CV-SincNet: Learning Complex Sinc Filters From Raw Radar Data for Computationally Efficient Human Motion Recognition

Sabyasachi Biswas, *Graduate Student Member, IEEE*, Cemre Omer Ayna, *Student Member, IEEE*, Sevgi Z. Gurbuz, *Senior Member, IEEE*, and Ali C. Gurbuz, *Senior Member, IEEE*

***Abstract*— The utilization of radio-frequency (RF) sensing in cyber-physical human systems, such as human-computer interfaces or smart environments, is an emerging application that requires real-time human motion recognition. However, current state-of-the-art radar-based recognition techniques rely on computing various RF data representations, such as range-Doppler or range-Angle maps, micro-Doppler signatures, or higher dimensional representations, which have great computational complexity. Consequently, classification of raw radar data has garnered increasing interest, while remaining limited in the accuracy that can be attained for recognition of even simple gross motor activities. To help address this challenge, this paper proposes a more interpretable complex-valued neural network design. Complex sinc filters are designed to learn frequency-based relationships directly from the complex raw radar data in the initial layer of the proposed model. The complex-valued sinc layer consists of windowed band-pass filters that learn the center frequency and bandwidth of each filter. A challenging RF dataset consisting of 100 words from American Sign Language (ASL) is selected to verify the model. About 40% improvement in classification accuracy was achieved over the application of a 1D CNN on raw RF data, while 8% improvement was achieved compared to real-valued SincNet. Our proposed approach achieved a 4% improvement in accuracy over that attained with a 2D CNN applied to micro-Doppler spectrograms, while also reducing the overall computational latency by 71%.**

***Index Terms*— Radar, RF sensing, FMCW, micro-Doppler signature, CV-SincNet, human activity recognition, ASL.**

I. Introduction

RECENT development of affordable solid-state transceivers and computationally efficient graphics processing units (GPUs) along with novel deep learning (DL) approaches has increased the usability of radio frequency (RF) sensors in an ever expanding range of applications involving human activity recognition (HAR) [1], such as defense and security [2], [3], [4], mini-UAV classification [5], advanced driver assistance system (ADAS) [6], [7], indoor monitoring [8], and health monitoring [9], [10]. Recently, real-time sign language recognition has experienced significant advancements, partly attributed to the emergence and progress of the MediaPipe toolbox [11]. Sign language recognition falls within the realm of Human Activity Recognition (HAR) due to the movement-based kinematic components of the language. Although sometimes conflated with gesture recognition, sign language differs due to the communicative nature of the signal, which includes grammatical structure and linguistic parameters, such as prosody, suprasegmental components, and phrasing. Thus, comprehensive sign language recognition involves correlating such linguistic parameters with fine spatio-temporal dynamics. Towards this aim, RF sensing has recently been proposed as a novel modality for sign language recognition [12], [13] as well as gesture-based human-computer interaction [14]. Advancements in both these areas have also contributed more broadly the real-time, radar-based HAR.

Most conventional approaches for radar-based HAR require a two-level process because raw radar data are a time series of complex I/Q data. First, time-frequency analysis or other radar signal processing techniques are applied to generate 2D (or higher 3D and 4D) radar data representations, such as time varying range-Doppler or range-Angle maps or micro-Doppler (μ-D) signatures (μ-DS) [15]. While there have been some studies that propose joint domain classification [16], [17], most studies utilize the μ-DS for HAR. The μ-D is a time-frequency (TF) domain representation of the radar data that can be obtained by using transforms such as the short-time Fourier transform (STFT) [18], wavelet transform [19], or Gabor transform [20]. These approaches require the optimization of a number of parameters, including window type, window size, Fast Fourier Transform (FFT) length, and overlap size between successive windows. After the first step of processing, the magnitude of the μ-DS is obtained and given as the input to a deep neural network (DNN) for human activity classification [21], [22], [23], [24]. This multi-step processing with parameter optimization consumes a lot of computational resources, while also having high temporal

Manuscript received 22 April 2023; revised 10 July 2023; accepted 27 August 2023. Date of publication 31 August 2023; date of current version 14 September 2023. This work was supported in part by the National Science Foundation (NSF) Awards under Grant 1932547, Grant 1931861, Grant 2047771, and Grant 2238653; in part by the American Association of University Women (AAUW) via a Research Publication Grant in Engineering, Medicine, and Science; in part by the Air Force Office of Scientific Research (AFOSR) Award under Grant FA9550-22-1-0384; and in part by the U.S. Engineer Research and Development Center under Grant W912HZ-21-2-0053. *(Corresponding author: Ali C. Gurbuz.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Alabama's Institutional Review Board (IRB) under Approval No. 18-06-1271.

Sabyasachi Biswas, Cemre Omer Ayna, and Ali C. Gurbuz are with the Department of Electrical and Computer Engineering and the Information Processing and Sensing Laboratory, Mississippi State University, Mississippi State, MS 39672 USA (e-mail: gurbuz@ece.msstate.edu).

Sevgi Z. Gurbuz is with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: szgurbuz@ua.edu).

Digital Object Identifier 10.1109/TRS.2023.3310894

latency, which imposes limitations in real-time sensing applications.

Consequently, there has been some interest in the development of classification techniques that directly take as input the raw, complex RF data stream. In [25], the authors proposed a recurrent neural network (RNN) to decode the time sequence of 2D raw I/Q radar data, where the I and Q channels are jointly provided as inputs to a real RNN. Six activities were classified with an average accuracy of 93%. However, this accuracy came with high computational cost: training the network took over 240 hours using four Nvidia GeForce GTX 1080 GPUs. In [26], a similar approach was utilized but now the range profile in addition to the raw I/Q data streams were utilized with a Long Short Term Memory (LSTM) network. In [27], the authors used real 2D sinc filters and 2D wavelet filters as the first layer of the neural network model, which operates directly on 2D raw radar data and addresses a 6-class gesture recognition problem. Note, however, that in this work as well the complex-valued (CV) raw radar data is fed into a real-valued network through separation of the real and imaginary components of the data.

The complex-valued convolutional neural network (CVCNN) architectures are proposed for several radar-based HAR problems. In [28] and [29] the authors proposed a CVCNN architecture that trains on the complex-valued spectrograms instead of real-valued spectrogram images calculated in the logarithmic scale. The authors in [29] also introduced FourierNet where they used a windowed Fourier transform which remains as a fixed pre-processing and operates on the 1D-time series data before the STFT operation. In [30], the authors used CVCNN techniques on complex-valued 2D representations such as Range-time, Range-Doppler maps, or spectrograms.

Although deep learning (DL) and specifically CNNs have shown unprecedented performance levels in various applications for directly processing raw data, these models are often regarded as "black-box" and the high performance is often obtained at the cost of reduced interpretability of the model. In [31] more interpretable filters that can directly work on the raw audio waveform are learned for the speaker identification task. Instead of learning all elements of every filter as the standard CNNs, one-dimensional (1D) sinc functions are utilized in the first layer of the network that implements band-pass filtering operations. In this SincNet architecture, the low and high cutoff frequencies of the Sinc filters are the only learnable parameters for each filter. These parameters are optimally learned to capture specific frequency components in the input signal. Moreover, these sinc filters when applied to the input signals can capture important temporal information. As a result, these filters provide a more meaningful and contextually relevant representation of the input signals which allows the network to focus on higher-level learnable parameters that have clear physical meanings. However, since audio data is real, original SincNet [31] learns two-sided band-pass filters, which can not be directly applied to complex raw radar data.

In this paper, we introduce CV-SincNet, a complex DNN architecture for classifying various human activities by directly learning from 1D complex slow-time raw radar data after range processing. Different from the real SincNet [31], the CV-SincNet can directly work with the complex data and convolves the input waveform with a set of newly designed complex sinc-functions that implements one-sided bandpass filters instead of symmetric band-pass filters learned by real SincNet. We parametrize each complex Sinc filter by the center frequency and bandwidth and we directly learn these high-level tunable parameters for each filter that have clear physical meaning. The remaining DNN operations such as convolution, non-linear activations, pooling, or normalization are modified to work on complex raw data [32].

In preliminary work [33], we showed on a small dataset of 15 ASL signs that a CV-SincNet architecture with fewer layers can provide comparable classification accuracy relative to that attainable by applying a 2D CNN on the spectrogram of RF data, i.e. the micro-Doppler signature. In this paper, we expand upon this work to show that actually improved accuracy can even be obtained for quite challenging tasks, such as the recognition of as much as 100 American Sign Language (ASL) signs [34] using a slightly deeper variant of CV-SincNet. We also show that at the same time, CV-SincNet also results in a significant reduction in complexity and computational latency, rendering it suitable for real-time applications. We also propose and develop quantitative analysis of interpretability, which provides insight on the most important velocity-based features and ASL properties.

