**Proceedings of the Canadian Society for Mechanical Engineering International Congress**
**32nd Annual Conference of the Computational Fluid Dynamics Society of Canada**
**Canadian Society of Rheology Symposium**
**CSME-CFDSC-CSR 2025**
**May 25–28, 2025, Montréal, Québec, Canada**

# Quantifying Aleatoric Uncertainty for Mobile Robot Object Detection

William Ronan[1,2], Zhaokai Wang[1,2], Xian Wang[1,2,*]
[1]Department of Mechanical and Materials Engineering, Queen's University, Kingston, Canada
[2]Ingenuity Labs Research Institute, Queen's University, Kingston, Canada
* Corresponding author

***Abstract*—** Object detection in computer vision is inherently associated with uncertainty, which can be categorized into two primary types: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty represents the inherent randomness of real-world data and usually cannot be reduced. In the context of mobile robot perception, aleatoric uncertainty is prevalent in images with poor quality, compromised visibility, and increased visual complexity. These factors can significantly degrade object detection performance, thereby impairing robotic decision-making and functionality. Further, the mobile nature of the robots introduces increased aleatoric uncertainty due to variation in image backgrounds, motion blur, and other environmental influences. This study proposes a methodology to predict the effect of aleatoric uncertainty on object detection performance in real-time. Using a Quanser QCar experimental setup, we derived aleatoric metrics from image data – contrast, edge, brightness, velocity – to quantify uncertainty factors latent within the image. Prediction of the object detection algorithm's performance is conducted using supervised learning to train models on the relationship between our aleatoric metrics and object detection ability. The results demonstrate the ability to preemptively assess the limitations of object detection under varying environmental conditions, providing a predictive framework for improving the reliability and safety of mobile robotic systems.

***Keywords-component: object detection, mobile robotics, aleatoric uncertainty***

## I. INTRODUCTION

Object detection is a cornerstone of computer vision, enabling systems to identify and localize objects within a scene while predicting properties such as their location, size, and orientation. Unlike traditional image classification tasks, object detection involves both spatial and semantic understanding, making it critical for applications across domains such as surveillance [1], medical imaging [2], and inventory management [3]. Advancements in deep neural networks (DNNs), particularly convolutional neural networks (CNNs), have revolutionized object detection, improving accuracy and computational efficiency [4]. These breakthroughs have facilitated its adoption in mobile robotics, where object detection is pivotal for tasks such as sign recognition, obstacle avoidance, and decision-making [5]. Autonomous systems, including self-driving vehicles, unmanned aerial vehicles, and automated machinery, rely on object detection to perceive and interact with their environments effectively.

Despite these advancements, object detection is inherently accompanied with uncertainty. This uncertainty is often separated into two main subgroups: aleatoric uncertainty and epistemic uncertainty [6]. Epistemic uncertainty refers to uncertainty caused by the imperfection of the prediction model itself. This type of uncertainty is reducible by improving the model architecture, the learning process, and the training data set [7]. Aleatoric uncertainty is intrinsic within real-world data and cannot be reduced [8]. Typically, the output of an object detection CNN is a numerical measurement of prediction strength which is bounded by a function such as the softmax or sigmoid [9]. This provides some measurement entangling both types of uncertainty, however, CNNs tend to be overconfident in their prediction strength [10]. Studies of uncertainty in object detection have led to better quantification, for example, Bayesian neural networks (BNNs) which use probability distributions rather than fixed values for weights, are adept at capturing epistemic uncertainty in model parameters [11]. Approximation techniques like Monte Carlo Dropout are used in conjunction with BNNs to quantify uncertainty in a model [12]. While BNN's and the use of Monte Carlo methods are good metrics for understanding epistemic uncertainty within a model, they provide little understanding on the aleatoric uncertainty of a prediction.

