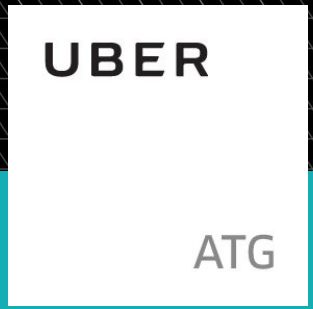
Investigating the Effect of Sensor Modalities in Multi-Sensor Detection-Prediction Models

Abhishek Mohta, Fang-Chieh Chou, Brian Becker,
Carlos Vallespi-Gonzalez, Nemanja Djuric



Introduction

- Detection and motion prediction are key components of a self-driving system
- Increased reliance on multiple sensors to achieve state-of-the-art performance
  - Increasing system complexity; model becoming more brittle
  - Higher chances of overfitting to single sensor; reduced generalization
- How do we handle cases when we have missing sensor modality?
  - Online latency or hardware issues; sensor noise
  - Gap between simulated and real sensor data

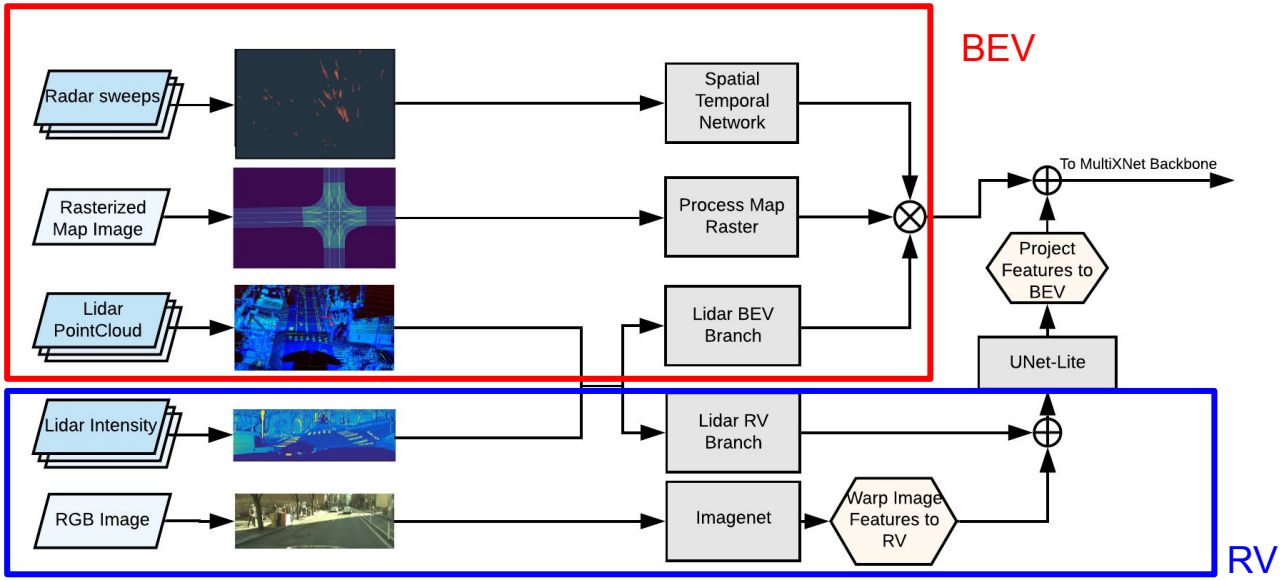
Key Contributions

In this paper:

- We analyze the contribution of each sensor modality through an ablation study
- Analyze how models perform when trained with multiple sensors, but evaluated without one
    - Modeling the missing sensor or sensor failure use case
- Propose a simple mechanism to build more robust, better performing models

Base architecture

- Modified version of the multi-view architecture [1] to include radar fusion [2]



UBER ATG

1. Fadadu, S., et al. “Multi-view fusion of sensor data for improved perception and prediction for autonomous driving”
2. Shah, M., et al. "Liranet: End-to-end trajectory prediction using spatio-temporal radar fusion"

