Cross-frequency training with adversarial learning for radar micro-Doppler signature classification

Sevgi Z. Gurbuz[a], M. Mahbubur Rahman[a], Emre Kurtoglu[a], Trevor Macks[a], and Francesco Fioranelli[b]

[a]The University of Alabama, Tuscaloosa, AL, USA
[b]Delft University of Technology, Delft, The Netherlands

ABSTRACT

Deep neural networks have become increasingly popular in radar micro-Doppler classification; yet, a key challenge, which has limited potential gains, is the lack of large amounts of measured data that can facilitate the design of deeper networks with greater robustness and performance. Several approaches have been proposed in the literature to address this problem, such as unsupervised pre-training and transfer learning from optical imagery or synthetic RF data. This work investigates an alternative approach to training which involves exploitation of "datasets of opportunity" – micro-Doppler datasets collected using other RF sensors, which may be of a different frequency, bandwidth or waveform - for the purposes of training. Specifically, this work compares in detail the cross-frequency training degradation incurred for several different training approaches and deep neural network (DNN) architectures. Results show a 70% drop in classification accuracy when the RF sensors for pre-training, fine-tuning, and testing are different, and a 15% degradation when only the pre-training data is different, but the fine-tuning and test data are from the same sensor. By using generative adversarial networks (GANs), a large amount of synthetic data is generated for pre-training. Results show that cross-frequency performance degradation is reduced by 50% when kinematically-sifted GAN-synthesized signatures are used in pre-training.

Keywords: Micro-Doppler classification, radar networks, transfer learning, generative adversarial networks

1. INTRODUCTION

In recent years, micro-Doppler classification has been exploited in a variety of applications, including border control and security, monitoring activities of daily living, sensing for smart environments, intruder detection, assisted living, and man-machine interfaces via gesture recognition. The primary focus of assisted living and remote health applications has been fall detection,[1–5] gait abnormality recognition,[6] concussion detection,[7] physical therapy and rehabilitation,[8] non-contact measurement[9] of heart rate[10, 11] and respiration,[12, 13] as well as detection of related conditions, such as sleep apnea[14] or sudden infant death syndrome.[15]

In all the aforementioned applications, Deep Neural Networks (DNNs) have been an enabling technology in the advancement of radar micro-Doppler classification algorithms. However, a common challenge in radar-based human activity recognition is how to deal with the limitations of small datasets, especially when training DNNs, which are quite data-hungry for supervised learning.[16] One approach for dealing with low sample support is to use sparsely connected layers, as opposed to fully connected layers, to reduce the total number of parameters in the network.[17] Convolutional autoencoders (CAE) have also been proposed as a DNN architecture that uses unsupervised pre-training to mitigate measured data requirements in micro-Doppler classification.[18] However, a recent study has[19] shown that when only a meager amount of training data is available, e.g. less than 650 training samples, CAEs have been outperformed by transfer learning from other domains, such as optical imagery.

In transfer learning, knowledge gained from a different domain is transfered to initialize the weights of the DNN. Pre-training the network usually gives a better starting point for optimizing network weights than random initialization, as it is more likely for the local minimum attained to be closer to the true minima in a global sense. Thus, transfer learning can be a powerful technique when the amount of training data is insufficient and have

---

Further author information: (Send correspondence to Sevgi Z. Gurbuz)
Sevgi Z. Gurbuz: E-mail: szgurbuz@ua.edu, Telephone: +1 (205) 348-4382

been effective in enabling numerous novel applications, including activity recognition,[20, 21] voice recognition,[22] and gesture recognition.[23] This, of course, presumes that the data used in pre-training relates to the classification task at hand. The phenomenology of radar signals, however, fundamentally differs from that of optical signals. While pre-training on optical imagery may guide the network in recognizing generic spatial information, it is does not inform the network on micro-Doppler specific features. Fine-tuning with real RF data significantly boosts performance after pre-training on optical imagery.

This has thus motivated interest in the generation of synthetic data for pre-training, which would offer better initialization than use of generic images. The use of motion capture (MOCAP) data to synthesize micro-Doppler data[24] has become more popular in recent years. MOCAP data can be used to animate a skeleton model of the human body and enable it to move in accordance to the MOCAP data. However, acquisition of MOCAP data still requires recruitment of participants with different body types, heights, weights, gender, and gait styles to obtain a sufficient amount of data with adequate statistical diversity. To minimize the amount of human effort required to generate a diverse training dataset, a method for applying data augmentation directly to the underlying skeletal model was recently proposed.[25] Scaling of the skeletal tracking data along the time axis was used to speed up or slow down motion, while scaling of physical dimensions used to represent people of different sizes. Perturbations were also applied to the motion trajectories of each body part to generate statistically independent variations in the resulting gait signature. In this way, just a small amount of measured MOCAP data was needed to generate a large amount of diversified MOCAP data, which better spanned expected target profiles. It has been shown that this method surpasses alternative methods using transfer learning on networks pre-trained with ImageNet, CAEs, and CNNs.[25, 26]

