RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ADEME
AGENCE DE LA TRANSITION ÉCOLOGIQUE

GENCI
Le calcul intensif au service de la connaissance

navya
be fluid

Real world data augmentations for autonomous driving

NAVYA INTRODUCTION

PASSENGERS TRANSPORT

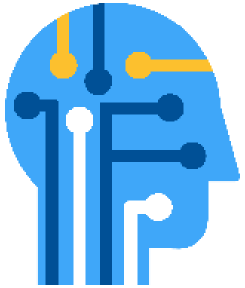
Autonom® Shuttle

Autonom® Shuttle Evo

GOODS TRANSPORT

Autonom® Tract AT135

CUSTOM SELF-DRIVING SOLUTION



DRIVEN BY NAVYA

Ambition level 4 for all our platforms

NAVYA ML TEAM

Deep learning modules on camera and LiDAR

ML team at Navya principally works on:

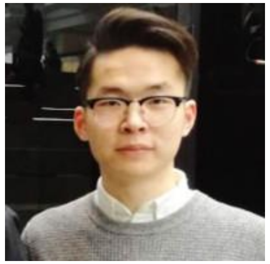
Camera :
- 2D Object detection and drivable zone segm. (2D-OD, MTL)
- Traffic light detection and relevancy (TLDR)
- 3D Monocular object detection (3D-MOD)

LiDAR :
- Large scale semantic segmentation on pointclouds
- Instance segmentation on pointclouds

Semantic Navya Dataset


Alexandre Almin


Thomas Gauthier


Hao Liu


B Ravi Kiran


Leo Lemarie


Anh Duong

OVERVIEW

- What is a data augmentation (DA)
  - Categories of data augmentation (classification, detection, segmentation)
  - Theoretical frameworks for DA
- Geometry preserving 2D-DA for 3D monocular detection
  - DA for 3D monocular object detection
  - Self supervision pretext tasks for 3D monocular detection
  - Evaluation on KITTI 3D detection dataset
- DA for data redundancy in Active learning (AL) pipeline
  - Semantic segmentation on pointclouds
  - Building an AL pipeline for mining informative samples
  - Evaluation on Semantic-KITTI dataset

DATA AUGMENTATION

A Brief review

https://albumentations.ai/docs/ ,
https://imgaug.readthedocs.io/en/latest/

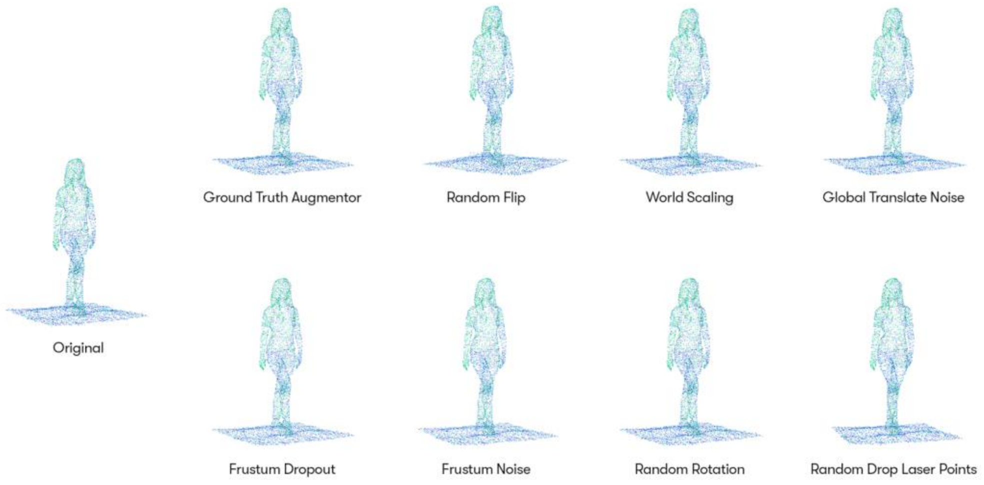
Generates augmented I/O pairs

- Performs model **regularization** & reduce the effect of overfitting in low dataset regime
- Increases the diversity of small datasets
- Fundamentally models invariance/equivariance to **real world transformations** that generate samples of a dataset

Image vs pointcloud augmentations

- Representations:
  - Images conserve an inherent matrix representation
  - Pointclouds are sets with arbitrary input domain size
- Instance transform:
  - Objects in pointclouds are **separable** from their background, enabling for easy 3D transformations (rotation translation).
  - Though this does require the transformation to account for change in pointcloud density with change in distance/orientation

https://blog.waymo.com/2020/04/using-automated-data-augmentation-to.html



WHY DATA AUGMENTATIONS WORK

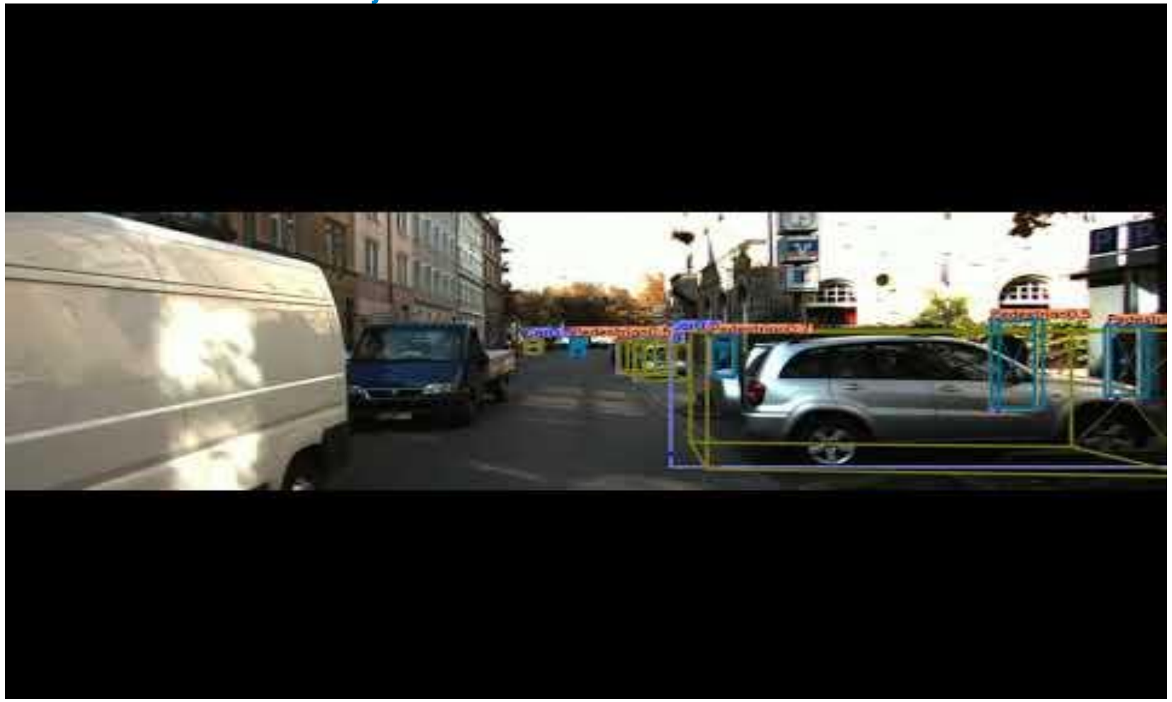
Reviewing theoretical understanding

- DA model invariances to represent data Anselmi, et al. 2016
  - Translation invariance already baked into CNNs due to convolutions
  - Rotations, Scaling, color transformations…
- DA connection with kernel theory **Dao, Tri, et al 2019**
  - DA augmentations are seen as a Markov process with transitions defined per sample
- DA are represented within a group G Jane H. Lee et al 2020
  - Augmented sample distribution gX are ***approximately invariant*** under the action of group elements g
  - X ~= gX, g from G
  - The probability of an augmentation sample is approximately equal to the original sample

- Anselmi, et al. "On invariance and selectivity in representation learning." Information and Inference: A Journal of the IMA 2016.
- Dao, Tri, et al. "A kernel theory of modern data augmentation." *International Conference on Machine Learning*. PMLR, 2019.
- Chen, Shuxiao, Edgar Dobriban, and Jane H. Lee. "A group-theoretic framework for data augmentation." *JMLR* 21.245 (2020): 1-71.

CASE STUDY 1 : 3D MONOCULAR OBJECT DETECTION(3D-MOD)

Exploring 2D Data Augmentation(DA) for 3D Monocular Object Detection



3D-MOD is a key component of obstacle detection pipeline

- Redundancy & Fusion with LiDAR 3D detection pipeline
- Stereo depth estimation pipelines are progressively replaced with monocular depth estimation

Motivation : DA for 3D-MOD

- Datasets for 3D-MOD are costly to create
- Data augmentations on 2D object detector's change image geometry
- View synthesis methods are robust, but heavy
- How to reuse existing annotations to be a self-supervised task ?

