3D Deep learning with pointclouds
Introduction

NAVYA INTRODUCTION

PASSENGERS TRANSPORT

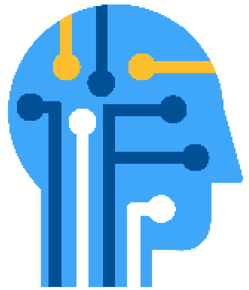
Autonom® Shuttle

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

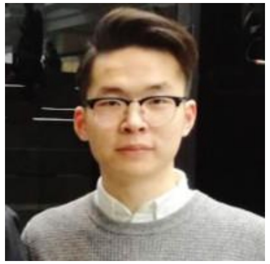
Deep learning modules on camera and LiDAR

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
  - 3D Object detection on pointclouds
- **Semantic Navya Dataset**


Alexandre Almin


Thomas Gauthier


Hao Liu


B Ravi Kiran


Leo Lemarie


Anh Duong

OVERVIEW

What is a Pointcloud

- LiDAR and 3D sensors
- Perception tasks: classification, detection, segmentation

Pointcloud representations for Deep Learning

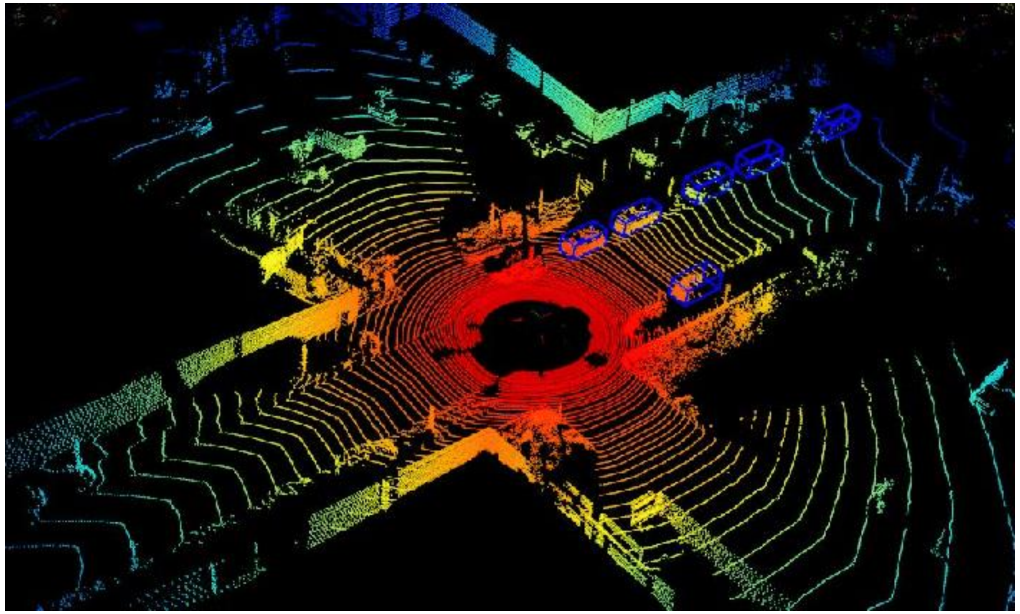
- Difference between pointclouds and images
- Pointcloud representations
  - Range images
  - Voxel based representations, Bird Eye View (BEV)
  - Continuous representations (KPconv)

Navya 3D Segmentation (N3DS) dataset

- Semantic segmentation on pointclouds
- Building an AL pipeline for mining informative samples
- Evaluation on Semantic-KITTI dataset

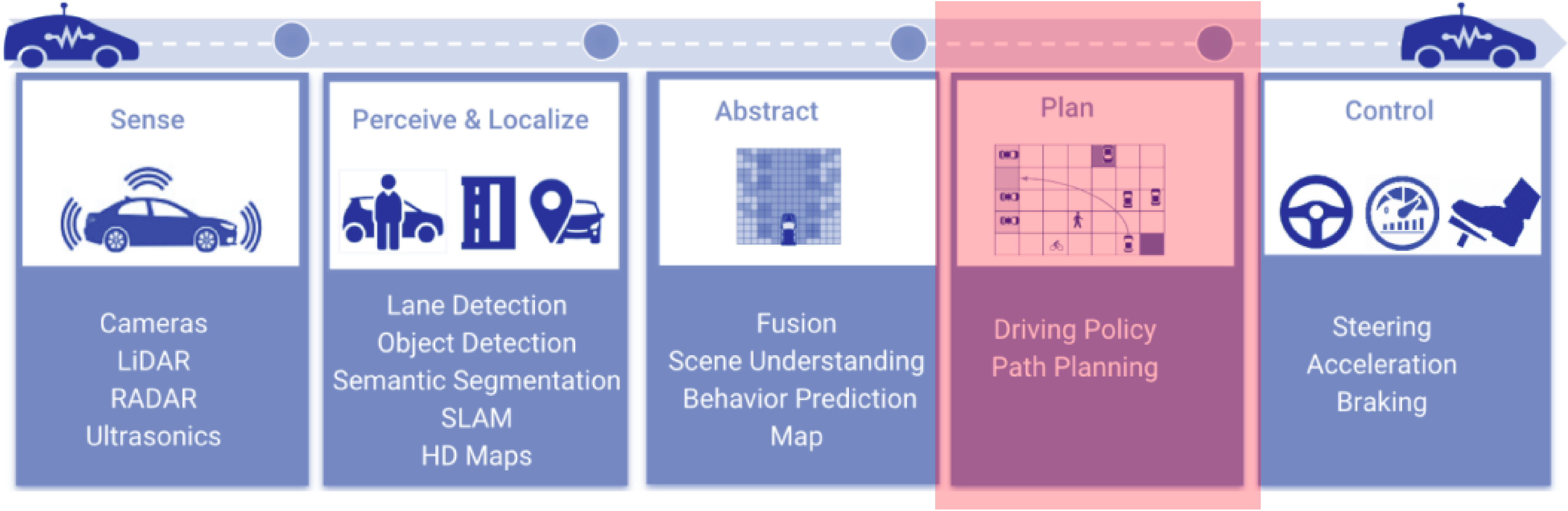
WHAT IS A POINTCLOUD

- A point-cloud is a set of points in 3D dimensions (cartesian)
  - Generated by LiDARS, Stereo Cameras, single layer proximity sensors, RADARs
- LiDARs : (Light detection and ranging)
  - Method for determining ranges (variable distance) by targeting an object or a surface with a laser and measuring the time for the reflected light to return to the receiver.
  - LiDARs also provide reflectivity or remission channel that measures the proportion of energy that was returned from a given laser fire
- Pointclouds types
  - Single scans at a single time instant t
  - Collection of scans that are aligned to create a Map
  - Single scans that are converted to occupancy grids (2D)