Experimental Setup

- Dataset: Proprietary large scale data set TCO12 in dense urban environment
  - 3 million frames of samples collected at 10Hz
  - 10 LiDAR sweeps, 3 radar sweeps and current image frame to predict 30 future states
- Metrics:
  - **Detection** - Average Precision (AP) metric
    - IoU threshold of 0.7, 0.1, 0.3 for vehicles, pedestrians and bikes
    - Additional detection metrics in FOV (Camera only in FOV)
  - **Motion Prediction** - Displacement Error (DE) at 3s
    - Operating point set at recall of 0.8

Sensor Ablation Study

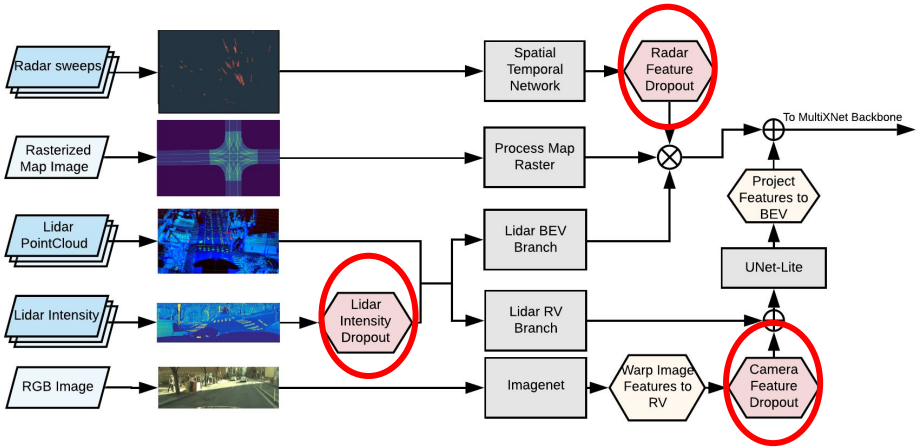
- Models trained and evaluated without particular sensor modalities
- Expected results:
  - “No camera” impacting results in FOV
  - “No radar” impacting results in DE for vehicles

Table 1: Comparison of AP (%) and DE (cm) for Sensor Ablation Study; impacted results in bold

	Vehicles			Pedestrians			Bicyclists		
	$AP_{0.7}$ ↑			$AP_{0.1}$ ↑			$AP_{0.3}$ ↑		
Method	**Full**	**FOV**	**DE↓**	**Full**	**FOV**	**DE↓**	**Full**	**FOV**	**DE↓**
Baseline	85.8	84.7	36.0	88.1	90.3	57.5	72.9	79.1	38.0
No camera	85.9	84.6	36.2	87.7	**89.0**	57.5	72.4	**74.5**	38.0
No radar	85.8	84.6	**37.3**	87.8	90.2	57.5	73.5	77.2	36.8
No intensity	85.8	84.8	36.0	87.3	89.9	58.3	**71.6**	77.4	**41.2**

Sensor Dropout

- Drop particular sensor inputs/features with some probability
  - Feature dropout for camera and radar; zero out feature vector
  - Input dropout for lidar intensity; replace with sentinel value (mean value)
- Making the model more robust
  - Better performance with missing sensor modalities



Sensor Dropout Results

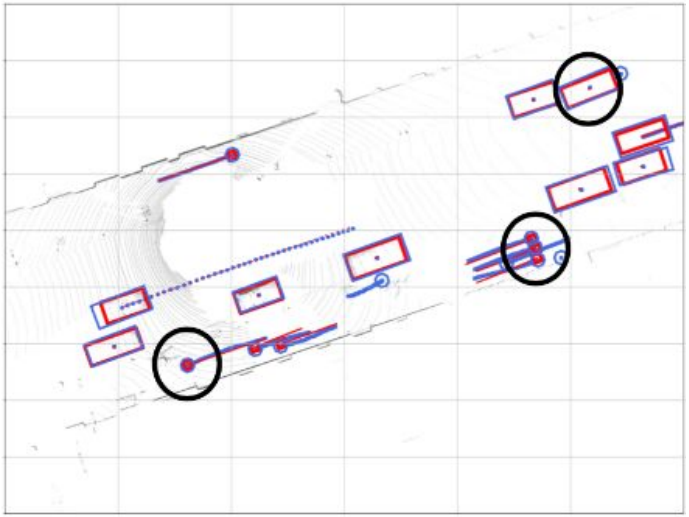
- Significantly improved performance with missing sensor modalities
- Minor impacts to model performance when all sensor modes are present

