Super-Resolution Radar Imaging With Sparse Arrays Using a Deep Neural Network Trained With Enhanced Virtual Data

CHRISTIAN SCHUESSLER (Graduate Student Member, IEEE),
MARCEL HOFFMANN (Graduate Student Member, IEEE), AND **MARTIN VOSSIEK** (Fellow, IEEE)

Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

CORRESPONDING AUTHOR: Christian Schuessler (e-mail:christian.schueslser@fau.de).

ABSTRACT This paper introduces a method based on a deep neural network (DNN) that is perfectly capable of processing radar data from extremely thinned radar apertures. The proposed DNN processing can provide both aliasing-free radar imaging and super-resolution. The results are validated by measuring the detection performance on realistic simulation data and by evaluating the Point-Spread-function (PSF) and the target-separation performance on measured point-like targets. Also, a qualitative evaluation of a typical automotive scene is conducted. It is shown that this approach can outperform state-of-the-art subspace algorithms and also other existing machine learning solutions. The presented results suggest that machine learning approaches trained with sufficiently sophisticated virtual input data are a very promising alternative to compressed sensing and subspace approaches in radar signal processing. The key to this performance is that the DNN is trained using realistic simulation data that perfectly mimic a given sparse antenna radar array hardware as the input. As ground truth, ultra-high resolution data from an enhanced virtual radar are simulated. Contrary to other work, the DNN utilizes the complete radar cube and not only the antenna channel information at certain range-Doppler detections. After training, the proposed DNN is capable of sidelobe- and ambiguity-free imaging. It simultaneously delivers nearly the same resolution and image quality as would be achieved with a fully occupied array.

INDEX TERMS Neural networks, super-resolution, sparse antenna arrays, compressed sensing, self-supervised learning, radar simulation.

I. INTRODUCTION

Radar sensors are among the most popular sensors for automotive driving applications. Compared to lidar and camera sensors, they are robust, cheap, and do not suffer due to harsh weather conditions, such as snow, rain, and fog [1]. Also, radar sensors can directly measure the radial speed of objects by exploiting the Doppler effect, which drastically improves the classification of moving entities such as bicycles and pedestrians [2], [3], [4].

However, one of the main disadvantages of radar sensors is the low angular resolution, which is typically below that of lidar sensors.

Therefore, the aim of many hardware and signal processing approaches is to obtain high-resolution angular information.

To limit both the hardware complexity and the amount of received data, one objective is to restrict the number of physical antennas. Among the most common ways to achieve this is by using multiple-input multiple-output (MIMO) antenna arrays, followed by subspace methods for direction of arrival (DoA) estimation, such as MUSIC [5], ESPRIT [6], or, more recently, compressed sensing (CS) methods [7], [8], [9], [10], [11], [12].

Another approach to reduce the number of antennas required is to sparsely sample the entire array aperture by not spacing the antennas equidistantly. This is done, for example, by deploying minimum redundancy arrays [13], co-prime arrays [14], or nested arrays [15]. A further approach is to design the antenna placement by optimizing certain properties
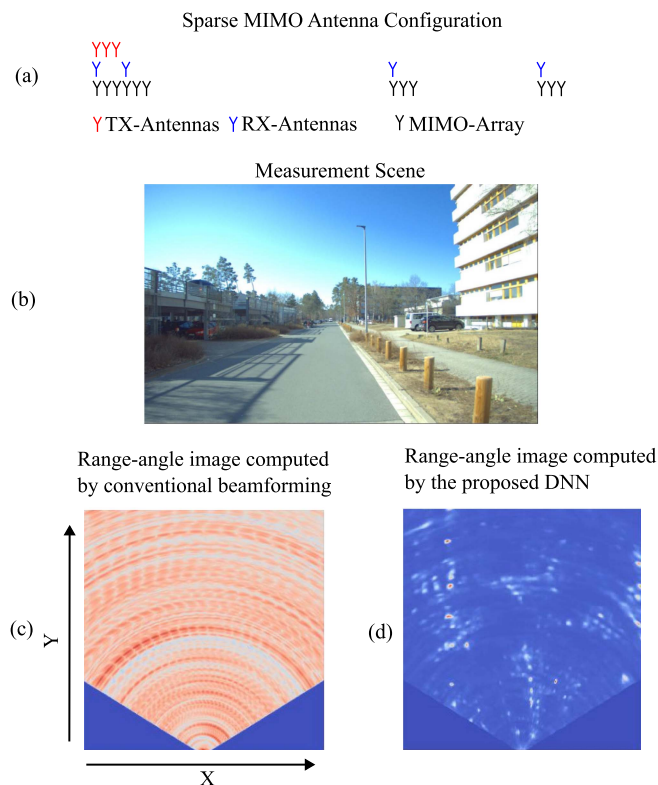
Sparse MIMO Antenna Configuration

FIGURE 1. Image (a) shows the sparse MIMO antenna arrray configuration that is derived from a small subset of the antennas of the AVR-QDM-110 radar system from Symeo GmbH, an indie Semiconductor company. The measurement scene used for evaluation is depicted in (b). The results for the conventional delay-and-sum (DaS) beamforming and the proposed approach using a deep neural network (DNN) applied to the sparse array configuration are presented in (c) and (d), respectively. As visible the DaS beamformer is not able to process the sparse data at all, wheras the DDN trained with our novel approach yield a nearly perfect result.

FIGURE 2. The image shows the overall concept of this approach. In (a), the complete DNN is trained using only simulated data and attempts to reproduce the high-resolution simulated and enhanced ground truth data as well as possible. After training, the DNN can be applied in real measurement scenarios (inference), as shown in (b).

of the expected signal, for example by minimizing the main-lobe width and at the same time constraining or minimizing the sidelobe level as demonstrated in [16] and [17], or more suitable for CS applications, the coherence of the signal [18].

In this work, high-resolution and aliasing-free radar images are generated from sparse antenna configurations that would lead to severe ambiguities and aliasing with conventional signal processing methods. The proposed deep learning–based signal processing scheme is not only applied to simulated data but also tested on real measurements, as shown in Fig. 1.

In our previous work [19], we applied a deep neural network (DNN) to radar range-angle images to drastically alleviate clutter and noise in real-world measurement scenarios. This was achieved by simulating both enhanced high-resolution noise and clutter-free data as the ground truth and radar data of an accurate digital twin of the radar sensor under test as input. As a consequence, the DNN, which was trained exclusively in a simulation environment, successfully learned to remove clutter and sharpen the image. The simulations were generated using the radar ray tracing implementation we proposed in [20].

Compared to other existing work, our previous work in [19] firstly utilized realistic radar simulations for radar image
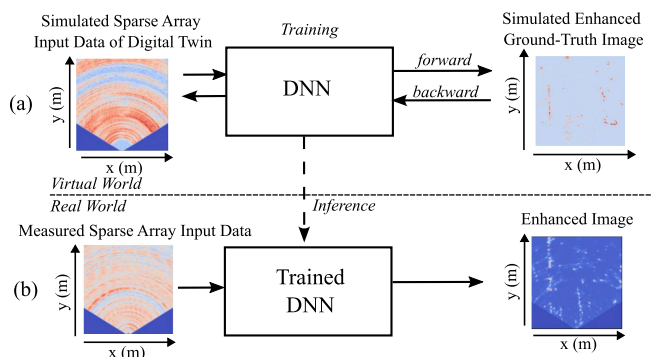
enhancement with respect of noise and clutter reduction. There and in this work, the simulations were conducted in a realistic environment, resembling a city landscape, see Fig. 6(a). This way, real-world structures, such as building, trees, and other objects are accurately mapped to the generated radar images. However, this was implemented for a fully occupied array and only the magnitude of the simulated range-angle image was used as input.

Here, our previous work is extended in two fundamental ways. Firstly, we consider severely thinned antenna arrays and show that even for extremely sparse array configurations, an image reconstruction is still possible. Secondly, a novel feature vector is proposed, which combines information by applying beamsteering on the sparse antenna array and utilizing the covariance matrix from raw channel data at the same time. Further, the training data is extended by synthetic data with closely spaced targets to improve the target separation performance even further. Compared to our previous work, an algorithm to extend the approach for multiple Doppler detections in the same range bin is proposed and evaluated. Also, the antenna radiation pattern is considered in the training data simulation, which was omitted in our previous work.

Compared to other existing work for DoA estimation, this work firstly utilizes the complete range-angle-image selected at specific Doppler detections instead of using only single range-Doppler detections. Also, the usage of high realistic simulation data for DoA estimation, especially for sparse arrays, is completely novel.

In detail, for the training of the DNN, two special datasets are generated with our realistic ray tracing simulator presented in [20]. The first simulated dataset is designed to resemble realistic measurement data as close as possible, serving as a digital twin of the real-world sensor. The second dataset is simulated using the corresponding enhanced virtual sensor with the same parameters but with a significant higher number of antennas, leading to a higher angular resolution. The proposed training concept is briefly summarized in Fig. 2.

Consequently, the first dataset can be used as the input for the DNN, whereas the second dataset can be used as ground

truth. This will be further described in Section III. Because of the easy scalability of simulation data, the DNN can be trained very thoroughly with large input and ground truth datasets. Since the simulated data is so close to reality, this DNN can then be directly applied to real measurement data.

The training and preprocessing code of the proposed approach is also published here: https://github.com/ChristianSchuessler/Sparse-Array-Radar-Imaging.