Problem formulation : Data augmentation
What transformations or augmentations could be performed to (image, 2D-BB) pair that do not change the depth, orientation or scale of the bounding box ?

Problem formulation : Pretext SSL task
What scalable auxilliary task along with a methodology to generate annotation could be added to the 3D-MOD detector primary task ?

Joint work: Sugirtha T, Sridevi M, Khailash Santhakumar, Hao Liu B Ravi Kiran, Thomas Gauthier, Senthil Yogamani Accepted at ICCV Workshop on Self-supervised learning for Autonomouas Driving (SSLAD 2021)

DATA AUGMENTATIONS FOR OBJECT DETECTION

Label independent image transforms

- Affine
- Color Tx
- Cutmix
- Blur
- Cutout
- Mixup

Weather based

- Shadow
- Rain , Snow
- Fog, Flare
- Gravel
- Autumn

Bounding box Instance based

2D boxes

- Mosaic
- Blur, Cutout
- Mask Cutpaste
- Flip
- Affine
- **Apply X to box**

Monocular 3D boxes

Geometry preserving

- **Mosaic-Tile**
- **Box Mixup**
- **Box Cutpaste**

Geometry Aware*

- Camera Postion
- Focal Length
- Receptive Field

View synthesis

- Differentiable Neural Rendering

Self-Supervised & Semi-supervised augmentations

- Random window classifier
- Jigsaw
- Rotation prediction
- Contrastive Losses
- ----

Learning based

- RandAugment
- FastAugment
- RL-Search
- GAN based
- Style Transfer
- Style Transfer
- Augmix
- Adeversarial

***Closest work** : Lian, Qing, et al. "Geometry-aware data augmentation for monocular 3D object detection." arXiv preprint arXiv:2104.05858 (2021).

GEOMETRY PRESERVING 2D DATA AUGMENTATIONS

For monocular 3d object detection

New data augmentations : Box-Mixup, Box-Cutpaste and Mosaic Tile

Geometry preserving augmentation: These transformations do not change the camera viewpoint or the 3D orientation of the objects in the scene

nauya

SELF-SUPERVISED LEARNING

For monocular 3d object detection

- Self-supervised learning aims at adding auxiliary/pretext tasks
  - Where the labels are either automatically generated either by another sensors (LiDAR) or are correlated task
  - The pretext task is correlated with the primary task and thus training on the pretext task provides better performance on the primary task

- Multi-object labeling (MOL)
  - Established pretext tasks for 2D object detection
  - Generate random windows covering existing foreground bounding boxes
  - Create soft label showing the proportion of areas of different classes in the random window

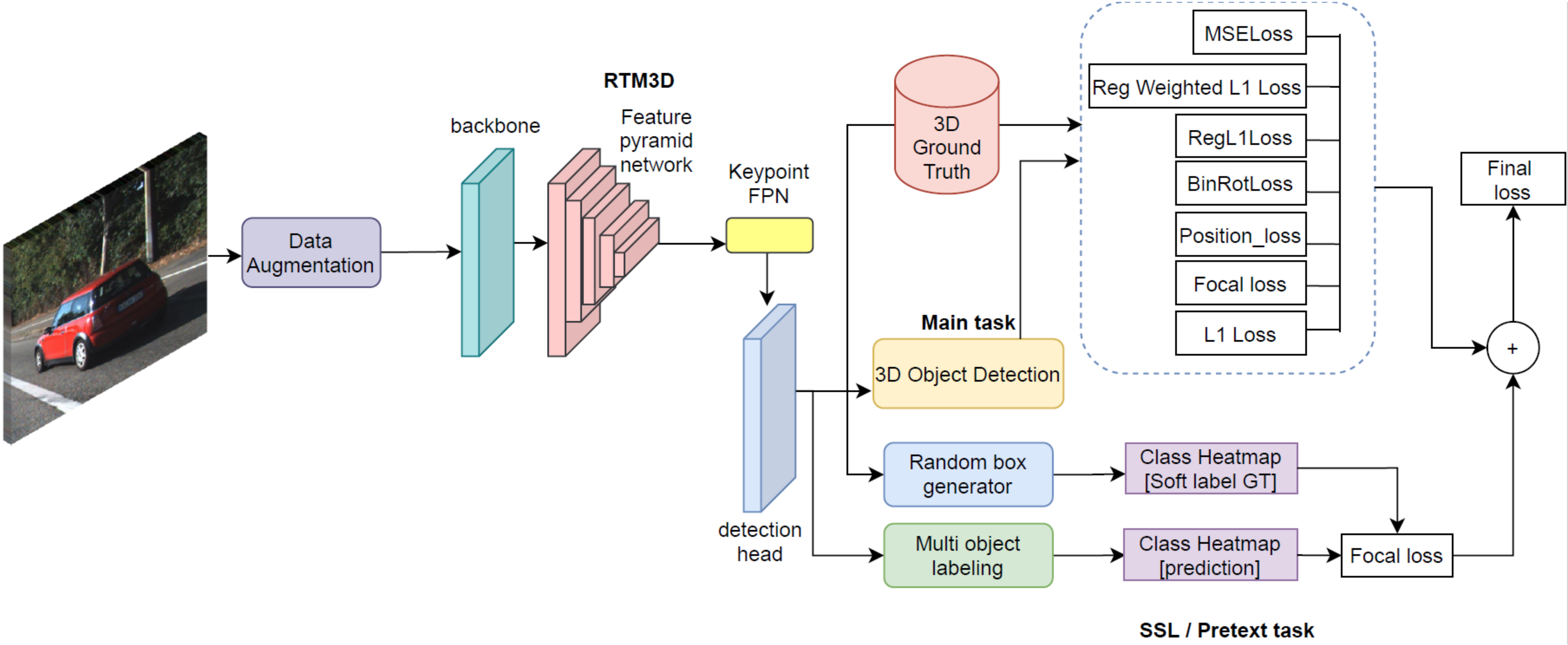
Soft-label over random window

Implentation from :
[Lee, Wonhee, Joonil Na, and Gunhee Kim. "Multi-task self-supervised object detection via recycling of bounding box annotations." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.]

nauya

SELF-SUPERVISED LEARNING

For monocular 3d object detection



Li, Peixuan, et al. "Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer International Publishing, 2020.

EVALUATION METRICS

Evaluating 3D object detection

Class	Car	Pedestrian	Cyclist
Frequency f_c	0.82	0.12	0.05
Inverted w_c	0.04	0.27	0.69

Normalized Class Frequency on validation set of KITTI 3D

- We evaluate performance using the mean Average Precision (mAP)
- We propose a new weighted ICFW mAP using the Inverse of the Class Frequency Weights to evaluate gains in non majority classes (especially in KITTI)
- We use the KITTI 3D object detection metrics
  - Average Precision (AP) per class
  - mAP 2D and ICFW mAP 2D
  - mAP 3D and ICFW mAP 3D
  - mAP BEV and ICFW mAP BEV
  - mAP AOS* and ICFW mAP AOS

$$\boxed{\mathrm{mAP}_{3D} = \frac{1}{|C|} \sum_{c \in C} \mathrm{AP}_c}$$

$$C = \{\text{car, pedestrian, cyclist}\}$$

$$\boxed{\text{ICFW mAP}_{3D} = \sum_{c \in C} w_c \mathrm{AP}_c}$$

New proposed metric

$$w_c := \frac{f_c^{-1}}{\sum_{c \in C} f_c^{-1}} \in [0,1] \quad \text{and} \quad \sum_{c \in C} w_c = 1$$

* AOS : Average orientation similarity

BEV mAP: Bird Eye View 2d Box Map

RESULTS : SELF SUPERVISED LEARNING WITH DA

IoU=0.5	mAP2D	mAPBEV	mAP3D	ICFW mAP2D	ICFW mAPBEV	ICFW mAP3D
Baseline (B)	41.44	21.17	19.12	33	15.1	14.65
Self-Supervised Learning (SSL) with MOL						
B + 8W	0.85	0.53	0.46	0.83	0.7	0.54
B + 16W	0.59	-0.75	-0.59	0.57	-1.88	-1.73
B + 32W	1.4	0.29	0.12	1.75	0.12	-0.17
Data Augmentation (DA)						
B + Cutout4	-0.91	0.11	-0.71	-2.79	0.15	-0.54
B + BoxMixup	0.39	0.29	0.21	0.53	0.12	0.04
B + Cutpaste	1.63	1.10	0.34	3.22	1.91	0.49
B + Mosaic	-2.61	-1.43	-0.26	-2.96	-0.17	0.09
SSL-MOL + DA						
B + 16W + Cutout	1.54	1.27	0.43	2.17	**2.81**	1.02
B + 16 W + box mixup	1.2	1.67	1.66	1.42	2.57	**2.59**
B + 16 W + boxmixup cutout	**3.51**	**1.84**	1.01	**5.57**	2.53	1.02
B +16 W + cutpaste cutout	2.87	1.38	**2.26**	5	1.13	1.19
B +16 W + cutpaste	0.98	0.67	0.72	1.61	0.65	0.73

