Multi-Camera Multiple 3D Object Tracking on the Move for Autonomous Vehicles

Pha Nguyen, Kha Gia Quach, Chi Nhan Duong, Ngan Le, Xuan-Bac Nguyen, Khoa Luu

https://cviu.uark.edu/

The 4th IEEE/CVF CVPR Precognition Workshop
CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

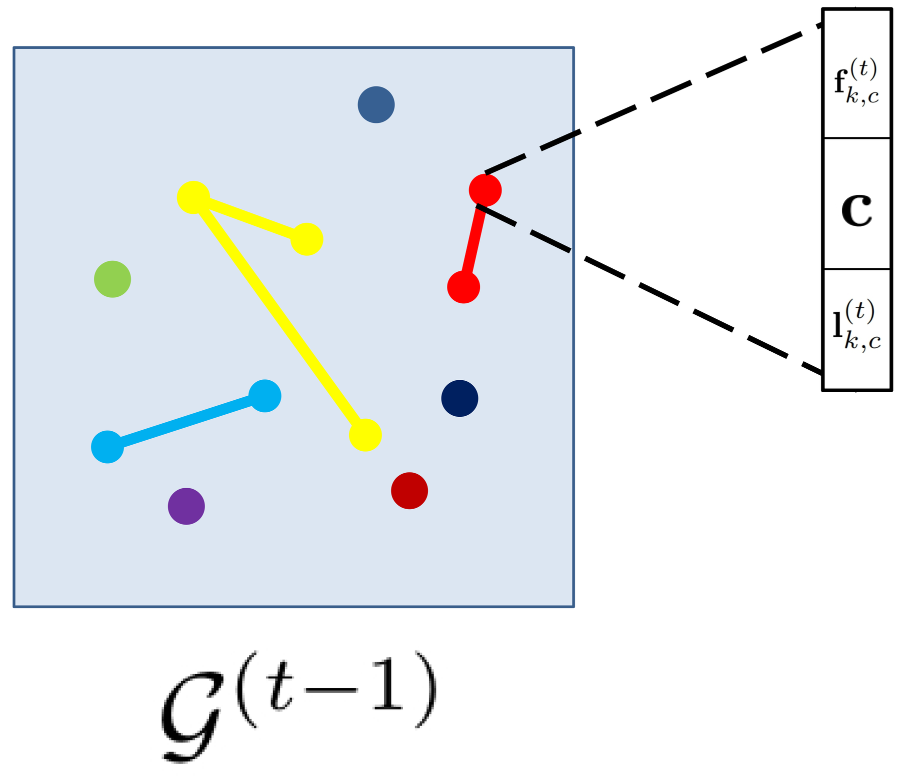
1. Motivation

- Existing two-stage MCMOT approaches firstly apply the MOT on each camera individually. Then links local tracklets together via global matching steps based on Re-ID features.
- Those approaches create more errors, i.e. fragmented local tracklets, and more computation, since the data association and the matching steps will perform multiple times both locally and globally.
- They are also unable to handle scenarios when the detector fails to detect objects from one of the cameras.

