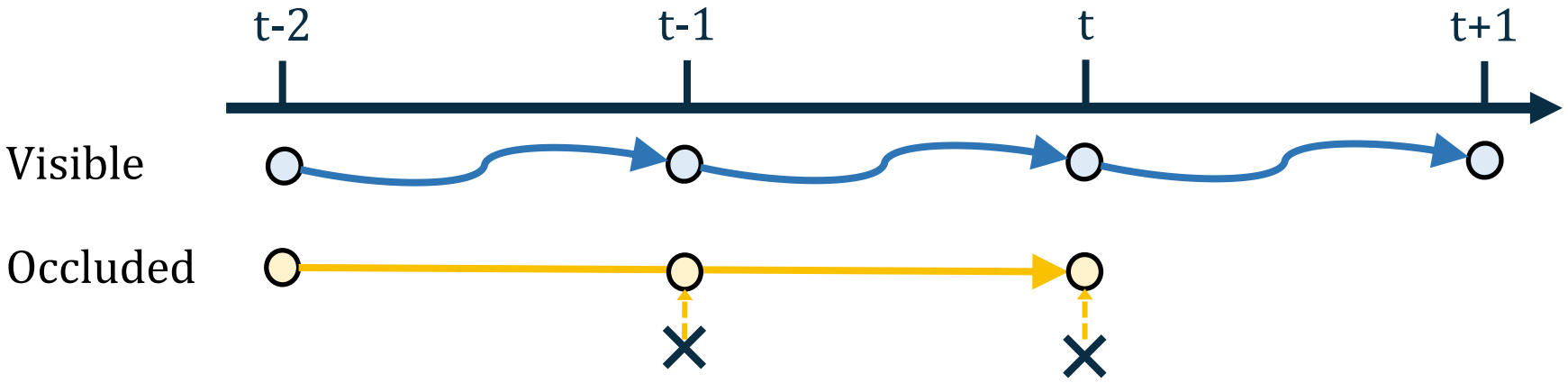
Method: Track Extension for Occlusion Reasoning

- Predictions on occluded frames are inaccurate.
- Solution: use trajectories from visible frames to propagate objects.



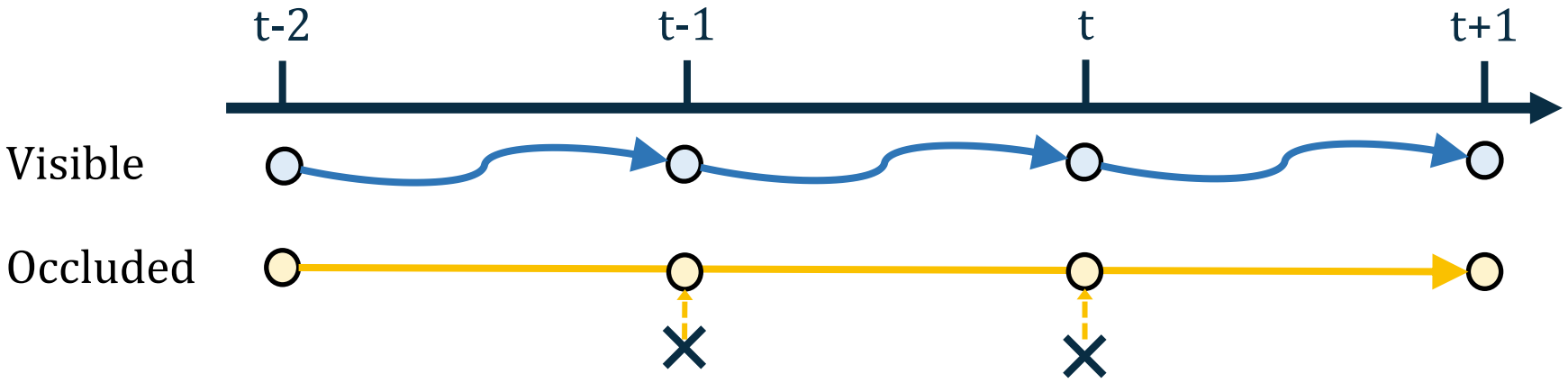
Method: Track Extension for Occlusion Reasoning

- Predictions on occluded frames are inaccurate.
- Solution: use trajectories from visible frames to propagate objects.