The proposed CV-SincNet architecture is compared with various neural network solutions that work on raw radar data or computed representations: i) standard CNN working on 2D μ-DS images, ii) real SincNet that works on raw radar data taking the real and imaginary parts of the complex radar data as a two-layer input, and iii) a 1D complex CNN that directly works on the complex raw radar data. Comparisons made with the 2D CNN models that operate on the μ-D spectrogram images are produced by the same raw radar data other models are working on. The proposed CV-SincNet provided about 8% higher classification accuracy than the real SincNet approach showing the importance of complex-valued networks working on the complex-valued dataset. Moreover, about 4.2% better classification accuracy is observed compared to standard 2D CNN models applied on the μ-D spectrogram images in addition to significantly reducing the time required for activity prediction. These outcomes demonstrate how well the proposed CV-SincNet performs when used for activity classification with complex-valued raw radar data. In addition, the obtained filters are more interpretable as the proposed sinc layer focuses only on physically relevant specific Doppler frequency bands.

The contributions of this paper can be summarized as

- A general complex Sinc filter learning architecture is formulated and the proposed CV-SincNet is applied to the complex raw radar data for human activity recognition.
- The proposed CV-SincNet is compared with the 2D-CNN model on μ-D images and real-CNN, real-SincNet and CV-CNN models on raw radar data.
- More interpretable band-pass filters are learned focusing on physically meaningful Doppler frequencies and bandwidths.

- The proposed CV-SincNet shows both high accuracy and compatibility for real-time classification with reduced overall latency as compared to other approaches.

The organization of the paper is as follows: The radar signal model used in multiple input multiple output (MIMO) frequency modulated continuous wave (FMCW) radars for data collection is provided in Section II. Section III discusses the proposed CV-SincNet architecture along with complex implementation details of all the neural network blocks. The experimental setup for the radar sensor and dataset collection is provided in Section IV, and Section V presents the different networks used, comparison between the proposed approach and other networks and characteristics of the proposed network. Finally, Section VI concludes the paper and discusses future directions.

II. Background

A. Radar Signal Model

An FMCW radar system transmits a radio-frequency (RF) signal that linearly sweeps over its bandwidth to enable the acquisition of both range and velocity measurements [35]. The instantaneous frequency of the chirp signal can be modeled as

$$f_{ch}(t) = f_{st} + \frac{B}{\tau}t, \quad 0 \le t \le \tau, \tag{1}$$

where f_{st} is the initial frequency at time $t = 0$, B is the bandwidth and τ represents the sweep time in seconds. The received signal from a target with time delay T_d is mixed with a copy of the transmitted signal and passed through a low-pass filter (LPF) resulting in the intermediate frequency (IF) signal that is modeled as

$$S_{IF}(t) = Aexp(2\pi(f_{st}T_d + \frac{B}{\tau}T_d t - \frac{B}{2\tau}T_d^2)), \tag{2}$$

where A represents the amplitude of the signal. The FMCW radar samples the IF signal and stores the received data in a three-dimensional (3D) structure known as a radar data cube (RDC). Each pulse consists of a fixed number of samples known as fast-time samples. The return from each pulse, which is repeatedly transmitted with a fixed pulse repetition interval (PRI), is stacked to form a two-dimensional (2D) structure, where the columns denote the slow-time samples. The number of receiver channels is represented by the third dimension of the RDC. Here, $I \times J \times K$ can be considered as the shape of the RDC consisting of I fast-time samples, J slow-time samples, and K receiver channels. Different RF data representations, such as Range-Doppler (RD), Range-Angle (RA) or μ-D signatures can be computed from the RDC.

B. Micro-Doppler Spectrogram Generation

The μ-D shift refers to modulations in frequencies caused by vibrations, oscillations, or rotations of parts of a target, such as the moving arms and legs of a walking person. The kinematic characteristics of any target in the radar field of view (FoV) can be determined by the modulated frequencies created from their movements [36]. A positive Doppler shift implies motion toward the radar and a negative shift indicates motion away from the radar. Currently, the most commonly used time-frequency transform for visualizing the μ-D signature of a target is the spectrogram, which estimates the instantaneous micro-Doppler frequency as a function of time by computing the square modulus of the windowed Short-Time Fourier Transform (STFT) across the slow-time data $x(t)$ as

$$S(n, \omega) = \left| \int_{-\infty}^{\infty} h(t-n)x(t)e^{-j\omega t}dt \right|^2. \tag{3}$$

where $h(t)$ is the window function. Fig. 1 shows samples of the μ-D signatures of ASL signs acquired from fluent ASL users using a 77 GHz radar system, where the time-varying fluctuations of μ-D reveal the distinctive characteristics of each ASL sign.

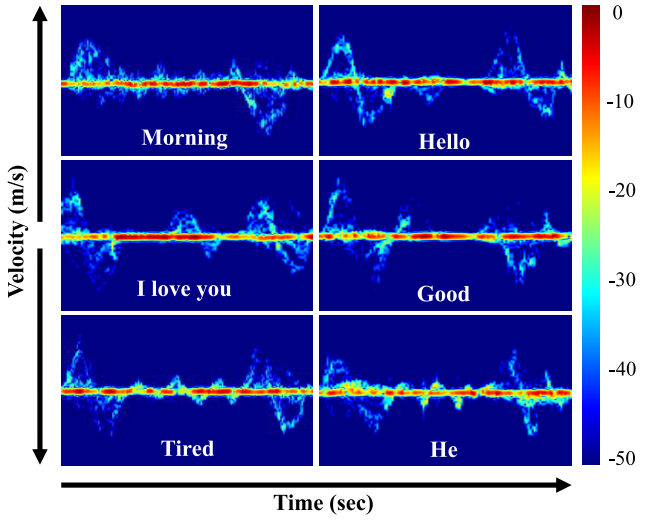


Fig. 1. Sample μ-D signatures of ASL signs.

III. Proposed Method

A. Learning Sinc Filters

Deep learning and modern neural networks have achieved transformative advancements in many fields, especially computer vision, but are based on the provision of real-valued inputs and often with the cost of reduced explainability of the underlying model. The filters learned by CNNs are generally noisy multi-band shapes that could make sense for a machine but not for human intuition. To help CNNs to learn more meaningful filters directly from raw data, [31] proposed SincNet, a new CNN architecture for speaker identification using real-valued, raw audio data that learns parametrized sinc functions in its first layer.

SincNet convolves the input waveform with a set of parametrized sinc functions that implement band-pass filters [37]. In this case, only the low and high cutoff frequencies of the band-pass filters are learned from the data instead of learning all taps of a filter. An ideal real bandpass filter response can be defined as the subtraction of two low-pass filters with cut-off frequencies f_ℓ and f_h in the frequency domain, where f_h is the higher and f_ℓ is the lower cut-off frequencies of the filter respectively. The ideal bandpass function, which is a rectangle function in the frequency domain, corresponds to a sinc function in the time domain. The definition of the real sinc

filter in time and frequency domains can be found in (4) and (5). Equation (5) shows that the filter is symmetric with respect to the origin in frequency domain; hence the time-domain filter in (4) is real, making it suitable for real-valued signals such as speech audio.

$$h[n, f_\ell, f_h] = 2f_h sinc(2\pi f_h n) - 2f_\ell sinc(2\pi f_\ell n) \quad (4)$$

$$H[f, f_\ell, f_h] = rect(\frac{f}{2f_h}) - rect(\frac{f}{2f_\ell}) \quad (5)$$

The trainable parameters for the original SincNet are the low and high cut-off frequencies f_ℓ and f_h respectively. Since derivatives with respect to these parameters can be calculated the sinc layer is compatible with the backpropagation operation, and thus the full model can be trained by the stochastic gradient descent algorithm. In order to guarantee that there is no aliasing in the convolution operation, the cut-off frequencies are first initialized in the range $[0, f_s/2]$, where f_s is the sampling frequency. There is no additional constraint for guaranteeing the range of the parameters during training since [31] reports that the parameters are observed not to move out of the range $[0, f_s/2]$. The layer also contains no additional learnable parameter for the amplitude since the consequent convolutional layers do the work of assigning importance over the filtered inputs. The sinc filter is multiplied with a fixed Hamming window before the convolution process to mitigate ripples near the edges of the pass-band.

