LiDAR 3D Object Detection

MODEL CARD

Author name/s and affiliation	Till Beemelmanns, RWTH Aachen University
Contact details	till.beemelmanns@ika.rwth-aachen.de
Link/s	Institute for Automotive Engineering
Affiliation logo	ika | RWTH AACHEN UNIVERSITY
Date	15/09/2025

Disclaimer

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for the

INTRODUCTION

MODEL DESCRIPTION

VoxelNeXt is a 3D object detector designed for autonomous driving. It leverages voxel-based feature representation with efficient sparse convolutional networks. VoxelNeXt improves detection accuracy and efficiency by combining voxelization with advanced backbone designs, making it robust under diverse driving conditions while still maintaining real-time performance.

- **Model version:** *Model ID: V0.136*
- **Model license:** Open source (Link)
- **Citation:** https://arxiv.org/abs/2303.11301
- **Repository link:** Github

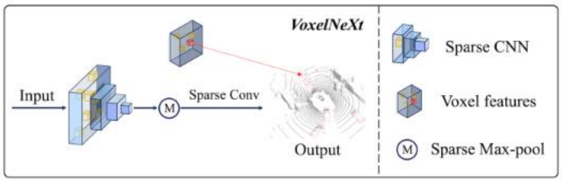

Figure 1 VoxelNeXt architecture overview. The model uses voxel-based feature representation of the point cloud with sparse 3D convolutions. The output is generated through classification and regression heads that predict object categories and 3D bounding box parameters.

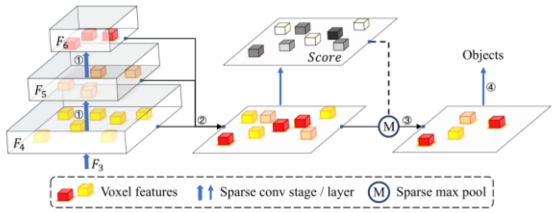

Figure 2 Detailed structure of VoxelNeXt framework.

MODEL DETAILS

MODEL ARCHITECTURE

- **Model type:** Sparse 3D Convolutional Network (voxel-based)
- **Model task:** 3D Object Detection
- **Inputs:** Voxelized 3D Point Cloud (X, Y, Z, Intensity)

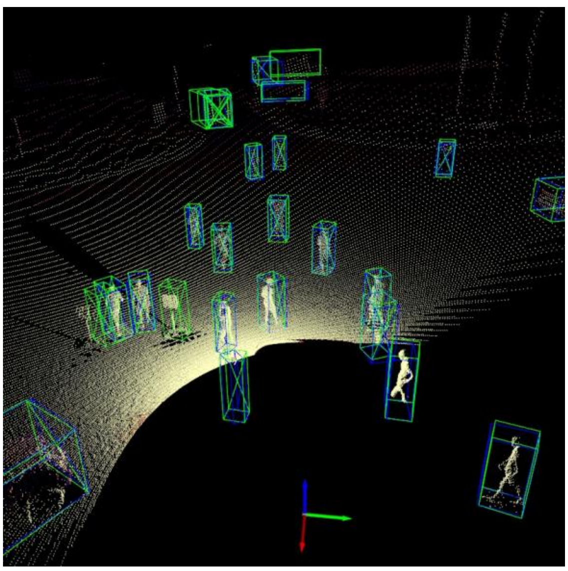

Figure 3 LiDAR Point Cloud with 3D Object Detections.

TRAINING DATA

- **Dataset description:** 3D Object Detection dataset with custom labelling using CVAT annotation tool.
- **Data splitting:** Train 70 % - Val 20% - Test 10%.
- **Data augmentation techniques:**
  - Random augmentations are applied during training. These include ground-truth sampling, random flipping along the x-axis, random rotations within ±45°, and random scaling in the range [0.95, 1.05].

EVALUATION DATA

- **Dataset description:** Test dataset prepared with custom labelling using CVAT annotation tool.

Outputs: 3D Bounding Boxes and Classes

- **Classification:** Predicted category for each detected object
- **Regression:** 3D bounding box parameters including position (x, y, z), dimensions (length, width, height), and orientation

INTENDED USE

USE CASES AND USERS

Intended use cases:

- **Urban traffic scenarios with mixed participants:** The model can detect cars, pedestrians, trucks, trailers, and two-wheelers in dense environments such as intersections, highways, and city streets, supporting collision avoidance and path planning.
- **Adverse conditions:** Since VoxelNeXt is designed for robustness, it can help perception under challenging situations like occlusions, partial visibility, or sensor noise.
- **Fleet monitoring / cooperative driving research:** It can be used in R&D settings for evaluating collaborative perception or improving autonomous driving safety.

Target audience or users of the model:

- **Self-driving car developers and researchers:** To integrate and benchmark voxel-based 3D object detection in perception stacks.
- **Automotive companies / suppliers:** To evaluate robustness in production LiDAR-based systems.

Users to be analyzed by the model:

- **Traffic participants:** Cars, pedestrians, trucks, trailers, and two-wheelers — as defined in the custom dataset.

