

Radar Based Joint Human Activity and Agility Recognition via Multi Input Multi Task Learning

Emre Kurtoglu

Dept. of Electrical & Computer Eng.
The University of Alabama
Tuscaloosa, USA
ekurtoglu@crimson.ua.edu

Moeness G. Amin

Center for Advanced Communications
Villanova University
Villanova, USA
moeness.amin@villanova.edu

Sevgi Z. Gurbuz

Dept. of Electrical & Computer Eng.
The University of Alabama
Tuscaloosa, USA
szgurbuz@ua.edu

Abstract—Radar-based recognition of human activities of daily living has been a focus of research for over a decade. Current techniques focus on generalized motion recognition of any person and rely on massive amounts of data to characterize generic human activity. However, human gait is actually a person-specific biometric, correlated with health and agility, which depends on a person's mobility ethogram. This paper proposes a multi-input multi-task deep learning framework for jointly learning a person's agility and activity. As a proof of concept, we consider three categories of agility represented by slow, fast and nominal motion articulations and show that joint consideration of agility and activity can lead to improved activity classification accuracy and estimation of agility. To the best of our knowledge, this work represents the first work considering personalized motion recognition and agility characterization using radar.

Index Terms—radar micro-doppler, human activity recognition, deep learning, deep neural networks, spectrograms

I. INTRODUCTION

Radar-based human activity recognition (HAR) [1], [2] has been an important focus of research for applications relating to defense, security, indoor monitoring [3], biomedical [4]–[6] and human-computer interaction [7], [8]. Current approaches for deep learning-based HAR [9] focus on training models such that they achieve high generalization performance to previously unseen users - in other words, a generic human mobility profile. To achieve this, models are trained so as to span all possible participant profiles: height, speed, and style. The more data that embodies all possible variations in the resulting radar micro-Doppler signature [10], the better the generalization.

However, in fact, human gait is actually a person-specific, biometric characteristic [11]. Every person walks differently. Depending on age, health, and mobility, the way a person kneels, lays down, or stands up is even unique. There may be a significant difference between the pace and posture of a healthy teenager versus an elderly person with fall risk. Thus, remote health monitoring applications require characterization and recognition of *person-specific* mobility profiles - not generalization across a generic profile. Moreover, in health applications the person being observed is generally known: we do not seek to identify anyone, but to instead better understand a *particular* person's mobility.

Beyond simply recognizing the activity, extracting characteristics of the gait has health significance. For example, gait abnormalities [12] and various gait parameters [13], [14] can be indicative of fall risk or other disorders. Physicians often use evaluations such as the Timed Up and Go (TUG) test [15] to also measure agility and balance. This is significant because it shows that not just persistent activities, but also transitions *between* activities have implications for health monitoring.

In behavioral psychology, the term *ethogram* [16], [17] refers to a quantitative description of normal behavior of an animal or a person. Similarly, an ethogram of human mobility [18], [19] is comprised of a state diagram that documents a person's daily mobility patterns: each state is reflective of a posture, such as standing (in place), ambulatory, laying down, or sitting, while activities themselves can be defined as being persistent within a state, such as walking, or transitory in nature. For example, falling is a transitory activity where a person suddenly goes from a "walking" state to a "laying down" state. This view offers an enriched perspective with which to view activities, as well as a scheme within which the likelihood of certain transitions (based on health and mobility) and the rate of transition (agility) can be logged. Combined, the states, in-state and transitional activities, transition likelihoods and agility comprise the ethogram - a more comprehensive, person-specific description of a person's mobility profile.