LARGE SCALE POINTCLOUDS

PERCEPTION TASKS IN AUTONOMOUS DRIVING



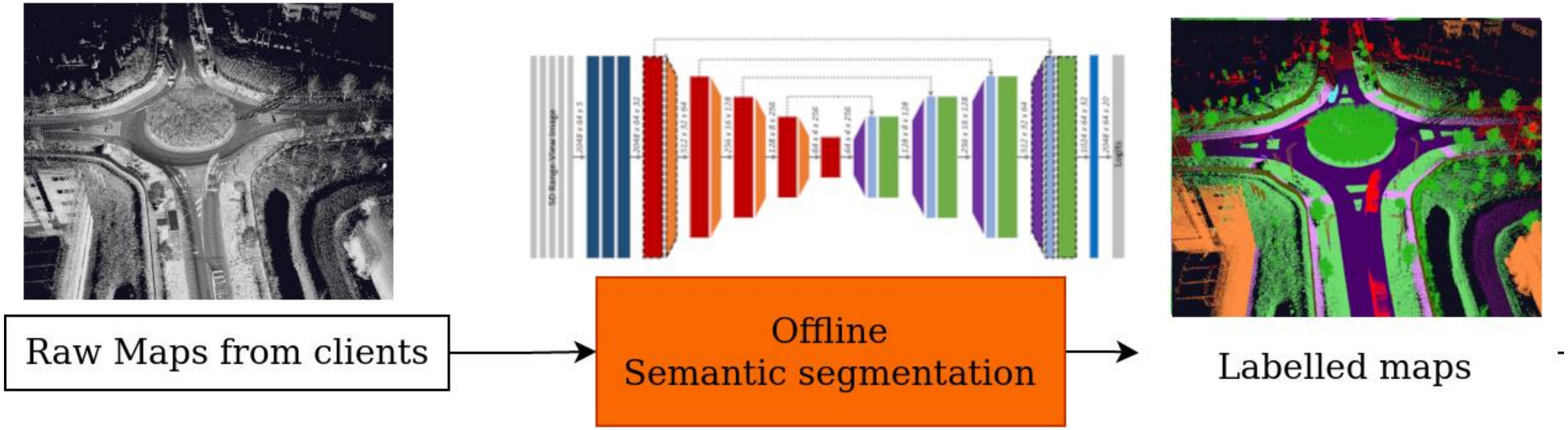
Scene interpretation tasks :
- 2D, 3D Object detection & tracking
- Traffic light/traffic sign
- Semantic segmentation
- Free/Drive space estimation
- Lane extraction
- HD Maps : 3D map, Lanes, Road topology
- Crowd sourced Maps

Fusions tasks:
- Multimodal sensor fusion
- Odometry
- Localization
- Landmark extraction
- Relocalization with HD Maps

Reinforcement learning tasks:
- Controller optimization
- Path planning and Trajectory optimization
- Motion and dynamic path planning
- High level driving policy : Highway, intersections, merges
- Actor (pedestrian/vehicles) prediction
- Safety and risk estimation

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, NeurIPS 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)

3D PERCEPTION TASKS

- Semantic segmentation of large-scale maps in 3D
- Object detection and tracking online and offline in 3D
- Pointcloud registration and SLAM (building maps)

KEY CHALLENGES POINTCLOUD PROCESS

- Pointclouds are sets : 3d points can arrive in different orders
  - There is no pixel grid or 3D grid that is inherently used to create pointclouds

- Pointclouds capture by LiDAR/3D scanners are 3D points sampled from 2D surfaces in a 3D world

- Pointclouds **vary in** density based on the sensor and its spatial resolution and position of the ego vehicle w.r.t surface

- Pointclouds are usually collected sequentially on the vehicle, this produces **motion ghosts**

navya

BEFORE DEEP LEARNING WAVE

Pointcloud semantic segmentation

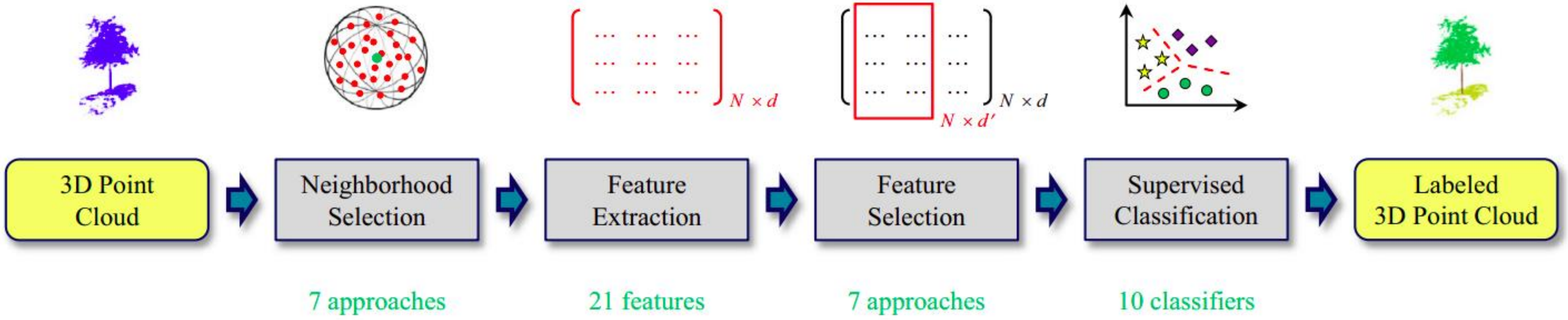
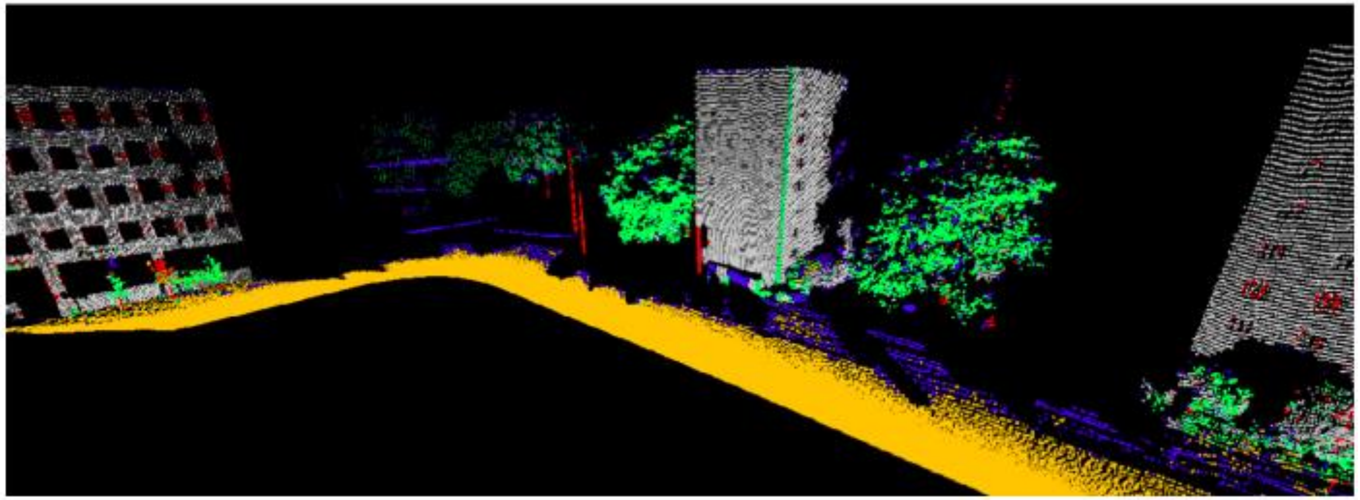


Fig. 1. The proposed framework and the quantity of attributes/approaches taken into account for evaluation.

$$L_\lambda = \frac{e_1 - e_2}{e_1}$$
$$P_\lambda = \frac{e_2 - e_3}{e_1}$$
$$S_\lambda = \frac{e_3}{e_1}$$
$$O_\lambda = \sqrt[3]{e_1 e_2 e_3}$$
$$A_\lambda = \frac{e_1 - e_3}{e_1}$$
$$E_\lambda = -\sum_{i=1}^{3} e_i \ln(e_i)$$
$$\Sigma_\lambda = e_1 + e_2 + e_3$$
$$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3}$$

1. Compute k-neighbourhood for each point in pointcloud
2. Evaluate eigen values
3. Calculate hand engineered geometric features
4. Classify point
5. Refine/post-process (MRFs/KNNs)



Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers Martin Weinmann, Boris Jutzi, Stefan Hinz, Clément Mallet ISPRS 2015

POINTCLOUD REPRESENTATIONS FOR DEEP LEARNING

- Represent them in an image format and then use classic semantic segmentation architectures
- Work with **spherical range images** (using a LiDAR's inherent structure)
- Introduce a grid artificially : **Voxel Grids** by partitioning the space into 3D cells
- Set based methods (To handle permutation invariance) : PointNet, PointNet++
- Define convolution in a continuous Space (KPConv/ConvPoint)
- Define a graph on pointclouds and work with Graph based CNN architectures (Superpoint Graphs)
- Hybrid architectures : Point-Voxel CNN for Efficient 3D Deep Learning(PVCNN), Cylinder3D (set based + voxel based)

MULTIVIEW REPRESENTATATION

A. Boulch et al./Computers & Graphics 000 (2017) 1–10

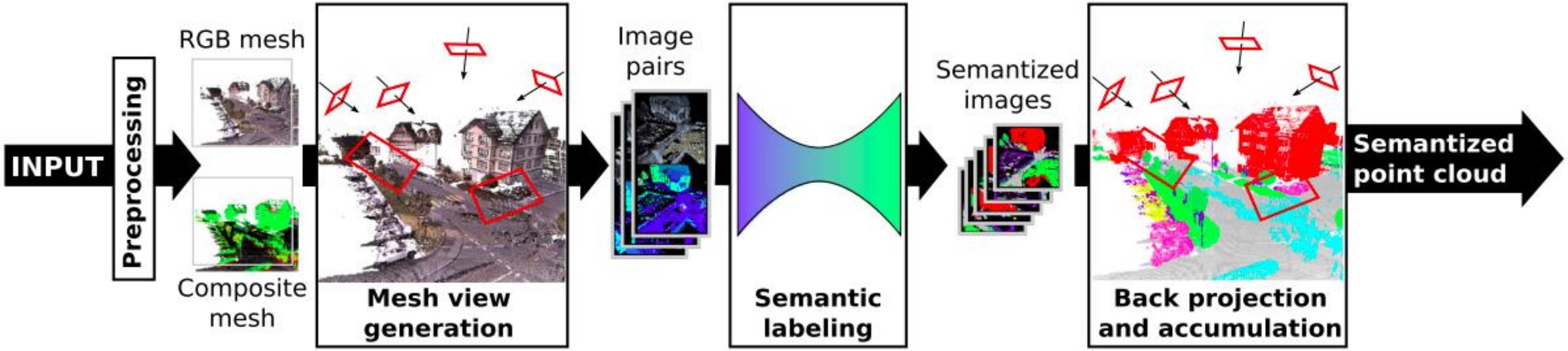


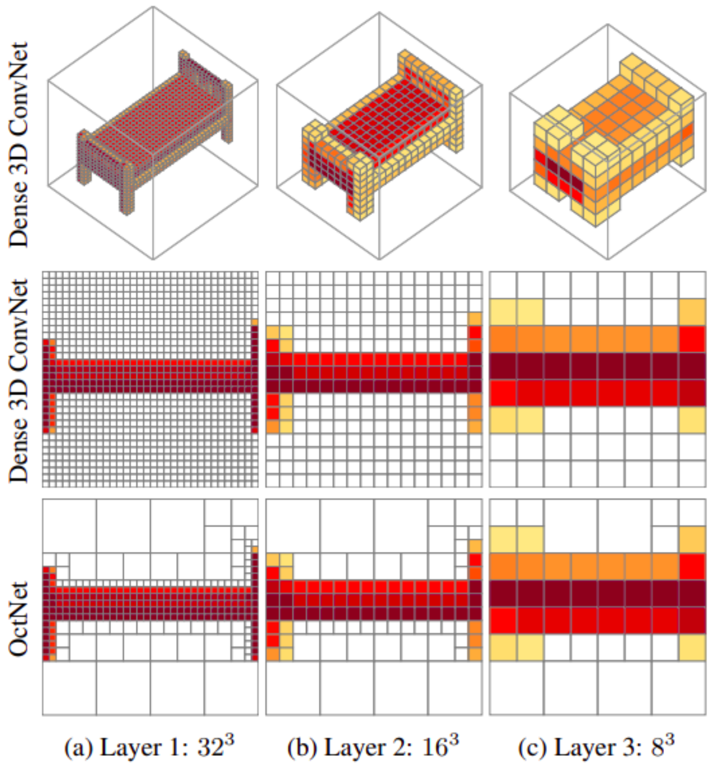
Fig. 2. Work-flow of the approach.

SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks, Alexandre Boulch, Joris Guerry, Bertrand Le Saux, Nicolas Audebert

VOXEL BASED METHODS

- Create a fixed discretization of 3d space by voxelization
  - Convolutional filters now operate in 3D space 3 strides
  - Feature maps are all 3D
  - Costly in memory even for small voxel sizes (memory explodes)
  - Rarely used in production

- More recent work
  - Sparse convolutions (SparseConv)



OctNet: Learning Deep 3D Representations at High Resolutions 2017
4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks 2020

RANGE IMAGE REPRESENTATION

Pointcloud representation

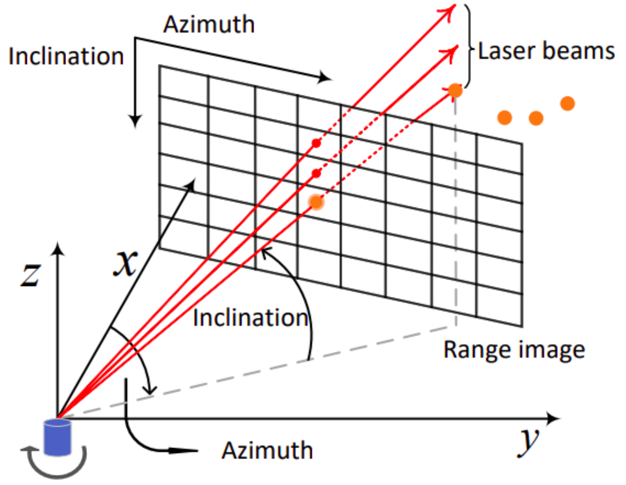
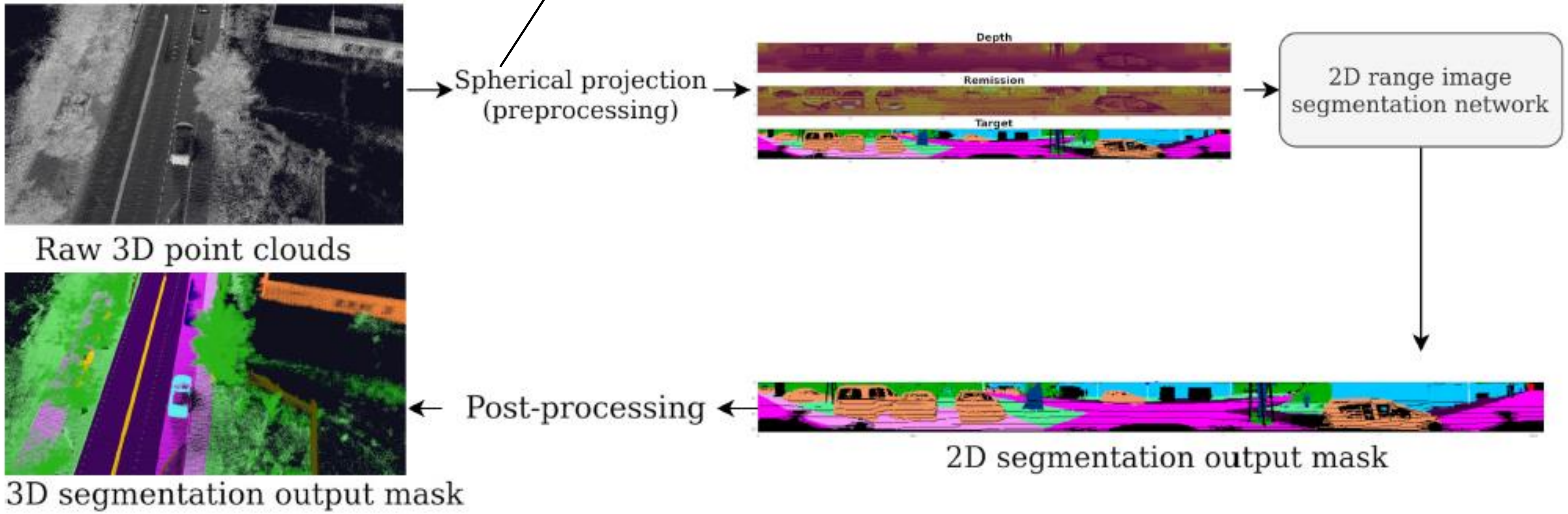