Aleatoric uncertainty intrinsic within real-world data is always present [13], and is worsened by reduced visibility, busy scenes, and camera motion effects. Studies exploring the effect of naturally occurring phenomena on object detection show a decrease in average precision when environmental influences such as rain, snow and fog are introduced within images [14]. Darkness was also shown to be a detriment to a model's detection ability [15]. Moreover, combining these scenarios led to reduction in precision due to uncertainty accumulating from multiple environmental influences [16]. Regarding the content of the scene, inclusion of different scene level objects can have varying effects on detection ability [17]. This can be due to added image complexity, as well as a model's use of contextual cues in prediction. For mobile systems, camera velocity was shown to deteriorate the precision of object detection models due to motion blur [18]. Specifically, rotational camera motion was shown to cause an increased distortion effect [19]. Many of the aforementioned factors are realistic in use cases for mobile robotics, which highlights limitations of mobile robot perception and raises concerns of safety of the intended functionality.

Current approaches to quantifying aleatoric uncertainty typically occur post-detection, relying on the detected output to infer uncertainty [20, 21]. However, the ability to predict aleatoric uncertainty preemptively, based solely on input image data, remains underexplored. Some studies have investigated CNNs' responses to image complexity, demonstrating that features such as human-perceived complexity or localized difficulty can be correlated with detection performance [22, 23]. These findings suggest that complexity metrics derived directly from images may serve as proxies for quantifying aleatoric uncertainty. Moreover, in real-time object detection, uncertainty can arise from temporal dynamics. Rapidly changing scenes and varying camera velocities further complicates the problem [24].

This study addresses the challenge of preemptively quantifying aleatoric uncertainty in mobile robot object detection. We propose a framework to derive aleatoric metrics directly from input images, encompassing factors such as contrast, brightness, edge detection length, and robot velocity. Using a controlled experimental setup with a Quanser QCar, we collected data under varying environmental conditions to train supervised models that predict the impact of aleatoric uncertainty on object detection performance. By enabling real-time prediction of uncertainty effects prior to detection, this approach aims to improve decision-making processes in mobile robotics and enhance the overall safety of the intended functionality.

## II. Methodology

Given that aleatoric uncertainty is intrinsic within data, it was hypothesized that within the context of a specific task, an image alone is enough to garner some understanding of the uncertainty it introduces. Moreover, the effect of an image's aleatoric uncertainty on the object detection algorithm can be predicted by quantifying aleatoric uncertainty metrics from image-data. We begin by outlining an experiment used to collect images from a mobile robot in a common scenario while varying scenic influences. Next, we describe the creation of an annotated dataset from the data collected in this experiment. Input features are derived from image-data, while labels are a measurement of detection ability. Finally, we introduce models which use the extracted image metrics to predict the performance of object detection on a given image.

### A. Hardware

The mobile robot used in the study was a Quanser QCar (Quanser Consulting Inc., Ontario, Canada) shown in Figure 1. The QCar is a prebuilt mechatronic car fashioned with a NVIDIA Jetson Nano embedded computer board, an Intel RealSense depth camera, as well as a range of other sensors [25]. The QCar has wheel encoders which were used for accurate distance and velocity measurements.

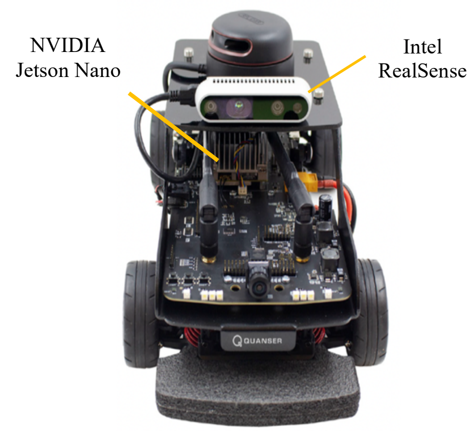


Figure 1: Quanser QCar adapted from [25] with permission

### B. Software

A pretrained SSD-Mobilenet architecture was the basis for object detection used in the experiment [26]. While the pretrained model was capable of detecting stop signs, it was further tuned to images of the specific stop sign used in the experiment. These tuning images were taken by the QCar at varied velocities, brightness levels, and background complexities to avoid bias towards any particular scene. An ROS2 system architecture was used to control the vehicle, while capturing and processing images in parallel. The system was programmed using python and the scikit-learn package was the basis for image processing.

