Diversified Radar Micro-Doppler Simulations as Training Data for Deep Residual Neural Networks

Mehmet S. Seyfioglu[1], Baris Erol[2], Sevgi Z. Gurbuz[3], and Moeness G. Amin[2]
[1]Dept. of Electrical-Electronics Eng., TOBB Univ. Economics and Technology, Turkey (msseyfioglu@etu.edu.tr)
[2]Center for Advanced Communications, Villanova University, USA (berol,moeness.amin@villanova.edu)
[3]Dept. Electrical and Computer Engineering, University of Alabama, Tuscaloosa, USA (szgurbuz@ua.edu)

***Abstract*—A key challenge in radar micro-Doppler classification is the difficulty in obtaining a large amount of training data due to costs in time and human resources. Small training datasets limit the depth of deep neural networks (DNNs), and, hence, attainable classification accuracy. In this work, a novel method for diversifying Kinect-based motion capture (MOCAP) simulations of human micro-Doppler to span a wider range of potential observations, e.g. speed, body size, and style, is proposed. By applying three transformations, a small set of MOCAP measurements is expanded to generate a large training dataset for network initialization of a 30-layer deep residual neural network. Results show that the proposed training methodology and residual DNN yield improved bottleneck feature performance and the highest overall classification accuracy among other DNN architectures, including transfer learning from the 1.5 million sample ImageNet database.**

I. Introduction

Deep neural networks (DNNs) have recently received great interest in radar applications, such as static object recognition, synthetic-aperture radar (SAR) target classification, and micro-Doppler based human activity [1]-[5], and hand gesture recognition [6]. A common challenge in all these applications is that the datasets available are typically quite limited, often on the order of a thousand samples or less in studies involving human activity or gesture recognition. This is primarily because collecting radio frequency (RF) measurements requires the test subject to repeat each trial separately to acquire independent samples - a process that can be time consuming and costly for operations and human resources. Moreover, the limitation in training sample size also limits the depth that can be attained with DNNs without suffering from overfitting. In fact, depth has been shown to be the single most important parameter influencing performance in feedforward neural networks [7].

This work aims to pave the way for the design of deeper, more accurate DNNs by proposing a novel method for generating a large, diversified training database of simulated signatures that improves the initialization of deep neural networks. The use of simulated signatures to classify real micro-Doppler measurements was first proposed in 2015 [8], where motion capture (MOCAP) data made available from the Carnegie Mellon University (CMU) Motion Capture Library [9] was used to classify measured data with only a 1% difference in performance as when only real data was used for training. More recently, MOCAP data from low-cost devices, such as Kinect, have also been utilized to simulate human radar returns [10] [11]. However, because these MOCAP simulations are tied to obtaining real infrared and optical sensor data from test subjects, as with radar data, the size of the dataset is still limited.

There are a number of methods that may be employed to mitigate overfitting in DNNs, such as regularization and dropout, which are routinely applied but insufficient, as well as unsupervised pre-training. Typically, convolutional neural networks (CNNs) are trained from scratch through random initialization. But, the objective function of CNNs is highly non-convex, meaning that the parameter space of the model contains many local minima. Random initialization may not lead, therefore, to a local minimum that is optimal in the global sense. Unsupervised pre-training tries to reach a better local minima by creating an autoencoder model in which unlabeled data is provided to an encoder-decoder pair to greedily pre-train the network. Afterwards, the decoder part is removed, and the encoder, which has now been initialized, is fine-tuned using supervised training with a small dataset. This approach can be combined with convolutional layers to yield a convolutional autoencoder (CAE) [12] [13].

