Deep Learning-based Object Classification on Automotive Radar Spectra

Kanil Patel[1,2], Kilian Rambach[1], Tristan Visentin[3,4], Daniel Rusev[3,4], Michael Pfeiffer[1], Bin Yang[2]

[1]Bosch Center for Artificial Intelligence, Renningen, Germany
[2]Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Germany
[3]Robert Bosch GmbH, Renningen, Germany
[4]KIT Faculty of Electrical Engineering and Information Technology, Karlsruhe, Germany

***Abstract*—Scene understanding for automated driving requires accurate detection and classification of objects and other traffic participants. Automotive radar has shown great potential as a sensor for driver assistance systems due to its robustness to weather and light conditions, but reliable classification of object types in real time has proved to be very challenging. Here we propose a novel concept for radar-based classification, which utilizes the power of modern Deep Learning methods to learn favorable data representations and thereby replaces large parts of the traditional radar signal processing chain. We propose to apply deep Convolutional Neural Networks (CNNs) directly to regions-of-interest (ROI) in the radar spectrum and thereby achieve an accurate classification of different objects in a scene. Experiments on a real-world dataset demonstrate the ability to distinguish relevant objects from different viewpoints. We identify deep learning challenges that are specific to radar classification and introduce a set of novel mechanisms that lead to significant improvements in object classification performance compared to simpler classifiers. Our results demonstrate that Deep Learning methods can greatly augment the classification capabilities of automotive radar sensors.**

I. Introduction

Autonomous vehicles rely on multiple sensors to obtain a reliable understanding of their environment. Simple localization of potential obstacles in the vehicle's path is insufficient; instead, a semantic understanding of the world in real time is crucial to take into account possible reactions of identified road users and to avoid unnecessary evasive/emergency brake maneuvers for harmless objects. At present, there is a strong focus on imaging sensors for scene understanding, because high-resolution color images contain substantial information that allows object classification [1]. However, vision is severely limited in difficult light or weather conditions, and automotive radar provides a particularly useful complementary source of information [2]. Typically, radar processing chains extract radar reflections and identify object classes by the shape of the point-cloud of reflections belonging to the same object [3].

This approach works well for larger objects such as cars, but distinguishing many object classes from small sparse point clouds has proven to be challenging. Here we propose that one reason for this difficulty is the loss of a substantial amount of information characteristic for the object type at the stage of point-cloud representation. It is therefore advantageous to base radar classification on a more informative data representation. For computer vision, a similar trend has been observed in recent years and by far the most successful approach has been the introduction of Deep Learning methods [4] for object classification, object detection, and semantic segmentation. Deep Learning is able to learn favorable representations of raw input data in deeper layers of neural networks, which capture the crucial features necessary for object classification, but also exhibit invariances to viewpoints, light conditions, noise, and other transformations.

Radar spectra generated by multi-dimensional Fast Fourier Transform (FFT) not only preserve all information available in the raw signal but also yield a data representation on which powerful Deep Learning methods such as Convolutional Neural Networks (CNNs) can be applied in a very similar fashion as in vision. In this article, we identify a number of radar-specific challenges that require adaptations of the neural network input, and which lead to significant improvements in classification performance. In particular, we observe that combining the spectra with information about the range and direction of arrival (DOA) of the object is beneficial for classification. Furthermore, integration of classification results over time can lead to substantial improvements.

In order to demonstrate our results, a real-world dataset was recorded in which multiple static objects were placed on a test track and recordings were done with an automotive radar sensor mounted in a vehicle driving between the different objects. The objects can be expected to be found on real streets and are relevant for scene understanding.The results demonstrate that prediction of object classes in real time with neural networks works reliably for all classes, and filtering classification results over time can greatly improve the performance.

A. Related Work

Unlike in computer vision, applying Deep Learning to radar is still at an early stage. The most direct application of CNNs is possible on occupancy radar grids [5], [6] and images generated via Synthetic Aperture Radar (SAR)[7], [8] for remote sensing. However, both require long integration times to first generate a map before Deep Learning can be applied, and are thus difficult to apply in rapidly changing environments, which is the default case for automotive applications.