### C. Standardized Experiement

To study the effect of external influences on image aleatoric uncertainty, a common setting was used in experimentation. A 2-meter-long track was built with a stop sign placed on the end of it as shown in Figure 2. With each trial, the QCar travelled the length of the track collecting image data as shown in Figure 3.

The images were subject to two different processing pipelines. First, image-data would be mined for aleatoric uncertainty metrics. Then, the image would be processed by the object detection algorithm which would detect the stop sign. Object detection ability varied with proximity to the stop sign,

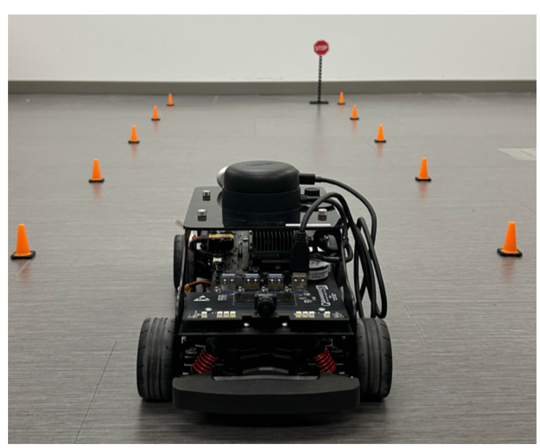

Figure 2: Standardized track

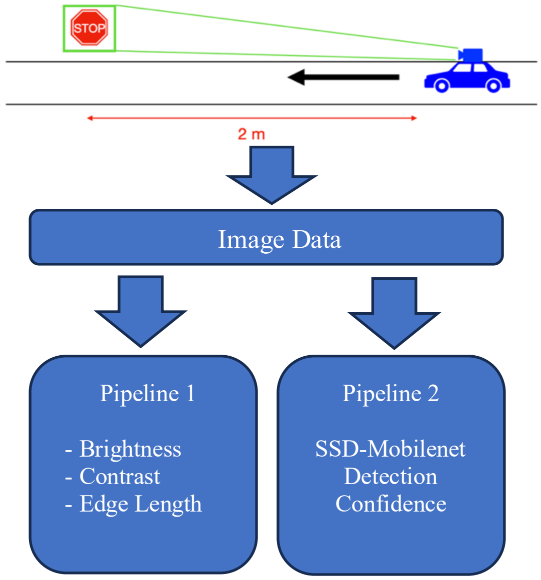


Figure 3: Experimental setup

thus, it was necessary to average the collected image-data over 10 cm intervals. This allowed image-data from different trials to be compared on a by-distance-interval basis. This experiment was conducted while systematically varying environmental influences. Various props were added to the background to change the aleatoric uncertainty of the images. As well, the brightness of the room, and the velocity of the robot were varied, also changing the aleatoric uncertainty of the images. Figure 4 shows a particular scene with increased background complexity, and reduced brightness. The data collected from experiment was used to compare the aleatoric metrics to the performance of the object detection algorithm for each given image.

### *D. Metrics*

Four different metrics were created to quantify the aleatoric uncertainty of a given image: Contrast, edge detection length, brightness, and the velocity of the mobile robot when the image was captured; $C$, $L$, $B$ and $V$ respectively. These metrics were chosen based on their representation of mutually exclusive indicators of aleatoric uncertainty to decrease parameter overlap. Figure 5 shows six different images, displaying examples low and high C, L and B respectively.

Contrast was used to gauge variation in colour data. This image contrast metric was taken as the combined Root Mean Square (RMS) contrast of the RGB channels shown as in (1). $N$ refers to the total number of pixels in a channel and $i$ refers to the $i$th pixel in each channel.

$$\mathrm{C} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(R_i-\bar{R})^2+(G_i-\bar{G})^2+(B_i-\bar{B})^2\right]} \tag{1}$$

Brightness was used to gauge the luminosity of the scene. It was calculated by converting the image to grayscale and calculating the mean of the pixel intensities (2). $Ii$ refers to the intensity of pixel $i$ in the grayscale image.