Another approach is to use transfer learning, in which the network is initialized using the weights of a model trained using data from another 'related' problem. Transfer learning has been proposed for applications in SAR image classification [14], and moving target recognition [15]. More recently, Park, et al. [16] used transfer learning from AlexNet and VGGnet, pre-trained with the ImageNet dataset, comprised of 1.5 million RGB images, to classify 5 different types of swimming. A classification accuracy of 80.3% was achieved from an overall micro-Doppler dataset of just 625 measurements. Transfer learning has been shown to be more effective than CAEs for extremely small training datasets [17], but its overall classification accuracy has been limited by the difference in underlying phenomenology between RF and optical data.

This paper bridges this gap by proposing a novel method for generating a large number of simulated signatures from a small number of MOCAP measurements. Three transformations are employed to increase in-class variance: 1) scaling of the time axis, to emulate slower or faster motion; 2) scaling of the underlying skeletal dimensions, to emulate smaller or larger body size, and 3) applying randomized perturbations to a parametric model of the kinematic trajectories of each body part to emulate individualized gait style. Although it is possible that these transformations may result in some physically impossible animations, we utilize a two-step approach to training the DNN that allows for compensation of such discrepancies. In the first step, the DNN is trained using the diversified simulated dataset; however, afterwards in the second step these initial weights are fine-tuned by using supervised training on a small number of real measurements. Residual learning is used as a means for overcoming the degradation in training accuracy that is incurred as depth increases.

Our results show that the proposed approach not only enables the design of deeper networks, but is also able, through the added in-class variance, to better generalize to generic targets never before seen by the radar, and can even facilitate class generalization. Improved generalization is a feature critical to the development of ground surveillance radar (GSR) for border control and security. Ultimately, the proposed approach will improve accuracy and make fielding of micro-Doppler based GSR easier, less costly, and more feasible.

The paper is organized as follows. In Section II, the methodology for the diversified human micro-Doppler simulations is described in detail. The measured data used as the test set is presented in Section III. In Section IV, the proposed DivNet architecture and other transfer learning methods VGGnet and ResNet are described in detail. In Section V, the performance of the proposed architecture, DivNet, is contrasted with that of a randomly initialized baseline CNN and with VGGnet and ResNet. Finally, in Section VI, key conclusions are presented.

II. Simulated Radar Micro-Doppler Database

There are two main approaches to simulating human micro-Doppler signatures [11]: kinematic modeling and MOCAP-based animation. Both methods revolve around the idea of decomposing the human body into a finite number of parts modeled as point targets, and summing the radar returns from each point target [18], as modeled from the radar range equation. However, unlike kinematic models, which typically encompass just one activity, MOCAP-based animations can be used to simulate almost any motion, and have thus gained in prevalence for micro-Doppler simulations.

A. Kinect-Based Micro-Doppler Simulator

In this work, the Kinect sensor is used as a markerless system for capturing the time-varying coordinate information of human joints needed for simulation of human micro-Doppler signatures. First, the radar return from the human body is represented as the sum of reflected signals from a finite number (K) of point targets representing various body parts. The Kinect measurements are used in lieu of the time-varying range measurements typically obtained from radar. Mathematically, the return signal from K point targets for a continuous wave (CW) radar is

$$s_h(t) = \sum_{i=1}^{K} a_{t,i} e^{-j[(2\pi f_0)t + \frac{4\pi}{\lambda} R_{t,i}]}, \qquad (1)$$

where f_0 is the transmit frequency, λ is the wavelength, t is time, $R_{t,i}$ is the time-varying range of each point target, as captured from the Kinect sensor, and $a_{t,i}$ is the amplitude as computed from the radar range equation.

