Reducing redundancy in Semantic-KITTI: Study on data augmentations within Active Learning

Ngoc Phuong Anh Duong, Alexandre Almin, Léo Lemarié, B Ravi Kiran
Machine Learning, Navya
firstname.lastname@navya.tech

Abstract

Active learning has recently gained attention in deep learning tasks dedicated to autonomous driving, such as image classification. However, semantic segmentation for point clouds remains a largely unexplored task in active learning, mainly due to the heavy computational cost of such work. In this paper, we present an analysis to reduce data redundancy in the large-scale dataset Semantic-Kitti [5], thanks to active learning uncertainty-based methods and data augmentation. We are able to demonstrate that data augmentation techniques are helping our active learning cycles, and achieve baseline accuracy with only 60% of the dataset.

1 Introduction

Autonomous driving has witnessed a recent increase in research and industry-based large-scale datasets in the point cloud domain such as Semantic-KITTI and Nuscenes. These datasets enable diverse driving scenarios and lighting conditions, along with variation in the poses of on-road obstacles. The collection procedure frequently involves recording temporal segments with key frames that are manually selected. These large-scale point clouds datasets have high redundancy, mainly due to the temporal correlation between point clouds scans, the similar urban environments and the symmetries in the driving environment (driving in opposite directions at the same location). Hence, data redundancy can be as seen missing information to improve the model due to the similarity between point clouds resulting from geometric transformations as a consequence of ego-vehicle movement along with changes in the environment. Data augmentations (DA) are transformations on the input samples that enable DNNs to learn invariances and/or equivariances to said transformations [1]. DA provides a natural way to model the geometric transformations to point clouds in large-scale datasets due to ego-motion of the vehicle.

Active Learning (AL) aims at interactively annotating unlabeled samples guided by a human expert in the loop. For large datasets, AL can be used to find a core-subset with equivalent performance w.r.t a full dataset. This involves sequentially selecting subsets of the dataset that greedily maximises model performance. AL helps distills an existing dataset to a smaller subset, thus enabling faster training times in production, while preserving high accuracy. It uses uncertainty scores obtained from predictions of a model or an ensemble to select informative new samples to be annotated by a human oracle. This paper studies the dataset distillation or reduction of redundant samples on point clouds from the Semantic-KITTI dataset. Contributions of the current study include:

1. Evaluating existing heuristic function, BALD [14] for the semantic segmentation task within a standardized AL library [3]. BALD in conjunction with DA techniques shows a high labeling efficiency on a 6000 samples subset of the Semantic-KITTI dataset.
2. Key ablation studies on informativeness of dataset samples vs data augmented samples that reflect how DA affect the quality of AL based sampling/acquisition function.
3. A competitive compression over the baseline accuracy while using only 60% of the dataset.

Bayesian Deep Learning workshop, NeurIPS 2021.

Like many previous studies on AL, we do not explicitly quantify the amount of redundancy in the datasets and purely determine the trade-off of model performance with smaller subsets w.r.t the original dataset.

1.1 Related work

The reader can find details on the major approaches to AL in the following articles: uncertainty-based approaches [10], diversity-based approaches [19], and a combination of the two [17][2]. Most of these studies were aimed at classification tasks. Adapting diversity-based frameworks usually applied to a classification, such as [19], [17], [2], to the point cloud semantic segmentation task is computationally costly. This is due to the dense output tensor from DNNs with a class probability vector per pixel, while the output for the classification task is a single class probability vector per image. Various authors in [16][12], Camvid and Cityscapes propose uncertainty-based methods for image and video segmentation. However, very few AL studies are conducted for point cloud semantic segmentation. Authors [22] evaluate uncertainty and diversity-based approaches for point cloud semantic segmentation. This study is the closest to our current work.

Authors [6] demonstrate the existence of redundancy in CIFAR-10 and ImageNet datasets, using agglomerative clustering in a semantic space to find redundant groups of samples. As shown by [8], techniques like ensemble active learning can reduce data redundancy significantly on image classification tasks. Authors [4] show that diversity-based methods are more robust compared to standalone uncertainty methods against highly redundant data. Though authors suggest that with the use of DA, there is no significant advantage of diversity over uncertainty sampling. Nevertheless, the uncertainty was not quantified in the original studied datasets, but were artificially added through sample duplication. This does not represent real word correlation between sample images or point clouds. Authors [13] uses DA techniques while adding the consistency loss within a semi-supervised learning setup for image classification task.

