Free Space Segmentation using Automotive Radar

Mujtaba Hassan[#*], Andras Palffy[^], Francesco Fioranelli[#], Alexander Yarovoy[#], Satish Ravindran[$], Dariu Gavrila[^]

[#]MS3 Group, EEMCS Faculty, Delft University of Technology, The Netherlands
[*]Advanced Radar Solutions, NXP Semiconductors, Germany
[^]IV Group, ME Faculty, Delft University of Technology, The Netherlands
[$]Advanced Radar Solutions, NXP Semiconductors, USA
{s.m.hassan, a.palffy, f.fioranelli, a.yarovoy, d.gavrila}@tudelft.nl, satish.ravindran@nxp.com

***Abstract*— A data driven method is proposed to obtain free space segmentation using automotive radar point clouds. It aggregates automotive radar detection points from multiple timestamps, projects them into a Birds-Eye-View grid-based representation, and applies a semantic segmentation Neural Network (NN) to classify each grid for free space segmentation. A lidar based supervision is used to generate the ground truth for training. Moreover, debris objects are manually annotated to enable the NN model to learn to detect these uncommon objects. Experimental results on a proprietary 4D Imaging Radar dataset demonstrate that the proposed method gives improved free space segmentation as compared to other baseline methods.**

***Keywords*— Free Space Segmentation, Automotive Radar, Neural Networks.**

I. INTRODUCTION

Automotive radar is an essential sensor in modern vehicles, due to its robustness in harsh weather conditions, its capacity to directly measure targets' velocity, and its cost-effectiveness. Recent developments in radar technologies, coupled with the availability of open-source radar datasets, have attracted increased attention to radar for perception tasks in autonomous driving using deep learning. Free Space Segmentation is a vital task that can allow an understanding of the drivable regions.

Techniques exist to classify free space regions using automotive radar. These can be broadly categorized into those using spectrum data and point cloud. Although using radar spectrum data provides detailed information that can help in free space segmentation [1], obtaining this is difficult and according to the authors' knowledge, an open-source radar spectrum data with annotated debris objects is not available. Hence, the focus of this work is on free space segmentation using automotive radar point clouds.

Free space segmentation is typically estimated using inverse sensor modeling (ISM) [2]. [3] obtained free space segmentation using Bresenham's line algorithm [4].

Other approaches utilize deep learning for performing free space segmentation. [1] proposed for the first time to learn the ISM function using a data-driven learning approach on radar spectrum data. [5] proposed to perform this learning based on clustered radar points. Ground truth occupancy labels are generated using lidar for both these methods [1], [5]. [6] proposed to perform semantic classification for the entire occupancy grid. [7] performed joint obstacle detection and free space segmentation using a Neural Network (NN).

Although the techniques published in literature improve free space segmentation, certain limitations exist in these approaches that are targeted in this work. These include:

- It is difficult to detect small targets, specially debris objects, e.g., plastic cones, since there can be little or no radar point detections received from these targets.
- Lidar methods to obtain annotations for debris objects are inaccurate because common lidar segmentation NN models are not trained for these objects.
- Most radar-based free space segmentation NNs do not provide an analysis on the architectures suitable for this purpose and typically use simple models.

In this paper, a signal processing pipeline to obtain free space segmentation using automotive radar is proposed for scenarios that include small debris objects. A NN based approach is used taking inspiration from [5]. Fig. 1 shows the proposed pipeline, with its three main modules. First is the ground truth generation. A pseudo lidar based method based on ray-casting is used to obtain the ground truth annotations which are further refined using manual annotations of debris objects and the output of a pretrained lidar segmentation NN. Second is the radar pre-processing. Radar points from multiple timestamps are aggregated and converted into BEV tensors with the point features as channels in the tensor. Third is the model itself. It is a semantic segmentation NN that takes these BEV radar tensors as input and outputs free space segmentation as three classes: 'occupied', 'free', 'unobserved'.