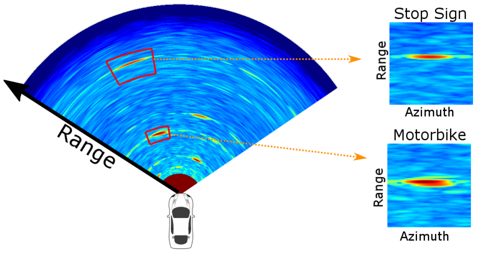


Fig. 1. Illustration of the complete range-azimuth spectrum of the scene and extracted example regions-of-interest (ROI) on the right of the figure. The figure depicts 2 of the detected targets in the field-of-view

Most approaches for automotive object classification work with radar reflections, which first requires applying a statistical detection algorithm (e.g. Constant False Alarm Rate (CFAR)) in order to curtail the information of the power spectrum to a set of detection points. Reflections belonging to the same object are then typically grouped via clustering algorithms, before classifying based on the shape of the resulting point cloud. An alternative, deep learning based, approach was recently presented by [9], which yields a semantic segmentation of the point cloud, meaning that a potential class label is assigned to every detected reflection. Their approach, however, does not resolve individual instances, but merely provides an indication of how many reflections belong to a certain class in a given scene, and where those classes are located. Overall, point-cloud based methods work well for distinguishing object classes with distinctive shapes, but for harder tasks they are limited by the loss of information due to CFAR detection.

In [10] the processing of raw radar spectra for pedestrian detection was suggested, but no machine learning was applied. Deep Learning methods that work on the radar spectrum after multi-dimensional FFT have been successfully applied in tasks such as human fall detection [11], human pose estimation [12], [13] and human-robot classification [14]. These approaches operate on the full radar spectrum, whereas our approach first extracts ROIs, which are classified by a CNN. To the best of our knowledge, this is the first Deep Learning based approach for automotive radar spectra that allows classification of multiple objects within a real-world scene.

II. Setup and Methodology

A. Experimental Setup

The goal of our study is to recreate a realistic scenario for classification with automotive radar sensors. A radar sensor is mounted on the front bumper of the test vehicle which drives between different objects on a test track, approaching them from several different directions. As objects we selected a single instance of each of the following seven different objects (commonly found in urban scenarios): car, construction barrier, motorbike, baby carriage, bicycle, garbage container, and stop sign. Although these objects are visually easy to distinguish, they pose a greater challenge for classification algorithms when working in the radio frequency spectrum. Furthermore, the scene is static and thus does not allow identifying the objects solely through the Doppler spectrum. For radar, dynamic objects with different Doppler spectra are easier to classify, due to their micro Doppler signature, but harder to record and annotate. For example, a moving bicycle with different moving parts may have an idiosyncratic signature in the Doppler domain, which would facilitate classification from permanent static objects such as stop signs.

For every object, the Differential GPS (DGPS) position is measured. During the measurement, the DGPS position of the test vehicle as well as its velocity is recorded. This allows computing the relative coordinates of the different targets with respect to the radar sensor, i. e. the range r, the relative radial velocity v, and the DOA (azimuth angle) ϑ which serves as the ground truth in the following data processing.

The radar system is a multiple input multiple output (MIMO) radar. The carrier frequency is $77\,\mathrm{GHz}$ and the bandwidth is $1\,\mathrm{GHz}$. It uses a chirp sequence modulation, i. e. a sequence of frequency modulated continuous wave (FMCW) chirps. The measurement time of one coherent processing interval is $15\,\mathrm{ms}$. The fully polarimetric sensor is described in detail in [15], [16]. We only use 4 transmitting (Tx) and 4 receiving (Rx) horizontally polarized antennas. The resulting virtual array of the MIMO radar is a linear array of 16 antennas with an aperture of $8.5\ \lambda$, with λ being the carrier wavelength. The cycle time of the radar system is approximately $57\,\mathrm{ms}$.

B. Data preprocessing

The data preprocessing consists of the following steps: first, a range-velocity spectrum is computed via a 2D-FFT. A non-coherent integration of the range-velocity spectrum is performed, and an ordered statistics constant false alarm detector (OS-CFAR) [17] is used to detect potential targets. For every range-velocity bin, the azimuth spectrum is calculated and magnitude is taken, which results in a 3D range-velocity-azimuth spectrum. For each *detected* object, we cut out a region-of-interest (ROI) of the 3D-spectrum with the range, velocity and azimuth extent of $5\,\mathrm{m}$, $0.7\,\mathrm{m/s}$, and $0.5\,\mathrm{rad}$[1], respectively, where the highest detected peak of the object is in the center of the ROI. The ground truth information is combined with the preprocessed data in order to automatically label each object.