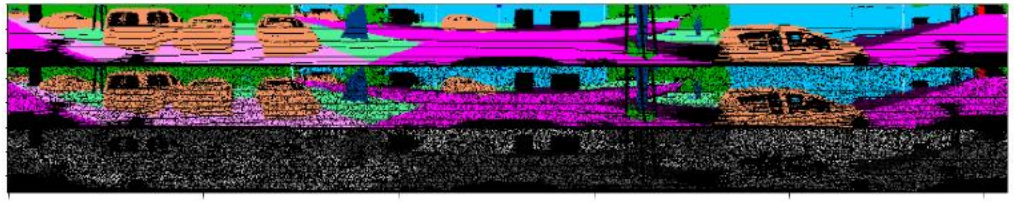
Figure 2. The illustration of the native range image.

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

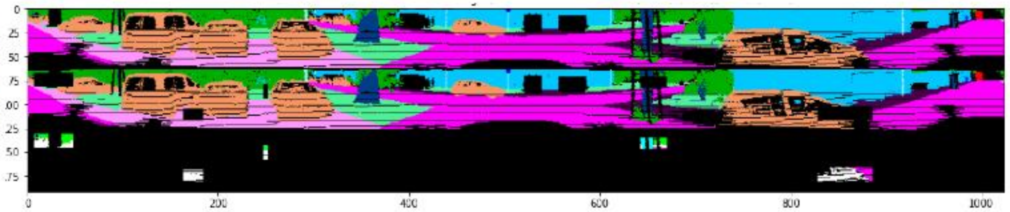


DATA AUGMENTATION IN POINTCLOUDS

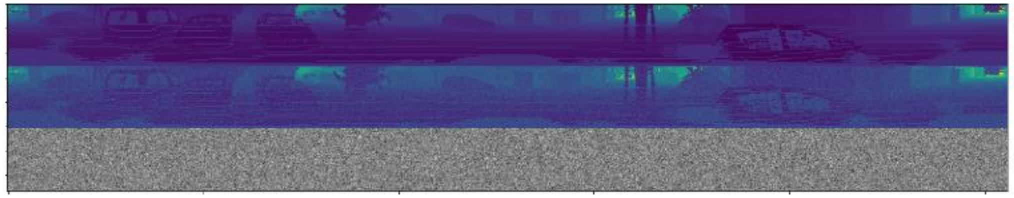
Using the range image representation



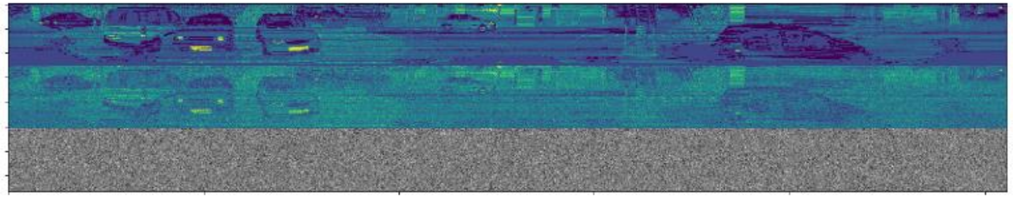
(a) Random dropout mask applied on range image and its target target



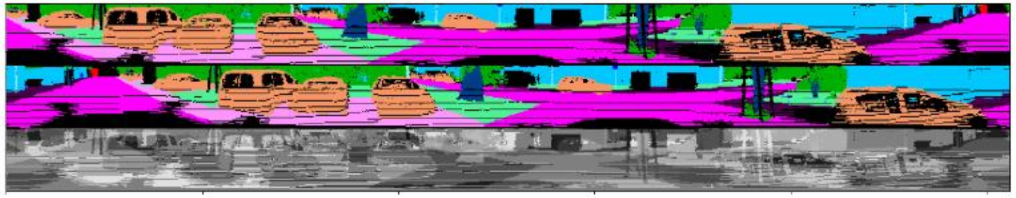
(b) Random masks out rectangle regions.



(c) Gaussian noise applied on depth of range image



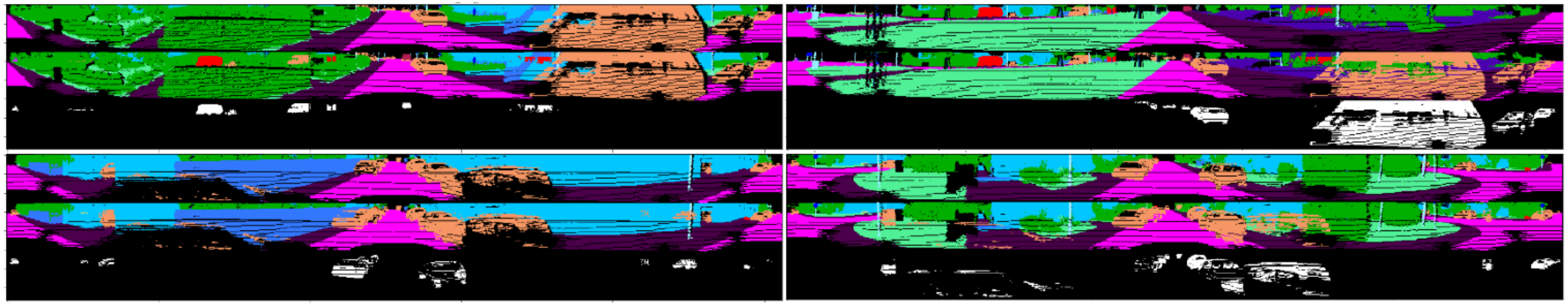
(d) Gaussian noise applied on remission channel of range image



(e) Random rotate range image and its target

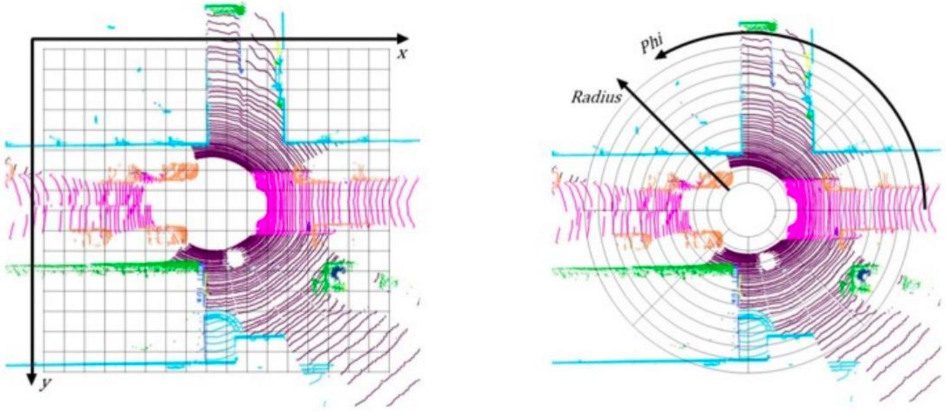
DATA AUGMENTATION IN POINTCLOUDS

Using the range image representation



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

BIRD EYE VIEW REPRESENTATION



(a) Cartesian BEV (b) Polar BEV

Two BEV quantization strategies. Each grid cell on the image denotes one feature in a feature map

