Annotating Automotive Radar efficiently: Semantic Radar Labeling Framework (SeRaLF)

Simon T. Isele[1,3,4]*, **Marcel P. Schilling**[1,2]*, **Fabian E. Klein**[1], **J. Marius Zoellner**[3,4]

[1]Dr. Ing. h.c. F. Porsche AG, Weissach, Germany
{simon.isele1,fabian.klein1}@porsche.de
[2]Institute for Automation and Applied Informatics
[3]Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute of Technology, Karlsruhe, Germany
marcel.schilling@kit.edu
[4] FZI Research Center for Information Technology, Karlsruhe, Germany
zoellner@fzi.de

Abstract

Research on localization and perception for Autonomous Driving (AD) is mainly focused on camera and LiDAR data sets, rarely on radar data. Transferring data-driven development processes to the radar domain, the obvious need for radar data set generation is the first step to take. We propose the modular twofold cross sensor Semantic Radar Labeling Framework (SeRaLF) for the automated offline generation of semantic labels for radar raw detections. First, a CNN is applied to a 360 degree view of camera to generate semantic information in the near-field of the scene. In parallel, a roof mounted LiDAR, covering both near and long range, serves as alternative perception source on which a CNN is applied to semantically segment the LiDAR point cloud data. Considering different mounting positions and Field of Views (FoV), the labels of optical perception are radar-associated via a projection into 2D image representation. In a subsequent fusion algorithm, each detection is associated with a final label describing its semantic class. The fusion algorithm considers consistency of both labels as well as uncertainties of the deep neural networks based on Monte Carlo Dropout. The automated pipeline is tested on real world data measurements. Subsequently, the proposed semantic radar labels are subject of a semi-manual inspection step to correct erroneous labeled points.

1 Introduction

Environmental perception is a key challenge in the research field of AD and mobile robots. The development of an automated labeling process based on vehicle sensors enables a competitive, unbiased, and efficient data-enrichment pipeline to succeed in this field. Therefore, we aim to boost the potential of radar sensors. Radar sensors are simple to integrate and reliable also in adverse weather conditions [1]. They provide 3D coordinates and additional information about raw detections, e.g. signal power or relative velocity. Both sparsity [2] and characteristic artifacts [3] pose challenges for the perception task. Furthermore, the lack of completely annotated publicly available data sets of radar data limits research on data-driven approaches for radar [4]. Classically, Adaptive Cruise Control (ACC) [5] and state-of-the-art object detection rely on radar, e.g. [2]. But, to the authors

*Both authors contributed equally

Machine Learning for Autonomous Driving Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

best knowledge, radar raw signals are rarely used in direct 3D fashion for AD or Advanced Driver Assistance Systems (ADAS).

In this workshop, we utilize the on-board sensors in a generic method to automatically annotate radar detections. The Semantic Radar Labeling Framework (SeRaLF) enables a competitive, unbiased and efficient data-enrichment pipeline as an automated offline process to generate comparable radar semantic data sets. Our framework SeRaLF applies the benefits of cross-modal sensors and is composed of two pipelines: camera and LiDAR processing. Piewak [6] describes the availability of a large amount of automatically labeled data as valuable start to work on machine learning approaches.

Our key **contributions** are the following:

1. A joint method to label real-world radar data from image processing of surround view cameras as well as LiDAR scans.
2. A fusion proposal for radar label predictions from independent camera and LiDAR semantic segmentation CNNs, considering label consistency and epistemic uncertainty of each CNN.
3. A simplified semi-manual annotation procedure using predictions of SeRaLF.

2 Related Work

To elaborate machine learning on radar, publicly available data sets comparable to KITTI [7] or Waymo Open [8] lack radar raw detections. Interest to work on radar point cloud data gains attention since the release of nuScenes [9] and Astyx [10] data set. As the only two available data sets containing both radar detections and object instances respectively, they provide only 2D objects [9], consider only front facing view [10], and do not focus on semantic labels for each detection.

Investigations e.g. on semantic scene understanding of radars by the means of neural networks in supervised fashion e.g. [11, 4, 12] or other radar applications for perception in AD, require currently expensive, non-scaleable manually labeled data sets of raw data [4].

For example the Convolutional Neural Network (CNN) HarDNet [13] yields very promising results on camera data sets like Cityscapes [14]. The release of SemanticKITTI [15] boosted LiDAR based semantic scene understanding, resulting in architectures as CNN RangeNet++ [16]. The transfer from regular structures as image or LiDAR data to the radar domain introduces to deal with unordered data. O. Schumann et al. [11] investigate semantic segmentation on radar data based on recurrent neural networks and random forests. In continuation, Schumann et al. [12] also propose a radar semantic segmentation approach using an adapted PointNet++ [17]. In addition to the above mentioned approaches for semantic scene understanding, also other input forms such as birds-eye-view fashion are applied for CNNs e.g. [4]. However, such approaches have in common that labeled data is required e.g. [11, 4, 12].