The principle disadvantage of this approach, however, is that the MOCAP data only models target motion. Thus, only the target component of the real measurement can be synthesized - sensor artifacts, clutter, and noise are not included in the simulation. An alternative approach for generating synthetic data, which can potentially bridge this deficiency, is adversarial learning.[27–29] Recent studies[27, 30] show that auxiliary conditional GANs (ACGANs) are effective in modeling both sensor imperfections and clutter in a through-the-wall sensing scenario. However, the main challenge in application of GANs has been the lack of kinematic fidelity in a significant percentage of signatures generated. The resulting discrepancy between the target profiles of the synthetic versus real signatures causes a significant amount of degradation - as much as 10% - in classification accuracy. This is because GANs view the data as real images, and do not have any means for knowing the underlying physical constraints that limit possible human motion. It is possible, therefore, for GANs to generate synthetic samples that may appear visually similar, but are in fact incompatible with possible human motion.

In this paper, we consider yet another alternative for training micro-Doppler signatures, which involves exploitation of "datasets of opportunity" - data collected at a different location, with different participants, under different experimental scenarios (e.g. perspectives and distances), for different classes, and with different sensors. As low-cost, embedded RF sensors have become more ubiquitous, an increasing number of researchers are making RF data they have acquired publicly available.[31] This paper specifically considers the problem of performance degradation due to mismatch in the frequency and bandwidth of the RF sensors, when other factors, such as the motion classes, are the same.

Cross-frequency compatibility is a key aspect sensing that has seen limited discussion in the literature,[32, 33] as most works focus on single frequency systems. In a smart environments, however, multiple RF sensors may be deployed, while operating at different transmit frequencies to prevent interference between sensors. Vishwakarma et. al.[33] proposed a dictionary learning approach for classification across multiple frequency, which was motivated by cognitive radar applications. However, only four easily discernable classes were considered with transmit frequencies varying by just 2.5 GHz. When the difference between transmit frequency is great, important differences in micro-Doppler signatures measured can be observed. While the general shape of motion may be the same, the signatures will differ in resolution, Doppler spread, electronic noise, clutter, and sensor artifacts due to the change in scattering mechanism, e.g. how moving targets, objects, and clutter reflect waves differently at different wavelengths.

In this paper, we aim at evaluating the extent of performance degradation incurred due to such discrepancies, and examining whether simply increasing the amount of training data can mitigate these performance losses. Section 2 describes the experimental setup, data collection and the data processing methods employed. The

network architecture and classification performances under a variety of cross-frequency training scenarios are detailed in Section 3, while key conclusions are summarized in Section 4.

2. EXPERIMENTAL DATASET

This section describes the RF sensors utilized for acquisition of micro-Doppler signatures for activities of daily living, as well as the pre-processing steps used on the data prior to classification.

2.1 Test Setup

Three different RF sensors were used in this study:

- 77 GHz FMCW radar: The Texas Instruments IWR1443 Frequency Modulated Continuous Wave (FMCW) transceiver was set at a 77 GHz center frequency and 750 MHz bandwidth.

- 24 GHz FMCW radar: An Ancortek SDR-Kit was set to transmit an FMCW waveform at 24 GHz center frequency and 1500 MHz bandwidth.

- XeThru UWB impulse radar: The XeThru X4 sensor transmits across a band of roughly 7 GHz - 10 GHz.

All sensors were placed side-by-side on top of a table of 1-meter height from the ground, with the test subject moving about 0.5-3 meters from the sensors. Figure 1 illustrates the sensors positioning. The Xethru and 24 GHz sensors were operated from one laptop, whereas a separate, dedicated laptop was used to operate 77 GHz sensor. Each sensor has its own graphical user interface and a common graphical user interface has been used to enable synchronization across the sensors and simultaneous data collection.

2.2 Activity Datasets Acquired

Six participants of various ages, heights and weights were involved in this study. A total of 11 different activities and ambulatory gaits were considered, as listed in Figure 1. The choice of these activities were motivated by smart environment applications, where monitoring of activities of daily living are required to support health monitoring and gesture recognition. Each participant conducted 10 repetitions of each activity, resulting in a total of 60 samples per class per sensor. Additionally, data previously acquired at an earlier date from the 24 GHz sensor was used to enrich the dataset with an additional 180 samples per class. All activities were conducted along the radar line-of-sight.

2.3 Data Processing

The data acquired by each RF sensor is a time-stream of complex I/Q data. After reshaping the data matrix according to the pulse repetition frequency (PRF), phase corrections on I & Q channels are applied. Afterwards, a moving target indication (MTI) filter is employed to suppress stationary returns, such as from ground clutter. The magnitude of the Short-Time Fourier Transform (STFT) of the signal is then applied to generate a 2D time-frequency representation of the signal.
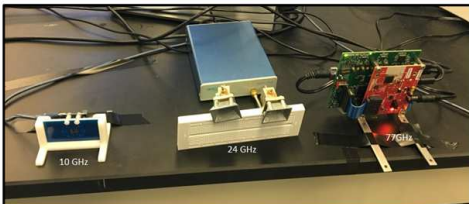
RF Sensors   **Human Motion Classes**

ID	Activity	ID	Activity
WLKT	Walking towards the radar	WLKA	Walking away from the radar
PICK	Picking up an object from the ground	BEND	Bending
SIT	Sitting on a chair	KNEEL	Kneeling
CRWL	Crawling towards the radar	LIMP	Limping with right leg stiff
WTOES	Walking on both toes	SHSTEP	Walking with short steps
SCSSR	Scissor gait		

Figure 1: Test setup and description of experiments.

