Improved Target Detection and Feature Extraction using a Complex-Valued Adaptive Sinc Filter on Radar Time Domain Data

Thomas Stadelmayer[1,2], Avik Santra[2], Markus Stadelmayer[3], Robert Weigel[1], Fabian Lurz[4]
[1]Friedrich Alexander University Erlangen-Nuernberg, Germany
[2]Infineon Technologies AG Munich, Germany
[3]Johannes Kepler University Linz, Austria
[4]Hamburg University of Technologie, Germany

2021 29th European Signal Processing Conference (EUSIPCO) | 978-9-0827-9706-0/21/$31.00 ©2021IEEE | DOI: 10.23919/EUSIPCO54536.2021.9616250

***Abstract*—In state-of-art radar classification tasks the raw time domain ADC data is transformed to frequency domain wherein the target is detected. Feature data, such as the micro-Doppler signature, is extracted for classification of e.g. a performed gesture or a person based on its gait. However, the transformation into frequency domain using short-time Fourier transform resulting in erroneous target detection due to super-position of target components thus leading to sub-optimal feature for subsequent classification task. Thus, in this paper we propose a target feature extraction approach that operates directly on 2D time domain radar data by using a complex-valued adaptive 2D sinc filter. The proposed approach tracks the target's slow-time and fast-time frequencies by a regulation loop, which progressively adjusts the filter's center frequency to the target location. We demonstrate that the proposed approach extracts better features in both single- and multi-target scenarios, leading to improved classification accuracy in gait classification problems using radars. Furthermore, the proposed time domain feature extraction approach facilitates the use of a parametric neural network that works directly on time domain data.**

***Index Terms*—adaptive sinc filter, feature extraction, radar classification, time domain processing**

I. Introduction

In the last few years, technical progress in radar technology has led to miniaturized and highly accurate radar sensors. This development leads to a number of new possibilities especially for short-range consumer and industrial applications [1]. Automatic light and air conditioning based on the presence and position of people help to save energy. Moreover, activity detection and elderly people supervision are emerging smart home technologies. The usage of camera based solutions lead to major invasions in the personal privacy and thus are unsuitable for such applications. Recent advancements in silicon—antenna in-package together with deep learning—has made radars a promising technology for such applications.

The vast majority of papers dealing with various radar based classification tasks, e.g. human activity recognition, gesture sensing or elderly fall detection, use pre-processed data such as range, Doppler or angle spectrograms [2], [3], range-Doppler map (RDM) [4] or points clouds [5], [6] typically along with a deep learning classifier. When using spectrograms or RDMs, the radar signal can be interpreted as images and advanced deep learning methods from computer vision can be applied. Nonetheless, treating the radar signals as images, instead of as time domain signals as they are by nature, is not efficient. Moreover, the pre-processing itself can be computationally intensive and is a static and manually designed feature extraction chain that can not be optimized during training. Therefore, there has been an increasing interest in training neural networks with time domain data in recent years [7]–[9].

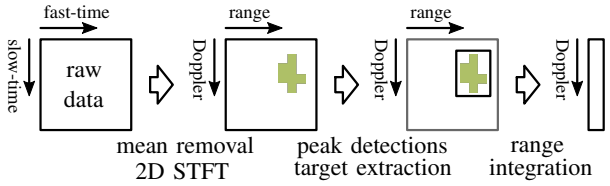
Furthermore, feature extraction on time domain radar data leads to superior features for classification compared to feature extraction on frequency domain data where erroneous features caused by a super-position of target components can occur. Moreover, the large body of radar classification works dealing with gesture or activity recognition do not address the issue of interference, although it is essential for real world systems. In [10] the impact of interference was discussed, but no solution provided. Other papers aim to mitigate interference in the deep learning model itself. In [11] the model was trained using interfered data, which in turn makes it not robust against unseen disturbances. In [12] a self-attention mechanism using unguided spatio-temporal learning was used to focus only on the target of interest for classification. However, extracting the information of the right target and removing the surrounding before feeding the signal into a deep neural network would further reduce the chance of miss-classification.

To address the aforementioned challenges, complex-valued 2D adaptive sinc filtering on radar time domain data by regulation loop for target detection and feature extraction is proposed. The proposed approach while comparing it to the state-of-art approach for target detection and feature extraction is outlined in Fig. 1. We demonstrate the superior target detection and feature extraction performance of the proposed solution by reporting the classification accuracy using a radar gait classification task.