The number of windows hyper-parameter with composition of data augmentation has been optimized for in this study and requires either a grid search or DA-search.

nauya

RESULTS : EXAMPLES

SSL with data augmentations performs better at detecting pedestrians/cyclists as well as cars at larger distances

2D DA for 3D-MOD

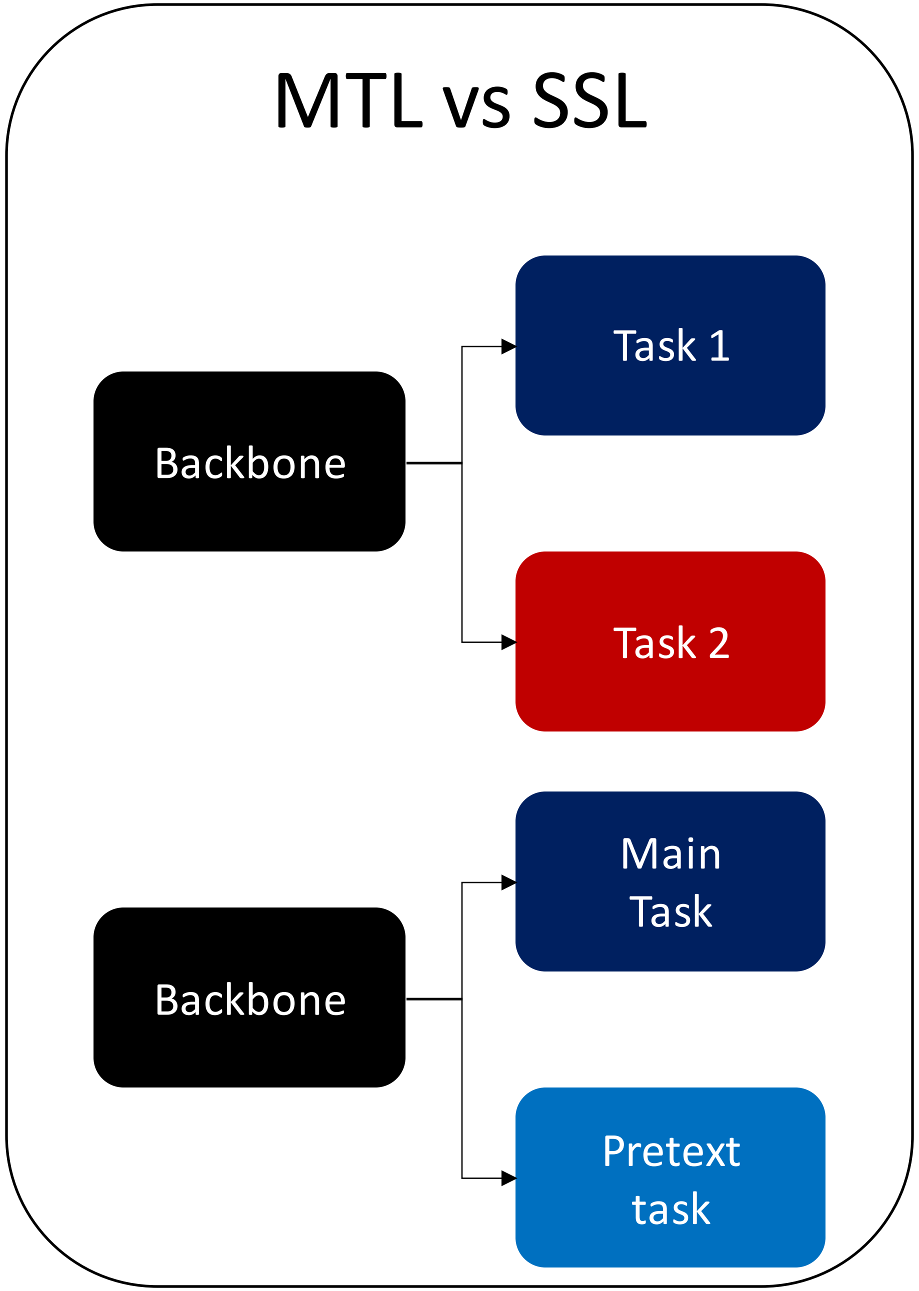
- DA stand-alone improves the performance of the main task
- **Box mixup/cut-paste** augmentation performs well

MOL-SSL task

- Enables the RTM3D network to classify foreground regions with different classes and background proportions better
- This is correlated with the box localisation task

DA-SSL Synergy :

- Data augmentation also helps the main task by providing representations that generalize for both main and augmented pretext-tasks
- SSL pretext tasks and their augmentations are both good regularizers (inductive bias) and can be combined fruitfully
- SSL-pretext task provides a soft-label and makes training with DA generalize better (hypothesis)
- **Cons** : Seperating augmentations between main and pretext task is not possible.



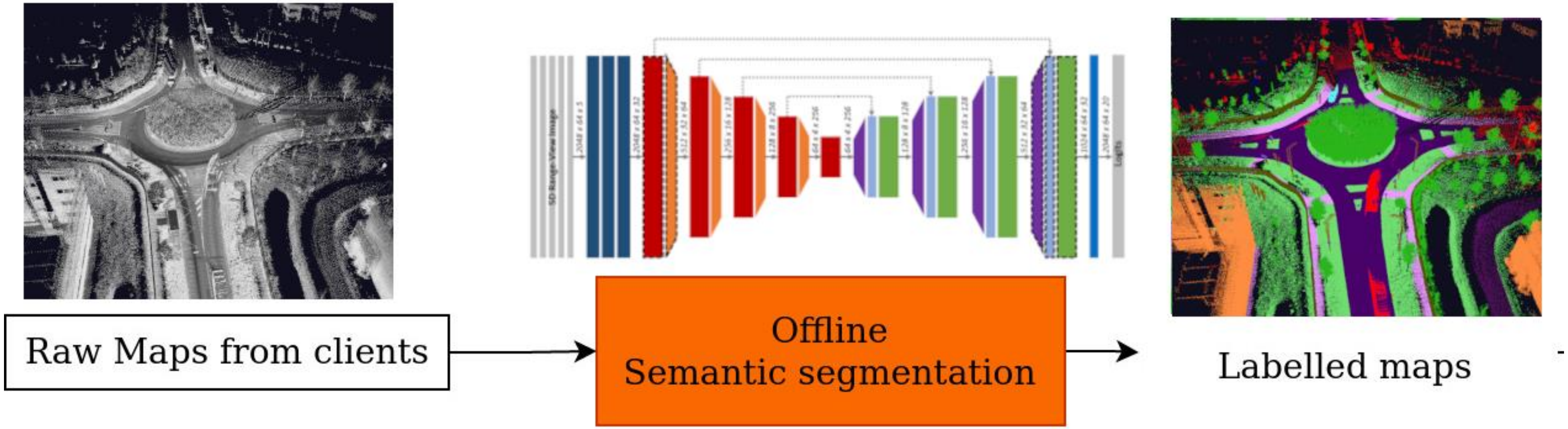
MTL : https://ruder.io/multi-task/

CASE STUDY 2 : DATA REDUNDANCY ON LARGE DATASETS

Studying data augmentation under Active learning setup to reduce data redundancy

Large scale pointcloud semantic segmentation are fundamental building blocks in modern AD perception stacks:

- Semantic Map layer in modern HDMaps
- Drivable zone extraction & Path planning
- Semantic re-localization and others…



Compressing Semantic-KITTI: Reducing dataredundancy on pointclouds by Active learning,
Anh Duong, Alexandre Almin, Leo Lemarie, B Ravi Kiran, In Submission 2021

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)

CASE STUDY 2 : DATA REDUNDANCY ON LARGE DATASETS

Studying data augmentation under Active learning setup to reduce data redundancy

Pointcloud semantic segmentation datasets have large amounts of redundant information

