Deep Learning Based Image Enhancement for Automotive Radar Trained with an Advanced Virtual Sensor

CHRISTIAN SCHÜSSLER1, (Graduate Student Member, IEEE), MARCEL HOFFMANN 1 (Graduate Student Member, IEEE), INGRID ULLMANN1, RANDOLF EBELT2, AND MARTIN VOSSIEK1 (Fellow, IEEE)

1Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
2BMW Group, Munich, Germany

Corresponding author: Christian Schüßler (e-mail: christian.schuessler@fau.de).

ABSTRACT This paper introduces a novel deep learning based concept for image enhancement and distortion suppression in automotive radar signal processing. The deep neural network (DNN) is trained solely on virtual data that is generated by an automotive MIMO radar ray tracing simulator. The simulator mimics the raw data that would be provided by a specific automotive MIMO radar for which the signal processing is envisaged. This virtual radar sensor, which has the same properties as the real radar, creates the DNN training input data. The unique feature of our approach is that an advanced virtual radar sensor is used to create the target values for the DNN training. The advanced virtual sensor is a simulated radar that works on the same basic principle as the real radar and its resembling virtual sensor, but is significantly upgraded in terms of performance. In our experiments the advanced virtual sensor, e.g., has a notably larger array, i.e., a much better lateral resolution, than the real radar. In addition, our advanced virtual sensor is not affected by multi-path effects, clutter, noise, beamforming sidelobes and other typical automotive radar distortions. The paper shows that a DNN trained on this advanced virtual sensor training data can deliver outstanding automotive radar signal processing results not only on simulated data but also on field data. The presented real-world automotive radar measurements show that multi-path, clutter and noise are efficiently suppressed by this DNN and that it has a remarkable ability to sharpen images and suppress sidelobes. The results presented in the paper suggest that the novel advanced DNN training concept based on virtual sensors offers opportunities that also go far beyond the radar signal processing capabilities illustrated. Hence, the concept put forward here is an attractive option for placing DNN-based radar signal processing on a novel and easy-to-implement foundation.

INDEX TERMS Radar Imaging, Automotive Radar, Deep Neural Networks, Image Enhancement, Radar Simulation

I. INTRODUCTION

Together with cameras and lidars, radar sensors are one of the most important sensors for advanced driver assistance systems (ADAS). When it comes to range resolution and operations in harsh weather conditions, they show superior behavior [1], [2]. Another important feature is that these sensors are able to directly measure the target’s radial velocity exploiting the Doppler effect. They can therefore separate closely located targets and differentiate between entities, such as pedestrians, cars and bicycles by their micro Doppler signature [3].

Despite these properties, radar sensors have several problems when it comes to angular resolution, noise, clutter or interference with other radar sensors. For all of these errors it was shown that they can be alleviated by applying machine learning algorithms. Angular resolution, or direction of arrival (DoA) estimation can be improved by multiple kinds of neural networks as for example shown in [4], [5], [6], and [7]. Also in the domain of radar imaging it was shown that DNNs trained by Compressed Sensing (CS) images could improve the image quality, see [8], [9] and [10]. Similar image enhancement approaches can be found for (inverse)
synthetic aperture radar (I)SAR applications in [11], [12], and [13].

Interference between sensors is becoming increasingly challenging the more vehicles are being fitted with radar sensors. These disturbances can also be mitigated by DNNs, as shown in [17].

Furthermore, ghost targets caused by multi-path or clutter are also a common source of false object detection in the automotive domain. Current approaches are conventionally based on annotated detection point clouds, as shown in [14], [15], and [16].

All of these examples often outperform conventional methods and show that DNNs are a vital approach for mitigating typical disturbances and errors in the radar domain.

