

Reducing Ghost Detections Through Uncertainty Modeling for Automated Driving

Ahmed Hammam^{*‡} Frank Bonarens^{*},
Seyed Eghbal Ghobadi[†] und Christoph Stiller[‡]

Abstract: Deep neural networks (DNN) have demonstrated remarkable performance in various tasks related to automated driving. However, one significant obstacle hindering their application in automated driving systems is the occurrence of false positive detections. In our context, false positive detections are referred to as ghost detections, wherein the DNN mistakenly identifies parts of the scene as objects. In this work, we explore the prospect of leveraging uncertainty modeling to effectively minimize ghost detections. We propose a method that builds on an instance segmentation framework that better separates true positive from false positive distributions than state-of-the-art methods. This method integrates the Intermediate Layer Variational Inference (ILVI) approach and Dirichlet Distributions into an instance segmentation network. Our experimental results demonstrate that our proposed method not only enhances instance and semantic segmentation performance but also improves uncertainty estimation. Leveraging significantly improved uncertainty estimation, we investigate the potential of thresholding on uncertainty to reduce the occurrence of ghost detections, thereby enhancing both precision and recall performance.

Keywords: Deep Neural Networks, Uncertainty Estimation, Instance Segmentation

1 Introduction

Deep learning has revolutionized computer vision, offering groundbreaking advancements in various domains including medical imaging [1] and automated driving (AD) [2]. In the field of AD systems, deep neural networks (DNNs) have emerged as the predominant method, finding widespread applications in sensor fusion [3], path planning [4] and image semantic segmentation [5].

Even though DNNs deliver high performance on their trained tasks, DNNs are overconfident by delivering unreliable high confidence on incorrect predictions [6]. This drawback becomes a serious limitation for AD systems. This typical insufficiency of a DNN leads to false or ghost detections, reducing the overall performance of an AD system [7].

In recent years, a common solution to this limitation involves enhancing DNNs with the capability to explicitly express uncertainty regarding their output predictions [8, 9]. The introduction of uncertainty modeling stands as a fundamental advancement in mitigating

^{*}Stellantis, Opel Automobile GmbH, Rüsselsheim am Main.

(ahmedmostafa.hammam@external.stellantis.com)

[†]Technische Hochschule Mittelhessen, Gießen.

[‡]Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, Karlsruhe.

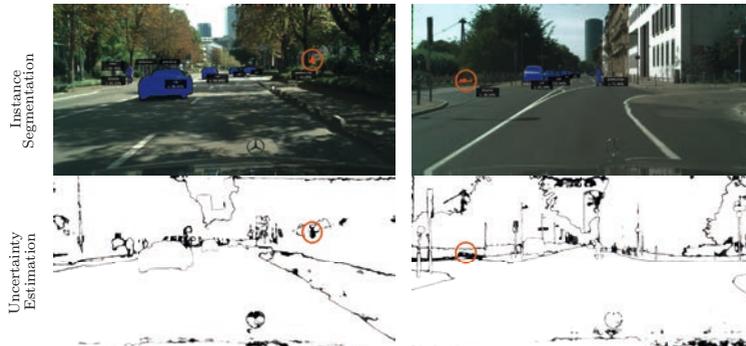


Figure 1: Sample DNN Outputs: Instance segmentation results (Top Row) and corresponding uncertainty estimates (Bottom Row). True positive detections are color-coded in blue, denoting high certainty; false detections are color-coded in red, indicating low certainty. The DNN effectively distinguishes between both detections, associating high certainty to true positives and low certainty to false ghost detections.

the insufficiencies of DNNs, particularly their tendencies toward overconfidence or lack of precision [10].

Building upon prior research [11–13], our study introduces a method centered around an enhanced architecture incorporating two key layers: Dirichlet Maximum Likelihood Estimation (MLE) and Intermediate Layer Variational Inference (ILVI). The integration of these two layers is to empower the DNN for accurate instance segmentation within scenes, while concurrently ensuring reliable uncertainty estimation.

By achieving reliable uncertainty estimates, our approach addresses the issue of ghost detections by being able to distinguish between true positives from false positives. This separation enables the incorporation of thresholding techniques, which play a crucial role in reducing ghost detections. Consequently, our methodology enhances uncertainty estimation capabilities and contributes to more accurate and robust results.