The original SincNet learns only real sinc filters which have corresponding two-sided band-pass filter responses. This is effective for real inputs, such as audio waveforms; however, radar raw data are complex I/Q time streams, while micro-Doppler frequencies can be positive or negative, depending on the underlying physical movement. Thus, real-valued SincNet is not well-matched for radar applications. In contrast, this paper develops a new DNN, CV-SincNet, that learns from the radar data complex sinc filters with a one-sided frequency domain band-pass representation.

B. Complex SincNet Building Blocks

One of the most important benefits of using complex-valued neural network models for radar signals is that it allows us to use the complex raw radar data directly without requiring exhaustive pre-processing, which can have steep computational complexity and latency. The proposed CV-SincNet architecture is composed of a complex sinc layer followed by complex convolution and complex fully-connected layers. Complex activations, pooling, and normalization operations are also required within the architecture. The theoretical basis of the complex-valued neural network operations for convolution or activations are provided by several studies [32], [38]. Next, details of the CV-SincNet architecture are provided.

1) Complex Sinc Layer: We propose a newly designed complex-valued sinc layer with structural changes in filter definitions in order to adapt to complex-valued raw radar input. We learn a complex filter-bank composed of single-sided rectangular band-pass filters that are modeled with their center frequency f_c and bandwidth B parameters. Different from the real-valued sinc filter in (4), the time-domain representation

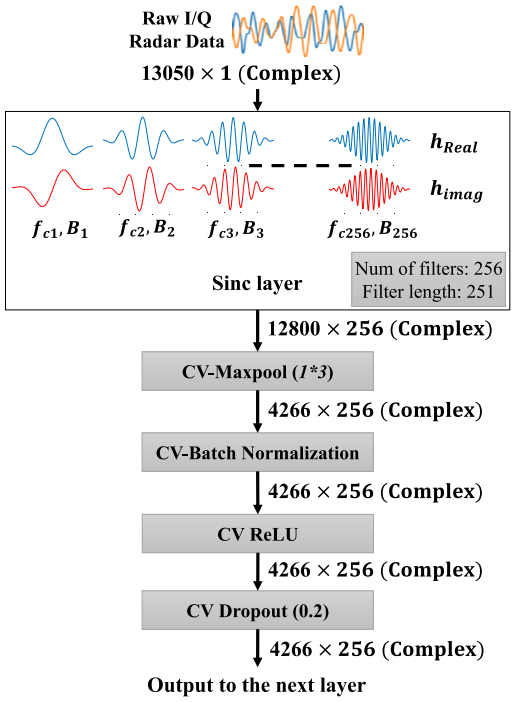


Fig. 2. The CV-Sinc block.

of the complex sinc filter is formulated as

$$h[n, B, f_c] = 2Bsinc(2\pi Bn) \times e^{j2\pi f_c n} \quad (6)$$

The complex filter is differentiable with respect to its learnable parameters: the center frequency, f_c, and the bandwidth, B; hence, it is also compatible with backpropagation. Each complex-valued sinc filter has an equally-sized real and imaginary part as $h = h_r + jh_i$.

$$h_r = 2Bsinc(2\pi Bn) \times cos(2\pi f_c n) \quad (7)$$

$$h_i = 2Bsinc(2\pi Bn) \times sin(2\pi f_c n) \quad (8)$$

We apply windowing by multiplying learned filters with a fixed Hamming window as $h_\omega[n, B, f_c] = h[n, B, f_c] * \omega[n]$ to smooth out the discontinuities at the edges of the band-pass filters. The Hamming window utilized is defined as

$$\omega[n] = 0.54 - 0.46\cos(\frac{2\pi n}{L}) \quad (9)$$

where L is the length of the window and the learned filters. The window is fixed and has no learnable parameters; hence, the only learnable parameters in the complex SincNet layer as shown in Fig. 2 are $f_{c,k}$ and B_k for each learned filter k. While the neural architecture in Sinc layer updates $f_{c,k}$ and B_k parameters of neurons, the input data is passed through sinc filters corresponding to the updated parameters defined by equations (7) and (8). As shown in Fig. 2 a pipeline of pooling, normalization, ReLU activation and a dropout with complex operations are employed after the initial CV-Sinc layer. This sequence of layers is considered as the CV-Sinc block. After the CV-Sinc block, the model is comprised of three parallel convolutional blocks (CB) followed by six series CBs. Each CB consists of a convolution, maxpooling, batch normalization, a ReLU activation function, and a dropout

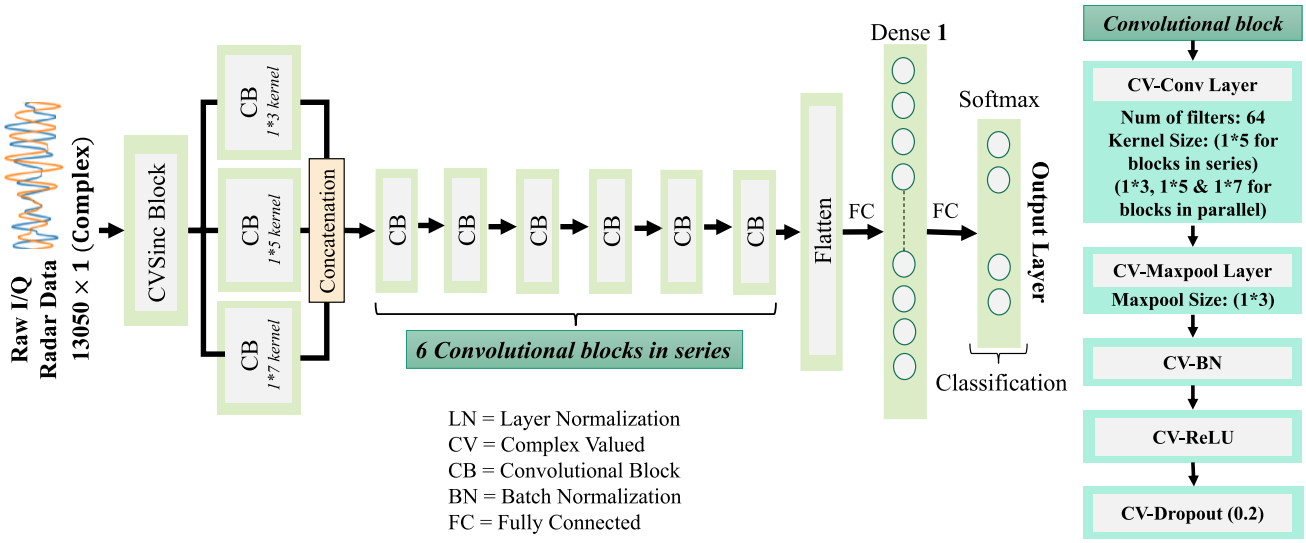


Fig. 3. Flow diagram of the CV-SincNet architecture.

layer. The number of convolutional blocks can differ based on the dimensionality of the input data. Then the tensor is flattened and fed into a dense layer, after which classification is done with a softmax layer. The general architecture of the CV-SincNet model is given in Fig. 3.

The proposed CV-SincNet model has several advantages. In addition to being directly operable on complex raw radar data, the proposed complex sinc layers have a fewer number of learnable parameters as compared to conventional layers. A 1D sinc filter layer with K number of filters, regardless of the length of the filter L, has only $2K$ learnable parameters, whereas a 1D convolutional layer with K filters has $L * K$ learnable parameters. This characteristic helps reduce the model size and computational complexity. Another advantage is the interpretability of the sinc layer. The parameters learned in the sinc layer have direct physical significance by showing which specific Doppler frequencies and bands of the neural network are important to the radar classification task. This makes it much easier to interpret what the model has learned as compared to traditional neural network architectures.

2) Complex Convolution Layer: To utilize complex-valued representations in deep neural networks [32], complex convolution is defined using real-valued representations of the complex-valued input and the kernel weights. The complex kernel matrix is defined as $W = W_R + jW_I$, where $W \in \mathbb{C}^{L\times K}$ for K number of kernels of length L; W_R and W_I are the real and imaginary parts of W, respectively The distributive property of the convolution operation allows for the definition of the complex convolution operation of W with the complex input $x = x_R + jx_I$ in the real-domain as in (10).

$$W * x = (W_R * x_R - W_I * x_I) + j(W_R * x_I + W_I * x_R) \tag{10}$$

Note that the real and imaginary parts of the complex convolution can be computed separately and the result is completely different than just doing real convolutions separately on real and imaginary channels. We can represent a complex-valued tensor with N channels as a tensor with $2N$ channels, where the first N channels are the real values and the rest of the N channels are the imaginary values. Such a representation is helpful for the computations over different layers.