Table 2: Comparison of AP (%) and DE (cm) on TCO12 data; improved results shown in bold

		Vehicles			Pedestrians			Bicyclists		
		$AP_{0.7}$ ↑			$AP_{0.1}$ ↑			$AP_{0.3}$ ↑		
Method	**Eval mode**	**Full**	**FOV**	**DE↓**	**Full**	**FOV**	**DE↓**	**Full**	**FOV**	**DE↓**
Baseline		85.8	84.7	36.0	88.1	90.3	57.5	72.9	79.1	38.0
No camera		85.9	84.6	36.2	87.7	89.0	57.5	72.4	74.5	38.0
No radar		85.8	84.6	37.3	87.8	90.2	57.5	73.5	77.2	36.8
No intensity		85.8	84.8	36.0	87.3	89.9	58.3	71.6	77.4	41.2
Sensor Dropout		85.9	84.9	36.8	88.0	90.2	57.5	73.5	78.7	38.2
Baseline	[-Camera]	84.9	**84.1**	**36.8**	86.6	88.0	59.6	68.9	**74.6**	**39.1**
Sensor Dropout	[-Camera]	**85.6**	**84.5**	**37.2**	**87.2**	**88.6**	**58.3**	**71.2**	**74.8**	**38.8**
Baseline	[-Radar]	81.2	83.6	41.3	86.7	89.4	**57.5**	70.9	**77.9**	44.1
Sensor Dropout	[-Radar]	**85.3**	**84.7**	**37.7**	**87.8**	**90.1**	**57.3**	**73.3**	**78.4**	**39.6**
Baseline	[-Intensity]	**85.5**	**84.7**	36.2	84.9	88.7	60.9	63.7	75.1	40.2
Sensor Dropout	[-Intensity]	**85.8**	**84.7**	36.8	**87.0**	**89.6**	**58.6**	**72.2**	**77.9**	**38.4**