Automatizing the labeling pipeline for LiDAR point cloud data sets by the means of camera processing, Piewak et al. [6] achieve improvements of model training for semantic segmentation tasks on LiDAR point clouds. With the application of deep learning models, uncertainty estimation [18, 19] gains importance to detect failures or other limitations of the models.

3 Method

The proposed framework in Figure 1 consists of parallel camera and LiDAR semantic segmentation CNNs. The fusion considers each CNNs' uncertainty and label consistency of the independent label proposals for each radar detection. SeRaLF is implemented in the Robot Operating System [20].

Firstly, radar detections from various sensors are transformed into a common reference coordinate system and concatenated to one point cloud consisting of points $\mathbf{p}_{i,t} \in \mathbb{R}^3$ at time t. To enable comparison of the multi-modal sensors in parallel segmentation branches, time synchronization [21] is necessary. After fusion, SeRaLF annotates each detection $\mathbf{p}_{i,t}$ with a proposed semantic class label $\hat{y}_{\text{sem}}(\mathbf{p}_{i,t})$. To simplify the problem of annotating radar detections with semantic labels, the number of classes in Cityscapes [14] and SemanticKITTI [15] are clustered to a reduced label set $\hat{y}_{\text{sem}}(\mathbf{p}_{i,t}) \in \{$ `flat`, `human`, `vehicle`, `construction`, `nature`, `pole`, `unknown` $\}$.

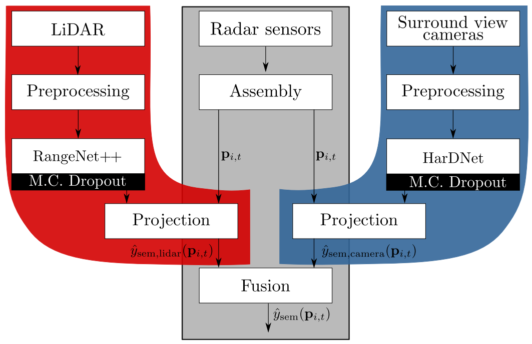

Figure 1: SeRaLF (red: LiDAR branch, blue: camera branch).

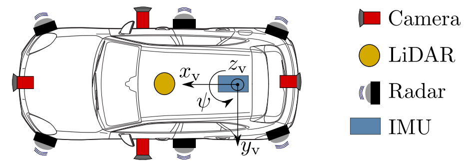

Figure 2: Sensor setup.

Table 1: Sensor setup details.

Name	Details
Camera	4 x Monocular surround view camera (series equipment)
LiDAR	Rotating Time-of-Flight LiDAR (centrally roof-mounted, 40 channel)
Radar	77 GHz FMCW Radar (160 degree hor. FoV, ±10 degree hor. FoV)

Camera branch In order to apply a state-of-the-art segmentation CNN on fisheye camera images, image preprocessing is necessary to reduce the domain shift from training images to the applied domain. We propose to apply distortion correction using OCamCalib [22], and a perspective transform subsequently to change the FoV to a straight view on the scene. Applying this preprocessing, the CNN HarDNet [13], which is originally trained on the Cityscapes data set [14], yields sound segmentation results. Considering the camera in- and extrinsics, each radar detection $\mathbf{p}_{i,t}$ is projected into the corresponding camera image from which $\hat{y}_{\text{sem,camera}}(\mathbf{p}_{i,t})$ is derived. Note, `sky` is an image segmentation label which is not appropriate to radar objects. Thus, in case a radar detection is projected onto a `sky` area or not visible in a camera image the radar label is changed to `unknown`.

LiDAR branch For the purpose of using RangeNet++ [16], pretrained on SemanticKITTI [15], to segment LiDAR scans, we apply the proposed preprocessing to generate an image-like LiDAR representation as input for the CNN inference. In the following, sensor in- and extrinsics help to project each radar detection $\mathbf{p}_{i,t}$ into the corresponding LiDAR image. Hence, $\hat{y}_{\text{sem,lidar}}(\mathbf{p}_{i,t})$ can be obtained. Similar to the procedure in the camera branch, detections out of direct LiDAR FoV or objects inside the LiDAR blind spot close to the vehicle, receive the label `unknown`. Same applies for radar objects e.g. obscured by other objects in the foreground and only visible for radar sensors, not for visual sensors.

