Driving confidence: An uncertainty-aware framework for optimal sensor deployment in autonomous vehicles

Dawid Migacz[1,2] [0000−0001−8607−3443]
Maciej Zięba[2] [0000−0003−4217−7712]
Bartosz Zieliński[1,3] [0000−0002−3063−3621]
Kamil Adamczewski[1] [0000−0002−2917−4392]

[1]*IDEAS NCBR*
`name.surname@ideas-ncbr.pl`
[2]*Wrocław University of Science and Technology*
`name.surname@pwr.edu.pl`
[3] *Jagiellonian University in Kraków*
`name.surname@uj.edu.pl`

Abstract. *Safety and confidence of autonomous systems guided by a range of sensors remains the crucial aspect of their development. We present an uncertainty-aware decision-making framework for optimising sensor usage in autonomous systems, specifically focusing on the choice between LIDAR and RGB cameras. We show that uncertainty-based training improves the object detection task. Moreover, we present uncertainty as a guiding metric to select the most suitable modality for challenging samples. Experiments demonstrate the effectiveness of the proposed framework in enhancing energy efficiency in computationally-contrasting environments.*
Keywords: *decision-making framework, sensor optimisation, modality-switching strategy, uncertainty estimation, energy efficiency, convolutional neural networks, Monte Carlo dropout*

1. Introduction

LIDAR (Light Detection and Ranging) and RGB cameras are two types of sensors commonly used in various applications, including autonomous driving, robotics, and environmental monitoring. Each technology, with its unique strengths and weaknesses, is tailored for specific applications, particularly in the context of autonomous driving systems.

The RGB cameras, which capture colours and textures, play an important role in identifying traffic signals, road signs, lane markers, and other vehicles. These high-resolution visual data are essential for object recognition and contextual understanding. However, LIDAR excels at measuring distances and operating effectively in low light conditions or in challenging weather, making it indispensable for certain scenarios. The most sophisticated autonomous vehicles employ a combination of sensors, including both LIDAR and RGB cameras. This multi-sensor approach capitalises on the advantages of each technology, enhancing perception accuracy and improving safety.

Given the computational limitations of autonomous systems, there might be situations where prioritising one sensor over another is necessary to minimise computational demands. In such cases, the decision on whether to prioritise LIDAR or RGB cameras hinges on specific requirements.

We explore an uncertainty-based scenario to guide this decision-making process. We consider the importance of uncertainty, a critical factor in high-stakes applications like autonomous driving. Lower uncertainty means more reliable decisions can be made, emphasising the need for precise and reliable sensor data. The decisions are made through a routing network capable of predicting sensor uncertainty.

To sum up, firstly, we notice the importance of uncertainty-aware training and incorporate Dropblock thus improving the baseline object detection accuracy. Secondly, we build router-based uncertainty-aware framework that assesses the confidence of autonomous driving systems and, given computational constraints, allows to select a sensor with the highest confidence. The experimental results show the ability to select appropriate scenes that enhance model performance and optimise its computational resources.

2. Preliminaries

2.1. Object detection

We base our work on CenterNet [1], which is an object detection network that describes the detecting objects as a triplet of key points (two corners and a central keypoint). This approach eliminates the necessity for anchors and the resourceintensive Non-Maximum Suppression (NMS). This change allows the network to output a regressed map of possible central points, avoiding the discontinuity inherent in the traditional approach.

2.2. Uncertainty

It is known that logit values do not provide a reliable measure of uncertainty [2]. Various methods have been proposed to estimate uncertainty, including deep ensembling [3] and Monte Carlo dropout [4]. Deep ensembling, requiring multiple runs of randomly initialised full models, is computationally demanding. Hence, we opt for uncertainty estimation using Monte Carlo Dropout, with the caveat that we employ DropBlock [5] due to its high performance in convolutional neural networks.

3. Method

We train two identical networks separately: one using LIDAR flattened images as input and the other using RGB images. Both networks incorporate DropBlock, which remains enabled during inference, rendering the networks nondeterministic. Subsequently, we compare the inferred central point maps, calculating the variance pixel by pixel. The obtained variances are then averaged to determine the uncertainty of prediction for each modality.

