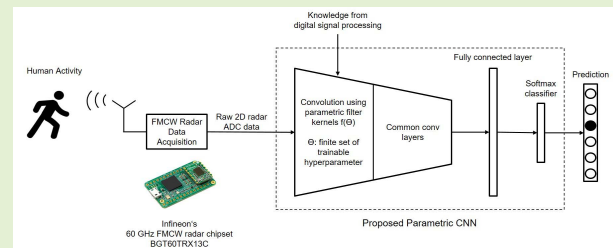


Data-Driven Radar Processing Using a Parametric Convolutional Neural Network for Human Activity Classification

Thomas Stadelmayer¹, Avik Santra², *Senior Member, IEEE*, Robert Weigel¹, *Fellow, IEEE*, and Fabian Lurz³, *Member, IEEE*

Abstract—The paper proposes a data-driven pre-processing optimization for radar data using a parametric convolutional neural network. The proposed method is applied on human activity classification as a use case. Present radar-based activity recognition system exploit micro-Doppler signature by generating Doppler spectrograms or a temporal series of range-Doppler maps, followed by deep neural networks or machine learning approaches for classification. Those radar data representations are typically generated on the basis of short-time Fourier transformations. A Fourier transformation equally resolves the frequency space, which may be sub-optimal in some applications. Although deep convolutional neural networks (DCNN) have been shown to implicitly learn features from raw sensor data in other fields, such as speech recognition, yet, for the case of radar-based DCNNs, pre-processing is required to develop a scalable and robust classification or regression application. In this paper, we propose a parametric convolutional neural network that mimics the radar pre-processing across fast-time and slow-time radar data through 2D sinc filter or 2D wavelet filter kernels to extract features for classification of various human activities. During training only the filter parameters of the 2D sinc filters or 2D wavelets are learned, leading to optimized feature representation for the classification task. It is demonstrated that our proposed solution shows improved results compared to equivalent DCNN architectures that rely on Doppler spectrograms or radar data cubes as input data.

Index Terms—2D sinc filters, 2D wavelets, parametric CNN, deep learning, human activity classification.



I. INTRODUCTION

PEOPLE sensing and activity classification have increasing application potential in various areas, such as physical security, defense and surveillance. In industrial and consumer space, human activity recognition finds applications in smart homes, human-machine interfaces and elderly fall-motion monitoring systems. Knowledge of the performed activity in a room can enable smart control of the energy consumption,

such as HVAC and lighting [1]–[3]. Furthermore, knowledge of the performed human activity facilitates true ubiquitous smart home solution by discerning the user's intent.

Most of the human activity recognition systems are based on cameras and computer vision approaches. These systems have the advantage that they are quite easy to implement and benefit from several years of research. As a result the accuracy of such systems is quite high. However, camera systems suffer from lack of privacy and are sensitive to illumination conditions, thus they are not a favorable choice for smart home solutions. On the other hand, radar sensors have shown to be an effective sensing modality for human activity classification [4], [9]–[17], [26]. The radar sensors offer privacy-preserving, illumination-invariance properties and are capable of being aesthetically concealed in the operating environment. Recent innovation in the semiconductor technologies have facilitated integration and antenna-in-package solutions making the radar sensor into a small form factor [18].

Radar sensors can sense and recognize human activities utilizing micro-Doppler signatures [24] that are generated by non-rigid-body motions of moving targets. To efficiently classify different human activities, meaningful features have to be extracted from the micro-Doppler signatures. In [4] the authors

Manuscript received May 17, 2021; accepted June 18, 2021. Date of publication June 24, 2021; date of current version August 31, 2021. The associate editor coordinating the review of this article and approving it for publication was Dr. Chirasree Roychaudhuri. (Corresponding author: Thomas Stadelmayer.)

Thomas Stadelmayer is with the Institute of Electronics Engineering, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany, and also with Infineon Technologies AG, 85579 Neubiberg, Germany (e-mail: thomas.stadelmayer@fau.de).

Avik Santra is with Infineon Technologies AG, 85579 Neubiberg, Germany (e-mail: avik.santra@infineon.com).

Robert Weigel is with the Institute of Electronics Engineering, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany (e-mail: robert.weigel@fau.de).

Fabian Lurz is with the Institute of High-Frequency Technology, Hamburg University of Technology, 21073 Hamburg, Germany (e-mail: fabian.lurz@tuhh.de).

Digital Object Identifier 10.1109/JSEN.2021.3092002

extract features, such as the total bandwidth of the Doppler signal and the normalized standard deviation of the Doppler signal strength among others, from the Doppler-spectrogram of different human activities. Also, in more recent works handcrafted features are used. The authors in [5] propose the extraction of median Doppler frequency, total bandwidth of the Doppler signal and the standard deviation in signal strength to name a few, whereas in [6] the envelope of micro-Doppler signatures is used as basis for classification. The classification result using handcrafted features highly depends on their discriminative characteristics and requires expert knowledge in designing them. Extracting such features from Doppler spectrograms inevitably removes information from the signal. In order to remove as little information as possible while compressing the data to meaningful features, the usage of a principle component analysis (PCA) was proposed in [7] and [8]. Whereas the PCA is a linear transformation, in recent years neural networks, which provide a non-linear transformation, were widely used for feature extraction in the radar domain. In [22], authors propose a novel deep auto-encoder based solution to sense elderly fall-motion from Doppler spectrograms and in [17], authors have used various deep convolutional neural network architectures to learn from range spectrogram, Doppler spectrogram and radar data cubes for different activity classification.

However, a predominant number of papers extracts features on the basis of short-time Fourier transformed (STFT) radar data such as Doppler spectrograms or radar data cubes. This holds for the handcrafted feature approaches as well as the dimensionality reduction approaches using a PCA as well as for the deep learning based approaches. The fact that for radar-based human-motion recognition and classification nearly all publications so far are based on pre-processed data is underlined by the survey of Gurbuz and Amin [25].

A STFT resolves the frequency space with equal resolution. However, when classifying between different human activities, the frequency regions representative of the different activities are of higher interest. Therefore, in [23], authors propose a deep deformable convolutional to focus on certain time-frequency areas in the Doppler spectrograms, which helps to handle the small inter-class differences and large intra-class variations of human fall-motion in a real-world situation. Further, in recent years different approaches feeding the raw radar data provided by the ADC, which is referred to as raw ADC data in this paper, directly to the neural network came up. In [27], the authors train a long-short term memory (LSTM) network directly with time series of complex-valued raw ADC data, which on the contrary leads to a very high network complexity. An approach based on the in-quase/quadrature (I/Q) trajectories is presented in [28]. First the I/Q trajectories are transformed into low-resolution images, which are then classified using a DCNN. Thus, the network extract features from the image rather than from the I/Q data directly. Further, in [29], a DCNN including a Fourier layer and using raw ADC data as input is proposed. The Fourier layer is a convolutional layer with its kernels initialized to the Fourier coefficients. However, those weights are adapted during training and thus after training the transformation in the Fourier layer most

likely differs from a Fourier transformation. In audio signal processing, where the information is also encoded in the time-frequency domain, a neural network called SincNet operating directly on the raw data is proposed in [30]. The core idea of the SincNet is to restrict the first layer of the neural network to the usage of 1D bandpass sinc filters and only allow the optimization of their cutoff frequencies during training. By doing this restriction and therefore putting prior knowledge into the system, the network converges faster to its optimum solution. Further, the outcome of this layer remains physically interpretable since only the parameters of the filters, and not the single filter weights independently, are optimized.

Inspired from the 1D SincNet in speech processing, we propose two DCNN architectures based on 2D sinc filters and 2D wavelet filters that directly learn joint fast-time-slow-time-frequency features from the raw ADC radar data. The proposed model classifies among different human activities using radar data captured by *Infineon's* 60-GHz frequency modulated continuous wave (FMCW) radar chipset *BGT60TR13C*. We demonstrate that the proposed trained DCNN architectures are able to achieve classification accuracy equal or better than equivalent DCNNs using STFT-based pre-processing techniques in form of Doppler spectrograms or radar data cubes as input. This is the first paper to the best of author's knowledge that deals with data-driven radar pre-processing optimization, thus allowing the network to learn implicitly better representation for classification of the particular task, which is human activity classification in this paper.

The rest of the paper is organized as follows, we present the radar system design and system parameters in Section II, the conventional signal processing involving Doppler spectrogram and range-Doppler-time or data-cube processing is presented in Section III, the contribution is also presented in Section III. The proposed solution is presented in section IV, the proposed architecture & learning in section V. The results and discussion along with the associated setup is presented in section VI, and we conclude in Section VII.