Fusion To select the applicable label, $\hat{y}_{\text{sem,lidar}}(\mathbf{p}_{i,t})$ and $\hat{y}_{\text{sem,camera}}(\mathbf{p}_{i,t})$ are fused. The Monte Carlo Dropout (MCD) from Kendall et al. [18, 19], yields an epistemic or model uncertainty estimate of the CNNs, which we use to detect failures. By randomly deactivating certain neurons in repeated inference steps on a single sample, MCD proposes small deviations among the inference outputs to indicate low CNN model uncertainty. In contrast, the aleatoric uncertainty characterizes the variation due to the input data of camera and LiDAR. However, following the argument of Kiureghian et al. [23] that for ideal models aleatoric uncertainty merges into epistemic uncertainty, a clear distinction remains open. Experiments with MCD proofed the difficulty to seperate model from input uncertainty.

By thresholding the results of MCD in Figure 1, we obtain binary images to describe whether a prediction is reliable or not. Comparing the independent labels plus their reliability score, the rule based final fusion decides the resulting annotation $\hat{y}_{\text{sem}}(\mathbf{p}_{i,t})$. If there is only one semantic prediction not equal to `unknown`, we adopt it. Moreover, consistent predictions of both sensors increase a label's reliability. In case of label conflicts, the more reliable sensor overrules. At conflicting labels with equal reliability, the label `unknown` is considered. Other rule-based approaches are pending.

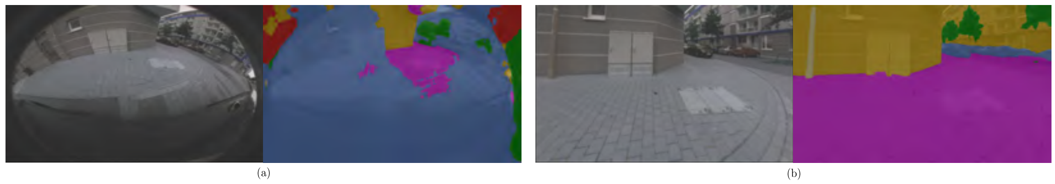

Figure 3: Results of semantic segmentation on fisheye camera (a) and after preprocessing (b).

Figure 4: Results of semantic segmentation on LiDAR image (a) and scene overview (b).

Semi-manual labeling The labels of the automated pipeline SeRaLF are not perfectly correct. To obtain ground truth labels for evaluation of algorithms, we transfer the predictions of SeRaLF to the point labeler [15] described as an initial guess. Thus, to correct a few falsely predicted, mostly overestimated labels is more comfortable compared to a labeling from scratch. Furthermore, Piewak et al. [6] demonstrate that automatically generated data, also containing errors, help to boost the training of deep neural networks. Even though errors may be learned by training on loosely labeled data sets, a fine-tuning with correct ground-truth data overcomes this errors [6].

4 Experiments

Sensor setup and experimental design We apply our approach on real world data and hardware. The vehicle test setup is introduced in Figure 2 and sensor set details are found in Table 1. We evaluate SeRaLF on three reference test tracks (urban area, small village, industrial park) in qualitative fashion.

Results Figure 3 shows the potency of our proposed preprocessing of raw fisheye images to feed regular image CNNs. Moreover, Figure 4 illustrates promising LiDAR results on the depicted scene.

Figure 5 shows the results of the binarized uncertainty images. Promising results in Figure 5 (c) are indicated as reliable blank areas. In image regions of Figure 5(a) in which by manual inspection, or in Figure 5(b) for unknown scenes, one expects higher uncertainty, MCD correctly predicts uncertainty in form of black pixels. Thus, we rely on MCD to detect failure of both CNNs. Figure 6 illustrates the results of the annotated radar data after fusion. The rich visualization in Figure 6 with LiDAR underlay facilitates the subsequent manual inspection and correction of misinterpreted labels. Note, that the radar raw point cloud is shown in Figure 6. Since Holder et al. [3] describe radar artifacts to lead to wrong labels, a combination with de-nosing of radar data is recommended.

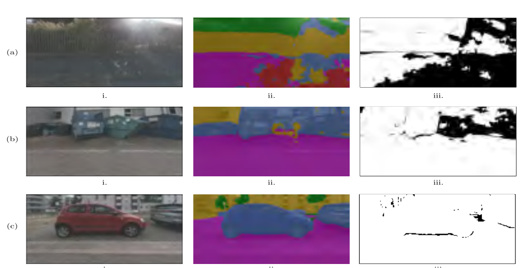

Figure 5: Input (i.), semantic segmentation (ii.) and binarized uncertainty (white: reliable, black: uncertain, iii.) on scenes (a-c).