In this study we are mainly interested in the range-azimuth spectrum, therefore we take the velocity slice which contains the maximum intensity in the 3D ROI which results in a 2D ROI containing 64 range and 66 azimuth bins. See Fig. 1 for examples of ROIs from the range-azimuth spectrum.

The dataset is split into independent training, validation and test sets. Training trials use data recorded from horizontal and diagonal driving patterns through the test track. The test set, for evaluation, follows unique and special driving patterns involving a series of curves through the test track. Therefore, the test dataset has a non-identical distribution to the training dataset which increases the difficulty of the classification task,

[1] In fact we use electrical angle $sin(\vartheta)$. For $\vartheta = 0$ the azimuth extent is $0.5\,\mathrm{rad}$ and for $\vartheta \neq 0$ it increases. For convenience we refer to it as azimuth.

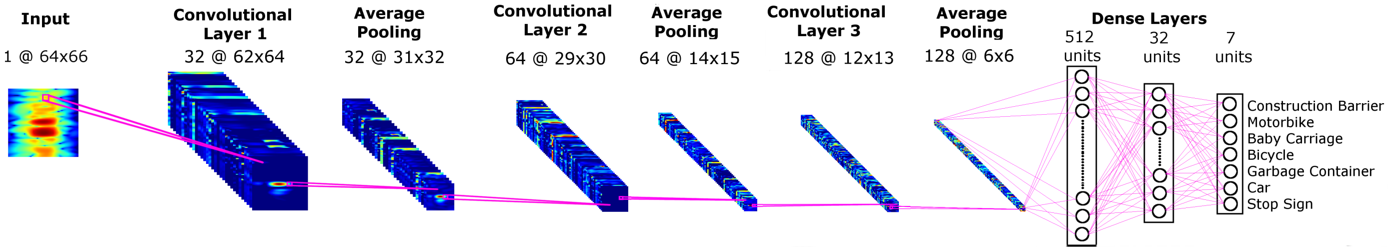


Fig. 2. CNN Architecture for ROI input. This CNN architecture with 3 convolutional and 2 fully-connected layers is kept identical for all CNN experiments.

though reliably evaluates the generalization capability of the algorithms. The validation set was created from independent measurements involving both straight and curvy driving patterns. The training, validation and test datasets contain 39126, 12212 and 8376 data points, respectively.

C. Network Architectures

Each ROI, in linear scale, forms the input to a Convolutional Neural Network (CNN) [18], which consists of 3 convolutional layers using $3x3$ filters, with 32, 64 and 128 filters. Each convolution layer is followed by a 2×2 average-pooling layer, reducing the dimensionality of the feature maps by a factor of 2. The final feature map is flattened and processed by 2 fully-connected layers with 512 and 32 neurons, respectively, before classification is performed by a softmax layer. Each layer uses rectified linear unit (ReLU) activation functions. The CNN architecture and the feature maps at different layers can be seen in Fig. 2.

During training, Batch Normalization [19] and Dropout [20] with a drop-out probability of 0.40 are used after each of the 2 fully-connected layers. Training of the network weights uses the Adam [21] optimizer with a batch size of 64. Hyper-parameters were optimized on an independent validation set.

III. Method

The ROIs represent only a portion of the entire field-of-view (FOV), i.e. full range-velocity-azimuth spectrum. As the sensor operates in the polar coordinate system, the physical area covered by the ROI expands with range. Therefore, the ROI may capture reflections from multiple targets and their side-lobes, presenting themselves as pernicious noise. Hence, most periphery parts of the ROI present distractions to the classification algorithm.

Without any prior information about the location of the ROI in the FOV, a machine learning algorithm that should generalize to real-world scenarios needs to learn to deal with these distortions from the data, as well as learn to ignore reflections from other objects. This would require a much larger and more diverse dataset than we have available, and importantly would require measurements of ROIs from all regions of the FOV, and with different objects distorting the measurements. Alternatively, data augmentation by selected transformations applied to ROIs could aid for such a task, though data augmentation for radar spectra is an open problem. In the following, we propose two novel radar-specific approaches to incorporate prior information about the location of the ROI in the FOV as an additional input to the networks, and to suppress interfering reflections from nearby objects.