This work represents the first work, to the best of our knowledge, to investigate the use of radar to jointly characterize agility in addition to transitional activity. In particular, as a proof of concept, we consider three categories of agility as represented by slow, fast and nominal motion articulations. We propose joint estimation of agility and activity using a multi-task [20] deep neural network with a cumulative loss function comprised of individual losses with adjustable importance factors. We show that the proposed approach improves the activity recognition performance while maintaining the agility score estimation accuracy with much fewer trainable parameters and memory requirements than if two separate DNNs were utilized. Additionally, we also assess the utilization of range information along with the Doppler information in a multi-input network. Towards this aim, this study proposes a multi-input multi-task learning (MIMTL) method for simultaneous recognition of human activity and agility. The proposed ap-

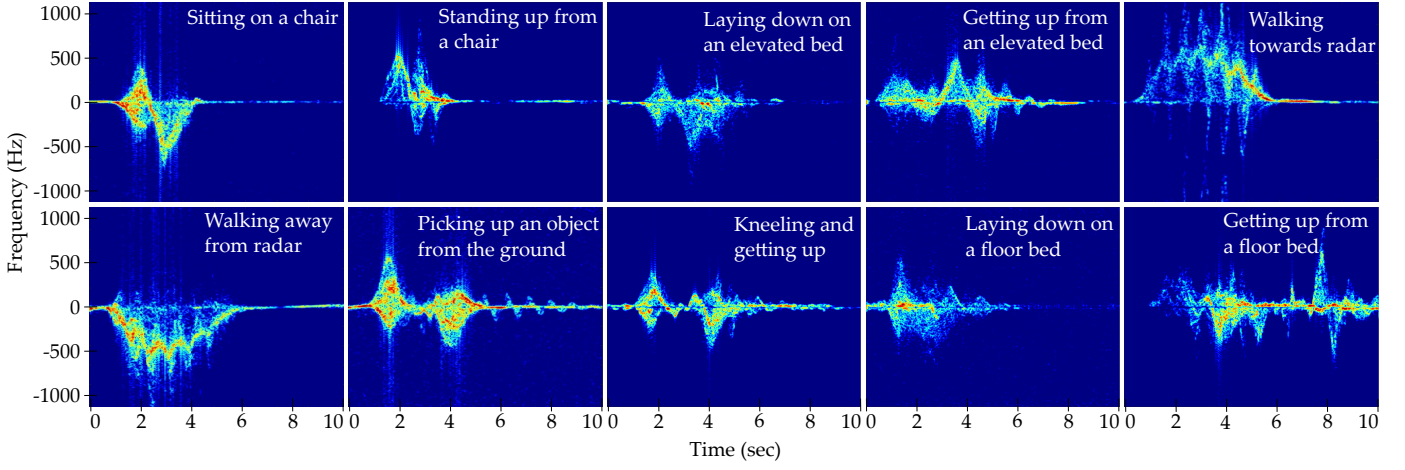


Fig. 1: μ D spectrograms of different activities.

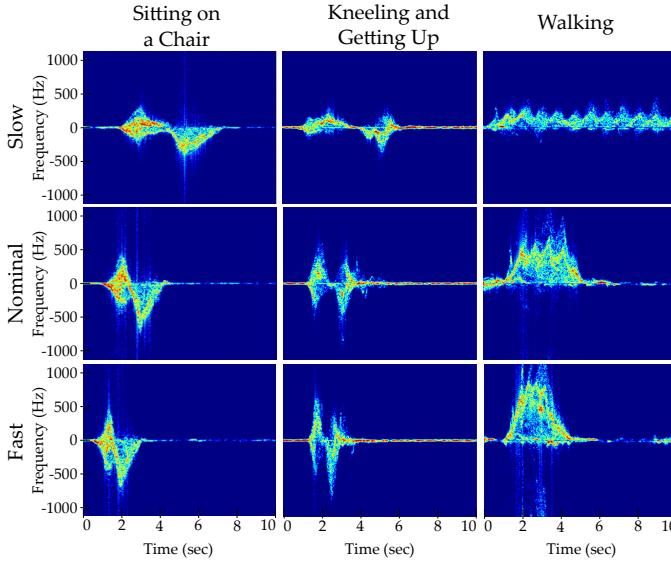


Fig. 2: μ D spectrograms of activities for varying speeds.

proach improves the activity recognition performance around 6% (from 89.7% to 95.6) and 3% (from 95.2% to 98.2%) for joint prediction and multi-task learning cases, respectively.