II. RADAR SYSTEM DESIGN

The work in this paper is based on *Infineon's* *BGT60TR13C* FMCW radar chipset. Its operating frequency ranges from 57 GHz to 64 GHz with an adjustable chirp duration. Its block diagram is shown in Fig. 1. The transmit path consists of a voltage controlled oscillator (VCO) that is regulated by a phase locked loop (PLL) to a reference frequency of $f_{\text{ref}} = 80$ MHz. Highly linear frequency chirps between 57 GHz and 64 GHz are produced by adjusting the divider value and an additional tuning voltage ranging from 1 V to 4.5 V. In the receive path the echo returning from the target object is down-converted with a replica of the transmitted frequency chirp. Therewith the baseband frequency spectrum can be sampled by the 12bit analog-digital converter (ADC). Moreover the receive path contains an intermediate frequency (IF) buffer amplifier and an analog IF-filter that can be adjusted corresponding to the received frequency range. The radar chip is package in an embedded waver level ball grid array package including 4 integrated patch antennas realized by a metal redistribution layer. Three of them are receive antennas having

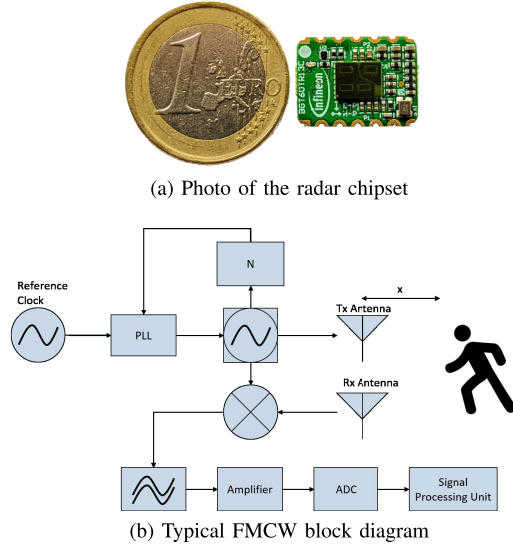


Fig. 1. (a) Infineon's BGT60TR13C 60-GHz radar sensor. (b) Functional block diagram of FMCW radar RF signal chain depicting 1TX, 1RX channel.

an antenna gain of 10dBi and one is the transmit antenna with a gain of 6dBi. Consequently, the radar sensor contains three identically structured receive paths and one transmit path. The radio frequency (RF) signal is distributed by an active RF distribution network to the receive paths.

The transmitted up-chirp from the FMCW radar's ramp generator is reflected by a moving object and is received at the receiver after round trip delay caused by the target's range from the radar and the velocity of the target. The received signal is mixed at the receiver with the transmitted signal and the resultant signal is low-pass filtered, thus performing the matched filtering operation. The phase of the resultant intermediate frequency or IF signal due to single point target can be expressed as

$$\Phi(t) = 2\pi \left(f_{\min} \tau + \frac{B}{T_c} t \tau - \frac{B}{2T_c} \tau^2 \right) \quad (1)$$

where f_{\min} is the ramp start frequency, B and T_c denote the chirp bandwidth and the chirp time respectively, and $\tau = \frac{2(x + \sum_{i=1}^1 v_i t)}{c}$ is the round trip propagation delay between the transmitted and received signal after reflection from the point target with range x and radial velocity components v_i . The Doppler frequency relates with radial velocity v_i as $v_i = 2v_i/\lambda = 2v_{\min}/c$, the Doppler is represented by the centroid Doppler of the Doppler components due to the target $v^c = \sum_{i=1}^1 v_i$, while micro-Doppler components are represented as $\{v_i - v^c\}_{i=1}^1$. The signal of an extended target is defined by the super-position of point target signals.

For our demonstrator, we configured the chip to transmit chirps using a pulse repetition interval (PRI) of $T_{\text{PRI}} = 1$ ms resulting in unambiguous maximum velocity of $v_{\max} = 1.25 \text{ m s}^{-1}$ which is sufficient in most indoor activity sensing applications. Figure 2 presents the chirp configuration. From the equidistant stream of chirps individual slices across slow-time can be extracted seamlessly. This provides maximum flexibility in further processing.

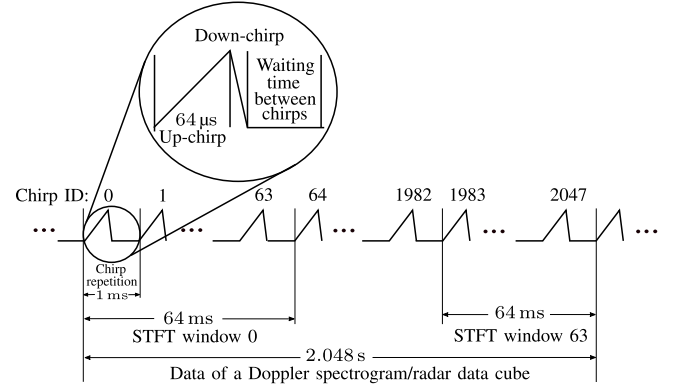


Fig. 2. Chirp sequence in the proposed system configuration.

TABLE I
OPERATING PARAMETERS

Parameters, Symbol	Value
Ramp start frequency f_{\min}	59.5 GHz
Ramp stop frequency f_{\max}	60.5 GHz
Bandwidth B	1 GHz
Range resolution δr	15 cm
Number of samples per chirp N_s	128
Maximum range R_{\max}	9.6 m
Sampling frequency f_s	2 MHz
Chirp time T_c	64 μs
Chirp repetition time T_{PRI}	1 ms
maximum Doppler v_{\max}	1.25 m s^{-1}
Elevation θ_{elev} per radar	70°
Azimuth θ_{azim} per radar	120°

The bandwidth is set to $B = 1$ GHz and the up-chirp time is set to $T_c = 64 \mu\text{s}$ accounting for a range resolution of 15 cm. The maximum detectable unambiguous range is 9.6 m. The set system parameters are provided in Tab. I.

III. CONVENTIONAL PIPELINE & CONTRIBUTIONS

The conventional signal pre-processing involves 1D moving target indication (MTI) filtering to remove the response from static targets and also Tx-Rx leakage, which effects the first few range bins. The reflections from stationary objects such as chair, tables and wall, etc. can overwhelm the reflections from the other moving targets limiting their visibility on the RDM or Doppler spectrogram. Thus, MTI filter is used to suppress the contribution of these stationary objects and leakage. Among several MTI filters, a simple 1D MTI filter subtracts the mean along the fast-time to remove the Tx-Rx leakage that perturbs the first range bins, followed by mean subtraction along the slow-time to remove the reflections due to static or zero-Doppler targets.

The range information of the target is extracted by performing the first FFT after applying 1D windowing along fast-time, which is the intra-chirp time. The Doppler information of the target is extracted by monitoring the change of target peak along slow-time, which is the inter-chirp time. One common approach is applying FFT along the fast-time as well as

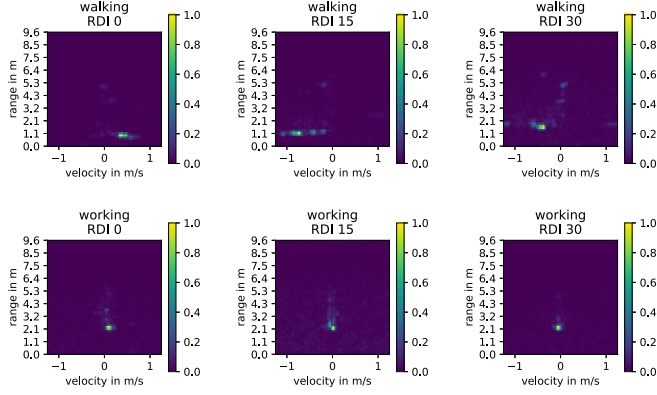


Fig. 3. Example radar data cube representation for walking and working activities at time steps 0, 15 and 30.

slow-time dimension. The outcome of this operation is a two-dimensional matrix representing the received power spectrum over range and velocity, also known as RDM. The received and deramped IF data is stored in matrices of size $N_c \times N_s$, where N_c being the number of chirps considered in a slice across slow-time and N_s is the number of transmit samples per chirp.

In a conventional processing pipeline, the above pre-processing is followed by feature image generation, such as a Doppler spectrogram or a range-Doppler-time radar data cube, which are then inputs to deep convolutional neural networks (DCNNs) or LSTM networks for classification.

