Deep Learning-based High-Resolution Radar Micro-Doppler Signature Reconstruction for Enhanced Activity Recognition

Sabyasachi Biswas, Ahmed Manavi Alam, and Ali C. Gurbuz
Dept. of Electrical and Computer Engineering, Mississippi State University, MS, USA
{sb3682, aa2863}@msstate.edu, gurbuz@ece.msstate.edu

2024 IEEE Radar Conference (RadarConf24) | 979-8-3503-2920-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/RADARCONF2458775.2024.10548969

***Abstract*—Micro-Doppler signatures (μ-DS) play a crucial role in activity classification using radar. However, conventional methods for μ-DS generation, such as the Short-Time Fourier Transform (STFT), suffer from several limitations, such as the resolution limit, sensitivity to noise, and the need for parameter tuning. To overcome these challenges, we introduce a novel deep learning (DL) based approach that directly reconstructs high-resolution μ-DS from 1D complex time-domain signals. Our deep learning architecture comprises three key components: an autoencoder block to enhance the signal-to-noise ratio (SNR), a Convolutional STFT block to acquire the knowledge of frequency transformations necessary for generating pseudo-spectrograms, and a UNET block for the reconstruction of high-resolution spectrogram images. We conducted evaluations of the proposed method using both synthetic and real-world datasets. In the case of synthetic data, we generated 1D complex time-domain signals with multiple time-varying frequencies and assessed the network's performance in generating high-resolution μ-DS under different SNR levels. For real-world data, A radar-based American Sign Language (ASL) dataset, consisting of 20 ASL signs are used, to assess the classification performance achieved with μ-DS generated by the proposed approach. Our results demonstrated a 3.34% increase in classification accuracy compared to traditional STFT-based μ-DS. Both synthetic and experimental μ-DS revealed that our approach effectively learns to reconstruct higher-resolution and sparser spectrograms, showcasing its potential for improving radar-based activity recognition applications.**

***Index Terms*—Radar, STFT, micro-Doppler signature, Autoencoder, U-Net, HAR, time-frequency analysis.**

I. Introduction

Recent advances in affordable solid-state transceivers, efficient graphics processing units (GPUs), and deep learning techniques have expanded the practicality of radio frequency (RF) sensors for applications such as human activity recognition (HAR) [1]–[3], defense and security [4], [5], mini-UAV classification [6], advanced driver assistance systems (ADAS) [7], anomaly detection [8], indoor monitoring [9] and health monitoring [10]. These advancements have opened up new possibilities for integrating RF sensors into an expanding range of practical HAR applications.

In the context of radar-based HAR, time-frequency (TF) analysis is crucial for capturing time-varying kinematic information about dynamic activities, facilitating accurate classification [11]. While some machine learning (ML) approaches can classify activities directly from complex RF data [12]–[15], traditional radar-based HAR methods typically involve a two-step process. First, TF analysis or other radar signal processing techniques are used to generate 2D (or higher-dimensional) radar data representations, like time-varying range-Doppler maps or μ-DS [16]. Although some studies have explored joint domain classification [17], most studies utilize μ-DS for HAR [18]–[20].

The Short-Time Fourier Transform (STFT) is a commonly used method for TF analysis, decomposing signals into constituent frequencies at different time windows, highlighting time-varying characteristics. However, STFT has trade-offs and limitations:

- Parameter Tuning: Selecting the right window length, overlap size, and number of frequency bins can be challenging, as these parameters must be adjusted based on signal characteristics. This tuning can be time-consuming and may not yield optimal results for different signal sizes and types.
- Resolution Limits: STFT has inherent resolution limits due to the fixed window size. This means it may not capture fine details in signals with rapidly changing frequencies.
- Frequency Resolution vs. Time Resolution: A trade-off exists between frequency and time resolution. Smaller window sizes provide better time resolution but poorer frequency resolution, while larger windows offer better frequency but poorer time resolution.