The proposed method is evaluated on a proprietary 4D Imaging radar dataset and compared with three baseline approaches: traditional ray-casting method and two NN based methods [5] and [7]. The proposed solution outperforms these baseline methods, showing an increase in mean Intersection-over-Union of 14.5%, 5.6%, 13.2%, respectively. The main contributions of the paper are as follows:

- Designing a NN-based free space detector using automotive radar point clouds which shows a significant improvement in mean IoU as compared to baselines.
- A pseudo lidar based annotation approach using ray-casting that takes the output of a lidar NN segmentation model to refine ground truth annotations.
- Ablation studies to illustrate the impact of different neural network architectures and frame aggregation.

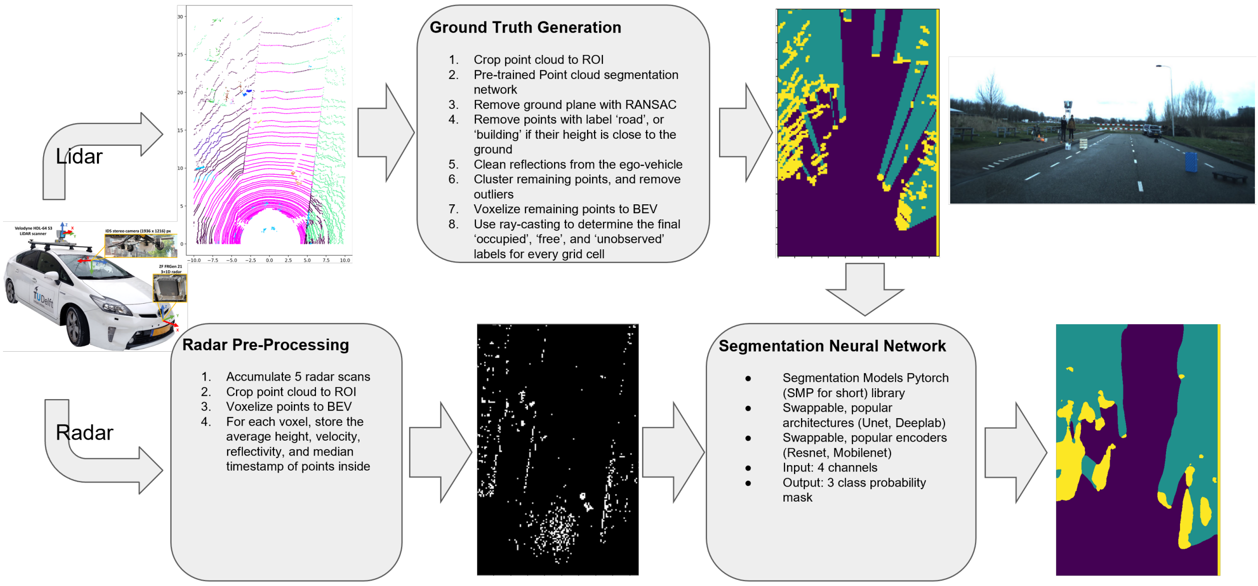

Fig. 1. Proposed Pipeline for free space segmentation using automotive radar.

The rest of the paper is organized as follows. Section II illustrates the proposed approach. Section III describes the experiments performed with an analysis on the results. Section IV concludes this work with a statement on future directions.

II. Proposed Approach

The three main components of the approach are described.

1) Ground Truth Generation

Manual annotations of free space is a cumbersome task as each region needs to be annotated. A lidar based ray-casting approach is thus used for generating the ground truth [1], [5]. The idea is to remove noise points and apply Bresenham's line algorithm [4] to obtain accurate free space annotation. However, unlike [1], [5], a pretrained lidar segmentation NN [8] is used in this work to refine the ground truth generation. The top-performing SPVNAS variant is chosen as the model, pre-trained on large-scale SemanticKITTI dataset [9].

Lidar data is in the format of a point cloud with 4 values per point: the spatial location (x,y,z) and intensity (i) of lidar reflections. The spatial location is used to crop the point cloud to Region Of Interest (ROI). Then, the remaining point cloud is processed through a pretrained point cloud segmentation network which gives an estimate of the segmentation value for each point. It must be noted that this is an automatic process, as the lidar model is not trained on data used in this work.