A. Radar Data Cube

The radar data cube represents a time series of RDMs. Before the single RDMs can be generated the static targets must be removed from the signal as described in the previous section. This is done by first removing the mean within each chirp or fast-time and then removing the mean across multiple chirps or slow-time. A RDM computed at the k^{th} slice across slow-time can be obtained by applying a 2D STFT on the mean removed ADC data and is expressed as

$$v_{\text{RDM}}(p, l, k) = \left| \sum_{m=1}^{N_{\text{st}}} \sum_{n=1}^{N_{\text{ft}}} w(m, n) s(m, n, k) e^{-j2\pi \left(\frac{mp}{N_{\text{st}}} + \frac{nl}{N_{\text{ft}}} \right)} \right| \quad (2)$$

where $w(m, n)$ is the 2D weighting function along the fast-time and slow-time, $s(m, n, k)$ is the mean removed ADC data on the k^{th} slow-time data slice. The index n, m sweep along the fast-time and slow-time axis respectively, while l, p sweep along the range and Doppler axes respectively. N_{st} and N_{ft} are the FFT size along the slow-time and fast-time respectively. Figure 3 presents the radar data cube for walking and working activity at the slow-time data slice 0, 15 and 30. For the RDMs showing the walking activity a person is approaching the radar and then moving away from it. Most interesting for the activity classification is that the motion in range-Doppler domain and the spread of the target can clearly be seen. On the other hand for a person working on the laptop, the signal is mainly visible

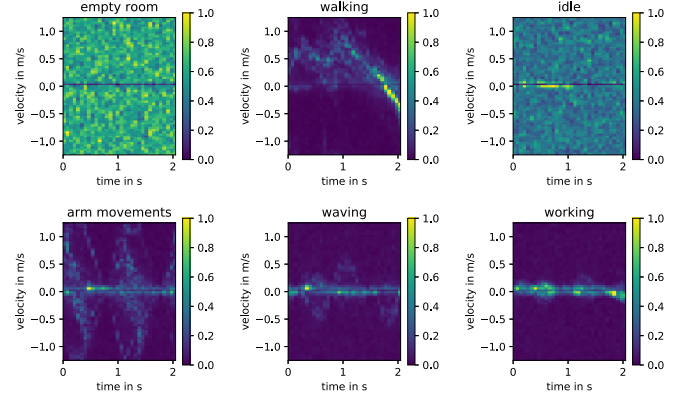


Fig. 4. Example Doppler spectrogram.

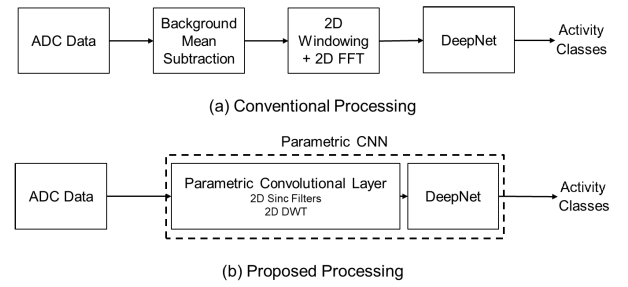


Fig. 5. (a) Conventional processing pipeline, involving explicit pre-processing, feature generation and neural network (b) Proposed processing pipeline, involving 2D Sinc filter DCNN or 2D Wavelet filter DCNN for implicit pre-processing, feature generation and classification.

in the zero Doppler bin and hardly a variation can be detected, which makes it hard to assign it to a certain activity class.

B. Doppler Spectrogram Features

The Doppler spectrogram is generated by marginalizing the radar data cube v_{RDM} across the range dimension and can be expressed as

$$S(p, k) = \sum_{l=1}^{N_{\text{ft}}} v_{\text{RDM}}(p, l, k). \quad (3)$$

where l is sweeps along the range axis, N_{ft} is the number of range bins and p, l, k are the Doppler, range and slow-time data slice indices respectively. The Doppler spectrum at slow-time data slice k contains both the Doppler components as well as micro-Doppler components due to hand and leg movements while performing an activity. The stacked Doppler spectrum across consecutive slow-time data slices is referred as Doppler spectrogram that captures information about the instantaneous Doppler spectral content and the variation of the Doppler spectral content over time. Figure 4 presents the Doppler spectrogram of the different activities, namely empty room, walking, standing idle, arm movement, waving and working on laptop.

Figure 5(a) presents the conventional pipeline that involves explicit pre-processing and feature generation followed by a neural network such as DCNN or LSTM for classification.

The novel aspect of the proposed architecture implicitly performs the pre-processing and feature generation in the neural network itself. Thus the input to the neural network is the raw ADC data directly as depicted in Fig. 5(b). The initial layer of the proposed DCNN learns 2D Sinc filter kernels or 2D wavelet filter kernels, which is representative of the pre-processing and feature extraction. The capability of DCNN to directly operate on the raw ADC data helps in reducing the computation complexity dramatically as well in practical implementation eliminates the need for digital signal processor (DSP) for pre-processing.

IV. PROPOSED PARAMETRIC CONVOLUTIONAL LAYER

Different activities can be distinguished by analyzing their unique range-velocity profiles. Some activities may have a very different profile such as walking and standing idle, but some may have only slight differences. For example working on the laptop and sitting idle on a chair only differ from slight hand movements to control the laptop. Thus, a higher resolution on specific frequency bands is required in order to accurately distinguish these actions. However, when applying a 2D STFT the whole observable range-velocity space is discretized in equal bins. By feeding the raw ADC radar data directly to a neural network, the neural network can learn filter kernels that extract more meaningful features than what can be achieved by fixed pre-processing steps. However, unlike as in computer vision or other domains, the feature is not spatially present in the raw ADC radar data. Therefore, a small filter kernel size of (3×3) or (5×5) , which is typically used in convolutional layers, is not able to extract meaningful features. This is why larger filter sizes such as (64×32) have to be considered. As a result the number of trainable filter weights increase drastically leading to overfitting and getting stuck in local minimum. However, since it is known that the information of the radar signal is in its time-frequency characteristic, the filter kernel can be restricted to a specific filter types, that, known from classical signal processing, are able to extract meaningful time-frequency features. Such a filter can e.g. be a Morlet wavelet. A Morlet wavelet is uniquely defined by two parameters, which are the frequency and the standard deviation of the Gaussian window. Thus, the neural network only has to learn the filter parameter instead of each single filter weights. We refer to a convolutional layer whose filter kernels are defined by a parametric filter function as parametric convolutional layer. In each training step the filter parameters are optimized and the filter kernels re-initialized accordingly. In this way, by using prior knowledge of the radar signal, pre-processing is integrated into the neural network and can be optimized according to the training data. The benefit of integrating pre-processing into the neural network itself by learning the filter parameters of a set of time domain filters was already shown in [30] for 1D audio signal processing. In the paper constraining the first convolutional layer to the use of 1D band-pass filters defined as the difference of two low-pass sinc filters with different cutoff frequencies is proposed. While training the lower and higher cutoff frequencies are optimized for the needs of the application. Replacing classical pre-processing by this layer shows improved results in speaker

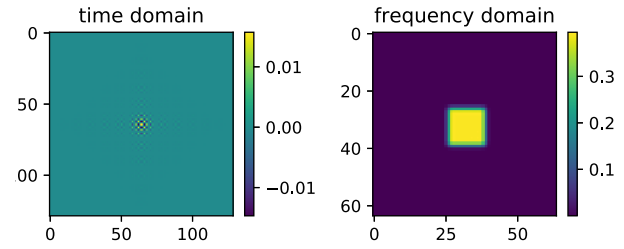


Fig. 6. Exemplary 2D sinc filter in time and frequency domain.

recognition. In this paper the extension from 1D audio signals to 2D radar signals is proposed. Additionally, the Morlet wavelet is evaluated as a second possible parametric filter function.