To address these limitations, there is a growing interest in developing DL methods for high-resolution frequency estimation and TF representation. Notable approaches include Deepfreq, HRFreqNet are DNN architectures for multi-sinusoidal complex-valued signal frequency estimation [21], [22], and Cresfreq, a complex-valued neural network (CVNN) for 1D frequency representation of multi-sinusoidal signals [23]. These methods can also be used to create 2D TF representations, similar to STFT's use of the Fast Fourier Transform (FFT). However, these DL methods have drawbacks, including computational inefficiency, fixed signal length, and sensitivity to noise, limiting their performance at lower signal-to-noise ratios (SNR). To address these challenges, TFA-net, another CVNN architecture, was introduced to generate high-resolution TF representations for complex multi-sinusoidal signals [24]. While TFA-net provides high-resolution TF rep-

resentations, it's computationally very intensive, especially for high-length signals, which are common in radar data due to high sampling rates and also susceptible to noise as it is trained with only clean dataset.

Expanding on the insights inspired by Deepfreq, Cresfreq, and TFA-net, this paper presents HRSpecNet, a novel end-to-end deep neural network (DNN) architecture designed to generate high-resolution 2D TF representations from 1D complex-valued signals. The proposed approach offers superior computational efficiency and enhanced noise robustness. HRSpecNet includes a convolutional auto-encoder for noise reduction, a Convolutional STFT block for intermediate frequency representation, and a UNET block for high-resolution TF generation. With the help of autoencoder block and UNET block, the HRSpecNet can effectively suppress noise, resulting in cleaner TF representations.

The contributions of this paper can be summarized as follows:

- Development of an end-to-end DNN architecture with a weighted loss function to ensure both better noise suppression and precise TF representation generation.
- Investigation of HRSpecNet's advantages over traditional STFT, highlighting its high-quality TF representations without extensive parameter tuning.
- Demonstration of HRSpecNet's generalization capability on real-life radar data, showing higher accuracy in classifying μD spectrograms compared to other methods.

The paper is organized as follows: the RF signal model and existing techniques are summarized in Section II, discussion of HRSpecNet in Section III, performance analysis in Section IV, and conclusion and discussion of future directions in Section V.

II. Theoretical Background

A. Radar Signal Model

Frequency-modulated continuous wave (FMCW) radar systems are commonly used for measuring both the range and velocity of objects. In such systems, the transmitted signal can be a linearly swept RF signal, where the instantaneous frequency $f_i(t)$ can be stated as,

$$f_i(t) = f_0 + \frac{B}{\tau}t, \quad 0 \le t \le \tau, \tag{1}$$

where, f_0 is the initial frequency at time $t = 0$, B is the bandwidth, and τ represents the sweep time in seconds. When this signal is reflected back from a target with a time delay T_d and mixed with a copy of the transmitted signal, it passes through a low-pass filter (LPF) to obtain the Intermediate Frequency (IF) signal. The IF signal can be modeled as:

$$s_{\text{IF}}(t) = A\exp\left(2\pi(f_0T_d + \frac{B}{\tau}T_dt - \frac{B}{2\tau}T_d^2)\right), \tag{2}$$

where, A represents the amplitude of the received signal. The received data is then sampled, and the FMCW radar stores this data in a three-dimensional (3D) radar data cube (RDC) of size $P \times N \times M$, where P is the number of fast-time samples, N is the number of slow-time samples, and M is the number of receiver channels. Various RF data representations can be derived from the RDC, such as Range-Doppler (RD), Range-Angle (RA), and Micro-Doppler Signatures (μ-DS).

B. Classical Radar Signal Processing for Micro-Doppler Signature Generation

Radar can be used to observe moving objects by measuring the Doppler shift in the reflected signals caused by the motion of the target. These Doppler shifts, which change over time, create a signature in the TF domain that can be observed in the TF representation of the radar signals.