3) Complex Fully Connected Layer: A complex-valued fully connected layer with M number of neurons has a complex weight matrix $W \in \mathbb{C}^{M\times N}$ and a complex bias vector $\beta \in \mathbb{C}^M$. The fully connected layer's output to a complex input $x \in \mathbb{C}^N$ is simply $Wx+\beta$ and can be computed in real domain as

$$Wx + \beta := \begin{bmatrix} \Re(W) & -\Im(W) \\ \Im(W) & \Re(W) \end{bmatrix} \begin{bmatrix} \Re(x) \\ \Im(x) \end{bmatrix} + \begin{bmatrix} \Re(\beta) \\ \Im(\beta) \end{bmatrix} \tag{11}$$

where the output's first M channels are the real part and the second M channels are the imaginary part.

4) Activation Functions With Complex Features: Two activation functions are used in the real SincNet model; ReLU within the architecture layers and SoftMax at the final layer. The proposed CV-SincNet utilizes the complex versions of ReLU within the architecture layers and SoftMax at the final layer. The complex ReLU function ($\mathbb{C}$ReLU) proposed in [31] applies the ReLU function over the real and imaginary parts of the input separately, as defined by

$$\mathbb{C}ReLU(x) := \begin{bmatrix} ReLU(\Re(x)) \\ ReLU(\Im(x)) \end{bmatrix} \tag{12}$$

The complex SoftMax function $\sigma_{\mathbb{C}}(x)$, on the other hand, is defined as the standard SoftMax function $\sigma(x)$ applied on the absolute value of the complex input as $\sigma_{\mathbb{C}}(x) = \sigma(|x|)$.

5) Complex MaxPooling and Batch Normalization: The complex MaxPooling layer is defined as a filtering operator over the absolute value of the complex input. In the first step, the absolute values of the complex input tensor are calculated. Then, the traditional MaxPooling with a specified filter length and stride is applied over the absolute values and the index of the highest absolute value is recorded. The actual output is then defined as the complex value that was located in the saved index.

Batch normalization is generally useful to accelerate learning. The batch normalization for complex data is achieved

TABLE I
TI IWR1143 RADAR PARAMETERS

Parameter	Value
Number of ADC samples	256
Number of TX channels	1
Number of RX channels	1
Start frequency	77 GHz
Stop frequency	81 GHz
Bandwidth	4 GHz
RX gain	45 dB
Periodicity	40 ms
Pulse repetition interval (PRI)	312.5 μs
Pulse repetition frequency (PRF)	3200 Hz
Number of ADC samples per chirp	256
Number of Chirp loops per Frame	128
Total number of frames	700
Total number of chirps	89,600
Total time	28s

following the process proposed in [32]. For this, a whitening of 2D vector is done by multiplying the zero-centered data with the inverse square root of the covariance matrix of real and imaginary data components.

IV. EXPERIMENTAL SETUP AND DATASET

A. Experimental Setup

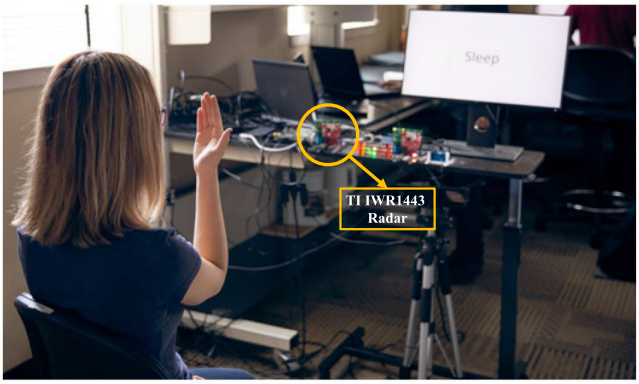
For RF data collection, a 77GHz TI IWR1443 automotive short-range radar has been used. The radar has three transmitters and four receivers in total. However, in this experiment, the data from only one transmitter and one receiver have been used. The radar system parameters selected for the data collection are given in Table I.

For the experimental setup, the radar was positioned on top of a table placed against a wall in a lab environment at 0.91 meters of height. ASL signers were sitting on a chair in front of the radar at a distance of 1.5m. A computer monitor was placed exactly behind the radar so that it is outside the radar field of view (FOV). The monitor continuously guides the participants about the signs that are needed to be articulated. The experimental setup is shown in Fig. 4a.

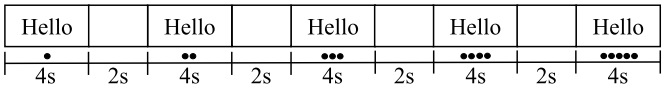
B. Dataset

Five people including three professional ASL signers, two deafs, one CODA (Child of Deaf Adults), and two lab members participated in conducting the IRB-approved data collection. 100 different ASL signs were selected from the database ASL-LEX2 [39] for this experiment. In order to make a diverse dataset, high-frequency signs which are not phonologically related to each other were chosen. To evaluate the performance of proposed CV-SincNet as compared to other approaches, this challenging 100-class ASL radar dataset is used. The corresponding English words for the utilized 100 ASL signs are shown in Table II. More information about the utilized dataset can be found in [34].

The dataset consists of a total of 2500 radar sign samples from five participants having 25 samples for each class. As illustrated in Fig. 4b, five repetitions of each sign were collected from each participant with 4s durations each followed by an inter-stimulus interval of 2s. In total, the data



(a) Experimental setup for data collection with radar [17]



(b) Example of sequential prompts given to user.

Fig. 4. Experimental setup and timing of sign articulation.

was collected for 28s for each class from each participant. Number of chirps captured during this 28s period was 89,600. Thus, the number of chirps collected during one sample of data (4s) is 12,800. Because of some small inconsistencies between the start and finish time of each sample, additional 125 chirps from both sides were also taken into consideration. For each chirp 256 fast-time samples are acquired, making the size of the 2D complex-valued raw radar data 256×13050 for a single sign. After range processing, the range bins corresponding to the target location are summed, forming a 1D slow-time complex raw radar data vector of 1×13050, which can be used for computation of the μ-D signature or as the input to a deep neural network. The proposed complex SincNet architecture uses this 1D complex slow-time radar data after range processing.

V. PERFORMANCE ANALYSIS OF THE PROPOSED METHOD

To evaluate the performance of the proposed CV-SincNet architecture, comparisons have been made with several current state-of-the-art techniques:

- CNN-2D: Conventional 2D real convolutional neural network taking 2D spectrogram images as their input.
- CNN-1D: A real 1D convolutional neural network taking raw radar data as real and imaginary in two channels.
- CVCNN-1D: A complex 1D convolutional neural network directly working on complex raw radar data.
- SincNet: The real SincNet applied on raw radar data as real and imaginary in two channels.
- CV-SincNet: The proposed complex SincNet architecture working directly on the complex raw radar data.

The next sections will discuss the details of each architecture and their training process. For analysis, the dataset was randomly split into a single split of 80% and 20% training and testing datasets respectively. The same set of training and testing datasets have been used for all the architectures compared.

TABLE II
DATASET CONSISTING OF 100 ASL SIGNS

100 Class ASL Dataset					
YOU	YES	ME	HOME	HOLD	FATHER
MY	MORNING	THREE	WRONG	TOILET	THERE
LONG	I lOVE YOU	DEAF	SLEEP	THANK YOU	TIRED
HELLO	OK	THIS	GOOD	MUST	HE
TIME	BETTER	TOMORROW	WHY	LIKE	YOUR
ONE	DON'T LIKE	FINE	SOMETHING	MOTHER	SEE
HOT	BREAKFAST	WATER	EAT	OH I SEE	LET ME SEE
SOON	WHERE	PLEASE	SHOULD	ALWAYS	TABLE
BOOK	MORE	BED	HELP	HAVE	CITY
GO AHEAD	SUMMON	LICENSE	THRILLED	WANT	WELL
FRIEND	READ	CHANCE	READY	BRING	PET
MONTH	GAS	AGAIN	WEEK	GO	NIGHT
TIE UP	CAN	RIGHT	FAMILY	KITCHEN	WINTER
WORK	TEACH	CAR	EVENING	EXPLANATION	PAPER
WHAT	TODAY	SSCHOOL	COFFEE	NOTHING	SHOP
TECHNOLOGY	WALK	COOK	SHOES	TEACHER	MAYBE
DOESN'T MATTER	EXCITED	MONEY	PEOPLE		

A. Detailed Descriptions of DNNs Compared

1) CNN-2D Architecture: Using Eq. (3), μ-D signatures are generated and saved as 128×128 images, which are then supplied as input to a 2D CNN. Based on earlier work [8], a CNN architecture of four convolutional layers with 32, 32, 64, and 64 filters, respectively, was utilized. Each convolutional filter uses a kernel size of 5×5, which is then followed by 2×2 maxpooling, batch normalization, ReLU activation function, and dropout of 0.3. After the convolutional blocks, the tensor is flattened and fed into a dense layer of size 256×1, after which dropout of 0.5 is applied and input to a softmax classifier. The CNN-2D architecture has a total of 469,060 parameters.