Comparing Camera, LiDAR-Spherical, LiDAR-BEV views

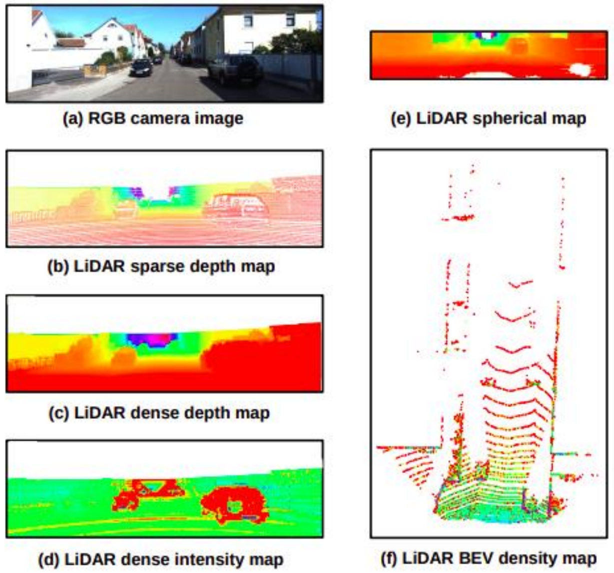
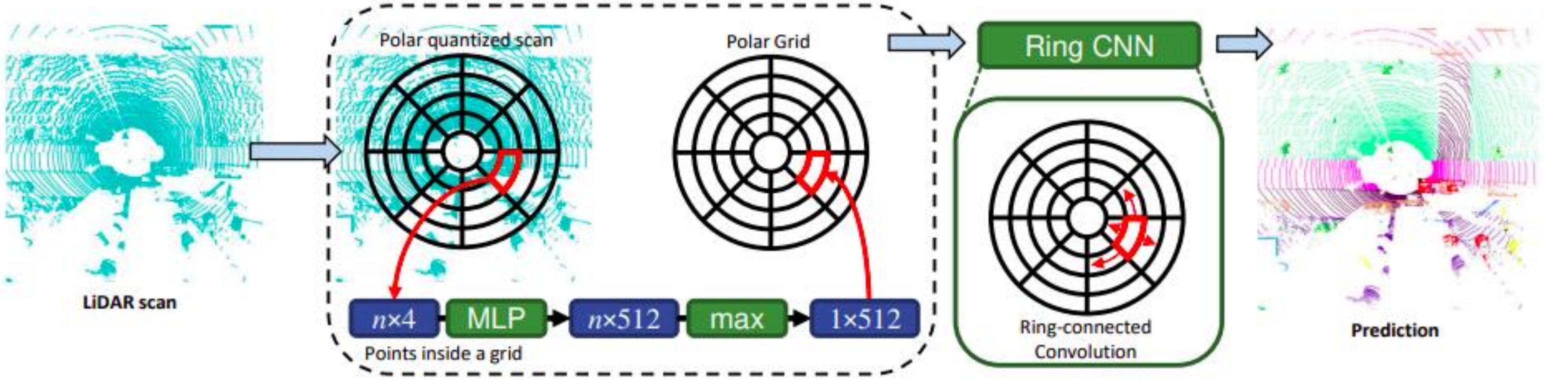


Fig. 6: RGB image and different 2D LiDAR representation methods. (a) A standard RGB image, represented by a pixel grid and color channel values. (b) A sparse (front-view) depth map obtained from LiDAR measurements represented on a grid. (c) Interpolated depth map. (d) Interpolation of the measured reflectance values on a grid. (e) Interpolated representation of the measured LiDAR points (surround view) on a spherical map. (f) Projection of the measured LiDAR points (front-facing) to bird's eye view (no interpolation).

[Ref:https://arxiv.org/pdf/1902.07830.pdf](https://arxiv.org/pdf/1902.07830.pdf)