Once the required ranges are estimated from the Kinect data, (1) can be computed for any human activity or radar parameters, such as center frequency, bandwidth, and sampling frequency. Finally, the simulated micro-Doppler signatures are computed as the spectrogram (S) – modulus squared of short-time Fourier Transform (STFT) – of the radar return:

$$S = |STFT(n,\omega)|^2 = \left| \sum_{m=-\infty}^{\infty} s[n+m] w[m] e^{-j\omega m} \right|, \quad (2)$$

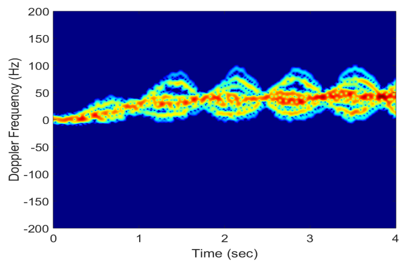
where n is discretized time, and $w[m]$ is a window function. In this work, the data was sampled at 2.4 kHz, while a Hanning window of length 256 and 128 overlap samples with 1024 total frequency points is applied in computing the spectrogram. Spectrograms were generated at 15 GHz and cropped to generate images that showed the internal structure of the micro-Doppler as clearly as possible. Figure 1 shows the resulting simulated micro-Doppler signatures for seven different activity classes.

B. Diversification Methodology for μD Signatures

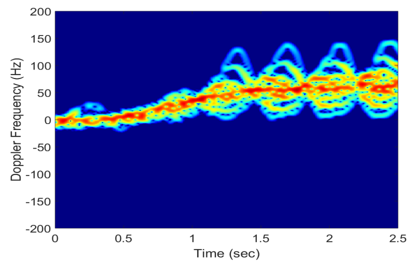
In the Kinect-based radar micro-Doppler simulator, the 3-D coordinate measurements of 17 joints acquired from the Kinect sensor are used. By changing this coordinate information, it is possible to form a large activity database with sufficient intra and inter class variations that emulates the diversity of human signatures caused by differences in height, speed and individual gait.

1) Height and Speed Modifications: The Kinect-based radar simulator permits modification of subject height and speed by scaling the time-varying joint position data along different axes. For example, scaling along the z-axis, while keeping the x and y axes unchanged, modifies the subject height which results in different dimensions of body parts. These changes influence the radar cross sections (RCS) computations of individual point scatterers, thus also affecting the received signal power. Due to the fact that scaling is performed only along the z-axis, the stride rate, speed, and style of the motion remain unmodified in the first step of the diversification methodology. However, in real scenarios, a tall person typically walks or runs faster than a shorter person, when all other body structure factors assumed equal, such as body mass, flexibility, and proportionality. Therefore, the y-axis position data (direction of motion) is also scaled by a small amount relative to the change in height to alter the speed of the subject.

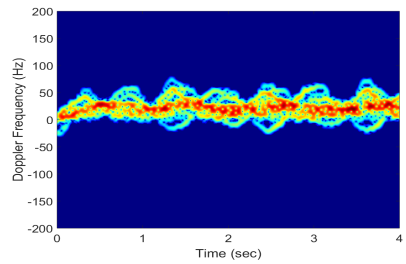
Snapshots from animations derived from the same subject, but with different heights, are provided in Figures 2-(a) and 2-(b). In Figure 2-(a), the subject's height is scaled down to 1.55 m, while in Figure 2-(b), the height is scaled up to 1.9 m. Corresponding micro-Doppler signatures are depicted in