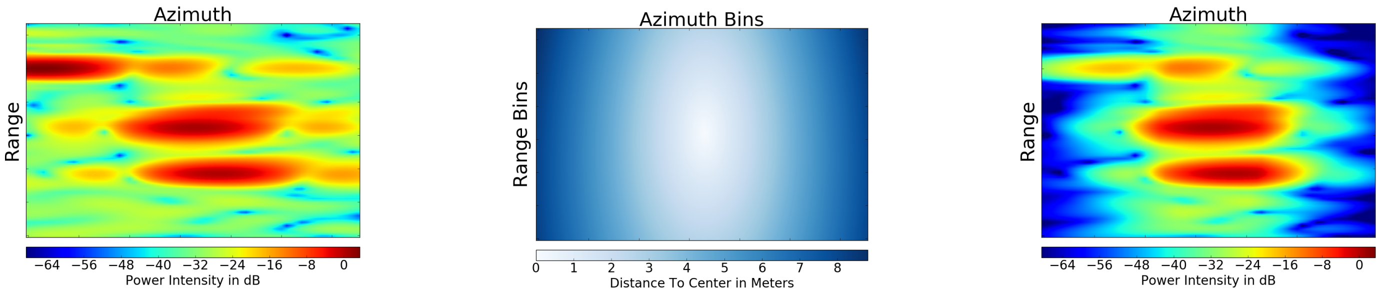
A. Incorporating Range-Azimuth Information

Additional information about the geometry of the ROI can be provided to the CNN in the form of an additional input channel. This so-called distance-to-center (DTC) map contains the *physical* distance (in meters) of each bin of the ROI to the center bin of the ROI, thereby implicitly encoding both the range and azimuth information. The CNN can learn to leverage this information (by applying the filters across the two input channels) to extract object-specific features, such as its size and reflectivity, as well as efficiently learn how the signal is distorted according to its relative location.

In order to explicitly attenuate all reflections and noise (including most but not all side-lobes from other object reflections in the FOV) from bins which do not originate from the direct vicinity of the object, the DTC map can be fused with the radar spectrum by decaying the intensity of each bin as a function of its distance to the center bin. The linear scale spectrum in the ROI is exponentially decayed by multiplying with $e^{-a \cdot (d - d_{min})}$, where d is the distance to the center bin obtained from the DTC map, and a is a hyper-parameter determining the rate of decay (empirically optimized as $a = 0.5$). In order to avoid attenuating reflections caused by the object, a pre-selected minimum decaying distance, d_{min}, is set where bins with $d < d_{min}$ are considered important and kept unaffected. This hyper-parameter is easily exchangeable and can be set in order to capture all possible reflections originating from the largest object class to be predicted. For all experiments, we set $d_{min} = 2.5\,\mathrm{m}$ which captures most reflections from objects in this study. The exponential decay generates a combined single input channel, which implicitly contains the location information (which can be learned from the decay rate of the signal), and pronounces object reflections by attenuating the periphery signals.

Overall, this allows us to compare three variants that incorporate radar-specific knowledge in the input:

- I_1: ROI spectrum
- I_2: ROI spectrum + DTC map (2 input channels)
- I_3: Decayed ROI spectrum outside d_{min}