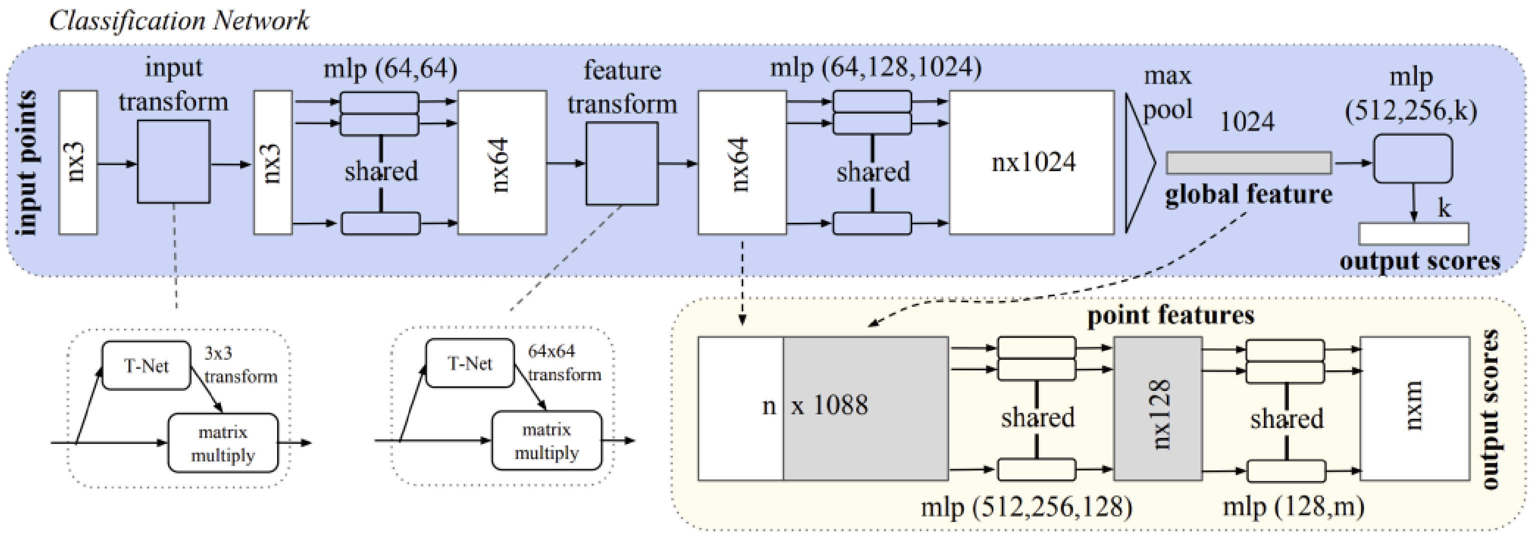


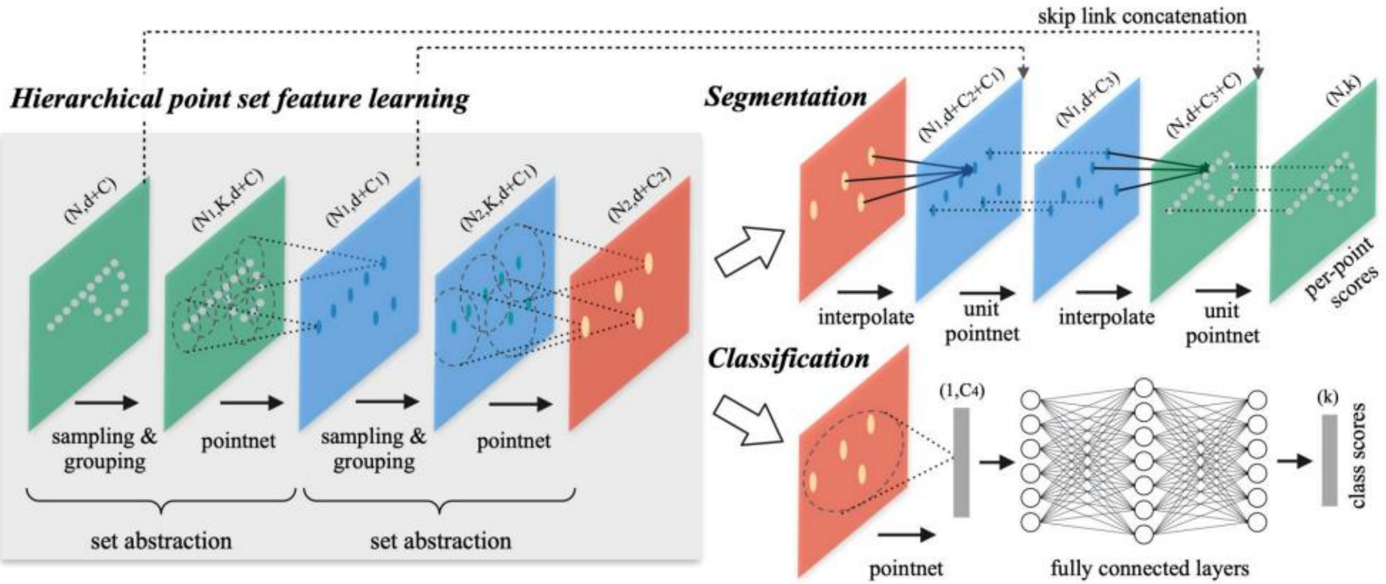
PolarNet: An Improved Grid Representation for Online LiDAR Point Clouds Semantic Segmentation CVPR 2020

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation 2017
PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space

SET BASED METHODS

POINTNET/POINTNET++





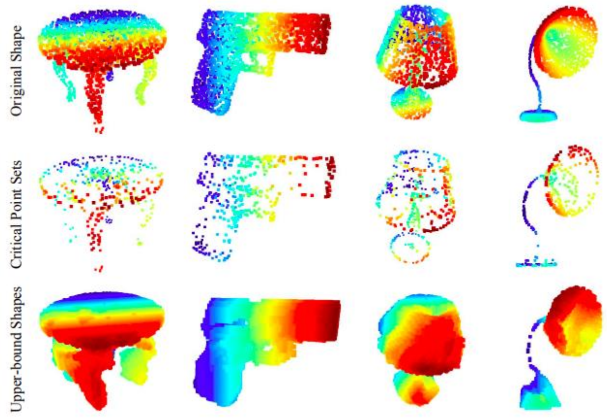


Figure 7. **Critical points and upper bound shape.** While critical points jointly determine the global shape feature for a given shape, any point cloud that falls between the critical points set and the upper bound shape gives exactly the same feature. We color-code all figures to show the depth information.

CONTINOUS CONVOLUTION ARCHITECTURES

KPConv

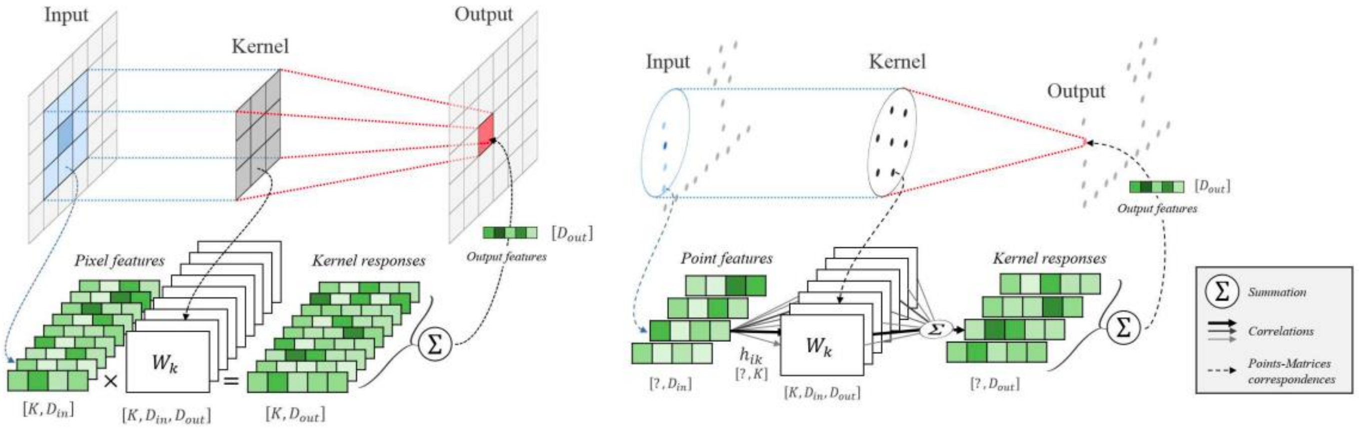


Figure 2. Comparison between an image convolution (left) and a KPConv (right) on 2D points for a simpler illustration. In the image, each pixel feature vector is multiplied by a weight matrix $(W_k)_{k<K}$ assigned by the alignment of the kernel with the image. In KPConv, input points are not aligned with kernel points, and their number can vary. Therefore, each point feature f_i is multiplied by all the kernel weight matrices, with a correlation coefficient h_{ik} depending on its relative position to kernel points.

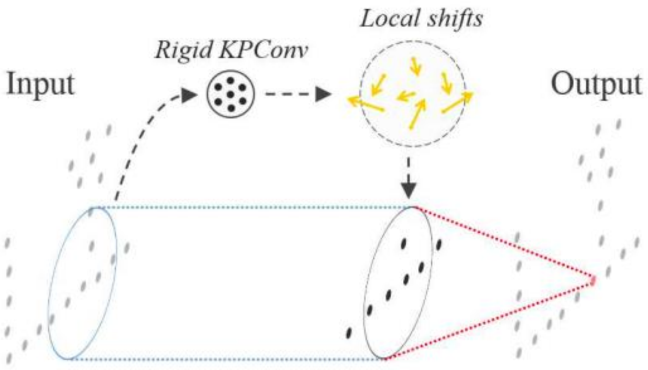
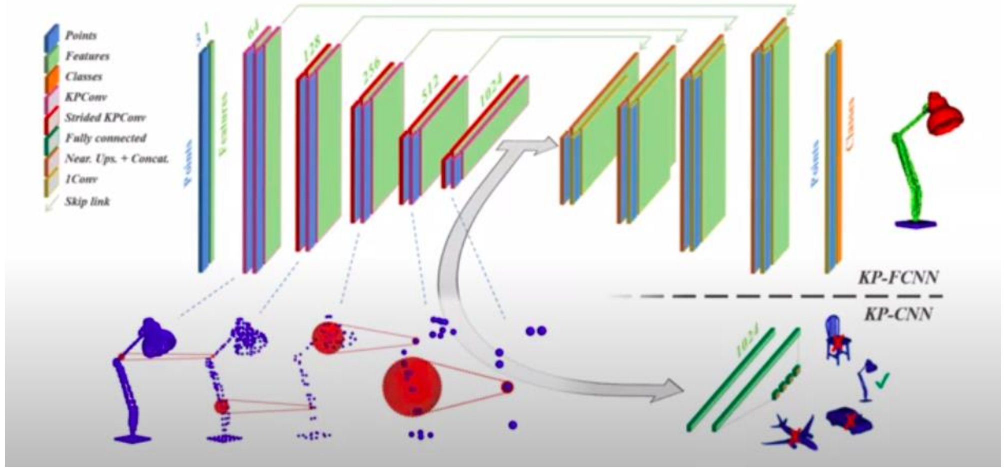