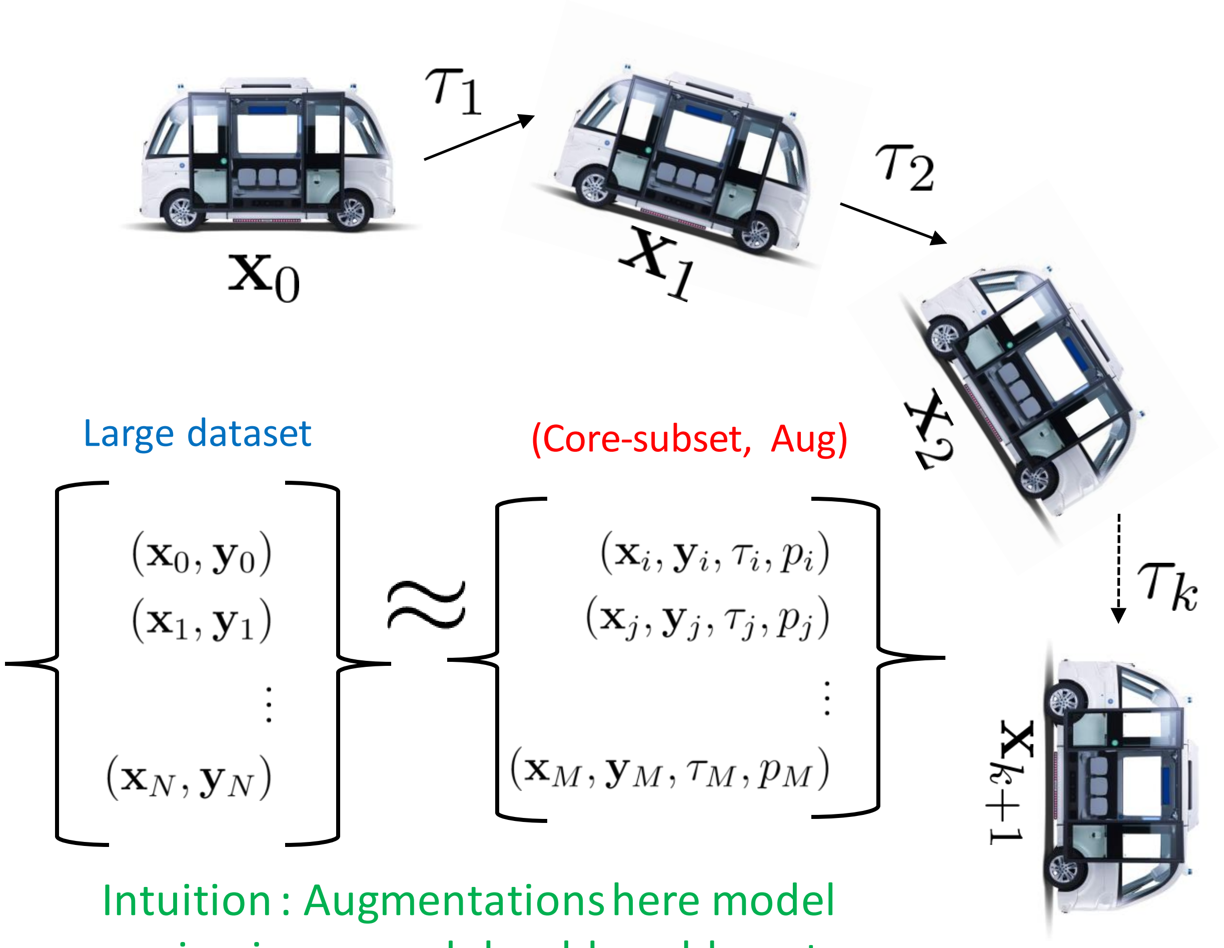
- Similar scans due to temporal correlation
- Similar scans due to similar urban environments
- Similar scans due to symmetries

Motivation :

- Data augmentations on large dataset had little gains
- How do we reduce redundancy or similar samples (pointcloud, GT) by selecting
- Full Dataset = A core subset + Augmentations

Approach :

- Study the effect of data augmentations on the active learning sampling



Intuition : Augmentations here model equivariances and should enable us to compress the dataset

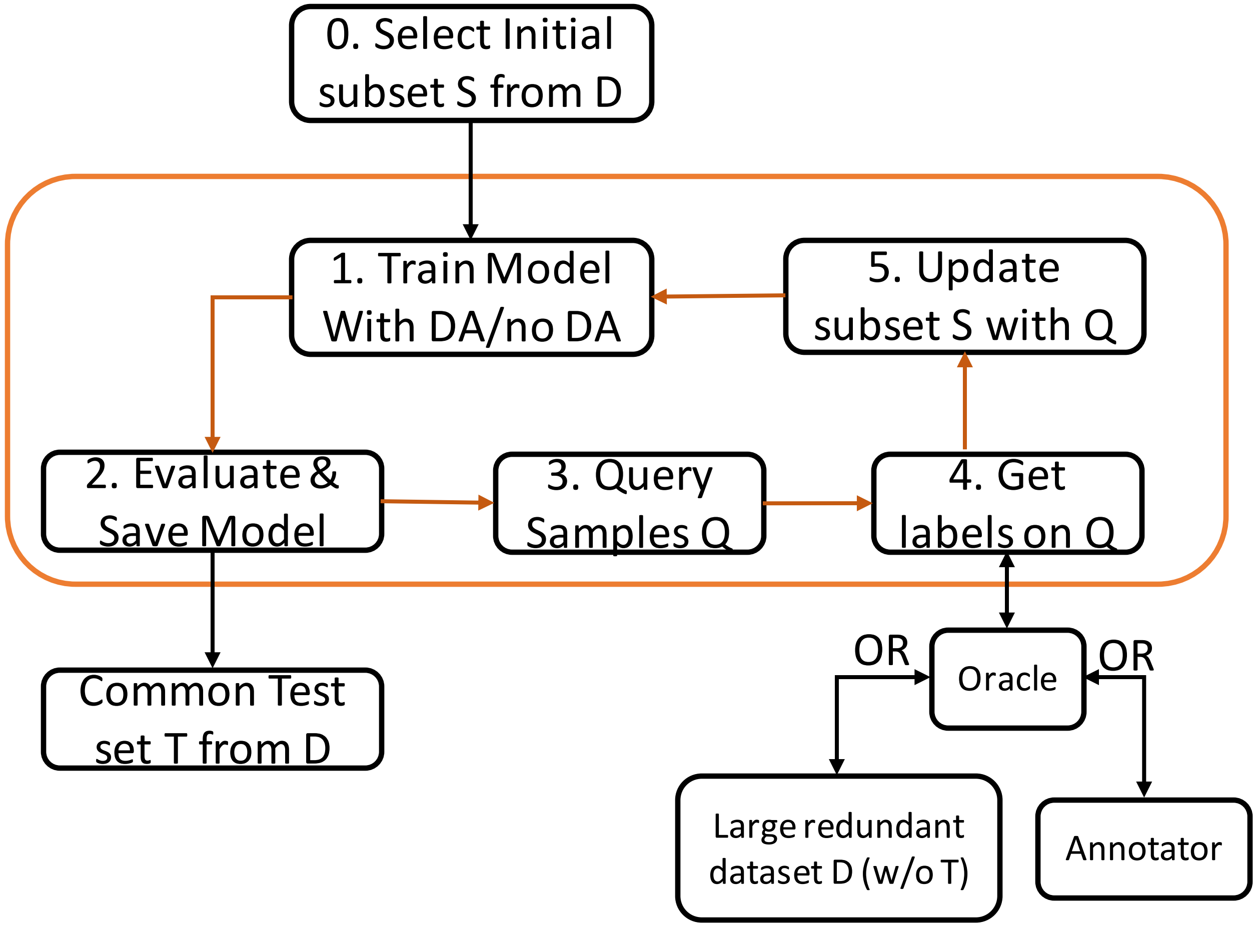
ACTIVE LEARNING

Background

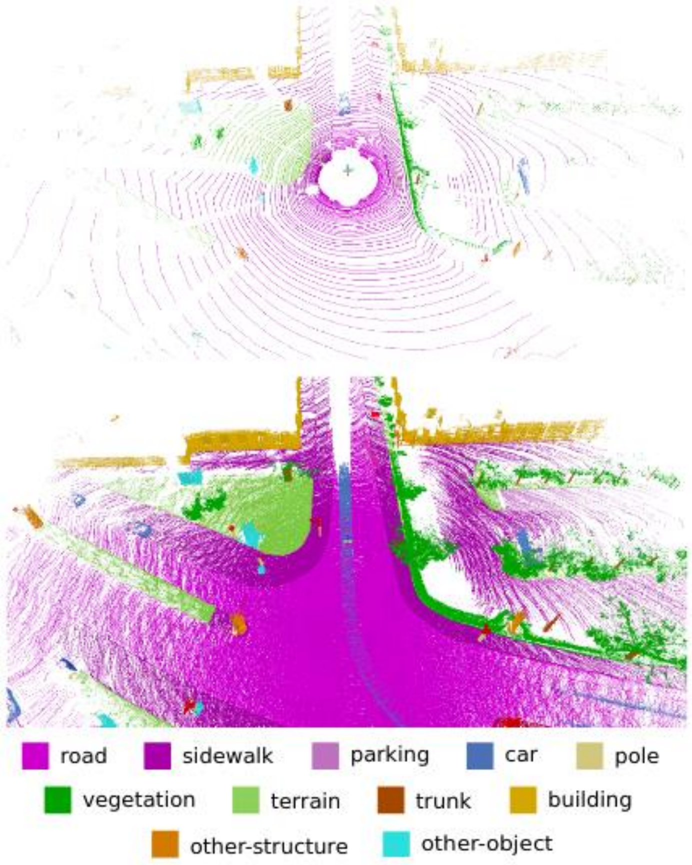
- Active learning (AL) is an interactive learning procedure
  - That greedily samples the most informative sample(s) to maximize the performance of the model.
  - Multiple ways to decide the most informative subset to label : likelihood P(x|M), uncertainty P(y|x)