2 Method

In this section, we will describe our setup used to evaluate the performances of active learning for point clouds semantic segmentation, including details on the dataset and model used, the chosen data augmentations techniques, and the most important, details on our active learning experiments.

2.1 Dataset

Although there are many open-source datasets for image semantic segmentation, not many of them are dedicated to semantic segmentation on point clouds. The Semantic-KITTI dataset & benchmark [5] provides more than 43000 point clouds of 22 annotated sequences, acquired with a Velodyne HDL-64 LiDAR. Semantic-KITTI is by far the most extensive dataset with sequential information. All available annotated point clouds, from sequences 00 to 10, for a total of 23201 point clouds, are later used for our experiments.

2.2 Model

Among different deep learning models available, we choose SqueezeSegV2 [21], a spherical-projection-based semantic segmentation model, which performs well with a fast inference speed compared to other architectures, thus reduces training and uncertainty computation time. We apply spherical projection [21] on point clouds to obtain a 2D range image as an input for the network shown in figure 1. To simulate Monte Carlo (MC) sampling for uncertainty estimation [9], a 2D Dropout layer is added right before the last convolutional layer of SqueezeSegV2 with a probability of 0.2 and turned on at test time.

2.3 Spherical projection: Converting pointcloud to range images

Rangenet++ architectures by authors [18] use range image based spherical coordinate representations of point clouds to enable the use of 2D-convolution kernels. The relationship between range image and LiDAR coordinates is the following:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[1 - \arctan(y, x)\pi^{-1}] \times w \\ [1 - (\arcsin(z \times r^{-1}) + f_{up}) \times f^{-1}] \times h \end{pmatrix},$$

where (u, v) are image coordinates, (h, w) the height and width of the desired range image, $f = f_{up} + f_{down}$, is the vertical fov of the sensor, and $r = \sqrt{x^2 + y^2 + z^2}$, range measurement of each point. The input to the DNNs used in our study are images of size $W \times H \times 4$, with spatial dimensions W, H determined by the FOV and angular resolution, and 4 channels containing the x, y coordinates of points, r range or depth to each point, i intensity or remission value for each point.

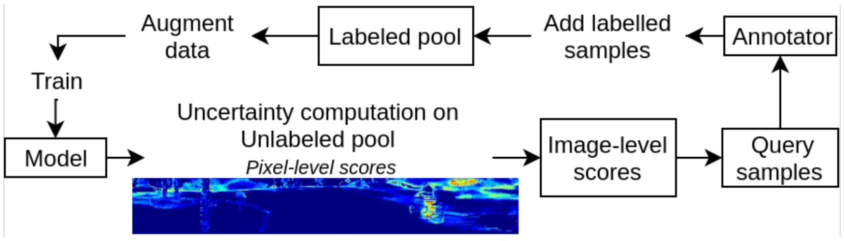


Figure 1: Global flow of active learning on range images from point clouds using uncertainty methods.

2.4 Bayesian uncertainty-based approach of active learning

In a supervised learning setup, given a dataset $\mathcal{D} := \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$, the DNN is seen as a high dimensional function $f_\omega : \mathcal{X} \to \mathcal{Y}$ with model parameters ω. A simple classifier maps each input x to outcomes y. A good classifier minimizes the empirical risk $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, which is defined with the expectation $R_{\text{emp}}(f) := \mathbb{P}_{X,Y}[Y \neq f(X)]$. The optimal classifier is one that minimizes the above risk. Thus, the classifier's loss does not explicitly refer to sample-wise uncertainty but rather to obtain a function which makes good predictions on average.

Predictive uncertainty [15] estimates uncertainty over each prediction $\hat{y} = f_\omega(\mathbf{x}) = p(y|\mathbf{x})$ given its input $\mathbf{x}$. A model's predictive uncertainty is a combination of the *aleatoric uncertainty*, irreducible uncertainty due to intrinsic randomness of underlying process, and the *epistemic uncertainty*, reducible uncertainty caused due to missing knowledge, and could be reduced given additional information.

