Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it
Bonn-Aachen
International Center for
Information Technology

R&D Project Proposal

Benchmarking Uncertainty Estimation of Deep Learning Models Using Synthetic Dataset

Sathwik Panchangam

Supervised by

Prof. Dr. Nico Hochgeschwender
MSc. Deebul Nair

May 2022

1 Introduction

Estimating uncertainty in the predictions of the deep learning models has gained popular interest in recent years. Basic neural networks provide only predictions and do not deliver uncertainty estimates and they suffer from over or under confidence [4]. Inabilities like providing reliable uncertainty estimates, over confident predictions, distinguishing between in-domain and out-of-domain samples can be considered as some of the examples for the limitations of deploying Deep neural networks (DNN)s in real world applications and safety critical systems like Robotics and medical image analysis etc [4]. Making a wrong decision in applications like self driving cars, autonomous systems and health care does not only result in the failure of the task but might even put human lives at risk.

Deep learning algorithms can be fully integrated into robotic systems only if a measure of prediction uncertainty is available. Also for a safe decision making, estimating the predictive uncertainty alone is not sufficient, but a good evaluation is required to check the quality of the uncertainty estimates. The current state of the art uncertainty estimation in deep learning has provided many uncertainty estimation techniques like, dropout, ensembles, Bayesian networks, deterministic approaches. Assessing the quality of these uncertainty estimates is not straightforward, as there are no direct uncertainty ground truth labels available for evaluation. Until recently, identifying the ground truth for uncertainty estimation had received little attention and in recent works researches have used Bayesian learning methods to measure the fidelity of approximate inference procedures in deep learning [10].

However these methods do not provide any ground truth uncertainty labels for good evaluation and comparison of the DNN models confidence. Collecting real world data sets with human defined ground truth uncertainty labels can be very difficult, time consuming and expensive. Synthetic data also called as artificial data has been used widely in the state of the art deep learning applications like self driving cars, robotics and health care.Generating such synthetic data sets along with the ground truth uncertainty labels might provide some of the answers to the challenges in the state of the art uncertainty estimation in deep learning models.

Hence, in this project we would like to address the problem of defining the ground truth for uncertainty estimation based on human/rule based criteria and conduct experiments using the state of the art uncertainty estimation techniques on different synthetic data sets generated using a simulation software like blender.By addressing the problem of defining the ground truth we are interested to set a benchmark for the uncertainty in Deep learning using synthetic data sets generated in a simulation software like blender.

1.1 Problem Statement

Defining ground truth for the evaluation of predictive uncertainty estimation is one of the challenging problems in state-of-the-art Deep learning. The quality of the uncertainty estimates mainly depends on the method used for estimating the uncertainty. The state of the art uncertainty estimation techniques are more relied on running the neural network many times or sampling the neural network to determine the DNN model's confidence but do not use any ground truth labels. Also, research in the state of the art uncertainty estimation in deep learning shows that defining ground truth uncertainties is challenging and there is a lack of ground truth for uncertainty estimates [4] [5]. In Recent work researches have used Bayesian learning methods to measure the fidelity of approximate inference procedures in deep learning [10]. However, these current state of the art uncertainty estimation methods did not produce any ground truth uncertainty labels for good evaluation and comparison. In this R&D we would like to address the problem of defining the ground truth for uncertainty estimation using human/rule based criteria. By addressing the problem of defining the ground truth for the uncertainty we are interested to set a benchmark for the uncertainty in deep learning models using synthetic data sets generated in a simulation software like blender.

1.2 Research Questions

1. Why is it difficult to define ground truth for uncertainty estimation?
2. Which method can be used for determining the ground truth for uncertainty?
   - How do we generate ground truth values from the rule based criteria?
   - How do we map the uncertainties to the synthetic data sets?
3. What are the requirements for generating uncertainty in different scenarios?
4. Which method can be used to determine the threshold region for the uncertainty in different scenarios?

2 Related Work

2.1 Uncertainty

Uncertainty in a deep learning model is mainly caused due to two types of uncertainty. First uncertainty is caused due to the the lack of knowledge of the neural network which is termed as model uncertainty or epistemic uncertainty and the second is type of uncertainty is caused due the presence of uncertainty in the training data, which is termed as data uncertainty or aleatoric uncertainty.

One of the most common way to estimate the predictive uncertainty is based on separately modelling the epistemic and aleatoric uncertainties. The current state-of-the-art uncertainty estimation methods uses Bayesian inference [3], ensembles [5], test time data augmentation [9] or single deterministic networks [8] to estimate the uncertainty of the predictions in DNN models.