II. RELATED WORK

This section provides an overview of the existing work on sparse array processing and deep learning approaches for radar imaging.

In [21], [22], [23], signal processing techniques were presented to interpolate the signal between antenna channels of sparse array configurations or to extrapolate small antenna arrays to larger ones. Recently, machine learning algorithms were applied to the array interpolation or extrapolation problem in [24], [25], and [26]. In [24], this was done by using real measurement data in a self-supervised way. For the input data, certain channels of the fully occupied array were ignored and the remaining channels have to be predicted by the neural network. This work uses a similar network architecture (U-Net) as our approach, but our work does not require full range-Doppler-channel tensors, instead only 2D images as input turned out be sufficient. Further, it directly predicts the reconstructed image by realistic, but high resolution simulated ground-truth data. This ground-truth data is completely noise free and has a much higher resolution compared to the deployed fully occupied array, which cannot be achieved by common self-supervised approaches based on real measurement data.

In comparison, [25] was purely based on synthetic data in combination with a smaller neural network for training. The authors suggest that interpolating the signal requires neural networks with fewer parameters compared to estimating the DoA directly. Interestingly, even the target separation performance could be improved, the results seem to be to some extend unstable for single target scenarios at different bearing angles.

The DoA can also be estimated directly by neural networks, as shown in various works, such as [27], [28], [29], [30], [31], [32], [33]. These methods typically outperform traditional methods, such as MUSIC or ESPRIT, especially in environments with a low signal-to-noise ratio (SNR) [30]. Most works in this context have relied on training data from synthetic point-like targets, requiring only a single snapshot of the antenna channels. However, in [32], real measurement data and a region of interest in the fast-time were additionally used as input for the neural network. In [34] and [35], DoA approaches were also applied to distributed sparse arrays. The performance of our approach is compared to [30] in the results and shows a significant better performance, due to more realistic training data and by utilizing complete range-angle images.
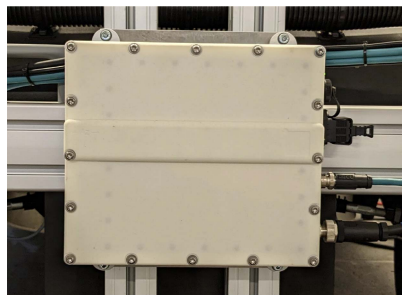

FIGURE 3. Image of the deployed automotive FMCW MIMO radar device AVR-QDM-110 from Symeo GmbH, an indie Semiconductor company.

Similar approaches can be found in the radar imaging domain, where the DoA of a signal is only considered implicitly, since a complete image has to be reconstructed at once. The authors in [36] used CS reconstructions of fully sampled simulated and measured data to teach the network to reproduce these results from low-resolution data. They consequently called their approach *CS-DNN*, since it can produce similar or even better results compared to traditional CS methods. Other similar approaches using DNNs in the radar imaging domain can be found in [37], [38], [39], [40], [41], [42]. Our approach adopts the concept of using 2D images as input by utilizing range-angle images, which are suited for automotive radar applications.

The reason for the popularity of NN architectures is their often better performance compared to traditional subspace or compressed sensing methods. Also, typically most DoA NN architectures do not need an estimation of the number of targets, e.g. model order as compared to subspace methods such as MUSIC or ESPRIT. For CS methods, the number of targets does not have to be estimated directly, but also hyperparameters for different loss terms and the number of iterations have to be tuned. However, tasks like proper feature engineering, collecting and generating enough training data can make NN architectures hard to build.

III. RADAR SYSTEM SETUP AND SIGNAL PROCESSING

In this section, the radar system, the frequency modulated continuous wave (FMCW) modulation scheme, and the corresponding signal processing workflow are presented. Following that, the input and ground truth data generation for the DNN is described in more detail.

Here, matrices are denoted in upper-case bold letters ($\mathbf{M}$), vectors in lower-case bold letters ($\mathbf{v}$), and scalars in lower-case letters (a).

A. RADAR SYSTEM SETUP

An image of the radar unit attached to a test vehicle is depicted in Fig. 3. The radar system consists of a MIMO FMCW radar device with the parameters stated in Table 1. Since the radar operates in MIMO mode, a virtual array can be constructed by combining the TX and RX antenna channels [43]. Sparse virtual arrays can be constructed by omitting certain TX and RX channels in the subsequent processing chain. Certain

IEEE Journal of Microwaves

TABLE 1. Radar Parameters

Symbol	Parameter	Value
f_c	RF carrier frequency	77 GHz
B	RF bandwidth	1 GHz
N_{chirp}	number of chirps	128
T_c	Chirp duration	80.6 μs
N_{TX}	Number of TX antennas	3
N_{RX}	Number of RX antennas	16
d_{TX}	TX antenna element distance	2 mm
d_{RX}	RX antenna element distance	6 mm
δ_a	Angular resolution	1.83 deg.

(a) 16 RX Antennas

(b) 6 RX Antennas

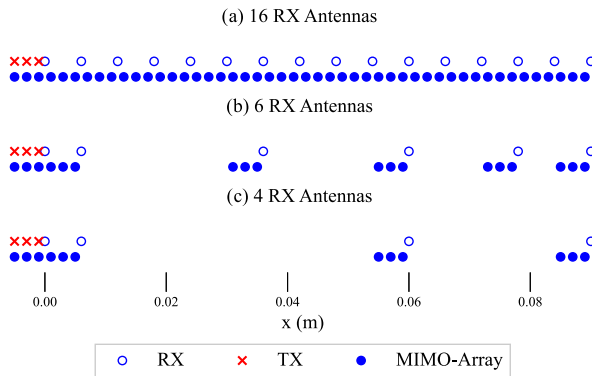(c) 4 RX Antennas

○ RX   × TX   ● MIMO-Array

FIGURE 4. The full antenna array of the chosen array configuration of the AVR-QDM-110 with 3 TX and 16 RX antennas, and consequently 48 virtual elements in the MIMO array, is shown in (a) with an virtual element spacing of 0.58λ. Two sparse variants are formed by omitting several RX antennas, in (b) and (c).

antenna configurations with their corresponding virtual arrays, which are evaluated in this work, are depicted in Fig. 4. As can be seen, while the fully occupied array can be constructed well with the MIMO configuration of the investigated radar sensor, it is difficult to construct meaningful sparse array configurations. Therefore, we tried to keep the overall aperture size large, while restricting the number of RX antennas drastically. Fig. 4 shows that there are still pairs of three virtual antennas, which is due to the original MIMO configuration of the radar sensor. However, this article shows that even under this suboptimal configuration high resolution ambiguity-free radar images can be obtained.

B. FMCW SIGNAL MODEL

The intermediate frequency (IF) signal, also called beat signal, of an FMCW radar can be given as

$$s_{IF}(i_{TX}, i_{RX}, t)$$
$$= \sum_{k=0}^{N-1} A_k \exp(2\pi j(\mu t \tau_k(i_{TX}, i_{RX}) + f_c \tau_k(i_{TX}, i_{RX}))), \quad (1)$$

with each transmit (TX) and receive (RX) antenna channel index denoted as i_{TX} and i_{RX}, respectively. The complete signal is composed as a coherent summation of the modulation

term including the round trip delay τ_k of all targets N in the scene with amplitude A_k, as defined in the equation below:

$$\tau_k(i_{TX}, i_{RX}) = \frac{||\mathbf{x}_{i_{TX}} - \mathbf{x}_k||_2 + ||\mathbf{x}_{i_{RX}} - \mathbf{x}_k||_2}{c}, \quad (2)$$

with c being the speed of light and $\mathbf{x}_{TX}$, $\mathbf{x}_{RX}$, and $\mathbf{x}_k$ representing the position of the antennas and targets, respectively. The frequency slope μ, also called the chirp-rate of the FMCW radar, is given by the bandwidth B divided by the chirp duration T_c

$$\mu = \frac{B}{T_c}. \quad (3)$$

A more detailed description of FMCW radar signal processing can be found in various review and tutorial papers, such as [44].

For the ground truth data simulation, the identical radar parameters to those described in Table 1 are used. However, in contrast to the sparse input data generation, a single TX is combined with 256 RX antennas that are placed as a uniform linear array (ULA). The 3-dB beamwidth (i.e. resolution) for linear arrays can be estimated as [45, p. 11]

$$\theta_3 = \frac{51.05\lambda}{D}, \quad (4)$$

for wavelength λ and aperture length D. The resolution of the full not sparse input antenna array is then:

$$\theta_3^i \approx \frac{51.05\lambda}{48 \cdot 0.58\lambda} = 1.83 \text{ deg.}, \quad (5)$$

and for the enhanced ground truth radar sensor

$$\theta_3^g \approx \frac{51.05\lambda}{256 \cdot 0.5\lambda} = 0.4 \text{ deg.} \quad (6)$$

C. INPUT DATA PROCESSING

For all the following computations, we assume far-field conditions, meaning that spherical waves originating from point-scatterers can be approximated by plane waves, given that the scatterer is sufficiently far away. This assumption implies that each target lies in the same range bin, which is typical in automotive radar applications [43].

First, an FFT is applied along the fast- and slow-time dimension of s_{IF} to obtain range and Doppler information for each TX–RX channel combination. In the next step only the magnitude of all range-Doppler images is taken to be processed into a single *mean* image using all range-Doppler images. This does improve the SNR for the following Doppler bin selection, and also ensures that for each antenna channel the same range-Doppler pixel is taken in all further processing steps.