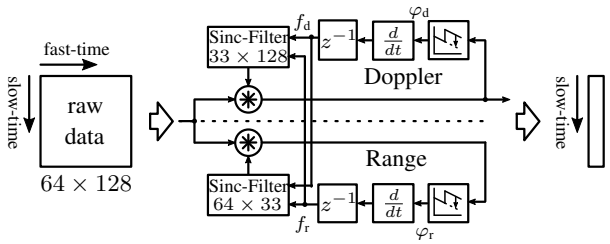
II. Radar System and Signal

A. Radar System

The proposed approach is evaluated using *Infineon*'s *BGT60TR13C* chipset. It is a miniaturized and power opti-



(a) Conventional target extraction algorithm.



(b) Proposed target tracking algorithm using adaptive Sinc-filters.

Fig. 1: Block diagram of (a) the conventional and (b) the proposed target feature extraction algorithm.

mized frequency modulated continuous wave (FMCW) radar including on-chip antennas resulting in a small size. In Fig. 2 the radar chipset is depicted in comparison to an one euro coin.

The radar is configured to send chirps within a frequency range starting from $f_{\text{min}} = 59.5\,\text{GHz}$ to $f_{\text{max}} = 61.5\,\text{GHz}$. Thus, a frequency bandwidth $B = 2\,\text{GHz}$ is covered, which results in a range resolution of $\delta r = 7.5\,\text{cm}$. Moreover, the radar sends bursts of chirps, which are referred as data frames. A data frame consists of $N_\text{c} = 64$ chirps with a chirp repetition time T_CRT or $400\,\text{ms}$. This allows to unambiguously observe a maximum absolute frequency v_max of $3.125\,\text{m}\,\text{s}^{-1}$. The back-scattered signal is down converted to baseband by mixing with an replica of the transmit signal. The anti-aliasing filter has a bandwidth of $600\,\text{kHz}$. A sampling rate of $f_\text{s} = 2\,\text{MHz}$ is applied for the analog-to-digital converter (ADC).

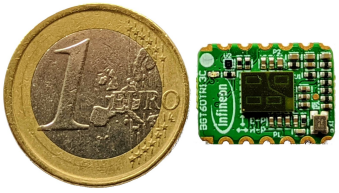


Fig. 2: *Infineon*'s *BGT60TR13C* radar chipset in comparison to an one euro coin

B. Radar Signal Description

As derived in [13] the two dimensional time domain signal for a single receiving antenna is described as

$$S_\text{n}(d_\text{n}, t_\text{ft}, t_\text{st}) = A\cos(2\pi(\frac{2Bd_\text{n}}{cT_\text{c}}t_\text{ft} + \frac{2f_0(d_\text{n} + v_\text{n}t_\text{st})}{c})). \quad (1)$$

where A is the voltage amplitude, B is the chirp bandwidth, d_n is the radial distance of target n to the radar, f_0 is the center frequency, c is the speed of light, T_c is the chirp time and t_st and t_ft are the slow- and fast-time, respectively.

The actually received signal is a superposition of the back-scattered signal of multiple targets. Thus, the final signal representation is defined as

$$S = \sum_{\text{n}\in\text{N}} S_\text{n}(d_\text{n}, t_\text{ft}, t_\text{st}) \quad (2)$$

where N is a set of targets. Since the processing is performed on the ADC signal, the discretized signal is defined as

$$S = \sum_{\text{n}\in\text{N}} S_\text{n}[d_\text{n}, k, l] \quad (3)$$

where k and l are swiping from 0 to N_s and N_c respectively.

C. Conventional Signal Processing

For indoor applications an important processing step is to apply a moving target indication, which means that the back-scattered signal of static targets is removed. Moreover, there is a leakage between transmit and receive antenna which adds low frequency perturbations within the chirps. A simple but effective approach to mitigate both issues is to remove first the mean within each chirp and then subtract the mean in slow-time direction, which means across the chirps.

