Influence of Radar Signal Processing on Deep Learning-based Classification

Sean Kearney
Dept. of Electrical and Computer Eng.
University of Alabama
Tuscaloosa, Alabama
sjkearney@crimson.ua.edu

Sevgi Z. Gurbuz
Dept. of Electrical and Computer Eng.
University of Alabama
Tuscaloosa, Alabama
szgurbuz@ua.edu

***Abstract*—As radar technology becomes more readily available to researchers and users, it is thus being explored how to better process this data for real-time implementations. To process this radar data, the short time Fourier transform (STFT) has been implemented to then find the micro-Doppler spectrogram. When computing the STFT, there are parameters which can be adjusted to alter the size of the resulting micro-Doppler spectrogram. In this work, these parameters were adjusted to find the optimal representation of micro-Doppler radar returns of human activities, which were recorded using a 77 GHz Frequency Modulated Continuous Wave (FMCW) millimeter wave radar. To determine these optimal combinations, the resulting micro-Doppler spectrograms were used to train and test a Convolutional Autoencoder (CAE). The t-Distributed Stochastic Neighbor Embedding (t-SNE) and k-Nearest Neighbor Classification (kNN) were also utilized to find the nearest representations in a low-dimensional space of the spectrograms.**

***Index Terms*—radar, micro-Doppler, deep learning, time-frequency transform**

I. Introduction

In recent years, low-cost and low-power radars have become more accessible to researchers and other users thanks to the increasing research and development for commercial radar uses. With this increasing accessibility, the opportunities for how to implement radars has become endless. A particularly interesting area of implementation has been attempts to create real-time radar systems. There are currently limitations to implementing a fully real-time radar system, but this type of implementation could be of great use for detection systems, such as that for drones [1], in applications where the radar returns are used for control such as in self-driving cars [2], or in health monitoring systems [3]. For each of these systems having real-time feedback is critical for system implementation.

A key challenge to real-time radar system implementation lies in the complexity of the data representation, and resulting dimensionality, of micro-Doppler returns [4]. The generation of time-frequency transforms (TFT), e.g. the short-time Fourier transform (STFT), from the radar data is dependent on a variety of parameters, such as the number of FFT points, windowing size, window overlap, and overall dwell time. Each of these parameters impact the dimensionality of the resulting TFT, which influences the resulting complexity and time to process the radar returns. While the generation of spectrograms comprised of a large number of pixels, e.g. smaller windows and overlaps, may reveal finer features of the signature, high dimensional inputs increases the time required to compute the spectrogram while not necessarily having a proportional benefit in terms of deep neural network (DNN) based classification. Indeed, due to the *curse of dimensionality*, typically utilization of spectrograms beyond a certain size in fact do not result in greater classification accuracy.

Current research into the design of real-time radar systems have implemented a variety of methods for dealing with the computational challenges of TFT generation. In [5], the fast Fourier transform (FFT) was utilized for signal processing, with the number of FFT points being determined as the number of samples per chirp. While the FFT is able to convert a signal into the frequency domain, it cannot be used to find the change in time of the signal as can be done when using a TFT such as the short-time Fourier transform. Because of this, there is a limit on the variety of uses for the transformed data from the FFT. Instead of using the FFT alone, [6] and [7] each implement the STFT for processing, with Hamming windows of 0.2s and 0.3s and an overlap of 95%. Another option for detection using the STFT is seen in [8], where the STFT is implemented to then generate the cadence velocity diagram.

Another approach instead of initially processing the radar data is to input the raw radar data directly into a deep neural network. This process was implemented in [9], where a tracker was implemented to achieve Bayesian classification. While this approach can save time on processing the data, the network needed to process this data is a complex network, which will increase the processing complexity. Also, the raw data has a significantly larger dimensionality than the processed spectrogram, which also increases the processing complexity and thus the processing time for the complex network. Additionally, DNNs designed to directly processes raw RF data are often limited to classification of a few number of classes and may be more susceptible to clutter or other artifacts that degrade performance. Computation of data representations such as the spectrogram allow for the application of signal processing techniques to boost signal-to-noise ratio (SNR), such as moving taget indication (MTI) filtering, or physics-aware machine learning techniques [10].