Lidar data contain many reflections from the ground which need to be removed. The lidar segmentation network already gives an estimate of the points generated from the road. Moreover, a ground plane estimation technique is used to remove these points. Here, RANSAC [10] is applied to find the ground points to be removed from the point cloud. Also, points labeled as 'road' and those with a 'building' label and a height close to the ground are removed. Reflections received from the ego vehicle are also removed from the point cloud. Noise points are removed by first clustering all the points using an algorithm such as DBSCAN [11], and afterwards removing the points which do not belong to any cluster.

Once the unnecessary points are removed, the remaining points are voxelized into a Birds-Eye-View (BEV) projection. This allows the ground truth to be in the form of a grid so that segmentation NN models can be used to determine the free space areas. Afterwards, a ray-casting method based on Bresenham's line algorithm [4] is used to obtain the final labels which include three regions 'occupied', 'free', 'unobserved'. Occupied regions include grids where a point detection is found; free regions include areas which are free to drive; unobserved regions represent regions behind the line of sight of the first lidar reflection which cannot be observed by lidar. Additionally, regions corresponding to non-reflective objects that yield no lidar returns are also treated as free space.

Moreover, debris objects are also annotated manually to ensure they are not missed. This is because these are uncommon objects which may not be detected by the lidar segmentation NN model and may be removed as noise points during the clustering process. The grid cells containing debris objects are therefore labeled as 'occupied'.

2) Radar Pre-Processing

The automotive radar used in this work is a 4D Imaging radar point cloud which has 5 features: spatial location (x,y,z), radar cross section (RCS) and velocity (v_r). Since the radar point cloud is very sparse, data is accumulated for a set of T frames. This allows to obtain a denser point cloud and is reasonable since most of the region in space contains static objects which will not change location for different timestamps. However, to maintain latency, T is limited to 5 frames which corresponds to 0.5 s. The point cloud is cropped to a RoI for evaluation on only these regions. The remaining points are voxelized into a BEV grid so that a segmentation NN can be applied. For each grid, RCS, velocity, height, and timestamp is stored as the features.

3) *Segmentation Neural Network*

Obtaining the input and ground truth in the form of grid allows usage of a grid-based segmentation NN to classify each grid as 'free', 'occupied' or 'unobservable'. Here, DeepLab [12] is used based on evaluation of different available network architectures. This is a popular architecture for semantic segmentation that can obtain multi-scale features at fine resolution. Since the dataset contains highly imbalanced class labels with different pixel counts, robust Dice loss [13] is used for training to use the class re-balancing properties of the Generalized Dice Overlap for obtaining accurate performance.

III. Experimental Results and Discussion

A. *Dataset and Performance Metrics*

A proprietary 4D Imaging Radar was used to assess the performance of the method. 4D Radar gives fine resolution and height information that can help in detecting small objects and enhancing free space segmentation. A ego-vehicle setup similar to the VOD Dataset [14] was used for data collection. Most scenes include the ego-car moving towards different debris objects which are placed in different locations for different scenes. This resulted in a total of 24 scenes with 22 training and 2 testing scenes, resulting in a total of 10017 train and 815 test frames respectively. A total of 16 different debris objects were annotated to provide variety to the dataset. Fig. 2 shows the number of points detected for different debris objects. It can be observed that most objects generate few detections, especially the smaller objects such as plastic cones. However, in automotive context it is useful to detect these objects for obtaining a better perception of the environment.

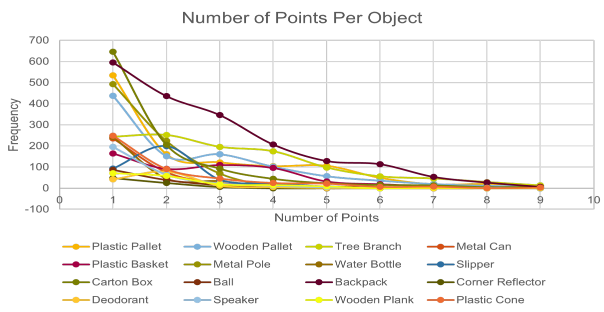

Fig. 2. Frequency of the number of detections for different debris objects.