A. 2D Sinc Filters

Besides applying a STFT, time domain bandpass filters can be used to analyze the frequency composition of a signal. Time domain bandpass filters yield the ability to adjust the cutoff frequencies according to the needs of the application and can therefore be learned within a neural network. Thus, similar to the proposal of Ravanelli *et al.* a sinc filter is chosen as parametric filter function. However, it has to be extended to the 2D radar domain. The 1D sinc filter is defined as

$$h_{K,f_s}(k, f_l, b) = 2(f_l + b) \text{sinc}(2(f_l + b) \frac{k - \lfloor \frac{K}{2} \rfloor}{f_s}) - 2f_l \text{sinc}(2f_l \frac{k - \lfloor \frac{K}{2} \rfloor}{f_s}) \quad (4)$$

where K is the filter length, f_s the sampling frequency of the signal, f_l the lower cutoff frequency, b the bandwidth and k the filter parameter index. The parameters of this filter are the lower cutoff frequency f_l and the bandwidth b that implicitly defines the higher cutoff frequency. By defining a lower cutoff frequency and bandwidth in slow-time as well as in fast-time direction, a 2D bandpass filter that is able to extract joint range and velocity features can be created. The 2D sinc filter is defined as

$$\text{sinc}_{2D}(n, m; f_l^{\text{st}}, b_{\text{st}}, f_l^{\text{ft}}, b_{\text{ft}}) = w(n, m) h_{N, f_s^{\text{st}}}(n, f_{l,\text{st}}, b_{\text{st}}) h_{M, f_s^{\text{ft}}}(m, f_{l,\text{ft}}, b_{\text{ft}}) \quad (5)$$

where N and M are the filter-lengths, f_s^{st} and f_s^{ft} the sampling frequencies, f_l^{st} and f_l^{ft} the lower cutoff frequencies, b_{st} and b_{ft} the filter bandwidths respectively in slow-time and fast-time direction. Furthermore, $w(n, m)$ is a 2D cosine weighting function. n is sweeping along slow-time and m along fast-time. An exemplary 2D sinc filter is shown in Fig. 6 in time as well as in frequency domain. In frequency domain the rectangular shape with clear cutoff frequencies can be seen. The first layer of a CNN is initialized according to the definition of 2D sinc filters and only the filter parameters are allowed to be learned during training.

B. 2D Morlet Wavelets

The range-velocity profile is not only defined by its composed frequencies but also by the change of frequencies

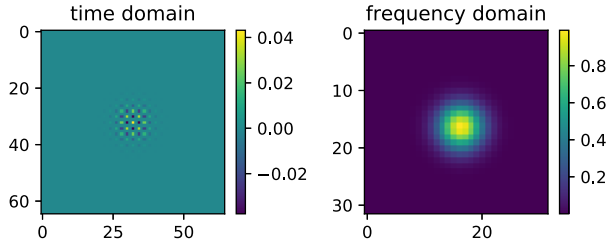


Fig. 7. Exemplary 2D Morlet wavelet in time and frequency domain.

over time. When transforming a signal to frequency domain the time information is lost. This can be overcome by windowing the time domain signal. However, smaller window sizes mean higher time resolution but to the cost of worse frequency resolution and vice versa. Especially for time varying signals wavelets have several advantages over fourier transformations as they provide a time and frequency resolution [31]. Due to the fact that radar signals are highly time varying, the usage of a 2D wavelet transformation using Morlet wavelets is proposed. The 2D Morlet wavelet is defined as

$$g_{N,M}(n, m; \sigma_{st}, \sigma_{ft}) = \frac{1}{2\pi \sigma_{st} \sigma_{ft}} \exp\left(-\left(\frac{(n/N - \lfloor N/2 \rfloor)^2}{2\sigma_{st}^2} + \frac{(m/M - \lfloor M/2 \rfloor)^2}{2\sigma_{ft}^2}\right)\right) \quad (6)$$

$$\phi[n, m; f_{st}, \sigma_{st}, f_{ft}, \sigma_{ft}] = g_{N,M}(n, m; \sigma_{st}, \sigma_{ft}) \cdot \cos\left(2\pi f_c^{st} \frac{n - \lfloor N/2 \rfloor}{f_s^{st}}\right) \cos\left(2\pi f_c^{ft} \frac{m - \lfloor M/2 \rfloor}{f_s^{ft}}\right) \quad (7)$$

where N and M are the filter-lengths, σ_{st} and σ_{ft} the standard deviations, f_c^{st} and f_c^{ft} the center frequencies, f_s^{st} and f_s^{ft} the sampling frequencies respectively in slow-time and fast-time direction. The filter parameters that can be optimized by the neural network are the center frequency and the standard deviation of the wavelet. Similar to the previous introduced 2D sinc filters the frequency area of interest can be adjusted by the center frequency. But, additionally also the time-frequency resolution can be optimized by changing the standard deviation of the gaussian part of the wavelet. Due to the fact that the defined wavelet is the product of a cosinus and a gaussian window function also the frequency response has the shape of a gaussian. That means that it has no clear cutoff frequencies as it can be seen in Fig. 7, where an exemplary 2D Morlet wavelet in time and frequency domain is depicted. The standard deviations of the gaussian in time domain and in frequency domain are indirect proportional. Consequently, decreasing the width of the gaussian in time domain will lead to an increased width of the frequency response, which in turn shows the time-frequency resolution trade-off.

V. ARCHITECTURE AND LEARNING

In this paper two state-of-the-art DCNNs evaluating the pre-processed data, a state-of-the-art DCNN evaluating raw input data and two novel DCNN architectures using raw ADC data as input are proposed. All architectures have several characteristics in common. First, all networks finishing with

a softmax classifier layer of size 6 as six different actions should be classified. Second, categorical crossentropy is used as loss function. Third, RMSprop optimizer is used with a learning rate $lr = 0.0001$, $\rho = 0.9$, $\epsilon = 10^{-8}$ and batches of size 128. Fourth, all unconstrained convolutional and dense layers are initialized using the 'Glorot' initialization scheme with an uniform weight distribution. Fifth, the common convolutional and dense layers are using a rectifier linear unit as activation. And sixth, after every common convolutional and dense layer a dropout with a rate of 0.2 is implemented in order to prevent over fitting.

A. 2D SincNet

The 2D SincNet uses 2D sinc filter convolutions in the first convolutional layer as described in chapter IV-A. As parameter this layer takes the filter lengths, the number of filters, the sampling frequencies, the padding mode and the stride for the slow- as well as fast-time direction respectively. Although there are no separated filters for slow- and fast-time, it is required to explicitly provide the number of filters in slow-time N_{st} and the number of filters in fast-time N_{ft} . According to this, 2D sinc filters are generated in a way that they form an equal grid of size $N_{st} \times N_{ft}$ covering the complete observable range Doppler domain. The trainable weights in this layer are the lower cutoff frequencies and the bandwidths in slow- as well as fast-time direction. In order to guarantee equal training in both filter dimensions the bandwidths and cutoff frequencies are normalized.

The 2D sinc filter layer is followed by a MaxPool layer with a pooling size of 8×2 . Afterwards a common two dimensional convolutional layers using 50 filters of size 3×3 is implemented. As already mentioned above, the convolutional layer follows a dropout layer with a rate of 0.2. Moreover, after the dropout a max pooling of size of 4×2 is applied to decrease dimensionality. Then the tensor is flattened and fed into a dense layer of size 32 followed by the softmax classifier layer. The proposed network is depicted in Fig. 8

B. 2D WaveConvNet

The 2D WaveConvNet (WCN) is designed similar to the 2D SincNet. Only the first convolutional layer is initialized by 2D Morlet wavelets as described in IV-B instead of using 2D sinc filters. Required parameter for this layer are the filter lengths, number of filters, sampling frequencies, padding mode and stride for the slow as well as fast-time direction respectively. Similar to the 2D sinc filters the number of filters in slow-time direction N_{st} as well as the number of filters in fast-time direction N_{ft} have to be explicitly provided in order to distribute the frequency response of the wavelets equally as a grid in the 2D frequency domain. Both time axis were normalized as already discussed in the previous section. As a result, the standard deviations is chosen to be 0.06 in both filter dimensions. In a total N_{st} times N_{ft} 2D wavelets are created. Trainable weights of this layer are the center frequencies and standard deviations in slow- as well as in fast-time dimension. Also in the 2D WCN the learnable weights are normalized.

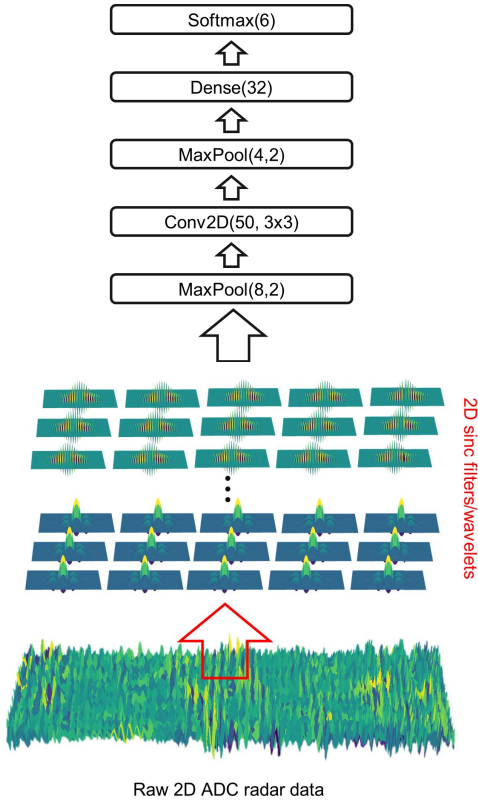


Fig. 8. Proposed Parametric CNN learning the filter parameters of 2D sinc filters or wavelets.

