Operating cost of camera vs. LiDAR-based object detection

Aaron Xiang
Mentored by Long Quang

Introduction

Three-dimensional (3D) object detection is essential in modern sensing systems, providing spatial information for applications such as autonomous navigation and robotics. Light Detection and Ranging (LiDAR) is a remote sensing method that generates 3D data by emitting laser pulses and measuring return time to create point cloud representations. Unlike cameras, which depend on visible light and are affected by shadows or glare, LiDAR generally performs consistently across varying lighting conditions.

PointPillars, a deep learning architecture for point cloud data, converts LiDAR inputs into vertical pillars and uses a two-dimensional convolutional backbone for efficient 3D object detection (Lang et al., 2019). Prior research shows that lighting, measured in lux (the number of lumens per square meter), can impact detection reliability, emphasizing the need for robust detection systems (Dong et al., 2023).

For this investigation of 2D image detection, the You Only Look Once, version 8 (YOLOv8) served as the camera-detection baseline, and PointPillars was selected as the LiDAR based model for 3D image detection to evaluate performance under different lighting conditions.

The goal of this project was to create an integrated system to simultaneously evaluate the performance of a camera, and LiDAR based system under varying light conditions, while also accounting for power.

Methods and Materials

For this investigation, the Velodyne VLP 16 LiDAR and a standard HD1080P camera were used as the sensing system. The Nvidia Jetson Xavier NX ran PointPillars and YOLOv8. A mount was created in Autodesk Fusion to mount the LiDAR, Nvidia Jetson Xavier NX, and camera into fixed positions (Figure 1). The LiDAR generated 3D point cloud data for PointPillars, while the camera captured image frames for YOLOv8. PointPillars was set up in Robot Operating System 2 (ROS2) Foxy by building the package in a ROS2 Foxy workspace and updating it to use the correct message types and topics. A custom ROS2 node subscribed to the LiDAR point cloud and published detections, with code to count inference attempts and detected objects. YOLOv8 was integrated through a similar ROS2 Foxy node that processed camera frames and produced detection messages with total detections and counts.

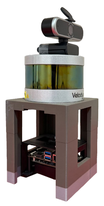


Figure 1 (right): Velodyne VLP-16, Jetson NX, camera, and mount can be seen together. The dimensions are 119.40 mm × 119.40mm × 215 mm.

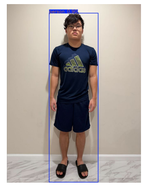
Methods and Materials (continued)

Lighting conditions were changed using the Philips WiZ Smart LED A23 E26, a brightness controlled light bulb through a mobile app. The bulb had a maximum output of 2550 lumens and five brightness levels. The levels were 10%, 25%, 50%, 75%, and 100% from the WiZ app.

Testing was performed on the system indoors, in a dark environment. YOLOv8 was first run to detect objects using a camera (Figure 2), and PointPillars was then run on point clouds from LiDAR (Figure 3).

The subject was placed at a distance of five feet from the device and moved to the right for 10 feet, approximately three seconds, maintaining a distance of five feet from the device. The lamp was positioned one foot behind and to the right of the LiDAR. Lux level, model type, and detections were recorded for each trial. After 20 trials at each lighting condition, the detection rate was calculated as the number of detections divided by the total number of inference attempts. Yearly energy use in kilowatt-hours/year (kWh/year) was calculated for the system by multiplying kilowatts with number of hours in a year. The camera system calculation included the Nvidia Jetson Xavier NX, camera, and lightbulb at 23.4% (intersection of the detection curves). The LiDAR system calculation included the Nvidia Jetson Xavier NX and LiDAR.

Figure 2 (right): YOLOv8 running on a static person can be seen with the bounding box



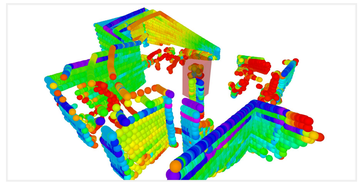
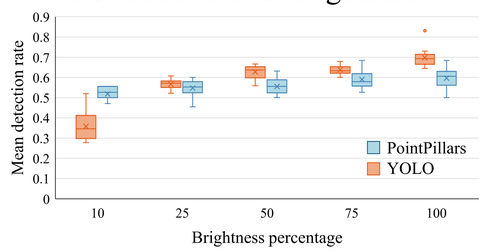


Figure 3 (left): Point cloud and bounding box visualized in rviz2.

Results

A one-way analysis of variance (ANOVA) showed that lighting level had a significant effect on YOLOv8 detection rate, $F(4, 95) = 208.78$, $p < .001$, with performance increasing from 10% brightness ($M = 0.36$) to 100% brightness ($M = 0.70$). A separate one-way ANOVA for PointPillars also indicated a significant effect of lighting level, $F(4, 95) = 13.21$, $p < .001$, although the changes across brightness levels were smaller, ranging from $M = 0.52$ at 10% to $M = 0.60$ at 100%. To directly compare the two models at a mid-range illumination level, a two-sample t-test was conducted at 25% brightness. When the systems were tested together, the difference between YOLOv8 ($M = 0.57$, $SD = 0.02$) and PointPillars ($M = 0.55$, $SD = 0.03$) was not statistically significant, $t(32) = 1.98$, $p = .056$, indicating that the two models performed similarly at this 25% lighting level (Graph 1).

Results (continued)



Graph 1 (left): PointPillars and YOLO detection rate vs. brightness graph displaying how PointPillars performs better under low light conditions, while YOLO performs better at higher brightness and surpasses PointPillars at higher brightness. Means are denoted by ×.

Assuming continuous use, the LiDAR system consumed 148.9 kWh/year, while the camera system consumed 135.0 kWh/year, meaning the LiDAR system used about 10% more energy then the camera system.

Conclusion

The findings show that both sensing approaches were affected by lighting, but in different ways. Camera-based detection dropped sharply at low brightness, while LiDAR-based detection remained more stable across lighting levels, demonstrating stronger low-light reliability despite showing variation. However, LiDAR detection rate should have stayed the same, displaying that other factors impacted testing, like reflections from surfaces and speed of walking. Additionally, less power was needed for camera system compared to LiDAR.

There were some limitations that contributed to the variation in LiDAR results. The models contain inherent uncertainty, introducing noise and reducing accuracy. The short testing duration also meant each detection had a large effect on the proportion. Future work should measure epistemic uncertainty, which is the lack of information about the model to better quantify reliability. The aleatoric uncertainty can also be modeled to see uncertainty that is inherent to the data.

References

Dong, Y., Kang, C., Zhang, J., Zhu, Z., Wang, Y., Yang, X., Su, H., Wei, X., & Zhu, J. (2023). Benchmarking robustness of 3D object detection to common corruptions in autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22236–22245. https://doi.org/10.1109/CVPR52729.2023.00105

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). PointPillars: Fast encoders for object detection from point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12697–12705. https://doi.org/10.1109/CVPR.2019.01298