LIMITATIONS

- **Limitations or restrictions of the intended use:**
  - The model performance depends on the characteristics of the **custom-labelled training dataset**. Generalization to very different environments (e.g., rural scenes, tunnels, or drastically different traffic distributions) may be limited.
  - The LiDAR sensor specifications (e.g., FoV, range, number of beams) should be close to the sensor used during data collection.
  - If the mounting position of the lidar sensor is changed the accuracy of the model will suffer.
- **Guidance on how to address limitations or restrictions:**
  - Use a Ouster-128 LiDAR sensor or similar.
  - Do not change the sensor mounting position.
- **Identification of potential biases:**
  - The model may be biased because of limitations of diversity in data collection where some types of scenarios and objects are not well represented

EVALUATION AND PERFORMANCE

METRICS

- **Performance metrics:** We report **3D APR40** as the sole metric. APR40 is **Average Precision computed over 40 recall positions** (R = {0.0, 0.025, ..., 0.975}) using one-to-one matching between predictions and ground truth based on **IoU thresholds**. A detection is correct if its 3D IoU with a ground-truth box exceeds the class-specific threshold. Precision–recall is computed per class and averaged over the 40 recall points to obtain APR40; **mAP** is the unweighted mean of class APR40 values.

- **Relevant thresholds or criteria: n/a**

- **Evaluation method:** We follow the KITTI evaluation scheme, including the exact definition of the APR40 Metric.

RESULTS

- **Results of the performance evaluation**

  Measured on custom dataset test split:
  mAP@R40: 84.57

Class	APR40@0.5	# Samples
Car	96.53	3218
Pedestrian	96.09	1720
Truck	99.50	937
Trailer	82.00	225
Two-wheeler	48.75	56

- **Performance on specific subsets of dataset: n/a**

- **Performance metrics interpretation: n/a**

Althena

ETHICAL CONSIDERATIONS

FAIRNESS

- **Fairness metrics:** No fairness measures or metrics were used to evaluate the model, and no data reweighting was performed during training. The model was not tested for fairness, and it is possible that it may exhibit bias against certain groups of people.
- **Remaining biases or limitations:** The reason is that the dataset does not contain additional meta data, which could be used for such analysis. The dataset contains only point clouds and 3D objects, but no information about certain groups or population.

PRIVACY

- **Data collecting and storing:** Private information, such as a person's identity, gender, or ethnicity, cannot be retrieved from a 3D point cloud with low resolution. This is because the point cloud is too sparse, meaning that there are not enough data points to accurately represent the fine details of the object/person.
  For example, a 3D point cloud of a person's face might not have enough detail to identify their individual features, such as their eyes, nose, and mouth. As a result, it would be impossible to tell who the person is or what their gender or ethnicity is.
- **Users' access to data:** Dataset is private and is not publicly available.

ACCOUNTABILITY

- **Model's decisions transparency through explainability techniques:** The model does not currently have any explainability mechanisms implemented. However, this feature will be implemented in the course of the project. Two potential explainability techniques that could be used are perturbation tests and attention maps. These techniques will be evaluated and implemented in the model in the future.
- **Reviewing and auditing processes:** n/a
- **Limitations or uncertainties:** n/a

USAGE AND LIMITATIONS

USAGE AND RECOMMENDATIONS

- **Model usage instructions:**
  The model can run **as a ROS2 node** in an automated-driving stack **or as a standalone inference service/script**. A point-cloud preprocessing step (including voxelization) is required; see the inference repository/configs for parameters (ranges, voxel sizes, thresholds)..
- **Input data format and preprocessing:**
  - Point Cloud Array of a Ouster-128 or similar
  - Fields: (X, Y, Z, Intensity)
  - The coordinates must be in the LiDAR Frame
- **Output interpretation:**
  - *The model outputs **3D bounding boxes** with **class scores**.*
  - ***Regression:** box **center (x, y, z)**, **dimensions (l, w, h)**, and **orientation (yaw)**.*
  - ***Classification:** per-box class probabilities (car, pedestrian, truck, trailer, two-wheeler, ...).*
  - *Post-processing uses **confidence thresholds** and **NMS** to filter overlapping boxes*

LIMITATIONS AND RISKS

- **Technical limitations:**
  - Domain Shift: A domain shift w.r.t. to the shape of vehicles might hurt the performance (North American Cars, European Cars); Special Types of vehicles (On Rails, Trains) might not be detected.
  - Different Sensor: The model might not be applicable to other sensor types, sensor configurations or sensor orientations; A similar input distribution as the train dataset is necessary.
  - Bad weather conditions (snow, rain, fog): might disturb the LiDAR sensor and thus the model might not be able to make reliable detections.
  - Small objects: Small objects have only a few data points on their surface, hence the detection becomes difficult or impossible.
- **Ethical or legal limitations and restrictions:**
  - The model does not process any personal/private information.
  - The identification of personal data from the input sensor data is not possible.
  - It is not necessary to perform any anonymization of data prior to the input of the model.
  - People who use this model cannot capture sensitive data from pedestrians. The sensor data is very sparse and does not capture textures. Hence, it is almost impossible to identify pedestrians or capture personal information from them.