The most interesting information within a radar signal when it comes to activity classification, as which also the gait classification can be seen in a wide sense, is the velocity profile. As derived in sec. II-B the velocity information is encoded in the frequency of the signal. Thus, a very common processing step is to apply a 2D short-time Fourier transform (STFT). The 2D STFT is defined as

$$\text{v}_\text{RDM}[d, r] = \left| \sum_{m=1}^{N_\text{c}} \sum_{n=1}^{N_\text{s}} w[m,n]\text{s}[m,n] e^{-j2\pi(\frac{md}{N_\text{c}} + \frac{nr}{N_\text{s}})} \right| \quad (4)$$

where $w[m,n]$ is a 2D window function, $s[m,n]$ is the mean removed ADC data of a data frame. The indices n, m sweep along the fast-time and slow-time axis, while r, d sweep along the range and Doppler axes. N_c and N_s are number of chirps and samples respectively.

For classifying different activities or gaits it is irrelevant at which range the person walks. Thus, in case there is only a single target in the field of view, the range information can be removed by integrating the RDM along range direction without losing any information, but significantly reducing the data size. The integration is defined as

$$DS[d] = \sum_{r=1}^{N_\text{s}} \text{v}_\text{RDM}[d, r]. \quad (5)$$

III. Proposed Solution

The classification accuracy can highly be impacted by distortions like a person walking by or a curtain waving in the background. By following the conventional processing steps the entire range-Doppler space is equally present in the RDM.

Consequently, also undesired distortions are clearly visible in the signal. Furthermore, target detection via peak detection or constant false alarm rate (CFAR) on RDM lead to spurious peaks and erroneous peaks due to transformation of superposed target components leading to sub-optimal feature extractions. To address the above issues of multi-target interference and sub-optimal feature extractions, we propose in this paper a regulation mechanism to track the target's fast-time and slow-time frequencies directly on time domain data using a 2D sinc filter.

A. Initial Target Detection

To detect a target initially a simple thresholding on the mean removed signal is applied. If the signal power surpasses the threshold a STFT is applied and the bin with the highest amplitude is selected as target bin. The center frequencies (fast-time and slow-time or range and Doppler) of this range-Doppler bin is then chosen as initial center frequency of the 2D sinc filter. The 2D sinc filter is the outer product of two complex valued 1D sinc filters. A complex valued 1D sinc filter is defined as

$$h[n, f_c, b] = 2b\ \mathrm{sinc}(2bt)\left(\cos(f_c nT) + j\sin(f_c nT)\right) \tag{6}$$

with $n \in \left[-\lfloor\frac{N}{2}\rfloor, \lfloor\frac{N}{2}\rfloor\right]$, where N is the filter length, T the sampling interval, f_c the center frequency and b the bandwidth. Thus the complex valued 2D sinc filter is defined as

$$\mathrm{sinc}_{2D}[n, m, f_{c,st}, f_{c,ft}, b_{st}, b_{ft}] = h[n, f_{c,st}, b_{st}]h[m, f_{c,ft}, b_{ft}] \tag{7}$$

whith $n \in \left[-\lfloor\frac{N_c}{2}\rfloor, \lfloor\frac{N_c}{2}\rfloor\right]$ and $m \in \left[-\lfloor\frac{N_s}{2}\rfloor, \lfloor\frac{N_s}{2}\rfloor\right]$. The bandwidths are constantly set to a normalized frequency bandwidth of 0.4 for Doppler frequency and 0.2 for range frequency, which is an empirically found value that is large enough to cover the range and Doppler variation of a human while walking.

B. Adaptive 2D Sinc Filters

By walking the person is changing its range position relative to the radar. Moreover, the walking speed might change when stopping, turning around and moving into the opposite direction. Hence, the 2D sinc filter has to follow the target's movements. To do so, the fast-time center frequency as well as the slow-time center frequency is adapted according to the current observed signal. This is done for each dimension independently, but in the same manner.

First, a complex valued 2D sinc filter is defined with an filter length of 128 in range direction, which is equal to the number of samples per chirp, and a filter length of 32 in Doppler direction, which is smaller than the number of chirps within a data frame. The raw signal of a data frame is convolved with this filter without using zero padding in range direction. Thus, only a single value in range direction is obtained. However, along slow-time zero padding is used and thus multiple points are obtained in slow-time dimension. Consequently, the 2D convolution results in a one dimensional slow-time signal within the range frequency region defined by the 2D sinc filter. The same procedure is done by using a 2D sinc filter with filter length of 32 in range direction and 64 in Doppler direction. Now zero padding is only applied in range direction to obtain a 1D range signal within the Doppler frequency region defined by the fast-time center frequency and bandwidth of the 2D sinc filter.