The remainder of the paper is organized as follows: Section 2 presents related work to Dirichlet modeling and its application in uncertainty estimation, whilst the architecture is discussed and explained in Section 3. The experiments conducted to test our approach are displayed in Section 4 and a conclusion of the work is discussed in Section 5.

2 Related Work

In recent years, efforts have concentrated on enhancing the credibility of generated outputs by modeling uncertainty within DNNs [14, 15]. Specifically, the focus has shifted towards modeling DNN outputs as Dirichlet distributions to refine uncertainty estimation. One approach has been Dirichlet Prior Networks, which builds upon the framework introduced in [16] by modeling the predicted logits from the DNN as the concentration parameters of a Dirichlet distribution that serves as a prior for the categorical distribution. Similar to Prior networks, the authors in [17] proposed a method that combines Prior networks with

the likelihood to maximize the whole posterior. Inspired by the Dempster-Shafer theory of evidence (DST) [18], they treat the predictions of the DNN as subjective opinions and train the DNN to gather evidence supporting these opinions. Additionally, a penalty term is introduced to penalize the DNN for incorrect detections and encourage it to exhibit high uncertainty in such cases.

Inspired by previous studies of [16, 17], our goal is to utilize Dirichlet models to enhance the reliability of uncertainty estimation and maintain segmentation performance. Optimizing the reliability of uncertainty estimation in the Dirichlet DNN by formulating its loss function using KL divergence is often considered challenging [16].

3 Methodology

In this section, we outline the core components of our architecture, presented in Figure 2, emphasizing their distinct roles. Subsequently, we explore the semantic segmentation decoder in-depth, showcasing its integration with the Dirichlet layer for improved uncertainty estimation. We then explain the ILVI approach, followed by a description of the applied thresholding methodology aimed at mitigating false positive occurrences.

3.1 Dirichlet DNN Architecture

The architecture, presented in Figure 2a, comprises a shared backbone that takes the input image and passes the extracted features to the ILVI module. The ILVI module, presented in Figure 2b, acts as a regularizer by adding stochasticity in the DNN avoiding overfitting and overconfidence. The output of the ILVI is passed on to the semantic segmentation decoder and the instance segmentation decoder.

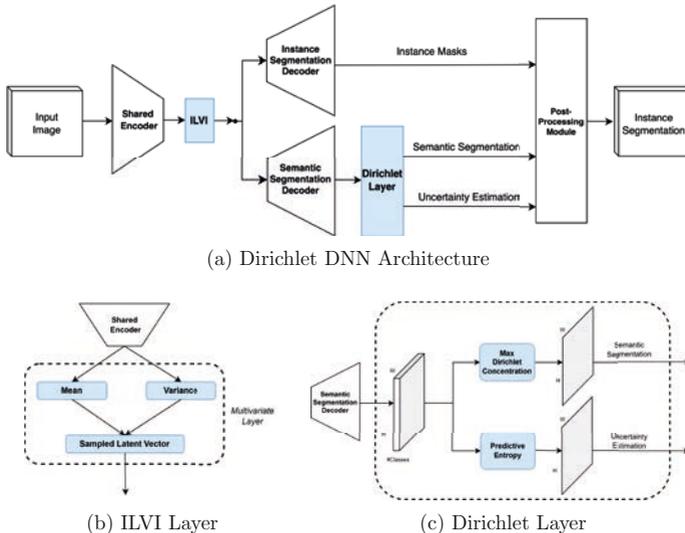
The semantic segmentation decoder and the Dirichlet layer, shown in Figure 2c, are trained together to model the semantic segmentation output as a Dirichlet distribution, which enables the uncertainty estimation. This decoder generates two results; semantic segmentation and uncertainty estimation based on the per-pixel Dirichlet distribution. The instance segmentation decoder generates the center points and the masks of the instances. The generated outputs from both decoders are passed onto the post-processing module. For every instance identified by the instance decoder, the instance mask, instance class, and the average uncertainty estimate are provided.

The architecture integrates the lightweight MobileNetV3 [19] as its foundational backbone and encompasses semantic and instance segmentation decoders influenced from Panoptic Deeplab [20].