This study systemically investigates the optimal parameters

of the short-time Fourier transform and resulting spectrogram for micro-Doppler radar returns as inputs to DNNs for classification. In Section II, which parameters of the short-time Fourier transform that are adjusted are discussed as well as how they change the resulting spectrogram. In Section III, the means for testing the spectrogram is described, with details on the Convolutional Autoencoder as well as the t-Distributed Stochastic Neighbor Embedding and k-Nearest Neighbor Classification which were utilized. Lastly, the results of the testing methods are presented in Section IV, with the optimal hyper-parameters of the short-time Fourier transform discussed.

II. Micro-Doppler Signature Generation

The received signal for an FMCW radar is a complex time-stream of I/Q samples that are comprised of a time-delayed and frequency-shifted version of the transmitted signal. This data can be re-shaped into a radar data cube that has dimensions of fast-time (range), slow-time (velocity), and channel (number of antenna elements). Typically, the micro-Doppler signature is computed by finding the TFT of a slow-time slice of the data matrix for a given channel at the target location.

A. Short-Time Fourier Transform (STFT)

The most commonly applied TF transform is the spectrogram, denoted by $S(t,\omega)$, which is the square modulus of the STFT. It can be expressed in terms of the employed window function, $w(t)$, as

$$S(t,\omega) = \left| \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-iwt}\,dx \right|^2 \tag{1}$$

The spectrogram reveals the micro-Doppler frequency modulations which are centered about the main Doppler shift related to translational movement [11]. While reaping the benefits of linearity, STFTs though must trade-off time and frequency resolution due to fixed length windows [12].

When computing the STFT, there are a number of parameters which can be adjusted to trade-off resolution in time and frequency. These parameters include the number of fast Fourier transform points (NFFT), window size, and window overlap. Use of longer windows improves frequency estimation, while lowering temporal resolution. Reducing the overlap between windows reduces the time between frequency estimates. Increasing the number of FFT points increases the fineness of the frequency representation. Thus, these parameters have an important impact not just on the ability of the TFT to capture fine micro-Doppler features, but also the resulting dimensionality (number of pixels comprising the image) of the micro-Doppler signature.

B. Parameter Variations

In this study, the NFFT was defined to be 2^9, or 512, and the window size was varied as either 128, 256, or 512. The pixel size may be computed based on the length of the input data, window size, and window overlap according to the following relation:

$$N = \frac{DataLength - WindowSize - 1}{WindowSize - Overlap} \tag{2}$$

With this equation, it is found that the pixel size directly corresponds to the length of the second dimension of the STFT while the first dimension corresponds to the number of FFT points. This in turn means that the Pixel size dictates the resolution in the time domain of the resulting spectrogram.

Using Eq. 2, the window overlap was found for the window sizes corresponding to pixel sizes of 128, 256, 512, and 1024. It was found that for a pixel size of 128, only a window size of 512 could be used, and for a pixel size of 256, only window sizes of 256 and 512 could be used since there was no possible window overlap size that would result in the desired pixel size for any of the defined representations. Also, as the desired pixel size increases, the variance of the actual pixel size increases as well since for a given window size there will be no window overlaps that result in the exact pixel size desired. This can be seen in Table I, where the desired pixel size and actual pixel size is listed for each parameter combination. For the purpose of this paper, the NFFT was kept fixed and therefore the dimension corresponding to the pixel size was adjusted. The variance in this dimension is tested for the rest of this paper, with the intention of finding combinations that reduce the pixel size while retaining decent accuracy for testing with a DNN.