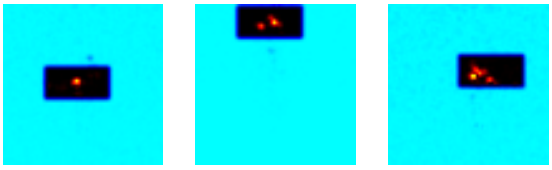
The average angular velocity for each complex 1D signal s is determined using

$$\omega_{\text{avg}} = \frac{1}{N-1}\sum_{n=1}^{N-1}\left[\phi(s[n]) - \phi(s[n-1])\right] \tag{8}$$

where $\phi(.)$ denotes the phase. Based on the average angular velocity the center frequency for the new time step can be updated by

$$f_{\text{c, t+}\Delta\text{t}} = \frac{\omega_{\text{avg}}}{2\pi} \tag{9}$$

for slow and fast-time independently in the same manner. By this the 2D sinc filter is basically hunting the target in range-Doppler domain by completely operating on the raw time domain data. In Fig. 3 the adaptive sinc filtering approach is visualized. The RDMs at three different time steps from a recording of a single person approaching the radar and walking away from it along with the frequency response of the adaptive tracking sinc filter is shown.



(a) RDM 1 (b) RDM 2 (c) RDM 3

Fig. 3: Target tracking visualized. A person is approaching the radar and moves away again. The figures shows the RDM together with the frequency response of the tracking sinc filter at three different time steps.

The 1D Doppler signal after filtering contains rich features about the targets movements. By applying a STFT on each 1D Doppler signal and stacking the signals of multiple time steps together the Doppler spectrogram (DS) can be generated.

IV. Results and Discussion

To evaluate the proposed solution, two different experiments were performed. First, a gait classification with 8 different persons was done and second the advantage of the proposed method compared to a conventional state-of-art processing using peak detection for two-person scenarios was evaluated.

A. Experimental Setup

The sensor was placed on a tripod at a height of about $1\,\mathrm{m}$ and configured with the settings shown in sec. II-A. The recorded persons were walking up and down in front of the

sensor in a range from 1 m to 5 m. For the gait classification in total 8 different volunteers were recorded. It was ensured that there are no distortions in the background during the recordings. If the signal power exceeds a certain threshold, then a target is detected and the recording starts. If the signal power drops below the threshold or if a maximum number of 200 data frames is reached, then the recording stops.

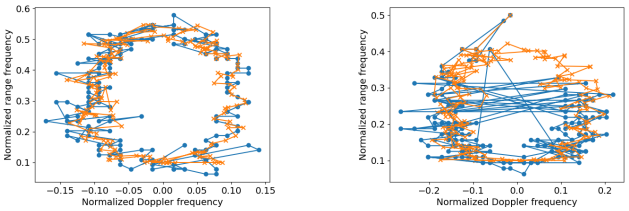
B. Signal Quality

By first filtering the signal in time domain and then extracting the features, the signal quality can be improved compared to transforming the signal first to frequency domain and then extracting the features. This is shown by evaluating the peak signal to noise ratio (PSNR) on the target extracted DSs for the conventional peak detection and the proposed adaptive sinc filter approach. To evaluate the PSNR, the 10% highest amplitude values are chosen as signal values, and the 80% lowest amplitude values are used for estimating the noise level. Therefore, a gap of 10% of the samples is used as safety margin between signal or noise signal parts. The average PSNR of the conventional algorithm on the entire gait classification data set is 36.1 dB. With pre-filtering the time domain signal and then transforming it to frequency domain a PSNR increase of 2.6 dB up to 38.7 dB was achieved.