Most of the state of the art uncertainty estimation techniques rely on running the neural network multiple times or sampling the neural network to determine the DNN model's confidence increasing the computation and training time. Research shows that single model approaches are more computationally efficient in training and evaluation [4] compared to the state-of-the-art uncertainty estimation approaches in deep learning.

Within a deterministic network, single model approaches provide a prediction based on a single forward pass. Only one network can be used for training and often these approaches can even be applied on pre-trained networks. Depending on the actual approach, only a single or at most two forward passes have to be fulfilled for evaluation [4].

2.2 Blender for generating data-sets

In recent years synthetic data set generation has become a popular alternative to the real world data sets and there is a vast usage of this synthetic data sets in the state of the art deep learning for various applications like self driving cars, robotics, segmentation and other computer vision tasks. Blender an open-source software [1] can be used to generate synthetic data sets with the help of Blender API allowing the user for programmatic control of environment, scenes and render parameters of the software to automatically generate images and create data sets for different applications. The usage of blender for generating synthetic data sets has been increasing in the recent years due to it real time rendering capabilities. It is used for generating synthetic data sets for various applications like instance segmentation [7], Ioannis Mariolis et al. used blender to generate depth images for the pose and category recognition of objects [6].

3 Proposed Solution

.

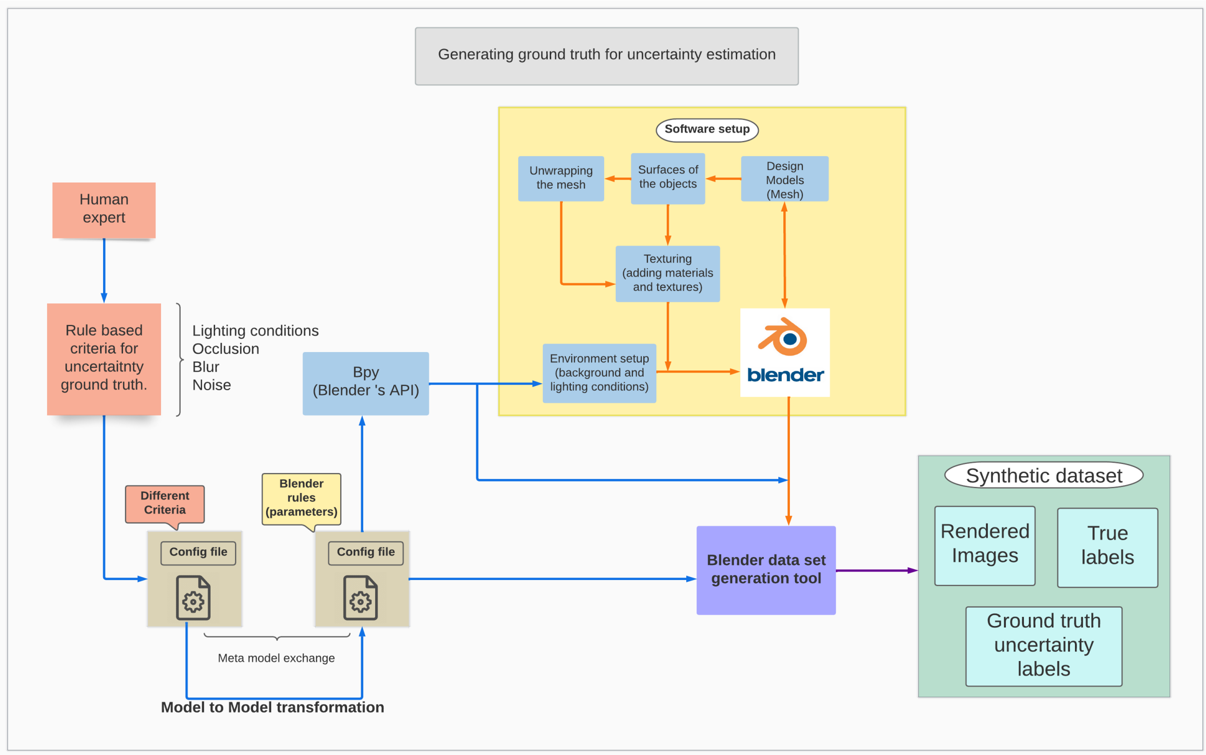

Figure 1: Proposed method

- In this R&D project we would like to address the problem of defining the ground truth for uncertainty estimation in deep learning models.
- To achieve this we would like to define some rule based criteria for the uncertainty in different scenarios like lighting, occlusion and blur using fuzzy rules and human expertise.
- Based on this human/rule based criteria we would like to map the uncertainty values to the synthetic data set generated in blender.
- For this we are interested to create different configuration files for the rule based criteria and modifiable parameters in blender and perform a model

to model transformation to obtain the parameters which can be used to change the scene conditions in blender software and also to generate the corresponding ground truth uncertainty labels.