The remainder of the paper is organized as follows: Section II presents the experimental setup, data collection and pre-processing stages, Section III describes and benchmarks different approaches on human activity and agility score estimation, Section IV presents the proposed MIMTL approach for joint estimation of activities along with agility scores and compares it with other methods, and finally, Section V concludes the paper with principal findings.

II. DATA COLLECTION AND EXPERIMENTAL SETUP

A. Radar Sensor

In this work, Texas Instrument's AWR2243BOOST frequency-modulated continuous-wave (FMCW) multiple-input-multiple-output (MIMO) radar is used for data acqui-

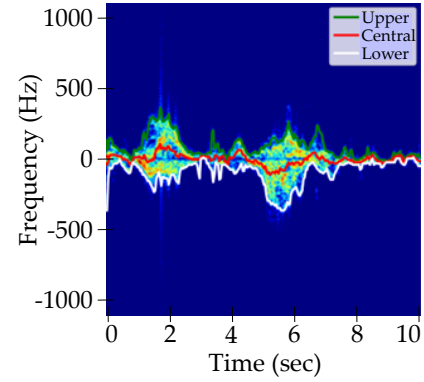


Fig. 3: Extraction of upper, central and lower envelopes from μ D spectrograms.

sition. The device operates at 77 GHz with 4 GHz bandwidth which yields <4 cm range resolution. The RF sensor is attached to a DCA1000EVM data capture card to stream the raw I&Q data. The RF antenna has $\pm 70^\circ$ azimuth and $\pm 15^\circ$ elevation beamwidths.

B. Experimental Setup and Data Acquisition

We consider ten human activities along with their three different agility scores to assess the recognition performance of the proposed approach. The defined activities and their abbreviations throughout the paper are listed below.

- **SITC**: Sitting on a chair
- **STDC**: Standing up from a chair
- **LAYB**: Laying down on an elevated bed
- **GETB**: Getting up from an elevated bed
- **WLKT**: Walking towards radar
- **WLKA**: Walking away from radar
- **PICK**: Picking up an object from the ground with a cane support
- **KNEE**: Kneeling and getting up
- **LAYF**: Laying down on a floor bed
- **GETF**: Getting up from a floor bed

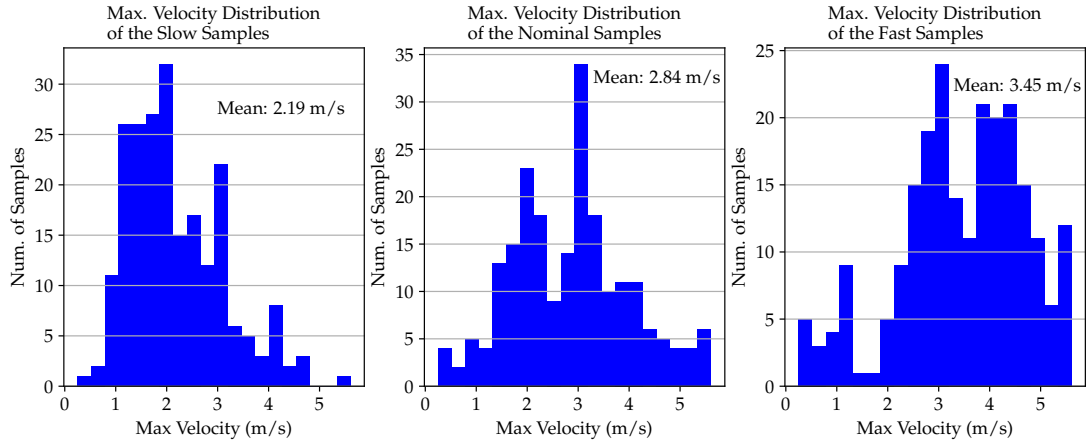


Fig. 4: Maximum velocity distributions of different agility scores.