A simplified block diagram summarizing the pre-processing steps is shown in Figure 2. The micro-Doppler signatures from all the three RF sensors are illustrated in Figure 3.
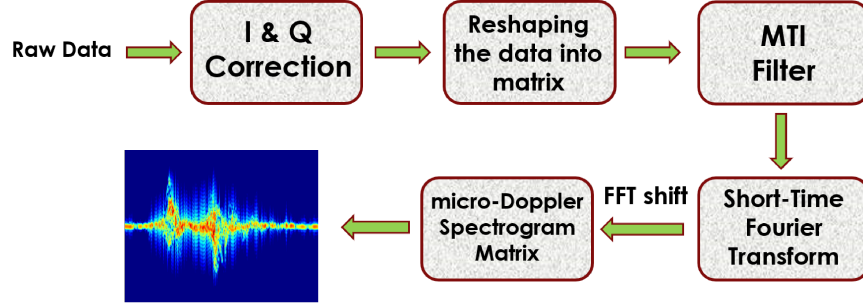
Figure 2: Pre-processing steps for micro-Doppler signature generation.
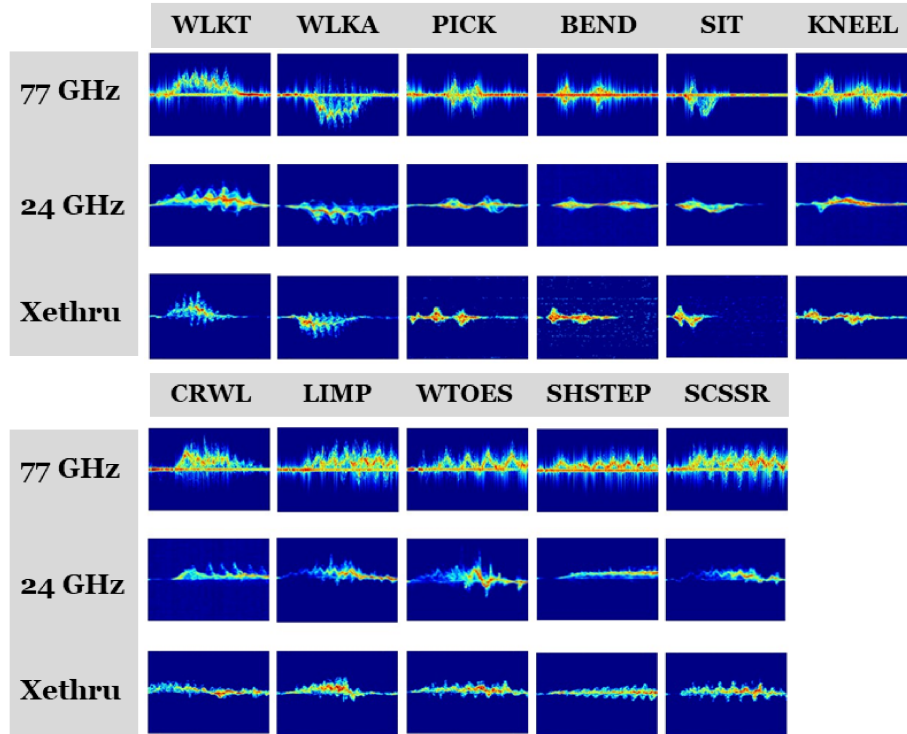
Figure 3: Micro-Doppler signatures for each activity class as acquired by each RF sensor.

3. CROSS-FREQUENCY CLASSIFICATION RESULTS

In this paper, the performance of several different deep neural networks (DNNs) and pre-training approaches is considered when comparing cross-frequency classification performance.

3.1 Transfer Learning from Optical Imagery

With transfer learning (TL), instead of starting the learning process from scratch, we start with patterns that have been learned from solving a different problem. TL utilizes the weights generated by a deep neural network pre-trained on a large dataset[34] from a different domain, which allows the detection of edges, curves and other

image patterns/properties. This network can then be exploited for classification of data from a different domain, by re-training the networks weights using a smaller set of labeled data from the desired, target domain. This is referred to as fine-tuning the network. In this way, TL mitigates the necessity of a large amount of target domain data for training, as well as reducing the amount of time required to train the classifier.