III. Experiment Design

A. Dataset Utilized

The dataset used in this study was acquired with a TI AWR1642 single-chip 76-GHz to 81-GHz automotive radar in an indoor environment. The radar was placed on a table 1.5 meters up from the ground facing a walkway that was 6 meters long and 3 meters wide. Data was acquired from 10 participants for 13 different ambulatory and 2 in-place activities:

1) Normal walking
2) Short steps
3) Walking with a cane
4) Walking with crutches
5) Skipping
6) Scissors gait
7) Walking on toes
8) Marching
9) Limping
10) Vacuuming the floor
11) Dragging furniture
12) Walking while carrying a load in one hand
13) Walking while carrying a load in both hands
14) Putting books on a bookshelf
15) Folding laundry

Each activity was articulated while moving towards the radar and a total of 1436 samples were acquired. Examples of the micro-Doppler spectrograms of the human activity classes

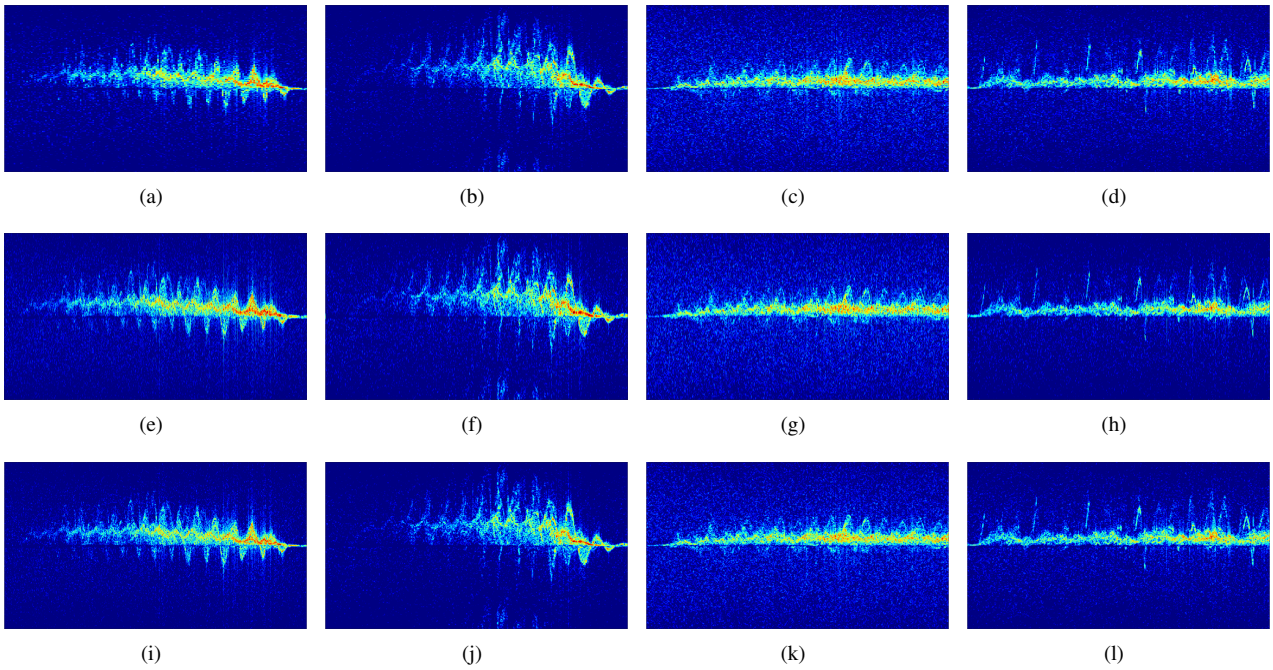


Fig. 1: Micro-Doppler spectrograms for Vaccuming, Normal walking, Short steps, and Walking with a cane. Row 1: window size of 512 and pixel size of 256. Row 2: window size of 256 and pixel size of 512, and Row 3: window size of 512 and pixel size of 1024.