One commonly used TF transform for visualizing Micro-Doppler Signatures is the spectrogram. It 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 radar data, denoted as $x(t)$, denoted as:

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

where, $h(t)$ represents a windowing function, such as a rectangular, Hamming, or Hanning window. The spectrogram provides a visual representation of how the Micro-Doppler frequency content of a target changes over time.

III. Proposed Method

This section presents the structure and specifics of our proposed ML approach for generating high-resolution μ-DS. It covers dataset creation, introduces HRSpecNet architecture, and explains the training process with a weighted loss function.

A. Dataset Generation

The dataset creation process consists of two phases. First, we generated multi-component 1D complex time-domain signals as inputs for our architecture. Then, we created label TF images to represent the TF characteristics of these input signals.

1) 1D Multi-component Complex Signal Generation: We utilize multi-component sinusoidal FM signals denoted as $s(k)$ as the dataset inputs. These signals and their corresponding instantaneous frequencies (IFs) are expressed as follows:

$$\begin{aligned} s(k) = \sum_{q=1}^{Q} A_q \exp(2\pi f_q k) \\ \times \exp(j2\pi B_q \times \sin(2\pi(a_q k^2 + b_q k + \theta_q))), \end{aligned} \tag{4}$$

$$\begin{aligned} IF_q(k) = f_q + 2\pi B_q(2a_q k + b_q) \\ \times \sin(2\pi(a_q k^2 + b_q k + \theta_q)), \end{aligned} \tag{5}$$

In this context provided, the component number Q adheres to a discrete uniform distribution $\mathcal{U}(1, 10)$. The intensity of the qth component, denoted as A_q, is calculated as $0.5 + 8|\sigma_q|$, where σ_q follows a uniform distribution $\mathcal{U}(0, 1)$. The vibration