Each activity is performed with three different speeds (i.e., agility scores): *slow*, *nominal* and *fast*. *Slow* activities are characteristic of the elderly population or persons with mobility restrictions, and performed by the participants consistent with these manners. Slow motion activities can also be a result of persons distracted by reading or carrying or holding items. *Nominal* activities are articulated by a wide age group and are performed with a normal speed. In this case, each participant performs the motion using his/her own personal pace. *Fast* activities, on the other hand, are typically exhibited with people in a hurry and are generally affiliated with the younger population. These activities are performed by the participants in a quick manner. While these agility scores are subject to the understanding of the participants about relative speed of execution of each activity as well as the interpretation of the radar operator who labels them, the participants are instructed to perform the activities with certain differences in pace. In Section II-D, distributions of the activity speeds are discussed in detail. Different speeds for each activity are denoted by preceding the activity name with the speed category's initial. For instance, *slow*, *nominal* and *fast* performing of SITC are denoted as S_SITC, N_SITC and F_SITC, respectively.

The radar was placed approximately 1 m above the ground and in-place motions were performed in the direct line-of-sight of the radar and at around 2 m distance while transient motions such as walking have varying distances. Each recording lasted for 10 seconds and 6 people with varying ages and genders participated the study. In total, 900 samples are acquired which later partitioned into 70% and 30% portions for training and testing, respectively.

C. RF Data Pre-Processing

The signal received by the radar receiver, $S_R(t)$, is a time-shifted and frequency-modulated version of the transmitted FMCW signal, $S_T(t)$, which can be mathematically expressed as:

$$S_T(t) = \exp(j2\pi f_c t + j\pi \alpha t^2) \quad (1)$$

where, f_c is the carrier frequency, α is the chirp rate defined as the ratio of bandwidth, B , to the sweep duration, T , as $\alpha = \frac{B}{T}$. Assuming the signal reflects back from a set of K targets, the received signal, $S_R(t)$, in one channel of the radar system can be expressed as:

$$S_R(t) = \sum_{i=1}^K A_i \exp(j2\pi f_c(t - t_i) + j\pi \alpha(t - t_i)^2) \quad (2)$$

where, A_i is a complex constant related to target's radar cross section, and t_i is the round trip time delay for the i^{th} target.

The raw I&Q data can be reshaped into a 3D-array named radar data cube (RDC) for further processing. First, a Fast Fourier Transform (FFT) is applied across analog-to-digital converter (ADC) samples for each chirp (i.e., pulse), providing the range information over time, also called range profile (RP). The activity time-frequency profiles, known as micro-Doppler signatures μD [10] are generated by performing windowed Fourier transform over slow-time or a number of chirp radar returns. This is known as spectrograms which are computed as the square modulus of the Short-Time Fourier Transform (STFT) of the continuous time input signal:

$$S(t, w) = \left| \int_{-\infty}^{\infty} h(t - u) x(u) e^{-jwu} du \right|^2 \quad (3)$$

where $h(\cdot)$ is the windowing function, $x(\cdot)$ is the received signal. Spectrograms are implemented by FFT. Figure 1 shows a sample μD spectrogram for each activity.

D. Agility Profiles

Varying speeds of performing an activity can appear in the μD spectrogram as compression or expansion behavior in time and frequency. Whereas *fast* movements exhibit expansion in time and compression in frequency, *slow* motions show the opposite. Figure 2 illustrates this phenomenon for

three activities: SITC, KNEE and WLKT. It can be observed that as the activities are performed faster (top to bottom), the frequency band enlarges while the overall activity time shortens. Such changes in activity pace can be an indicator of mobility restrictions, abnormalities or health measures. Hence, it is important to accurately recognize the person’s agility score in health applications.