2. Contributions

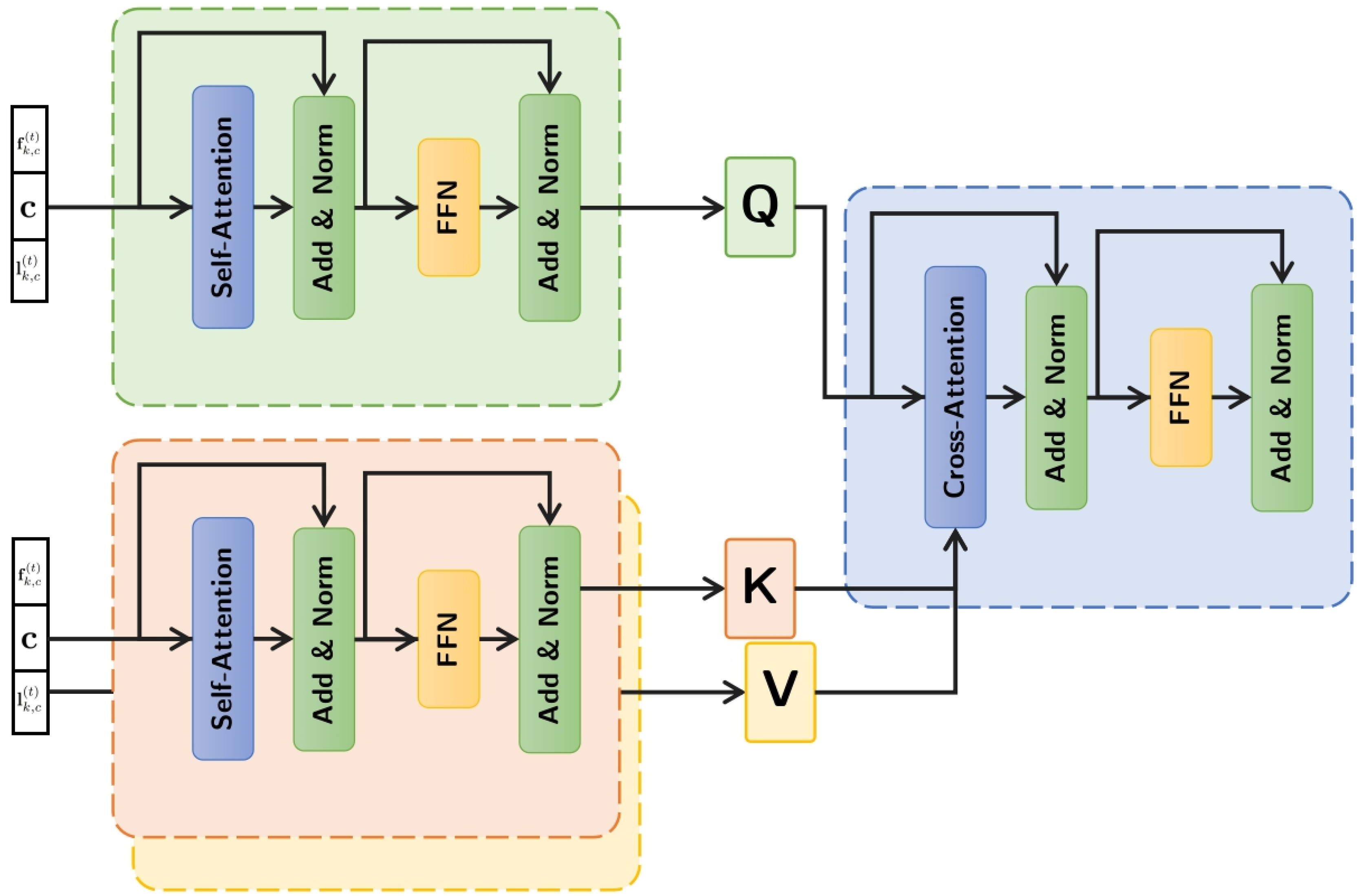
- Propose a global graph whose nodes represent tracked objects and edges represent the relation between the nodes.
- Present an *Auto-regressive Graph Transformer* network including self-attention layer to transform appearance features and cross-attention to predict the motion features of objects.