Afterwards, one single Doppler bin, typically the bin with maximum amplitude, is selected for each range bin. This leads to the 2D signal S_{IF} in the frequency domain, that has dimensions in channel and range. In the second stage, the TX and RX antenna channels are sorted to form a virtual antenna array, which corresponds to a non-uniform linear array, as in Fig. 4(b) and (c).

Essentially, two features are generated out of the signal S_{IF} as described below:

1) Typically, for ULA configurations, a fast fourier transform (FFT) is also applied in the channel direction to obtain angular information. However, this is not immediately possible for sparse arrays, since there are gaps in the signal. Instead, a common delay-and-sum (DaS) beamforming method that applies steering weights for an angle hypothesis is chosen [45]. The steering weights for an arbitrary array configuration are shown below:

$$\mathbf{v}(\theta) = \begin{pmatrix} w_0 \cdot e^{-j\frac{2\pi}{\lambda}\sin(\theta)\mathbf{x}_0} \\ \vdots \\ w_{N_v-1} \cdot e^{-j\frac{2\pi}{\lambda}\sin(\theta)\mathbf{x}_{N_v-1}} \end{pmatrix}, \qquad (7)$$

where $\mathbf{x}_i$ denotes the one dimensional virtual array position, and θ is the current steering angle. The number of virtual antennas is denoted by N_v, which is the product of N_{TX} and N_{RX}. To reduce the sidelobe level, a window function implemented via the weights w can be applied. However, for simplicity all weights w are set to 1, since this approach already lead to good results. Moreover, we apply the beamsteering in the sine domain, such as for $u = \sin(\theta)$, so that this approach is also suited to extrapolate linear arrays where the beamsteering processing can be replaced by an often faster FFT.

The complete beamforming process for all angles θ, resulting in the vector $\mathbf{i}_s$, can then be stated as a matrix multiplication for each range index r_i with the processed signal $\mathbf{S}_{IF}$ with steering matrix $\mathbf{V}$, as shown below:

$$\mathbf{i}_s(r_i) = \mathbf{S}_{IF}(r_i)\mathbf{V}. \qquad (8)$$

The shape of $\mathbf{V}$ is $N_v \times N_\theta$, with N_θ being the number of steering angles. The entries of $\mathbf{V}$ are

$$\mathbf{V} = \begin{pmatrix} e^{-j\frac{2\pi}{\lambda}u_0 x_0} & \cdots & e^{-j\frac{2\pi}{\lambda}u_{N_\theta} x_0} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{\lambda}u_0 x_{N_v-1}} & \cdots & e^{-j\frac{2\pi}{\lambda}u_{N_\theta} x_{N_v-1}} \end{pmatrix}. \qquad (9)$$

The computed vector $\mathbf{i}_s(r_i)$ has the length N_θ, which is equal to the required angular bins for the DNN. For ULA configurations, this matrix multiplication is equal to a discrete Fourier transform [45] and can therefore be replaced by an FFT. This feature was selected, because it already worked very well in our previous work for radar image enhancement [19]. Instead of only using the magnitude, additional phase information of the signal was selected as feature in this work, as it showed good results in [36].

2) Secondly, the covariance matrix is used as an input for the DNN. It has a shape of $N_v \times N_v$ and is computed for each range bin as described in the equation below:

$$\mathbf{\Sigma}(r_i) = \frac{\mathbf{S}_{IF}(r_i)\mathbf{S}_{IF}^H(r_i)}{||\mathbf{S}_{IF}(r_i)\mathbf{S}_{IF}^H(r_i)||_2}. \qquad (10)$$

The symbol $||\cdot||_2$ denotes the Frobenius norm of the matrix. Since the resulting matrix is Hermitian, only the upper or lower triangular is utilized for further processing, as already shown by [33]. This approach saves memory for larger antenna arrays, since our network is only designed for 450 angular bins. Furthermore, in order to fit a single angular line for each range index, the covariance matrix has to be unrolled to a vector and padded with zeros in case the resulting vector is smaller than the number of angular bins. The complete operation is summarized in the equation below:

$$\mathbf{i}_{cov}(r_i) = \text{pad}(\text{unroll}(\text{triu}(\mathbf{\Sigma}(r_i)))). \qquad (11)$$

For large antenna arrays (e.g., full array), the resulting vector would be too large for the designed angular bins. In this case, the matrix is cropped symmetrically before unrolling. This discards some information but still improves the results considerably.

The real and imaginary part and as well as the phase information of the covariance was chosen since it lead to good results in the work [30] for DoA estimation.

The final input matrix $\mathbf{I}_{in}(r_i)$ is real valued and composed of the following N_{feat} features:

$$\mathbf{I}_{in}(r_i) = \begin{pmatrix} \text{Abs}(\mathbf{i}_s(r_i)) \\ \text{Phase}(\mathbf{i}_s(r_i)) \\ \text{Real}(\mathbf{i}_{cov}(r_i)) \\ \text{Imag}(\mathbf{i}_{cov}(r_i)) \\ \text{Phase}(\mathbf{i}_{cov}(r_i)) \end{pmatrix}. \qquad (12)$$

Computing these features for each range bin leads to the final DNN input image of shape $N_r \times N_\theta \times N_{feat}$.

In comparison, in a conventional FMCW radar signal processing pipeline, the beamforming is only applied to detections in the range-Doppler map. In this work, the beamforming is applied for each range bin to obtain image-like data. In this way, the DNN can also exploit the spatial relation between neighbouring range bins.

The complete signal processing workflow starting from the raw, 3D radar cube is depicted in Fig. 5(b) and compared to the common signal processing in Fig. 5(a).

D. GROUND TRUTH DATA PROCESSING

The generation of the ground truth signal is described in detail in our earlier work [19]. It is based on a matched filter approach and resembles the description in [46]. The advantages of this technique are that this reconstruction approach is also valid for targets that are not located in the far field. This is especially relevant for the deployed large array. Also, using matched filters typically results in a better resolution with a lower sidelobe level. These advantages come at the price of a longer reconstruction runtime. However, since this reconstruction technique is only used to generate ground truth data, it does not have to be executed in real-time during the actual measurement process. Doppler and range information are obtained in an identical way to the input data procedure.
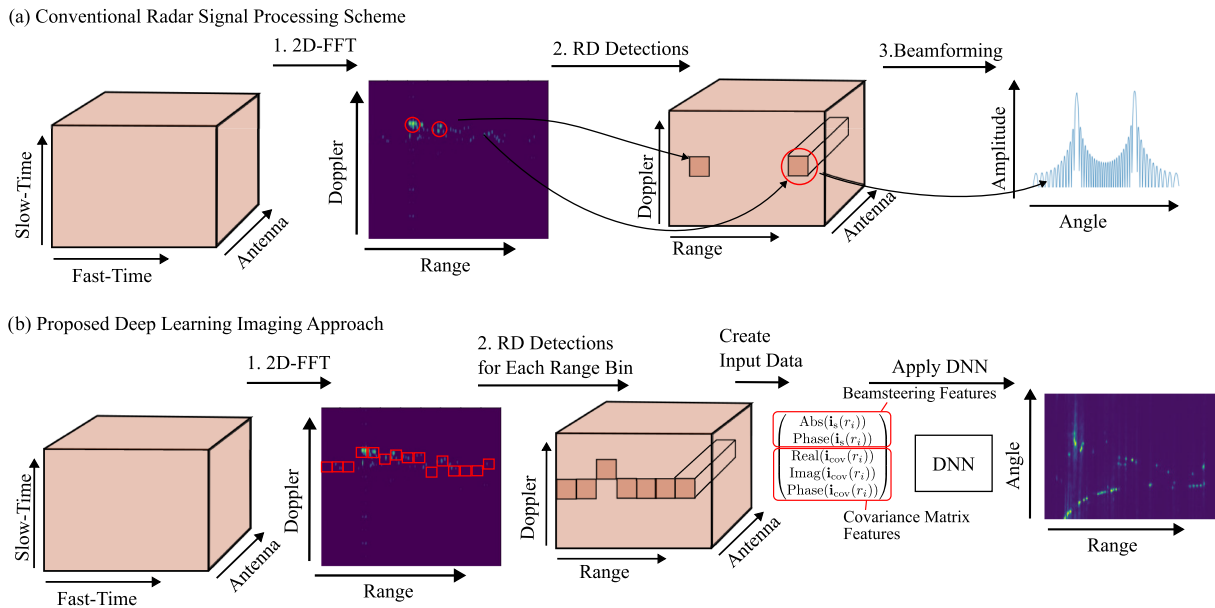
(a) Conventional Radar Signal Processing Scheme

(b) Proposed Deep Learning Imaging Approach

FIGURE 5. The upper image describes the normal radar signal processing scheme. In the first step, the slow and fast-time is converted to range and Doppler information via a 2D-FFT. Then, targets are detected in the range-Doppler map, and for each detection, the angle is estimated. In contrast, in the proposed approach (b), after computing the range-Doppler image for each virtual channel, the maximum value of each range bin in Doppler direction is taken. Afterwards, different features are extracted from the range-channel data. These are then used as an input for the DNN, resulting in high-resolution range-angle maps.

Examples for input and ground truth signals are shown in Fig. 6(a), along with virtual camera images of the simulated scene.