C. Tracking Evaluation

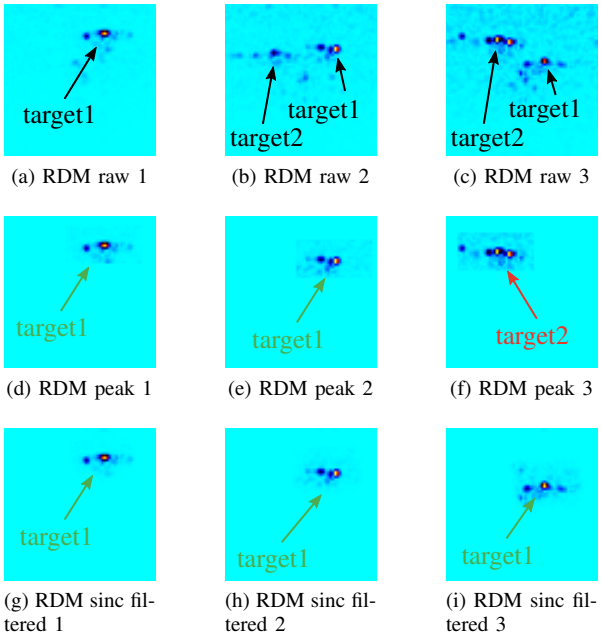
A comparison of a conventional peak detection on the pre-processed RDM with the proposed 2D adaptive sinc filter tracking is performed. For this the center frequencies of the peak target detected in the RDM are plotted along with the center frequencies followed by the proposed adaptive sinc filter approach in Fig. 4. In Fig. 4 (a) the tracks are shown for a single target in the field of view. For this setting both approaches can track the target well. However, in Fig. 4 (b) the tracks are shown for a two target scenario, where the two persons were walking inversely up and down. The proposed adaptive sinc filter is able to follow the intended target, whereas the simple peak detection switches back and forth between the two targets depending on which one back-scatters the most energy, which is usually the closer target. The lower variation of the detected target peaks along the 2D frequency represents an improved target detection by the proposed sinc filter approach.

The target switching problem is additionally visualized using RDMs in Fig. 5. The top row shows the raw RDMs. In the first time step target 1 is close to the radar and moving away from it. In the second time step a second target approaching the radar can be seen in the RDM. Finally, in the third time step target two is closer to the radar than target two. The second row shows the target detected by a peak detection mechanism on the RDM. The bottom row shows the RDM on the basis of the pre-filtered signal using the adaptive sinc filter. While the area of interest determined by peak detection jumps to the second target as soon as it is closer to the radar and therefore back-scattering more signal power, the adaptive sinc filter successfully continues to track target 1.



(a) One person in the field of view (b) Two persons in the field of view

Fig. 4: Tracked center frequencies in range-Doppler domain for a single (a) and two (b) person scenario. The Blue line shows the track of the peak detection approach and orange the proposed approach.



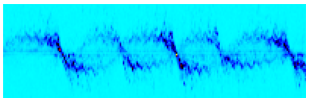
(a) RDM raw 1 (b) RDM raw 2 (c) RDM raw 3

(d) RDM peak 1 (e) RDM peak 2 (f) RDM peak 3

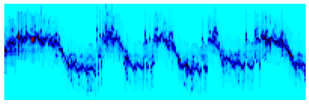
(g) RDM sinc filtered 1 (h) RDM sinc filtered 2 (i) RDM sinc filtered 3

Fig. 5: Target tracking with a second person approaching the sensor. Target 1 is the intended target to track. Correct target detection is visualized in green and the miss-detection of target 2 is visualized with red color.

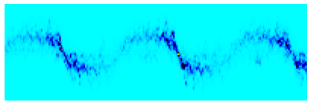
Additionally, in Fig. 6 the same scene is visualized with DSs. In Fig. 6 (a) the DS generated from the unfiltered signal is given. The two persons walking up and down in opposite directions can clearly be seen. In Fig. 6 (b) the DS generated from the extracted signal of the target found by peak detection is shown. Obviously, the target region switches to the closer target, which can be seen in the DS. In comparison, in Fig. 6 (c), the DS generated from the adaptive sinc filtered signal is depicted. Only the signal of the intended target is within the DS and the signal of the second target was successfully filtered out. Moreover, the update step is a continuous function, the center frequencies are continuous. Due to this and the fact that the filter edge is also continuous, the resulting DS is much smoother than the one obtained from peak detection where the center frequencies are limited to the discrete range-Doppler bins.



(a) Unfiltered Doppler spectrogram



(b) Doppler spectrogram generated from peak extracted RDM signal



(c) Doppler spectrogram generated from sinc-filered signal

Fig. 6: Doppler spectrograms comparison of the different processing approaches.

D. Gait classification

A gait classification task it used to show the advantage of keeping the data in time domain during target extraction. This gives the possibility of integrating the pre-processing into the neural network.

There are 200 recordings of each person with unequal length. However, since the neural network requires a fixed input size, all samples were zero padded to a equal number of 200 time steps. The final input shape is therefore $(200, 64, 1)$, where 200 is the number of data frames or time steps, 64 is either the number of complex valued Doppler signal points or the number of Doppler bins depending on whether raw time domain data or pre-processed DS are used as network input. For training $80\,\%$ or 1280 samples are randomly chosen and for validation the remaining $20\,\%$ of the samples are used.