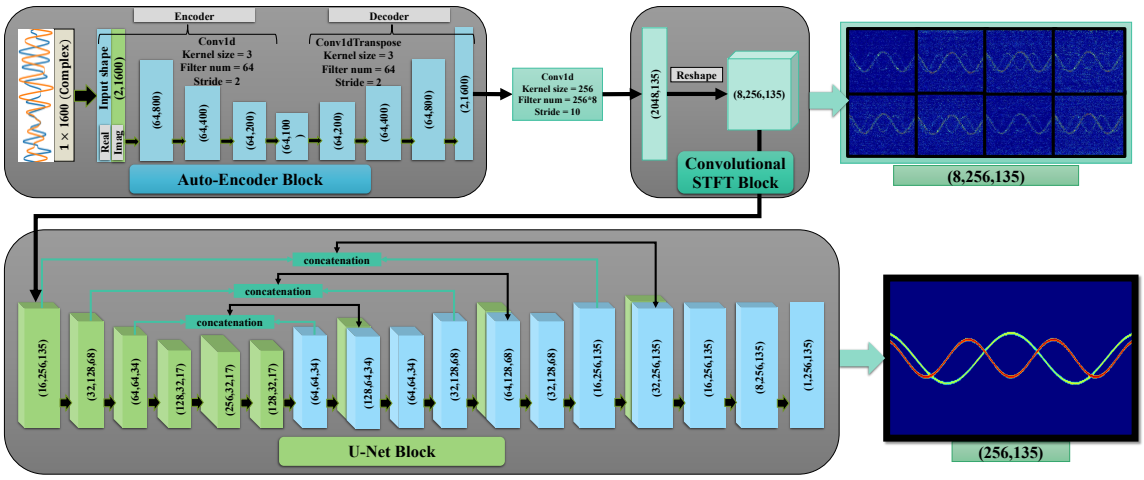

Fig. 1: The HRSpecNet Architecture

amplitude of the qth component, represented by B_q, is sampled from $\mathcal{U}(0.2, 16)$. Both a_q and b_q fall within the ranges of $[-4, 4)$ and $[-2.4, 2.4)$, respectively. The parameter θ_q is drawn from a uniform distribution $\mathcal{U}(0, 2\pi)$, and the Doppler shift f_q follows a uniform distribution $\mathcal{U}(-1000, 1000)$. In order to limit the computational challenges while training the DL model, the signal length L is set to 1600 with a sampling frequency of 3200. The sampling frequency is determined by the pulse repetition frequency (PRI) of our radar system. We set the number of frequency bins, N_f, to 256. It's important to note that our model architecture is designed to be versatile, allowing it to generate TF representations for inputs of varying lengths and a wide range of sampling frequencies during testing even though the training is done over signal length of L.

2) Ground truth TF representations: To generate the ground truth data, we start by capturing each of the IF signal components as per (5). Then, we apply a moving average operation with a window size L_w and shift S_o to establish the label frequencies. This relationship can be expressed as follows:

$$\widehat{IF}_q(i) = \frac{1}{L_w} \sum_{j=S_o*i}^{S_o*i+(L_w-1)} IF_q(j); \quad i = 0, 1, 2, \ldots, L_t \tag{6}$$

This operation bears resemblance to the STFT process, resulting similar size time index for the labeled frequencies as in the STFT process.

Given that we have set the number of frequency bins to N_f, the resulting shape of the initial ground truth TF will be a matrix of size $N_f \times L_t$, where $L_t = (\left\lfloor \frac{L-L_w}{S_o} \right\rfloor + 1)$. This initial 2D ground truth matrix, denoted as $GT_{initial}(f, i)$, can be represented as:

$$\text{GT}_{initial}(f, i) = \begin{cases} A_q, & (\exists \text{q})(f\Delta f \leq (\widetilde{IF}_q(i) < (f+1)\Delta f) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where the time index, denoted as i, spans the range from 0 to $L_t - 1$, and the frequency index, denoted as f, ranges from 0 to $N_f - 1$. In this context, $\Delta f = \frac{f_s}{N_f}$ represents the frequency interval, with f_s being the sampling frequency. Lastly, $\widetilde{IF}_q(i) = \text{mod}(\widehat{IF}_q, f_s/2)$ signifies the modulo operation of $\widehat{IF}_q$ with respect to $f_s/2$. Afterwards, we convolve the 2D $GT_{initial}$ with a 1D Gaussian kernel with a kernel size of 3 and a standard deviation of 1 along the frequency dimension in order to compensate for averaging effects and smooth out our final ground truth, denoted as GT_{final}.

B. Proposed HRSpecNet Architecture

The HRSpecNet is a novel DL architecture for high-resolution TF representation and it consists of three main components: an auto-encoder, a convolutional network resembling the STFT to separate proxy TF, and a U-Net block for generating high-resolution 2D TF representations. The input comprises complex signals with $C \times L$, where $C = 2$ represents real and imaginary parts, and L iss on the input signal length. During training, the network processes noisy signals with varying SNR levels. The auto-encoder reduces noise, improving overall performance. The output of the auto-encoder feeds into the STFT-like convolutional network, which generates multiple proxy TF representations. The U-Net block refines these proxies into high-resolution TF representations. The model is trained with a weighted loss, encouraging it to generate high-resolution TF images close to the ground truth while reducing noise in the input signal. Detailed explanations of each block follow are given next.

1) Auto-Encoder Module: The autoencoder architecture employs multiple Conv1D layers to capture intricate patterns in the input data while reducing the feature space dimension. Each Conv1D block has 64 filters, a kernel size of 3, and

a stride of 2 for local convolution operations, resulting in reduced spatial dimensions. This encoding process enables the autoencoder to learn a compact representation of the input data. The input data is of shape 2×1600, and the autoencoder preserves the same output shape during training. The model minimizes the sum of squared error (SSE) loss, denoted as L_1 in Figure 2, by comparing its output to a noise-free version of the input signal, effectively reducing noise and enhancing the final TF representation.