C. State-of-Art Networks

1) **DSNet**: The DSNet is a typical state-of-art 2D DCNN architecture followed by a dense and softmax layer. It uses already pre-processed 2D Doppler spectrograms as input. Furthermore, it contains three common 2D convolutions with 4, 8 and 16 filters respectively. Each filter uses a kernel size of 3×3 . After each convolution a dropout layer with dropout rate of 0.2 is used. Moreover, after the first two convolutional layers a MaxPooling of size 2×2 is used. Afterwards the tensor is flattened and fed into a dense layer of size 64 followed by the softmax classifier. The DSNet architecture is shown in Fig. 9 (a).

2) **RDCNet**: The RDCNet has a temporal sequence of RDMs in the form of a 3 dimensional radar data cube as input. Therefore three 3D convolutional layers are used to extract information from the radar data cube. They all have a kernel size of $3 \times 3 \times 3$ and use 4, 8 and 16 filter kernels respectively. Also here a dropout layer with a rate of 0.2 is added after each convolutional layer to prevent overfitting. After the first two dropout layers a maxpooling of size $2 \times 2 \times 4$ is performed. Afterwards the tensor is flattened and further processed by a dense layer of size 64 before it is classified by the final softmax layer. The RDCNet is sketched in Fig. 9 (b).

3) **2D ConvNet**: The 2D ConvNet uses the same architecture as the 2D SincNet and 2D WCN. Only the first layer is substituted by a unconstrained 2D convolutional layer with 'Glorot' weight initialization. No predefined time domain

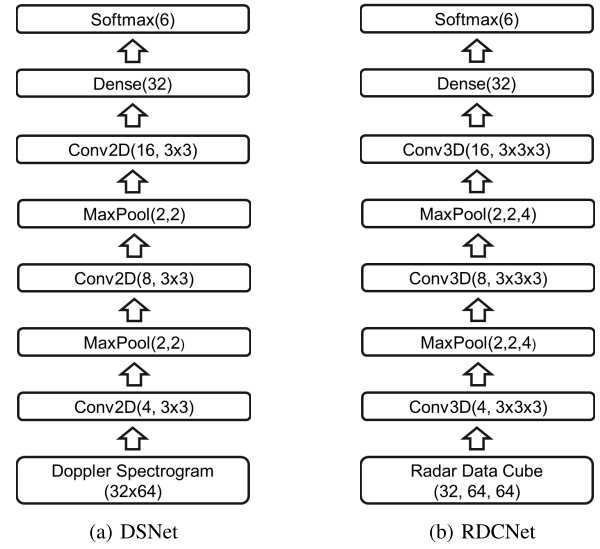


Fig. 9. State-of-art DCNN architectures for evaluating (a) Doppler spectrograms and (b) radar data cubes.



Fig. 10. Experimental setup for data recording with test person performing activity "working".

filters are used. Therefore each filter parameter can be learned individually.

VI. RESULTS AND DISCUSSION

A. Experimental Setup and Dataset

To evaluate the approach presented in this paper, a dataset was recorded in a real world environment. The radar was mounted on a tripod at a height of 1.20m and was placed in the corner of the room. The room has about $20m^2$ with a table and chairs inside. The experimental setup is shown in Fig. 10.

The dataset is chosen in a way that it covers fast moving activities such as walking, as well as slow moving activities like standing idle or working on the laptop. Thus, the challenge is to cover a large Doppler velocity range and simultaneously yield a high Doppler resolution in certain regions in order to differentiate similar activities. The dataset contains five different human activities plus additionally a recording of an empty room. To record the class "walking" a single human was allowed to randomly walk around. The class "idle" is split up in two recordings. First, a person was standing in front of the radar and in the second recording the person was sitting at the table facing towards the radar. As third activity

TABLE II
SAMPLES PER CLASS

empty room	walking	idle	arm movements	waving	working
579	695	834	694	687	686

TABLE III
MODEL SIZES OF DSNET AND RDCNET

Layer	DSNet	Layer	RDCNet
2D Conv (4, 3x3)	40	3D Conv (4, 3x3x3)	112
2D Conv (8, 3x3)	296	3D Conv (8, 3x3x3)	872
2D Conv (16, 3x3)	1168	3D Conv (16, 3x3x3)	3472
Dense (32)	65569	Dense (64)	102432
Softmax (6)	198	Softmax (6)	198
Total	67,270	Total	529,198

random arm movements while standing were recorded. This class is called “arm movements”. In order to record the fourth class called “waving” a person was waving with its hand at different positions in the room facing towards the radar. As last class working at the laptop while sitting at the table was recorded. The data was acquired in a clean environment, since the objective is mainly to differentiate activities, that cover a wide Doppler frequency bandwidth and simultaneously require high Doppler resolution in certain frequency bands, rather than to handle disturbances. Each activity was performed by the same person and recorded for about 18 minutes in total.

Samples containing 2048 chirps with an overlap of 512 chirps are cut out from the recordings. Given the fact that the chirp repetition time is 1ms each sample captures 2.048s. For each sample a Doppler spectrogram and a radar data cube as described in sec. III-B and III-A is created. Therefore a dataset with raw ADC data, Doppler spectrograms and radar data cubes based on exactly the same chirps per sample is obtained. For each activity about 700 samples are available. Due to slightly different recording times for each activity, the number of samples per class varies. In table II the exact number of samples per class is stated.

To evaluate the proposed approach, a filter length of 65 in slow- and 33 in fast-time dimension was chosen for the 2D sinc filters. The same filter size is used for the constrained convolutional layer that substitutes the 2D sinc filter layer in the 2D ConvNet. For the 2D wavelet convolution the filter length in slow-time direction was doubled in order to allow the gaussian window function of the Morlet wavelet to expand. Moreover “valid” padding is used in both dimensions to obtain the same output shape after 2D sinc and 2D wavelet layer. This provides the possibility of substituting both layers one by one while keeping the remaining network the same. To reduce computational intensity a stride of 4 and 8 is used in slow- and in fast-time dimension respectively.

As the filter sizes are specified, the final size of the networks can be stated. The composition of parameters per layer is shown in tab. III for the DSNet and the RDCNet and in tab. IV for the 2D SincNet, 2D WCN and 2D ConvNet.

TABLE IV
MODEL SIZES OF 2D SINCNET, 2D WCN AND 2D CONVNET

Layer	2D SincNet	2D WCN	2D ConvNet
2D Sinc (64, 65x33)	256	-	-
2D Wave (64, 129x33)	-	256	-
2D Conv (64, 65x33)	-	-	137344
2D Conv (50, 3x3)	28850	28850	28850
Dense (32)	102432	102432	102432
Softmax (6)	198	198	198
Total	131,736	131,736	268,824

Given the fact that when evaluating Doppler spectrogram only velocity information has to be evaluated the network can be designed accordingly smaller. The RDCNet uses a radar data cube as input data. Thus, 3D convolutions have to be used which result in a higher number of parameters. Moreover, the first layer of the 2D SincNet as well as of the 2D WCN have significantly less parameters as the corresponding unconstrained convolutional layer of the 2D ConvNet. This results from the fact that only the four filter parameter are trained in the parametric convolutional layer. Since 64 2D sinc filters or wavelets are used, the parametric layer has just $64 \cdot 4 = 256$ parameters to optimize. The unconstrained convolutional layer in the 2D ConvNet in contrast has to learn each single filter weight. As a result the network size is reduced by more than 50%. The proposed approach is a data-driven pre-processing optimization. Hence, besides the evaluation on the clean dataset, the proposed approach is also evaluated regarding the impact of limited amount of training data and the impact of pre-known disturbances, such as a static 50Hz frequency of a power line, which can be considered during training.

B. Clean Dataset

In order to give a proof-of-concept, the proposed idea was evaluated on the dataset as described in the previous section. The dataset does not contain disturbances and yields enough training data.