However, these approaches still have some drawbacks. Existing radar image enhancement applications especially are trained solely on point-like targets and evaluated only on very limited measurement data. Clutter originating from multi-path effects is hard to model or to annotate, especially on raw radar signals or images, but plays a significant role in automotive applications. Moreover, a simple solution that can mitigate multiple or even all disturbances at the same time without the need for time-consuming manual annotation is still missing.

Recently, it became possible to simulate realistic automotive radar data by utilizing ray tracing techniques in an acceptable amount of time, as shown in [18] and [19]. In this work the simulator described in [19] is taken, which was designed for realistic automotive scenarios. The simulator is also suitable for very large MIMO arrays.

In this work realistic radar simulations are utilized to train a DNN, so that it is able to enhance real world measurements even if it was trained solely on synthetic data. This is done with two virtual sensors that both resemble a physical sensor. One of the sensors shares almost all properties with the original sensor and the other one – an advanced virtual sensor – shares only some of the properties with the original sensor, but otherwise has superior properties itself. The proposed approach does not need any manual annotations and can sharpen the input image, suppress sidelobes, remove electronic noise, and clutter all at the same time.

This article is structured as follows: In Section II a brief overview of the overall approach is given. The simulation data generation is described in Section III. Section IV describes the DNN used for the image enhancement approach. The results on how the proposed DNN approach performs on simulated and real measurement data are presented in Section V. In the last section, the work is summarized and an outlook for future research is given.

II. CONCEPT

The idea behind this concept is to simulate a virtual sensor of the real radar sensor, which resembles its counterpart as much as possible. To enhance the sensor data, another advanced virtual sensor is simulated. This virtual sensor does not actually resemble a physical radar, instead it generates data that would be desired from a perfect radar sensor (e.g., clutter free, high resolution, ...). The simulated data of both virtual entities is then used to train a DNN. This DNN tries to enhance the input images of the virtual sensor so that they resemble the ground truth images produced by the advanced virtual sensor.

With this approach we intend to increase the angular resolution, improve the SNR and reduce clutter all at the same time. Since the virtual sensor resembles the real radar sensor, the trained DNN, which has never seen any real measurement can be applied on real measurements directly.

The complete workflow is depicted in Fig. 1. In the following, the name of the images generated by the virtual sensor is set to input images, the images produced by the advanced virtual sensor are named ground truth images, and the images that are measured by the radar sensor are named measured images. Enhanced images describe the output data of the DNN, no matter whether the input comes from simulation or real radar measurements.

The parameters of the real and virtual radar sensors are shown in Table 1. The virtual array of the real radar generated by the MIMO scheme is a uniform linear array (ULA) using a time division multiplexing (TDM) approach. For the sake of simplicity and runtime performance no multiplexing scheme was simulated for the virtual sensor, since in the simulation all channels are separated automatically, as described in [19]. All antennas are placed solely in azimuth direction and therefore no elevation information is available in this measurement setup.

III. DATA GENERATION

To generate suitable training data for the DNN, two main challenges have to be addressed. The first one is to model the environment in the simulation accurately enough so that the trained DNN can also be applied to real-world problems. Secondly, high-resolution and high-quality ground truth data...
TABLE 1. Radar sensor parameters for different types of radars. The virtual sensor should resemble the real radar sensor as much as possible.

| Symbol | Parameter | Real Radar | Virtual Sensor | Advanced Virtual Sensor |
|--------|-----------|------------|----------------|-------------------------|
| $f_c$  | RF carrier frequency | 77 GHz | 77 GHz | 77 GHz |
| $B$    | RF bandwidth | 1 GHz | 1 GHz | 1 GHz |
| $N_{chirp}$ | number of chirps | 1 | 1 | 1 |
| $T_c$  | chirp duration | 400 $\mu$s | 400 $\mu$s | 400 $\mu$s |
| $N_{TX}$ | number of TX antennas | 3 | 3 | 1 |
| $N_{RX}$ | number of RX antennas | 16 | 16 | 256 |
| $d_{TX}$ | TX antenna distance | 2 mm | 2 mm | — |
| $d_{RX}$ | RX antenna distance | 6 mm | 6 mm | 1.9 mm |
| $\alpha$ | angular resolution | 5.3 deg | 5.3 deg | 0.4 deg |
| $n$    | electronic noise | yes | yes | no |

has to be generated so that the DNN can enhance the input images accordingly. Both tasks are addressed in the following subsections.