In past work,[19] we have found the VGG16 architecture to be more effective in classification of radar micro-Doppler than other networks, such as AlexNet or GoogleNet. Thus, in this paper, the VGG16 architecture,[35] pre-trained on the ImageNet database, is exploited with modification of the top layer to include a global average pooling layer followed by two dense layers and a classification layer with SoftMax activation function. The two dense layers use the Relu activation function, with each of them followed by a dropout of 0.5. Among other tuning parameters, we have used a batch size of 32, 40 epochs, learning rate of 1e-4, momentum of 0.9 and Adam optimizer. The results in Table 1 shows that the highest classification accuracy of 89% is obtained when the network is trained and tested with data from the same RF sensor. When fine-tuned on real data at a different frequency, the classification drastically drops by over 70%!

Table 1: VGG16 Classification Results

Fine-Tuning	Testing	Accuracy
77 GHz	77 GHz	89.23
	24 GHz	16.66
	10 GHz	14.28
24 GHz	77 GHz	11.13
	24 GHz	85.57
	10 GHz	15.55
10 GHz	77 GHz	9.00
	24 GHz	14.21
	10 GHz	83.01

3.2 Convolutional Autoencoder and Convolutional Neural Network

Convolutional Autoencoders (CAE) have demonstrated their effectiveness in various applications like image denoising, dimensionality reduction and image search. Whereas convolutional neural networks (CNNs) are initialized using random weights, CAEs utilize unsupervised pre-training for network weight initialization. CAEs have been shown to be effective when small, yet reasonable, amounts of real data are available for training.[18]

To ascertain the best CAE architecture for our datasets, we evaluated the performance of a combination of different parameters, such as the number of convolutional layers, number of filters, filter sizes and layer concatenation. The performance attained for various combinations of these parameters is listed in Table 2 for the case where the CAE is trained and tested with real data from the same sensor. Each model was trained for 100 epochs with mini-batch size of 16 and max pooling stride of 2 in each convolutional block. After training the CAE model, the decoder was removed and two fully connected layers with 128 neurons in each were added after flattening the output of the encoder. At the end, a softmax layer with 11 nodes is employed for classification. It may be observed that a 5-layer CAE with 64 9x9 & 64 3x3 concatenated filters gives the best performance for all datasets.

The cross-frequency training performance for unsupervised pre-training with the CAE is compared with that of a randomly initialized, comparable CNN with the same architecture as just the encoding layers of the CAE. Hyperparameters, such as the number of hidden nodes in dense layers, dropout rate, mini-batch size, learning rate are further optimized using grid search method. To assess cross-frqeuency performance, the CNN is trained on Sensor A while tested with Sensor B. In the case of the CAE, all combinations of pre-training, fine-tuning, and testing are considered: the CAE can be pre-trained on Sensor A, fine-tuned with Sensor B, and tested on Sensor C. The classification accuracies attained for these training scenarios are summarized in Table 3 and 4.

Notice that for both the CAE and CNN, the best performance was achieved, not unsuprisingly, for the case when the training and test data were from the same RF sensor. In this case, a classification accuracy of 91.5% was achieved. More interesting, however, are the cross-frequency training results of the CAE:

Table 2: Parameter Optimization Table for CAE

Filter Size	Depth	Num. of Filters	Accuracy (%)			Depth	Num. of Filters	Accuracy (%)		
			10 GHz	24 GHz	77 GHz			10 GHz	24 GHz	77 GHz
3x3 − 9x9	3	8	79.6	77.8	78.5	5	16	81.6	83.3	83.1
	3	16	77.6	79.2	85.4	5	32	78.9	84.3	84.6
	3	32	79.6	79.4	83.8	5	64	89.1	86.9	88.5
	4	8	76.2	81.6	83.1	6	16	82.3	84	81.5
	4	16	84.4	82.1	85.4	6	32	83	81.1	86.9
	4	32	82.3	84.1	83.1	6	64	83	82.9	81.5
3x3 − 5x5	3	8	78.9	75.2	82.3	5	16	76.2	83	84.6
	3	16	76.9	78.1	80.7	5	32	86.4	84.1	87.7
	3	32	79.6	78.6	76.2	5	64	85.3	85	87.7
	4	8	84.4	76.7	85.4	6	16	78.9	85.4	84.6
	4	16	76.9	81.6	85.4	6	32	86.4	86.6	85.3
	4	32	87.1	84.4	86.9	6	64	83.7	86.6	86.9
3x3	3	16	74.8	70.9	79.2	5	16	78.9	76.4	78.5
	3	32	76.9	72.6	79.2	5	32	83	78.8	84.6
	3	64	78.9	75.6	76.2	5	64	83.7	83.8	83.1
	4	16	75.5	74.1	80	6	16	73.5	76.3	82.3
	4	32	79.6	80.5	83.1	6	32	80.3	80.3	82.3
	4	64	80.3	82.5	82.33	6	64	80.3	84.1	83.1

1. **Fine-tuning with data that is from the same sensor as the test data if of great importance.** In cross-frequency training cases, where only the pre-training involved data from a different RF sensor, the classification accuracy dropped by 7% - 15%, as opposed to the 70% drops seen in TL and when there is mismatch between fine-tuning and testing datasets.