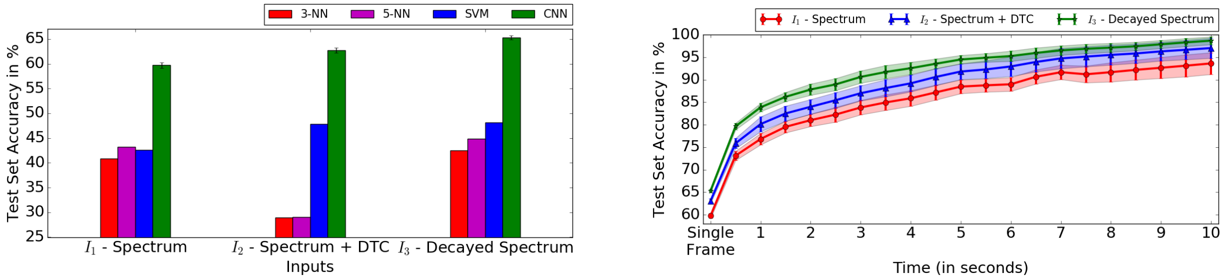


(a) I_1

(b) I_2

(c) I_3

Fig. 3. Example ROI: Construction Barrier at range $30.36\,\mathrm{m}$ and azimuth $0.56\,\mathrm{deg}$ with reflections from another object (top left region). (a) ROI spectrum (I_1) (b) Distance-to-Center (DTC) Map (c) Decayed ROI spectrum (I_3). It can be seen that the peripheral reflections are attenuated and important reflections are pronounced when using the proposed pre-processing operator. Best viewed in color



(a)

(b)

Fig. 4. (a) Single-frame classification accuracies on the test set for different classifiers and input representations. The CNN approaches significantly outperform the baseline algorithms. The decayed ROI spectrum (input variant I_3) yields the best accuracy. (b) Mean and standard deviation of accuracies for temporally filtered predictions of increasing window size (over 30 networks). A clear improvement for longer time windows over single frame approaches is visible.

IV. Experimental Results

In order to evaluate the performance of different classifiers, we compute the average class-weighted accuracy.This metric incorporates the class-imbalance in the test dataset in order to provide a representative performance metric. This is achieved by calculating the accuracy on a per-class basis and taking the average across all classes:

$$\mathcal{A} = \frac{1}{C} \sum_{c=1}^{C} \frac{p_c}{N_c} \tag{1}$$

where C is the number of classes, p_c the number of correctly classified samples of class c, and N_c the number of samples belonging to class c.

A. Baseline Algorithms

There are no public comparable datasets or algorithms for radar spectrum classification, hence we have to create our own baseline to put the accuracy of the CNN classifier into context. We compare against other machine learning classification algorithms, in particular, K-Nearest Neighbor (KNN) for $k = 3$ and $k = 5$ with the standard Euclidean distance as the metric and Support Vector Machines (SVM) with a Radial Basis Function kernel, which can both operate directly on the 2D ROI images.

B. Assimilating Location Information

The first experiments evaluate the effect of the different input representations I_1, I_2, and I_3, as presented in Section III-A. For this experiment, the network architecture and hyper-parameters are kept identical and only the network inputs and weight initializations are changed. The results in Fig. 4a report the mean and standard deviation of the classification accuracy on an independent test set over 30 networks (each network having different random weight initializations).

Fig. 4a shows that CNNs clearly outperform the baseline algorithms on all tested input representations. Furthermore, the CNNs, but not necessarily the other algorithms, benefit from the implicitly encoded geometrical information in the DTC map and the decayed spectrum. The input representation with distance-dependent exponential decay in peripheral parts of the ROI (I_3) consistently leads to the best classification accuracy for all tested algorithms. For KNN, which is known to have difficulty with high-dimensional data, the second input channel in representation I_2 leads to a drop of performance, but KNN does benefit from representation I_3. Table I provides the detailed results for CNNs for the three input representations, indicating clear advantages of I_2 over I_1, and I_3 over I_2. The sample means shown in Table I are statistically different at a significance level of $p < 0.01$.

In summary, CNNs exhibit the best classification performance, and the use of radar-specific input representations has

TABLE I
MEAN AND STANDARD DEVIATION OF CNN PERFORMANCES FOR THE 3 DIFFERENT INPUT REPRESENTATIONS.

I_1 - Spectrum	I_2 - Spectrum + DTC	I_3 - Decayed Spectrum
59.73 +/- 0.56	62.75 +/- 0.50	**65.30 +/- 0.40**

a clearly beneficial effect.

C. Filtering Predictions over Time

Due to the sensitivity of radar reflections to the aspect angle to the objects, radar spectra and their classification may abruptly change from one frame to the next. Filtering classification results over time is, therefore, an obvious way to improve classification performance. A simple and computationally efficient approach is to apply a temporal filter across the single-frame predictions and predict the class based on a majority voting scheme, where voting ties are broken at random. This operation improves performance by integrating previous predictions, and using the prior knowledge that classification results for *static* objects should not change abruptly across time.