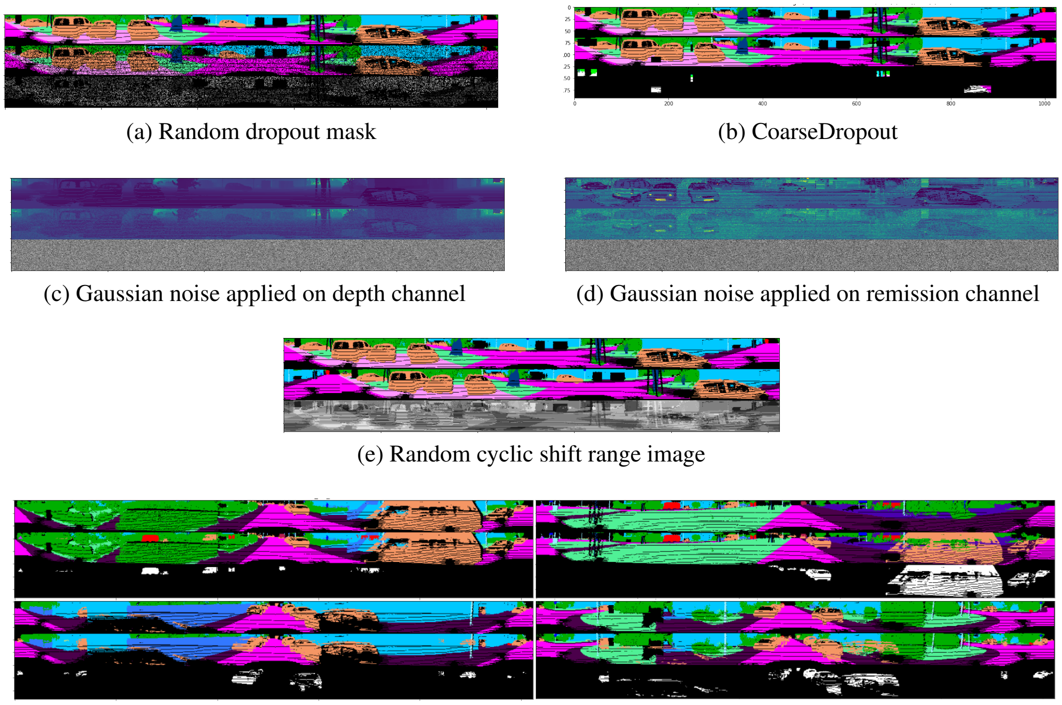
Authors [9] propose generation of MC samples for a given model and input, by activating standard dropout layers at inference time. This provides an uncertainty estimation by sampling different values of DNN weights. Readers can consult work by [11] for uncertainty estimation in DNNs.

We summarize here the key components of the bayesian AL framework:

1. *Labeled dataset* $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ where $\mathbf{x}_i \in W \times H \times 4$ are range images with 4 input channels, W, H are spatial dimensions, and $y_i \in W \times H \times C$ are one-hot encoded ground truth with C classes. The output of the DNN model is distinguished from the ground truth as $\hat{y}_i$ with the same dimensions.
2. Partition of the dataset into *Labeled pool* $L \subset D$ and a unlabeled pool $U \subset D$ considered as a data with/without any ground-truth, where at any AL-step $L \cup U = D$, the subsets are disjoint and restore the full dataset.
3. Query size B, also called a *budget*, to fix the number of unlabeled samples selected for labeling
4. Acquisition function, known as heuristic, providing a score for each pixel given the output $\hat{y}_i$ of the DNN model, $f : \mathbb{R}^{W \times H \times C} \to \mathbb{R}^{W \times H}$
5. Including the usage of MC iterations the output of the DNN model could provide several outputs given the same model and input, $\hat{y}_i \in W \times H \times C \times T$ where T refers to the number of MC iterations.
6. *Subset model* f_L is the model trained on labeled subset L
7. *Aggregation function* $a : \mathbb{R}^{W \times H \times C \times T} \to \mathbb{R}^{+}$ is a function that aggregates heuristic scores across all pixels in the input image into a positive scalar value, which is used to rank samples in the unlabeled pool.

