Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

Out-of-Distribution Detection in 3D Semantic Segmentation

Master Thesis

September 02, 2022

Lokesh Veeramacheneni

Out-of-Distribution detection

- Tesla autonomous driving system detects the moon as yellow traffic light
- These faulty predictions might result in autonomous agent behaving unpredictably
- An ideal trustworthy visual recognition system
  - Produce accurate predictions on known examples
  - Detect and reject unknown examples

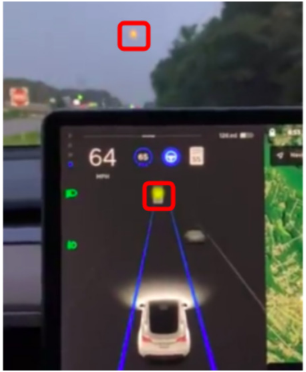


Figure 1: Caption

Out-of-Distribution detection

- Deep Neural Networks (DNNs) are trained based on closed world assumption
- closed world assumption - test data is assumed to be drawn from same distribution as training data which is called In-Distribution (ID)
- When deployed in real world (open world scenario) the test samples can be Out-of-Distribution (OOD) i.e. the test samples can be,
  - from different class
  - from different domain
- Moon is calssified as an OOD object in the above example

Importance of OOD detection

- Figure 2 depicts the pipeline of modules in Apollo driving platform.
- Prediction and motion planning module are dependent on perception module.
- A misdetection of an OOD sample will propogate the error to motion planning and affects the total vehicle control and this might lead to unfortunate consequence

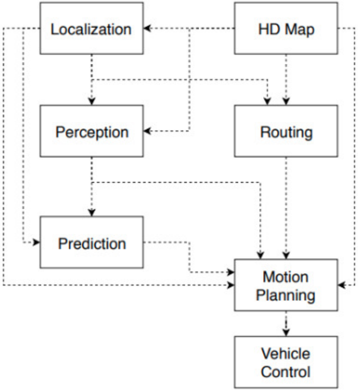


Figure 2: Caption

3D Light Detection And Ranging (LiDAR)

- Uses pulsed lasers to find the range to the objects
- Unlike images, LiDAR is insusceptible to illumination and provide rich 3D information.
- Figure 3 depicts the sample point cloud with LiDAR is placed in round white circle found at the center of point cloud
- Typically, features of each point in point cloud include
  - spatial features (XYZ)
  - Colour (RGB)
  - Intensity

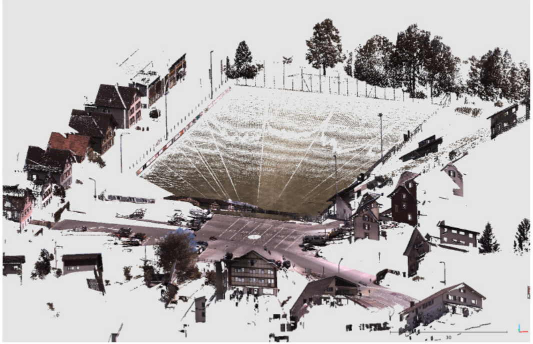


Figure 3: Caption

3D Semantic Segmentation

- An important task in computer vision because of its use in scene understanding
- Further helps in navigation and planning of robots
- Objective - Assign each point in the point cloud a specific class

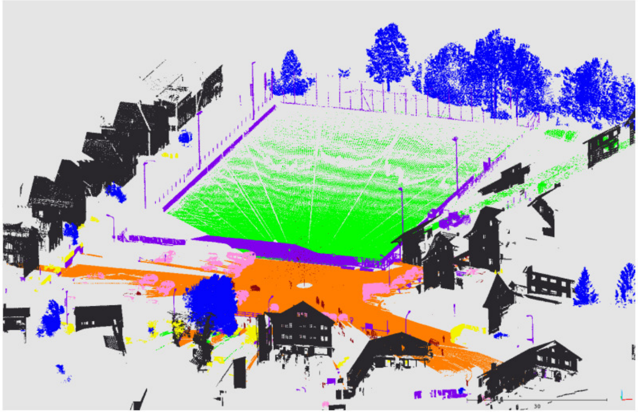


Figure 4: Caption

Thesis objective

- OOD detection in the 3D semantic segmentation setting
- Create a benchmark datasets for OOD detection among existing 3D LiDAR datasets. We define OOD data based on two categories
  - if the point is from different class than training data
  - if the point has inferior quality
- We also study whether uncertainty estimation is a practical approach for OOD detection in 3D domain