2) Convolutional STFT Module: The noise-reduced output from the auto-encoder block feeds directly into the convolutional STFT block. This block employs a 1D convolutional unit to generate multiple proxy TF representations. The hyper-parameters of this convolutional layer, such as the kernel size and stride, correspond to the window size and shift in a typical STFT operation. The number of filters in this layer determines how many TF representations are created for the subsequent U-Net block. For an input signal of length L, the STFT module uses L_w as the kernel size and $N_f B$ filters, and the convolution shifting is controlled by a stride of n_0. The output feature maps of the STFT layer have dimensions of $N_f B \times L_t$. These feature maps are reshaped within the STFT block to produce $B \times N_f \times L_t$, illustrating B TF representations. For instance, in Figure 1, an example signal with a length of 1600 samples is depicted, divided into real and imaginary parts in a 2-channel configuration. In the convolutional STFT block, the kernel size, filter numbers, and stride are set to 256, 256×8, and 10, respectively. After reshaping, the output feature maps of the STFT block become $8 \times 256 \times 135$, with 8 representing the number of TF representations, 256 as the number of frequency bins, and 135 as the time index, L_t. Importantly, the framework is not dependent on a specific signal length, and the time index L_t within the convolutional STFT block is determined directly by the number of input samples. Therefore, the trained model with these parameters can generate TF representations for varying input lengths.

3) UNET Module: The U-Net module's main purpose is to combine multiple TF feature maps from the convolutional STFT module into a high-resolution 2D TF representation. It follows an encoder-decoder architecture, with the encoder using convolutional layers for hierarchical feature extraction, and the decoder using transposed convolutional layers to upsample feature maps. Skip connections, achieved through concatenation, help blend low-level and high-level features for detailed TF analysis, and the final convolutional layer generates a single-channel output for the 2D TF representation.

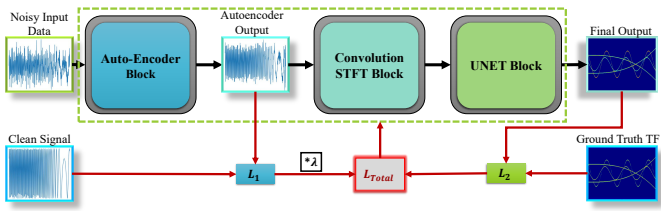

Fig. 2: Flow-diagram of the proposed architecture.

4) Training the proposed architecture: Our training dataset is constructed using 2×10^5 noisy signals, with uniformly random generated SNR levels ranging from 0 to 15 dB. For validation purposes, we employed a different set of 2×10^3 noisy signals, with SNR levels following the same variation. The loss in the U-Net module L_2 is computed as SSE loss between the final 2D TF output of the model and the labeled TF. This L_2 loss is combined with the loss in the auto-encoder block L_1. The total loss utilized in training the whole model is given as

$$L_{\text{total}} = \lambda \times L_1 + L_2 \tag{8}$$

where L_{total} denotes the aggregate training loss, while the parameter λ is a hyperparameter of the model and is systematically tuned to a value of 3 through an iterative process, aimed at achieving optimal performance. The flow diagram of the proposed model is shown in Fig. 2 for better visualization.

IV. Performance Analysis of the Proposed Method

In this section we will evaluate the performance of the proposed HRSpecNet model.