Figure 4: Complex scene with reduced brightness

$$\mathrm{B} = \frac{1}{N}\sum_{i=1}^{N} I_i \tag{2}$$

Edge detection length was used to quantify the amount of fine-grained detail in an image. The edge length returned by using a Sobel filter (or, $S_{length}$) was used for edge detection length (3). The Sobel filter catches sharp changes in gradient commonly associated with feature detection.

$$\mathrm{L} = S_{\mathrm{length}} \tag{3}$$

The velocity of the QCar at the time of the image was used to quantify both temporal aleatoric effects and potential motion blur. Velocity was calculated during the experiment using the QCar's wheel encoders (4). $D$ represents wheel diameter, $P$ the number of pulses per revolution and $\Delta t$ the time interval.

$$\mathrm{V} = \frac{\mathrm{Pulses}\times \pi D}{P\times \Delta t} \tag{4}$$

The effect of the aleatoric uncertainty was quantified to provide labels for training the predictive model. The SSD-MobileNet's outputted detection confidence was the basis for quantifying detection ability. To measure the effect of aleatoric uncertainty, first, baseline detection confidence was found by performing the experiment in the best possible conditions. This control study was conducted in the scene shown in Figure 2. The control study used a blank background, the brightest light setting, and slowest possible velocity of the QCar. The detection confidence for images in the control study were averaged to find the baseline confidence for each distance interval. After quantifying the baseline detection confidence, $c_{baseline}$, Loss of Detection Confidence (LDC) could be calculated as the difference between detection confidence for a new image, and the baseline confidence at the same distance. Shown in Equation (5), $LDC$ represents the Loss of Detection Confidence while $c_{baseline}$ and $c$ represent the baseline detection confidence and the detection confidence using the new image respectively.

$$LDC = c_{baseline} - c \tag{5}$$

### *E. Models*

The data collected during experimentation was used to create a training set for supervised regression models. The aleatoric metrics calculated from image-data were used as input features

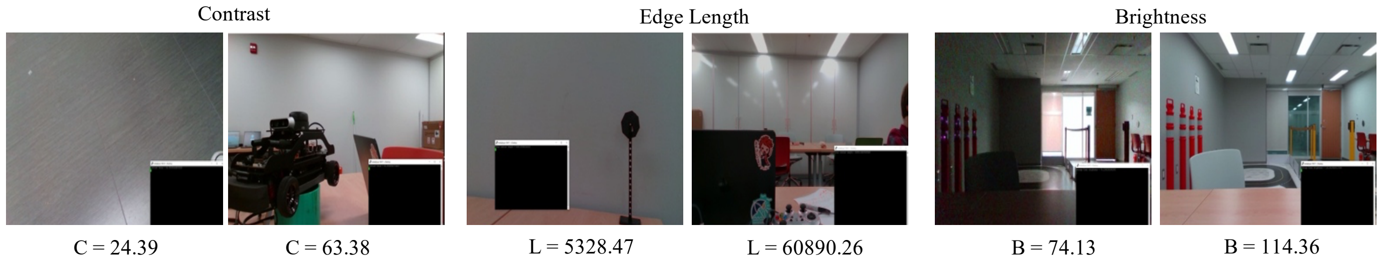

Figure 5: Visualization of contrast, edge length, and brightness metrics prior to standardization