- Through this approach we can generate a data set containing images, true labels and the ground truth uncertainty labels.
- Obtaining such data sets for real world images with labels is time consuming and very costly.
- Hence in this project we are interested to generate different synthetic data-sets using an open source software blender [1] [2] by applying textures and providing necessary environment conditions to the CAD designed models of objects present in YCB benchmark object data-set and Robocup@work components data-set.
- Using these generated data-sets we would like to estimate the uncertainty and compare it with the real world data-set using single model approaches.
- These uncertainty estimates are desired to be computed using single model approaches as they are computationally efficient in training and evaluation.
- One of the main focus of this RnD project is to create a benchmark for estimating and comparing the uncertainty for different criteria like lighting, occlusion and blur.
- This Rnd also focuses on estimating uncertainties using different uncertainty estimation methods like dropout, ensembles and also on different DNN models like Resnet, MobileNet and SqeezeNet.
- Finally we would like perform a comparative evaluation of all the predicted uncertainties of different blender generated data-sets and the real world data-sets.

4 Project Plan

4.1 Work Packages

The bare minimum will include the following packages:

WP1 Literature search

- Literature search on single model approaches for estimating uncertainty in classification tasks.
- Literature search on evaluation methods for uncertainty estimation.
- Literature search on using blender for data-set generation.
- Literature search on defining ground truth for uncertainty estimation.

WP2 Methodology to define ground truth and mapping uncertainty.

- Defining the human/rule based criteria for different conditions like lighting, occlusion, blur.
- Defining a methodology to map the ground truth uncertainties to the synthetic data set.

WP3 Design of objects.

- Design the cad models for the components present in Robocup@work dataset (18 classes) in a 3d modelling software
- Design the cad models for the objects present in YCb object dataset (10 classes) in a 3d modelling software.
- Design the cad models for the objects Hope dataset (20 classes) in a 3d modelling software.
- Add materials and textures for all the cad models using blender software.

WP4 Data set generation

- Setup different environment and lighting conditions of the scene with cad models using blender software for generating dataset.

- Generate data-set of textured objects in different environments and lighting conditions using blender for the objects present in Robocu@work components, YCB objects and Hope data set.
- Generate data-set of objects with different textures on a same background or same environment conditions for all the objects.
- Generate data-set of objects with same textures on a different background or different environment conditions for all the objects.

WP5 Mapping uncertainty to the synthetic dataset.

- Generate the ground truth values from the rule based criteria and map the values to the synthetic datasets generated using blender's API(Bpy).

WP6 Mid term report

- Train a DNN model for estimating uncertainty in classification task using evidential loss on Robocup@work components real world dataset.
- Train a DNN model for estimating uncertainty in segmentation task using evidential loss on Robocup@work components real world dataset.
- Generating synthetic datasets for Robocup@work components dataset using blenderproc code.
- Train a DNN model for estimating uncertainty in classification task using evidential loss on Robocup@work components synthetic dataset.
- Train a DNN model for estimating uncertainty in segmentation task using evidential loss on Robocup@work components synthetic dataset.

WP7 Uncertainty estimation.

- Estimating uncertainty using single model approaches for classification task on the real world data-sets of YCB benchmark objects and Robocup@work components.
- Estimating uncertainty using single model approaches for classification task on blender generated data-sets of YCB benchmark objects and Robocup@work components.

- Estimating uncertainty using dropout, ensembles for classification task on blender generated data-sets of YCB benchmark objects and Robocup@work components.
- Performing all the experiments of uncertainty estimation using Resnet, Mobilenet and squeeze net architectures for all the available datasets.
- Comparative evaluation of the estimated uncertainties of different uncertainty estimation methods on both real world and blender generated datasets and also on three deep learning architectures like ResNet, MobileNet and SquezeNet.