- AL Components:
  - Training subset S
  - Query subset Q
  - **Heuristic func. h** (random, entropy, BALD)
  - **Aggregation func.** (aggregates pixel scores to scalar)
  - **Redundant/large** dataset D
  - Test set T

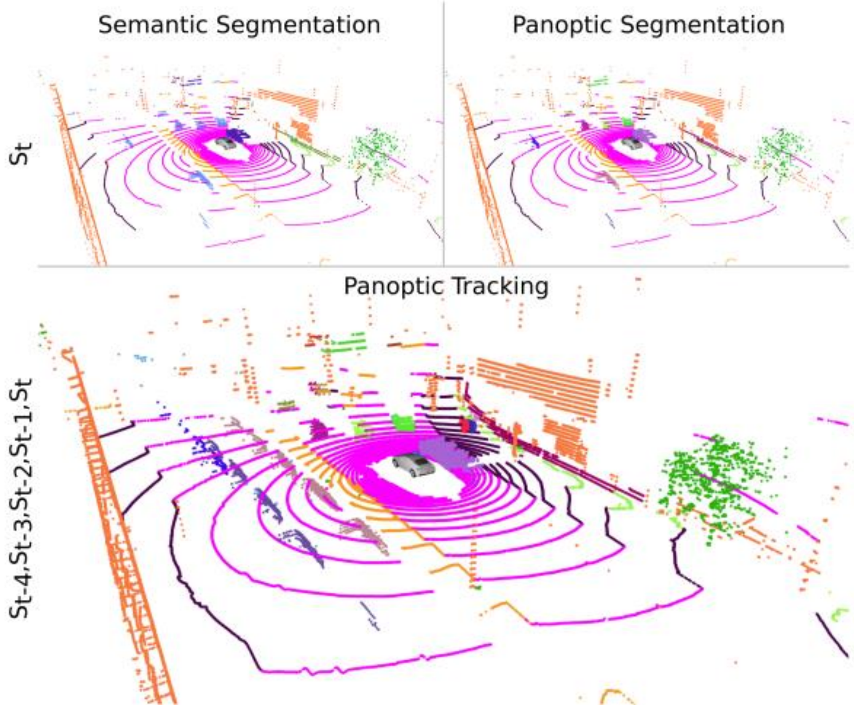
- Goals :
  - AL goal : Reduce annotation requests to Oracle
    - Reduces Labeling Cost
  - Our goal : Reduce redundancy large datasets
    - Reduces Training Time



POINTCLOUD SEMANTIC SEGMENTATION DATASET



[Semantic KITTI]



[Panoptic nuscenes]

- Large scale pointcloud sequences with semantic labels per point
  - Annotations include semantic class along with instance ID information
  - Panoptic-Nuscenes provides panoptic tracklet level labels which are temporally consistent across pointcloud scans
  - Established Architectures : Rangenet++, Salsanext, Cylinder3D

DATASET SIZE

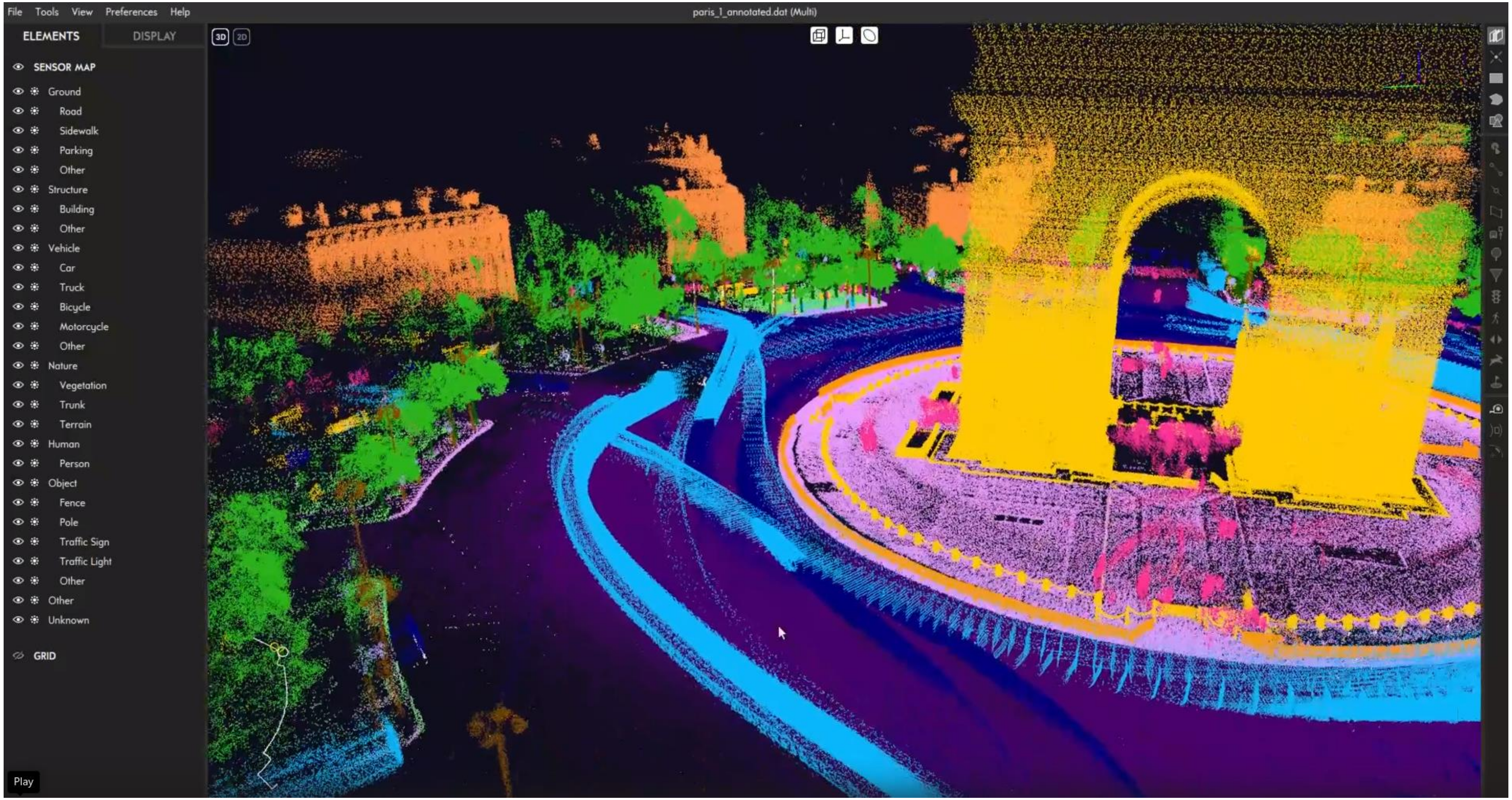
Semantic segmentation on pointclouds

Dataset	Cities	Sequences (Or Points)	#classes	Annotation	Sequential
Semantic KITTI	1x Germany	22 (long)	28	Point, Instance	Yes
Panoptic Nuscenes	Boston Singapore	1000 40K scans	32	Point, Box, Instance	Yes
PandaSet	2x USA	100	37	Point, Box	Yes
Semantic Navya (ours*)	22x Cities in 10 Countries France, Swiss, US, Denmark, Japan, Germany, Australia, Israel, Norway, New Zealand	22 (long) 50K scans	24	Point, Instance	Yes

*in construction

SEMANTIC NAVYA

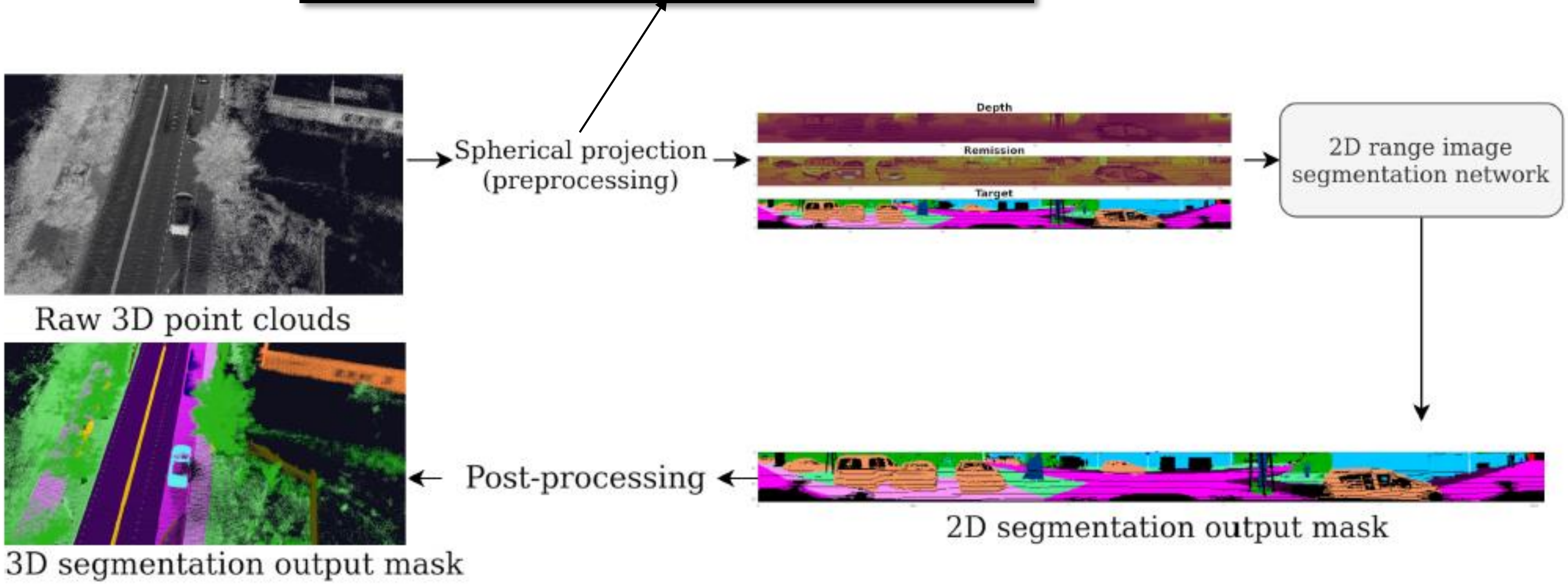
Large scale semantic segmentation dataset