Setup

- 3D Semantic Segmentation model
- Uncertainty methods
- OOD score methods
- Datasets

RandLA-Net

- Lightweight, efficient computation, meomry usage and inputs 3D point cloud directly
- Random point sampling and local feature aggregation module are most important modules
- Local feature aggregation module is subdivided into local spatial encoding, attentive pooling and dilated residual block
- Encoder-Decoder style architecture as depicted in Figure6

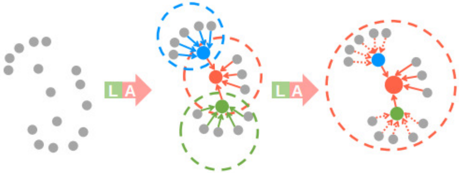


Figure 5: caption

RandLA-Net

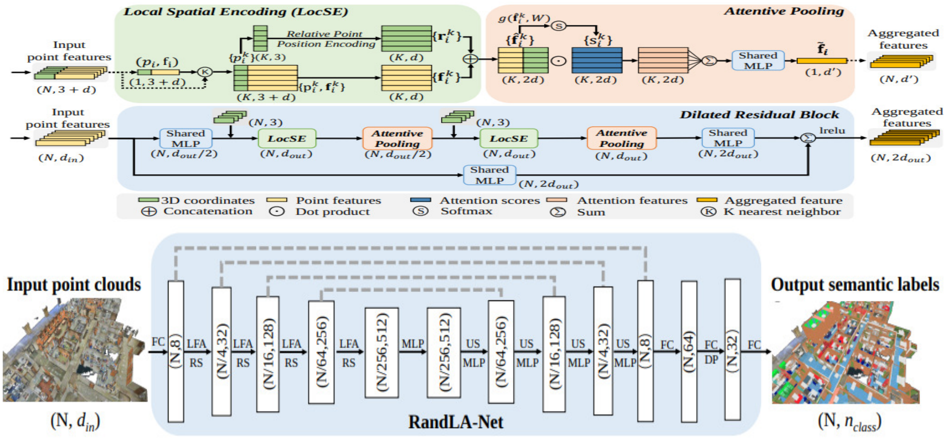


Figure 6: caption

Setup

- 3D Semantic Segmentation model - RandLA-Net
- Uncertainty methods
- OOD score methods
- Datasets

Deep Ensembles

- Ensemble learning technique - train N randomly initialized models with same data
- Reulting N predictions are then averged
- Perormance boosting along with uncertainty value for a prediction
- Requires more computation power

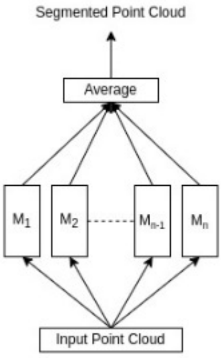


Figure 7: caption

Flipout

- Introduced as a method to decorrelate gradients in a mini batch of examples
- Add independent weight perturbations sampled from a distribution
- The output of Flipout versioned neuron is
- Train single instance of Flipout versioned network and then perform multiple forward passes for same input

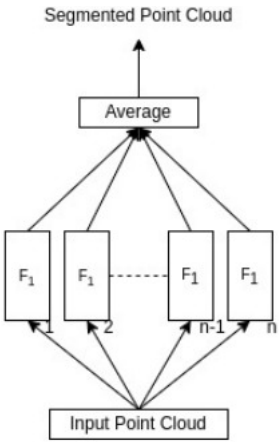


Figure 8: caption

Setup

- 3D Semantic Segmentation model - RandLA-Net
- Uncertainty methods - Deep Ensembles & Flipout
- OOD score methods
- Datasets