E. EXTENSION FOR MULTIPLE DOPPLER DETECTIONS IN A SINGLE RANGE BIN

Obviously, the proposed DNN is only able to process a range-angle image for a single Doppler detection in each range bin. In most cases this might be sufficient, as for example shown in [47], where the sum projections on each dimension of the radar cube are utilized to create detections in an automotive radar scenario. But in general, more than one Doppler detection for a single range bin should be considered. The proposed approach can be extended by running the same DNN multiple times for different Doppler detections in the following way:

1) Create Doppler detections for each range bin and sort them by their strength
2) Run the DNN for the strongest Doppler detection in each range bin
3) Run the DNN for the second strongest Doppler detection. If no second strongest detection does exists take the strongest one.
4) Repeat the procedure for the remaining Doppler detections.

IV. DEEP LEARNING APPROACH

In this section, the deployed DNN is briefly described and a description of the training data generation is given.

A. DEEP NEURAL NETWORK

The DNN used for this approach is a conventional U-Net [48] augmented with an attention mechanism, adapted from [49]. This network was chosen because it showed good results in previous work related to radar and especially to radar imaging tasks [19], [24], [41], [42]. The complete DNN is depicted in Fig. 7, featuring the $N_{feat} = 5$ input channels as described in the previous section. A softmax activation function can be added as the last activation function for a classification task. It should be mentioned, that the same network architecture as in our previous work [19] was adapted. The difference in the performance is achieved by choosing more suitable input features as described in the previous section.

In this setup, there are two ways to formulate the proposed problem. Firstly, it can be stated as a detection problem. In the machine learning context, this is often described as a binary classification problem, where the decision has to be made if a target is present at a specific pixel or not. Secondly, the problem can be viewed as a (potentially ill-posed) inverse problem. The first problem formulation is almost always chosen for the DoA deep learning approaches mentioned in Section II.

The second problem formulation is very common in compressed sensing and in the deep learning image reconstruction methods mentioned in Section II.

Depending on how the problem is formulated, a different preprocessing of the ground truth data, a different loss function, and different activation functions are chosen. If the problem is formulated as a detection problem, every pixel in the output image is set to either zero or one depending on whether a detection is present or not. Furthermore, the binary cross-entropy (BCE) loss is commonly chosen as the loss

TABLE 2. Problem Formulations

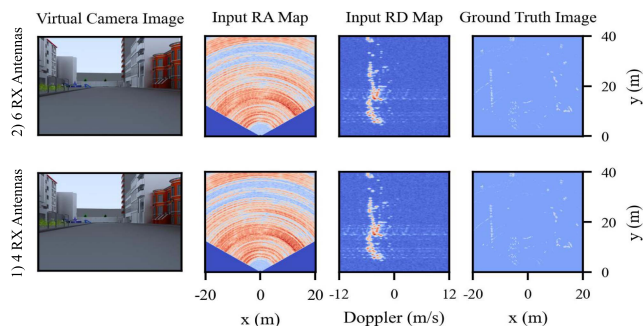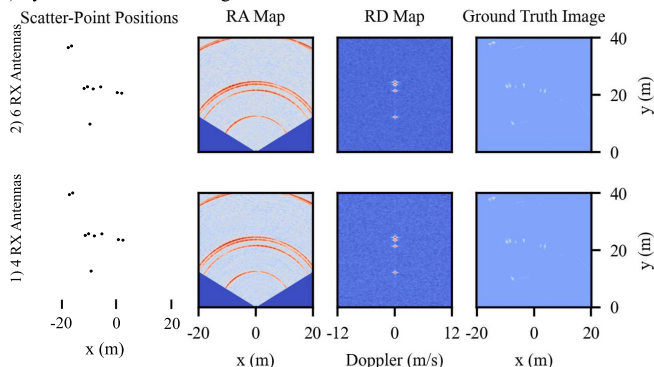|	Binary Classification Task (Detection)	Regression Task
Loss	$l_{BCE} = \mathrm{mean}[-\mathbf{Y_{tru}} \cdot \log(\mathbf{X_{est}}) + (\mathbf{1} - \mathbf{Y_{tru}}) \cdot \log(\mathbf{1} - \mathbf{X_{est}})]$	$l_{MSE} = \mathrm{mean}[\|\mathbf{X}_{est} - \mathbf{X}_{tru}\|_2 + \alpha \|\mathbf{X}_{est}\|_1]$
Pre-processing	$\mathbf{Y_{tru}}(x,y) = \begin{cases} 1, & \text{for } \hat{\mathbf{Y}}_{\mathbf{tru}}(x,y) > 0 \\ 0, & \text{otherwise} \end{cases}$	$\mathbf{Y}_{tru} = \beta \hat{\mathbf{Y}}_{tru}$
Last layer	Softmax activation function	No activation function

(a) Ray Tracing Simulations

(b) Synthetic Point-Like Targets

FIGURE 6. The upper images in (a) show example data for the two different antenna configurations generated by ray tracing simulations. The image in the first column depicts a virtual camera image of the simulated scene, and the second column image shows range-angle (RA) maps produced by the DaS beamforming technique. Since the array is sparse, sidelobes dominate and no angular information can be perceived by the human eye. The range-Doppler (RD) map is depicted in the third column image, and the ground truth image, based on the ULA with 256 elements, is presented in the right column. The lower images in (b) show the same information, although without a simulated camera image. Here, synthetically generated static point-like targets were generated, which are closely spaced in angular direction.

function. Other possible loss functions for imbalanced class distributions, such as focal loss [50] have also been deployed for DoA estimation [27].

Nevertheless, if the problem is stated as an inverse problem, the L2-Norm will typically be chosen with an optional regularization term to incorporate prior information. In compressed sensing, the signal is often expected to be sparse, and therefore, the L1-Norm on the estimated signal itself is added. Table 2 summarizes the differences of both problem formulations for the proposed approach. Here, the variable X_{est} describes the estimated radar image and Y_{tru} describes the ground truth radar image. For the classification tasks, each

pixel of the simulated ground truth range-angle image is set to either zero or one.

The preprocessing in the regression task allows for a scaling of the ground truth image with the scaling factor β. This is required if the pixels are too close to the background, since the DNN might optimize them to a minimum with all pixels being set to zero. Scaling the values to from zero may lead the learning process to another more suitable minimum, where the foreground pixels are detected properly. In our approach, $\beta = 10$ has generated satisfactory results. We did not conduct a exhaustive grid search for this parameter, but setting the value to $\beta = 1$ resulted in a training process that did not converge.

In the loss term, the sparsity assumption of the output image can be adjusted via the parameter α. Setting this parameter to zero already leads to good results, since the ground truth data initially encodes the sparsity information. In our case, even relatively small values ($\alpha = 0.1$) degraded the results slightly. However, in scenarios with less training data, setting α to a value different to zero could be beneficial. Also, the use of the logarithmized input magnitude data, instead of the linear magnitude, improves the training performance in this scenario.

Both problem formulations were thoroughly investigated, but the training for the regression task achieved better results in respect to detection performance and aliasing artifacts. This is because the sidelobes could not be fully removed in the ground truth data and were therefore equally weighted as real targets in the classification task. Also, the amplitude in the ground truth image still contains useful information that is completely removed when each target is set to a fixed value of one.

B. TRAINING DATA GENERATION
Accurate generation of sufficient training data is crucial for the performance of the DNN. For this, two approaches are pursued.

The first approach uses data generated by realistic radar simulations in a detailed virtual world based on the approach presented in [20]. The procedure is identical to that in [19], where the virtual sensor was placed randomly at predefined positions on a virtual map imported from the open source simulator CARLA [51]. Additionally, for each virtual measurement, the position of the virtual sensor was moved in a random direction with a random velocity of between 0 and $6\,\frac{\mathrm{m}}{\mathrm{s}}$. In this way, 13 000 virtual measurements were generated.

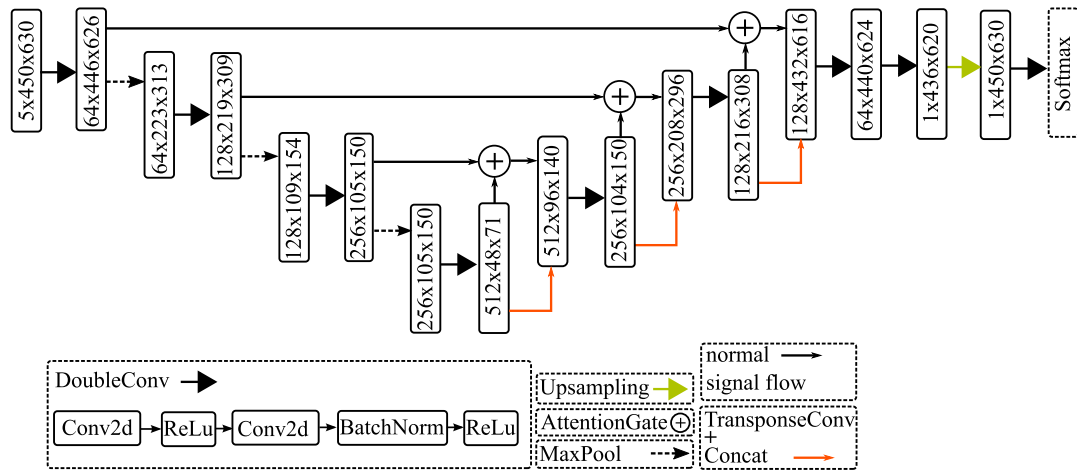However, although these data accurately resembles real-world measurements, it does not contain enough examples

FIGURE 7. The complete Attention U-Net architecture for the proposed problem, slightly adapted from [19]. Each input image consists of C channels with width W and height H, denoted as $C \times W \times H$ in the scheme. For example, the input image consists of two channels with a width of 450 (representing the angular dimension) and a height of 630 (representing the range dimension).

with closely spaced targets in the same range bin. Consequently, this dataset does not sufficiently challenge the DNN to produce high-resolution output but mainly reduces noise and sidelobes.