2. **The accuracy achieved with VGG16 pre-trained on ImageNet is greater than that obtained with the CAE in cross-frequency situations.** Looking at cases where the data used for fine-tuning and testing are from the same sensor, at 77 GHz, VGG16 yields 89% whereas the CAE attains 83.8% and 80% with pre-training on 24 GHz and 10 GHz, respectively. At 24 GHz, VGG16 achieves an accuracy of 85%, while the CAE yields 74% and 79% with pre-training on 77 GHz and 10 GHz, respectively. At 10 GHz (Xethru), VGG16 gives 83% accuracy, while the CAE yields 75% and 81% with pre-training on 77 GHz and 24 GHz, respectively.

3. **The amount of data used in pre-training VGG16 with ImageNet is orders of magnitude greater than the cross-frequency RF data used to pre-train the CAE.** The ImageNet database is comprised of a million images, whereas the the real RF data used in pre-training the CAE included 60 samples/class (for a total of 660 samples) of 77GHz or 10 GHz Xethru data and 240 samples/class (for a total of 2,640 samples) of 24 GHz data. Thus, even though the classification accuracy using pre-training of VGG16 using ImageNet surpasses that of pre-training on cross-frequency real RF data with the CAE, the performance improvement of just 2% - 9% is not that much considering the huge difference in pre-training sample size. One reason for this is that even though at a different frequency, pre-training on data with the same phenomenology as the test set has great benefits.

The question then arises, whether or not the cross-frequency training loss might not be overcome by pre-training on large amounts of synthetic RF data as opposed to optical imagery. To examine this possibility, we discuss synthetic data generation using generative adversarial networks (GANs), next.

Table 3: CAE Classification Results

Pre-training	Fine Tuning	Testing	Testing Accuracy
77 GHz	77 GHz	77 GHz	91.5%
		24 GHz	22.5%
		10 GHz	18.9%
	24 GHz	24 GHz	74.4%
	10 GHz	10 GHz	75.5%
24 GHz	24 GHz	24 GHz	91.2%
		77 GHz	28.5%
		10 GHz	40.5%

Pre-training	Fine Tuning	Testing	Testing Accuracy
24 GHz	77 GHz	77 GHz	83.8%
	10 GHz	10 GHz	81.6%
10 GHz	10 GHz	10 GHz	91.8%
		77 GHz	24.1%
		24 GHz	18.8%
	77 GHz	77 GHz	80.0%
	24 GHz	24 GHz	79.1%

Table 4: CNN Classification Results

Training	Testing	Accuracy
77 GHz	77 GHz	91.5
	24 GHz	27.4
	10 GHz	20.4
24 GHz	77 GHz	28.7
	24 GHz	91.5
	10 GHz	40.6
10 GHz	77 GHz	13.1
	24 GHz	32.7
	10 GHz	91.2

3.3 Data Synthesis with Auxilliary Conditional GANs (ACGANs)

In general, the architecture of GANs consists of two competing neural networks playing a minmax game. The generator network samples a predefined latent space and upsamples via transposed or deconvolutional layers to produce a synthetic image. The discriminator network takes in images as input and attempts to classify them as being real or fake. Both networks are linked through a combined loss function,

$$\min_{G} \max_{D} = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_{z}z}[\log(1 - D(G(z)))] \tag{1}$$

where $D(x)$ is the discriminator's estimate of the probability that the real instance x is real; $G(z)$ is the generator's output when given noise z; and $D(G(z))$ is the discriminator's estimate of the probability that a fake instance is real. This loss function is calculated from the discriminator's output, and seeks to minimize the generator's loss while maximizing the discriminators. As a result, the generator gets better at generating images that look real, and the discriminator gets better at classifying images as real or fake. In the ideal case the loss converges to 0.5, which represents a point where the discriminator can not tell synthetic and real images apart.

In an Auxiliary Conditional GAN (ACGAN),[36] the discriminator has an additional parallel network which attempts to classify images as their respective classes and allows for automatic separation of generated images. In this way each class of generated images can be studied separately and easily used to augment the existing training dataset. The generator and discriminator architecture used in this work is shown in Figures 4 and 5.
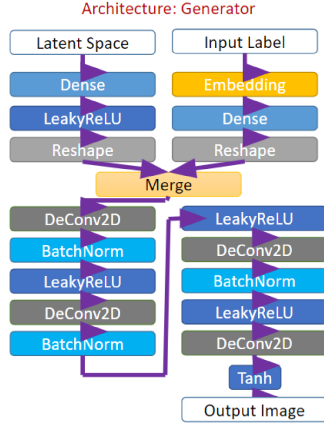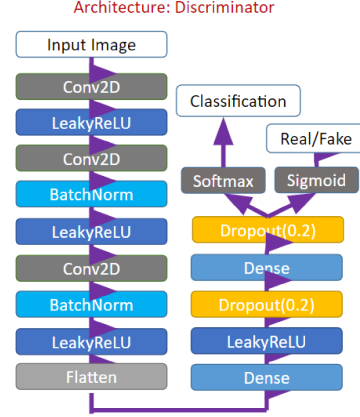
Figure 4: Generator Network Architecture