while LDC was used for labels. Each data instance corresponded to a particular distance interval of a particular trial. Prior to the data being used, an analysis was performed to remove outliers. For a given trial of the experiment, the z-score was determined for all distance intervals. The average z-score for LDC was taken for each distance interval across all trials, as shown in Equation (6). $\bar{z}_d$ represents the average z-score of a given interval. This was calculated with $N_d$ representing the number of trials with data for the given distance interval $d$, and $z_{i,d}$ representing the z-score at distance interval $d$ in trial $i$. This analysis revealed that a few distance intervals corresponded with z-score spikes. This was likely due to increased glare at certain distances. Distance intervals with an average z-score of the LDC greater than 0.3 were removed. The resulting dataset was comprised of 864 instances each corresponding to a particular distance interval of a particular trial. Input data was standardized to ensure all aleatoric metrics were considered equally in training. Both a random forest regression model and a forward feeding neural network (FNN) were chosen for training based on the complex relationships within the data.

$$\bar{z}_d = \frac{1}{N_d}\sum_{i=1}^{N_d} z_{i,d} \qquad (6)$$

## III. Results

### A. Loss Predicting Models

Supervised models were given the aleatoric metrics calculated using equations (1), (2), (3) and (4) as input to predict the LDC calculated using equation (5). The dataset was comprised of a total of 864 instances.

The evaluation results in Table I. compare prediction models in terms of average root mean squared error (RMSE). Each model was retrained 200 times using bootstrapping to create randomly sampled subsets of the data with replacement. For each iteration, the data was split into 80% training set and 20% testing set. The table includes the results for the random forest model and the FNN, as well as two additional models for comparison. The Always-Guess-0 model shows the RMSE when LDC was predicted to be 0 for all instances. The Mean Model shows the RMSE if the mean LDC of the dataset was predicted for all instances.

**TABLE I**. PREDICTIVE MODEL PERFORMANCE (N = 200)

| | **Always-Guess-0** | **Mean Model** | **FNN** | **Random Forrest** |
|---|---|---|---|---|
| Root Mean Squared Error | 0.0795 | 0.0540 | 0.0461 | 0.0449 |

The results show that the random forest performed better than the FNN on the dataset ($p < 0.0001$). However, both the FNN and the random forest outperformed the Mean Model ($p < 0.0001$, $p < 0.0001$), indicating that they both were equipped to predict the loss of detection confidence. Moreover, the random forest and FNN achieved an RMSE of 0.0449 and 0.0461 respectively. Both models achieved an RMSE lower than the standard deviation of the data, 0.053 ($p < 0.0001$, $p < 0.0001$), suggesting a causal link between the latent features of the image-data, and object detection ability. However, the margin between the models' RMSE and the standard deviation of the data is low, suggestion room for improvement.

### B. Predictive Contributions of Aleatoric Metrics

The random forest regression model and the FNN were also used to calculate the contribution of each aleatoric metric. Random forest models inherently provide a measure of feature importance based on how much each feature contributes to refining the model. For the FNN, permutation testing was used to determine feature importance. This process involved varying the values of each feature independently and measuring the impact on the model's importance. Figure 6a and Figure 6b show the contribution of the aleatoric metrics towards predicting LDC for the random forest and FNN respectively.

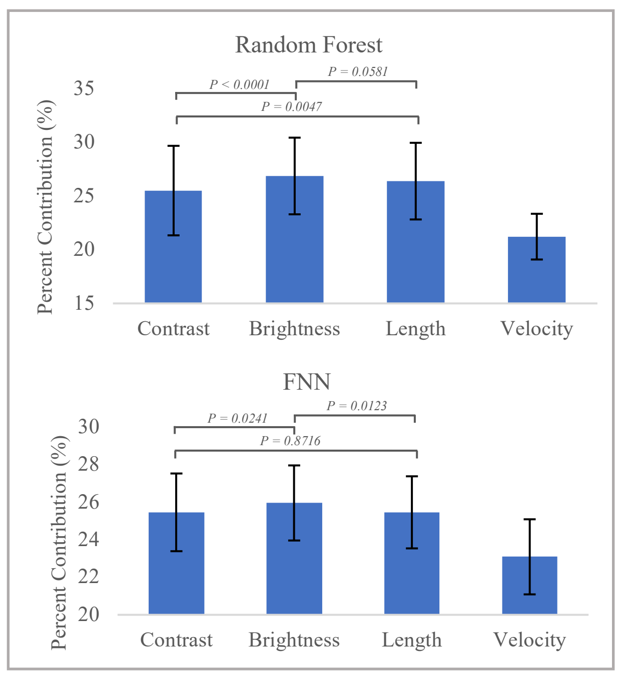

Figure 6: (a) Contribution of metrics in random forest (N = 200), (b) contribution of metrics in FNN (N =200)