Therefore, as a second approach for the generation of training data, we also simulated 5500 synthetic datasets composed of closely separated point-like targets. This synthetic dataset with point-like targets could not be used on its own to train the network. However, in combination with sufficient ray tracing simulation data, the results could be improved to a relevant extent.

The two datasets were mixed together randomly, and 17 500 entries were used for training, 500 for validation, and 500 for test purposes. Again, examples for both generated input images are shown in Fig. 6.

V. RESULTS

In this section, the proposed DNN technique is compared to a common convolutional neural network (CNN)-based DoA algorithm trained on the covariance matrix of the proposed sparse antenna array configurations.[1] This CNN DoA estimator proposed in [30] was chosen as the baseline, since it showed very good and robust performance in a comparison conducted in [27]. The training and pre-processing is described in more detail in Section V-A.

Furthermore, a conventional DaS beamforming technique using a hanning window, and the MUSIC algorithm were used for comparison. For the MUSIC approach, the number of targets were estimated using the AIC criterion [52] with exact formulas taken from [53].

A. REFERENCE IMPLEMENTATION

The baseline DoA estimator from [30] relies on the normalized covariance matrix Σ generated by the antenna channels

as input:

$$\Sigma = \frac{\mathbf{x}\mathbf{x}^{\mathrm{H}}}{||\mathbf{x}\mathbf{x}^{\mathrm{H}}||_2}. \tag{13}$$

Here, $\mathbf{x}$ denotes the MIMO antenna channels resorted as for the virtual array depicted in Fig. 4. Compared to the proposed approach, the beamforming technique takes only a single antenna channel line as input and no additional input is required. Before being passed on to the CNN, the complex covariance matrix is split and stacked to a tensor of the dimension $3 \times N_v \times N_v$ according to the following conversion:

$$\Sigma_{\mathrm{T}} = \begin{pmatrix} \mathrm{Re}(\Sigma) \\ \mathrm{Im}(\Sigma) \\ \arg(\Sigma) \end{pmatrix}. \tag{14}$$

The CNN is loosely trained following the instructions from [27], although adjusted to the fewer targets and a higher SNR, expected in our test data. The processing steps are as follows:

1) Generate a random number of targets N (between 1 and 3).
2) Assign an amplitude to each target, which is chosen from a specific range.
3) Create an IF signal with a single time-snapshot, as described above.
4) Add complex normally distributed noise to the signal with a random SNR value from a specific predefined range (from 3 to 30 dB in this work).

In this case, the SNR is defined as the ratio between the weakest scatterer and the noise power, as adapted from [54]. The data were generated online, and the neural network was trained on 20 million samples.
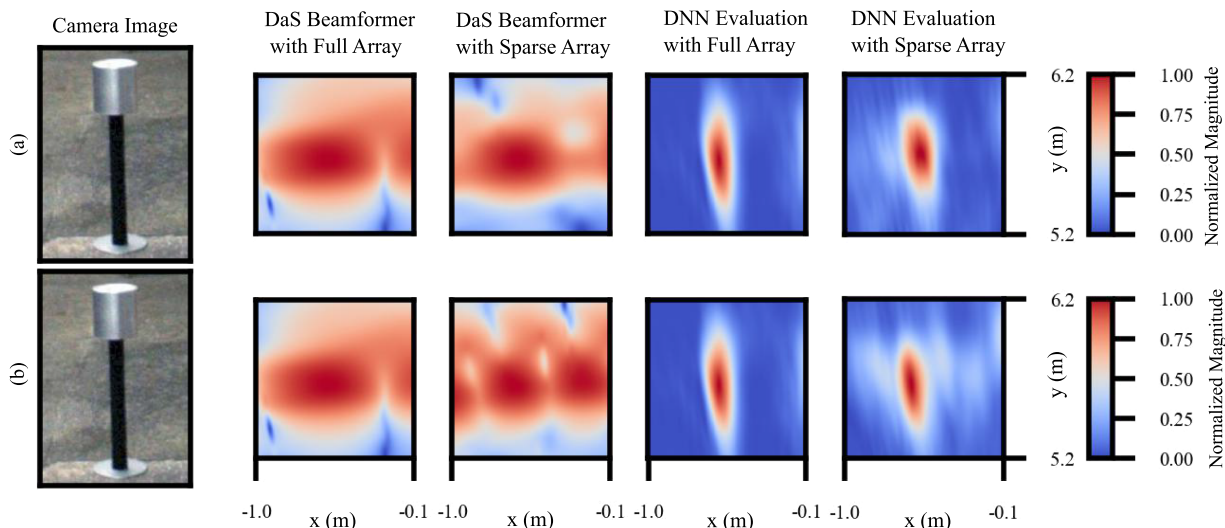
---

[1]Our DNN approach and the conventional CNN DoA estimator are both convolutional networks, but for the sake of simplicity, the former is abbreviated as DNN and the other as CNN in the following.

FIGURE 8. PSF of the proposed DNN approach compared to a conventional DaS beamforming. The first two columns show the PSF of a conventional DaS beamforming, the first (a) for a fully occupied array and the second (b) for the sparse array configuration. The same is shown for the DNN approach in the right two columns.
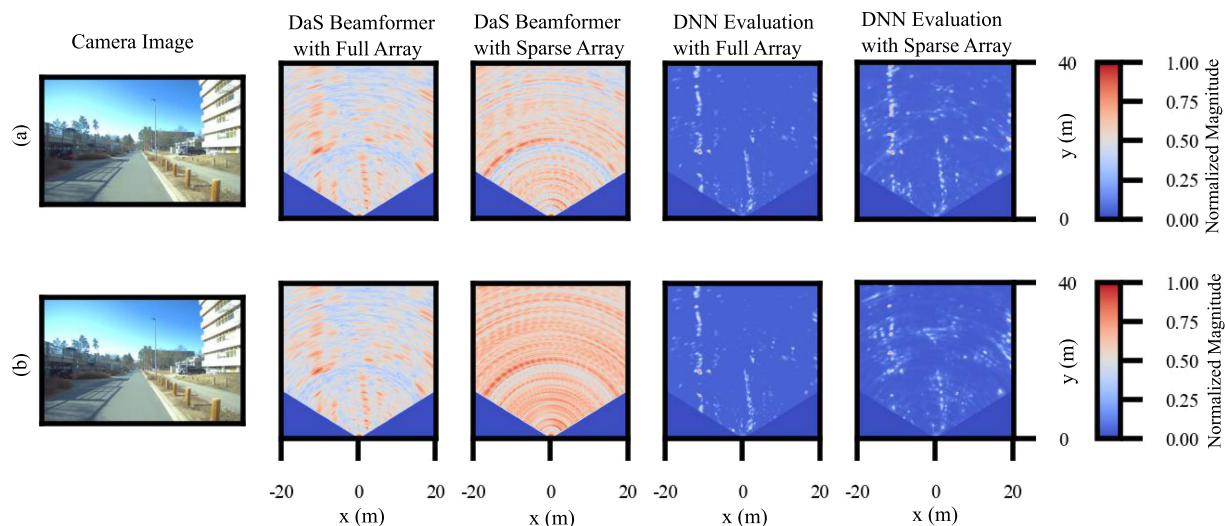

FIGURE 9. Evaluation of the proposed DNN approach on a realistic automotive radar scene. The first two columns show the results using a conventional DaS beamforming, the first (a) for a fully occupied array and the second (b) for the sparse array configuration. The same is shown for the DNN approach in the right two columns.

B. EVALUATION OF RESOLUTION AND POINT-SPREAD-FUNCTION

To evaluate the proposed DNN approach, several real-world measurements were conducted.

In the first test, a single metal cylinder at a distance of approximately 5.7 m was observed and the corresponding point spread function (PSF) of the DNN result compared to the conventional DaS beamforming approach. The results are presented in Fig. 8. Clearly, the DNN not only led to a much sharper PSF but also removed aliasing artifacts almost completely.

In the second test, the DNN results were compared to the DaS beamforming approach in realistic automotive scenarios, shown in Fig. 9. Similar to the previous results, the DNN

results were almost aliasing-free, even for 4 RX antennas, which are significantly more challenging.

As in comparison, the conventional DaS beamforming technique was not capable of evaluating useful angular information due to massive aliasing artifacts.