2.5 Data augmentation setup

We apply DA directly on the range image projection. We selected known effective transformations: (a) *Random dropout mask* on range image and its target by creating a binary mask with uniform dropout probability $p \in [0.1, 0.5]$; (b) *CoarseDropout* which randomly masks out rectangular regions by applying with the following parameters: max_height: 16, max_holes: 5, max_width: 64, min_height: 1, min_holes: 2, min_width: 1, from the Albumentations library [7]; (c) *Gaussian noise* on depth of range image with the following parameters $\mu = 0, \sigma^2 \in [0.05, 0.1]$; (d) *Gaussian noise* on remission channel of range image with the following parameters: $\mu = 0, \sigma^2 \in [0.5, 1.0]$; (e) *Random cyclic shift* on range image (corresponding to rotations on point cloud) and its target to left and right, from 0 to 22.5 degrees around the center; (f) *Instance Cut Paste* randomly copying and pasting instances from one scan to another within a batch. More description and experiment setup of these transformations are in figure 2.



(a) Random dropout mask

(b) CoarseDropout

(c) Gaussian noise applied on depth channel

(d) Gaussian noise applied on remission channel

(e) Random cyclic shift range image

(f) Instance Cut Paste

Figure 2: Original, Augmented and error images on Semantic-KITTI, from top to bottom. Transformations a, b, c, d are directly based on Albumentations library [7]

2.6 Heuristic functions

Heuristic functions are transformations over the model output probabilities $p(y|x)$ that define uncertainty-based metrics to rank and select informative examples from the unlabeled pool at each AL-step. We used the following uncertainty-based metrics in our experiments:

1. *Certainty* heuristic measures the least confident class probability across the highest confident prediction over T number of MC iterations:

$$\text{min}_y \text{max}_i \{f_\omega(\mathbf{x})\}_{i=1}^T$$

2. *Entropy* heuristic measures the entropy over predicted class probabilities

$$H(y|x, L) = -\sum_c^m p(y = c|x, L) log(p(y = c|x, L))$$

3. *Variance* computes the variance of predictions from model parameters for each class, then averages all variances from all classes to obtain the aggregated score for a sample in classification, or a pixel in image semantic segmentation. The heuristic selects the samples having the highest aggregated scores. The variance for each class over T number of MC iterations:

$$\sigma^2(p(y=c|x,L)) = \frac{1}{T}\sum_{i=1}^{T}(p(y=c|x,w_i|L) - p(y=c|x,L))^2$$

4. *BALD* [14] selects samples maximizing information gain between the predictions from model parameters, using MC Iterations. The expectation in the equation below is performed over model parameters ω. The information gain $I(y,\omega|x,L)$ is given by

$$H(y|x,L) - E_{p(\omega|L)}(H(y|x,\omega)) \tag{1}$$

Table 1: Common experiments settings to each AL run.

Data related parameters			AL Hyper parameters				
Range image resolution	**Total pool size**	**Test pool size**	**Init set size**	**Budget**	**MC Dropout**	**AL steps**	**Aggregation**
1024x64	6000	2000	240	240	0.2	25	sum
Hyper parameters for each AL step							
Max train iterations	**Learning rate (LR)**	**LR decay**	**Weight decay**	**Batch size**	**Early stopping**		
					Evaluation period	**Metric**	**Patience**
100000	0.01	0.99	0.0001	16	500	train mIoU	15

2.7 Experimental Setup

The pipeline (Figure 1) follows a Bayesian Active Learning (AL) using Monte Carlo Dropout (subsection on 2.4) for point clouds range image. The uncertainty-based acquisition function, also called heuristic, computes uncertainty scores for each pixel and uses *sum* as an aggregation method to combine all pixel-wise scores of an image into a single score. The samples query step selects a fixed amount of samples, called the budget or query size, following a ranking given by the heuristic function.

Based on this pipeline, we made active learning runs with different heuristics (random, BALD [14], entropy and certainty, which are fully described in subsection 2.6) and tested the effect of DA at training time. As mentioned in Table 1, we only use 6000 randomly chosen samples from Semantic-KITTI over the 23201 samples available, because every experiment is very time-consuming, and our resources were limited. At each training step, we reset model weights to avoid biases in the predictions, as proven by [4].

In order to evaluate the performances of our pipeline over each experiment, on test set we use labeling efficiency and mean intersection over union (mIoU) as our metrics, which are fully described in subsection 2.8. Finally, to speed up the training steps, we use early stopping based on training mIoU stability over *patience* $*$ *evaluation_period* iterations.