Once μ D signatures are generated, their upper, central and lower envelopes can be extracted using the percentile method [21] as shown in Figure 3. Upper and lower envelopes can then be used to compute the peak velocity of each sample indicating the maximum speed in an activity. Figure 4 presents the maximum velocity distribution of *slow*, *nominal* and *fast* samples. It can be observed that although there are overlapping regions, *slow* samples have much lower average maximum velocity of 2.19 m/s when compared to the *fast* samples which have average maximum velocity of 3.45 m/s. *Nominal* samples, on the other hand, have the mean maximum between the two with 2.84 m/s. It should also be noted that these distributions include all the classes and some activities naturally have lower radial motion than others which can potentially move the distributions closer. For instance, F_LAYB can have a similar radial speed as S_WLKT. Considering such similarities, a certain overlap between the distributions is expected.

III. ACTIVITY AND AGILITY SCORE ESTIMATION

Four different DL-based approaches for activity and agility score recognition are evaluated. These are: activity recognition only, agility recognition only, joint estimation of activity and agility, and simultaneous estimation of activity and agility in a multi-task manner. In order to have a fair comparison across the different approaches, a common backbone network is utilized for feature extraction. It was decided to use VGG-16 network [22] for its effectiveness on various RF datasets [23].

Activity-only and the agility-only networks are two separate networks which have the identical architecture except the number of neurons at the output classification layer. Joint activity and agility prediction network combines activity and agility recognition into the same task which estimates activity along with its agility score, hence resulting in 30 (10×3) neurons at the output layer. Finally, multi-task learning is evaluated where the network has two separate output layers with 10 and 3 neurons for activity and agility prediction, respectively. The total loss, $\mathcal{L}_{\text{total}}$, of the two tasks is:

$$\mathcal{L}_{\text{total}} = \lambda \mathcal{L}_{\text{act}} + \gamma \mathcal{L}_{\text{ag}} \quad (4)$$

where $\mathcal{L}_{\text{act}}$ and $\mathcal{L}_{\text{ag}}$ are the categorical cross-entropy losses for activity and agility recognition, and λ and γ are the weights of two losses. In this work, both λ and γ are chosen to be 1 in order to not have any bias towards a certain task.

Table I presents the classification results for the four aforementioned approaches. While activity-only and agility-only networks yield a good classification accuracy of 95.9% and 89.7%, respectively, they are targeted towards individual tasks.

TABLE I: Performance comparison of different activity and agility score recognition approaches.

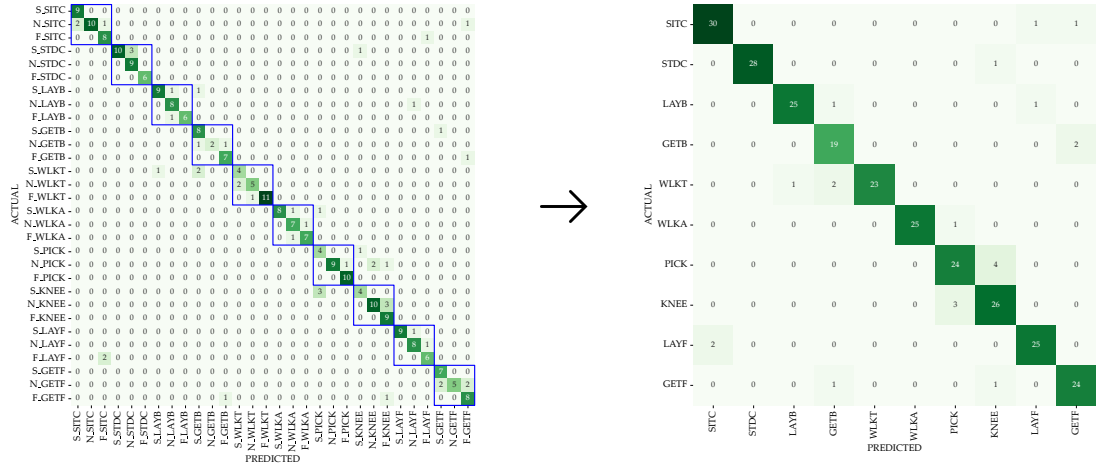
Recognition Method	Number of Output Classes	Acc. (%) Activity	Acc. (%) Agility
Activity Only	10	95.9	-
Agility Only	3	-	89.7
Joint Act. & Ag.	30	89.7	91.9
Multi-task Act. & Ag.	10 & 3	95.2	90.8