OOD Score calculation

- We use the following two methods to generate the OOD scores.
- Maximum Softmax Probability
  - $max(y_n), y_n = [P(C_1), P(C_2), ..., P(C_n)]$
- Entropy
  - $-\sum_i P(x_i)log(P(x_i))$ with i iterates across all the classes for point x

Setup

- 3D Semantic Segmentation model - RandLA-Net
- Uncertainty methods - Deep Ensembles & Flipout
- OOD score methods - Maximum Softmax Probability & Entropy
- Datasets

3D LiDAR datasets

acquisition mode	dataset	frames	total points (in million)	classes	scene type
static	Oakland[60]	17	1.6	44	outdoor
	Paris-lille-3D[71]	3	143	50	outdoor
	Paris-rue-Madame[74]	2	20	17	outdoor
	S3DIS[5]	5	215	12	indoor
	ScanObjectNN[85]	-	-	15	indoor
	Semantic3D[31]	30	4009	8	outdoor
	TerraMobilita/IQmulus[88]	10	12	15	outdoor
	TUM City Campus[26]	631	41	8	outdoor
	DALES[90]	40 (tiles)	492	8	outdoor
sequential	A2D2[27]	41277	1238	38	outdoor
	AIO Drive[96]	100	-	23	outdoor
	KITTI-360[100]	100K	18000	19	outdoor
	nuScenes-lidarseg[12]	40000	1400	32	outdoor
	PandaSet[99]	16000	1844	37	outdoor
	SemanticKITTI[7]	43552	4549	28	outdoor
	SemanticPOSS[62]	2988	216	14	outdoor
	Sydney Urban[19]	631	-	26	outdoor
	Toronto-3D[79]	4	78.3	8	outdoor
synthetic	SynthCity[30]	75000	367.9	9	outdoor

Table 1: caption

Semantic3D

- Huge 3D point cloud benchmark classification static dataset with 4 million points
- Scenes are taken in european streets around church, stations and fields
- Point features include XYZ, RGB and Intensity values.
- It has 8 classes with distribution of points represented in Figure9
- cite states that the scanning artefacts, hardscapes and cars are the most challenging classes

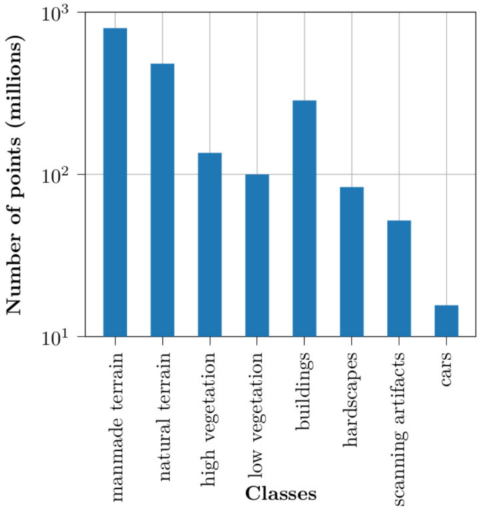


Figure 9: caption

Semantic3D

Figure 10: caption

S3DIS

- Indoor dataset with scans from various buildings
- Dataset include scans of personal offices, restrooms, open spaces, lobbies and hallways
- It has 12 classes, further subdivided into two types
  - structural elements
  - everyday items
- One of the most evaluated datasets for indoor semantic segmentation

Figure 11: caption

OOD Benchmark datasets

ID dataset	OOD dataset	OOD detection difficulty	Summary
Semantic 3D	S3DIS	Easy	1. No class overlap 2. Less structural similarity 3. Different domain(outdoor-vs-indoor)
	Semantic3D without color	Hard	1. Same structural properties 2. Difference in RGB values 3. Same domain as ID dataset 4. Same classes

Table 2: caption

Setup

- 3D Semantic Segmentation model - RandLA-Net
- Uncertainty methods - Deep Ensembles & Flipout
- OOD score methods - Maximum Softmax Probability & Entropy
- Datasets - Semantic3D, S3DIS & Semantic3D w/o colour

Experiments

- Semantic3D (ID) vs S3DIS (OOD)
  - Deep Ensembles
  - Flipout
  - Area Under Receiver Operating Characteristic (AUROC) score comparison
- Semantic3D vs Semantic3D w/o colour