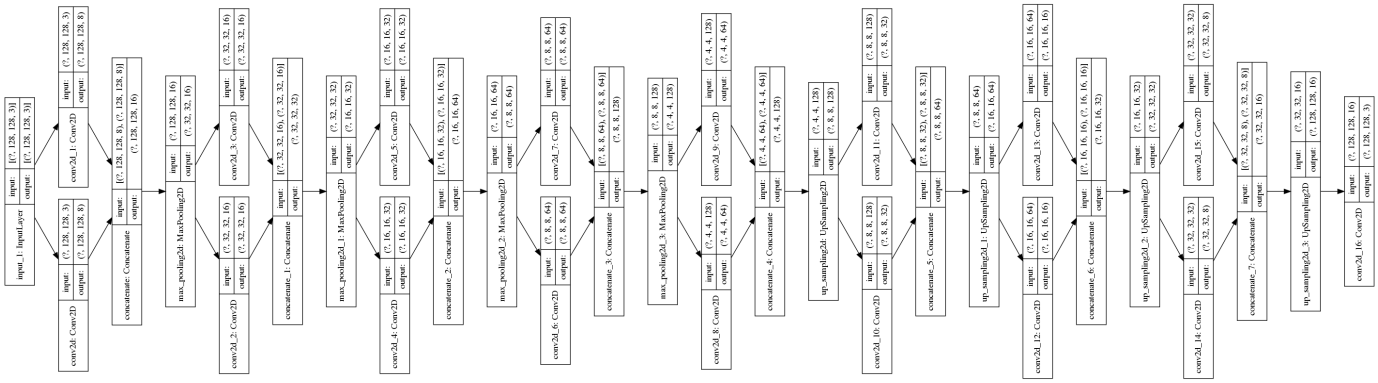


Fig. 2: CAE Network Structure.

is shown in Figure 1, with each individual class containing between 92 to 106 samples except for class 15 Folding laundry, which contains 47 samples.

B. Classification Approach

In this study, a convolutional autoencoder (CAE) was used to evaluate the accuracy resulting from varying the parameters of the spectrogram and resulting input data dimensionality. The CAE was trained on 80% of the spectrograms while the remaining 20% were used for testing. Additionally, the k-nearest neighbor classifier was found for the spectrograms. Examples of the spectrogram images used for training can be seen in Figure 1, corresponding to spectrograms generated with window sizes of 512, 256, and 512, and pixel sizes of 256, 510, and 1024, respectively.

The architecture of the CAE utilized is shown in Figure 2. CAEs are composed of an encoder and decoder. Initially, the encoder takes as input all images, which are then encoded through the first 16 layers. This encoded output is then sent to the decoder, where the goal of the decoder is to recreate the initial input to the encoder. After training the full CAE, the model for training and testing for classification of the images removes the decoder and instead inputs the output of the encoder directly to the classifier, which consists of a batch normalization layer, 2 dense layers each followed by a dropout layer, and then a dense layer followed by a softmax layer for prediction. The results from utilizing the CAE for classification is discussed in Section IV.

TABLE I: Results Per Window Size and Pixel Size

Window Size	512	256	512	128	256	512	128	256	512
Desired Pixel Size	128	256	256	512	512	512	1024	1024	1024
Actual Pixel Size	128	256	256	511	510	512	1014	1012	1024
Mean Accuracy	64.76	75.4	78.24	73.07	77.19	61.1	76.36	69.3	78.3
Best Accuracy	75.3	77.4	82.9	79.02	82.9	73.6	78.8	75.7	84.4