Figure 5: Discriminator Network Architecture

3.3.1 PCA-Based Sifting

When analyzing spectrograms of human motions, it is possible to infer certain information from the signal's shape and intensity. For example, when someone is walking towards the radar antenna the spectrogram will have overwhelmingly positive components and an increasing intensity as the person gets closer, which can be seen in Figure 3. Therefore, a spectrogram with largely negative frequency components would not be able to belong to the 'walking towards the radar' class, as it does not match the kinematics of the movement itself. Other examples of kinematic errors in synthetically generated data include disconnected components, leakage between target signature and sensor or clutter artifacts, blank space in the middle of the signature corresponding to stopping, and a blurring of normally distinct periodicities in the signature. When generating synthetic images, the inherent randomness of the generator's latent space input can lead to the generation of some kinematically unsound images, like the examples mentioned above. These kinematically inconsistent signatures can degrade classification performance when augmenting a training dataset. To maximize the potential of our data augmentation, a selection criteria needs to be in place for filtering out deviant samples.[30] Towards this aim, principal component analysis (PCA) is used to extract the feature space for each signature.

PCA is linear transform which is often used in data compression and as a machine learning (ML) classification method. The transform aims to reduce a collection of data samples into a set of principal components, each of which represents a varying level of the original data's variance. The first component contains the most variance, second has the second most, and so on. Because the majority of information in spectrograms comes from the sections with the most variance, it is possible to represent the majority of an image with a much smaller subset of values. Before processing, all data was standardized to have zero mean and scaled variance. By running 3 component PCA on the the real data for each class separately, we generated a collection of reduced feature points representing kinematically viable motions. Then a Delaunay Convex hull[37] was created in 3D space for each class' set of PCA components. The last step in the process involved running PCA on our generated synthetic data and checking if it falls within the convex hull boundaries for that given class. If it does not, it is considered to be unsound, and it is not saved for later augmentation. Of 10,000 images synthesized using the ACGAN, the PCA-based sifting algorithm rejected 4,000 of them, leaving a final synthetic dataset size of 6,000 images spread out among 11 classes with approximately 600 samples per class. Some sample synthetic images and their corresponding class' Delauney hulls are shown in Figure 6.

3.3.2 Cross Frequency Training with Synthetic Data

To evaluate the cross-frequency training degradation incurred when a large database of ACGAN-synthesized data is used for pre-training, a CAE was used to compare classification accuracies. First, our CAE model was pre-trained on 77 GHz synthetic data for 100 epochs. Next, fine-tuning was done using real training data from each sensor separately for 20 epochs. Lastly, the model was tested using all 3 frequencies of data.
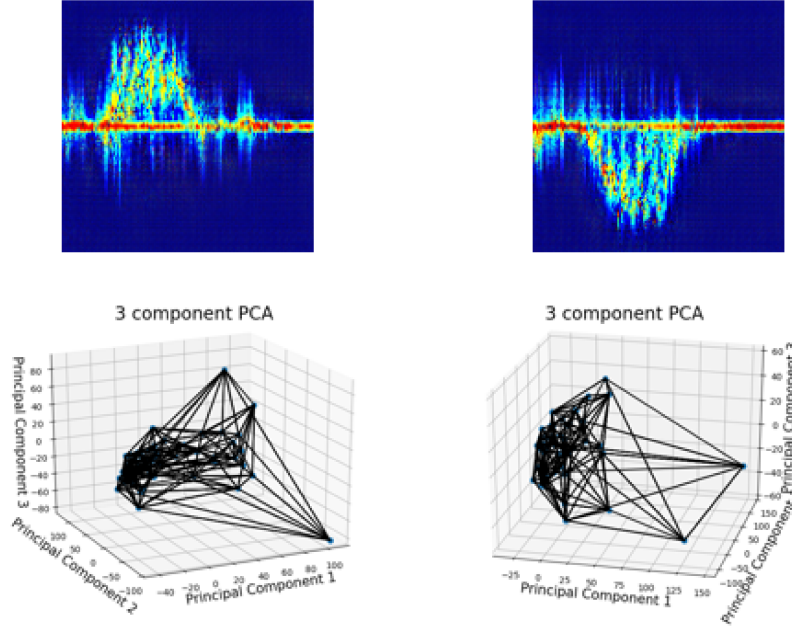
Figure 6: Generated Images and Their Respective Class' Delauny Hulls: Class 1 (left) and Class 2 (right)