2.8 Evaluation metrics

MeanIoU Intersection over Union (IoU) [20], known as Jaccard index, measures the number of common pixels between the target and prediction masks over the total number of pixels. MeanIoU (mIoU) is mean value of IoU over all classes. Given TP_c, FP_c, and FN_c as the number of true positive, false positive, and false negative predictions for class c, and C is the number of classes, MeanIoU can be formulated as

$$\text{MeanIoU} = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FP_c + FN_c} \quad \text{LE} = \frac{n_{\text{labeled\_baseline}}(\text{MeanIoU} = a)}{n_{\text{labeled\_others}}(\text{MeanIoU} = a)} \tag{2}$$

Labeling efficiency (LE) is used by [4] compare the amount of data needed among different sampling techniques with respect to a baseline. In our experiments, instead of accuracy, we use MeanIoU as the performance metric. Given a specific value of MeanIoU, the labeling efficiency is the ratio between the number of labeled samples, range images, acquired by the baseline sampling and the other sampling techniques.

3 Experiments & Analysis

Based on previously described AL configurations, we investigate which heuristic performs the best on semantic segmentation for point clouds (A), but also the benefit and the impact of DA techniques on dataset compression (B) and sample selection in AL steps (C), and finally the model's stability and efficiency for sample selection (D).

A. Evaluating heuristic function First we had to explore the performances of each heuristic over random, which can be seen as the most basic acquisition method. Every AL run can achieve the goal performance using fewer number of labeled samples (Figure 3). The most efficient heuristic is BALD as it outperforms the other heuristics, allowing the model to converge faster with the highest labeling efficiency ratio. Based on these results and because other heuristics are showing the same patterns in our researches, we choose to focus on BALD and random for the rest of the experiments.

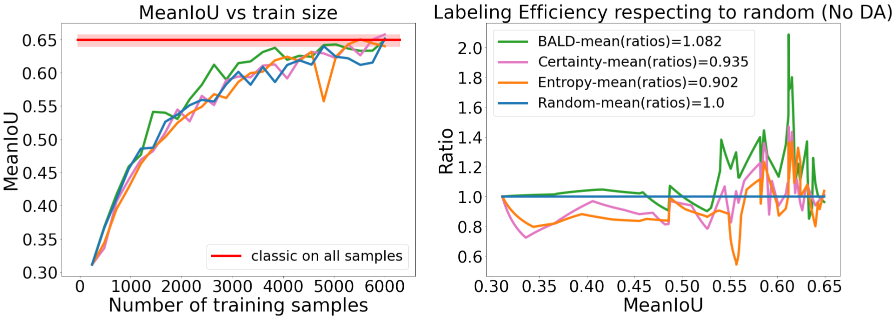


Figure 3: MeanIoU vs number of training samples and labeling efficiency evaluated on test set. Using 100% of available samples at the end of each run allows us to define an average top performance.

B. Effect of Data augmentation In this experiment, DA techniques are applied at training time. On both random and BALD heuristics, figure 4 shows that DA helps the model to reach the baseline accuracy on test set faster compared to runs without DA. As DA improves model generalisation, the elimination of similar examples learnt by invariance during sampling makes the selection pickier. In other words, with DA, the model tends to select samples different from the trained samples and their transformations, and so reduce redundancy. BALD with DA can achieve an important dataset compression, by using only 60% of the total sample pool and still achieving baseline accuracy.