Semantic3D vs S3DIS - Deep Ensembles

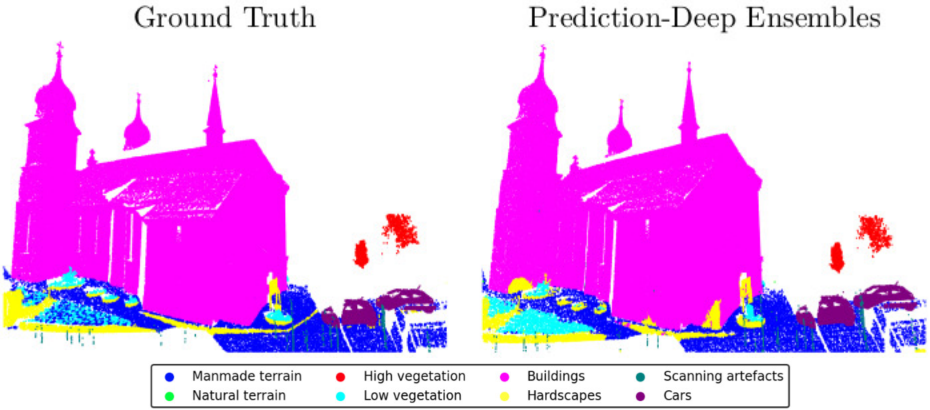


Figure 12: caption

Semantic3D vs S3DIS - Deep Ensembles

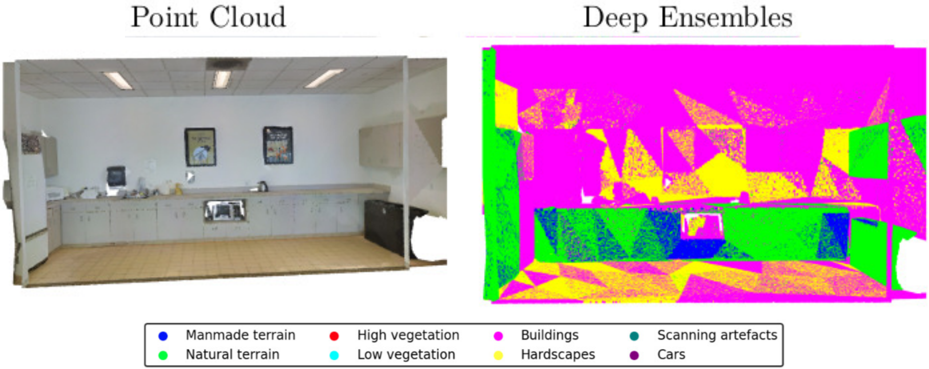
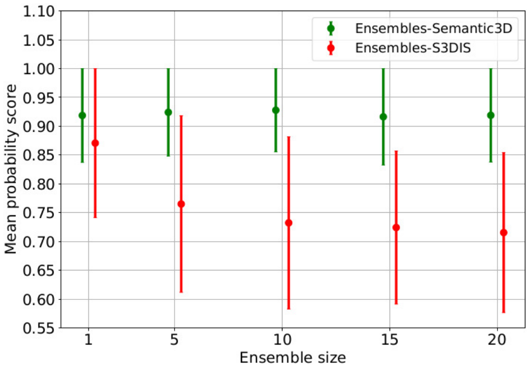
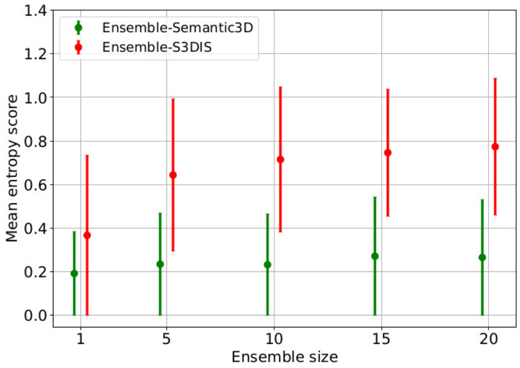