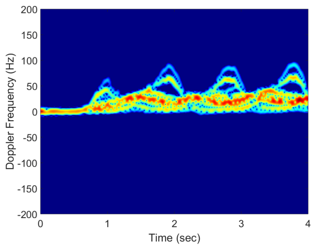
(a) Walking



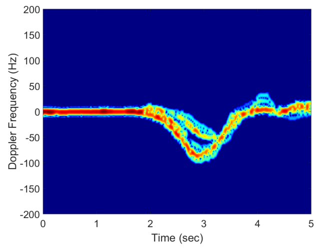
(b) Running



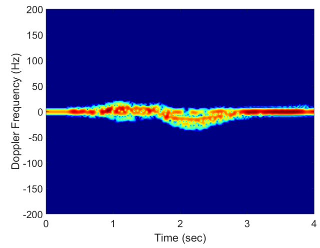
(c) Limping



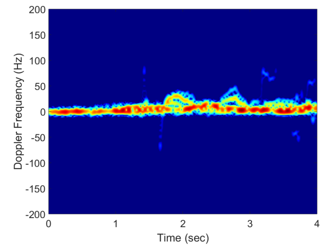
(d) Cane



(e) Falling



(f) Sitting



(g) Walker

Fig. 1: Kinect-based micro-Doppler signatures for a CW radar

Figure 2-(e) and (f) for the same deviations. From micro-Doppler signatures, it may be seen that reflections from the right and left foot are more visually distinguishable for a taller subject due to the increased lower leg dimensions. Moreover, speed of the entire motion is relatively increased due to scaling along y-axis relative to the height change.

A second parameter that greatly influences the micro-Doppler signature is the speed of the subject. Completing this task in Kinect-based MOCAP data can be challenging due to fact that the acquired position data already contains a speed factor. Therefore, another operation along y-axis is required with the manipulation of sample frequency of the raw Kinect data which changes the stride rate and speed of the motion. In Figure 2-(c) and (d), animations derived from one sample are provided for two different speeds. It is evident that a faster subject travels longer distance than a slower subject within the same time interval. The micro-Doppler signatures are also depicted in Figure 2-(g) and (h) for the same derivations. Note that faster subject exhibits shortened cycles than slower subject within the same time interval.

2) Parameterization of Individual Joints: The last step of the diversification methodology complements scaling and focuses on the individual joint data, such as left and right leg, right and left arm, and head. The main idea is to parameterize the Kinect raw range data of the different joints separately. Then, by just perturbing the coefficients of the constructed models, it is possible to create class variations. Limited alterations of the model parameters manifest itself in the style of how the motion is performed.

Parameterization of the joint can be done in several ways, most easily using curve fitting models. Given the periodic nature of the joint data and by examining the goodness of the constructed model, the Fourier series was determined to be the best suited parameterization model for the underlying problem. This model also provides a good fit for non-period range trajectories, which are mostly encountered in non-rhythmic motions, such as falling and sitting. The Fourier series model describes the given Kinect range data as a sum of sine and cosine functions. Resulting model can be represented in the trigonometric form as

$$y = a_0 + \sum_{i=1}^{n} a_i \cos(iwx) + b_i \sin(iwx), \tag{3}$$

where a_0 models a constant term in the data and is associated with the cosine term for $i = 0$, w is the fundamental frequency of the signal, and n $(0 < n < 9)$ is the number of terms (harmonics). This model provides $2n$ coefficients, which we can alter depending on the number of harmonics used. To prevent demolishing the joint trajectory entirely, and preserve the underlying information of the joint data, we only changed the n-dual harmonic coefficients $[(a_1, b_1), ..., (a_n, b_n)]$, one pair at a time, and within a 10% range value. Some constraints are also imposed to make the methodology more consistent with the kinematics of different motions. For example, when a harmonics's dual coefficient pair is changed for different joints, the algorithm automatically compares the altered arm and leg lengths with the original lengths to determine whether alteration is kinematically possible.

III. Experimental Test Dataset

The test set used in this work comprises entirely real radar measurements conducted in an indoor laboratory at TOBB University. An NI-USRP 2922 model software-defined radio, mounted 1 meter above the ground, was used to transmit a 4 GHz CW signal. Two SAS-571 horn antennas were placed on