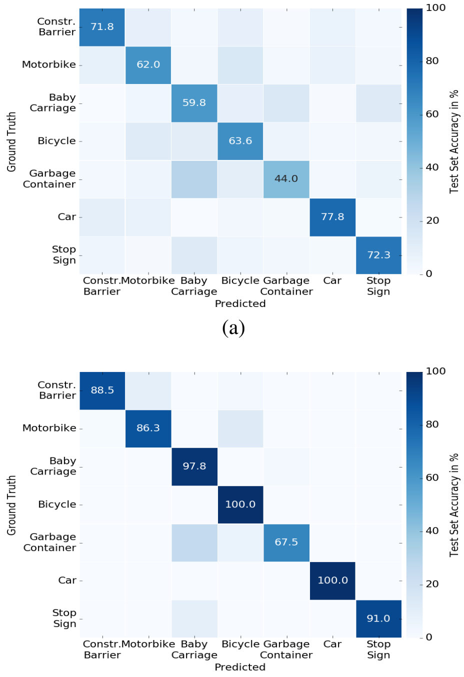
Fig. 4b shows that, as expected, increasing the filtering window T improves the classification performance for a static environment. It can further be seen that the performance order between input representations I_1 to I_3 is maintained, and again the distance-dependent exponential decay (I_3) shows the best performance and smallest variance for all filter lengths. As temporal filtering is a simple method to incorporate temporal information into the prediction stage, it relies highly on the single-frame classification performance. Therefore, applying the same temporal filter across the other weaker KNN and SVM predictions degraded classification performance severely.

We also experimented with presenting multiple frames simultaneously as inputs to the CNN, but no advantage over majority-voting was visible, and the resulting CNN is of significantly higher complexity.

In summary, temporal filtering of CNN predictions significantly improves the classification performance over single-frame approaches.

D. Per-Class Accuracies

Not all objects are equally difficult to classify, hence it is interesting to observe the per-class accuracies and confusion among the classes. The confusion matrices in Fig. 5 show how often each of the 7 objects (true class in rows) is classified into every other class (predicted class in columns). Fig. 5a shows the confusion matrix of the best single-frame CNN (for input I_3), whereas Fig. 5b shows the confusion matrix for a filter length of 4 seconds. A window size of 4 seconds was chosen here because it is a reasonable time frame for sequentially observing a single static object in the dataset. Both matrices show the desired concentration in the diagonal (indicating correct classifications). For the single frame case, large objects such as cars or barriers are best classified, whereas there is some confusion between garbage containers and baby carriages, or bicycles and motorbikes.



(a)

(b)

Fig. 5. Test set confusion matrices for (a) single-frame CNN trained with input variant I_3 and (b) prediction filtering method over best single-frame CNN predictions with a window size of 4 seconds.

With temporal filtering, many of the confusions are removed, although there are still several cases where garbage containers and baby carriages are mixed up because indeed their spectra look visually similar. Ultimately it will be more important to identify the functional relevance of the object, e.g. whether an emergency brake is necessary, rather than the exact identity. This will be the focus of further studies with a larger set of objects.

V. DISCUSSION

The above results indicate that deep learning applied to automotive radar spectra is a promising approach for object classification and scene understanding. Since all objects were measured from many different distances and aspect angles, and under real-world conditions, the high accuracies of CNNs show that deep networks are able to extract features from spectra that allow them to generalize well. This opens up possible research questions about interpreting the features which the network extracts. Furthermore, the architectures of the CNNs are small enough to be efficiently implemented in hardware.This means that our proposed system has a high potential for real-time radar-based object classification. Currently, the ROI extraction and classification only occurs for the detected objects for which the ground truth (i.e. label) exists. The system does not yet have the ability to classify reflections from an unknown object in the full spectrum which also produced a detection (e.g. road curb or false detections). The detection process currently uses a conventional detection algorithm (OS-CFAR) in order to extract ROIs to ensure that the target was detected; this step can potentially be replaced by a neural-network based detection approach modeled after successful region-proposal schemes used in the vision domain [22], [23].

