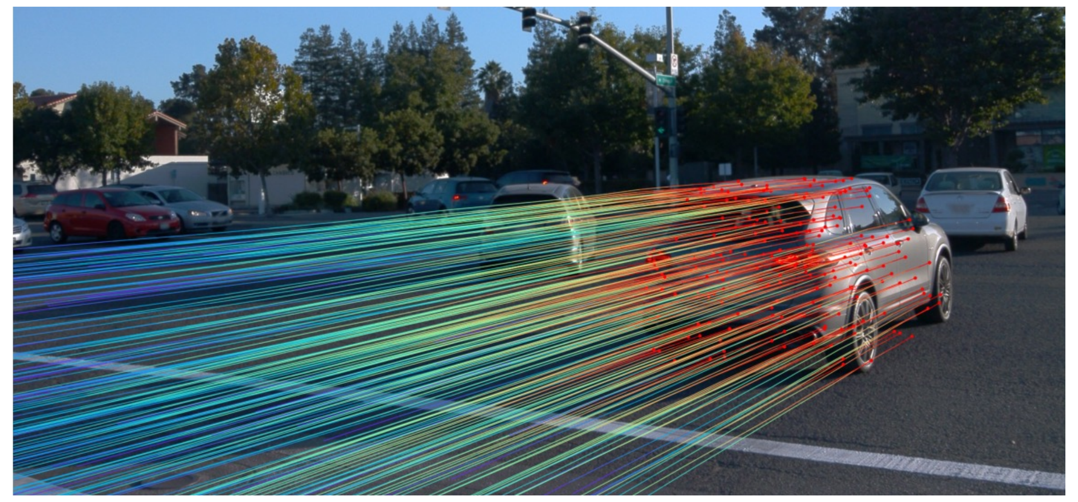
DriveTrack: A Benchmark for Long-Range Point Tracking in Real-World Videos

Arjun Balasingam, Joseph Chandler, Chenning Li, Zhoutong Zhang*, Hari Balakrishnan
MIT CSAIL and *Adobe Systems

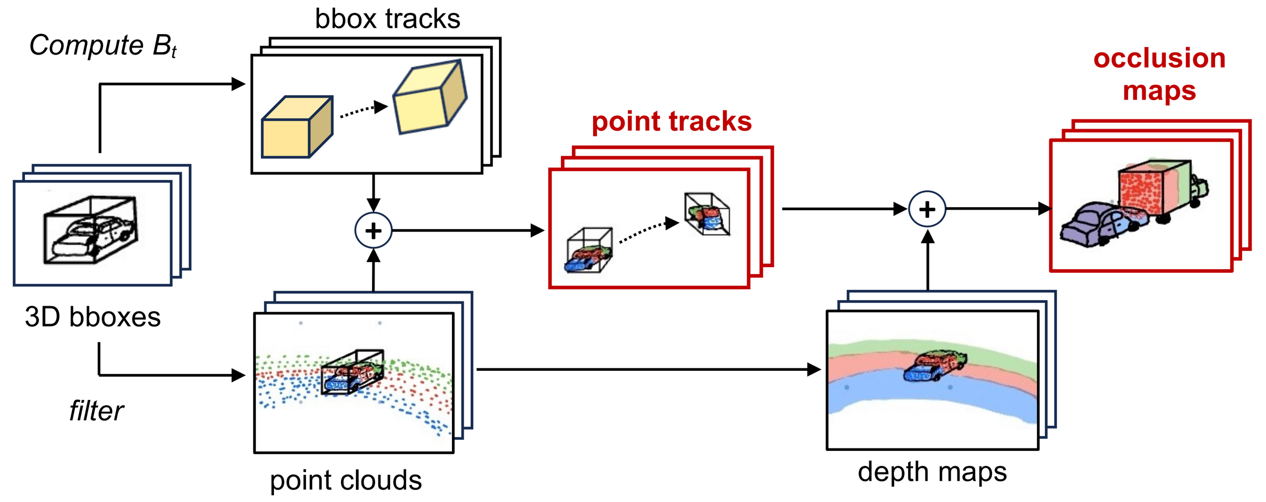
1 Introduction

- Track Any Point (TAP) = given a video + query points, estimate locations of those points in all other frames in which they are visible
- State-of-the-art trackers have poor performance on real-videos
- Noisy visual characteristics can hinder tracking performance
- Problem: modern trackers train on vast *synthetic* datasets



- DriveTrack is a large-scale benchmark for long-range point tracking
- Brings to real-world videos the density and fidelity of annotations available only for synthetic benchmarks today
- Uses autonomous driving videos + labels to annotate point tracks in real-world scenes
- DriveTrack has 1 billion point tracks across 24 hours of video

2 DriveTrack Overview



- Transform each timestamped point cloud according to the vehicle's camera pose and a target object's bounding boxes
- Estimate occlusions by computing depth maps
- Construct point tracks for all points in point clouds