In addition, they require separate networks which doubles the memory and the computational requirements. Joint estimation network, on the other hand, combines the two tasks into one network with a wider output layer with 30 neurons. It enables simultaneous optimization of the two task. While the overall testing accuracy for this network for 30 classes is 82.3%, when activities and agility scores are combined, individual metrics can be obtained as 89.7% and 91.9%. Figure 5a shows the transition from 30-class to 10-class for activity recognition evaluation. The blue squares indicate the predictions of the agility score within the same activity category and they can be merged to disregard the errors stemming from agility prediction. Figure 5b, on the other hand, shows the transition from 30-class to 3-class for agility recognition evaluation. The blue squares indicate the predictions of the activity within the same agility category and they can be merged to disregard the errors stemming from activity prediction. The off-diagonal red squares show the area where *slow* and *fast* activities are confused with each other. It can be seen that there is no such confusion as all of those cells are zero. Note that the numbers in the confusion matrices represent the number of samples.

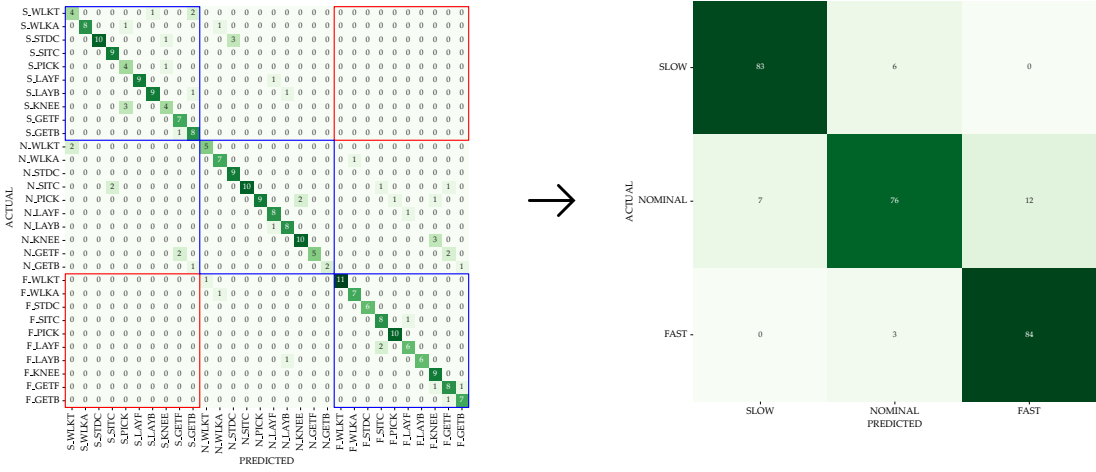
The accuracy for activities are significantly lower than the activity-only case, while the accuracy of agility is higher than the agility-only network. This indicates that the network is biased towards agility differentiation more than the activity differentiation. This is a consequence of the network having a single output for estimating both activity and agility, as we cannot put emphasis on a certain task. This calls for a less restricted approach — multi-task learning. Having separate outputs for different tasks allows putting arbitrary importance coefficient for each task. The multi-task approach yields 95.2% and 90.8% accuracy for the activity and agility recognition tasks. While the activity performance is much better than the joint case and almost as good as the activity-only case, the agility performance is in-between the joint and agility-only cases with subtle differences.

IV. MULTI-INPUT-MULTI-TASK LEARNING NETWORK

μ D Spectrograms are very effective in representing the body kinematics over time, but they do not provide any information regarding positions of the body parts. In this section, we evaluate the efficacy of introducing range information into the network using RPs. In order to have one-to-one comparison, the same VGG-16 backbone network is used. However, since both μ D spectrograms and RPs are simultaneously fed into the network, two parallel VGG-16 branches are constructed for



(a) Confusion matrix of 30-class activity and agility score estimation (left) and merging of different agility scores of the same activity into the same cell (right).