B. *Evaluation Metrics*

Performance of different methods is evaluated using Intersection-over-Union (IoU) per class metric as in [5]. This provides a detailed evaluation of performance for individual classes and helps in addressing the class imbalance problem. Moreover, a mean IoU value is also calculated to give a single performance measure. Equations (1), (2) give the formula for obtaining these metrics.

$$IoU_c = \frac{TP}{TP + FN + FP} \tag{1}$$

$$Mean\_IoU = \sum_{c=1}^{c=C} IoU_c \tag{2}$$

where c denotes Class; TP denotes True Positives; FN denotes False Negatives; FP denotes False Positives.

C. *Results*

1) *Comparison with other methods*

The proposed method is compared with three baseline methods. The first method used ray-casting where a similar procedure as done for ground truth generation is used. However, the removal of points based on the segmentation NN model is not performed. The other two baselines are NN-based corresponding to a pioneer free space segmentation method in [5], named OccupancyNet and a recent obstacle & free space segmentation method in [7], named Nvradarnet.

Table 1 shows performances of the different segmentation methods averaged over all test frames. The proposed method outperforms all baseline methods. Specifically, an improvement in mean IoU of 14.5%, 5.6%, and 13.2% is achieved as compared to ray-casting, OccupancyNet [5] and Nvradarnet [7], respectively. Compared to ray-casting, the proposed NN-based method is able to successfully segment regions by removing noise points from obstacles. Also, compared to other NN methods, the usage of DeepLab NN architecture results in obtaining useful features for improved free space segmentation.

Table 1. Comparison between different free space segmentation methods

IoU	Occupied	Free	Unobserved	Mean IouU
Proposed Method	55.07	73.77	32.29	53.71
Ray-casting	32.77	59.62	23.12	39.16
OccupancyNet [5]	43.79	67.24	33.11	48.05
Nvradarnet [7]	35.94	63.17	22.25	40.45

Figure 3 shows visualization of the output of proposed method on an interesting scene. Here, debris objects are present on the right and left corners of the road. The proposed model is able to achieve accurate free space segmentation by classifying the parts of road containing debris objects as occupied and the middle section of the road as free. Note also that the lidar based ground truth is not perfect: some parts of road are not labeled as 'free' because some ground points are not removed. Also, some unobserved cells are labeled as 'occupied'. This partly explains the relatively lower IoU for unobserved classes.

2) *Impact of NN Architectures*

An investigation of the impact of different NN architecture on free space segmentation was also performed. For this, segmentation architectures including DeepLab [12], UNet [15], FPN [16] and Segformer [17] were used. Table 2 provides a comparison of the results achieved using them. It is observed that using different architectures results in varying performance, and that the choice of correct architecture has an impact on results. DeepLab architecture 2 used in this work performs best. This is due to the dilated convolutions and spatial pyramid pooling layers which gives fine resolution multi-scale features to help in free space segmentation.

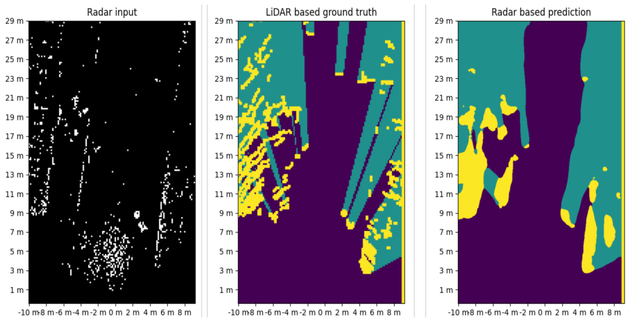

Fig. 3. Visualization of the performance of free space segmentation method for a selected scene: radar input (left column); ground truth (middle column); predicted output (right column). The proposed method can correctly detect the debris objects present on the corners of the road and mark them as occupied regions. Here, yellow denotes occupied; purple denotes free; cyan denotes unobserved regions.