In the third test, the angular resolution of the implemented DNN was evaluated and compared to the MUSIC algorithm and a state-of-the-art CNN beamforming approach introduced before. The results are depicted in Fig. 10. The evaluation of the CNN approach for a full array was not possible, since the full covariance matrix required for this approach was too large. Comparing the results for the sparse array configurations, the DNN approach outperformed the other algorithms in terms of sidelobe suppression and resolution. In particular, for
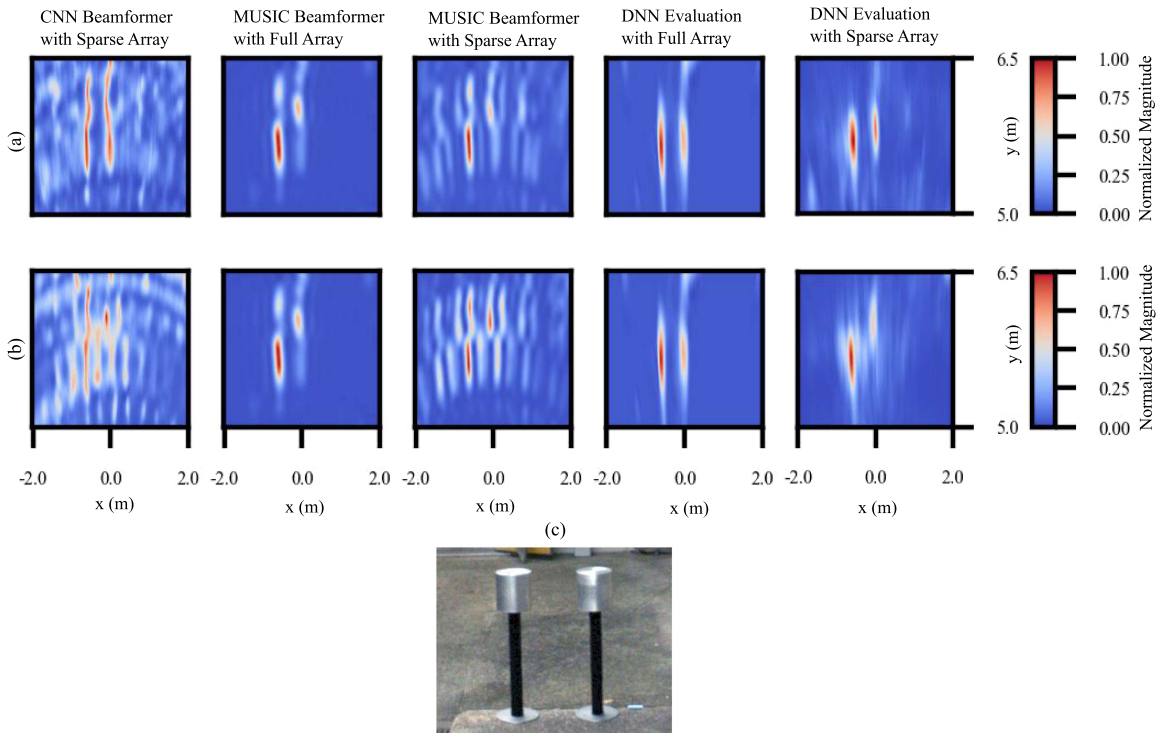
FIGURE 10. Evaluation of the proposed DNN approach compared to a conventional CNN approach as well as a MUSIC beamformer in a scenario testing the angular resolution. The first column shows the results of a state-of-the-art CNN for the sparse array. Since the input covariance matrix is too large, the results could not be generated for the full array configuration. The second and third column depict the imaging results using the super-resolution algorithm MUSIC for a fully occupied ULA and the sparse array configuration. The fourth and fifth column show the results of the DNN approach, again for the fully occupied and the sparse array, respectively. (a) presents the results for the sparse array from Fig. 4 using 6 RX antennas, and (b) gives the results for the second configuration using 4 RX antennas. In (c), the camera image of the two cylinders acting as point-like targets is shown.

the array configuration with only 4 RX antennas depicted in Fig. 10(b), the DNN is still capable of separating both targets. Due to the high aliasing, this is not feasible with the other approaches presented.

C. EVALUATION OF DETECTION AND FALSE-ALARM PROBABILITY

As a fourth test, the proposed DNN approach was evaluated in terms of precision, detection rate, and false alarm rate. These metrics are briefly described below.

1) The detection rate P_D, also called *recall*:

$$P_D = \frac{TP}{TP + FN}, \quad (15)$$

2) The false alarm rate P_{FA}:

$$P_{FA} = \frac{FP}{TN + FP}, \quad (16)$$

3) The precision, as the ratio between correct detections and all detections:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (17)$$

Here, TP and TN denote true positives and true negatives, respectively. FP and FN describe false positives and false negatives, respectively. The approach was again applied to

simulated data, which were neither used for training or validation. The test set consisted of 50 images, in which again each range bin with at least one detection in the ground truth data was considered.

The evaluation was conducted for each simulated dataset in the following way.

1) Create the range-Channel image and extract the necessary features for the deployed DoA estimators.
2) Apply the DNN approach on the complete test image and store the resulting image.
3) Iterate over all range bins and apply the beamforming algorithms (CNN and MUSIC). For the DNN approach take the selected line from the already processed image.
4) Apply a peak search algorithm for the currently selected range bin to find possible detections. As peak search algorithm, the `find_peaks` method from the python library scipy[2] was used. The sensitivity of the peak search was varied with the *prominence* parameter, which according to the documentation is *a measure of how much a peak stands out from the surrounding baseline*.[3] For more information about the algorithm and its

---

[2][Online]. Available: https://docs.scipy.org
[3][Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.peak_prominences.html
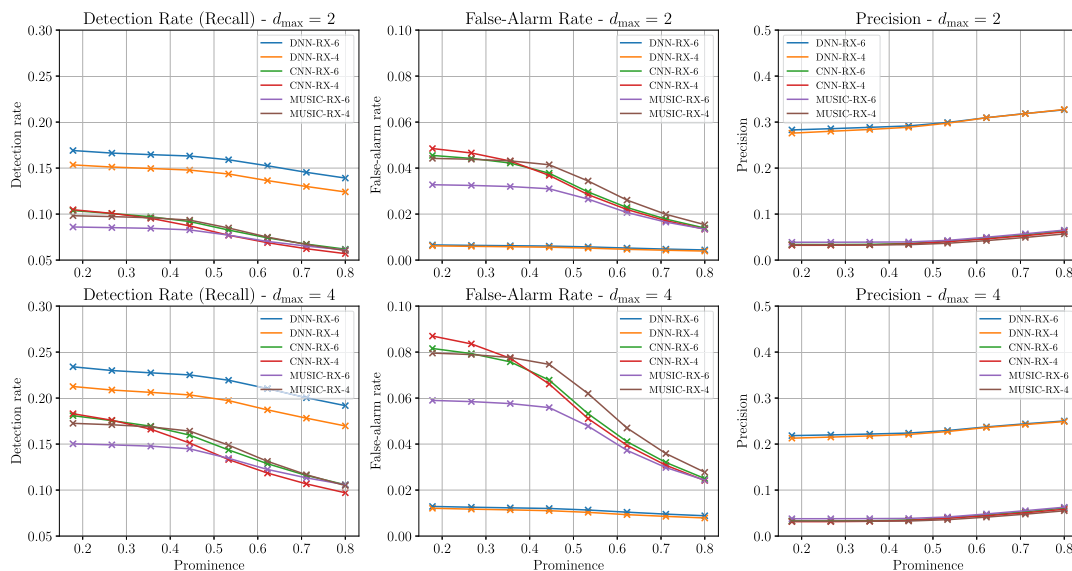
FIGURE 11. Evaluation of three metrics for the comparison of the DNN approach with the conventional CNN method. The columns indicate the specific metrics, which are detection rate, false alarm rate, and precision. The rows indicate the performance, for the specific metric, of the investigated algorithms. An estimated detection is counted if the distance to the ground truth detection is less than or equal to the maximum distance d_{max} in the angular bins.

parameters, the reader is refered to documentation of the algorithm.

5) A detected peak is counted as successfully detected (TP) if

$$p_{suc} = \begin{cases} TP, & |x_{est} - x_{true}| \leq d_{max} \\ FP, & \text{otherwise} \end{cases}, \quad (18)$$

where, x_{est} denotes the pixel position of the estimated target, x_{true} is the correct target position, and d_{max} describes the maximum tolerated distance in pixels between x_{est} and x_{true}. In the proposed evaluation scenario, $d_{max} = 2$ and $d_{max} = 4$ was tested. For the sake of simplicity, one estimated detection can be assigned to multiple ground truth detections. A false negative (FN) detection was counted if the a peak was present in the ground truth signal, but was not detected by the beamforming algorithm. In case no peak was detected and was not present in the ground truth signal, a true negative (TN) was counted.

Further, every spectrum assigned to the peak finding method was normalized to a range of 0 and 1. Since the CNN approach used for comparison in particular suffers from severe sidelobes, only peaks higher than 0.5 were counted. Even the proposed DNN beamforming did not considerably benefit from this additional threshold, this resulted in a fairer comparison.

The results are shown in Fig. 11. It is apparent that all algorithms exhibited a rather low detection rate. This may be due to the fact that the ground truth image was generated with a much higher angular resolution, which prevented smaller arrays from resolving every target. It is nonetheless obvious that the proposed DNN approach outperformed the

CNN method significantly. This is presumably due to the improved ratio between the mainlobe and the sidelobe magnitudes of the detected targets. Consequently, the peak search algorithm could more easily identify peaks as targets instead of noise or aliasing artifacts. Interestingly, the MUSIC approach for only 6 RX antennas showed a slightly worse detection performance compared by using only 4 RX antennas. This might be explained by the fact, that the aperture size for both antenna configurations is identical, but the MUSIC algorithm with only 4 antennas creates more sidelobes, which may be at the same position as real targets.

The same behavior could be observed for the false-alarm rate, which is considerably lower for the DNN approach compared to the MUSIC and CNN techniques. Even with an additional threshold during the peak search, the distortion from sidelobes was still very severe in the other approaches.