WP8 Project report

- Documentation of state of the art uncertainty estimation techniques like single model approaches,dropout and ensembles.
- Documentation of the methodology for defining the human/rule based criteria and generating ground truth.
- Documentation of the methodology for mapping the ground truth to the synthetic dataset generated using blender software.
- Documentation of using blender for data-set generation.
- Documentation of estimating uncertainty for Robocup@work components data-set and YCB benchmark object data-set using single model approaches and DNN models like ResNet,MobileNet and Squeeze Net.
- Documentation of estimating uncertainty for Blender generated datasets of YCB objects and Robocup@work components using single model approaches and DNN models like ResNet,MobileNet and Squeeze Net..
- Documentation of evaluation of the estimated uncertainties of both real world and blender generated datasets using different uncertainty estimation methods like ensembles, dropout and different DNN models like ResNet, MobileNet and Squeeze Net.
- Documentation of results, conclusion and future work.
- Documentation of draft copy of RnD report.
- Documentation of Final RnD report.

4.2 Milestones

M1 Literature search

- Literature search on uncertainty estimation using single model approaches.
- Literature search on evaluation methods for uncertainty estimation.
- Literature search on using blender for data-set generation.
- Literature search on defining ground truth for uncertainty estimation.

M2 Defining the method to generate ground truth and mapping uncertainty.

- Defining the human/rule based criteria for different conditions like lighting, occlusion, blur as a configuration file based on fuzzy rules.
- Defining a methodology to map the ground truth uncertainties to the synthetic data set.

M3 Design of objects.

- Design the cad models for the components present in Robocup@work dataset (18 classes) in a 3d modelling software
- Design the cad models for the objects present in YCb object dataset (10 classes) in a 3d modelling software.
- Design the cad models for the objects Hope dataset (20 classes) in a 3d modelling software.
- Add materials and textures for all the cad models using blender software.

M4 Mapping uncertainty to the synthetic dataset.

- Generate the ground truth values from the rule based criteria and map the values to the synthetic datasets generated using blender's API(Bpy).

M5 Data set generation

- Setup different environment and lighting conditions of the scene with cad models using blender software for generating dataset.
- Generate data-set of textured objects in different environments and lighting conditions using blender software for the objects present in Robocup@work components, YCB objects and Hope data sets.
- Generate data-set of objects with different textures on a same background or same environment conditions for all the objects using blender software.
- Generate data-set of objects with same textures on a different background or different environment conditions for all the objects using blender software.

M6 Mid term report

- Train a DNN model for estimating uncertainty in classification task using evidential loss on Robocup@work components real world dataset.
- Train a DNN model for estimating uncertainty in segmentation task using evidential loss on Robocup@work components real world dataset.
- Generating synthetic datasets for Robocup@work components dataset using blenderproc code.
- Train a DNN model for estimating uncertainty in classification task using evidential loss on Robocup@work components synthetic dataset.
- Train a DNN model for estimating uncertainty in segmentation task using evidential loss on Robocup@work components synthetic dataset.

M7 Determining the ground truth

- Creating a configuration file from the rule based criteria for different scenarios like lighting conditions, occlusion and blur.
- Generating the ground truths from the rule based criteria/configuration file and mapping the values to the synthetic datasets generated in blender.

M8 Uncertainty estimation.

- Estimating uncertainty using single model approaches for classification task on the real world data-sets of YCB benchmark objects and Robocup@work components using PyTorch.
- Estimating uncertainty using single model approaches for classification task on blender generated data-sets of YCB benchmark objects and Robocup@work components using PyTorch.
- Estimating uncertainty using dropout, ensembles for classification task on blender generated data-sets of YCB benchmark objects and Robocup@work components using PyTorch.
- Performing all the experiments of uncertainty estimation using Resnet, Mobilenet and Squeeze net architectures using PyTorch for all the available datasets.
- Comparative evaluation of the estimated uncertainties of different uncertainty estimation methods on both real world and blender generated datasets and also on three deep learning architectures like ResNet, MobileNet and SquezeNet.

M9 Final R&D report

- Documentation of state of the art uncertainty estimation techniques like single model approaches,dropout and ensembles.
- Documentation of the methodology for defining the human/rule based criteria and generating ground truth.
- Documentation of the methodology for mapping the ground truth to the synthetic dataset generated using blender software.
- Documentation of using blender for data-set generation.
- Documentation of estimating uncertainty for Robocup@work components data-set and YCB benchmark object data-set using single model approaches and DNN models like ResNet,MobileNet and Squeeze Net.

- Documentation of estimating uncertainty for Blender generated datasets of YCB objects and Robocup@work components using single model approaches and DNN models like ResNet,MobileNet and Squeeze Net..
- Documentation of evaluation of the estimated uncertainties of both real world and blender generated datasets using different uncertainty estimation methods like ensembles, dropout and different DNN models like ResNet, MobileNet and Squeeze Net.
- Documentation of results, conclusion and future work.
- Documentation of draft copy of RnD report.
- Documentation of Final RnD report.