1) *Classification Results*: For evaluation a 5-fold cross validation is performed. First, the dataset is split into 5 blocks, whereof each is used once as testing set. Thus, the model is trained 5 times and leaving out a different test set each time. Finally, the results of the five runs are averaged. In this way variations in the results due to unfortunate train and test data splits are reduced. All models are trained long enough to reach their saturation. Thus, the DSNet is trained for 100 epochs, the RDCNet is trained for 50 epochs, the 2D ConvNet is trained for 40 epochs and the proposed 2D SincNet and 2D WCN are trained for 20 epochs. Afterwards the accuracy as well as the F1 score are evaluated using the testing part of the dataset. The obtained accuracies and F1 scores are averaged over all runs. Additionally, the standard deviation is calculated for both matrices.

The results are shown in tab. V. The state-of-art approaches achieve an accuracy of about 90.7% and 95.6% and 98.2% respectively, whereas the proposed architectures show an

TABLE V
ACCURACIES (IN %) AND F1-SCORES (IN %) FOR THE
EVALUATED APPROACHES ON THE CLEAN DATASET

Model	Accuracy (\pm deviation)	F1-score
DSNet	90.7 (\pm 3.2)	90.8 (\pm 3.0)
RDCNet	95.6 (\pm 0.7)	95.9 (\pm 0.6)
2D ConvNet	98.2 (\pm 2.2)	98.2 (\pm 2.3)
2D SincNet	99.2 (\pm 0.6)	99.2 (\pm 0.6)
2D WCN	99.5 (\pm 0.3)	99.5 (\pm 0.3)

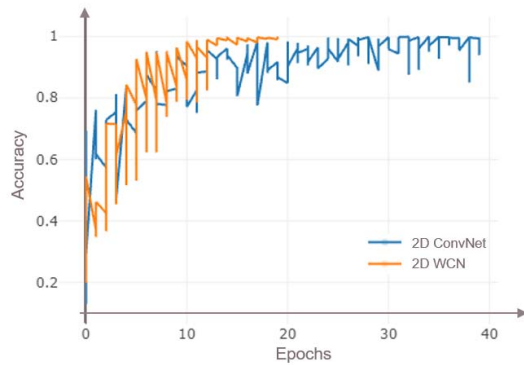


Fig. 11. validation accuracy over training epochs of five training runs of the 2D ConvNet and 2D WCN. Vertical lines indicate the lowest and highest achieved accuracy in a certain epoch.

improved accuracy of 99.2% and 99.5%. In order to analyze the classification results in more detail the confusion matrices of RDCNet representing the state-of-art approaches and the confusion matrix of 2D WCN representing the novel architectures are shown in Fig. 12. The limitation of the state-of-art approaches is unveiled by looking at the individual accuracy per class. While most actions are similar well classified by RDCNet and 2D WCN, a big uncertainty between the class “idle” and “working” exists. However, the proposed filter learning based approaches do not show this limitation and therefore achieves better accuracy scores.

In Fig. 11 the validation accuracy over training epochs for the 2D ConvNet in comparison with the 2D WCN is plotted. Both models were trained five times. The vertical lines visualize the lowest as well as highest accuracy achieved by the different training runs at a certain epoch and thus represent the dispersion of the accuracies in a certain epoch. As seen in the plot for the 2D ConvNet the mean accuracy as well as the dispersion does not improve anymore from about epoch 30 onward, which indicates that the training has already converged.

The high variance for ConvNet approach has two main reasons. First, the optimization problem is very high dimensional and thus has many local minima towards which the solution can converge to. And second, since the learned filter can have very specific frequency selectivity. Small changes of these filter kernels during training results in a different frequency selectivity and thus has large impact on the final classification result. However, our proposed filter kernels are chosen to have a controlled and continuous frequency selectivity and thus are



(a) RDCNet



(b) 2D WCN

Fig. 12. Confusion matrix of (a) RDCNet and (b) 2D WCN. The values are rounded to three decimals.

more stable. In contrary, the 2D SincNet as well as 2D WCN converge always to low-dimensional search space leading to very similar scores enabled by the guidance of parametric functions due to sinc filters and wavelets.

2) Learned Filters: Before training is started the sinc filters as well as the wavelets are initialized as described in sec. V-A and sec. V-B. Moreover, the first unconstrained convolutional layer of the 2D ConvNet is initialized using ‘Glorot’ scheme. For each approach, cumulating the initial filters leads to an approximately uniform range and velocity gain over the whole space. During training the filter parameters are iteratively optimized. As a result the initial grid structure is dissolved by individual shifts and shape changes of the filters.

The cumulative gain of all filters after training is depicted in Fig. 13. The 2D sinc filters as well as the 2D wavelets have a bandpass characteristic. Therefore the resulting gain of cumulative filters look similar except of the fact that the 2D wavelet gain is smoother caused by its smooth filter shape in frequency domain. However, the resulting weights of unconstrained convolutional layer are quite different and cannot be physically interpreted.

C. Limited Dataset

An essential factor in deep learning applications is the dataset. The neural network can only learn the information given in a dataset. Thus, data acquisition is an important task. However, it is also time consuming and challenging to acquire

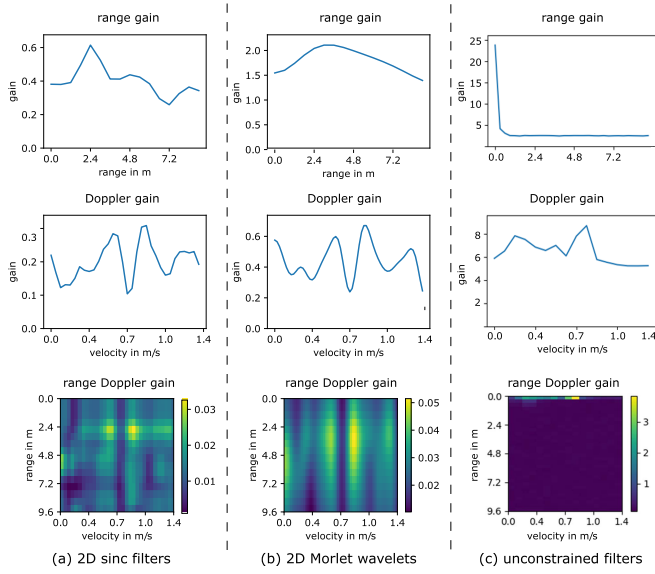


Fig. 13. Cumulative gain of learned (a) 2D sinc filters, (b) 2D wavelets and (c) unconstrained filters.

TABLE VI
F1-SCORES (IN %) OF THE EVALUATED APPROACHES UNDER LIMITED NUMBER OF TRAINING SAMPLES S

Model \ S	10	25	50	100	250
DSNet	74.5 (± 2.0)	80.9 (± 1.1)	83.1 (± 3.1)	86.6 (± 1.5)	86.5 (± 3.0)
RDCNet	86.4 (± 1.5)	93.1 (± 2.2)	94.1 (± 1.5)	94.5 (± 1.2)	95.2 (± 0.8)
2D ConvNet	91.8 (± 1.4)	96.2 (± 1.1)	97.2 (± 0.6)	95.7 (± 3.3)	98.9 (± 1.8)
2D SincNet	99.1 (± 0.7)	99.4 (± 0.3)	99.2 (± 0.5)	99.5 (± 0.3)	99.6 (± 0.2)
2D WCN	99.2 (± 0.8)	99.4 (± 0.3)	99.7 (± 0.2)	99.4 (± 0.2)	99.6 (± 0.2)

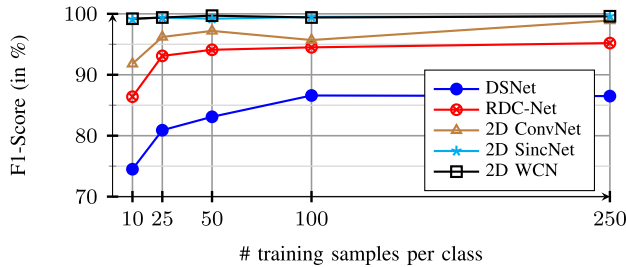


Fig. 14. F1-Scores over number of training samples per class.

representative data for the application. If the dataset is too small, neural networks tend to overfit. Hence, it is evaluated how the proposed networks perform under low number of training samples. Thus, the different networks are trained until convergence using 10, 25, 50, 100 and 250 samples per class. The training samples are randomly selected. In order to get more stable results, each network was trained five times for each number of training samples using different randomly selected training samples every time. In tab. VI the mean F1-scores are stated for the different models based on the number of training samples, which are additionally visualized in Fig. 14. Especially for low number of training samples the 2D SincNet and 2D WCN show very good results. Since their first layers are constrained to the usage of a set of parametric filters known from signal processing, less parameter have to be trained and therefore good results are already possible for