2) CNN-1D and SincNet Architectures: Two real networks, a 1D CNN and SincNet, were utilized to process the 1D complex raw radar data. Both real networks separate the raw complex RF data stream into its real and imaginary parts, forming a real input representation of size 13050×2. The first layer of SincNet is comprised of 256 real sinc filters, each with 251 data points. This sinc layer is then followed by a maxpooling layer of size 1×3, batch normalization, ReLU activation, and dropout of 0.2. This sequence of layers comprises the sinc block, which is shown in Fig. 2. The sinc block is followed by three parallel and five series convolutional blocks, as well as two dense layers. The kernel sizes of the parallel convolutional blocks were selected as 1×3, 1×5, and 1×7 to enable multi-resolution feature extraction, yielding more information and making the model more stable, while also enhancing the resulting accuracy. The kernel sizes for the series convolutional blocks were fixed at a size of 1×5. The outputs of the convolutional blocks are flattened before input to a dense layer with a size of 256. After application of dropout of 0.3, the network uses a softmax layer for classification.

In lieu of the sinc blocks, the CNN-1D architecture utilizes a convolutional block as its first layer, after which its architecture is the same as that of SincNet. Details about the architectures and a layer-by-layer comparison of the number of parameters for each network are given in Table III.

TABLE III
MODEL SIZES OF CVCNN-1D, CNN-1D, SINCNET AND CV-SINCNET

Blocks (Num. of filters, Kernel Size, Maxpool)	Parameters			
	CVCNN-1D	CNN-1D	SincNet	CV-SincNet
Sinc(256,Nan,3)			1024	1536
Conv 1 (256, 5, 3)	4096	3840		
Parallel Convs 2,3 & 4 (64, (3,5,7), 3)	493440	246720	246720	493440
Conv 5 (64, 5, 3)	123520	61760	61760	123520
Conv 6 (64, 5, 3)	41600	20800	20800	41600
Conv 7 (64, 5, 3)	41600	20800	20800	41600
Conv 8 (64, 5, 3)	41600	20800	20800	41600
Conv 9 (64, 5, 3)	41600	20800	20800	41600
Conv 10 (64, 5, 2)	41600			41600
Dense 1 (256)	66048	49408	49408	66048
Softmax (100)	51400	25700	25700	51400
Total Parameters	947,528	469,060	466,244	943,944

3) CVCNN-1D and CV-SincNet Architectures: In contrast with the real networks, the complex-valued networks, CVCNN-1D and proposed CV-SincNet, directly take as input the raw complex RF data. The principle differences between the complex DNNs and their real counterparts are that 1) complex convolutions, complex max-pooling, complex batch normalization, complex ReLU and complex dropout blocks replace their real counterparts; and 2) The complex sinc block in CV-SincNet have been completely redesigned to be one-sided, complex bandpass filters that separately focus on positive or negative Doppler frequencies. While the CVCNN-1D employs 10 convolutional blocks, in CV-SincNet the first convolutional block is replaced with a sinc block, while the remainder of the architecture is the same.

For better visualization, the flow diagram of the full CV-SincNet architecture is shown in Fig. 3. Detailed information about the complex networks and their parameters are given in Table III as well.

TABLE IV
PERFORMANCE OF THE COMPARED MODELS IN TERMS OF EVALUATION METRICS

Network	Input	Testing Accuracy	Precision	Recall	F1 Score
CNN-2D	2D μ-Ds	60.97	65.47	60.65	59.5
CNN-1D	1D slow-time raw radar data	17.51	18.56	17.55	16.36
CVCNN-1D		23.21	21.73	23.20	21.61
SincNet		57.05	62.96	57.55	56.73
CV-SincNet		65.19	73.29	65.45	64.95

TABLE V
TESTING ACCURACY FOR TOP-1,3 AND-5 FOR COMPARED APPROACHES

Network	Testing Accuracy		
	Top 1	Top 3	Top 5
CNN-2D (on μ-Ds)	60.97	81.01	87.76
CNN-1D	17.51	30.38	38.40
CVCNN-1D	23.21	35.44	41.98
SincNet	57.05	75.69	82.69
CV-SincNet	65.19	81.86	88.82

B. Performance Comparison

1) Performance for 100 Class ASL Data: For performance evaluation, the entire dataset was split into 80% training and 20% testing as mentioned earlier in this section. Each model has been trained for a long enough time to reach convergence. All of the compared models are trained for 150 epochs. Afterward, the trained models are tested on the same test data and confusion matrices were generated for each model. Testing accuracy, precision, recall and F1 scores were evaluated from the confusion matrices and shown in Table IV.

The proposed CV-SincNet architecture outperforms the compared approaches in all the metrics. First, among the methods that utilize radar raw data, SincNet and CV-SincNet significantly outperform the convolutional networks, CNN-1D and CVCNN-1D. The use of sinc filters as the initial layer of the network and learning frequency domain information has improved the performance as compared to trying to learn random convolutional filters directly from the raw radar data. Second, the proposed CV-SincNet performs higher in all evaluation metrics as compared to the real SincNet. This signifies the importance of complex network computational and learning complex Sinc filters that are physically more relevant for radar data. Utilizing complex networks also helps the performance of 1D CNN architectures since raw radar data is naturally complex. Third, the proposed CV-SincNet achieves improved accuracy relative to that obtained using a CNN-2D, a real-valued network that takes as input μ-D signatures after pre-processing with a time-frequency transform. This approach is one of the most applied approaches in the literature [8] and consistently provides high performance because target kinematics are explicitly revealed by the μ-D. Thus, it is significant that CV-SincNet can provide comparable or even slightly better results without the computational burden of explicitly computing the μ-D signature.

For further assessment, top-3 and top-5 accuracies were also computed and shown in Table V. Top-N accuracy indicates whether the model is able to predict the expected class within the top-N highest probabilities. Since the dataset consists of

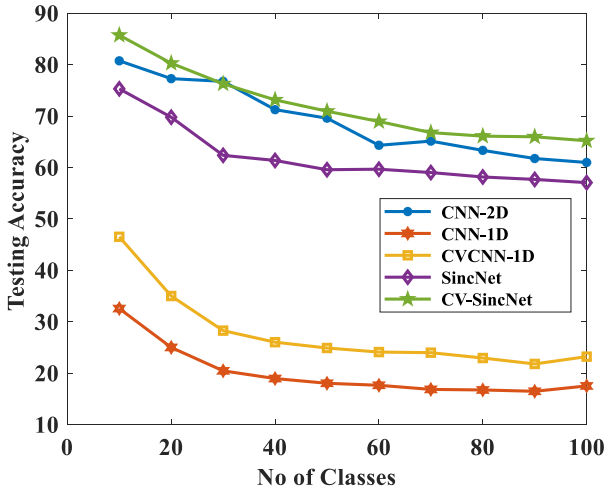


Fig. 5. Testing accuracy for all networks with respect to varying number of classes.

100 classes, Top-N accuracy measurement is an important performance-evaluating factor to consider. The top-3 and top-5 accuracies of CV-SincNet also surpass that of other methods, although the performance gap between CV-SincNet and CNN-2D is much lower in these two cases. While CNN-1D and CVCNN-1D are still showing poor performance both models have around 40% accuracy while estimating the top-5 case. Both complex-valued networks yielded better performance than their real counterparts.