(b) Rearranged confusion matrix of 30-class activity and agility score estimation (left) and merging of different classes with the same agility score into the same cell (right).

Fig. 5: Confusion matrices of joint activity and agility score estimation.

feature extraction, followed by task specific fully connected layers. After the multi-layer perceptron (MLP) layers, latent spaces of the two inputs are aggregated in a concatenation layer. Finally, two output layers are appended for activity and agility classification. The overall architecture of the MIMTL network is depicted in Figure 6.

The MIMTL network's performance is compared with a multi-input joint activity and agility estimation network which has the identical model architecture except having single output layer with 30 neurons instead of two output layers with 10 and 3 neurons for classification.

Figure 7 presents the classification results for activity and agility recognition for four different approaches: μ D-only joint estimation, μ D-only multi-tasking, multi-input joint estimation and MIMTL method. It can be observed that there is a significant improvement in activity recognition accuracy when the network is built as multi-task in lieu of joint estimation of activity and agility score for single and multi-input cases

(i.e., 89.7% to 95.2% and 95.6% to 98.2%). Inclusion of range information also improves the recognition capability of activities for both joint (89.7% to 95.6%) and multi-task (95.2% to 98.2%) approaches. For agility score recognition, only slight performance drop (91.9% to 90.8%) and increase (88.2% to 90%) are observed for single and multi-input cases, respectively.

V. CONCLUSION

This paper presented the first work on joint estimation of human activities of daily living and the activity agility score. We evaluated different approaches for simultaneous classification of both tasks. It was found that optimizing both tasks in separate output layers with a joint total loss function significantly improves the activity recognition performance while maintaining the agility score estimation accuracy. Importance of each task can be determined based on the end application by adjusting the individual loss weight factors accordingly.

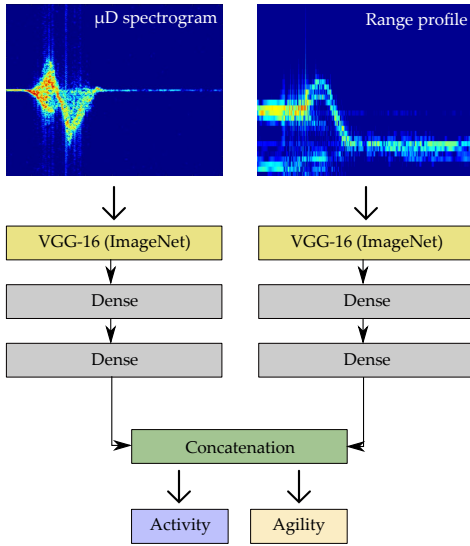


Fig. 6: Multi-input multi-task learning network for joint activity and agility score prediction.

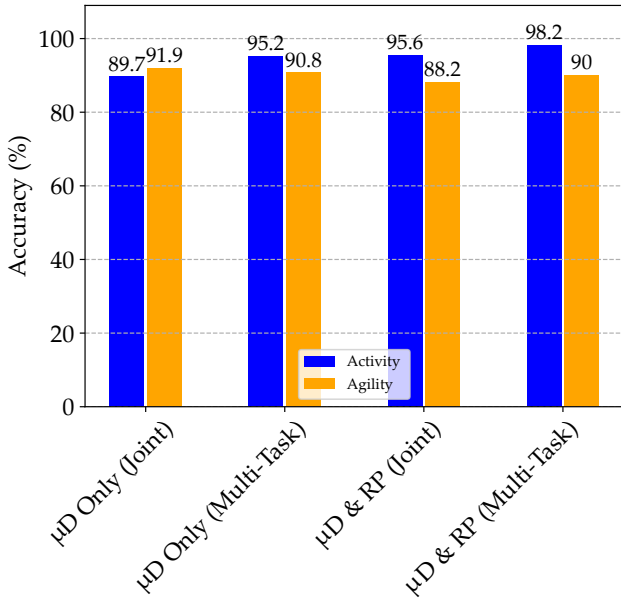


Fig. 7: Comparison of joint and multi-task estimation results for single and multi-input networks.