A. Micro-Doppler Signature Generation

To illustrate the performance of the proposed approach, an example signal consisting of three frequency components is considered. The signal expressions are given as follows:

$$\begin{aligned} \text{s(t)} = &\exp(j2\pi(\frac{1}{80}\sin(6\pi t)f_s)) \\ &+ \exp(j2\pi(\frac{1}{300}\sin(4\pi t)f_s)) \\ &+ \exp(j2\pi(\frac{1}{2.5}t - 0.4t^2)f_s) \end{aligned} \tag{9}$$

In this experiment, the SNR was 10 dB, two different sampling frequency, 4000 and 10000 Hz, were used, and the time duration was 1 seconds. First, the STFT of the data is computed using window lengths of 32, 64, 128, and 256 with a non-overlap length of 10. Figure 3 compares the TF representations of STFT and HRSpecNet with the same colormap. The STFT shows a considerable amount of noise, while the HRSpecNet gives a higher resolution, sparser, and cleaner representation. Additionally, the proposed HRSpecNet method was able to detect the intersection points of IF signals much more clearly than the STFT. For the best TF representations from STFT can be seen from window lengths 128 and 256 for a f_s of 4000 and 10000 Hz respectively. But HRSpecNet, without any parameter tuning, outputs fairly consistent TF representations for both tested sampling frequency cases. This signifies the robustness of HRSpecNET to exhaustive parameter tuning. Also, the result shows the high-resolution capability of the proposed approach.

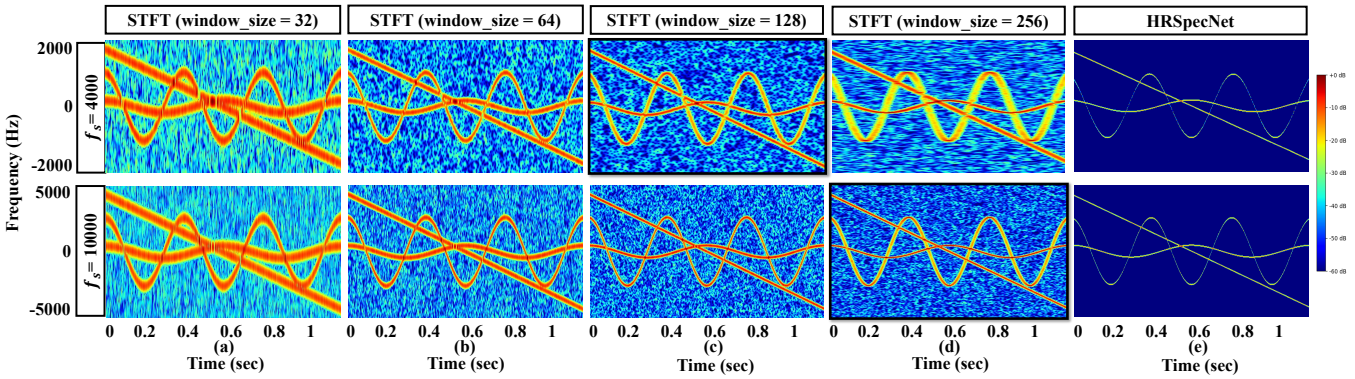

Fig. 3: TF representation of the Signal (9) by (a) to (d) STFT, (e) HRSpecNet. The highlighted black boxes indicate the best TF representations observed from STFT.

B. Quantitative Comparison on Classification Performance with STFT

One important analysis is how the reconstructed TF representations affect the final classification performance for STFT and HRSpecNet, in a real-world HAR scenario using a dataset of 20 ASL signs. The μ-DSs are generated with STFT and HRSpecNet. Notably, the HRSpecNet model, trained on synthetic data only, outperformed STFT by producing higher-resolution spectrograms with better noise suppression and clearer distinctions in subtle movements as shown in Fig. 4. These μ-DSs were then used as input for CNN classification.

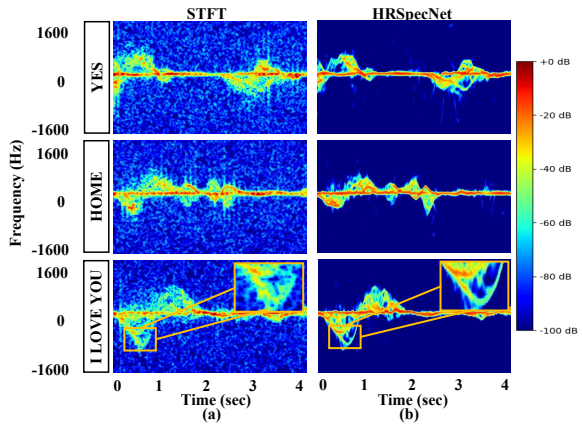

Fig. 4: Sample μD signatures for ASL signs for (a) STFT, (b) HRSpecNet

1) Experimental Setup and Dataset: For radio frequency (RF) data collection, a 77 GHz TI IWR1443 automotive short-range radar was used. The radar system parameters selected for the data collection are given in Table I.