RANGE IMAGE REPRESENTATION

Pointcloud representation

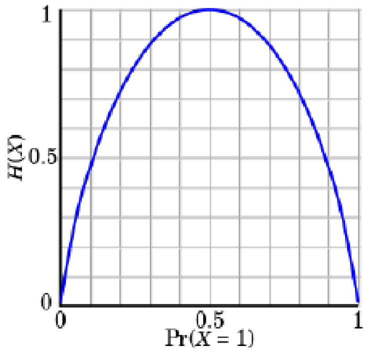
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[1 - \arctan(y, x)\pi^{-1}] & w \\ [1 - (\arcsin(zr^{-1}) + f_{up})f^{-1}] & h \end{pmatrix}$$



HEURISTIC FUNCTION

- Entropy heuristic :
  - Choose sample with the largest entropy
  - Samples that are most uncertain

$$H(y|x,L) = -\sum_{c}^{m} p(y^* = c|x,L) log(p(y^* = c|x,L))$$

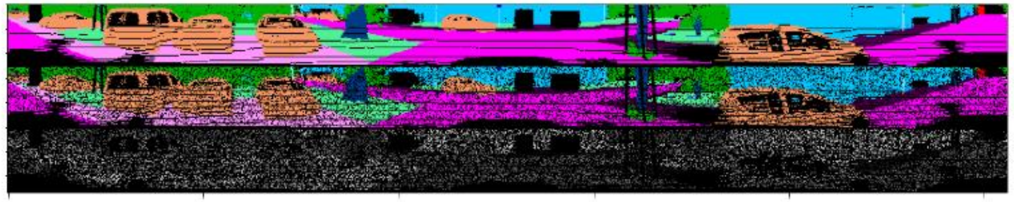


- BALD
  - Measures information gain between model predictions, and perturbed* model predictions
  - Select samples that maximize the information gain from model parameters

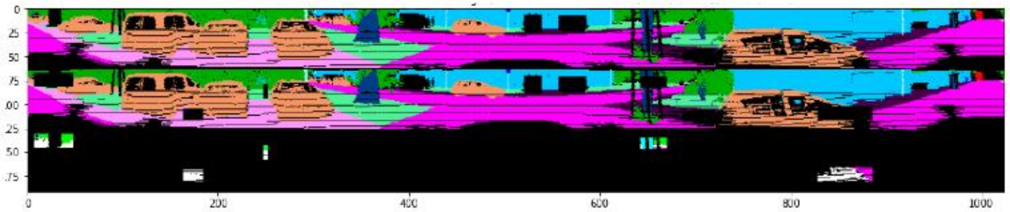
$$I(y^*,\omega|x^*,L) = H(y^*|x^*,L) - E_{p(\omega|L)}(H(y^*|x^*,\omega))$$

DATA AUGMENTATION IN POINTCLOUDS

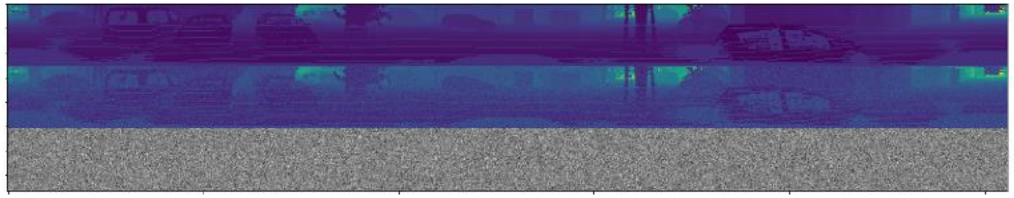
Using the range image representation



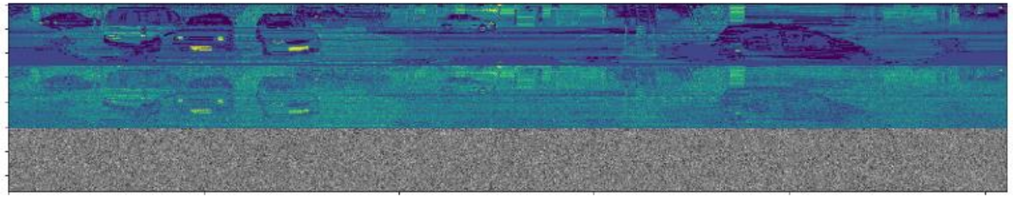
(a) Random dropout mask applied on range image and its target target



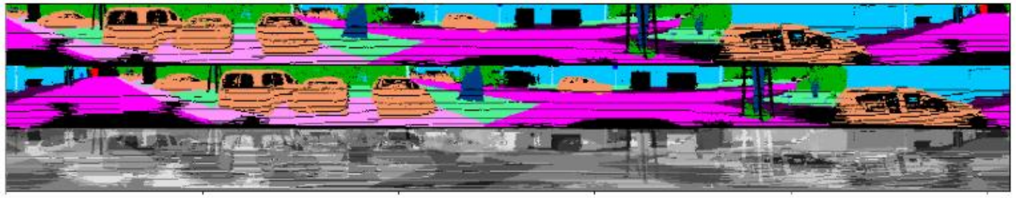
(b) Random masks out rectangle regions.



(c) Gaussian noise applied on depth of range image



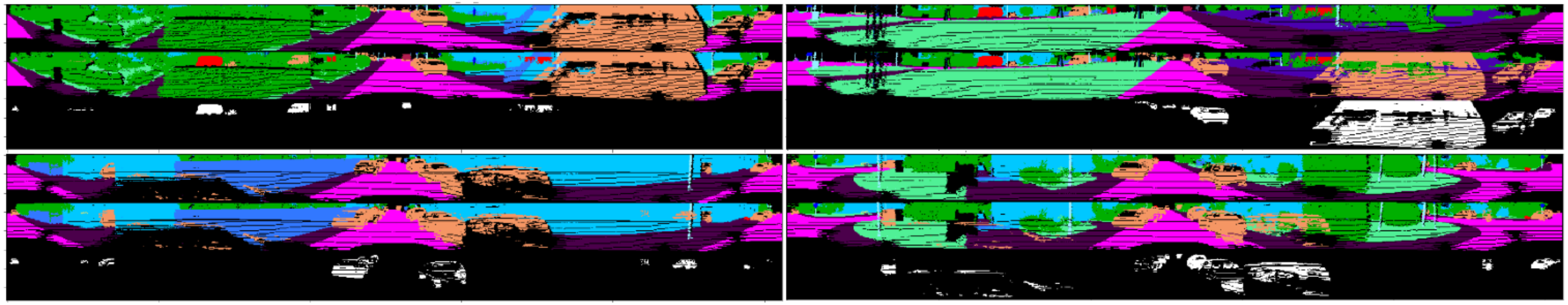
(d) Gaussian noise applied on remission channel of range image



(e) Random rotate range image and its target

DATA AUGMENTATION IN POINTCLOUDS

Using the range image representation



(f) Random copy and paste instances from one scan to another within a batch

EXPERIMENTAL SETUP ON SEMANTIC KITTI

AL training setup

AL steps : 25 **Budget** : 240 samples

Full Dataset D: 6000 samples
[Random 1/5th subset of Semantic KITTI]

