AUTONOMOUS TRUCKS: SENSOR FUSION FOR OBJECT CLASSIFICATION AND TRACKING

JASON XIE, HENRY HUNG, DALE SIMPSON, KUSHAGRA KUMAR

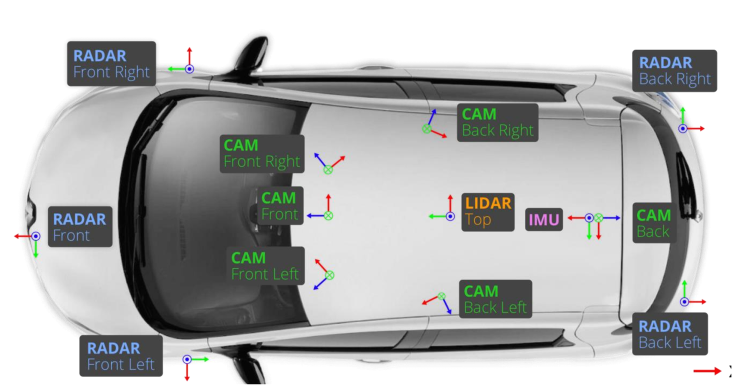
INTRODUCTION & BACKGROUND

- Reliable visual perception plays a critical role in enabling autonomous vehicles to safely navigate unseen, unstructured environments.
- In order to anticipate and avoid obstacles, such a perception system needs to detect, classify, localize, and track dynamic objects within range of the vehicle.
- Many perception systems in state-of-the-art autonomous vehicles rely on LiDAR (light detection and ranging) to produce an accurate geometric representation of the vehicle's environment; however, such systems can be costly to acquire and maintain.
- Our project focuses on object detection and tracking from 2D RGB camera inputs, owing to their relative low cost and capacity to capture dense representations of scene textures.

SYSTEM REQUIREMENTS

- Inputs to the system are a stream of RGB images captured by each camera: $\{\dots, I_{t-2}, I_{t-1}, I_t\}$, $I_t \in [0,1]^{H \times W \times 3}$
- Objects are represented as 3D cuboid "bounding boxes", parameterized by their center point $(x, y, z) \in \mathbb{R}^3$, size dimensions $(length, width, height) \in \mathbb{R}^3$, orientation in the plane of the ego vehicle $\theta \in \mathbb{R}^3$, and their class - one of $\{bicycle, motorcycle, pedestrian, bus, car, trailer, truck\}$ - and a unique identity signature s.
- The tracker attempts to associate the identity signatures of detected objects to the same objects detected at previous time steps, thereby generating a trajectory of each object's path across time.