Figure 13: caption

Semantic3D vs S3DIS - Deep Ensembles





Semantic3D vs S3DIS - Flipout

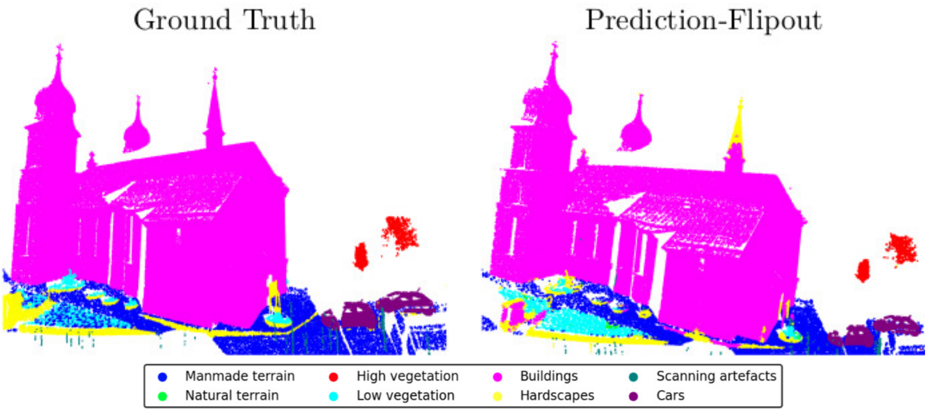


Figure 14: caption

Semantic3D vs S3DIS - Flipout

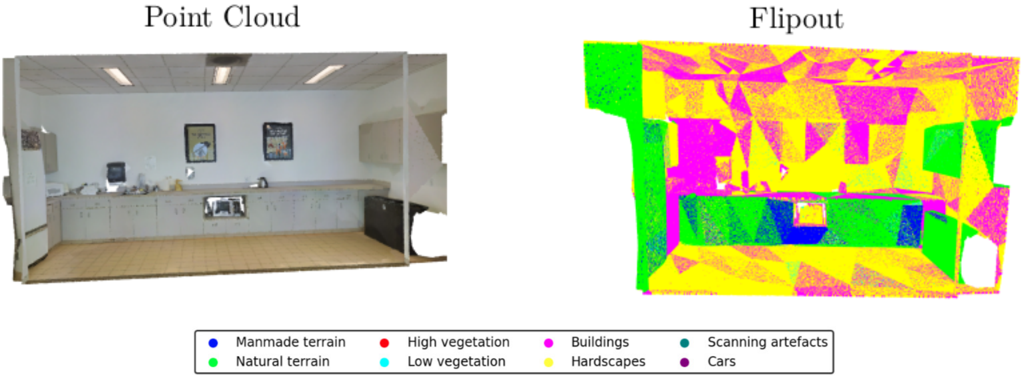
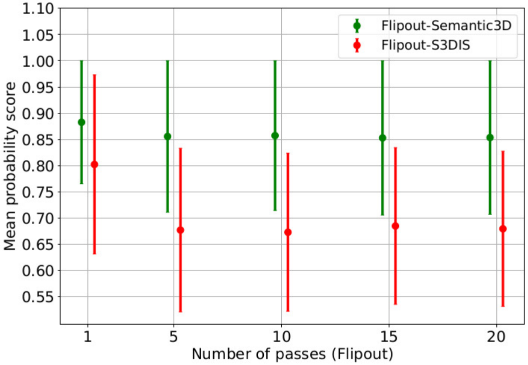
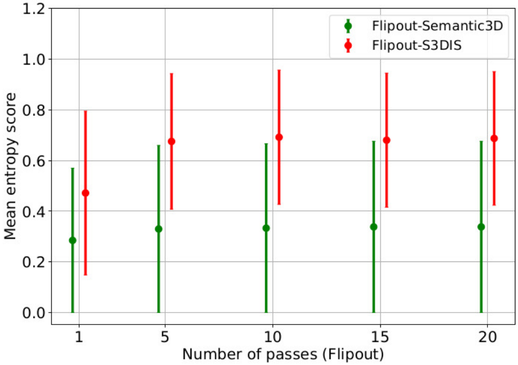