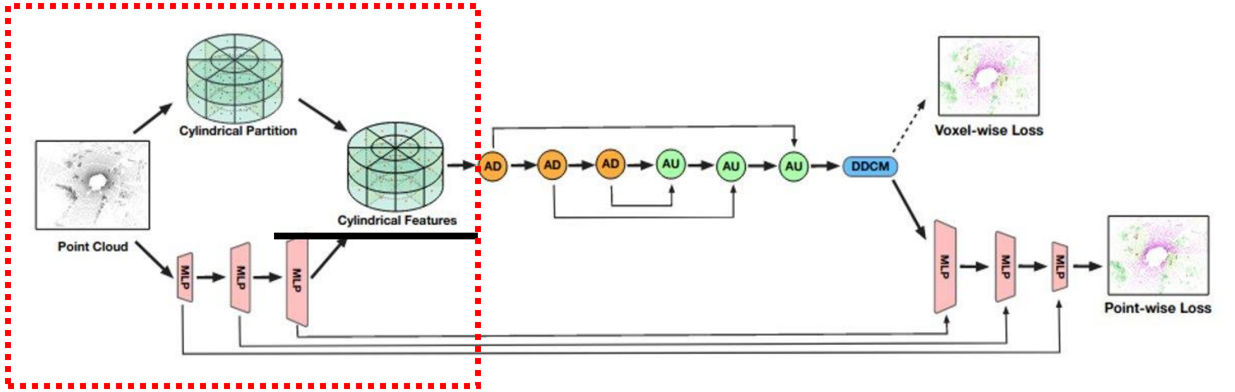
Figure 3. Deformable KPConv illustrated on 2D points.

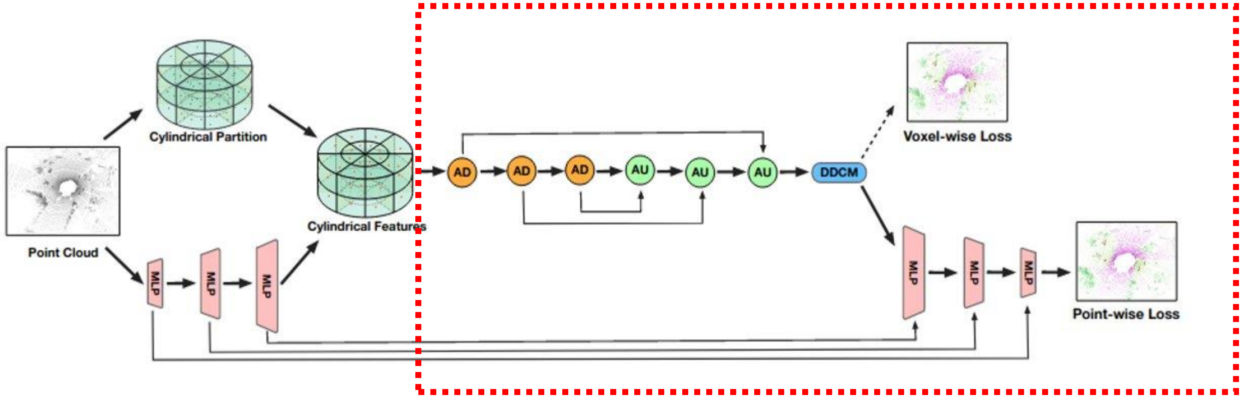


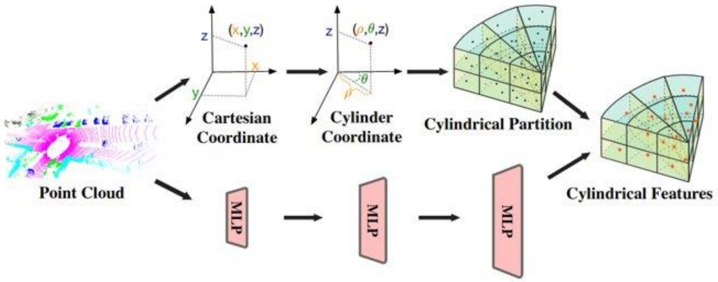
KPConv: Flexible and Deformable Convolution for Point Clouds ICCV 2019 Hugues Thomas et al
ConvPoint: Continuous Convolutions for Point Cloud Processing Boulch 2019