There are no directly comparable data sets or classification methods for this task, hence our results show relative comparisons between different machine learning methods and input representations. We find a clear advantage for deep learning methods over simpler methods such as KNN and SVM. In the future, a comparison with state-of-the-art reflection-based methods is necessary to evaluate whether the advantage of the method lies in the additional information available in the spectra, or in the powerful CNN classifier. We also plan to expand measurements to even more object classes and multiple instances of each class to evaluate the true generalization capabilities. Currently, our database contains only static objects, but the approach should easily transfer to dynamic scenes, where Doppler information could provide an additional cue to distinguish objects.

Two insights from our study are particularly interesting for future studies: First, the explicit integration of radar know-how into input pre-processing yields significant improvements. This suggests that additional insights from radar signal processing, e.g. for data augmentation or input normalization, could improve the performance even more. Second, the filtering of classification predictions over time provides a significant boost to the CNN classifiers. Our results show that even an integration over a single second can already improve the accuracy by 18%, and if the object is in the FOV long enough an almost perfect classification is possible. Ultimately this becomes a trade-off between accuracy and latency and depends on the available time until a decision is required. Our current approach uses a majority vote over multiple single-frame predictions, but it seems likely that a direct accumulation of evidence in a recurrent neural network architecture such as LSTM [24] yields similar or potentially better results.

VI. Conclusion

This article has presented the first evaluation of deep learning methods applied directly to radar spectra for scene understanding under real-world conditions. The approach presents a promising alternative to classical radar signal processing methods and outperforms other machine learning approaches on a novel dataset with realistic objects. The best results can be obtained by combining state-of-the-art deep learning with specific radar know-how and prior understanding of the task. This suggests that a hybrid between data-driven and model-based approaches may have the greatest chance for success, in particular with limited available real-world training data.

Acknowledgments

The authors thank Jürgen Hasch, Markus Gonser, Felix Müller, Michael Scherl, Gor Hakobyan, Eugen Schubert and Johannes Fink for discussions and help in data collection.

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[2] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle detection techniques for collision avoidance systems: A review," *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2318–2338, 2015.

[3] E. Schubert, F. Meinl, M. Kunert, and W. Menzel, "Clustering of high resolution automotive radar detections and subsequent feature extraction for classification of road users," in *International Radar Symposium*, 2015, pp. 174–179.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[5] J. Lombacher, M. Hahn, J. Dickmann, and C. Wöhler, "Potential of radar for static object classification using deep learning methods," in *IEEE Int. Conference on Microwaves for Intelligent Mobility*, 2016, pp. 1–4.

[6] R. Dubé, M. Hahn, M. Schutz, J. Dickmann, and D. Gingras, "Detection of parked vehicles from a radar based occupancy grid," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1415–1420.

[7] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[8] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, pp. 125–138, 2016.

[9] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic segmentation on radar point clouds," in *International Conference on Information Fusion*, 2018, pp. 2179–2186.

[10] A. Bartsch, F. Fitzek, and R. Rasshofer, "Pedestrian recognition using automotive radar sensors," *Advances in Radio Science*, vol. 10, no. B. 2, pp. 45–55, 2012.

[11] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 180–189, 2018.

[12] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[13] M. Zhao, Y. Tian, H. Zhao, M. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "RF-based 3D skeletons," in *ACM Special Interest Group on Data Communication*, 2018, pp. 267–281.

[14] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-doppler based human-robot classification using ensemble and deep learning approaches," *IEEE Radar Conference*, pp. 1043–1048, 2018.

[15] T. Visentin, J. Hasch, and T. Zwick, "Calibration of a fully polarimetric 8x8 MIMO FMCW radar system at 77 GHz," in *11th European Conference on Antennas and Propagation*, 2017, pp. 2530–2534.

[16] ——, "Analysis of multipath and DOA detection using a fully polarimetric automotive radar," *International Journal of Microwave and Wireless Technologies*, vol. 10, no. 5-6, p. 570–577, 2018.

[17] H. Rohling, "Ordered Statistic CFAR technique - an overview," in *International Radar Symposium*, 2011, pp. 631–638.

[18] Y. Lecun and Y. Bengio, "Convolutional networks for images, speech, and time-series," in *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995.

[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, vol. abs/1412.6980, 2014.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision*, 2016, pp. 21–37.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–1780, 1997.