3. Global Graph Constructing

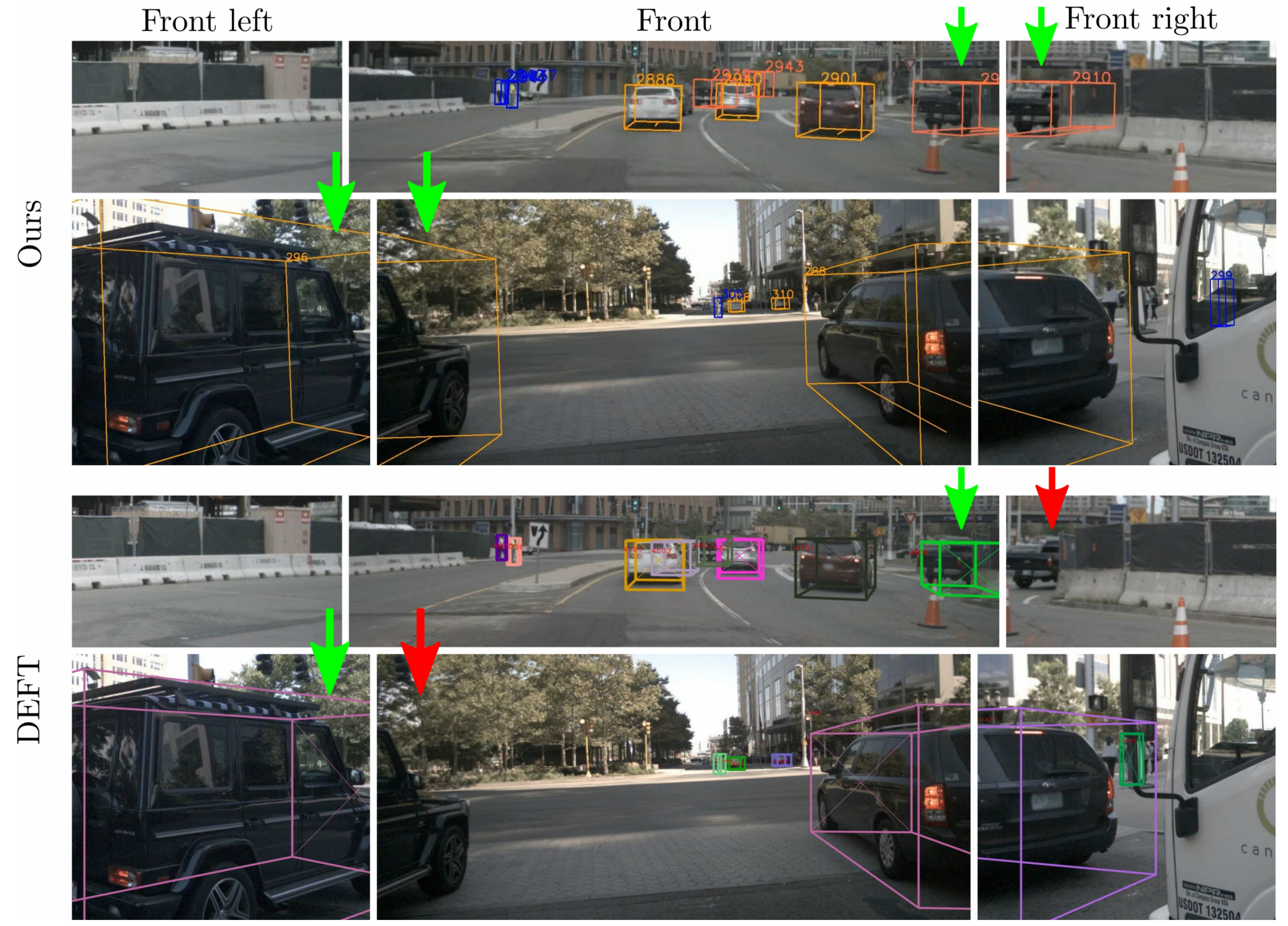


A single graph is constructed and maintained across time, whose input node features as the concatenation of embedding features with camera and location encoding.

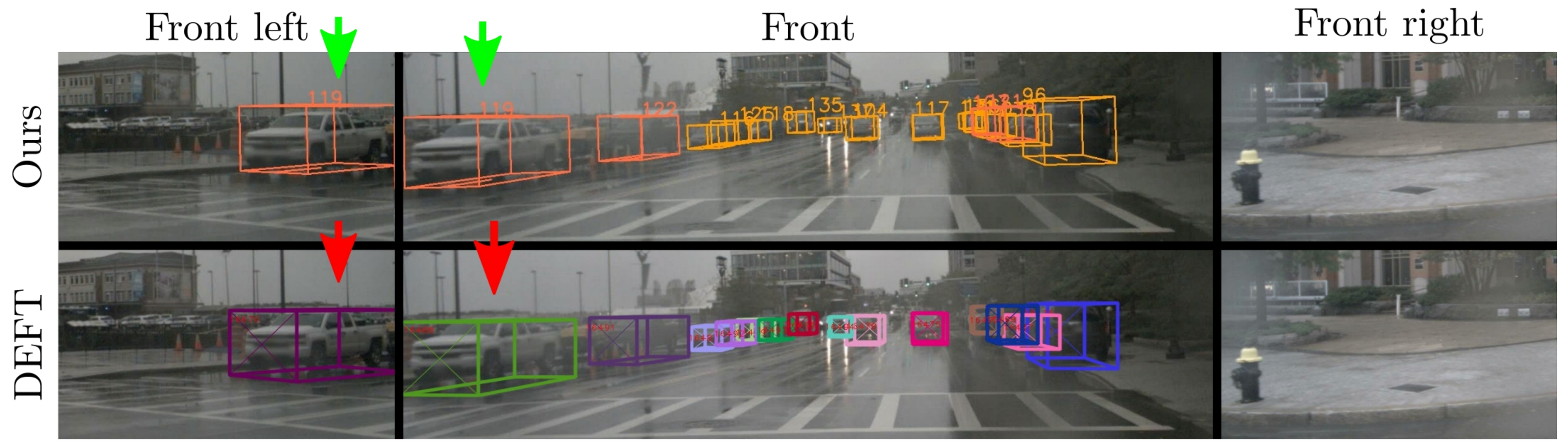
4. Auto-Regressive Graph Transformer Networks



5. Quantitative and Qualitative Results



Our proposed method (top) can recover a false negative detection case



Our proposed method can recognize a positive tracking case

Method	Glo. Assoc.	AMOTA	MT ↑	ML ↓	IDS ↓	FRAG ↓
MonoDIS [29]	✗	0.045	395	3961	6872	3229
CenterTrack [42]	✗	0.068	524	4378	2673	1882
DEFT [5]	✗	0.213	1591	2552	5560	2721
QD-3DT [15]	✗	**0.242**	1600	2307	5646	2592
Ours	✓	0.240	**1643**	**2162**	**1362**	**1462**

Comparison of 3D tracking performance on the nuScenes validation set

6. Acknowledgment

This material is based upon work supported in part by the US NSF Data Science, Data Analytics that are Robust and Trusted (DART) and NSF WVAR-CRESH Grant.