Prior to fine tuning, the model gave a 53.85% accuracy for 77 GHz data, 11.55% for 24 GHz, and 4.80% accuracy for the 10 GHz test data. After fine tuning, the 77 GHz test dataset gave an accuracy of 85.4%, which is comparable with previous results using training with real data. Table 5 shows all cross-frequency training results obtained for the case of pre-training on synthetic 77 GHz micro-Doppler signatures. As before, discrepancies between the data used for fine-tuning and testing result in over %70 drop in classification accuracy. However, pre-training with a large amount of synthetic data has reduced the cross-frequency training loss when there is only a difference in frequency between pre-training and fine-tuning/testing. When pre-training on real 77 GHz data, fine-tuning and training on 24 GHz and 10 GHz data incurred about 16%-17% performance degradation relative to all the data being from the same sensor. In contrast, when pre-training on synthetic 77 GHz data, but fine-tuning and testing on 10 GHz and 24 GHz data, degradation of only 5% and 7% were incurred. Thus, increasing the amount of training data, even if synthetic, did serve to significantly reduce the performance degradation due to cross-frequency training, while also increasing the overall cross-training classification accuracy by 3%-5%.

However, notice that the overall classification accuracy for the case when all data is from the same sensor has been degraded. This is mostly likely due to mismatch between the distributions of the synthetic and real data that remain despite sifting. Improving the kinematic fidelity of GAN-synthesized data[30] could be one way of mitigating this problem.

Pre-Training	Fine-Tuning	Testing	Testing Accuracy
Synthetic 77 GHz	77 GHz	77 GHz	85.40%
		24 GHz	8.85%
		10 GHz	7.20%
	24 GHz	24 GHz	77.15%
	10 GHz	10 GHz	80.80%

Table 5: Cross-Frequency Training Results with Synthetic 77 GHz Data

4. CONCLUSION

This work presented a detailed study of the exploitation of "datasets of opportunity," such as data acquired from RF sensors with different frequencies, bandwidth, and waveform. The accuracies obtained from three different network initialization approaches were compared: randomly initializaed CNN, VGG16 pre-trained on ImageNet, and CAE pre-trained on real RF data. In cases when the RF frequency of the data used for fine-tuning differed from that used in testing, drastic drops in accuracy over 70% were observed irrespective of the data used in pre-training. However, when the same data was used for fine-tuning and testing, the degradation incurred from pre-training on RF data of a different frequency was observed to be as high as 15%. Generation of a large amount of synthetic training data using a PCA-sifted ACGAN reduced the degradation by 50% while also increasing the overall classification accuracies by about 3% - 5%. Future work in improving the kinematic fidelity of GAN-synthesized data could help in further increasing classification accuracies.

5. ACKNOWLEDGEMENTS

REFERENCES

[1] MG, A., "Radar for indoor monitoring: Detection, classification, and assessment," (2017).

[2] Mercuri M, Soh PJ, P. G. et al., "Analysis of an indoor biomedical radar based system for health monitoring.," *IEEE Transactions on Microwave Theory and Techniques.* **61(5)**, 2061–2068 (May 2013).

[3] Garripoli C, Mercuri M, K. P. et al., "Embedded dsp-based telehealth radar system for remote in-door fall detection.," *IEEE Transactions on Biomedical and Health Informatics* **19(1)**, 92–101 (January 2014).

[4] Su BY, Ho KC, R. M. et al., "Doppler radar fall activity detection using the wavelet transform.," *IEEE Transactions on Biomedical Engineering* **62(3)**, 865–875 (March 2014).

[5] Amin MG, Zhang YD, A. F. et al., "Radar signal processing for elderly fall detection: The future for in-home.," *IEEE Signal Processing Magazine* **33(2)**, 71–80 (March 2016).

[6] Seifert A, Zoubir AM, A. M., "Radar classification of human gait abnormality based on sum-of-harmonics analysis.," *In: Proc. IEEE Radar Conference* , 940–945 (2014).

[7] Palmer JW, Bing KF, S. A. et al., "Detecting concussion impairment with radar using gait analysis techniques," *In: Proc. IEEE Radar Conference* , 22–225 (2011).

[8] Postolache O, Pereira JMD, V. V. et al., "Gait rehabilitation assessment based on microwave doppler radars embedded in walkers.," *IEEE Int. Symp. on Medical Measurements and Appl.* , 208–213 (2015).

[9] Li C., Lubecke V., B. O. et al., "A review on recent advances in doppler radar sensors for noncontact healthcare monitoring.," *IEEE Transactions on Microwave Theory and Techniques.* **61(5)**, 2046–2060 (May 2013).

[10] Massagram W, Lubecke VM, H.-M. A. et al., "Assessment of heart rate variability and respiratory sinus arrhythmia via doppler radar.," *IEEE Trans on Microwave Theory and Techniques.* **57(10)**, 2542–2549 (October 2009).

[11] Hu W., Zhao Z., W. Y. et al., "Noncontact accurate measurement of cardiopulmonary activity using a compact quadrature doppler radar sensor.," *IEEE Transactions on Biomedical Engineering.* **61(3)**, 725–735 (March 2014).

[12] Rahman A, Lubecke VM, B. O. et al., "Doppler radar techniques for accurate respiration characterization and subject identification.," *IEEE J Emerging and Selected Top in Circuits and Sys.* **8(2)**, 350–359 (June 2018).