Figure 15: caption

Semantic3D vs S3DIS - Flipout





Semantic3D vs S3DIS - AUROC Scores

Ensemble size/ #passes	Method	AUROC	
		MSP	Entropy
1	Dropout	0.53311	0.53041
	Flipout	**0.69988**	**0.69368**
	Deep Ensembles	0.62020	0.62529
5	Dropout	0.58439	0.57821
	Flipout	0.77885	0.76934
	Deep Ensembles	**0.84013**	**0.83665**
10	Dropout	0.60168	0.59925
	Flipout	0.78728	0.78327
	Deep Ensembles	**0.87929**	**0.87541**
15	Dropout	0.59773	0.59557
	Flipout	0.7667	0.76741
	Deep Ensembles	**0.88486**	**0.88246**
20	Dropout	0.59766	0.59661
	Flipout	0.77331	0.77237
	Deep Ensembles	**0.89338**	**0.89052**

Table 3: caption

Semantic3D vs S3DIS - AUROC Scores

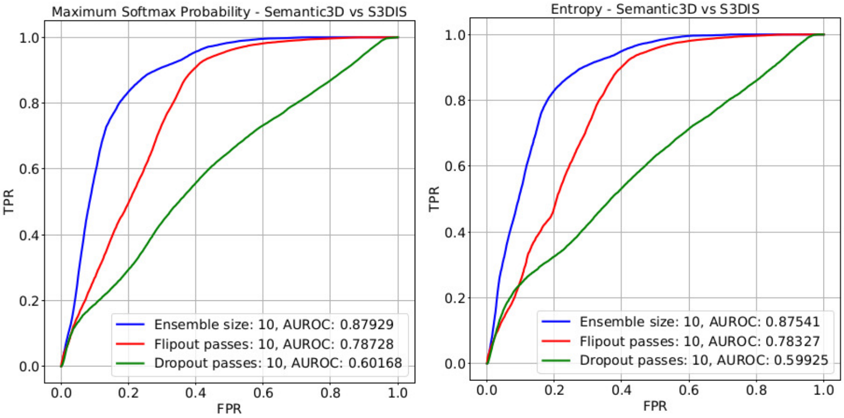


Figure 16: caption

Experiments

- Semantic3D (ID) vs S3DIS (OOD)
- Semantic3D vs Semantic3D w/o colour
  - Deep Ensembles
  - Flipout
  - Area Under Receiver Operating Characteristic (AUROC) score comparison

Semantic3D colour vs w/o colour - Deep Ensembles

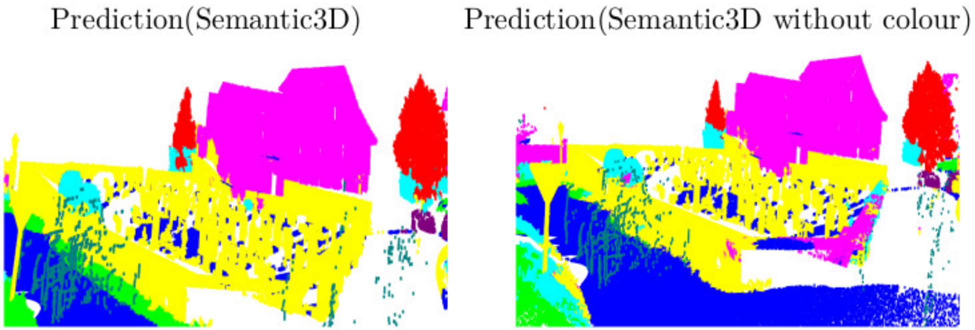
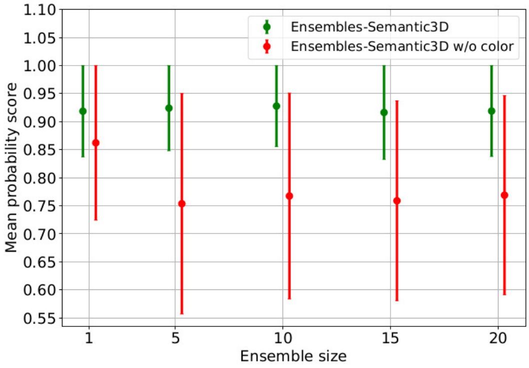
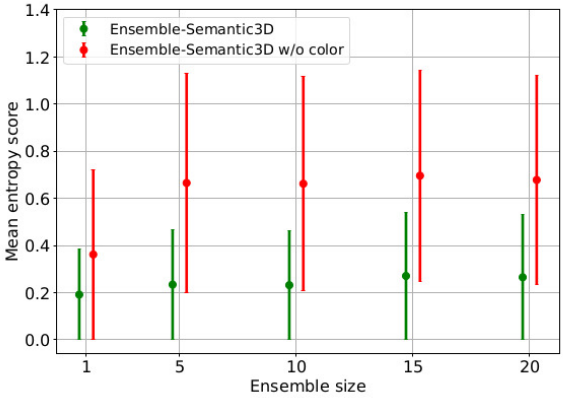