TABLE II: Distance of kNN

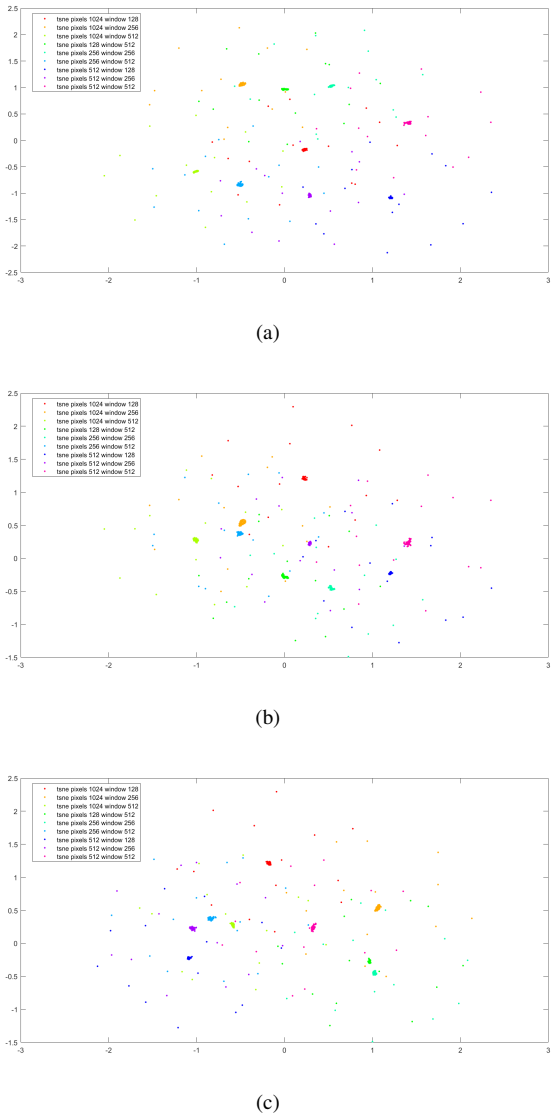
Pixel Size	**Window Size**	**Average distance from center of 1024 512**
128	512	6.1592
256	256	7.6355
256	512	2.5758
512	128	7.7816
512	256	4.3760
512	512	9.3176
1024	128	5.0756
1024	256	5.6929
1024	512	0.4758

C. Neighbor Classification

In addition to using a CAE for classification, neighbor classification methods were utilized to compare the similarities between micro-Doppler spectrograms created with the different window and pixel sizes. For neighbor classification, t-Distributed Stochastic Neighbor Embedding (t-SNE) and k-Nearest Neighbor Classification (kNN) was utilized. The t-SNE is used to preserve the local structure of data by matching pairwise similarity distributions in both the higher-dimensional original data space and the lower-dimensional projected space [13]. Initially, the t-SNE was implemented to map each 10-second data sample into a 1x3 low-dimensional space. Each of these 1x3 results were then stacked to form a 1436x3 matrix, where each row corresponds to a single data sample. This matrix was then utilized to find the distance between k-nearest neighbors of each window and pixel size combination with the samples created with a window size of 512 and a pixel size of 1024, since that combination was shown to have the highest classification accuracy in Section IV.

IV. Results

The classification results of the CAE can be seen in Table I. To attain this result, the model was trained and tested 5 times for each window and pixel combination. It was then found what the mean classification accuracy is based on the 5 trials as well as the highest classification accuracy during the 5 trials. From these results, it can be seen that for a window size of 512 and a pixel size of 1024 the best mean classification accuracy was observed to be 78.3% and the highest classification accuracy was observed as 84.4%. The next best combination was then observed for a window size of 512 and a pixel size of 256, where the mean classification accuracy was observed as 78.24% and the highest classification accuracy was observed as 82.9%. With this combination, the average classification accuracy and highest observed accuracy were each the second best among all of the combinations. The only other combination to observe a maximum classification accuracy above 80% was the combination with a window size of 256 and a pixel size of 512, which was only 1 percentage point off the average accuracy of the combination with a window size of 512 and a pixel size of 256. Example micro-Doppler spectrograms of



(a)

(b)

(c)

Fig. 3: t-SNE feature plots comparing layers (a) 1 and 2, (b) 1 and 3, and (c) 2 and 3.

these three window and pixel size combinations is displayed in Figure 1. With these results, it was then observed how these parameter combinations compared for the k-Nearest Neighbor classification.