In a lab setting, the radar was positioned on a table against a wall at a height of 0.91 meters, with ASL signers seated 1.5 meters away. To prevent signal interference, a computer monitor behind the radar displayed sign instructions. Six participants, including professional ASL signers, deaf individuals, CODAs, and lab members, collected data with IRB approval. The dataset consists of 100 diverse high-frequency ASL signs,

TABLE I: TI IWR1143 Radar Parameters

Parameter	Value
Number of TX & RX channels	1 & 1
Start & Stop frequency	77 & 81 GHz
Bandwidth	4 GHz
RX gain	45 dB
Periodicity	40 ms
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 time	28s

with 3,000 radar sign samples (30 per sign). Each participant performed five 4-second repetitions of each sign, with a 2-second interval, resulting in 28 seconds of data per sign per participant. For classification, the analysis will focus on the first 20 sign classes. Additional dataset details can be found in reference [25].

2) Classification Model: The μDSs generated from STFT and HRSpecNet were saved as 128×128 images and used as input for a 2D CNN classification model. This CNN architecture, based on previous work [9], consists of four convolutional blocks, as shown in Fig. 5. Each block includes two convolutional layers with 32 filters in the first two blocks and 64 filters in the remaining blocks. All convolutional layers use a 4×4 kernel size. Each block is followed by 2×2 maxpooling, batch normalization, ReLU activation, and a 0.3 dropout. After the convolutional blocks, the tensor is flattened and passed through a dense layer with a size of 256×1, followed by a 0.3 dropout, and then input into a softmax classifier.

3) Classification Performance for 20 Class ASL Data: For performance evaluation, two datasets were generated based on STFT and HRSpecNet respectively. Each dataset was split into 80% for training and 20% for testing. Models were trained for 150 epochs to ensure convergence. Subsequently, confusion matrices were generated for each case, and metrics like testing accuracy, precision, recall, and F1 scores were computed. Table II demonstrates that the classification acheived using the TF images generated by the proposed model outperformed

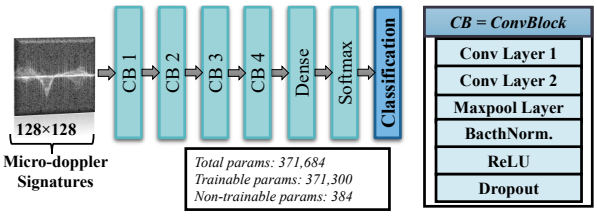

Fig. 5: CNN Architecture for ASL Sign Classification

the classification using the STFT-based spectrograms in all metrics. Notably, it achieved a 2.91% higher accuracy is achieved compared to the state-of-the-art STFT method.

TABLE II: Performance of the compared models in terms of evaluation metrics.

Network	Testing Accuracy	Precision	Recall	F1 Score
STFT	77.58	80.19	79.71	76.82
HRSpecNet	80.49	84.07	82.43	79.86

V. Conclusion

HRSpecNet, a novel deep learning architecture, exceling at reconstructing precise, high-resolution TF representations of complex signals is proposed. The proposed approach eliminats the need for parameter tuning, offering accurate, high-resolution spectrograms efficiently. The initial study shows the model's noise resilience and generalization capability in an experimental radar dataset, surpassing traditional methods in classification accuracy. It holds promise for RF-sensing-based activity recognition. For future work, an extensive analysis of the proposed approach as well as comparisons with other DL-based approaches will be provided.