[13] DellAversano A., Natale A., B. A. et al., "Through the wall breathing detection by means of a doppler radar and music algorithm.," *IEEE Sensors Letters.* **1(3)**, 1–4 (June 2017).

[14] Lee YS, Pathirana PN, S. C. et al., "Monitoring and analysis of respiratory patterns using microwave doppler radar.," *IEEE Journal of Translational Engineering in Health and Medicine* **2**, 1–12 (2014).

[15] Ziganshin EG, Numerov MA, V. S., "Uwb baby monitor.," *5th Int. Conf. on UWB and Ultrashort Impulse Signals* **61(3)**, 159–161 (2010).

[16] Yang Y, Hou C, L. Y. G. D. H. D. and J, X., "Open-set human activity recognition based on micro-doppler signatures," *Pattern Recognition* **85**, 60–69 (January 2019).

[17] Chen S., Wang H., X. F. and YQ., J., "Target classification using the deep convolutional networks for sar images.," *IEEE Trans. on Geosci. and Rem. Sens.* **54(8)**, 4806–4817 (August 2016).

[18] Seyfioglu, M. S., Ozbayoglu, A. M., and Gurbuz, S. Z., "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems* **54**(4), 1709–1723 (2018).

[19] Seyfioglu, M. S. and Gurbuz, S. Z., "Deep neural network initialization methods for micro-doppler classification with low training sample support," *IEEE Geoscience and Remote Sensing Letters* **14**(12), 2462–2466 (2017).

[20] Park, J., Javier, R., Moon, T., and Kim, Y., "Micro-doppler based classification of human aquatic activities via transfer learning of convolutional neural networks," *Sensors* **16**, 1990 (11 2016).

[21] Du, H., Jin, T., Song, Y., Dai, Y., and Li, M., "Efficient human activity classification via sparsity-driven transfer learning," *IET Radar, Sonar Navigation* **13**(10), 1741–1746 (2019).

[22] Khanna, R., Oh, D., and Kim, Y., "Through-wall remote human voice recognition using doppler radar with transfer learning," *IEEE Sensors Journal* **19**(12), 4571–4576 (2019).

[23] Alnujaim, I., Alali, H., Khan, F., and Kim, Y., "Hand gesture recognition using input impedance variation of two antennas with transfer learning," *IEEE Sensors Journal* **18**(10), 4129–4135 (2018).

[24] Erol, B. and Gurbuz, S. Z., "A kinect-based human micro-doppler simulator," *IEEE Aerospace and Electronic Systems Magazine* **30**(5), 6–17 (2015).

[25] Seyfioglu MS, Gurbuz SZ, E. B. et al., "DNN transfer learning from diversified micro-doppler for motion classification," *IEEE Transactions on Aerospace and Electronic Systems.* , 1–6 (2018).

[26] Gurbuz, S. Z. and Amin, M. G., "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine* **36**(4), 16–28 (2019).

[27] Erol, B., Gurbuz, S., and Amin, M., "GAN-based synthetic radar micro-doppler augmentations for improved human activity recognition.," *IEEE Radar Conferences.* , 1–6 (2019).

[28] Alnujaim, I., Oh, D., and Kim, Y., "Generative adversarial networks to augment micro-doppler signatures for the classification of human activity," in [*IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*], 9459–9461 (2019).

[29] Doherty, H. G., Cifola, L., Harmanny, R. I. A., and Fioranelli, F., "Unsupervised learning using generative adversarial networks on micro-doppler spectrograms," in [*2019 16th European Radar Conference (EuRAD)*], 197–200 (2019).

[30] Erol, B., Gurbuz, S. Z., and Amin, M. G., "Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures," *IEEE Transactions on Aerospace and Electronic Systems* , 1–1 (2020).

[31] Fioranelli, F., Shah, S. A., Li, H., Shrestha, A., Yang, S., and Kernec, J. L., "Radar signatures of human activities," (2019).

[32] Shrestha, A., Murphy, C., Johnson, I., Anbulselvam, A., Fioranelli, F., Le Kernec, J., and Gurbuz, S. Z., "Cross-frequency classification of indoor activities with dnn transfer learning," in [*2019 IEEE Radar Conference (RadarConf)*], 1–6 (2019).

[33] Vishwakarma S, S. S., "Dictionary learning with low computational complexity for classification of human micro-dopplers across multiple carrier frequencies," *IEEE Access.* **6**, 29793 – 29805 (2018).

[34] Bengio Y., Guyon G., D. V. o., "Deep learning of representations for unsupervised and transfer learning.," *International Conference on Unsupervised and Transfer Learning Workshop* **277**, 17–37 (2011).

[35] Karen, S. and Andrew, Z., "Very deep convolutional networks for large-scale image recognition," *ICLR* (2015).

[36] Odena., A. and Others, "Conditional image synthesis with auxiliary classifier gans," *International Conference on Machine Learning* **4** (2017).

[37] de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M., [*Computational Geometry: Algorithms and Applications*], Springer-Verlag (March 2009).