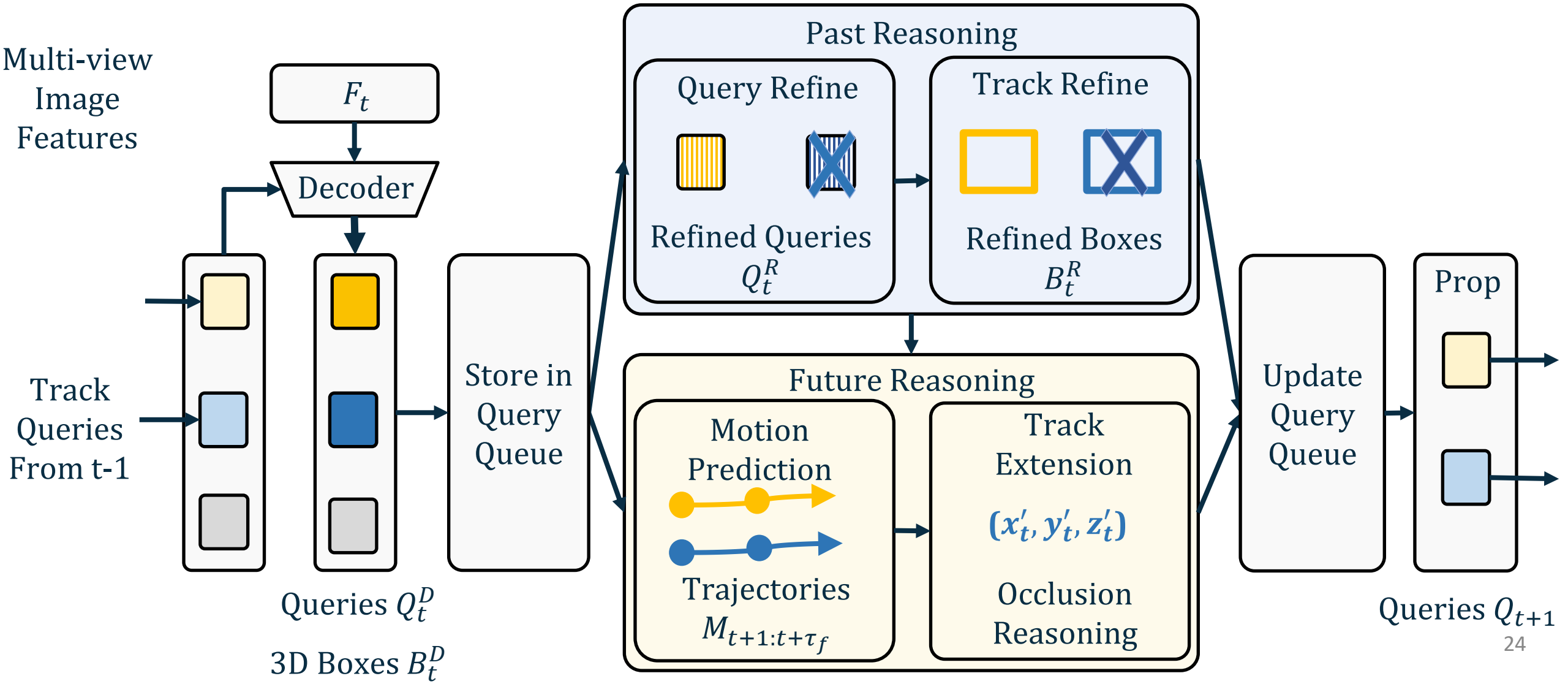


Method: Track Extension for Occlusion Reasoning

- Predictions on occluded frames are inaccurate.
- Solution: use trajectories from visible frames to propagate objects.



Overall Framework



Experiments of
PF-Track

Experiments (Comparison with SOTA)

Compared to monocular state-of-the-art, PF-Track is

- Higher in the major metric: AMOTA.
- **An order of magnitude less** in ID-Switch.

Method	AMOTA ↑	IDS ↓
MUTR3D (Zhang *et al.* 2022)	0.270	6018
CC-3DT (Fischer *et al.* 2022)	0.410	3334
PF-Track (Ours)	**0.434**	**249**

AMOTA: max(0, 1 - (IDS + FP + FN – (1 - r) * TP) / r * TP).

ID-Switch: Times of changed IDs per matched GT track.

Compared on nuScenes test split (Caesar et al.). Caesar, Holger, et al. "nuScenes: A multimodal dataset for autonomous driving." *CVPR*. 2020.
Fischer, Tobias, et al. "CC-3DT: Panoramic 3D Object Tracking via Cross-Camera Fusion." *CoRL* 2022.
Zhang, Tianyuan, et al. "MUTR3D: A multi-camera tracking framework via 3D-to-2D queries." *CVPR Workshop*. 2022.