For comparing the neighbor classification, the distance between k-Nearest Neighbors for k being 5 is displayed in Table II. Each combination of window size and pixel size was compared to the combination with a window size of 512 and a pixel size of 1024. This base parameter combination was used for all comparisons since it was observed to have produced the best classification accuracy when trained and tested on the CAE network, as demonstrated in Table I. The results here correlate with the classification results of the CAE, where the parameter combination with a window size of 512 and a pixel size of 256 had the closest distance from the comparison case. The next closest combination was once again the parameter combination with a window size of 256 and a pixel size of 512 as well. It is then observed that none of the other parameter combinations is within a distance of 5.0.

Lastly, the comparisons of the t-SNE spaces was plotted and labeled for each data point as shown in Figure 3. Three plots were made to compare each of the three channels from the RGB spectrogram images that was created as the t-SNE, which can be seen in Figures 3(a), 3(b), and 3(c). In each of these three plots, the closest cluster to the cluster corresponding to a window size of 512 and a pixel size of 1024 is once again the combination with a window size of 512 and a pixel size of 256, which reinforces the CAE and kNN results reported above.

V. Conclusion

This paper has demonstrated how adjusting the window size and pixel size of the short-time Fourier transform can result in optimal combinations for data processing. Each of the nine combinations were initially implemented in a Convolutional Autoencoder to find the classification accuracy, where it was observed that a window size of 512 and a pixel size of 1024 resulted in the best trained CAE. It was also observed that a window size of 512 and a pixel size of 256 had the next best results out of the 9 parameter combinations. Then, the t-SNE was implemented on the micro-Doppler spectrogram data matrices and the k-Nearest Neighbors was found between each combination, where it was confirmed that the combination with a window size of 512 and a pixel size of 256 was closest to the combination with a window size of 512 and a pixel size of 1024. This reduced pixel size will allow for micro-Doppler spectrograms to be generated with a significantly smaller size without significantly reducing the classification accuracy resulting from this micro-Doppler spectrogram, thus allowing for a sped-up processing time in real-time applications.

Acknowledgment

This work was funded in part by the National Science Foundation (NSF) Award #1932547 as well as AFOSR Award #FA9550-22-1-0384. Human studies research was conducted under UA Institutional Review Board (IRB) Protocol #18-06-1271.

References

[1] F. Hoffmann, M. Ritchie, F. Fioranelli, A. Charlish, and H. Griffiths, "Micro-doppler based detection and tracking of uavs with multistatic radar," in *2016 IEEE Radar Conference (RadarConf)*, 2016, pp. 1–6.

[2] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 20–31, 2019.

[3] S. Z. Gurbuz, E. Kurtoglu, M. M. Rahman, and D. Martelli, "Gait variability analysis using continuous rf data streams of human activity," *Smart Health*, vol. 26, p. 100334, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235264832200068X

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

[5] J.-W. Choi, C.-W. Park, and J.-H. Kim, "Fmcw radar-based real-time hand gesture recognition system capable of out-of-distribution detection," *IEEE Access*, vol. 10, pp. 87 425–87 434, 2022.

[6] F. Fioranelli, M. Ritchie, and H. Griffiths, "Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-doppler," *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1702–1710, 2016.

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

[8] R. Ricci and A. Balleri, "Recognition of humans based on radar micro-doppler shape spectrum features," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1216–1223, 2015.

[9] P. Vaishnav and A. Santra, "Continuous human activity classification with unscented kalman filter tracking using fmcw radar," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

[10] M. Rahman, S. Gurbuz, and M. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–15, 2022.

[11] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using rf sensing," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3763–3775, 2021.

[12] M. Mahbubur Rahman and S. Z. Gurbuz, "Multi-frequency rf sensor data adaptation for motion recognition with multi-modal deep learning," in *2021 IEEE Radar Conference (RadarConf21)*, 2021, pp. 1–6.

[13] D. M. Chan, R. Rao, F. Huang, and J. F. Canny, "T-sne-cuda: Gpu-accelerated t-sne and its applications to modern data," in *2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, 2018, pp. 330–338.