Contrary to the detection rate, the false-alarm rate of the MUSIC algorithm with 6 RX antennas is significantly better compared by using only 4 antennas. This fits our previous explanation, that the detections for 4 RX antennas mainly stems from sidelobes, which also lead to large a amount of false detections resulting in a high false-alarm rate. The same argumentation can be applied to the precision metric, which is similar for the MUSIC with the CNN technique and the DNN approach showing the best results. As expected, all beamforming algorithms exhibited a higher detection rate with increasing d_{max}.

Interestingly, in this comparison, the difference between both DNN approaches was not very significant. This may be explained by the dataset used for comparison, which contained simulated but quite realistic radar images, where the task of differentiating two closely spaced targets did not happen often. Therefore, even though the DNN trained with
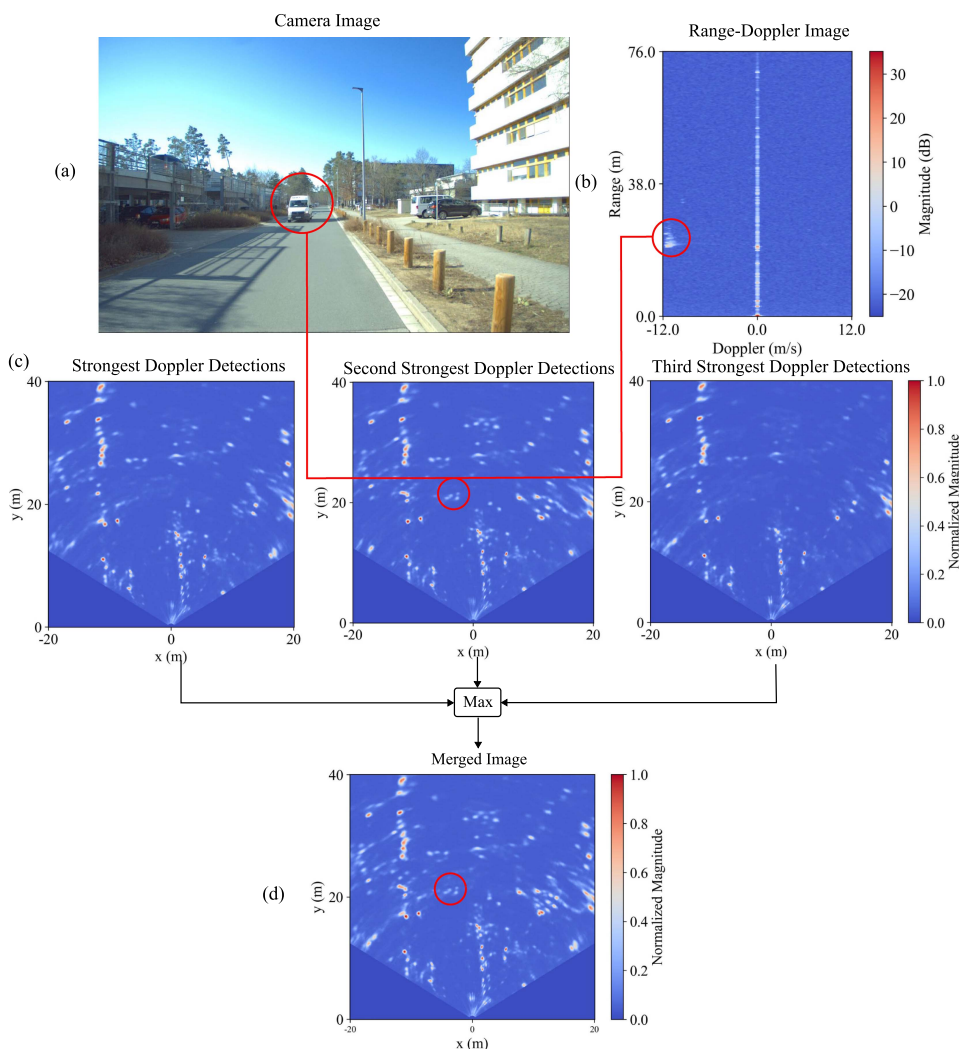
FIGURE 12. Image (a) shows the camera image with the approaching truck and image (b) shows the corresponding range-Doppler image, in which the truck can clearly be separated in the Doppler direction. The generated Doppler detections are sorted and used to generate three range-angle images by applying the DNN, as shown in image (c). All images are stacked onto each other and the maximum value for each pixel along these images is taken to generate a merged image , see image (d).

4 RX antennas had a lower resolution, it still showed a good sidelobe suppression, while maintaining a sufficient resolution for most of the tasks at hand.

D. EVALUATION OF A DYNAMIC SCENE

Here, the approach described in Section III-E with the DNN trained for 6 RX antennas is evaluated on a dynamic scene, in which a truck is moving towards the radar unit. In this case, not only the maximum Doppler is chosen, but the peak search algorithm utilized in the previous section was used to obtain up to three Doppler peaks for each range bin. The result is depicted in Fig. 12. The final image is created by stacking all three single images onto each other and taking the maximum projection along the new channel dimension, as shown in image d) in Fig. 12. Alternatively, also detections from each single image could be collected and written in a common detection list.

This example illustrates, that evaluating multiple Doppler peaks is a reasonable measure, since only in the second image

(Fig. 12(c) the detections originating from the truck are clearly visible.

VI. CONCLUSION AND FUTURE WORK

This work demonstrates that adequately preprocessed radar data from an extremely thinned aperture can be processed aliasing-free and with super-resolution using a DNN. For this, the DNN was specially trained with sufficient realistic simulation data obtained from advanced ray tracing simulations and with synthetically generated point-like targets. These targets were closely spaced to explicitly train the resolution capability of the DNN. With the two evaluated sparse array configurations, it could be seen that the proposed approach outperformed even state-of-the-art machine learning DoA algorithms and super-resolutions techniques such as MUSIC.

In the following, we discuss possible reasons why this approach is beneficial in the proposed scenarios and make some suggestions for further improvements.

The first advantage of the proposed approach is that the input data do not consist of only a single covariance matrix or a single snapshot of the antenna array data. Instead, the complete range-angle map and covariance information are given to the DNN, which can exploit specific spatial features encoded in the input data to improve the radar image result. In addition, due to advanced ray tracing simulation, the training data used are much more realistic than simple point-like targets used in other approaches.

However, there are some possible areas for further improvement. The proposed DNN is quite large, and it should be investigated if similar performance is possible with smaller architectures with fewer parameters.

Furthermore, optimized array configurations should also be analyzed, since the virtual antenna array configurations investigated were limited by the MIMO configurations, and in this work, only the RX channels were reduced. However, the fact that this approach performs well even on such unconventional array configurations opens up more flexibility to array design.

The evaluation of the proposed approach also revealed several promising aspects for future work. Because this approach works well for the unconventional array configurations presented, it provides an unprecedented flexibility for the array design, which often suffers from strict limitations in terms of spatial constraints, power consumption, calibration issues, and raw data size.

Additionally, since the virtual antenna positions are comparably widely spaced, this approach is promising for the fusion and joint DoA estimation of distributed radar sensors.

Finally, also, other more recent neural network architectures, such as vision transformer [55], [56], which have been already applied successfully to radar applications [57] could be investigated in respect of image quality and runtime performance in the future.

ACKNOWLEDGMENT

REFERENCES

[1] F. Rosique et al., "A systematic review of perception system and simulators for autonomous vehicles research," *Sensors*, vol. 19, no. 3, 2019, Art. no. 648.

[2] R. Prophet et al., "Pedestrian classification with a 79 ghz automotive radar sensor," in *Proc. IEEE 19th Int. Radar Symp.*, 2018, pp. 1–6.

[3] S. Heuel and H. Rohling, "Pedestrian classification in automotive radar systems," in *Proc. IEEE 13th Int. Radar Symp.*, 2012, pp. 39–44.

[4] A. Palffy, J. Dong, J. F. Kooij, and D. M. Gavrila, "CNN based road user detection using the 3D radar cube," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1263–1270, Apr. 2020.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[6] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[7] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar-algorithms and performance," in *Proc. IEEE Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput.*, 2009, pp. 464–468.

[8] F. Roos et al., "Compressed sensing based single snapshot DoA estimation for sparse MIMO radar arrays," in *Proc. IEEE 12th German Microw. Conf.*, 2019, pp. 75–78.

[9] A. Correas-Serrano and M. A. González-Huici, "Experimental evaluation of compressive sensing for DoA estimation in automotive radar," in *Proc. IEEE Int. Radar Symp.*, 2018, pp. 1–10.

[10] C. Alistarh, L. Anitori, W. L. van Rossum, S. K. Podilchak, J. Thompson, and M. Sellathurai, "Compressed sensing for MIMO radar using siw antennas for high resolution detection," in *Proc. IEEE 18th Eur. Radar Conf.*, 2022, pp. 485–488.

[11] B. Hu et al., "Off-grid DoA estimation based on compressed sensing on multipath environment," *Int. J. Antennas Propag.*, vol. 2023, 2023, Art. no. 9315869, doi: 10.1155/2023/9315869.

[12] A. De Maio, Y. C. Eldar, and A. M. Haimovich, *Compressed Sensing in Radar Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[13] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[14] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[15] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[16] T. Pavlenko, C. Reustle, Y. Dobrev, M. Gottinger, L. Jassoume, and M. Vossiek, "Design and optimization of sparse planar antenna arrays for wireless 3-D local positioning systems," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 7288–7297, Dec. 2017.