TABLE VII
ACCURACIES (IN %) AND F1-SCORES (IN %) UNDER THE PRESENCE OF A STATIC 50 HZ INTERFERENCE

Model	Accuracy	F1-score
DSNet	88.3 (± 2.7)	88.3 (± 3.7)
RDCNet	95.5 (± 0.6)	95.7 (± 0.6)
2D ConvNet	93.2 (± 4.2)	93.5 (± 3.8)
2D SincNet	98.8 (± 0.5)	99.5 (± 0.5)
2D WCN	98.8 (± 1.3)	98.9 (± 1.3)

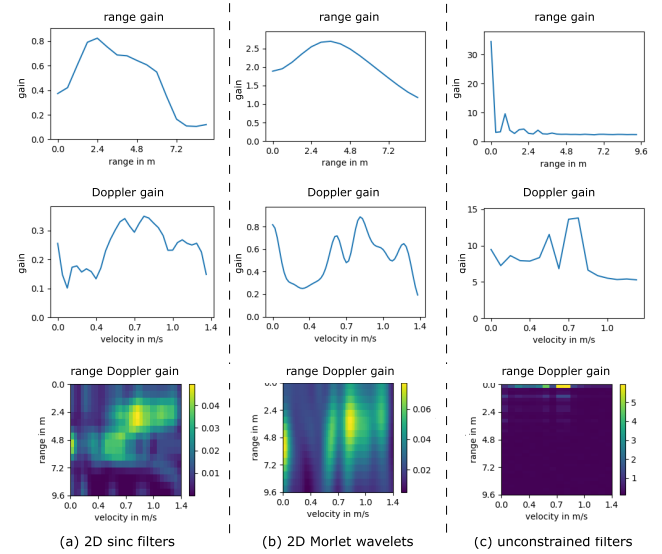


Fig. 15. Cumulative gain of learned (a) 2D sinc filters, (b) 2D wavelets and (c) unconstrained filters under the presence of a static 50Hz interference.

little amount of training data. Further, it is noticeable that the variance in the results of the 2D ConvNet is large, which shows that it may come up with very good solutions, but sometimes also gets stuck in a local minimum. In contrary, the 2D SincNet as well as 2D WCN converge always to very similar scores due to the guidance of sinc filters and wavelets.

D. Fixed Disturbance

There exist disturbances depending on the application that are known before and can therefore be included into the training process. An example is a static 50 Hz interferer due to the power grid when installing the radar into a ceiling light. Thus, a 50 Hz sinusoidal signal with -12 dB with respect to the maximum detectable signal power of the sensor was added to the dataset. The networks are then re-trained using the modified dataset.

In Tab. VII the final accuracies and F1-scores are shown. The models using pre-processing are only slightly effected by the disturbance. Since the 50 Hz interference is almost static within a chirp, its influence is almost completely removed when subtracting the mean of the signal before calculating the STFT. The architectures operating directly on the raw ADC data have to learn to suppress the disturbing frequency. The results show, that the constrained networks, namely 2D WCN and 2D SincNet, perform better in suppressing the disturbance. In Fig. 15 the adaption of the sinc filters and

wavelets respectively to suppress the 50 Hz interference, which corresponds to a velocity of 0.25 m s^{-1} , are depicted. The learned filters of the unconstrained 2D ConvNet still do not provide any physical interpretability.

E. Discussion

The limitation of the state-of-art approaches has its origin in pre-processing. The STFT for generating the RDMs equally discretizes the range as well as velocity domain. However, both activities “idle” and “working” contain very slight movements. As a result their features share similar range-Doppler bins and thus the STFT processed data of those actions is very similar. This complicates a classification. However, due to the ability of learning filters, the 2D WCN as well as the 2D SincNet is able to mitigate this limitation by adapting the parameter of its filters accordingly in order to clearly separate features of similar actions. Furthermore, the lack of range information gets noticeable when predicting classes based on Doppler spectrograms. The missing range information is expected to have an even higher impact when analyzing activities of multiple humans simultaneously.

The 2D ConvNet is not limited by using pre-processed data. Therefore, in theory 2D ConvNet can potentially achieve the same accuracy, though not guaranteed, as SincNet and WCN by increasing the training effort. The underlying issue for this is the fact that the learned sinc filters as well as wavelets are within the search space of the unconstrained convolutional layer. However, as the number of parameter is significant higher than the number of parameters of the proposed networks, the learning speed is decreased and more training samples are required to learn the extraction of meaningful features from the raw data. This was shown in the experiments with limited amount of training data. The 2D ConvNet achieves significant lower F1-scores as the 2D SincNet and 2D WCN and does not always converge to the same result, but shows high variance in training.

In signal processing static interferences can easily canceled out. That is why the approaches using pre-processing where only slightly influenced by the 50 Hz interferer. However, when the raw ADC data is directly fed into the neural network, the mitigation has to be learned during training. Due to the guidance of pre-defined filters, the proposed 2D SincNet and 2D WCN show a faster convergence and better mitigation of the interferer than the unconstrained counterpart network 2D ConvNet.

Using a predefined set of filters such as the proposed sinc filters or wavelets makes the outcome of the convolutional layer physically interpretable, which helps understanding the system. Moreover, due to filter learning application dependent meaningful range Doppler areas are focused.

VII. CONCLUSION

Human activity classification has several applications in surveillance, human-computer interfaces and smart home applications. We present an activity classifier based on parametric DCNN using 2D sinc filter kernels or 2D wavelet filter kernels that can seamlessly detect and classify human

activities directly using raw ADC radar data. We demonstrated the performance of the proposed DCNN in comparison to conventional DCNNs that use Doppler spectrograms or radar data cubes as input data and demonstrated the proposed solution offers better classification accuracy. While the pre-processing steps in the latter processing are fixed, the parametric DCNN is able to learn filter parameters specific to the activity classification task. Further, the filter parameters can adapt to suppress fixed disturbances in the signal. Additionally, we demonstrate that, due to guidance by the parametric filters, the proposed networks perform better for low number of training samples than the unconstrained DCNN. Our experiments show, that the conventional un-parametric DCNN is unable to mimic the pre-processing operations due to the large search space and thus is not suitable for a robust solution. As future work, we aim to extend the proposed DCNN to simultaneously classify multiple activities from multiple targets in the field of view.

REFERENCES

- [1] “Occupancy sensors: Positive on/off lighting control,” Electr. Power Res. Inst., Palo Alto, CA, USA, Tech. Rep. EPRIBR-100323, 1994.
- [2] V. Garg and N. K. Bansal, “Smart occupancy sensors to reduce energy consumption,” *Energy Buildings*, vol. 32, no. 1, pp. 81–87, Jun. 2000.
- [3] A. Santra, R. V. Ulaganathan, and T. Finke, “Short-range millimetric-wave radar system for occupancy sensing application,” *IEEE Sensors Lett.*, vol. 2, no. 3, pp. 1–4, Sep. 2018.
- [4] Y. Kim and H. Ling, “Human activity classification based on micro-Doppler signatures using a support vector machine,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
- [5] Y. Sun, T. Fei, F. Schliep, and N. Pohl, “Gesture classification with handcrafted micro-Doppler features using a FMCW radar,” in *IEEE MTT-S Int. Microw. Symp. Dig.*, Apr. 2018, pp. 1–4, doi: [10.1109/ICMIM.2018.8443507](https://doi.org/10.1109/ICMIM.2018.8443507).
- [6] M. G. Amin, Z. Zeng, and T. Shan, “Hand gesture recognition based on radar micro-Doppler signature envelopes,” in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6, doi: [10.1109/RADAR.2019.8835661](https://doi.org/10.1109/RADAR.2019.8835661).
- [7] B. Erol and M. G. Amin, “Radar data cube processing for human activity recognition using multisubspace learning,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019, doi: [10.1109/TAES.2019.2910980](https://doi.org/10.1109/TAES.2019.2910980).
- [8] J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. J. Soraghan, and S. Marshall, “Robust PCA micro-Doppler classification using SVM on embedded systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 2304–2310, Jul. 2014, doi: [10.1109/TAES.2014.130082](https://doi.org/10.1109/TAES.2014.130082).
- [9] D. P. Fairchild and R. M. Narayanan, “Human activity classification using Hilbert–Huang transform analysis of radar Doppler data,” *Proc. SPIE*, vol. 8021, Jun. 2011, Art. no. 80210F.
- [10] D. P. Fairchild and R. M. Narayanan, “Classification and modeling of human activities using empirical mode decomposition with S-band and millimeter-wave micro-Doppler radars,” *Proc. SPIE*, vol. 8361, May 2012, Art. no. 83610X.
- [11] Y. Kim and T. Moon, “Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.
- [12] R. J. Javier and Y. Kim, “Application of linear predictive coding for human activity classification based on micro-Doppler signatures,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, Oct. 2014.
- [13] R. M. Narayanan and M. Zenaldin, “Radar micro-Doppler signatures of various human activities,” *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1205–1215, 2015.
- [14] M. Zenaldin and R. M. Narayanan, “Radar micro-Doppler based human activity classification for indoor and outdoor environments,” *Proc. SPIE*, vol. 9829, May 2016, Art. no. 98291B.
- [15] Y. Lin, J. Le Kernec, S. Yang, F. Fioranelli, O. Romain, and Z. Zhao, “Human activity classification with radar: Optimization and noise robustness with iterative convolutional neural networks followed with random forests,” *IEEE Sensors J.*, vol. 18, no. 23, pp. 9669–9681, Dec. 2018.