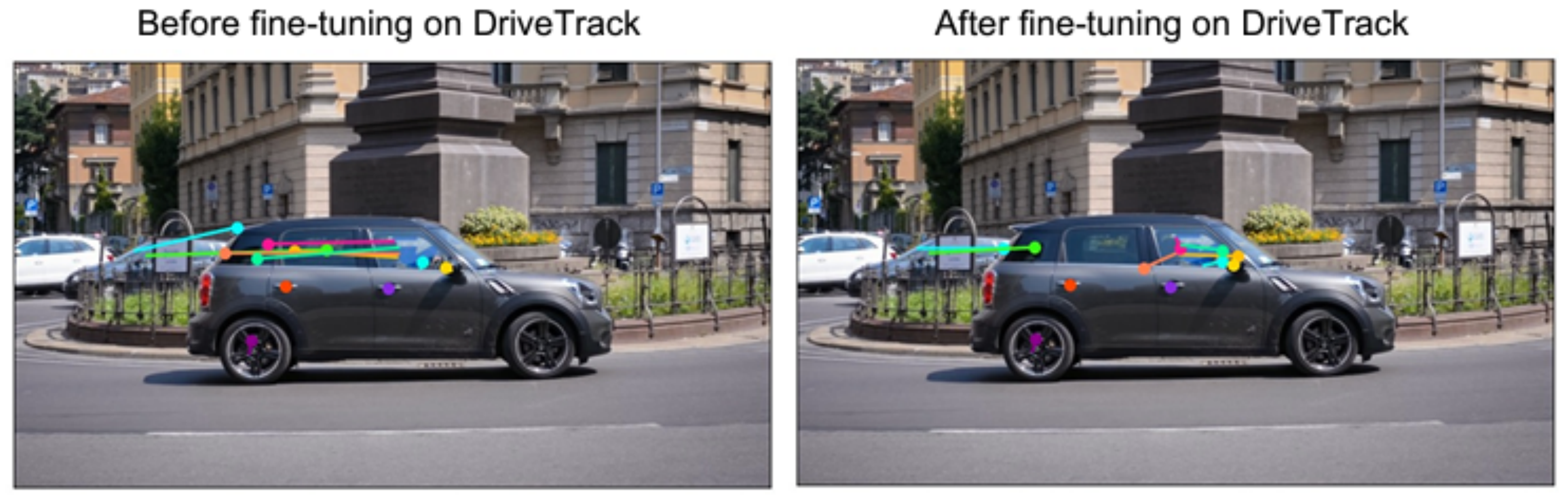
- Point tracks computed by DriveTrack on a scene in Waymo Open Dataset
- Visible points with ·, occlusions with x
- Point tracks + occlusions are extremely accurate

3 Fine-tuning Keypoint Trackers with DriveTrack

- Finetuned on 900 24-frame training annotations
- Used state-of-the-art trackers (PIPs++, TAP-Net, TAPIR)
- Trained for 5000 steps, stopped when performance reached peak
- Demonstrated up to 7% performance improvement on DriveTrack
- Performance improves on TAP-Vid DAVIS even when only fine-tuning on DriveTrack

Tracker	Training	Kubric [9]			DAVIS [19]			DriveTrack		
		AJ	$<\delta^x_{avg}$	OA	AJ	$<\delta^x_{avg}$	OA	AJ	$<\delta^x_{avg}$	OA
TAP-Net [4]	Kubric [9]	**65.4**	**77.7**	**93.0**	38.4	53.1	**82.3**	63.6	73.8	92.4
	+ DriveTrack	37.0	54.0	83.5	**39.2**	**54.7**	78.6	**70.3**	**80.4**	**93.2**
TAPIR [5]	Panning Kubric	**84.7**	**92.1**	**95.8**	62.8	74.7	**89.5**	78.8	87.1	94.4
	+ DriveTrack	80.8	89.6	93.8	**64.0**	**76.1**	88.0	**84.1**	**90.9**	**95.1**
PIPs++ [29]	PointOdyssey [29]	–	**25.8**	–	–	70.4	–	–	81.5	–
	+ DriveTrack	–	25.6	–	–	**71.2**	–	–	**85.3**	–

Quantitative performance improvements

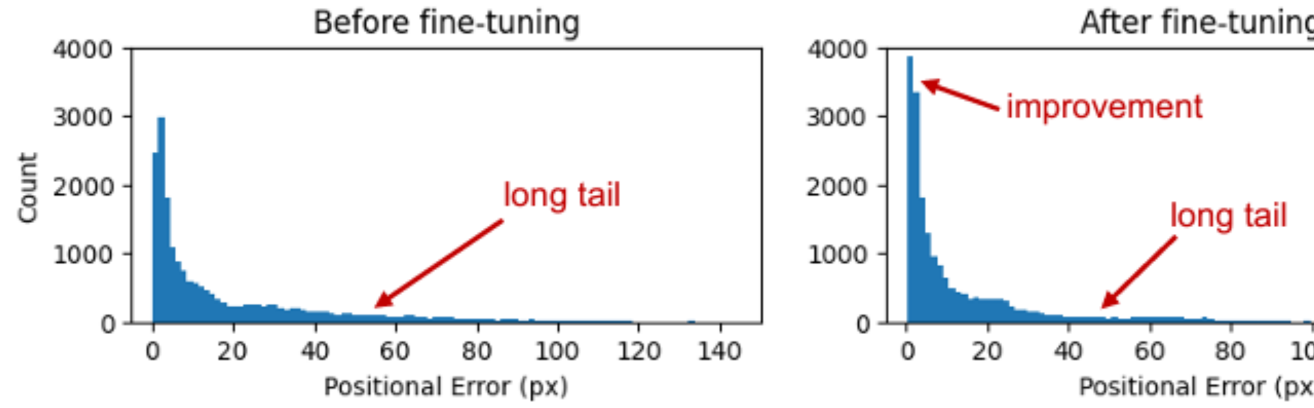


Visual performance before and after fine-tuning

Fine-tuning on DriveTrack improves tracking performance on other real-world datasets

4 Quantifying sensitivity of trackers to keypoints

- The distribution of tracking error is heavy-tailed even after fine-tuning
- High-error points correspond to visual imperfections in the scene
- These points are unlikely to see improvement even with rigorous training or fine-tuning



- To avoid high-error points, keypoint selectors are needed to choose the right query points for tracking
- Existing keypoint selectors are sparse or fail to avoid visual imperfections
- Therefore, point selectors should be developed in tandem with point trackers