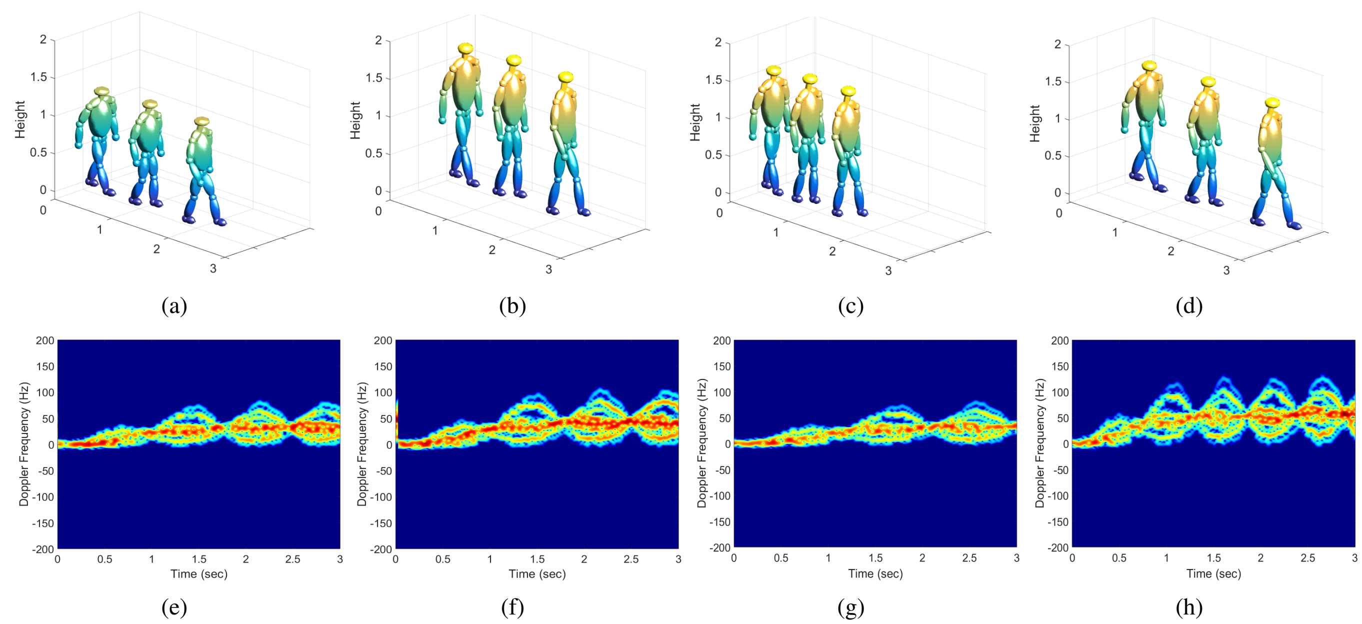


Fig. 2: Kinect-based animation results derived from one data for (a) A short subject (b) A tall subject (c) Subject with a slow stride rate, (d) Subject with a fast stride rate and resulting micro-Doppler signatures for a CW radar (e) Short (f) Tall (g) Slow, and (h) Fast

either side of the USRP to yield an approximately monostatic configuration. Test subjects conducted seven different human activities at a range from the radar varying between 1 meter and 5 meters. All experiments were conducted in alignment with the radar line-of-sight. The micro-Doppler signature for each measurement was computed for a total duration of 4 seconds.

Two datasets were formed for testing the proposed transfer learning approach: a 7-class data set, and an 11-class dataset. For the 7-class dataset, the activities enacted along with total number of measurements per class are as follows: walking (71), jogging (72), limping (104), walking while using a cane (123), walking while using a walker (121), falling forward from a standing position (53), and sitting (50). This data set was augmented with data from four additional classes to form the 11-class data set; namely, creeping (56), crawling (74), using a wheelchair (149), and walking with crutches while one leg is bent at the knee (74).

Micro-Doppler signatures are represented as spectrograms, computed using a hamming window with length of 2048 samples, 4096 FFT points, and 128 samples overlap. Each spectrogram was then cropped to a duration of 4 seconds, converted to grayscale and saved as an image. To reduce dimensionality, the resulting images were then downsampled from a size of 656x875 pixels to 90x120 pixels.

IV. DivNet: Residual Transfer Learning with Diversified Simulations

Transfer learning is a technique that has been proposed for domain adaptation when a little or no labeled data is available relating to the targeted classes. The network is first initialized using labeled data from a different source domain; then, the small set of task-related labeled data is used to fine tune initialized parameters prior to classification of the test set. VGGnet is a 16-layer CNN, which has been highly successfully in classification of the ImageNet database, as well as in classification RF datasets using transfer learning [16].