The training of this architecture is governed by the following composite loss function:

$$\mathcal{L} = \mathcal{L}_{sem} + \mathcal{L}_{ILVI} + \mathcal{L}_{ins} \quad (1)$$

Here, \mathcal{L}_{sem} represents the loss function for per-pixel classification in semantic segmentation, combined with the Dirichlet distribution modeling. \mathcal{L}_{ILVI} stands for the ILVI loss which introduces stochasticity to the DNN, thereby enhancing its uncertainty estimation capacity and generalization efficacy. The third component, \mathcal{L}_{ins} , pertains to the instance segmentation loss. The specifics of each loss term are further explained in the corresponding following sections. [21]



(a) Dirichlet DNN Architecture

(b) ILVI Layer

(c) Dirichlet Layer

Figure 2: The Dirichlet MLE DNN is illustrated, emphasizing its key components for enhancing OOD identification. The ILVI layer introduces a multivariate layer structure, while the Dirichlet layer handles semantic segmentation and uncertainty estimation.

3.2 Semantic Segmentation Decoder

In this section, we explain the fundamental concepts and methods at the core of our approach. We begin with an in-depth exploration of the Dirichlet distribution within the probability simplex, a foundational probabilistic structure characterized by concentration parameters.

A supervised network aims to predict the target value $y \in \mathcal{Y}$ for an input $x \in \mathcal{X}$, where the input space \mathcal{X} corresponds to the space of images. Accordingly, a supervised machine learning problem with the task of semantic segmentation has a target \mathcal{Y} consisting of a finite set of c classes where the task for the network is to predict the class of each pixel out of the set of classes K . For our purpose, a DNN is defined as a function $f_w : \mathcal{X} \rightarrow \mathcal{Y}$, parameterized by $w \in \mathbb{R}$, which maps an input $x \in \mathcal{X}$ to an output $f_w(x) \in \mathcal{Y}$.

Given the probability simplex as $\mathcal{S} = \{(\theta_1, \dots, \theta_k) : \theta_i \geq 0, \sum_i \theta_i = 1\}$, the Dirichlet distribution is a probability density function on vectors $\theta \in \mathcal{S}$ and categorized by concentration parameters $\alpha = \{\alpha_1, \dots, \alpha_K\}$ as:

$$\text{Dir}(\theta; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K \theta_i^{\alpha_i - 1} \quad (2)$$

where the normalizing constant $\frac{1}{B(\alpha)}$ denotes the multivariate Beta function $B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\alpha_0)}$, $\alpha_0 = \sum_{i=1}^K \alpha_i$ and Gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, and θ denotes the ground truth probability distribution [22]. To model the Dirichlet distribution, the con-

centration parameters α correspond to each class output from the semantic segmentation decoder as follows: $\alpha = f_w(x)$, where α changes with each input x .

To train the Dirichlet distributions, we propose a direct maximization of the likelihood, inspired by the works of [12, 13]. Unlike the Dirichlet Prior and Evidential approaches, our method eliminates the constraints of the KL-divergence term in the loss function, allowing the DNN to explore the weight space more freely. This leads to improved segmentation performance and enhanced reliability in uncertainty estimation by encouraging a sharper concentration of Dirichlet parameters for correct predictions and flatter distributions for incorrect predictions.

Training a Dirichlet DNN with maximum likelihood estimation (MLE) can be done by minimizing the negative log-likelihood [22] as follows:

$$F(\alpha; \theta) = \log \prod \text{Dir}(\theta; \alpha) = \log \Gamma \left(\sum_{i=1}^K \alpha_j \right) - \sum_{i=1}^K \log \Gamma(\alpha_i) + \sum_{i=1}^K (\alpha_i - 1) \log \theta_i \quad (3)$$

where θ represents the probability distribution to be maximized.

We aim to train the DNN to produce reliable uncertainty estimations by treating the DNN's correct and incorrect predictions separately. Our primary objective is to obtain accurate predictions with low uncertainties, while assigning high uncertainty to incorrect predictions.

To ensure high certainty for correct predictions, the DNN should exhibit a strong concentration toward the correct class, as shown in Figure 3a. This can be achieved by maximizing the likelihood using the ground truth label probability and employing a *one-hot vector*. Conversely, for incorrect predictions, high uncertainty is achieved by maximizing the likelihood using an equal probability vector with equal probabilities assigned to all classes, as shown in Figure 3b.

