Fairness in Autonomous Driving: Towards Understanding Confounding Factors in Object Detection under Challenging Weather

Bimsara Pathiraja[1], Caleb Liu[2], Ransalu Senanayake
Laboratory for Learning Evaluation of autoNomous Systems (LENS Lab), Arizona State University
1. bpathir1@asu.edu 2. calebliu@asu.edu

LENS Lab

CVPR JUNE 17-21, 2024 SEATTLE, WA

Problem: Fairness in Autonomous Driving

- Object detection is a key component of **autonomous driving**.
- We investigate the algorithmic bias and **fairness of transformer-based object detectors** [1].
- Most previous studies focus on fairness in image classification and other tasks rather than object detection [2].
- Previous research studies [3] performance in either varying weather or demographic group, not considering both together.

Main Contributions

- **Propose novel metrics** to evaluate the fairness of detection models to supplement existing mean average precision and recall (mAP, mAR).
- **Create novel datasets** derived from hi-fidelity simulations.
- **Evaluate SOTA object detection** models (DETR) under different confounding factors and with different demographic groups.

Confidence-Based Metrics

- Current metrics do not provide insights into **model confidence**
- **Average True Positive Confidence (ATPC)**
  How confident are the correct predictions
- **Average False Positive Confidence (AFPC)**
  How confident are the incorrect predictions
- **For fairness comparisons we use following disparity metrics**
  Worst-case difference Δ_{worst}s
  Best-case difference Δ_{best}s
  Wasserstein-2 metric W_s

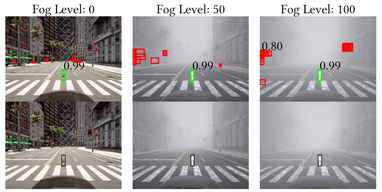
Datasets

- **FACET Dataset [4]**
  Publicly available fairness evaluation dataset containing 32,000 images.
  Perceived skin tone is annotated according to Monk Skin Tone (MST) scale.
  Lighting condition is categorized as well-lit and dimly-lit.
- **Simulate ambient darkness in FACET images**

FACET sample image across processed darkness levels 0.1, 0.4, 0.7, and 1.0 with 1.0 reflecting original darkness level and 0.0 representing total darkness. The darkness achieved through image processing techniques are intended to mimic natural lighting conditions.

- **Carla dataset**
  Carla dataset was created by authors simulating multiple weather conditions and pedestrian types using Carla simulator and an autopilot-enabled car.

- **Carla dataset images vs. DETR**



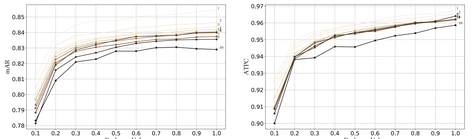
Black - ground truth
Green - true positives
Red - false positives.

The confidence score is shown only when it is > 0.5. With high levels of fog, it is possible to get false positives with confidences as high as 0.8.

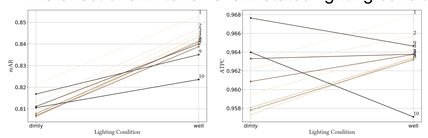
Results

- **Experiments on FACET Dataset**

The effect of skin tone vs. artificial darkness

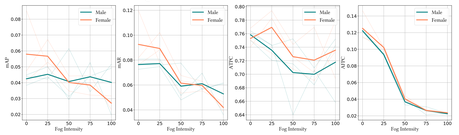


The effect of skin tone vs. annotated lighting condition

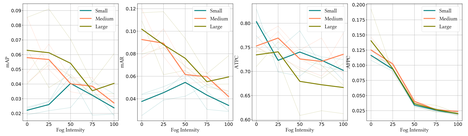


- **Experiments on Carla Dataset**

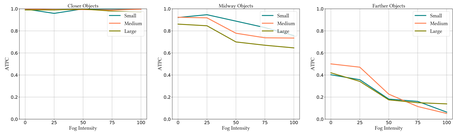
The effect of gender vs. weather conditions



The effect of gender vs. weather conditions



The effect of object distance



The effect using disparity metrics

	Gender			Body size		
Fog levels	$\Delta_{worst}mAR$	$\Delta_{best}mAR$	W_{mAR}	$\Delta_{worst}mAR$	$\Delta_{best}mAR$	W_{mAR}
0%	4.71	0.02	0.05	9.45	0.22	0.42
25%	2.73	0.16	0.02	7.85	0.28	0.21
50%	2.42	0.19	0.009	5.14	0.08	0.05
75%	1.43	0.11	0.002	3.24	0.03	0.03
100%	2.27	0.29	0.02	4.95	0.07	0.07

Conclusions

- **Weather conditions** impact demographic groups differently.
- **Confidence scores are important** to consider when evaluating detection models, as they reveal information about models that current metrics do not.
- Due to existence of many confounding factors, testing of object detection models should include both **simulation and real-word evaluation data.**

References:
[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. ECCV 2020.
[2] Sunnie SY, Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. Humans, ai, and context: Understanding end-users' trust in a real-world computer vision application. FAccT 2023
[3] Martim Brandao. Age and gender bias in pedestrian detection algorithms. arXiv 2019.
[4] Laura Gustafson, Chloe Rolland, Nikhila Ravi, Quentin Duval, Aaron Adcock, Cheng-Yang Fu, Melissa Hall, and Candace Ross. Facet: Fairness in computer vision evaluation benchmark. ICCV 2023.

Project Page