A. ENVIRONMENT MODEL

As mentioned before, the radar data simulation is based on the implementation in [19]. For the simulation environment a static 3D mesh of a virtual city from the open source automotive driving simulator CARLA [20] was exported. The simulated map can be seen in Fig. 2. Red cubes were distributed over the complete map to mark the radar positions. At each simulation position the radar was rotated by 5 degrees around its vertical axis to measure the simulation environment from different viewing angles. Each radar position and orientation produces a single radar image, which is used as a DNN input. By using this setup over 100,000 radar image samples can be generated. However, 10,000 randomly taken data sets from all possible positions and orientations turned out to be sufficient.

While buildings and entities such as traffic lights and even trash bins were modeled quite accurately in the simulation environment, vegetation, especially bushes are only sparsely distributed over the map and also often lack the detail to produce realistic results, see Fig. 3. Scattering models for vegetation have been available for a long time, see for example [21], [22] or [23]. Therefore, according to the volume scattering model in [21], small cylinders were distributed with random position and orientation inside a predefined box to emulate the reflectivity of bushes or grass patches. With this method, it is possible to satisfactorily simulate vegetation and since no Doppler is considered in the proposed approach, motion does not have to be considered.

B. SIGNAL MODEL

The simulator generates the intermediate frequency signal (IF signal or beat signal) of the frequency modulated continuous wave (FMCW) radar signal scheme as given by

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

(1)

A more detailed description of signal processing for automotive radar sensors can be found in [24] and [25].

Since the simulation is able to separate each TX and RX channel (MIMO scheme), the IF signal data is ordered in a three dimensional array with dimensions in fast time $i_t$, TX antenna index $i_{TX}$ and RX antenna index $i_{RX}$. The carrier frequency of the signal is denoted by $f_c$ and $\mu$ describes the frequency slope

$$ \mu = \frac{B}{T_c}, $$

(2)

which is the ratio between the bandwidth $B$ to the chirp duration $T_c$ of the signal. The round-trip delay $\tau$ between TX antenna $x_{i_{TX}}$, reflecting object position $x_k$, and RX antenna $x_{i_{RX}}$ is denoted as

$$ \tau_k(i_{TX}, i_{RX}) = \frac{||x_{i_{TX}} - x_k||^2 + ||x_{i_{RX}} - x_k||^2}{c}. $$

(3)

In the presented work, we are only interested in spatial dimensions, therefore no Doppler information is simulated and taken into account here.

C. SIMULATED INPUT IMAGE GENERATION

The simulated input data $s_{IF}$ is generated by a virtual sensor consisting of several TX and RX antennas in horizontal direction operating in MIMO mode.

Clutter and multi-path effects are already included in the simulated data, but electronic noise was not considered in the initial simulation, therefore a specific amount of normal distributed complex noise $n(i_c, t)$ was added to the initial signal

$$ s_{IF}(i_c, t) = s_{IF}(i_c, t) + n(i_c, t). $$

(4)

The amount of electronic noise varies with the number of transmitted chirps and antenna channels. Therefore, the noise power in the simulations is chosen according to real-world measurements.

To generate a radar image in range and angular directions, a two dimensional Fourier reconstruction over the fast time and the antenna channels is utilized. In order to apply a Fourier transform over the antenna channels, all TX and RX antenna combinations can be sorted to a continuous virtual array, resulting in a two dimensional data array (dimensions are channel and fast time). The virtual antenna channels are denoted by the index $i_c$. The complete IF signal is then padded and windowed by a 2D hanning window resulting in
FIGURE 2. In the left image (a) a screenshot of the complete simulation map is shown. Small red cubes were distributed over the complete 3D scene, marking the radar positions, as also shown in the right image (b).