To address these cases, we extend the formulation presented in Equation 3 for the semantic segmentation as follows:

$$\mathcal{L}_{sem} = F(\alpha_{correct}; \theta_{correct}) + F(\alpha_{incorrect}; \theta_{incorrect}), \quad (4)$$

where $\alpha_{correct}$ and $\alpha_{incorrect}$ are the network's concentration parameters representing the correct and incorrect DNN predictions respectively, and $\theta_{correct}$ and $\theta_{incorrect}$ represent the ground truth probability distribution for the correct classes and the equal probability vector to yield high uncertainty respectively.

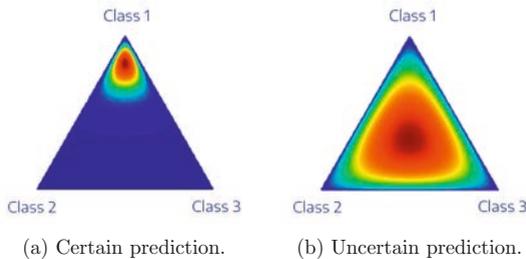


Figure 3: Dirichlet Plots

In this work, we model the uncertainty estimation of the Dirichlet distribution using the predictive entropy:

$$\hat{\mathbb{H}}[y|x] = - \sum_c (p(y = c|x, w)) \log(p(y = c|x, w)) \quad (5)$$

where y is the output variable, c ranges over all the classes K , $p(y = c|x, w) = \frac{\alpha_c}{\sum \alpha}$ is the probability of the input x being class c , and w are the model parameters. The class of each pixel for the semantic segmentation output is determined according to the highest concentration value of the Dirichlet distribution.

3.3 Intermediate Layer Variational Inference

The Intermediate Layer Variational Inference (ILVI) approach is designed to address the limitations of existing Bayesian deep neural network approximation techniques. The concept of ILVI, presented in Figure 2b, modifies a latent layer in the network to take the shape of a multivariate Gaussian distribution with mean and variance, instead of using single point estimates. Studies showed that by adopting this method, stochasticity is introduced allowing for the sampling of points from this layer and consequently improving the uncertainty estimation of the DNN and also its generalization performance [11].

The variational posterior of ILVI is modeled as a diagonal Gaussian distribution. The weight parameters w are sampled from this distribution using the reparametrization trick [23]. Specifically, each weight parameter is computed as: $w = \mu + \sigma \odot \epsilon$. Here, μ represents the mean of the Gaussian distribution, σ represents the standard deviation parameter, $\epsilon \sim \mathcal{N}(0, I)$ is a random variable, and \odot is the pointwise multiplication. The parameter σ is pointwise parameterized as $\sigma = \log(1 + \exp(\rho))$, ensuring its non-negativity. This formulation preserves the mean and log-variance vectors as learnable parameters while introducing stochasticity through the random variable ϵ .

The ILVI method utilizes Bayesian variational inference based on the Kullback-Leibler (KL) divergence, following:

$$\mathcal{L}_{ILVI} = KL(q(\phi) || p(w|X, Y)). \quad (6)$$

This loss comprises of approximating a new probability distribution that is close to the posterior distribution produced by the model. To achieve the new approximate distribution, the KL divergence is needed to minimize the new variational parameters $q(\phi) \approx p(w|X, Y)$ for approximation.

3.4 Instance Segmentation

The instance segmentation decoder is a crucial component of the architecture, designed to identify individual object instances within an image. This decoder operates in conjunction with the semantic segmentation decoder to provide comprehensive scene understanding.

The output from the ILVI layer passes to the instance segmentation decoder to refine the features and generate predictions for object instances. Specifically, the decoder produces two outputs: instance masks and center points. Instance masks define the spatial boundaries of separate objects, while center points identify the most prominent pixel

within each object. These central points function as reference markers for precisely determining the locations of objects within the scene. Accordingly, the instance segmentation loss is formulated as follows:

$$\mathcal{L}_{ins} = \mathcal{L}_{center} + \mathcal{L}_{mask}. \quad (7)$$

The center points loss \mathcal{L}_{center} employs the mean squared error loss to penalize the DNN’s center points with respect to the ground truth. In the case of the instance offset loss \mathcal{L}_{mask} , the L1 loss is utilized to penalize the difference between the DNN’s generated masks and the ground truth mask [20].

3.5 Thresholding for Uncertainty-Based Filtering

At its core, thresholding involves establishing a threshold value for the model’s uncertainty estimates. Detections with uncertainties below this threshold are discarded, effectively reducing false positive instances. The primary aim here is to enhance precision, the ratio of correctly predicted positive instances to all predicted positive instances. While this approach significantly improves precision, it may have effects on the recall, the ratio of correctly predicted positive instances to all actual positive instances. An inherent challenge with thresholding is the potential reduction in recall due to the elimination of detections below the threshold. This trade-off can result in missed true positive instances, which in turn may impact the model’s overall recall performance.