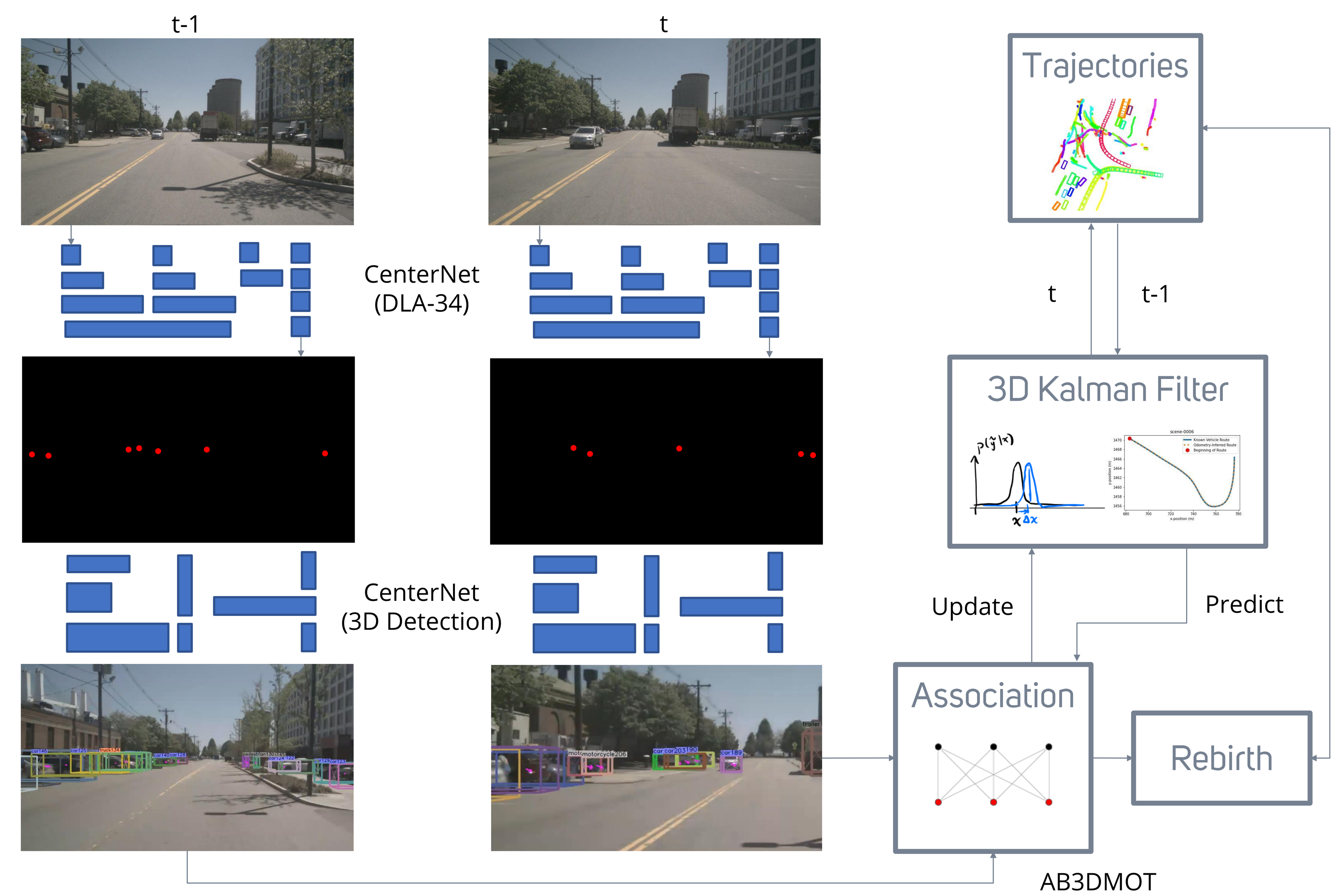
HARDWARE & DATA CONFIGURATION



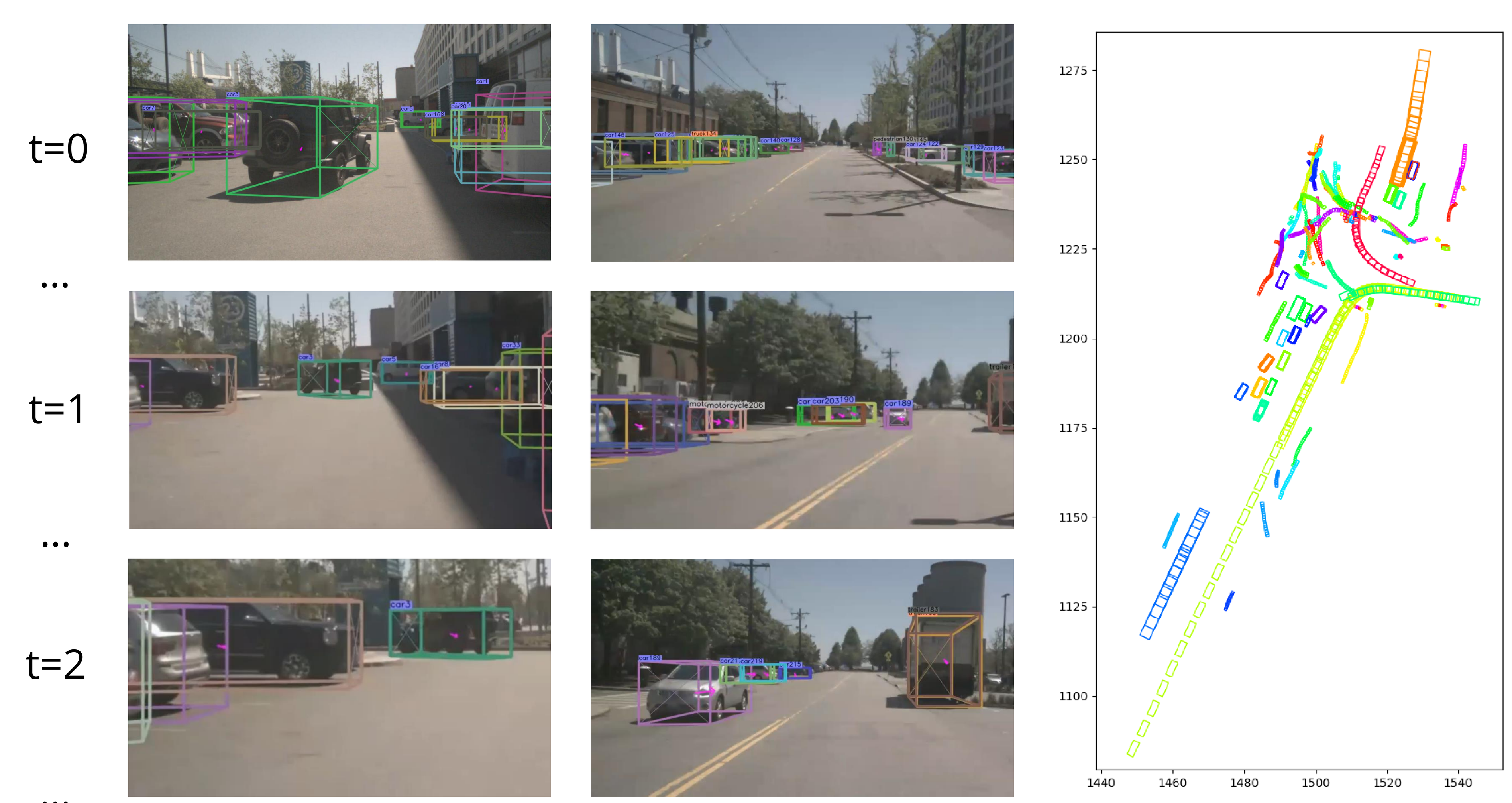
Vehicle Setup. From [0]