- [16] X. Li, Y. He, Y. Yang, Y. Hong, and X. Jing, "LSTM based human activity classification on radar range profile," in *Proc. IEEE Int. Conf. Comput. Electromagn. (ICCEM)*, Mar. 2019, pp. 1–2.
- [17] R. Hernangomez, A. Santra, and S. Stanczak, "Human activity classification with frequency modulated continuous wave radar using deep convolutional neural networks," in *Proc. Int. Radar Conf. (RADAR)*, Toulon, France, Sep. 2019, pp. 23–27.
- [18] A. Santra and S. Hazra, *Deep Learning Applications of Short-Range Radars*. Norwood, MA, USA: Artech House, 2020.
- [19] S. Tomii and T. Ohtsuki, "Falling detection using multiple Doppler sensors," in *Proc. IEEE 14th Int. Conf. e-Health Netw., Appl. Services*, Beijing, China, Oct. 2012, pp. 196–201.
- [20] B. Y. Su, K. C. Ho, M. J. Rantz, and M. Skubic, "Doppler radar fall activity detection using the wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 865–875, Mar. 2015.
- [21] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, "Radar-based fall detection based on Doppler time-frequency signatures for assisted living," *IET Radar, Sonar Navigat.*, vol. 9, no. 2, pp. 164–172, Feb. 2015.
- [22] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2016, pp. 1–6.
- [23] Y. Shankar, S. Hazra, and A. Santra, "Radar-based non-intrusive fall motion recognition using deformable convolutional neural network," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 1717–1724.
- [24] V. C. Chen, *The Micro-Doppler Effect in Radar*. Norwood, MA, USA: Artech House, 2011.
- [25] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019, doi: [10.1109/MSP.2018.2890128](https://doi.org/10.1109/MSP.2018.2890128).
- [26] P. Vaishnav and A. Santra, "Continuous human activity classification with unscented Kalman filter tracking using FMCW radar," *IEEE Sensors Lett.*, vol. 4, no. 5, pp. 1–4, May 2020.
- [27] S. Yang, J. L. Kerne, F. Fioranelli, and O. Romain, "Human activities classification in a complex space using raw radar data," in *Proc. Int. Radar Conf. (RADAR)*, Sep. 2019, pp. 1–4, doi: [10.1109/RADAR41533.2019.171367](https://doi.org/10.1109/RADAR41533.2019.171367).
- [28] T. Sakamoto, X. Gao, E. Yavari, A. Rahman, O. Boric-Lubecke, and V. M. Lubecke, "Hand gesture recognition using a radar echo I-Q plot and a convolutional neural network," *IEEE Sensors Lett.*, vol. 2, no. 3, pp. 1–4, Sep. 2018, doi: [10.1109/LSSENS.2018.2866371](https://doi.org/10.1109/LSSENS.2018.2866371).
- [29] W. Ye and H. Chen, "Human activity classification based on micro-Doppler signatures by multiscale and multitask Fourier convolutional neural network," *IEEE Sensors J.*, vol. 20, no. 10, pp. 5473–5479, May 2020, doi: [10.1109/JSEN.2020.2971626](https://doi.org/10.1109/JSEN.2020.2971626).
- [30] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with SincNet," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2018, pp. 1021–1028.
- [31] V. Pukhova, E. Gorelova, G. Ferrini, and S. Burnasheva, "Time-frequency representation of signals by wavelet transform," in *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (EIConRus)*, Saint Petersburg, Russia, Feb. 2017, pp. 715–718.



Thomas Stadlmayer was born in Regensburg, Germany, in 1994. He received the B.S. and M.S. degrees in computational engineering with focus on digital signal processing from Friedrich-Alexander University Erlangen-Nuremberg in 2016 and 2019, respectively. He is pursuing the Ph.D. degree with the Circuits, Systems and Hardware Test Group, Institute of Electronics Engineering. He is also working in cooperation with Infineon Technologies AG. His research interests include digital signal processing and deep learning for mmWave radar systems.



Avik Santra (Senior Member, IEEE) received the M.S. (Hons.) degree in signal processing from the Indian Institute of Science, Bengaluru, in 2010. He is currently leading research and development of signal processing and deep learning algorithm/solutions of radar and depth sensors for human-machine interface applications with Infineon, Neubiberg. Earlier in his career, he has worked as a System Engineer for LTE/4G modem with Broadcom Communications, and also has worked as a Research Engineer developing cognitive radars with Airbus. He is the author of the book titled *Deep Learning Applications of Short-Range Radars* published by ArTech House. He has filed over 50 patents and published more than 35 research articles related to various topics of radar waveform design, radar signal processing and radar machine/deep learning topics. He was a recipient of several outstanding reviewer awards. He is a reviewer at various IEEE and Elsevier journals.



Robert Weigel (Fellow, IEEE) was born in Ebermannstadt, Germany, in 1956. He received the Dr.-Ing. and Dr.-Ing.Habil. degrees in electrical engineering and computer science from the Technical University of Munich, Munich, Germany, in 1989 and 1992, respectively. From 1994 to 1995, he was a Guest Professor of SAW Technology with the Vienna University of Technology, Vienna, Austria. He was a Research Engineer, a Senior Research Engineer, and a Professor of RF Circuits and Systems with the Technical University of Munich until 1996. From 1996 to 2002, he was the Director of the Institute for Communications and Information Engineering, University of Linz, Linz, Austria, where he co-founded the company DICE in 1999, which split into an Infineon Technologies (DICE) and an Intel (DMCE) company, which are devoted to the design of RFICs and MMICs. In 2000, he was appointed as a Professor of RF Engineering with Tongji University, Shanghai, China. Since 2002, he has been the Head of the Institute for Electronics Engineering, University of Erlangen-Nuremberg, Erlangen, Germany, where he co-founded the companies eesy-id and eesy-ic in 2009 and 2012, respectively. He has authored or coauthored more than 900 articles. He has been engaged in research and development of microwave theory and techniques, electronic circuits and systems, and communication and sensing systems. He served in many roles for the IEEE MTT-S and UFFC-S. He has been the Founding Chair of the Austrian COM/MTT Joint Chapter, a Region 8 MTT-S Coordinator, a Distinguished Microwave Lecturer, an IEEE MTT-S AdCom Member, and the 2014 MTT-S President.



Fabian Lurz (Member, IEEE) received the B.Sc. and M.Sc. degrees in information and communication technology and the Dr.-Ing. degree from Friedrich-Alexander University Erlangen-Nuremberg (FAU), Erlangen, Germany, in 2010, 2013, and 2019, respectively. He joined the Institute for Electronics Engineering, FAU, as a Research Assistant, in 2013, and headed the Research Group Circuits, Systems and Hardware Test from 2017 to 2020. Since June 2020, he has been working as a Senior Engineer and a Research Group Leader with the Institute of High-Frequency Technology, Hamburg University of Technology, Hamburg, Germany. His current research interests include microwave circuits and system, especially for low-cost and low-power metrology applications. He was a recipient of the First Prize of the High Sensitivity Radar Student Design Competition of the IEEE International Microwave Symposium in 2014, 2017, and 2018, respectively, and the IEEE Microwave Theory and Techniques Society Graduate Fellowship Award in 2016.