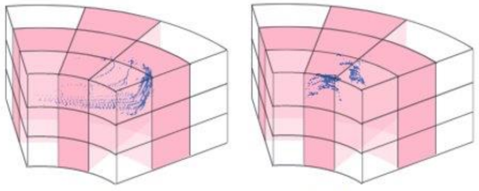
CONTINOUS CONVOLUTION ARCHITECTURES

Cylinder3d

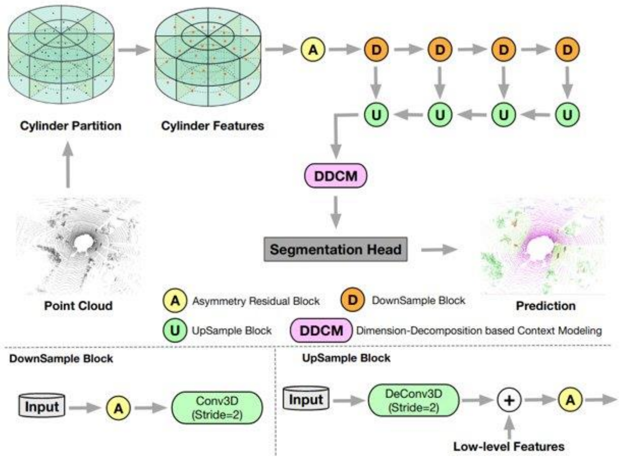


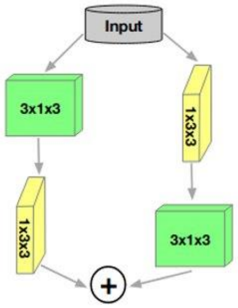




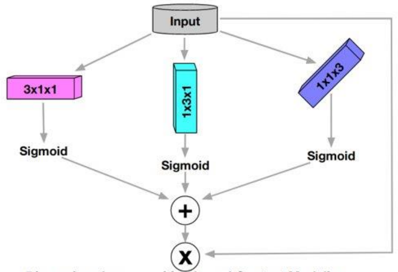


(a) Car (b) Motorcycle





Asymmetrical Residual Block

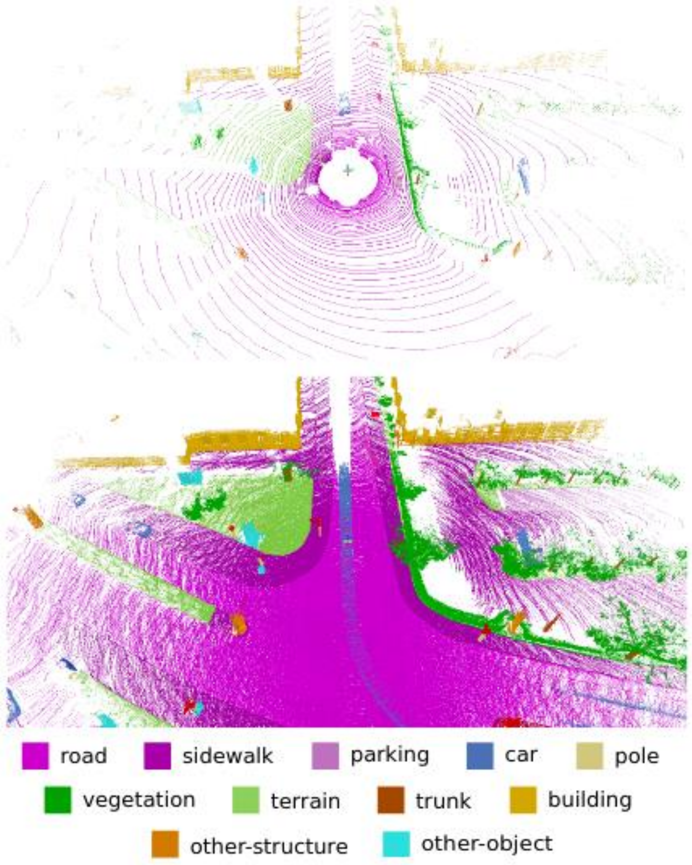


Dimension-decomposition based Context Modeling

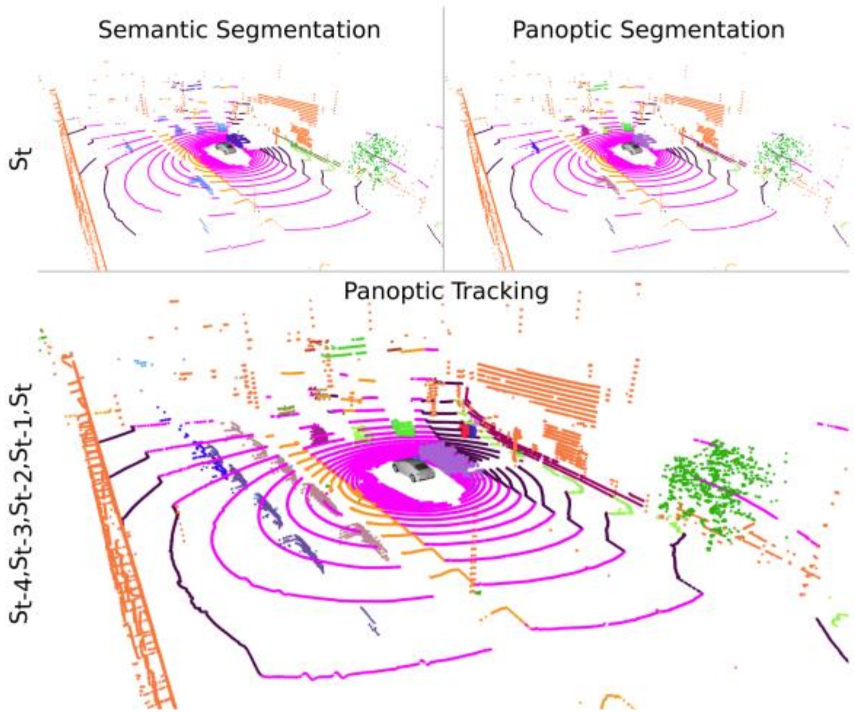
Cylinder3D: An Effective 3D Framework for Driving-scene LiDAR Semantic Segmentation

nauya

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

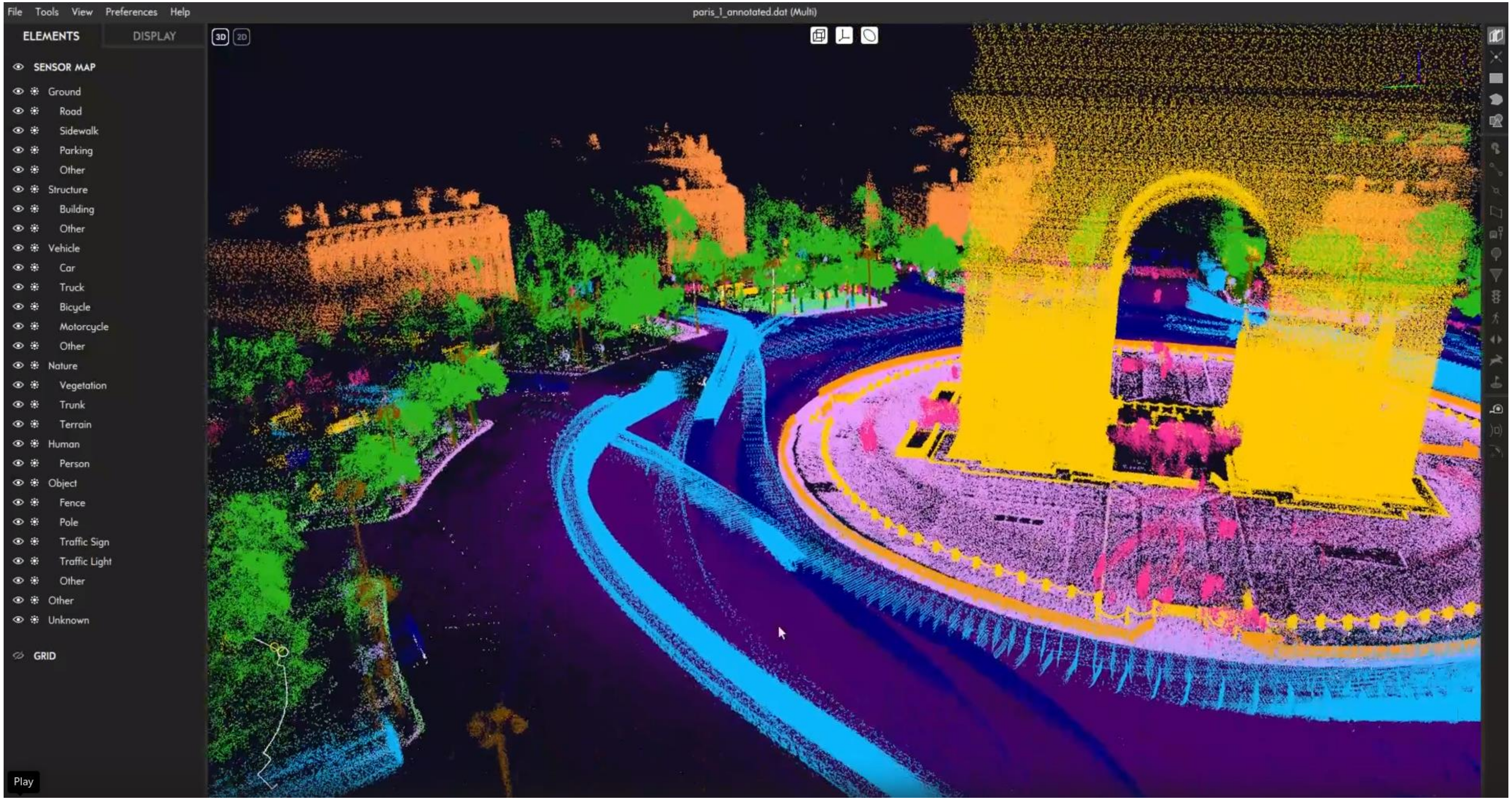
Semantic segmentation on pointclouds

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

NAVYA 3D SEGMENTATION(N3DS) DATASET

Large scale semantic segmentation dataset



REFERENCES

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.