The effectiveness of thresholding relies on selecting the appropriate threshold value. To choose the threshold value, the precision and recall values should be plotted against varying threshold values. This visually illustrates the trade-off between these two crucial performance indicators. The threshold is selected to maximize precision improvement without compromising recall performance.

4 Experiments and Results

In the following section, we present a comprehensive analysis of our methodology and the corresponding results. Experiments conducted encompass evaluation of segmentation performance, uncertainty estimation, and uncertainty thresholding.

We undertake a thorough evaluation of our methodology’s performance in comparison to two state-of-the-art approaches: Prior Network, and Evidential Network whilst having Cross Entropy (CE) as our baseline. In this study, DNNs are trained on the Cityscapes dataset [24] and evaluated using its validation set. Additionally, we test their adaptability on the KITTI dataset [25], which examines the models’ resilience and real-world applicability across varying environments evaluating its generalization capabilities. In this work, we model the uncertainty estimation of all approaches using predictive entropy.

4.1 Segmentation Performance

The results of the DNNs’ segmentation performance are shown in Table 1. In this table, we present a comprehensive comparison of instance and semantic segmentation performance across the different approaches on both the Cityscapes and KITTI datasets. The metrics used include precision and precision up to 50 meters (Precision 50m) for

Table 1: Performance comparison for instance and semantic segmentation. The Dirichlet MLE + ILVI method achieves the highest scores across multiple metrics on both datasets, highlighting significant performance improvement.

	Cityscapes			KITTI	
	Precision	Precision 50m	mIoU	Precision	mIoU
CE	38.2	35.8	65.2	39.1	47.6
Prior	39.8	37.5	66.7	38.7	48.2
Evidential	42.9	39.8	68.1	40.6	50.1
Dirichlet MLE + ILVI	45.1	43.5	69.1	42.1	51.3

instance segmentation and mean Intersection over Union (mIoU) for semantic segmentation.

The effectiveness of the Dirichlet MLE + ILVI approach in our work is clearly demonstrated in the presented results in Table 1. Across both the Cityscapes and KITTI datasets, the Dirichlet MLE + ILVI approach consistently outperforms other techniques, showcasing its effectiveness in enhancing instance and semantic segmentation performance.

Moreover, the Dirichlet MLE + ILVI approach shows improvements in precision for the Cityscapes dataset, surpassing other methods. It also excels in precision 50%, precision 100m, and mIoU. This superiority extends to the KITTI dataset, with the highest precision 50% and mIoU. Consistent cross-dataset performance underscores its strength in instance and semantic segmentation. This highlights its potential to enhance localization accuracy and semantic understanding, making it invaluable for scene analysis tasks.

4.2 Uncertainty Estimation Performance

Two key evaluation metrics, namely separation efficiency and accuracy vs. certainty, are utilized to assess the efficacy of the proposed method in enhancing uncertainty estimation and prediction accuracy. Table 2 presents the results and offers a comprehensive insight into the performance of different uncertainty estimation approaches.

Distributional Separation Efficiency

We aim to quantify the efficiency of the DNN to differentiate between correct and incorrect predictions by plotting their corresponding certainty distribution for both cases. The distributions are then compared using the Wasserstein distance metric, where a high value indicates dissimilar distinctive distributions and vice versa. Notably, our Dirichlet MLE + ILVI method exhibits a significantly higher Wasserstein distance surpassing the other approaches.

Accuracy vs. Certainty

An important factor for deep neural networks is not only to be accurate about their predictions but also to be certain about them. Proposed by [26], ratios between accuracy and certainty are defined and quantified to compare between DNNs' uncertainty performances.

Table 2: Separation efficiency and accuracy vs. certainty (\uparrow). These results collectively emphasize the efficacy of the Dirichlet MLE + ILVI method in achieving a balance between separation efficiency and accuracy across varying levels of certainty.