[17] D. Mateos-Núñez, M. A. González-Huici, R. Simoni, F. B. Khalid, M. Eschbaumer, and A. Roger, "Sparse array design for automotive mimo radar," in *Proc. IEEE 16th Eur. Radar Conf.*, 2019, pp. 249–252.

[18] B. Eisele, A. Bereyhi, I. Ullmann, and R. Müller, "A novel antenna placement algorithm for compressive sensing MIMO radar," in *Proc. IEEE Radar Conf. (RadarConf23)*, San Antonio, TX, USA, 2023, pp. 1–6.

[19] C. Schüßler, M. Hoffmann, I. Ullmann, R. Ebelt, and M. Vossiek, "Deep learning based image enhancement for automotive radar trained with an advanced virtual sensor," *IEEE Access*, vol. 10, pp. 40419–40431, 2022.

[20] C. Schüßler, M. Hoffmann, J. Bräunig, I. Ullmann, R. Ebelt, and M. Vossiek, "A realistic radar ray tracing simulator for large MIMO-arrays in automotive environments," *IEEE J. DoA*, vol. 1, no. 4, pp. 962–974, Oct. 2021.

[21] B. Friedlander, "Direction finding using an interpolated array," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1990, pp. 2951–2954.

[22] M. Hott, J. Mietzner, S. Lutz, M. Bockmair, and P. A. Hoeher, "Joint super-resolution and array interpolation for MIMO radar virtual arrays," in *Proc. IEEE 15th Eur. Radar Conf.*, 2018, pp. 126–129.

[23] H. Sim, S. Lee, S. Kang, and S.-C. Kim, "Enhanced DoA estimation using linearly predicted array expansion for automotive radar systems," *IEEE Access*, vol. 7, pp. 47714–47727, 2019.

[24] I. Orr, M. Cohen, H. Damari, M. Halachmi, M. Raifel, and Z. Zalevsky, "Coherent, super-resolved radar beamforming using self-supervised learning," *Sci. Robot.*, vol. 6, no. 61, 2021, Art. no. eabk0431.

[25] M. Eschbaumer, S. Achatz, and G. Balazs, "Analysis of a machine learning based virtual array augmentation technique for automotive radar," in *Proc. IEEE 19th Eur. Radar Conf.*, 2022, pp. 149–152.

[26] I. Roldan, F. Fioranelli, and A. Yarovoy, "Self-supervised learning for enhancing angular resolution in automotive MIMO radars," *IEEE Trans. Veh. Technol.*, early access, Apr. 21, 2023, doi: 10.1109/TVT.2023.3269199.

[27] J. Fuchs, M. Gardill, M. Lübke, A. Dubey, and F. Lurz, "A machine learning perspective on automotive radar direction of arrival estimation," *IEEE Access*, vol. 10, pp. 6775–6797, 2022.

[28] Y. Ma, Y. Zeng, and S. Sun, "A deep learning based super resolution DoA estimator with single snapshot MIMO radar data," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4142–4155, Apr. 2022.

[29] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DoA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[30] G. K. Papageorgiou, M. Sellathurai, and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low SNR," *IEEE Trans. Signal Process.*, vol. 69, pp. 3714–3729, 2021.

[31] M. Gall, M. Gardill, J. Fuchs, and T. Horn, "Learning representations for neural networks applied to spectrum-based direction-of-arrival estimation for automotive radar," in *Proc. IEEE/MTT-S Int. Microw. Symp.*, 2020, pp. 1031–1034.

[32] T.-H. Sang, F.-T. Chien, C.-C. Chang, K.-Y. Tseng, B.-S. Wang, and J.-I. Guo, "Doa estimation for FMCW radar by 3D-CNN," *Sensors*, vol. 21, no. 16, 2021, Art. no. 5319.

[33] Z.-M. Liu, C. Zhang, and P. S. Yu, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, Dec. 2018.

[34] S. R. Pavel, M. W. T. Chowdhury, Y. D. Zhang, D. Shen, and G. Chen, "Machine learning-based direction-of-arrival estimation exploiting distributed sparse arrays," in *Proc. IEEE 55th Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 241–245.

[35] S. R. Pavel and Y. D. Zhang, "Neural network-based compression framework for DoA estimation exploiting distributed array," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 4943–4947.

[36] Q. Cheng, A. A. Ihalage, Y. Liu, and Y. Hao, "Compressive sensing radar imaging with convolutional neural networks," *IEEE Access*, vol. 8, pp. 212917–212926, 2020.

[37] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, "Enhanced radar imaging using a complex-valued convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 35–39, Jan. 2019.

[38] Y. Dai, T. Jin, H. Li, Y. Song, and J. Hu, "Imaging enhancement via CNN in MIMO virtual array-based radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7449–7458, Sep. 2021.

[39] M. Salucci, M. Arrebola, T. Shan, and M. Li, "Artificial intelligence: New frontiers in real–time inverse scattering and electromagnetic imaging," *IEEE Trans. Antennas Propag.*, vol. 70, no. 8, pp. 6349–6364, Aug. 2022.

[40] P. Rostami, H. Zamani, M. Fakharzadeh, A. Amini, and F. Marvasti, "A deep learning approach for reconstruction in millimeter-wave imaging systems," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 1180–1184, Jan. 2023.

[41] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 1849–1860, Apr. 2019.

[42] K. Xu, L. Wu, X. Ye, and X. Chen, "Deep learning-based inversion methods for solving inverse scattering problems with phaseless data," *IEEE Trans. Antennas Propag.*, vol. 68, no. 11, pp. 7457–7470, Nov. 2020.

[43] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 98–117, Jul. 2020.

[44] X. Li, X. Wang, Q. Yang, and S. Fu, "Signal processing for TDM MIMO FMCW millimeter-wave radar sensors," *IEEE Access*, vol. 9, pp. 167959–167971, 2021.

[45] M. A. Richards, *Fundamentals of Radar Signal Processing*. New York, NY, USA: McGraw-Hill Educ., 2014.

[46] S. S. Ahmed, A. Schiessl, F. Gumbmann, M. Tiebout, S. Methfessel, and L.-P. Schmidt, "Advanced microwave imaging," *IEEE Microw. Mag.*, vol. 13, no. 6, pp. 26–43, Sep./Oct. 2012.

[47] B. Major et al., "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 924–932.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[49] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, Apr. 2018, pp. 1–10.

[50] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[51] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.

[52] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[53] V. Madisetti, *The Digital Signal Processing Handbook*. Boca Raton, FL, USA: CRC, 1997.

[54] M. Wang and A. Nehorai, "Coarrays, music, and the Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2017.

[55] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021.

[56] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[57] J. Giroux, M. Bouchard, and R. Laganiere, "T-fftradnet: Object detection with swin vision transformers from raw ADC radar signals," 2023, *arXiv:2303.16940*.

CHRISTIAN SCHUESSLER (Graduate Student Member, IEEE) was born in Fürth, Germany, in 1991. He received the B.Eng and M.Eng. degrees in electrical engineering and information technology from the Technische Hochschule Nürnberg Georg-Simon-Ohm, Nuremberg, Germany, in 2014 and 2015, respectively. Since 2019, he has been working toward his Ph.D. at the institute of Microwaves and Photonics (LHFT) at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. From 2015 to 2019, he was a software Engineer of the algorithmic Group of the Fraunhofer Center for X-Ray Development (EZRT) in the field of computer tomography. His research interests include radar simulation, signal processing, and machine learning, especially for large MIMO arrays in automotive applications.

MARCEL HOFFMANN (Graduate Student Member, IEEE) was born in Witten, Germany, in 1994. He received the B.Sc. and the M.Sc. degrees in electrical engineering in 2016 and 2018, respectively, from the Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, where he is currently working toward the Ph.D. degree. In 2018, he joined the Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, His research interests include radar signal processing and automotive radar applications including new SAR and SLAM approaches.

MARTIN VOSSIEK (Fellow, IEEE) received the Ph.D. degree from Ruhr-Universität Bochum, Bochum, Germany, in 1996. In 1996, he joined Siemens Corporate Technology, Munich, Germany, where he was the Head of the Microwave Systems Group from 2000 to 2003. Since 2003, he has been a Full Professor with Clausthal University, Clausthal-Zellerfeld, Germany. Since 2011, he has been the Chair of the Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. He has authored or coauthored more than 350 articles. His research has led to more than 100 granted patents. His research interests include radar, microwave systems, wave-based imaging, transponders, RF identification, communication, and wireless sensor and locating systems. Dr. Martin Vossiek is a Member of the German National Academy of Science and Engineering (acatech) and of the German Research Foundation Review Board. He is a Member of the IEEE Microwave Theory and Technology Technical Committees for MTT-24 Microwave/mm-wave Radar, Sensing, and Array Systems, MTT-27 Connected and Autonomous Systems (as founding Chair), and MTT-29 Microwave Aerospace Systems. He is also on the advisory board of the IEEE CRFID Technical Committee on Motion Capture & Localization. Dr. Martin Vossiek was the recipient of the numerous best paper prizes and other awards. In 2019, he was the recipient of the Microwave Application Award by the IEEE MTT Society for Pioneering Research in Wireless Local Positioning Systems. Dr. Vossiek is a Member of organizing committees and technical program committees for many international conferences and has served on the review boards of numerous technical journals. From 2013 to 2019, he was an Associate Editor for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES. Since October 2022, he has been an Associate Editor-in-Chief for IEEE TRANSACTIONS ON RADAR SYSTEMS.