Experiments (Ablation Study on Past & Future Reasoning)

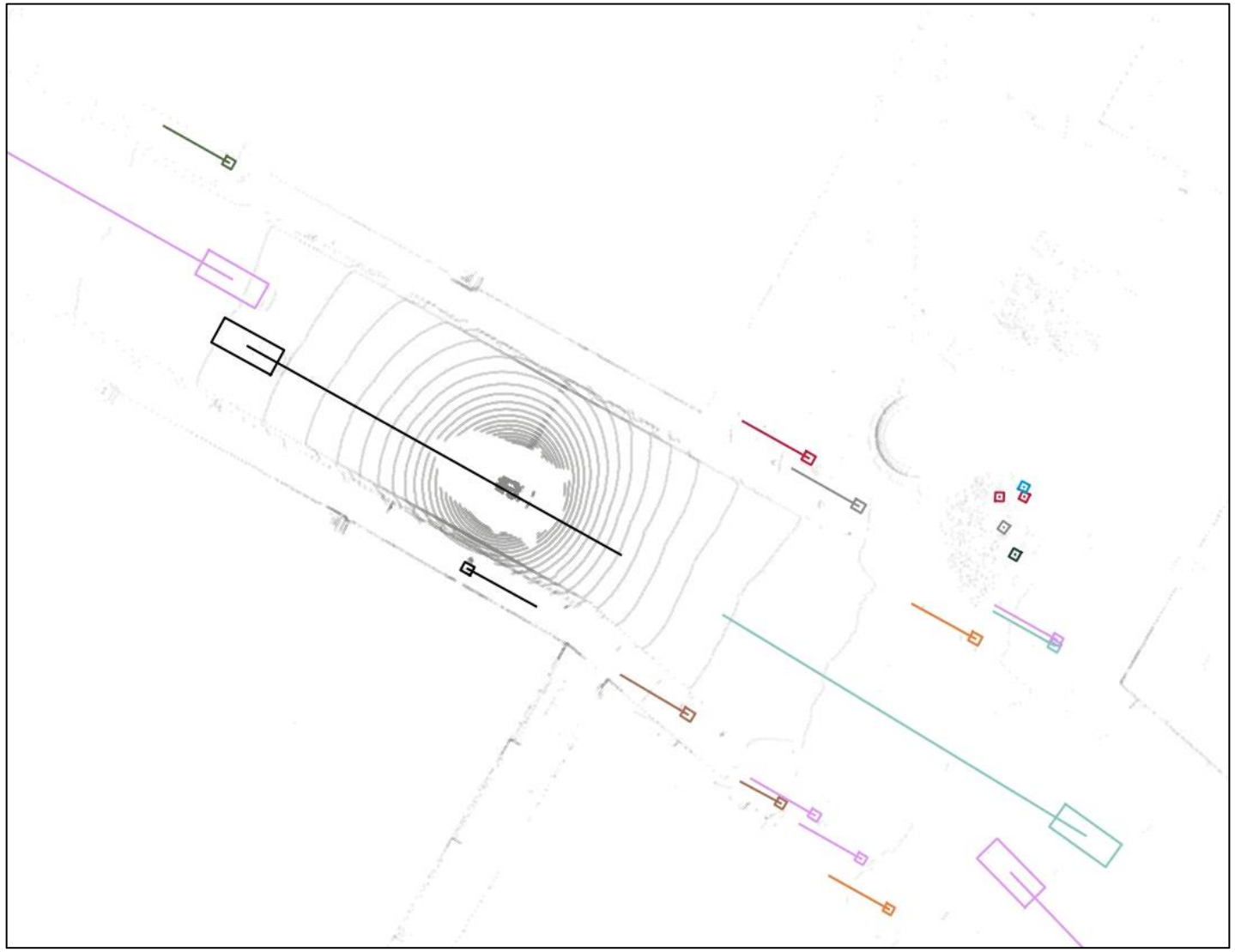
Past and future reasoning is:

- Beneficial separately.
- More beneficial jointly.

	Past Reasoning		Future Reasoning		AMOTA ↑	IDS ↓
	Query Refinement	Track Refinement	Motion Prediction	Track Extension		
1					0.368	507
2	✓				0.378	453
3	✓	✓			0.380	400
4			✓		0.374	469
5			✓	✓	0.391	**155**
6	✓	✓	✓	✓	**0.408**	166

Experiments (Prediction Video)

- Predict future trajectories end-to-end.



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thank you!

https://github.com/TRI-ML/PF-Track

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

TOYOTA RESEARCH INSTITUTE

woven planet