Let O_i^m denote the sequence of outputs of the network with modality $m \in \{\mathrm{RGB}, \mathrm{LIDAR}\}$ for image number i. Let also every individual output $O_i^m(n)$ be a set of heads (corresponding to centre-point coordinates, heading of detected object, dimensions of detected bounding boxes, etc.) and every head $O_i^m(n)(h)$ be a set of channels (central points probabilities of different classes, individual dimensions of bounding boxes, etc.). Let every channel be a set of pixels (of the same dimensions as the original input). We then denote the output at an individual pixel of a channel by $O_i^m(n)(h)(c)(x, y)$. The variance at this pixel is then defined as:

$$\mathcal{V}^m_{i,h,c,\langle x,y\rangle} = \mathrm{Var}_{n\in|O_i^m|}\left(O_i^m(n)(h)(c)(x,y)\right)$$

and the uncertainty of image i with modality m is defined as:

$$\mathcal{U}_i^m = \frac{1}{\left|O_i^m(0)\right|}\sum_{h\in O_i^m(0)}\frac{1}{\left|O_i^m(0)(h)\right|}\sum_{c\in O_i^m(0)(h)}\frac{1}{\left|O_i^m(n)(h)(c)\right|}\sum_{\langle x,y\rangle\in O_i^m(0)(h)(c)}\mathcal{V}^m_{i,h,c,\langle x,y\rangle}.$$

3.1. Modality switching

Our aim is to optimise the energy usage of an edge device (such as a drone or autonomous vehicle). We assume that the RGB camera is always enabled as it is necessary for the operator controls. We selectively activate the LIDAR sensor on the basis of necessity, specifically to handle the most challenging samples. Our strategy involves identifying samples where the difference $\mathcal{U}_i^{\text{LIDAR}} - \mathcal{U}_i^{\text{RGB}}$ is maximised and processing them using the LIDAR network. However, since this difference is unavailable at inference time (calculating it directly would incur prohibitive costs outweighing any potential savings), we estimate it using a shallow convolutional neural network. This network takes an RGB image as input.

4. Experiments

We conduct experiments using the KITTI dataset [6] and a single NVIDIA A100 GPU. The training process involves LIDAR and RGB networks, each trained for 300 epochs with DLA34 as the backbone [7]. Additionally, the router network, featuring two hidden layers, undergoes training for 40 epochs.

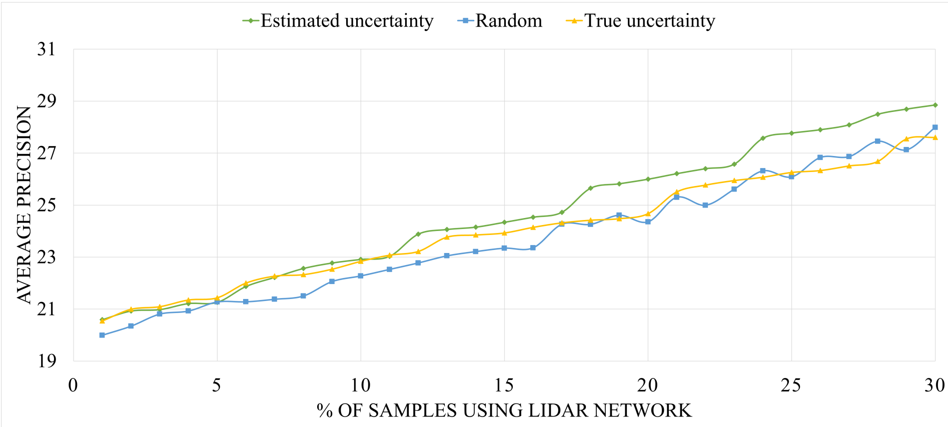
4.1. Dropblock-based object detection

First, we show that incorporating Dropblock-based training improves the baseline Centernet network. We report average precision (AP) at IoU=0.7 where scale = small, medium, large, denotes the scale of object. We use Dropblock=0.15 which provided the best results.