2) Performance for Varying Number of Classes: Radar-based activity recognition applications may require the recognition of a varying number of classes. To test the effectiveness of the proposed approach as a function of the number of classes, the dataset was divided into subsets by varying the number of classes from 10 to 100 with an interval of 10 between each of them. Each subset of data was then divided into 80% training and 20% testing. The same network architectures were used to train the 10 datasets separately. Figure 5 shows the testing accuracy results for each network as a function of the number of classes. While all models followed a downward trend in performance when the number of classes increased, the CV-SincNet has the highest accuracy for almost all cases. The CV-SincNet yielded 85.72% and 65.19% testing accuracy, whereas the 2nd best model, CNN-2D, yielded 80.74% and 60.94% accuracy for 10 and 100 classes, respectively. CV-SincNet also performed on average 10% higher accuracy as compared to real SincNet. The CVCNN-1D and CNN-1D models' performances were much lower than the rest of the networks for all cases. While the complex model CVCNN-1D showed approximately 8% higher performance than the CNN-1D, the significant improvement of utilizing sinc filters in lieu of convolutional layers is a testament to the significance of frequency-domain information and the efficiency of learning frequency-domain representations using the proposed approach.

3) Comparison of Computational Latency: A key metric that is as critical as classification accuracy when considering real-time classification applications, is the total computational latency for classification. The overall latency is the time required starting with data acquisition to generate a prediction.

TABLE VI
COMPUTATIONAL TIME REQUIRED FOR EACH NETWORK

Network	Total training time	Pre-processing time	Avg. pre-diction time	Total la-tency
CNN-2D (on μ-Ds)	8 min	2601 ms	29 ms	2630 ms
CNN-1D	33 min	377 ms	179 ms	556 ms
CVCNN-1D	42 min	377 ms	216 ms	603 ms
SincNet	2.25 hrs	377 ms	338 ms	715 ms
CV-SincNet	2.68 hrs	377 ms	379 ms	756 ms

Methods with low latency and high performance are desired to achieve near real-time human activity recognition. To compare the computational requirements of each approach, we observed the average pre-processing, training, and prediction times, which are listed in Table VI.

The pre-processing time is the duration for one data sample from data acquisition to be ready as an input to the deep neural network. First, the 2D fast-time slow-time data matrix is obtained from the corresponding TX-RX pair. Next, range processing is applied to the fast-time samples and 1D slow-time data is obtained through summing the range bins for each slow-time sample. This 1D data is the input for all networks other than CNN-2D. For CNN-2D, an additional short-time Fourier transform (STFT) operation is needed over the slow-time data to generate the μ-D spectrogram images. The computation of an STFT requires multiple windowed Fourier transforms in order to generate the spectrogram output. This is the reason why the pre-processing of the data takes much longer for CNN-2D as compared to the other 1D approaches. Next, each network is trained over the training dataset and after training the average prediction time represents the time for a trained network to generate the prediction result when given an input data sample. Both the data processing and network training have been done using an Alienware m15 R7 laptop with an NVIDIA 3060 GPU, Intel 11th Gen CPU, and 32 GB memory.

While training times for SincNets are higher than CNN-1D or CNN-2D approaches, since training is done only once, the total latency time, which is the sum of pre-processing and average prediction time, is the important metric for the real-time suitability of the approaches. For the total latency, all 1D raw data processing networks have much lower latency values compared to CNN-2D. A method is desired to have both low latency and high accuracy performance. Figure 6 is a visual representation of testing accuracy and prediction time for each network. It can be seen that real SincNet and the proposed CV-SincNet provide both high accuracy and low latency, while the proposed CV-SincNet provides the best performance for both considerations. While CNN-2D provides high accuracy, it also has a very long latency, which precludes it from being used in real-time applications. CNN-1D and CVCNN-1D have low latency, but also a very low accuracy, also making it an untenable approach. The performance of the proposed CV-SincNet shows that it is a good solution for near real-time and high-accuracy radar-based activity recognition problems.

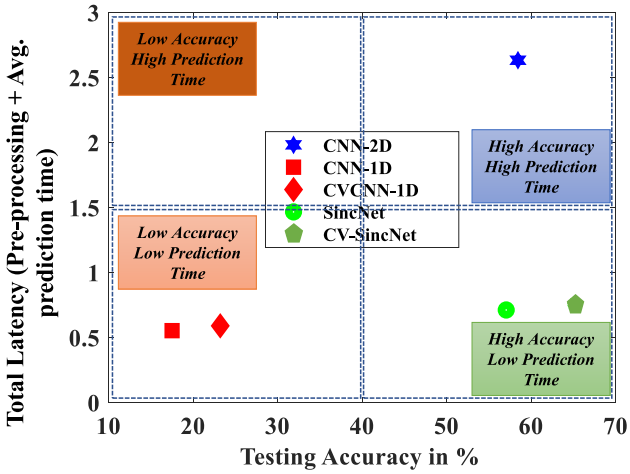


Fig. 6. Testing accuracy vs prediction time.

TABLE VII
CV-SINCNET PERFORMANCE V. # SINC FILTERS

Num of fil-ters	Num of parame-ters	Training time	Testing ac-curacy	Avg. prediction time
8	466,296	0.35 hrs	30.97%	83 ms
16	481,704	0.37 hrs	47.68%	100 ms
32	512,520	0.45 hrs	61.18%	111 ms
64	575,152	0.58 hrs	64.98%	167 ms
256	943,944	2.68 hrs	65.19%	379 ms

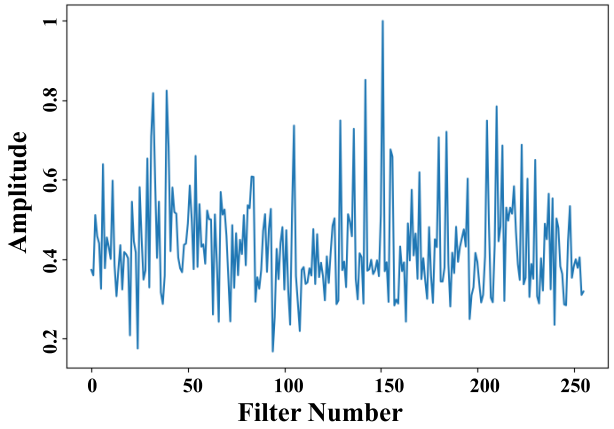


Fig. 7. Weights of each sinc filter.

We further analyze the proposed CV-SincNet in terms of the effect of the number of sinc filters on both accuracy and latency metrics. The model is trained with sinc layers with a varying number of sinc filters from 8 to 256 and the achieved testing accuracy and average prediction times for each case are shown in Table VII. Using fewer sinc filters reduces prediction times; but too few filters also reduces the accuracy. Using 64 filters, the average prediction time was reduced by about 56% while maintaining almost the same performance level as using 256 filters.

C. CV-SincNet Interpretability

An important property of the proposed CV-SincNet is its enhanced interpretability. Each sinc filter only depends on