Table 2. Impact of NN Architectures

IoU (%)	Occupied	Free	Unobserved	Mean IouU
DeepLab	55.07	73.77	32.29	53.71
UNet	42.45	69.65	23.37	45.16
FPN	37.36	69.96	27.28	44.86
Segformer	32.74	65.60	21.35	39.89

3) *Impact of Frame Aggregation*

An investigation on the impact of frame aggregation on performance is also performed. In this respect, the aggregated frames were changed from T=1 to T=5 frames. Table 3 shows the comparison of the performance using different number of aggregated frames. It can be observed that adding frames improves IoU especially for 'occupied' class. This happens as it results in more points to help in detection of smaller objects.

Table 3. Impact of Frame Aggregation

IoU (%)	Occupied	Free	Unobserved	Mean IouU
T=5	55.07	73.77	32.29	53.71
T=4	39.16	70.13	23.60	44.30
T=3	46.52	51.80	14.32	42.03
T=2	27.42	65.97	17.82	37.07
T=1	23.65	68.57	4.45	32.22

IV. Conclusion

A novel NN-based method to obtain free space segmentation using automotive radar point cloud is proposed. Ground truth is generated using a refined lidar ray-casting method. The proposed method showed an increase in free space segmentation performance compared to three baselines; one using traditional ray-casting and the other two using NNs. Moreover, it was shown that the choice of architecture has an impact on performance. Lastly, it was shown that aggregating multiple radar frames improves the method performance.

Although an improved free space segmentation is achieved, this method used radar point cloud as the input because of the unavailability of radar spectrum data. This results in a limitation for small objects with no point detections. As part of future work, it is planned to address this issue by collecting radar spectrum data and developing a method using this input.

Acknowledgment

This work is part of the IPCEI ME/CT and is funded by the European Union Next Generation EU, the German Federal Ministry for Economic Affairs and Climate Action, the Bavarian Ministry of Economic Affairs, Regional Development and Energy, the Free State of Saxony with the help of tax revenue based on the budget approved by the Saxon State parliament and the Free and Hanseatic City of Hamburg.

References

[1] R. Weston, S. Cen, P. Newman, and I. Posner, "Probably unknown: Deep inverse sensor modelling radar," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5446–5452.

[2] S. Thrun, "Probabilistic robotics," *Commun. ACM*, vol. 45, no. 3, 52–57, Mar. 2002.

[3] M. Li, Z. Feng, M. Stolz, M. Kunert, R. Henze, and F. Küçükay, "High resolution radar-based occupancy grid mapping and free space detection," in *International Conference on Vehicle Technology and Intelligent Transport Systems*, 2018.

[4] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems Journal*, vol. 4, no. 1, pp. 25–30, 1965.

[5] L. Sless, B. E. Shlomo, G. Cohen, and S. Oron, "Road scene understanding by occupancy grid learning from sparse radar clusters using semantic segmentation," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 867–875.

[6] R. Prophet, A. Deligiannis, J.-C. Fuentes-Michel, I. Weber, and M. Vossiek, "Semantic segmentation on 3d occupancy grids for automotive radar," *IEEE Access*, vol. 8, pp. 197 917–197 930, 2020.

[7] A. Popov, P. Gebhardt, K. Chen, and R. Oldja, "Nvradarnet: Real-time radar obstacle and free space detection for autonomous driving," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 6958–6964.

[8] H. Tang, Z. Liu, S. Zhao, *et al.*, "Searching efficient 3d architectures with sparse point-voxel convolution," in *European Conference on Computer Vision*, 2020.

[9] J. Behley, M. Garbade, A. Milioto, *et al.*, "Towards 3D LiDAR-based semantic scene understanding of 3D point cloud sequences: The SemanticKITTI Dataset," *The International Journal on Robotics Research*, vol. 40, no. 8-9, pp. 959–967, 2021.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, 381–395, 1981.

[11] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, 1996, pp. 226–231.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision – ECCV 2018*, Cham: Springer International Publishing, 2018, pp. 833–851.

[13] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, M. J. Cardoso, T. Arbel, G. Carneiro, *et al.*, Eds., Cham: Springer International Publishing, 2017, pp. 240–248.

[14] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1d radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4961–4968, 2022.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241.

[16] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6392–6401.

[17] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, ser. NIPS '21, 2021.