Increasingly deep networks have improved the state-of-the-art in image classification, with more recently the 152-layered ResNet architecture surpassing the performance of VGGNet on the ImageNet database. Assuming sufficient training sample support to prevent overfitting, very deep networks also suffer from a degradation of training accuracy that increases with depth. Residual learning [19] has been proposed as a method for overcoming this problem, and has also been shown to be more easily optimized with gains in accuracy due to increased depth. Deep residual networks are comprised of building blocks which rather than computing the original mapping of $y_l(x) := h(x_l)$, compute the residual mapping of

$$y_l = h(x_l) + F(x_l, W_l), \tag{4}$$

$$x_{l+1} = f(y_l), \tag{5}$$

where x_l is the input to the l^{th} residual unit (RU), $W_l = W_{l,k|1 \leq k \leq K}$ is a set of weights and biases for the l^{th} RU, K is the number of layers in a RU; F is a residual function, e.g. a stack of two convolutional layers; $h(x_l) = x_l$ is an identity mapping that is performed by using a shortcut path, and f is an activation function, in this work, a ReLU. In addition, ResNet contains batch normalization (BN) layers where batches are standardized (zero mean/unit variance) throughout the network to reduce the interval co-variance shift, which allows larger learning rates to be used [20]. Shortcut connections can thus be created within the network by simply implementing an *identity* mapping. By driving subsequent layers with the residual (i.e., the input x as well as output of previous layer), the network is forced to effectively learn something new beyond that already embodied by previous layers. Thus, another advantage of residual learning is that very deep networks can be constructed without worrying about whether the network is "too deep" - if adding layers gives no benefit, residual blocks can learn the identity mapping and thus do no harm to performance.

We propose classification with a 30-layer residual transfer network in which the transfer domain compromises simulated micro-Doppler signatures that have been diversified to span a wider range of potential human motion within each activity class. A modified residual unit (RU) is implemented, in contrast to that implemented by ResNet. This network structure has been reported to be more easily trained and has improved generalization properties [21]. In this case, the RU still comprises two convolutional layers, but the activation function is now placed before the final addition, as opposed to afterwards. The overall residual DNN architecture, which we have named DivNet, includes 3 residual units and is comprised of a total of 30 layers, as shown in Figure 3.

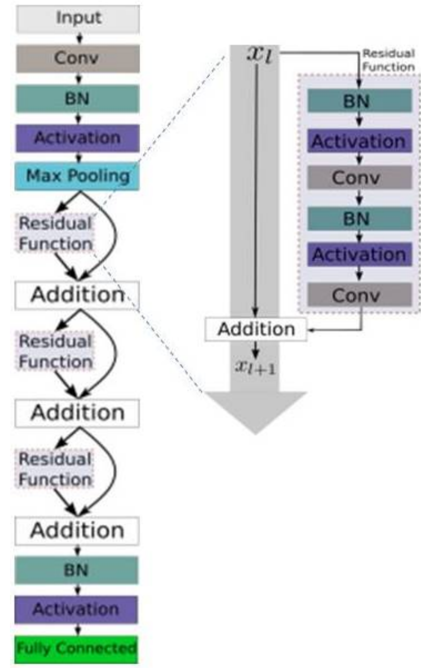


Fig. 3: Proposed DivNet architecture

V. Experimental Results

In this section, the performance of the proposed residual transfer learning network, DivNet, trained on diversified simulations of micro-Doppler signatures, is compared against transfer learning from ImageNet with VGGnet and ResNet-50, as well as a 4-convolutional layer, randomly initialized CNN with a convolutional filter size of 3x3, a ReLU activation function and two fully connected layers with 150 neurons in each layer. A 2x2 max pooling was used for all CNN depths except that with 11-layers. After the first fully connected layer, a dropout operation is applied with the probability of 50% to combat overfitting.