UBER ATG

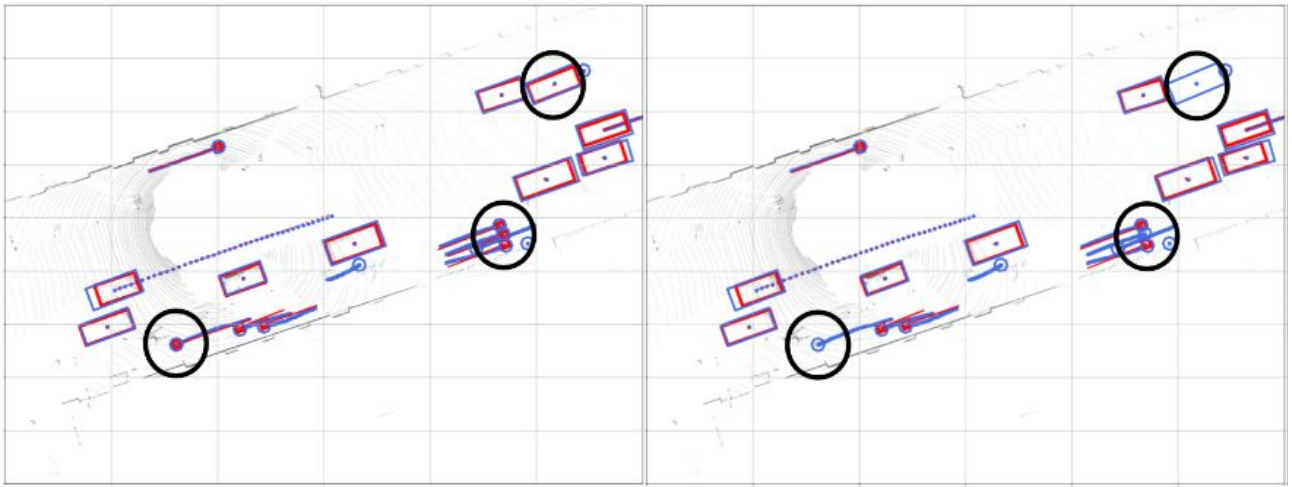
Qualitative Example



(a) Baseline

UBER ATG

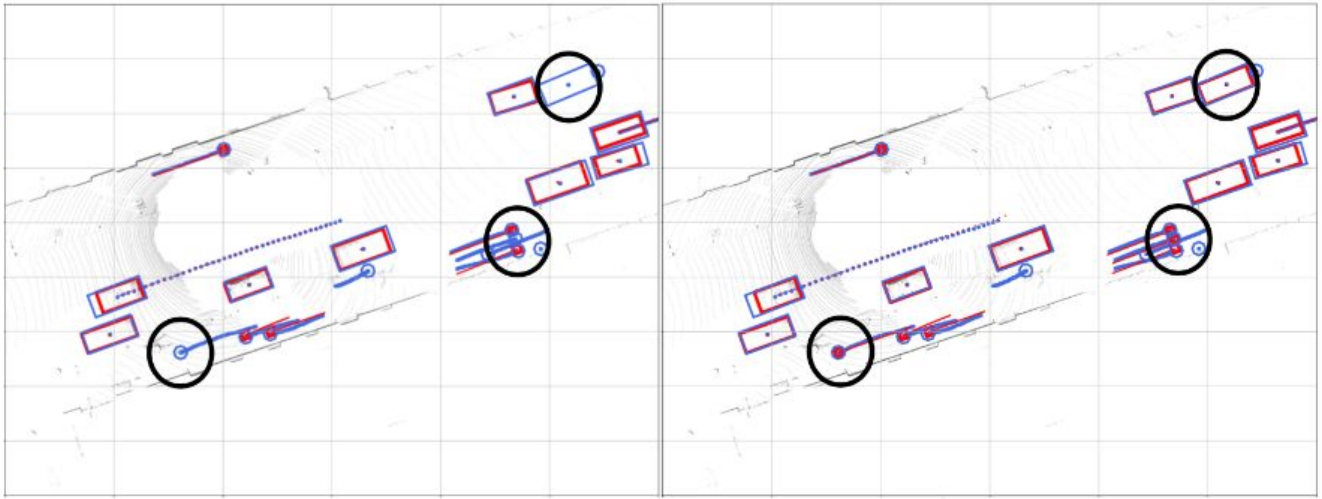
Qualitative Example



(a) Baseline

(b) Baseline [-Camera]

Qualitative Example



(b) Baseline [-Camera] (c) Sensor Dropout [-Camera]

Figure 1: Qualitative example showing improved performance with the Sensor Dropout model when the camera input is removed; ground-truth is shown in blue, model detections are shown in red.

Dropout Ablation Study

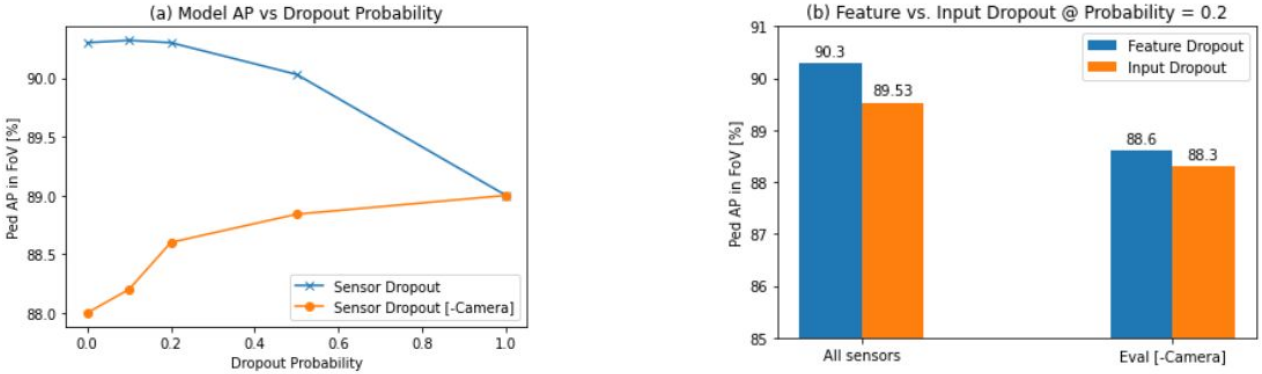


Figure 2: Effect of different settings for camera dropout: (a) sweep over dropout probability values, and (b) feature dropout vs input dropout.