- We evaluate the system on the nuScenes dataset [0], which consists of 15h of driving data in Boston and Singapore across a variety of urban traffic scenarios, times, and weather conditions.
- Ground truth annotations are provided 2 Hz.
- All objects that are not directly visible to cameras or greater than 40m away from the ego vehicle are removed prior to evaluation.
- Ego Vehicle: Renault Zoe
- Cameras: 6x Basler acA1600-60gc, Lens F1.8 f5.5mm 1/1.8" @ 12 Hz
- Radar: 5x Continental ARS 408-21 @ 13 Hz
- IMU & GPS: 1x Advanced Navigation Spatial
- LiDAR point clouds are not provided as input to the perception pipeline.

SYSTEM DESIGN



QUALITATIVE RESULTS



IMPLEMENTATION DETAILS

- We use the DLA-34 implementation of CenterNet [2] with deformable convolutions as the backbone feature extractor, with additional layers to infer the 3D bounding box characteristics of all objects in the environment.
- All object states are transformed and compensated by odometry estimates from the ego vehicle's CAN data.
- Overlapping objects from different views are resolved via NMS (non-maximum suppression).
- Data association is performed via the Hungarian algorithm method over 3D IOU (intersection-over-union) as in AB3DMOT [3]. Unmatched tracks are kept for $t = 3$ before they are deleted.
- Object states are tracked with a 3D Kalman Filter [3] with a velocity augmented state. Motion models and uncertainty estimates are tuned based on the class of the tracked object $\{vehicle, cyclist, pedestrian\}$.
- At inference time, full sweeps of the data at native sensor sampling rates are used for detection and tracking.

FUTURE WORK

- Depth estimation is challenging with a single camera - a stereo camera setup provides more robust 3D resolution and may be better suited to safety critical applications.
- Radar was underutilized in this project due to its sparsity, high prevalence of false positives; however, radar returns instantaneous radial velocity estimates which can be used to refine vehicle motion estimates.

CONCLUSION & ACKNOWLEDGEMENTS

- We present a camera-based perception pipeline for 3D object classification, detection, and tracking of dynamic objects in an autonomous driving context. The system runs in near real-time on a desktop GPU and generates object tracks that can be fed downstream to a motion planning and control system for navigation in autonomous driving.
- Acknowledgements: We thank Austin Thind, Prof. Blake Hannaford, Prof. Payman Arabhsahi, and Daniel King for their valuable feedback throughout this project.

REFERENCES

1. "nuScenes: A multimodal dataset for autonomous driving", H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, In arXiv 2019.
2. "Objects as Points", X. Zhou, D. Wang, and P. Krahenbuhl, In arXiv 2019.
3. "A Baseline for 3D Multi-Object Tracking", X. Weng and K. Kitani, In arXiv 2019.

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

ADVISORS: PROF. BLAKE HANNAFORD, PROF. PAYMAN ARABSHAHI, DANIEL KING

INDUSTRY MENTOR: AUSTIN THIND (PACCAR)

SPONSOR: PACCAR Inc.