The diversified training database of simulated signatures is generated from 55 original Kinect-based MOCAP measurements made of 5 test subjects in Radar Imaging Laboratory at Villanova University and diversification methodology applied to increase the total sample size to 32,000 signatures. For the 7-class problem, results are shown for DivNet trained on these 32,000 samples, as well as VGGnet and ResNet-50, trained on the 1.5 million optical images of ImageNet. The baseline CNN is trained on 474 measured signatures, while these same measured signatures are also used for supervised fine-tuning of the transfer learning networks. All methods are tested on 120 measured micro-Doppler signatures.

A confusion matrix performance for the 7 class scenario given by DivNet is shown in Table I. The primary source of misclassification is observed to be between the classes of using a cane versus a walker, with about 15% confusion. This is not surprising as these two classes are most similar, and because the current simulations of micro-Doppler include only reflections due to the human body, not due to walking aides.

In some cases, simulations of all target classes may not be available. In this part, we test the ability of the proposed method to facilitate class generalization: i.e., 7-classes of diversified simulated micro-Doppler signatures are available, but we desire to classify 11-classes of measured radar data. To proceed, we apply our proposed DivNet network by initializing with 7-classes of diversified data, and then fine-tuning with 11-classes of measured data. A confusion matrix performance for the 11-class scenario given by DivNet is shown in Table II. This is compared to applying VGGnet, ResNet, and the baseline 4-convolutional layer CNN to the 11-class problem. The results comparing test accuracy and complexity are given in Table III.

While the baseline CNN did not show significant deviation in performance between the 7-class and 11-class cases, transfer learning methods did show a drop of several percent in performance. Nevertheless, all transfer learning methods surpassed that of the baseline CNN, while the proposed DivNet method maintained the highest performance of all methods with a test accuracy of 95.7%.

VI. Conclusion

This work considers in detail the design of deep neural networks under small training sample support, which is typical in RF applications. A method for diversifying motion-capture based simulation of human motion is proposed to generate an arbitrarily large training database of micro-Doppler signatures that better capture variations in human motion due to height, speed, and individual gait. It is shown that training on diversified simulated signatures using a residual transfer learning network, DivNet, yields results that surpass transfer learning from the domain of optical imagery, such as ImageNet. Moreover, the proposed diversification methodology enables a degree of class generalization, as well as generalization to previously

TABLE I: Confusion matrix for DivNet with overall test accuracy of 97%.

%	Walking	Jogging	Limping	Falling	Sitting	Cane	Walker
Walking	100	0	0	0	0	0	0
Jogging	0	100	0	0	0	0	0
Limping	0	0	97.2	0	0	2.8	0
Falling	0	0	0	100	0	0	0
Sitting	0	0	0	0	100	0	0
Cane	0	0	0	0	0	96.4	3.6
Walker	0	0	0	0	0	15.41	84.59

TABLE II: Confusion matrix for DivNet with overall test accuracy of 95.7%.

%	Walking	Jogging	Limping	Falling	Sitting	Cane	Walker	Crutches	Wheelchair	Crawling	Creeping
Walking	100	0	0	0	0	0	0	0	0	0	0
Jogging	0	100	0	0	0	0	0	0	0	0	0
Limping	0	0	100	0	0	0	0	0	0	0	0
Falling	0	0	0	100	0	0	0	0	0	0	0
Sitting	0	0	0	0	100	0	0	0	0	0	0
Cane	0	0	0	0	0	100	0	0	0	0	0
Walker	0	0	0	0	0	13.14	86.86	0	0	0	0
Crutches	0	0	0	0	0	0	0	100	0	0	0
Wheelchair	0	0	4	0	0	0	0	0	96	0	0
Crawling	0	0	0	0	0	0	0	0	0	88.4	11.6
Creeping	0	0	0	0	0	0	0	0	0	18.47	81.53