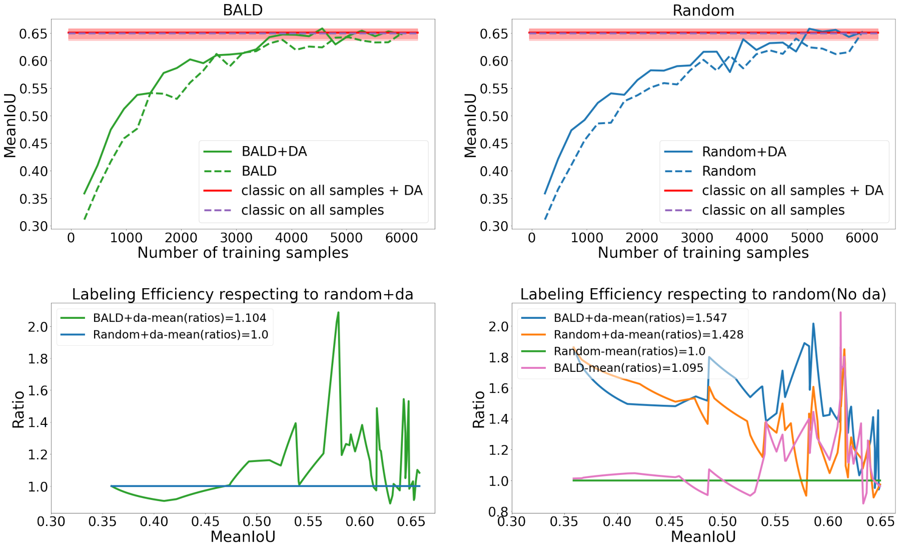


Figure 4: MeanIoU vs number of training samples and labeling efficiency evaluated on test set.

C. Uncertainty study on sample selection with data augmentation In an effort to understand how data augmented samples affect the heuristic function we evaluated the heuristic function using models trained without DA while predicting on test time augmented images. We evaluated the aggregated heuristic scores for BALD over firstly the labeled and unlabeled pools, secondly we use Test-Time Data Augmentations (TT-DA) on both labeled and unlabeled pool samples (Figure 5) at different AL steps. To be clear, we used models with no DA during training for this experiment. *(TT-DA(L))* is generated by applying DA at test time on the Labeled pool *(L)* at each training epoch. *(TT-DA(U))* contains augmented samples from the Unlabeled pool *(U)*. We ensure that the combined sizes of *(TT-DA(U))* and *(U)* is always equal to 6000 samples.

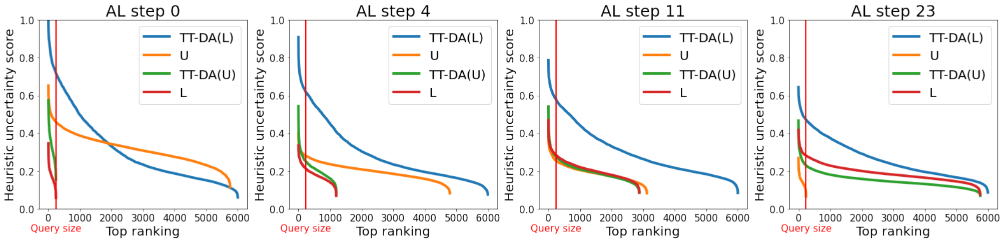


Figure 5: Heuristic score of samples sorted by decreasing value, to simulate the uncertainty scores of the samples that would have been selected at each AL step

In figure 5 the sorted aggregated scores a to the left of the red line which defines the budget of each AL-step, we notice the following ordering : These results show that in the early AL steps:

$$a(TTDA(L)) > a((U)) > a(TTDA(U)) > a((L))$$

and in final AL steps:

$$a(TTDA(L)) > a(TTDA(U)) > a((L)) > a((U))$$

We observe that during the initial AL step, the heuristic uncertainty is low as expected on *(L)*, which has been used to train the model. Because the model has been trained on only 240 samples from *(L)*, the uncertainty score is very high on *(U)*, *(TT-DA(U))* and *(TT-DA(L))*. As the AL steps goes on, the uncertainty score is globally decreasing, this can be explained by the growing pool of selected data *(L)* used to train the model which reduces the uncertainty score based on the model prediction. For one of the final AL step, *(U)* has the smallest uncertainty scores as the model is now well trained and able to correctly generalize on unseen samples. The highest uncertainty scores are related to data augmented samples from the labeled *(TT-DA(L))* and unlabeled *(TT-DA(U))* pool. This could be because the DA is providing transformed samples that are now outside the support of the dataset distribution.

D. Model stability and effectiveness for sample selection In this part, we study the model stability, based on the mean variance computed on class probabilities across all MC iterations. We also measure the model sampling effectiveness by computing mean BALD metric. Across all AL steps (Figure 6), models with DA become certain on their predictions (impacted by dropout), and are able to select sample that maximise information gain sooner across all MC iterations than models without any DA. This experiment shows that DA enhances the stability of models and allows a better and faster sample selection by reducing the uncertainty over heuristic functions.