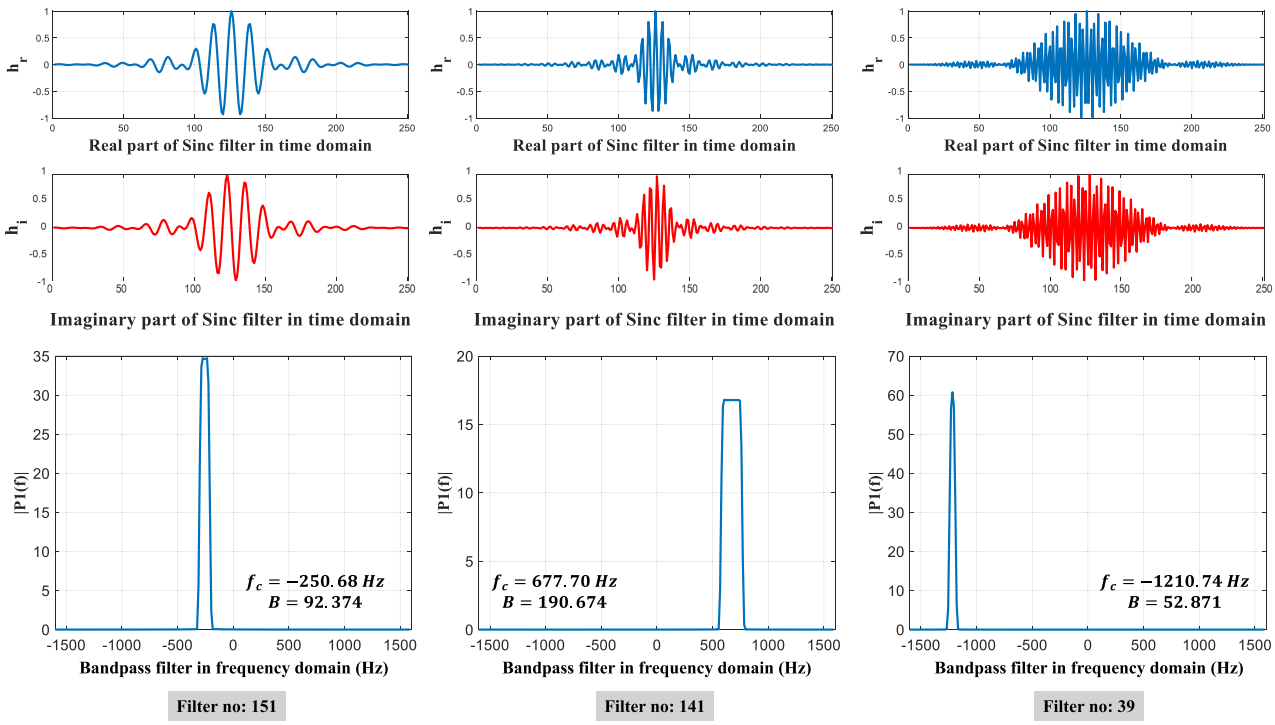


Fig. 8. Three most important sinc filters based on weights.

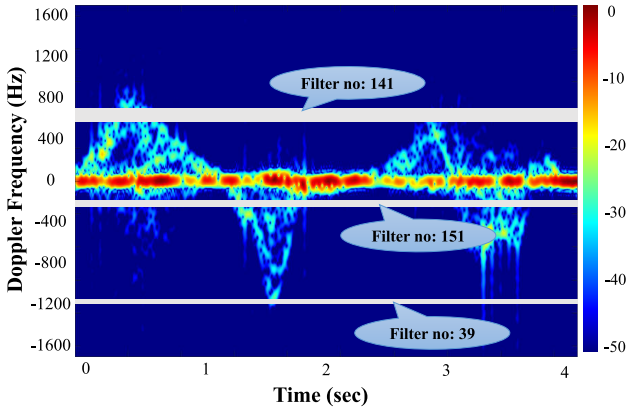


Fig. 9. Three most important filters on a μ-D spectrogram.

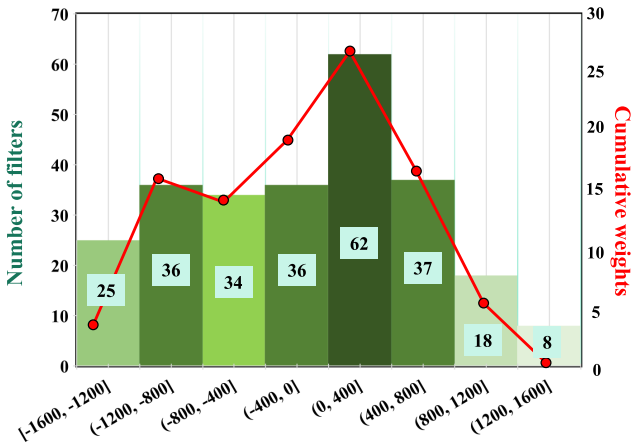


Fig. 10. Filter and weight distribution in frequency domain of the learned complex sinc filters.

parameters with a clear physical meaning, such as Doppler frequencies and bandwidths. In order to understand which frequency bands are spanned by the learned filters, the cumulative weights of the three parallel convolutional filters subsequent to the sinc blocks have been calculated. The normalized magnitude of the weights corresponding to each of the 256 sinc filters is shown in Figure 7. It may be observed that the model gives higher importance to some of the learned filters. Figure 8 indicates the time and frequency domain representation of the top three learned sinc filters based on their calculated weights. The corresponding center frequencies and bandwidths for these filters are also illustrated. For better visualization, these three filters were projected over a μ-D spectrogram image as shown in Fig. 9. It is interesting to see that two of the top learned filters mainly focus on the lower and upper Doppler frequencies. The third filter learns a band around -250 Hz, near zero Doppler.

Furthermore, insights about sign language can be obtained from results revealed by plotting the histogram and the cumulative weight distribution of the learned sinc filters in the frequency domain, as shown in Figure 10. It may be observed that 98 out the 256 learned filters lie between -400 Hz and 400 Hz. This shows that the model focuses on this part of the Doppler spectrum by using more filters and weights, while some of the top learned filters focus on upper and lower envelope frequencies. While the model doesn't give importance to some frequency regions such as [1200,1600), where not much of the activity response resides, other bands such as (-1200 to -800] or [400,800) are utilized with a greater number of learned filters and corresponding weights. This also makes sense since intervals span frequency bands containing much of the higher-frequency micro-Doppler

components, which correspond to the movement of the hands during sign language articulation. In fact, it is interesting to note that the distribution of filters across frequency is asymmetric, with a larger number of filters spanning positive frequencies rather than zero or negative frequencies. This is not a coincidence. The signs comprising our database of 100 signs are more kinetic in nature, with intentional inclusion of a distribution of signs articulated using 2, 3, or 4 strokes of movement. Signs that start on the head or the body are more likely to embody forward movements. Such movements toward the radar result in positive Doppler frequencies observed by the radar, which accounts for the increased number of filters learned over positive frequencies.

VI. Conclusion and Future Work

In this work, we introduced complex-valued SincNet (CV-SincNet), where the network learns more interpretable filters and directly works on complex raw radar data. In order to make our model efficient and compatible with the raw complex radar data, we designed a complex sinc layer, where the model learns sinc filters having a center frequency and bandwidth as trainable parameters for each filter. The remaining aspects of the CV-SincNet architecture utilized complex versions of common DNN blocks, such as convolution, maxpooling, and batch normalization, ReLU and dropout. Furthermore, the efficacy of the proposed CV-SincNet architecture was demonstrated on a challenging 100-sign dataset using complex radar data as input. The results of our current approach show promise for ubiquitous, real-time RF sensing applications. Not only does the CV-SincNet architecture provide better accuracy result than the widely used state-of-the-art CNN-2D that works on μ-D spectrogram images or other CNN variants working on 1D radar data, but it also provides much lower latency results showing a high potential for near real-time human activity recognition. The learned sinc filters depend only on meaningful physical parameters, such as the center Doppler frequency and bandwidths; hence, they are more interpretable. Future work will focus on embedding the proposed models on edge computing platforms to demonstrate RF-sensing based cyber-physical human system applications.

References

[1] S. Gurbuz, *Deep Neural Network Design for Radar Applications*. London, U.K.: IET, 2020.

[2] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the NetRAD multistatic radar for micro-Doppler and singular value decomposition features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1933–1937, Sep. 2015.

[3] Z. Ni and B. Huang, "Gait-based person identification and intruder detection using mm-wave sensing in multi-person scenario," *IEEE Sensors J.*, vol. 22, no. 10, pp. 9713–9723, May 2022.

[4] S. Björklund, T. Johansson, and H. Petersson, "Target classification in perimeter protection with a micro-Doppler radar," in *Proc. 17th Int. Radar Symp. (IRS)*, May 2016, pp. 1–5.

[5] A. Huizing, M. Heiligers, B. Dekker, J. de Wit, L. Cifola, and R. Harmanny, "Deep learning for classification of mini-UAVs using micro-Doppler spectrograms in cognitive radar," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 11, pp. 46–56, Nov. 2019.

[6] G. Hakobyan and B. Yang, "High-performance automotive radar: A review of signal processing algorithms and modulation schemes," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 32–44, Sep. 2019.

[7] S. Biswas, B. Bartlett, J. E. Ball, and A. C. Gurbuz, "Classification of traffic signaling motion in automotive applications using FMCW radar," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2023, pp. 1–6.

[8] 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.

[9] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 71–80, Mar. 2016.

[10] F. Fioranelli and J. L. Kernec, "Contactless radar sensing for health monitoring," in *Engineering and Technology for Healthcare*. 2021, pp. 29–59.

[11] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.

[12] S. Gurbuz, A. Gurbuz, C. Crawford, and D. Griffin, "Radar-based methods and apparatus for communication and interpretation of sign languages," U.S. Patent. 11 301 672 B2, Mar. 16, 2021.

[13] S. Z. Gurbuz et al., "American sign language recognition using RF sensing," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3763–3775, Feb. 2021.

[14] J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, p. 142, Jul. 2016.