TABLE III: Performance comparison for 7 and 11 classes

DNN	D	T/E	Init.	F.T.	7-cl.	11-cl.
DivNet	30	1 s	32k Sim μD	real μD 7-cl.: 474 11-cl.: 757	97%	96%
ResNet	50	20 s	ImageNet		95%	91%
VGGnet	16	9 s	ImageNet		94%	91%
CNN	4	0.4 s	Random		86%	87%

D: depth; Sim: simulated; T/E: time/epoch; Init.: Initialization; F.T.: fine-tuning

unseen targets, as can be observed in that the subjects for the MOCAP measurements used in training are different from those used in the real radar measurements of the test set.

REFERENCES

[1] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE Geosci. and Remote Sensing Lett.*, vol. 13, no. 1, pp. 8–12, 2016.

[2] R. Trommel, R. Harmanny, L. Cifola, and J. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms," in *Radar Conference (EuRAD), 2016 European*. IEEE, 2016, pp. 81–84.

[3] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *Radar Conference (RadarConf), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[4] M. S. Seyfioglu, S. Z. Gurbuz, A. M. Ozbayoglu, and M. Yuksel, "Deep learning of micro-doppler features for aided and unaided gait recognition," in *2017 IEEE Radar Conference (RadarConf)*, May 2017, pp. 1125–1130.

[5] J. Kwon and N. Kwak, "Human detection by neural networks using a low-cost short-range doppler radar sensor," in *2017 IEEE Radar Conference (RadarConf)*, May 2017, pp. 0755–0760.

[6] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, May 2015, pp. 1–8.

[7] R. Eldan and O. Shamir, "The power of depth in feedforward neural networks," in *Proc. Conference on Learning Theory*, June 2016, pp. 907–940.

[8] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson, "Knowledge exploitation for human micro-doppler classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2125–2129, Oct 2015.

[9] Carnegie melon university motion capture library. [Online]. Available: http://mocap.cs.cmu.edu/

[10] B. Erol and S. Z. Gurbuz, "A kinect-based human micro-doppler simulator," *IEEE Aerospace and Electronic Systems Magazine*, vol. 30, no. 5, pp. 6–17, May 2015.

[11] S. S. Ram, S. Z. Gurbuz, and V. C. Chen, "Modeling and simulation of human motions for micro-doppler signatures," in *Radar for In-Door Monitoring: Detection, Classification, and Assessment*, M. Amin, Ed. CRC Press, September 18, 2017.

[12] M. S. Seyfioglu and S. Z. Gurbuz, "Deep convolutional autoencoder for radar-based human activity recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. PP, no. 99, February 2018.

[13] K. N. Parashar, M. C. Oveneke, M. Rykunov, H. Sahli, and A. Bourdoux, "Micro-doppler feature extraction using convolutional auto-encoders for low latency target classification," in *2017 IEEE Radar Conference (RadarConf)*, May 2017, pp. 1739–1744.

[14] C. Kang and C. He, "Sar image classification based on the multi-layer network and transfer learning of mid-level representations," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 1146–1149.

[15] Z. Yang, J. Deng, and A. Nallanathan, "Moving target recognition based on transfer learning and three-dimensional over-complete dictionary," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5671–5678, July 2016.

[16] J. Park, R. J. Javier, T. Moon, and Y. Kim, "Micro-doppler based classification of human aquatic activities via transfer learning of convolutional neural networks," *Sensors*, vol. 16, no. 12, 2016.

[17] M. S. Seyfioglu and S. Z. Gurbuz, "Deep neural network initialization methods for micro-doppler classification with low training sample support," *IEEE Geoscience and Remote Sensing Lett.*, vol. 14, no. 12, December 2017.

[18] P. van Dorp and F. C. A. Groen, "Human walking estimation with radar," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 150, no. 5, pp. 356–365, Oct 2003.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 630–645.