(a) Overview of the simulation
(b) Simulated radar sensor positions

FIGURE 3. The upper bush (a) is imported from CARLA and only consists of very few triangles. The lower bush (b) is modeled by distributing randomly orientated cylinders in a predefined box.

(a) Bush mesh taken from CARLA
(b) Volume scatter model

$S_{IF,\text{pad}}$, before a two dimensional DFT is applied to generate the final input image

\[
\hat{I}(i_r, i_c) = \text{DFT}_{2d}(S_{IF,\text{pad}}(i_c, i_t))
\]

\[
= \sum_{i_c=0}^{N_c-1} \sum_{i_t=0}^{N_t-1} \exp(-2\pi j \left( \frac{i_r i_t}{N_t} + \frac{i_r i_c}{N_c} \right)) S_{IF,\text{pad}}(i_c, i_t).
\]

The final input image shows the targets in range and angular directions, denoted by the indices $i_r$ and $i_c$, respectively. Whereby the angle $\alpha$ is encoded in a sine function (27, p. 515).

This approach is common in automotive applications as the fast Fourier transform is a comparatively fast algorithm. However, it suffers from inaccuracies for close range objects and large apertures, where the electromagnetic waves cannot be approximated as plane waves anymore (no far field condition). Furthermore, it suffers from wider main lobes and higher sidelobes compared to techniques such as backprojection or compressed sensing (CS).

D. GROUND TRUTH IMAGE GENERATION

The idea is to provide the network with Fourier based reconstructions as input images, which are common in automotive radar applications [26]. Afterwards, the images will be improved by a DNN, which is trained by high resolution, clutter free radar images, obtained by a more time-consuming but more accurate reconstruction algorithm. In order to obtain these high accurate ground truth images a backprojection algorithm [28] augmented with some pre-processing (zero-padding and windowing) was chosen.

These images are generated by a single TX antenna and 256 RX antennas, resulting in an almost five times higher angular resolution than in the input images. All other parameters are equal to the input image simulation. Since only a single chirp is simulated, the signal can be flattened into a two dimensional array of size $N_t \times N_c$. In the first step, the signal is windowed and zero-padded by an integer factor of $n_{\text{pad}}$ along the time dimension

\[
S_{IF}(i_r, i_c) \xrightarrow{\text{Window}} S_{IF,\text{pad}}(i_r, i_c).
\]

After that, a DFT operator is applied in order to obtain the spectrum

\[
S_{IF}(i_r, i_c) = \sum_{i_c=0}^{N_c} S_{IF,\text{pad}}(i_t, i_c) \exp(-2\pi j i_r i_t / N_t).
\]

The applied hanning window is required to reduce the sidelobes in range direction. Sidelobes in angular direction would still cause remarkable imaging artifacts, which may significantly degrade the performance of the trained DNN. Therefore another window in antenna (channel) direction is applied

\[
S_{IF}(i_r, i_c) = w(i_c) S_{IF}(i_r, i_c).
\]

Since the ground truth image must be directly compared with the input image, the final reconstruction grid has to be
in the same polar space. Note that the azimuth angle $\alpha$ is encoded into a sine function to resemble the Fourier based reconstruction used for the input images. Every sample point $x_s$ is defined as

$$x_s = \left( \frac{r}{\sin(\alpha)} \right). \quad (8)$$

Each sample point $x_s$ is converted directly into a delay $\tau$ by firstly converting it into Cartesian coordinates and applying equation (3) afterwards. As can be seen, the delay depends on the sample point and the selected antenna combination. The range index $i_r$ for a given $\tau$ can be computed by

$$i_r = \left\lfloor \frac{\tau(x_s, i_c)\mu T_c n_{pad}}{c} \right\rfloor, \quad (9)$$

whereby $\lfloor \cdot \rfloor$ is the floor operator. The backprojection algorithm can then be stated as a matched filter between the signal $S_{IF}$ and the hypothetically expected phase of a single point scatter at $x_s$.

$$I(x_s) = \sum_{i_c=0}^{N_{RX}} S_{IF}^{\mathbb{R}}(i_r, i_c) \cdot \exp\left(-\frac{2\pi j f_c \tau(x_s, i_c)}{c}\right), \quad (10)$$

resulting in the final ground truth image $I(x_s)$.