For the random forest, brightness and edge length were too close in importance to discern which was the best metric. However, both brightness and edge length outperformed the next most important metric, contrast. Finally, while velocity still contributed an average importance of 21.22%, it was less important than the three other metrics.

For the FNN, brightness was most important, outperforming contrast and edge length. The contributions of contrast and edge length were too close to statistically discern which was the next best metric. Similar to the random forest, the velocity contributed an importance of 23.09%, however, this was less important than the three other metrics.

## IV. DISCUSSION

Our results found that loss of detection confidence can be estimated from aleatoric metrics collected prior to a detection being made. Compared with methods found in literatures [20, 21], this provides the advantage of a predeterminant prediction of the effects of aleatoric uncertainty. Given the low computation complexity of deriving the aleatoric metrics and predicting loss of predation strength, supervised models of this type could be used for real-time monitoring of the limitations of an object detection algorithm in a mobile robot.

In this study, the random forest classifer outperformed the FNN. While the difference in perfomrance was proven, both methods were able to provide useful predictions of LDC given the aleaotric metrics. The dataset used in this experiment was relatively small with 864 instances, and only included 4 parameters. For larger datasets with a greater number of aleaotric metrics, FNNs may provided an increased advantage due to their ability to genrealize better with large amounts of data and perform internal feature extraction.

While predictability for loss of detection confidence was shown, the prediction accuracy was modest. One area for improvement for the predictive aleatoric uncertainty model is the selection of aleatoric metrics. First, while the four metrics used were sufficient to discern a correlation with detection confidence, this study was limited in the exploration of other possible metrics. Second, different combinations of metrics in prediction models may yield greater precisions. Further exploration into component analysis might yield useful insights into the relationships of aleatoric metrics. Third, experimenting with new metrics that involve converting image-data into the frequency domain could add greater dimensionality to a prediction model of this type. Finally, studying the effect of angular velocity would provide a better understanding for robotic systems with multiple degrees of motion.