Acknowledgment

The presented work was funded in part by the National Science Foundation (NSF) Awards #2047771. Human studies research was conducted under UA Institutional Review Board (IRB) Protocol #18-06-1271. We thank the team of the University of Alabama for providing the data for this research work.

References

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

[2] E. Kurtoğlu, S. Biswas, A. C. Gurbuz, and S. Z. Gurbuz, "Boosting multi-target recognition performance with multi-input multi-output radar-based angular subspace projection and multi-view deep neural network," *IET Radar, Sonar & Navigation*, vol. 17, no. 7, pp. 1115–1128, 2023.

[3] B. Debnath, I. A. Ebu, S. Biswas, A. C. Gurbuz, and J. E. Ball, "Fmcw radar range profile and micro-doppler signature fusion for improved traffic signaling motion classification," in *Proc. 2024 IEEE Radar Conference (RadarConf24)*, 2024.

[4] 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 Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933–1937, 2015.

[5] S. Björklund, T. Johansson, and H. Petersson, "Target classification in perimeter protection with a micro-doppler radar," in *2016 17th International Radar Symposium (IRS)*, 2016, pp. 1–5.

[6] 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 Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 46–56, 2019.

[7] S. Biswas, J. E. Ball, and A. C. Gurbuz, "Radar-lidar fusion for classification of traffic signaling motion in automotive applications," in *2023 IEEE International Radar Conference (RADAR)*, 2023, pp. 1–5.

[8] A. M. Alam, M. Kurum, and A. C. Gurbuz, "Radio frequency interference detection for smap radiometer using convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 10 099–10 112, 2022.

[9] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

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

[11] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

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

[13] S. Biswas, C. O. Ayna, S. Z. Gurbuz, and A. C. Gurbuz, "Cv-sincnet: Learning complex sinc filters from raw radar data for computationally efficient human motion recognition," *IEEE Transactions on Radar Systems*, vol. 1, pp. 493–504, 2023.

[14] T. Stadelmayer and A. Santra, "Data-driven radar processing using a parametric convolutional neural network for human activity classification," *IEEE Sensors Journal*, vol. 21, pp. 19 529–19 540, 2021.

[15] S. Biswas, C. O. Ayna, S. Z. Gurbuz, and A. C. Gurbuz, "Complex sincnet for more interpretable radar based activity recognition," in *2023 IEEE Radar Conference (RadarConf23)*, 2023, pp. 1–6.

[16] V. Chen, *The Micro-Doppler Effect in Radar, 2nd Ed.* Artech House, 2019.

[17] E. Kurtoğlu, 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 Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 699–712, 2022.

[18] S. Biswas, B. Bartlett, J. E. Ball, and A. C. Gurbuz, "Classification of traffic signaling motion in automotive applications using fmcw radar," in *2023 IEEE Radar Conference (RadarConf23)*, 2023, pp. 1–6.

[19] 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 & Navigation*, vol. 11, pp. 107–115, 1 2017.

[20] E. Kurtoglu, A. C. Gurbuz, E. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "Sequential classification of asl signs in the context of daily living using rf sensing," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

[21] G. Izacard, S. Mohan, and C. Fernandez-Granda, "Data-driven estimation of sinusoid frequencies," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[22] S. Biswas and A. C. Gurbuz, "Deep learning based high-resolution frequency estimation for sparse radar range profiles," in *2024 IEEE Radar Conference (RadarConf24)*, 2024.

[23] P. Pan, Y. Zhang, Z. Deng, and G. Wu, "Complex-valued frequency estimation network and its applications to superresolution of radar range profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[24] P. Pan, Y. Zhang, Z. Deng, S. Fan, and X. Huang, "Tfa-net: A deep learning-based time-frequency analysis tool," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2022.

[25] 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 Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 2732–2745, 2022.