	Separation Efficiency Wasserstein Distance	Accuracy vs. Certainty (%)		
		P(A C)	P(U I)	AvU
CE	1.1	50.9	24.4	51.8
Prior	2.3	72.7	17.5	38.5
Evidential	3.3	53.2	63.1	61.2
Dirichlet MLE + ILVI	4.9	85.4	78.1	70.1

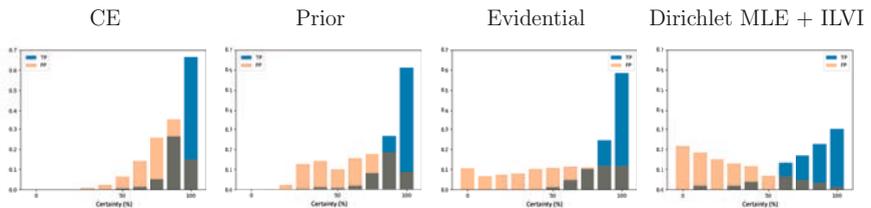


Figure 4: Distribution plots of true positives (TP) and false positives (FP) showing improved separation capability of Dirichlet MLE + ILVI method.

After attaining the DNN predictions and their respective uncertainty estimations, they are compared with the ground truth.

Three conditional probabilities are needed for this evaluation test, following the work of [26]: $p(\text{accurate}|\text{certain}) = \frac{n_{ac}}{n_{ac}+n_{ic}}$, $p(\text{uncertain}|\text{inaccurate}) = \frac{n_{iu}}{n_{ic}+n_{iu}}$ and $AvU = \frac{n_{ac}+n_{iu}}{n_{ac}+n_{au}+n_{ic}+n_{iu}}$, where n_{ac} are the accurate and certain predictions, n_{au} are the accurate and uncertain predictions, n_{ic} are the inaccurate and certain predictions, and n_{iu} are the inaccurate and uncertain predictions. AvU stands for accuracy vs. uncertainty, describing the probability of getting a good prediction out of the network either accurate and certain or inaccurate and uncertain.

In Table 2, our proposed method stands out with high accuracy vs. certainty percentage indicating the ability in generating predictions that are both accurate and certain. Additionally, it can be observed that there is a high improvement in the P(U|I) whilst still maintaining high performance on the other two metrics. This is not frequently observed as any method trying to improve uncertainty representation would come to a cost of reduced performance on the other two metrics. This finding underscores the approach’s capability to effectively recognize uncertain predictions.

4.3 Uncertainty Estimation Thresholding

By setting a certainty threshold, the DNN can effectively reduce the occurrence of false positives, leading to an improvement in its overall precision. The threshold acts as a cutoff value where detections less than the thresholded certainty are omitted.

For that, to choose the threshold we first plot the average precision and recall at varying thresholds, as shown in Figure 5. Average precision and recall are plotted for

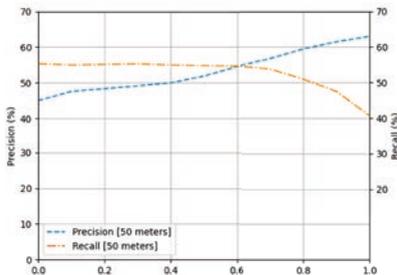


Figure 5: Average precision and recall values up to 50 meters are plotted with varying uncertainty thresholds.

Table 3: Precision and recall performance comparison before and after thresholding (in % (\uparrow)).

	Precision	Precision 50m	Recall	Recall 50m
CE	38.2	35.8	31.8	45.6
Prior	39.8	37.5	34.5	48.9
Evidential	42.9	39.8	37.1	51.3
Dirichlet MLE + ILVI	45.1	43.5	39.4	55.3
Dirichlet MLE + ILVI (Thresholded)	57.2	54.5	37.5	54.2

detections upto only 50 meters, as most false positive detections are observed within 50 meters range of the vehicle. This plot shows the precision and recall values at varying thresholds in steps of 10%. A threshold of 0% signifies the absence of thresholding, implying the utilization of all detections. Conversely, a threshold of 100% signifies the inclusion of solely those detections exhibiting 100% certainty.

With increasing the threshold it can be observed that precision increases, indicating that false positives are correctly being eliminated. As for the recall, it remains almost steady until threshold of 60% and begins decreasing. This reflects the elimination of true positives after this value hence reducing the recall performance of the DNN.

For that, taking the threshold at 60% gives a good trade-off between improved precision whilst maintaining recall performance. A higher threshold would risk omitting correct detections, and a lower threshold would keep unneeded false positives.