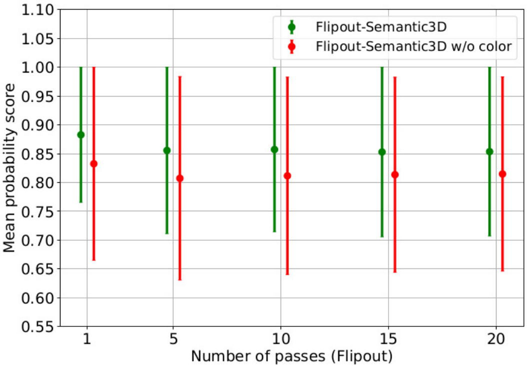
Figure 17: caption

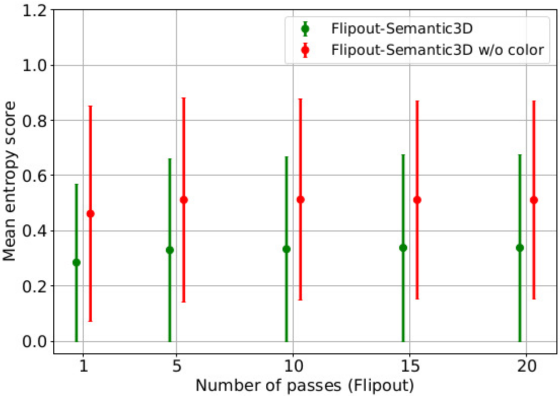
Semantic3D colour vs w/o colour - Deep Ensembles





Semantic3D colour vs w/o colour - Flipout





Semantic3D colour vs w/o colour - AUROC Scores

Ensemble size/ #passes	Method	AUROC	
		MSP	Entropy
1	Dropout	0.66349	0.65908
	Flipout	0.64221	0.66157
	Deep Ensembles	**0.67855**	**0.67866**
5	Dropout	0.69448	0.68507
	Flipout	0.63743	0.66536
	Deep Ensembles	**0.76769**	**0.77120**
10	Dropout	0.68568	0.68004
	Flipout	0.63712	0.66535
	Deep Ensembles	**0.77837**	**0.78142**
15	Dropout	0.68975	0.68347
	Flipout	0.63022	0.65976
	Deep Ensembles	**0.77302**	**0.77881**
20	Dropout	0.68447	0.68199
	Flipout	0.63017	0.65934
	Deep Ensembles	**0.77031**	**0.77584**