The network consists of four Inception v1 modules [14] followed by 2 fully connected layers of size 128 and 64. After each fully connected layer, a dropout with a rate of 0.4 is applied and after each inception module the data is compressed using max pooling of size $(2, 2)$. Per path in the four inception modules 4, 8, 16 and 32 filter kernels are applied respectively. When the time domain data is fed into the network, additionally a complex frequency extraction layer (CFEL) as proposed in [13] is used at the beginning of the neural network. The network was initialized with the 'glorot' inizialization scheme and focal loss in combination with rectified adam as optimizer was used for training. A 5-fold cross validation was performed to get more stable or meaningful results. The evaluated approaches are using the unfiltered DS, the DS computed from the previous sinc-filtered signal and the sinc-filtered signal directly in combination with a CFEL at the beginning of the network. In tab. I the accuracy and standard deviation obtained from the 5 runs of the different approaches is shown. The approach integrating the pre-processing into the neural network itself shows the best results and thus proves the advantage of using raw data directly as input.

TABLE I: Gait classification results

	Unfiltered DS	**Unfiltered raw**	**sinc-filtered DS**	**sinc-filtered raw**
Accuracy (± std dev)	89.8% (±1.6%)	90.4 (±1.2)	91.1% (±0.5%)	**92.4%** (±2.4%)

V. Conclusion

The paper presents a complex-valued adaptive 2D sinc filter applied directly on the radar time domain data, which leads to better target detection and feature extraction. We demonstrate that the adaptive sinc filter via the regulation loop can progressively track the target's fast-time and slow-time frequencies even in the presence of multiple targets. The superior performance of the proposed solution compared to state-of-art processing using peak detection is shown with respect to a radar gait classification task.

References

[1] A. Santra and H. Souvik, *Deep Learning Applications of Short-Range Radars*. Artech House Books, 2020.

[2] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from fmcw radar with bi-lstm networks," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 607–13 619, 2020.

[3] Y. Sun, T. Fei, X. Li, A. Warnecke, E. Warsitz, and N. Pohl, "Real-time radar-based gesture detection and recognition built in an edge-computing platform," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10 706–10 716, 2020.

[4] H. R. Lee, J. Park, and Y.-J. Suh, "Improving classification accuracy of hand gesture recognition based on 60 ghz fmcw radar with deep learning domain adaptation," *Electronics*, vol. 9, no. 12, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/12/2140

[5] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019, pp. 33–40.

[6] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, ser. mmNets'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 51–56. [Online]. Available: https://doi.org/10.1145/3349624.3356768

[7] T. Stadelmayer, A. Santra, R. Weigel, and F. Lurz, "Parametric Convolutional Neural Network for Radar-based Human Activity Classification Using Raw ADC Data," 9 2020. [Online]. Available: https://www.techrxiv.org/articles/preprint/Parametric_Convolutional_Neural_Network_for_Radar-based_Human_Activity_Classification_Using_Raw_ADC_Data/12896108

[8] W. Jiang, Y. Ren, Y. Liu, and J. Leng, "A method of radar target detection based on convolutional neural network," *Neural Computing and Applications*, pp. 1–13, 2021.

[9] M. Arsalan, A. Santra, and C. Will, "Improved contactless heartbeat estimation in fmcw radar via kalman filter tracking," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

[10] Z. Wang, Z. Yu, X. Lou, B. Guo, and L. Chen, "Gesture-radar: A dual doppler radar based system for robust recognition and quantitative profiling of human gestures," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 1, pp. 32–43, 2021.

[11] Y. Liu, Y. Wang, H. Liu, A. Zhou, J. Liu, and N. Yang, "Long-range gesture recognition using millimeter wave radar," in *Green, Pervasive, and Cloud Computing*, Z. Yu, C. Becker, and G. Xing, Eds. Cham: Springer International Publishing, 2020, pp. 30–44.

[12] S. Hazra and A. Santra, "Radar gesture recognition system in presence of interference using self-attention neural network," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 1409–1414.

[13] M. Stephan, T. Stadelmayer, A. Santra, and G. Fischer, "Radar image reconstruction from raw adc data using parametric variational autoencoder with domain adaptation," in *25th International Conference on Pattern Recognition*. IEEE, 2021.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.