To further improve the result, the signal is linearly interpolated between the two adjacent signal values $S_{IF}(i_r, i_c)$ and $S_{IF}(i_r+1, i_c)$, whereby magnitude and phase are interpolated separately

$$\text{mag}\{S_{IF}^{\mathbb{R}}(i_r, i_c)\} = \text{mag}\{S_{IF}(i_r, i_c)\} + T_n_{pad}(\text{mag}\{S_{IF}(i_r+1, i_c)\} - \text{mag}\{S_{IF}(i_r, i_c)\}) \cdot (\mu \tau - \frac{i_r}{T_n_{pad}}), \quad (11)$$

$$\text{arg}\{S_{IF}^{\mathbb{R}}(i_r, i_c)\} = \text{arg}\{S_{IF}(i_r, i_c)\} + T_n_{pad}(\text{arg}\{S_{IF}(i_r+1, i_c)\} - \text{arg}\{S_{IF}(i_r, i_c)\}) \cdot (\mu \tau - \frac{i_r}{T_n_{pad}}). \quad (12)$$

For phase interpolation, $2\pi$ periodic jumps have to be considered. In contrast to the input data, the output data is simulated without noise and clutter. The clutter caused by multi-path effects is avoided by allowing only two ray hits in the ray tracer program, namely one hit at the measured object’s surface and one hit at the RX antenna.

The remaining sidelobes can be reduced even further by applying a simplistic static image processing pipeline on the logarithmized reconstructed image as depicted in Fig. 5. Before a histogram based threshold is applied, the image is weighted in radial direction to take the reduced signal strength in range direction into account as indicated by the radar equation. The final ground truth image is a clean and sharp radar image. A comparison between some selected simulated input and ground truth images is shown in Fig. 4.

IV. DEEP NEURAL NETWORK

For the purpose of this work, a neural network which already showed good performance in radar applications [29] and was originally proposed by [30] for medical image processing is used. In comparison to [29] almost the same version as proposed in the original work was implemented, since it already showed satisfying results. The complete network with the corresponding image sizes is shown in Fig. 6.

There are only two differences. Firstly, the network is augmented with an up-sampling layer to ensure that the final image has the same shape as ground truth image. A bi-linear interpolation is used to re-sample the images. Secondly, the final image has only one output channel and the network is used for a regression not a segmentation task. Therefore, the Square Error (SE) between the ground truth $X_{true}$ and the interfered or enhanced image $X_{proc}$ is optimized as shown below.

$$l = \|X_{proc} - X_{true}\|_2. \quad (13)$$
The logarithmized magnitude of the complex valued radar images was taken as DNN input. Additionally, the network was trained by an ADAM optimizer [31] with a learning rate of $10^{-4}$ and a batch size of 2 for 7 iterations. The training was stopped after 7 iterations, because afterwards the DNN started to overfit (i.e. the train loss was reduced but the test loss started to increase). This issue was alleviated by using more training data and a number of 10 000 radar images lead to visually appealing results and therefore the amount of training data was not further increased. All our implementations are realized with the deep learning framework PyTorch.

V. RESULTS
Firstly, the trained DNN is evaluated on two simulated metal spheres to investigate the resolution of the DNN-based image reconstruction. Both spheres were simulated with a radius of 0.5 m at a distance of 20 m from the radar unit. Random noise was added as described in Section III. The result is shown in Fig. 7. Clearly the DNN removes the background noise completely and is able to distinguish between the two targets, even when they are inseparable in the simulated input image.