It was also shown that inclusion of range information along with Doppler information in a multi-input fashion boosts the activity recognition accuracy while having slight detrimental effect on agility score recognition performance. This results implies that the MIMTL approach can be effectively utilized for recognition of human activities and the current state of agility of a person. Sudden or over-time changes in agility score can be indicators of certain gait abnormalities or healing processes depending on the change direction of agility score. While agility scores can be described by more than three labels, as considered in our analysis, or even assume a continuous scalar, our work paves the way for future generalization

and advanced joint activity and agility state recognition.

ACKNOWLEDGMENT

This work was funded in part by the National Science Foundation (NSF) Awards #2233503 and #2238653. Human studies research was conducted under UA Institutional Review Board (IRB) Protocol #18-06-1271 and #23-04-6553.

REFERENCES

- [1] 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.
- [2] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE Geosci. Rem. Sens. Lett.*, vol. 13, no. 1, pp. 8–12, 2016.
- [3] M. Amin, *Radar for Indoor Monitoring: Detection, Classification, and Assessment*. CRC Press, 2018.
- [4] C. Li, J. Cummings, J. Lam, E. Graves, and W. Wu, "Radar remote monitoring of vital signs," *IEEE Microwave Magazine*, vol. 10, no. 1, pp. 47–56, 2009.
- [5] S. A. Shah and F. Fioranelli, "Rf sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, 2019.
- [6] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 180–189, 2018.
- [7] J. Lien, N. Gillian, M. E. Karagozler, P. Amihoud, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.
- [8] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "Asl trigger recognition in mixed activity/signing sequences for rf sensor-based user interfaces," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 699–712, 2022.
- [9] S. Z. Gurbuz, Ed., *Deep Neural Network Design for Radar Applications*. London: IET, 2020.
- [10] V. Chen, *The Micro-Doppler Effect in Radar, Second Edition*, 2019.
- [11] E. Isaac, S. Elias, S. Rajagopalan, and K. Easwarakumar, "Trait of gait: A survey on gait biometrics," 03 2019.
- [12] A.-K. Seifert, A. M. Zoubir, and M. G. Amin, "Radar classification of human gait abnormality based on sum-of-harmonics analysis," in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 0940–0945.
- [13] A.-K. Seifert, M. Amin, and A. Zoubir, "Toward unobtrusive in-home gait analysis based on radar micro-doppler signatures," *IEEE Transactions on Biomedical Engineering*, vol. PP, pp. 1–1, 01 2019.
- [14] 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.
- [15] D. Podsiadlo and S. Richardson, "The timed 'up & go': A test of basic functional mobility for frail elderly persons," *Journal of the American Geriatrics Society*, vol. 39, no. 2, pp. 142–148, 1991.
- [16] B. Jesness, "A human ethogram: Its scientific acceptability and importance (now new, because new technology allows investigation of the hypotheses) (an early must read)," 02 1985.
- [17] L. K. Jones, B. M. Jennings, M. K. Higgins, and F. B. M. de Waal, "Ethological observations of social behavior in the operating room," *Proceedings of the National Academy of Sciences*, vol. 115, no. 29, pp. 7575–7580, 2018.
- [18] M. G. Amin, "Micro-doppler classification of activities of daily living incorporating human ethogram," in *SPIE DCS Symp.*, 2020.
- [19] M. G. Amin and R. G. Guendel, "Radar classifications of consecutive and contiguous human gross-motor activities," *IET Radar, Sonar & Navigation*, vol. 14, no. 9, pp. 1417–1429, 2020.
- [20] M. Crawshaw, "Multi-task learning with deep neural networks: A survey," *CoRR*, vol. abs/2009.09796, 2020.
- [21] P. V. Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigation*, vol. 2, no. 2, pp. 135–145, 2008.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.