[15] V. Chen, *The Micro-Doppler Effect Radar*, 2nd ed. Norwood, MA, USA: Artech House, 2019.

[16] 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.

[17] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "ASL trigger recognition in mixed activity/signing sequences for RF sensor-based user interfaces," *IEEE Trans. Hum.-Mach. Syst.*, vol. 52, no. 4, pp. 699–712, Aug. 2022.

[18] V. Chen, "Analysis of micro-Doppler signatures," *IEE Proc.-Radar, Sonar Navigat.*, vol. 150, no. 5, pp. 271–276, Aug. 2003.

[19] T. Thayaparan, S. Abrol, E. Riseborough, L. Stankovic, D. Lamothe, and G. Duff, "Analysis of radar micro-Doppler signatures from experimental helicopter and human data," *IET Radar, Sonar Navigat.*, vol. 1, no. 4, pp. 289–299, Aug. 2007.

[20] F. H. C. Tivive, S. L. Phung, and A. Bouzerdoum, "Classification of micro-Doppler signatures of human motions using log-Gabor filters," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1188–1195, Dec. 2015.

[21] 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.

[22] W. Taylor, K. Dashtipour, S. A. Shah, A. Hussain, Q. H. Abbasi, and M. A. Imran, "Radar sensing for activity classification in elderly people exploiting micro-Doppler signatures using machine learning," *Sensors*, vol. 21, no. 11, p. 3881, Jun. 2021.

[23] M. Chakraborty, H. C. Kumawat, S. V. Dhavale, and A. A. B. Raj, "Application of DNN for radar micro-Doppler signature-based human suspicious activity recognition," *Pattern Recognit. Lett.*, vol. 162, pp. 1–6, Oct. 2022.

[24] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar Navigat.*, vol. 11, no. 1, pp. 107–115, Jan. 2017.

[25] S. Yang, J. L. Kernec, F. Fioranelli, and O. Romain, "Human activities classification in a complex space using raw radar data," in *Proc. Int. Radar Conf. (RADAR)*, 2019, pp. 1–4.

[26] C. Loukas, F. Fioranelli, J. Le Kernec, and S. Yang, "Activity classification using raw range and I & Q radar data with long short term memory layers," in *Proc. IEEE 16th Int. Conf. Dependable, Autonomic Secure Comput.*, Aug. 2018, pp. 441–445.

[27] T. Stadelmayer, A. Santra, R. Weigel, and F. Lurz, "Data-driven radar processing using a parametric convolutional neural network for human activity classification," *IEEE Sensors J.*, vol. 21, no. 17, pp. 19529–19540, Sep. 2021.

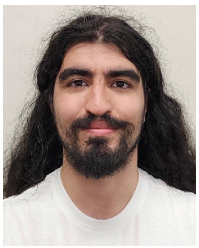
[28] X. Yao, X. Shi, and F. Zhou, "Human activities classification based on complex-value convolutional neural network," *IEEE Sensors J.*, vol. 20, no. 13, pp. 7169–7180, Jul. 2020.

[29] D. A. Brooks, O. Schwander, F. Barbaresco, J.-Y. Schneider, and M. Cord, "Complex-valued neural networks for fully-temporal micro-Doppler classification," in *Proc. 20th Int. Radar Symp. (IRS)*, 2019, pp. 1–10.

[30] X. Yang, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Radar-based human activities classification with complex-valued neural networks," in *Proc. IEEE Radar Conf. (RadarConf)*, Mar. 2022, pp. 1–6.

[31] 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.
[32] C. Trabelsi et al., "Deep complex networks," 2017, *arXiv:1705.09792*.
[33] S. Biswas, C. O. Ayna, S. Z. Gurbuz, and A. C. Gurbuz, "Complex SincNet for more interpretable radar based activity recognition," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2023, pp. 1–6.
[34] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for ASL recognition with RF sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2732–2745, Aug. 2022.
[35] J.-J. Lin, Y.-P. Li, W.-C. Hsu, and T.-S. Lee, "Design of an FMCW radar baseband signal processing system for automotive application," *SpringerPlus*, vol. 5, no. 1, pp. 1–16, Dec. 2016.
[36] V. C. Chen, D. Tahmoush, and W. J. Miceli, *Radar Micro-Doppler Signatures*. London, U.K.: Institution of Engineering and Technology, 2014.
[37] A. T. Basokur, "Designing frequency selective filters via the use of hyperbolic tangent functions," *Yerbilimleri*, vol. 32, pp. 69–88, Jan. 2011.
[38] C. Lee, H. Hasegawa, and S. Gao, "Complex-valued neural networks: A comprehensive survey," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1406–1426, Aug. 2022.
[39] N. K. Caselli, Z. S. Sehyr, A. M. Cohen-Goldberg, and K. Emmorey, "ASL-LEX: A lexical database of American sign language," *Behav. Res. Methods*, vol. 49, no. 2, pp. 784–801, Apr. 2017.

Sabyasachi Biswas (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2019. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Mississippi State University, Mississippi State, MS, USA. He is a Research Assistant with the Information Processing and Sensing (IMPRESS) Laboratory. His research interests include radar signal processing, human activity recognition using radar, camera and lidar, and developing machine learning algorithms for activity classification using raw radar signals. He is a member of the IEEE Signal Processing Society. He was the Winner of the Graduate Research Symposium held in Mississippi State University in Fall 2022.



Cemre Omer Ayna (Student Member, IEEE) received the B.S. degree in electrical and electronics engineering from Koç University, Istanbul, Turkey, in 2021. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Mississippi State University, Mississippi State, MS, USA. He is a Research Assistant with the Information Processing and Sensing (IMPRESS) Laboratory, Mississippi State University. He is involved in image reconstruction and demosaicing, compressive sensing, image segmentation, hyperspectral image processing, RADAR signal processing, model-informed machine learning, and adversarial machine learning. His main research interests include machine learning models for measurement learning and signal compression problems.

Sevgi Z. Gurbuz (Senior Member, IEEE) received the B.S. degree in electrical engineering with a minor in mechanical engineering and the M.Eng. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1998 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2009. From February 2000 to January 2004, she was a Radar Signal Processing Research Engineer with the U.S. Air Force Research Laboratory, Sensors Directorate, Rome, NY, USA. She was an Assistant Professor with the Department of Electrical and Electronics Engineering, TOBB University, Ankara, Turkey, and a Senior Research Scientist with the TUBITAK Space Technologies Research Institute, Ankara. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA. She has recently received a patent in April 2022 relating to radar-based American Sign Language (ASL) recognition. Her current research interests include RF sensor-enabled cyber-physical systems, radar signal processing, physics-aware machine learning, sensor networks, human motion recognition for biomedical, automotive autonomy, and human–computer interaction (HCI) applications. She is a member of the SPIE and the Association for Computing Machinery (ACM). She also serves as a member of the IEEE Radar Systems Panel. She was a recipient of the 2023 NSF CAREER Award, the 2022 American Association of University Women Research Publication Grant in Engineering, Medicine and Science, the IEEE Harry Rowe Mimno Award for the Best IEEE AES Magazine Paper of 2019, the 2020 SPIE Rising Researcher Award, an EU Marie Curie Research Fellowship, and the 2010 IEEE Radar Conference Best Student Paper Award. She is an Associate Editor of the IEEE TRANSACTIONS OF AEROSPACE AND ELECTRONIC SYSTEMS (T-AES) and the IEEE TRANSACTIONS ON RADAR SYSTEMS (T-RS). She is an Editorial Board Member of the *IET Radar, Sonar, and Navigation* (RSN) journal.

Ali C. Gurbuz (Senior Member, IEEE) received B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 2003, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2005 and 2008, respectively. In 2009, he was a Post-Doctoral Researcher with Georgia Tech, where he researched compressive sensing-based computational imaging problems. From 2009 to 2017, he was a Faculty Member with TOBB University and The University of Alabama, where he pursued an active research program on the development of sparse signal representations, compressive sensing theory and applications, radar and sensor array signal processing, and machine learning. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Mississippi State University, where he is the Co-Director of the Information Processing and Sensing (IMPRESS) Laboratory. He was a recipient of the Best Paper Award for *Signal Processing* Journal in 2013, the Turkish Academy of Sciences Best Young Scholar Award in Electrical Engineering in 2014, and the NSF CAREER Award in 2021. He has served as an Associate Editor for several journals, such as *Digital Signal Processing*, *EURASIP Journal on Advances in Signal Processing*, and *Physical Communication*.