In Table 3, a comparison between before and after thresholding for average precision and recall for overall detections and detections upto 50 meters are displayed. We can see a significant improvement for the average precision over all other approaches and also for our method without thresholding. As for recall, we can see an increase over the baseline and state-of-the-art approaches, but slightly less than the Dirichlet MLE + ILVI without thresholding.

This improvement of precision stems from the fact that the DNN is able to associate high certainty to true positives and low certainty to false positives, hence achieving an efficient separation between both categories, as shown in Table 2 and Figure 4.

5 Conclusion

This paper presents a novel approach to enhance instance segmentation performance and uncertainty estimation in deep neural networks for automated driving applications. The proposed architecture combines the Dirichlet Maximum Likelihood Estimation approach, Intermediate Layer Variational Inference, and uncertainty-based thresholding to achieve more accurate instance segmentation and reliable uncertainty estimates. This integration introduces a new perspective on countering ghost detections through uncertainty thresholding. By thresholding the uncertainty, we observe a boost in the precision performance of the DNN whilst maintaining high recall performance. The demonstrated performance improvements establish the potential of this methodology for advancing the capabilities of AI-driven systems in real-world applications. Future research can further explore the optimization of threshold selection to enhance the reliability and effectiveness of uncertainty estimation in DNNs for automated driving applications.

ACKNOWLEDGMENT

This work is partly funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) and partly financed by the European Union in the frame of NextGenerationEU within the project "Solutions and Technologies for Automated Driving in Town" (FKZ 19A22006P).

References

- [1] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.
- [2] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020.
- [3] Keli Huang, Botian Shi, Xiang Li, Xin Li, Siyuan Huang, and Yikang Li. Multi-modal sensor fusion for auto driving perception: A survey. *arXiv preprint arXiv:2202.02703*, 2022.
- [4] Szilárd Aradi. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [5] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [6] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

- [7] Oliver Willers, Sebastian Sudholt, Shervin Raafatnia, and Stephanie Abrecht. Safety concerns and mitigation approaches regarding the use of deep learning in safety-critical perception tasks. In *International Conference on Computer Safety, Reliability, and Security*, pages 336–350. Springer, 2020.
- [8] Alex Kendall. *Geometry and uncertainty in deep learning for computer vision*. PhD thesis, University of Cambridge, UK, 2019.
- [9] Yarin Gal. Uncertainty in deep learning. 2016.
- [10] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 2021.
- [11] Ahmed Hammam, Seyed Eghbal Ghobadi, Frank Bonarens, and Christoph Stiller. Real-time uncertainty estimation based on intermediate layer variational inference. In *Computer Science in Cars Symposium*, pages 1–9, 2021.
- [12] Ahmed Hammam, Frank Bonarens, Seyed Eghbal Ghobadi, and Christoph Stiller. Predictive uncertainty quantification of deep neural networks using dirichlet distributions. In *Proceedings of the 6th ACM Computer Science in Cars Symposium*, pages 1–10, 2022.
- [13] Ahmed Hammam, Frank Bonarens, Seyed Eghbal Ghobadi, and Christoph Stiller. Towards improved intermediate layer variational inference for uncertainty estimation. In *European Conference on Computer Vision*, pages 1–14. Springer, 2022.
- [14] Jakob Gawlikowski, Cedrique Rovile Njietcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*, 2021.
- [15] Hao Wang and Dit-Yan Yeung. A survey on bayesian deep learning. *ACM Computing Surveys (CSUR)*, 53(5):1–37, 2020.
- [16] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.
- [17] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.
- [18] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 57–72. Springer, 2008.
- [19] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF*

- International Conference on Computer Vision (ICCV)*, pages 1314–1324, October 2019.
- [20] Bowen Cheng, Maxwell D Collins, Yukun Zhu, Ting Liu, Thomas S Huang, Hartwig Adam, and Liang-Chieh Chen. Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12475–12485, 2020.
- [21] Ahmed Hammam, Frank Bonarens, Seyed Eghbal Ghobadi, and Christoph Stiller. Identifying out-of-domain objects with dirichlet deep neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2023.
- [22] Thomas Minka. Estimating a dirichlet distribution, 2000.
- [23] Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28, 2015.
- [24] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.
- [25] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.
- [26] Jishnu Mukhoti and Yarin Gal. Evaluating bayesian deep learning methods for semantic segmentation. *arXiv preprint arXiv:1811.12709*, 2018.