Similar results were also obtained in a trial with two metal cylinders, as shown in Fig. 8. However, as the two cylinders became completely inseparable in the input image, the DNN was not able to distinguish between both objects. Therefore, even though there seems to be a gain in resolution to some extent, we cannot claim that the trained DNN achieves super-resolution e.g., comparable performance to subspace methods, such as MUSIC [32] or ESPRIT [33]. This is not surprising, since the DNN was only trained on magnitude images and was not designed explicitly for this task. Nevertheless, the reduction in sidelobes and clutter is clearly visible.

In the next step, the DNN is evaluated on simulated test data, which was generated with the same parameters and simulation environment as the training data, but was not used to train the DNN. A selection of results can be seen in Fig. 13.

To quantify the clutter and noise removal performance of the DNN, the signal-to-noise ratio (SNR) was computed for the images shown in Fig. 13. In typical image processing applications the SNR is defined as the power of the ground

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{The complete DNN, very similar to the original work in [30]. The first dimension represents the image channels, the second angular indices, and the third range indices.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Two simulated metal spheres with radius 0.5 m at a distance of 20 m from the radar unit. In image (a) with angular distance of 10 deg, in image (b) 5 deg, and in image (c) 3 deg from the sphere's center, respectively. Note image (c) where the DNN picked up two targets, which cannot be separated in the initial input image. The magnitude of all images was normalized between 0 and 1.}
\end{figure}
domain is the detection rate key figures to evaluate the detection performance in the radar task in the image processing domain. The most common two can be considered as a detection or as a binary segmentation separate targets from noise and clutter ([27], p.295), which seen.

The results for the detection and false alarm rates are depicted in Fig. 9 and Fig. 10, respectively. The detection and false alarm rates were computed for 50 simulated images and averaged. It can be seen that with acceptable false alarm rates the detection performance of the enhanced images is considerably higher. Even considering that a significant number of false alarms in the input image segmentation originates from the higher mainlobe width, the segmentation of the enhanced images leads to better detection performance especially for small and distant targets, as shown in Fig. 11.

TABLE 2. SNR of the input- and enhanced images for simulated data in dB. All images were scaled between 0 and 1 before the SNR was evaluated.

| Dataset       | 1    | 2    | 3    | 4    | 5    | 7    |
|---------------|------|------|------|------|------|------|
| SNR Input     | -17.0| -17.3| -19.8| -19.1| -20.3| -16.9|
| SNR Enhanced  | -4.4 | -1.7 | -5.7 | -2.9 | -5.56| -4.1 |

truth image divided by the noise of the corrupted image ([34], p. 164)

\[
\text{SNR}(I, \bar{I}) = 10 \log_{10} \frac{\sum_{x_s} I^2(x_s)}{\sum_{x_s} |I(x_s) - \bar{I}(x_s)|^2} . \tag{14}
\]

In our case, the ground truth image is represented by the simulated advanced virtual sensor data and the corrupted image is represented by the input- and enhanced images, respectively. The SNR values for the example datasets are stated in Table 2, showing an average SNR improvement of 14 dB.

Clearly, the network performs very well on simulated data, it seems to remove the clutter completely and improves the resolution of the image in general. Almost no difference between interfered and simulated ground truth data can be seen.

Another primary task in the radar signal processing is to separate targets from noise and clutter ([27], p.295), which can be considered as a detection or as a binary segmentation task in the image processing domain. The most common two key figures to evaluate the detection performance in the radar domain is the detection rate \( P_D \) (true positive rate) and the false-alarm rate \( P_{FA} \) (false positive rate).