Table 4: caption

Semantic3D colour vs w/o colour - AUROC Scores

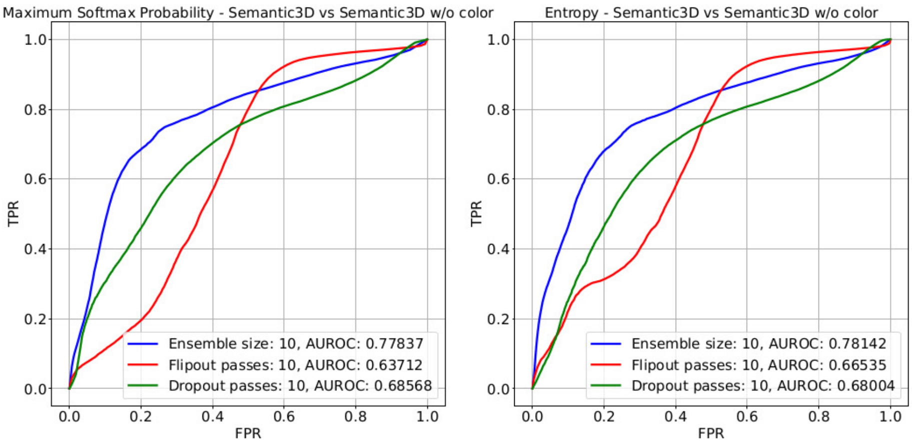


Figure 18: caption

1. Introduction

2. Experimental Setup & Methodology

3. Experiments & Results

4. Conclusion

Conclusion

- We propose two benchmark datasets
  - Semantic3D-vs-S3DIS (Outdoor-vs-Indoor) - Easy OOD identification
  - Semantic3D-vs-Semantic3D without colour - Hard OOD identification
- The second case is hard becuase of same point geometry between ID and OOD datasets
- Both Maximum Softmax Probability and Entropy are able to identify OOD points
- Deep Ensembles outperform Flipout and Dropout in both the benchmark datasets

Lessons Learned

Learning's during the duration of the thesis are

1. Training and evaluation of 3D DNNs are time consuming and resource intensive.
2. Finding the proper prior for Flipout layers is hard and currently we use brute force to find the best fitting prior.
3. OOD benchmarking require in depth analysis of datasets like studying the structural similarties in the datasets and also colour spectrum.
4. LiDAR datasets have large memory requirements especially for the preprocessing and metric computation.
5. Getting 100% OOD detection performance is not possible with the post-hoc methods used as some points in the ID dataset also have low probability scores.

Future Work

This thesis can be extended in the following ways.

1. This thesis is limited to only point based models, this can be extended to graph and projection based models.
2. The datasets involved are only static datasets and this thesis study can be further extended to other type of datasets such as synthetic and sequential datasets.
3. Since this thesis utilzes post-hoc threshold methods for OOD detection. Other methods such as Mahalanobis distance based OOD detection [1] or MetaSeg [2] can be added as an extension to this thesis.

References

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin.
A simple unified framework for detecting out-of-distribution samples and adversarial attacks.
Advances in neural information processing systems, 31, 2018.

Philipp Oberdiek, Matthias Rottmann, and Gernot A. Fink.
Detection and retrieval of out-of-distribution objects in semantic segmentation.
In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops*, pages 1331–1340. Computer Vision Foundation / IEEE, 2020.

What is OOD Detection?

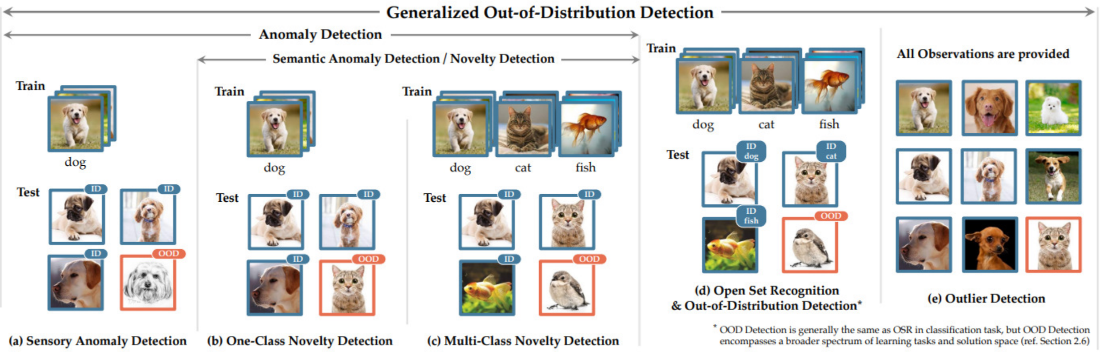


Figure 19: Generalized Out-of-Distribution Detection: A Survey