	small	medium	hard
LIDAR Pretrained	58.3248	42.8515	37.1175
LIDAR Dropblock	61.0089	46.4143	40.4691

4.2. Modality switching

We calculate the $\text{AP}|_{40}$ of 3D bounding box predictions at IOU = 0.7, referring to the LIDAR network for a specified percentage of samples (checking every percentage from $\{1, \ldots, 30\}$). Decision-making approaches are compared on the basis of random selection, calculated uncertainty, and estimated uncertainty. For clarity we present results for large objects, however they are similar for the other classes.

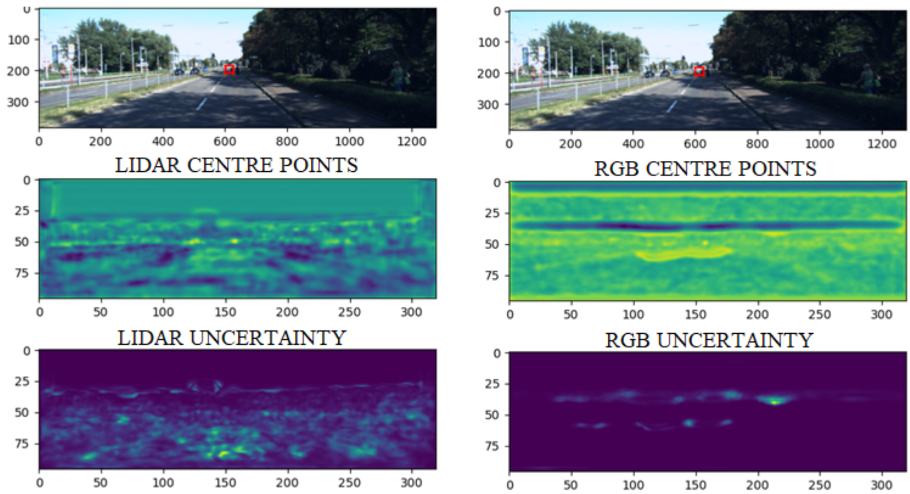


To enable comparing various methods, we propose using the area under the curve (AUC) as a metric of success.

method	random	true uncertainty	estimated uncertainty
AUC	712.12	721.50	**740.01**

4.3. Uncertainty visualisation

The graphic illustrates predictions and uncertainties from both RGB and LIDAR networks. The top-left and top-right corners show RGB and LIDAR images with respective bounding-box predictions. In the middle, probability maps indicate the likelihood of specific points being the centre of a car's bounding box for both networks. The bottom part displays pixel-wise uncertainty maps for RGB and LIDAR predictions, offering a visual understanding of the models' confidence across the images.



5. Conclusions

It can be observed that uncertainty enables the device to decide on using a second modality. Furthermore, using a shallow network introduces a regularising effect, effectively denoising the true uncertainty. Importantly, the low computational cost of the router network incurs minimal energetic overhead during inference, making it a practical means of achieving savings in this regard.

Acknowledgement

We gratefully acknowledge the Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2023/016814.

References

[1] Zhou, X., Wang, D., and Krähenbühl, P. Objects as points, 2019.

[2] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

[3] Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6405–6416. Curran Associates Inc., Red Hook, NY, USA, 2017. ISBN 9781510860964.

[4] Gal, Y. and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059. PMLR, New York, New York, USA, 2016. URL `https://proceedings.mlr.press/v48/gal16.html`.

[5] Ghiasi, G., Lin, T.-Y., and Le, Q. V. Dropblock: a regularization method for convolutional networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 10750–10760. Curran Associates Inc., Red Hook, NY, USA, 2018.

[6] Geiger, A., Lenz, P., and Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012.

[7] Yu, F., Wang, D., and Darrell, T. Deep layer aggregation. *CoRR*, abs/1707.06484, 2017. URL `http://arxiv.org/abs/1707.06484`.