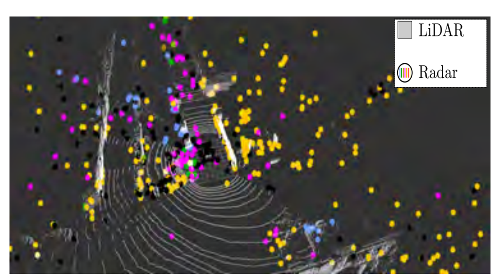

Figure 6: Results of semantic radar labels after fusion of scene depicted in Figure 4 (b).

5 Conclusion

SeRaLF provides a generic tool to annotate radar data with semantic labels from camera and LiDAR sensor. This is achieved by fusing the predictions of the camera and LiDAR branch, considering consistency between the branches predictions, and the epistemic uncertainty of each CNN evaluated by the MCD procedure. However, the annotation accuracy highly depends on the results of the two branches and may contain errors. Providing increased data sets for initial training of neural networks leverages generalization. SeRaLF helps to accelerate the generation of ground data utilizing a subsequent semi-manual annotation correction procedure. Furthermore, SeRaLF enables an enrichment of existing data sets like nuScenes [9] with semantic labels.

References

[1] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*, 8:58443–58469, 2020.

[2] Di Feng, Christian Haase-Schutz, Lars Rosenbaum, Heinz Hertlein, Claudius Glaser, Fabian Timm, Werner Wiesbeck, and Klaus Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, page 1–20, 2020.

[3] Philipp Rosenberger Christoph Popp Prof. Dr. Hermann Winner Martin F. Holder, Clemens Linnhoff. Modeling and simulation of radar sensor artifacts for virtual testing of autonomous driving. In *9. Tagung Automatisiertes Fahren*, München, 2019. Lehrstuhl für Fahrzeugtechnik mit TÜV SÜD Akademie.

[4] O. Schumann, J. Lombacher, M. Hahn, C. Wöhler, and J. Dickmann. Scene understanding with automotive radar. *IEEE Transactions on Intelligent Vehicles*, 5(2):188–203, 2020.

[5] L. H. Eriksson and B. . As. Automotive radar for adaptive cruise control and collision warning/avoidance. In *Radar 97 (Conf. Publ. No. 449)*, pages 16–20, Oct 1997.

[6] Florian Piewak, Peter Pinggera, Manuel Schäfer, David Peter, Beate Schwarz, Nick Schneider, Markus Enzweiler, David Pfeiffer, and J. Zöllner. *Boosting LiDAR-Based Semantic Labeling by Cross-modal Training Data Generation: Munich, Germany, September 8-14, 2018, Proceedings, Part VI*, pages 497–513. 01 2019.

[7] Andreas Geiger, Philip Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32:1231 – 1237, 2013.

[8] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. pages 2443–2451, 06 2020.

[9] Holger Caesar, Varun Kumar Reddy Bankiti, Alex Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving, 03 2019.

[10] M. Meyer and G. Kuschk. Automotive radar dataset for deep learning based 3d object detection. In *2019 16th European Radar Conference (EuRAD)*, pages 129–132, 2019.

[11] O. Schumann, C. Wöhler, M. Hahn, and J. Dickmann. Comparison of random forest and long short-term memory network performances in classification tasks using radar. In *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6, 2017.

[12] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler. Semantic segmentation on radar point clouds. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 2179–2186, 2018.

[13] P. Chao, C. Kao, Y. Ruan, C. Huang, and Y. Lin. Hardnet: A low memory traffic network. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3551–3560, 2019.

[14] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. 06 2016.

[15] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9296–9306, 2019.

[16] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss. Rangenet ++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019.

[17] C. R. Qi, L. Yi, H. Su, and L. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NIPS*, 2017.

[18] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 5574–5584. Curran Associates, Inc., 2017.

[19] A. Loquercio, M. Segu, and D. Scaramuzza. A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2):3153–3160, 2020.

[20] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Ng. Ros: an open-source robot operating system. volume 3, 01 2009.

[21] Josh Faust and Vijay Pradeep. Message filter: approximate time. `http://wiki.ros.org/message_filters/ApproximateTime`, 2020. Access: 10.05.2020.

[22] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. A toolbox for easily calibrating omnidirectional cameras. 10 2006.

[23] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105 – 112, 2009. Risk Acceptance and Risk Communication.