The detection rate \( P_D \) is defined as below

\[
P_D = \frac{TP}{TP + FN}, \tag{15}
\]

whereby \( TP \) is the number of true positives (hits) and \( FN \) is the number of false negatives (misses). With \( FP \) being the number of false positives, \( P_{FA} \) is defined as

\[
P_{FA} = \frac{FP}{TN + FP}. \tag{16}
\]

For radar measurements that include Doppler information, a constant false alarm rate (CFAR) algorithm is commonly applied in range Doppler direction ([35] and targets are resolved in angular direction afterwards. Since this work focus mostly on angular resolution omitting Doppler processing, the detection algorithms are adapted and directly applied in the range angular domain.

For the input radar images, the detection task was performed with two ordered statistic CFAR (OS-CFAR) algorithms with different kernel sizes. The noise level was computed by taking the center value of the ordered noise values (median). The final threshold is then chosen by multiplying the noise estimation by a factor alpha. Also a constant threshold over the complete image was analyzed by taking the median of the complete image and multiplying it with the factor alpha.

For the enhanced radar image, no CFAR was applied since noise in these images is almost non-existent and therefore no useful noise statistic can be calculated. Instead, a threshold using Otsu’s method ([36] suitable for segmentation tasks was computed and varied by factor alpha. Unlike as for the enhanced image, using a threshold based on Otsu’s method for the input image segmentation was not considered in the evaluation since it performed considerably worse than the other methods.

The real radar measurement results are in Fig. 14. The measurements were taken at the university campus and are therefore not as dense as a crowded city environment. Hence, most of the clutter is expected to come from vegetation and not from multi-path between buildings and other traffic objects. Still, from a visual point of view the DNN performs almost as well as for the simulated data. Even weakly reflecting objects seem to be preserved to a certain extent, as shown in image (b) in ellipse (c). Sidelobes are, nevertheless, not eradicated for strongly reflecting targets (traffic signs). They even seem to be generated artificially as in ellipse (d) in image (b). This is because even in the generated ground truth images sidelobes cannot be suppressed completely and are therefore learned to some extent. Other artifacts appear
VI. CONCLUSION

We showed that deep neural networks trained with two virtual sensors, simulating complex urban automotive radar data, are able to improve real-world radar measurements. For simulated input data, the trained network was able to almost resemble the ground truth image obtained by an advanced virtual sensor with a much higher angular resolution. Even on real measurement data, the DNN is not only able to remove background clutter almost completely, but it also sharpens the image in general. The occurrence of clutter or ghost detections can be alleviated by this approach and real objects can also be separated more easily.

However, while the DNN tries to resemble high resolution data, it adds some artificial structures, especially connecting single targets into lines resembling borders. Even though the conducted test measurements do not show any misleading artifacts, a much broader investigation has to be conducted before deploying this concept commercially. This is necessary to guarantee that this concept does not produce safety-critical ghost targets.

This approach could be further improved by incorporating elevation information and extending 2D images to 3D data. This would give the neural network even more information to separate objects from clutter. Especially for large antenna beam widths in elevation, objects above the road (tree tops, bridges, ...) might be projected onto the 2D image and might confuse or mislead the DNN.

In the future Doppler information can be incorporated in the simulation. With this approach it would be possible to train and evaluate the proposed algorithm on a more common signal processing chain [4] i.e. range-Doppler detection followed by an angle estimation for each target. However, this would not only increase the simulation effort, but also the amount of radar data considerably and therefore the design of the DNN might be reconsidered.

Finally, we want to outline that this concept enables not only the removal of single errors, such as lack of resolution or disturbances, but it is also possible to train a DNN, which removes all simulated errors at once, by modeling them in a virtual sensor and ignoring them in an advanced virtual sensor.

Therefore, this approach is not limited to our example, but may be extended to correct multiplexing artifacts, calibration errors or interference disturbances caused by other radar sensors.