4.3 Deliverables

Minimum Viable

- Literature search on using simulation tools for generating data-sets.
- Literature search on evaluation methods for uncertainty estimation.
- Literature search on defining ground truth for uncertainty estimation.
- Defining a method for generating the ground truth values from the rule based criteria.
- Generating different data-sets of Robocup@work components dataset by varying the textures and background conditions in blender.
- Estimating uncertainty using single model approache sand different deep neural networks like ResNet, MobileNet and Squeeze Net on both real world data-sets and blender generated data-sets.

Expected

- Defining a method to map uncertainty ground truth values to the generated synthetic data-set.
- Mapping the uncertainty ground truths to the synthetic datasets generated in blender.
- Estimating uncertainty using single model approaches and different deep neural networks like ResNet, MobileNet and Squeeze Net on the real world data-sets and blender generated data-sets of YCB objects.
- Comparative evaluation of the estimated uncertainties from single model approaches for all the data-sets and in three different deep learning architectures like ResNet, MobileNet and SqueezeNet.

Desired

- Generating different synthetic datasets for YCB object dataset and Robocup@work dataset by changing the lighting, occlusion parameters in blender software.
- Estimating the uncertainty for the real world and synthetic data sets using the state-of-the-art uncertainty estimation methods like ensembles and dropout.
- Comparative evaluation of the predicted uncertainties for three deep learning models like ResNet, MobileNet and SqueezeNet using single model approaches, dropout and ensembles for different data sets.

4.4 Project Schedule

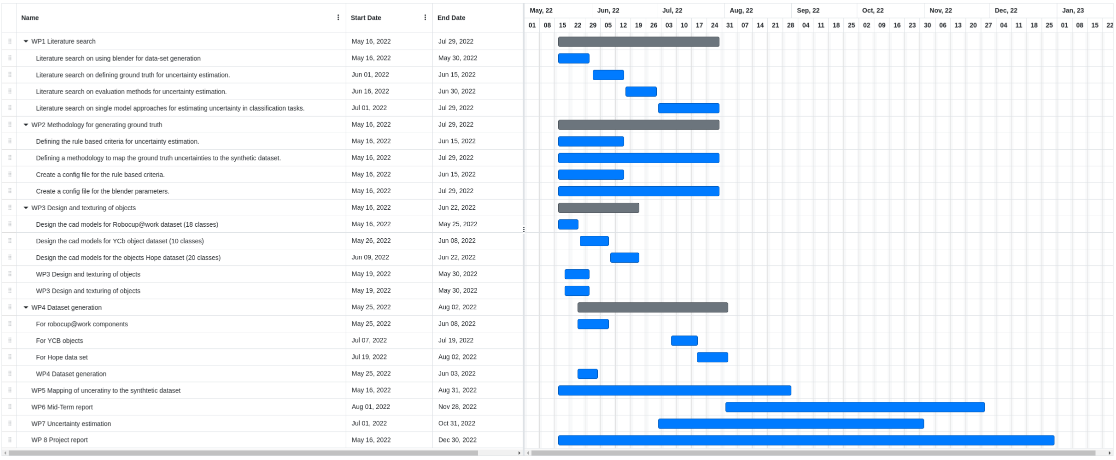

Figure 2: Project Timeline

References

[1] Blender is free software. `https://www.blender.org/about/license/`.

[2] Maximilian Denninger. Dlr-rm/blenderproc. `https://github.com/DLR-RM/BlenderProc`.

[3] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[4] Jakob Gawlikowski, Cedrique Rovile Njieutcheu 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.

[5] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[6] Ioannis Mariolis, Georgia Peleka, Andreas Kargakos, and Sotiris Malassiotis. Pose and category recognition of highly deformable objects using deep learning. In *2015 International Conference on Advanced Robotics (ICAR)*, pages 655–662, 2015. doi: 10.1109/ICAR.2015.7251526.

[7] Deokhwan Park, Joosoon Lee, Junseok Lee, and Kyoobin Lee. Deep learning based food instance segmentation using synthetic data. In *2021 18th International Conference on Ubiquitous Robots (UR)*, pages 499–505, 2021. doi: 10.1109/UR52253.2021.9494704.

[8] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in Neural Information Processing Systems*, 31, 2018.

[9] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[10] Andrew Gordon Wilson, Pavel Izmailov, Matthew D. Hoffman, Yarin Gal, Yingzhen Li, Melanie F. Pradier, Sharad Vikram, Andrew Foong, Sanae Lotfi, and Sebastian Farquhar. Evaluating Approximate Inference in Bayesian Deep Learning. 2021.