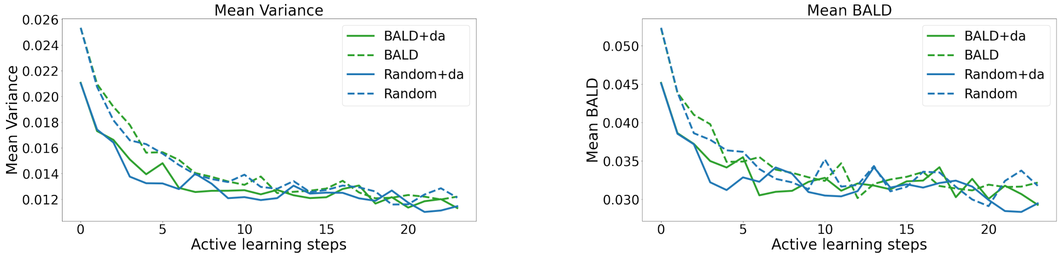


Figure 6: Mean of variances computed on class probabilities over MC predictions and mean of BALD for all pixel across all classes in the test set (subsection 1).

4 Conclusion

Our work demonstrates the benefits of data augmentation in active learning for point clouds semantic segmentation task. It confirms the conclusion of [4] made on image classification tasks, that BALD combined with data augmentation techniques provides a robust and label efficient heuristic for sample selection. It not only select more uncertain samples at each active learning step, but also increase the heuristic's stability. With only 60% of the samples, we reach the same accuracy as a supervised training with the full selected subset. Data augmented samples reduce heuristic scores over redundant samples when parametrized well, and enable us to compress the dataset.

Acknowledgements This work was granted access to the HPC resources of [TGCC/CINES/IDRIS] under the allocation 2021- [AD011012836] made by GENCI (Grand Equipement National de Calcul Intensif). This publication is also part of the Deep Learning Segmentation (DLS) project financed by ADEME (Agence de la transition écologique).

References

[1] Fabio Anselmi, Lorenzo Rosasco, and Tomaso Poggio. On invariance and selectivity in representation learning. *Information and Inference: A Journal of the IMA*, 5(2):134–158, 2016. 1

[2] Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds, 2020. 2

[3] Parmida Atighehchian, Frédéric Branchaud-Charron, and Alexandre Lacoste. Bayesian active learning for production, a systematic study and a reusable library, 2020. 1

[4] Nathan Beck, Durga Sivasubramanian, Apurva Dani, Ganesh Ramakrishnan, and Rishabh Iyer. Effective evaluation of deep active learning on image classification tasks, 2021. 2, 5, 8

[5] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Juergen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences, 2019. 1, 2

[6] Vighnesh Birodkar, Hossein Mobahi, and Samy Bengio. Semantic redundancies in image-classification datasets: The 10don't need, 2019. 2

[7] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020. 4

[8] Kashyap Chitta, Jose M Alvarez, Elmar Haussmann, and Clement Farabet. Training data subset search with ensemble active learning. *arXiv preprint arXiv:1905.12737*, 2019. 2

[9] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2016. 2, 3

[10] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017. 2

[11] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A survey of uncertainty in deep neural networks, 2021. 3

[12] S. Alireza Golestaneh and Kris M. Kitani. Importance of self-consistency in active learning for semantic segmentation, 2020. 2

[13] SeulGi Hong, Heonjin Ha, Junmo Kim, and Min-Kook Choi. Deep active learning with augmentation-based consistency estimation. *arXiv preprint arXiv:2011.02666*, 2020. 2

[14] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning, 2011. 1, 5

[15] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 2021. 3

[16] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision?, 2017. 2

[17] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning, 2019. 2

[18] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220. IEEE, 2019. 2

[19] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach, 2018. 2

[20] Shuran Song, Fisher Yu, Andy Zeng, Angel X. Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image, 2016. 5

[21] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4376–4382. IEEE, 2019. 2

[22] Tsung-Han Wu, Yueh-Cheng Liu, Yu-Kai Huang, Hsin-Ying Lee, Hung-Ting Su, Ping-Chia Huang, and Winston H. Hsu. Redal: Region-based and diversity-aware active learning for point cloud semantic segmentation, 2021. 2