ACKNOWLEDGMENT

The authors would like to thank the Symeo team from indie Semiconductor (Jannis Groh, Martin Schütz, and Mark Christmann) for their support with the radar system used for tests.
FIGURE 11. Different segmentation results for an alpha value of 1.3. For all tested approaches the segmentation using the input radar image leads to inferior results compared to the segmentation of the enhanced image. Besides having larger sidelobes in the input segmentation, especially weak targets at greater distances cannot be detected with a low false alarm rate as marked by red ellipses.

FIGURE 12. The red labeled ellipses point to specific region of interests (ROIs) in the camera image, the input measurement image, and the enhanced image, respectively. The DNN is able to the enhance dynamic objects in a similar way as static objects. Even for weak radar responses as can be seen in image (a) and (b), the objects are clearly detected as objects and are not removed as clutter. Further, the contour of the truck in image (c) is clearly detected and enhanced.
FIGURE 13. The trained DNN evaluated on simulated data. The input images from the virtual sensor could be drastically improved in terms of noise, clutter and image sharpness. Besides still looking slightly more blurred than the ground truth images (advanced virtual sensor) they resemble them very well in general. Some examples of multi-path and clutter are marked by red ellipses. In most cases the clutter could be eliminated or at least attenuated considerably.
FIGURE 14. The red labeled ellipses point to specific region of interests (ROIs) in the camera image, the input measurement image, and the enhanced image, respectively. The red arrows point to artifacts that seem to originate from antenna calibration errors, which cannot be eradicated by the DNN. The trained DNN can reduce most of the noise and clutter in the image, while making the image considerably sharper.
MARCEL HOFFMANN (Graduate Student Member, IEEE) was born in Witten, Germany, in 1994. He received the B. Sc. and the M. Sc. degree in Electrical Engineering from the Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany in 2016 and 2018, respectively. In 2018, he joined the Institute of Microwaves and Photonics (LHFT) at the Friedrich-Alexander University Erlangen-Nürnberg, where he is currently working toward his Ph.D. His research focuses on radar signal processing and automotive radar applications including new SAR and SLAM approaches.

INGRID ULLMANN received the M.Sc. degree and Ph.D. degree from Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany in 2016 and 2021 respectively. She is currently a post-doc at the Institute of Microwaves and Photonics at Friedrich-Alexander University. Her research interests include radar imaging and radar signal processing for nondestructive testing, security, medical and automotive applications. She is group leader of the research group “Radar and Imaging Systems” at the Institute of Microwaves and Photonics and serves as a reviewer for the European Radar Conference (EuRAD) and various journals in the field of microwaves. Ms. Ullmann received the Argus Science Award (sponsored by Airbus Defense and Space, now Hensoldt) for her master thesis in 2016 and the Best Paper Award of the European Radar Conference in 2019.

RANDOLF EBELT received the Ph.D. degree from the Institute of Microwaves and Photonics, University of Erlangen–Nuremberg, Erlangen, Germany, in 2014. He is currently the product owner of objects, freespace and computer vision at BMW Group and is responsible for the environmental model for autonomous driving. He is the author of several publications on wireless positioning and radar signal processing.

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 (LHFT), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. He has authored or coauthored more than 250 articles. His research has led to more than 90 granted patents. His current research interests include radar, transponder, RF identification, communication, and locating systems. Dr. Vossiek has been a member of organizing committees and technical program committees for many international conferences. He is a member of the German IEEE Microwave Theory and Techniques (MTT)/Antennas and Propagation (AP) Chapter Executive Board and the IEEE MTT-S Technical Coordinating Committees MTT-24, MTT-27, and MTT 29. He was the Founding Chair of the MTT IEEE Technical Coordinating Subcommittee MTT-27 Wireless-Enabled Automotive and Vehicular Application. He was the recipient of several international awards. Recently, he was awarded the 2019 Microwave Application Award from the IEEE MTT Society (MTT-S). He has served on the review boards for numerous technical journals. From 2013 to 2019, he was an Associate Editor for IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.