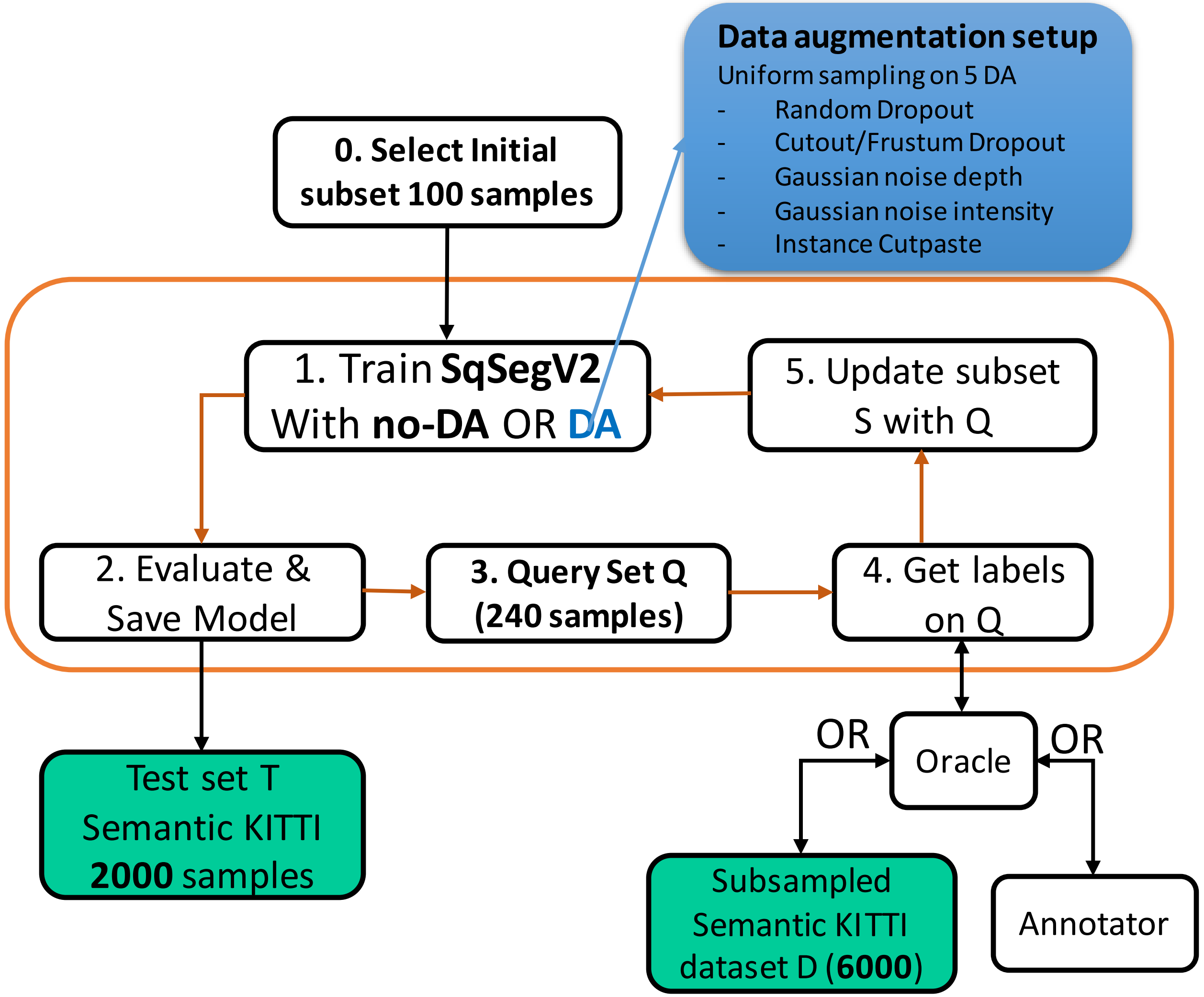
Model : SqueezeSegV2

Pointcloud representation : Spherical Range image

Test Set T : 2000 samples

Heuristics : **Random, Entropy, BALD**

Data-augmentation : with and without



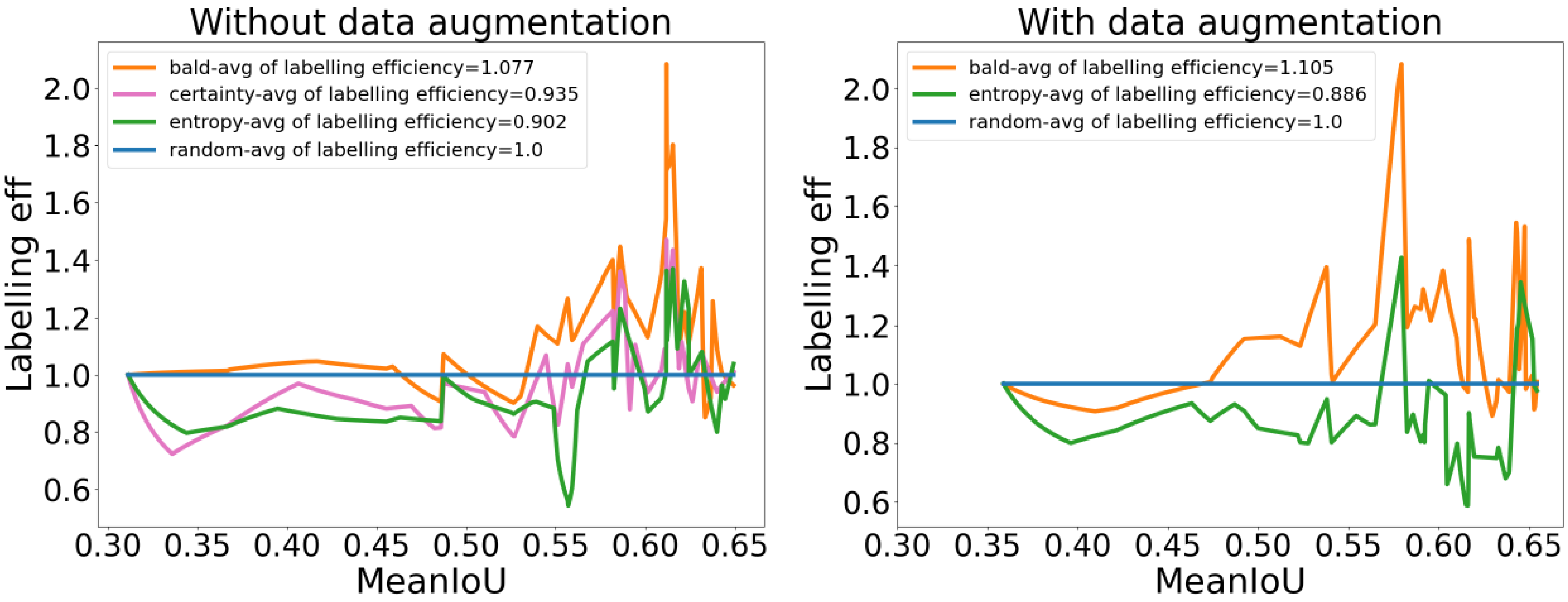
RESULTS

Label Efficiency

DA provides gains in performance at each AL step on Full dataset of 6000 samples (Semantic KITTI subset)

LABEL EFFICIENCY

With and without data augmentation

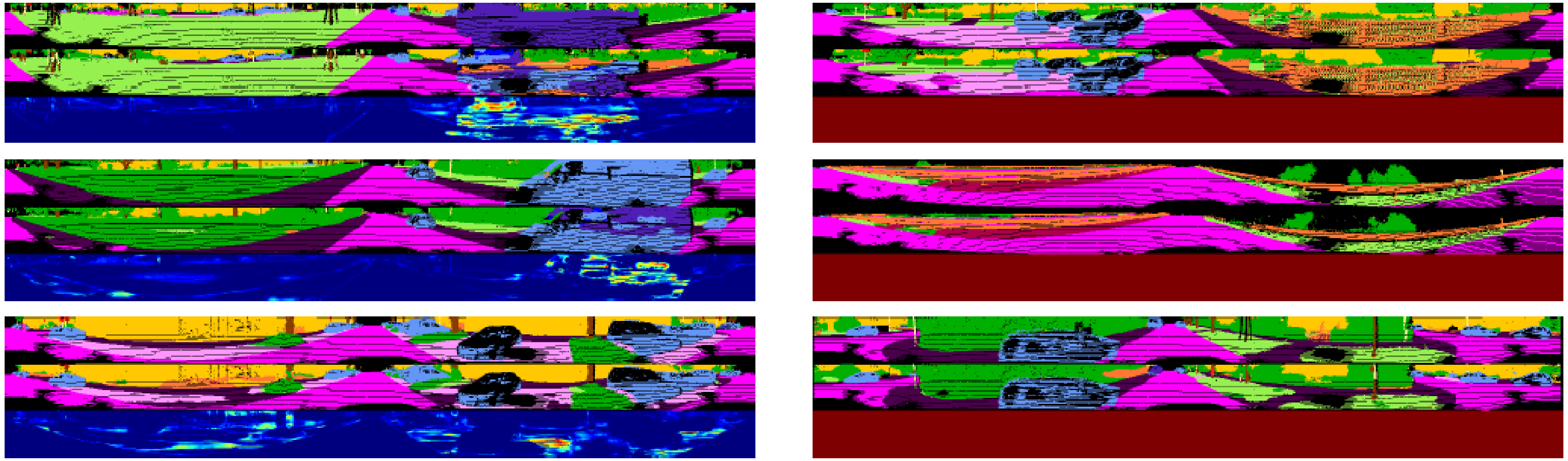


BALD with DA has best performance on SemanticKITTI.