## REFERENCES

1. P. Viola and M. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004, doi: 10.1023/B:VISI.0000013087.49260.fb
2. A. Frangi, W. Niessen, R. Hoogeveen, T. van Walsum, and M. Viergever, "Model-based quantitation of 3-D magnetic resonance angiographic images," IEEE Transactions on Medical Imaging, vol. 18, no. 10, pp. 946–956, 1999, doi: 10.1109/42.811279.
3. N. K. Verma, T. Sharma, S. D. Rajurkar and A. Salour, "Object identification for inventory management using convolutional neural network," 2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2016, pp. 1-6, doi: 10.1109/AIPR.2016.8010578.
4. Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 10, pp. 2242–2262, 2020.
5. S. Syntakas, K. Vlachos and A. Likas, "Object Detection and Navigation of a Mobiles Robot by Fusing Laser and Camera Information," 2022 30th Mediterranean Conference on Control and Automation (MED), Vouliagmeni, Greece, 2022, pp. 557-563.
6. A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," Advances in Neural Information Processing Systems, vol. 30, 2017.
7. X. Zhou, H. Liu, F. Pourpanah, T. Zeng, and X. Wang, "A survey on epistemic (model) uncertainty in supervised learning: Recent advances and applications," Neurocomputing, vol. 489, pp. 449–465, 2022, doi: 10.1016/j.neucom.2021.10.119.
8. M. Valdenegro-Toro, "I Find Your Lack of Uncertainty in Computer Vision Disturbing," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 1263-1272.
9. J. Gawlikowski et al., "A Survey of Uncertainty in Deep Neural Networks," Artificial Intelligence, vol. 56, pp. 1513-1589, 2023.
10. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in International Conference on Machine Learning, PMLR, 2017, pp. 1321–1330.
11. A. Harakeh, M. Smart, and S. L. Waslander, "BayesOD: A Bayesian approach for uncertainty estimation in deep object detectors," in 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 87-93.
12. K. Deepshikha, S. H. Yelleni, P. K. Srijith, and C. K. Mohan, "Monte Carlo DropBlock for modelling uncertainty in object detection," arXiv preprint arXiv:2108.03614, 2021.
13. A. Der Kiureghian and O. Ditlevsen, "Aleatory or Epistemic? Does It Matter?" Structural Safety, vol. 31, no. 2, pp. 105–112, 2009, doi: 10.1016/j.strusafe.2008.06.020.
14. L. Peng, H. Wang, and J. Li, "Uncertainty Evaluation of Object Detection Algorithms for Autonomous Vehicles," Automotive Innovation, vol. 4, no. 3, pp. 241–252, 2021, doi: 10.1007/s42154-021-00154-0.
15. L. Ren, H. Yin, W. Ge, and Q. Meng, "Environment Influences on Uncertainty of Object Detection for Automated Driving Systems," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1–5, doi: 10.1109/CISP-BMEI48845.2019.8965948.
16. N. Bunzel, M. Geibler and G. Klause, "Measuring the Effects of Environmental Influences on Object Detection," in 2024 54th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), Brisbane, Australia, 2024, pp. 29-31.
17. L. Vonder Haar, T. Elvira, L. Newcomb, and O. Ochoa, "Measuring the Impact of Scene Level Objects on Object Detection: Towards Quantitative Explanations of Detection Decisions," arXiv preprint arXiv:2401.10790, 2024.
18. X. Yang, F. Sang, T. Wang, X. Pei, H. Wang and T. Hou, "Research on the Influence of Camera Velocity on Image Blur and a Method to Improve Object Detection Precision," 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI), Beijing, China, 2021, pp. 1-6, doi: 10.1109/ICCSI53130.2021.9736224.
19. X. Weng, S. Wu, F. Beainy and K. M. Kitani, "Rotational Rectification Network: Enabling Pedestrian Detection for Mobile Vision," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 2018, pp. 1084-1092, doi: 10.1109/WACV.2018.00124.
20. P. Cui, G. He, D. Zhang, Z. Deng, Y. Dong, and J. Zhu, "Exploring Aleatoric Uncertainty in Object Detection via Vision Foundation Models," doi: 10.48550/arXiv.2411.17767.
21. J. Liu, J. Zhang, and N. Barnes, "Modeling Aleatoric Uncertainty for Camouflaged Object Detection," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022.
22. R. T. Ionescu, B. Alexe, M. Leordeanu, M. Popescu, D. P. Papadopoulos and V. Ferrari, "How Hard Can It Be? Estimating the Difficulty of Visual

Search in an Image," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2157-2166, doi: 10.1109/CVPR.2016.237.

23. E. Saraee, M. Jalal, and M. Betke, "Visual complexity analysis using deep intermediate-layer features," Computer Vision and Image Understanding, vol. 195, p. 102949, 2020, doi: 10.1016/j.cviu.2020.102949.
24. T. Liu, C. Wang, Z. Yin, Z. Mi, X. Xiong, and B. Guo, "Complexity Quantification of Driving Scenarios with Dynamic Evolution Characteristics," Entropy, vol. 26, no. 12, p. 1033, 2024, doi: 10.3390/e26121033.
25. Quanser, "QCar: The Quanser Self-Driving Car," [Online]. Available: www.quanser.com/products/qcar/. [Accessed: 29-Dec-2024].
26. "Dusty-Nv, 'Jetson Inference.' [Online]. Available: https://github.com/dusty-nv/jetson-inference. [Accessed: 29-Dec-2024]."