RANKED HARD EXAMPLES BY THE HEURISTIC

BALD vs Random

Ground truth
Prediction
Heuristic function
Ground truth
Prediction
Heuristic function
Ground truth
Prediction
Heuristic function



(a) BALD (b) Random

Figure: Top 3 **hardest** samples selected at step 2/25. Each sample includes, from top to bottom, ground truth, prediction, and image scores of that sample.

- Larger diversity in samples from the BALD heuristic.
- No Heuristic scores are available for Random heuristic since they are sampled uniformly

EFFECT OF DATA AUGMENTATION FOR CLASSIFICATION

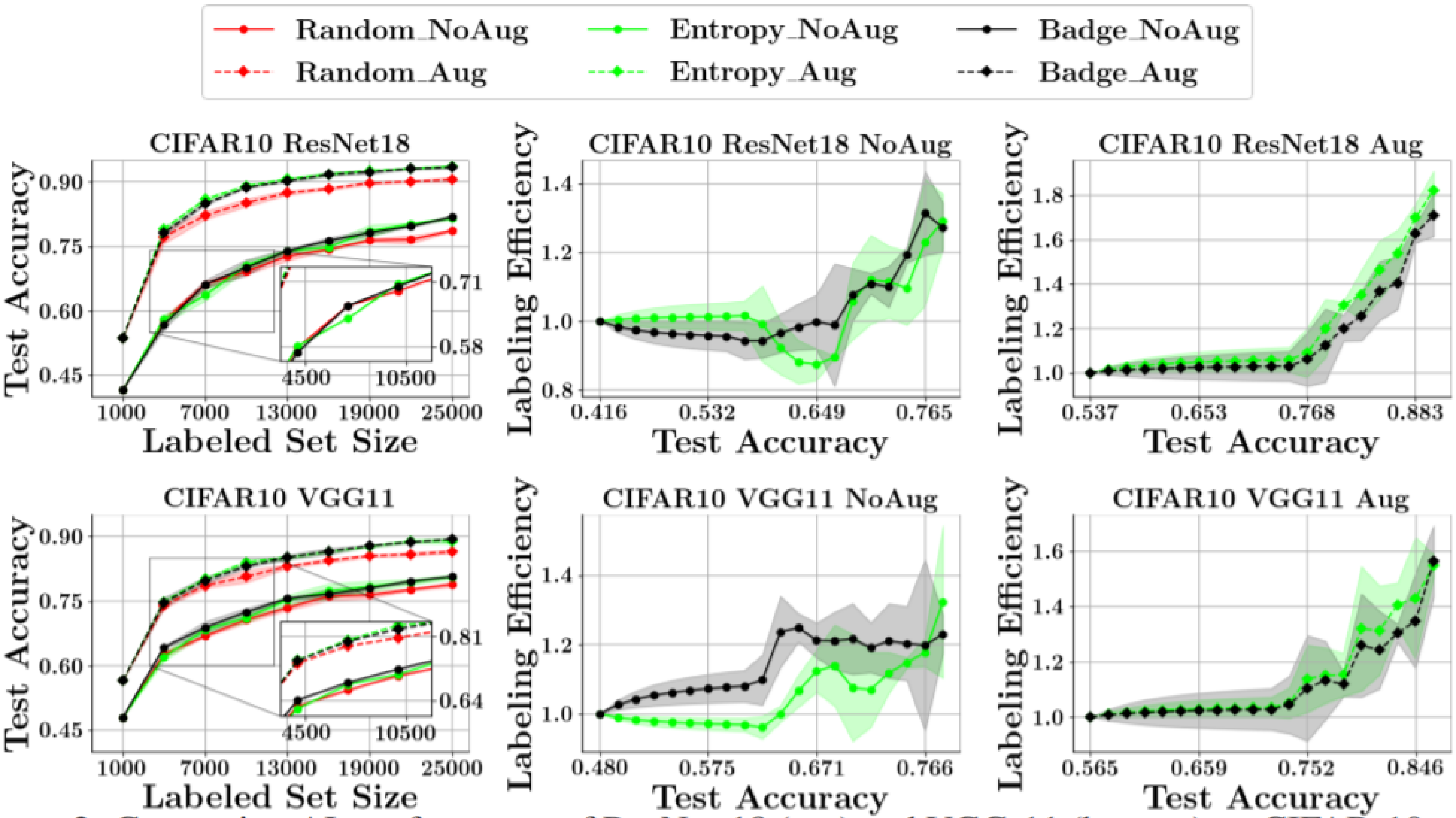


Figure 2: Comparing AL performance of ResNet-18 (top) and VGG-11 (bottom) on CIFAR-10 with and without augmentation. Data augmentation not only increases test accuracy but also improves the labeling efficiencies of AL. Furthermore, BADGE outperforms entropy sampling without data augmentation, but BADGE loses its advantage over entropy sampling when data augmentation is used.

Beck, Nathan, et al. "Effective Evaluation of Deep Active Learning on Image Classification Tasks." arXiv preprint arXiv:2106.15324 (2021).

CONCLUSIONS & FUTURE WORK CASE STUDY 2

Conclusions

- DA enable better accuracies at each AL loop step
  - DA provides better label efficiency by sampling pointclouds that are different from dataset+DA samples
- A heuristic's performance with DA applied depends on the task
  - Entropy with DA was better than a sophisticated heuristic function such as BADGE
  - BALD along with DA was better than Random which performed better than Entropy
  - Tasks : Classification vs Semantic Segmentation
- Recent work on Semi-Supervised learning to AL framework
  - Similar effect of Data augmentations while working with unsupervised DA (consistency loss)

Future Work

- Complete benchmark on full semantic KITTI and Semantic Navya dataset
- Aggregation maps uncertainty scores across a whole PC/image into a scalar
  - Require a way to sample regions/volumes of PCs
  - Heuristic functions are scalars and confound multiple regions of the image
- Find the **most informative set** of data augmentations for a dataset to reduce redundancy

REFERENCES I

Data augmentations for monocular 3d detection

1. Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." Journal of Big Data 6.1 (2019): 1-48.
2. Li, Peixuan, et al. "Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer International Publishing, 2020.
3. Hu, Hou-Ning, et al. "Monocular Quasi-Dense 3D Object Tracking." arXiv preprint arXiv:2103.07351 (2021).
4. Santhakumar, Khailash, et al. "Exploring 2D data augmentation for 3D monocular object detection." *arXiv preprint arXiv:2104.10786* (2021).
5. Lian, Qing, et al. "Geometry-aware data augmentation for monocular 3D object detection." *arXiv preprint arXiv:2104.05858* (2021).
6. Ning, Guanghan, et al. "Data Augmentation for Object Detection via Differentiable Neural Rendering." arXiv preprint arXiv:2103.02852 (2021).
7. Chen, Shuxiao, Edgar Dobriban, and Jane H. Lee. "A group-theoretic framework for data augmentation." Journal of Machine Learning Research 21.245 (2020): 1-71.
8. Beck, Nathan, et al. "Effective Evaluation of Deep Active Learning on Image Classification Tasks." arXiv preprint arXiv:2106.15324 (2021).

REFERENCES II

Data augmentations within Active learning for data redundancy

1. Behley, Jens, et al. "Semantickitti: A dataset for semantic scene understanding of lidar sequences." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
2. Fong, Whye Kit, et al. "Panoptic nuScenes: A Large-Scale Benchmark for LiDAR Panoptic Segmentation and Tracking." *arXiv preprint arXiv:2109.03805* (2021).
3. Milioto, Andres, et al. "Rangenet++: Fast and accurate lidar semantic segmentation." 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019.
4. Cortinhal, Tiago, George Tzelepis, and Eren Erdal Aksoy. "SalsaNext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds for autonomous driving." *arXiv preprint arXiv:2003.03653* (2020).
5. Zhou, Hui, et al. "Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation." arXiv preprint arXiv:2008.01550 (2020).
6. Hahner, M., Dai, D., Liniger, A., & Van Gool, L. (2020). Quantifying data augmentation for lidar based 3d object detection. arXiv preprint arXiv:2004.01643.
7. Ash, Jordan T., et al. "Deep batch active learning by diverse, uncertain gradient lower bounds." arXiv preprint arXiv:1906.03671 (2019), ICLR 2020. BADGE
8. https://baal.readthedocs.io/en/latest/
9. https://decile-team-distil.readthedocs.io/en/latest/index.html
10. Gao, Mingfei, et al. "Consistency-based semi-supervised active learning: Towards minimizing labeling cost." European Conference on Computer Vision. Springer, Cham, 2020.