Activity classification using raw range and I & Q radar data with Long Short Term Memory layers

1st Charalampos Loukas
School of Engineering, University of Glasgow, G12 8QQ
University of Glasgow
Glasgow, United Kingdom
c.loukas.1@research.gla.ac.uk

2nd Francesco Fioranelli
School of Engineering, University of Glasgow, G12 8QQ
University of Glasgow
Glasgow, United Kingdom
Francesco.Fioranelli@glasgow.ac.uk

3rd Julien Le Kernec
School of Engineering, University of Glasgow, G12 8QQ
University of Glasgow
Glasgow, United Kingdom
Julien.LeKernec@glasgow.ac.uk

4th Shufan Yang
School of Engineering, University of Glasgow, G12 8QQ
University of Glasgow
Glasgow, United Kingdom
Shufan.Yang@glasgow.ac.uk

***Abstract*—This paper presents the first initial results of using radar raw I & Q data and range profiles combined with Long Short Term Memory layers to classify human activities. Although tested only on simple classification problems, this is an innovative approach that enables to bypass the conventional usage of Doppler-time patterns (spectrograms) as inputs of the LSTM layers, and adopt instead sequences of range profiles or even raw complex data as inputs. A maximum 99.56% accuracy and a mean accuracy of 97.67% was achieved by treating the radar data as these time sequences, in an effective scheme using a deep learning approach that did not require the pre-processing of the radar data to generate spectrograms and treat them as images. The prediction time needed for a given input testing sample is also reported, showing a promising path for real-time implementation once the LSTM network is properly trained.**

***Index Terms*—radar, deep learning, Human Activity Recognition (HAR), LSTM**

I. Introduction

Automatic human monitoring, to discriminate between different activities of daily life (ADL) and detect critical events such as falls, attracted significant interest in the research community due to the rise of health conditions and multimorbidity related to the general ageing of the population. Assisted living technologies address this problem by making use of different devices, such as wearable, camera based, RGB-Depth and radar sensors [1]. Radar has been recently suggested in this context, as it can provide the advantage of contactless and non-intrusive monitoring, with no compliance required from the end-users and reduced risks for privacy as plain optical images are not recorded [2].

In the context of assisted living, classification of different human indoor activities is important to monitor the wellbeing of vulnerable people (for example the presence of important activities such as personal hygiene, or food preparation and intake), and detect possible anomalies in activity patterns that may be linked to worsening health conditions. This classification is covered extensively in the literature using in most of the cases analysis of micro-Doppler signatures, which originate from the periodic micro-motions of the limbs and torso of the humans, inducing frequency modulation in the main backscattered radar signal [3]. Each activity will generate a characteristic micro-Doppler signature, as the person will perform specific movements of torso, arms, and legs for different activities. Short-Time Fourier Transform (STFT) is used commonly to analyse these signatures in a time frequency domain; a process which requires to choose different parameters such as the type of window to calculate individual Fast Fourier Transforms (FFTs), the percentage of window overlap, and the number of FFT points. This type of pre-processing requires computational resources and fine-tuning of the related parameters, as well as imposing trade-offs in time and frequency resolutions when STFT is used. Furthermore, handcrafted feature samples are usually extracted from the spectrograms to be used as inputs to the classification algorithms, requiring even more fine-tuning of the parameters involved, and the selection of different set of features for various scenarios to achieve optimal classification results [4], [5].

Deep Learning techniques have shown great potential in classifying human activities bypassing the manual tuning of feature extraction algorithms. Previous works [6], [7] used Convolutional Neural Networks (CNN) and Deep Auto-Encoders (AEs) to process spectrograms of different human activities to classify, essentially casting this classification problem as an image recognition problem. Basically, each spectrogram is treated as a matrix of pixels and used as inputs to CNN and AE based networks. In order to reduce the radar-data pre-processing load and extract information from other radar domains, range profiles were used to identify seven different motions using CNNs [8]. However, even in this case the range profiles were organised as Range-Time matrices, and treated as images (matrices of pixels) before feeding them into neural networks.

In this paper, we present initial results for a binary classification problem with a new prediction for every 0.5 s of recorded activities, using Long Short Term Memory (LSTM) layers that use range profiles and raw I & Q radar data as inputs to the network. This innovative approach enables to bypass both generation of spectrograms or other Doppler-time matrices through the application of time-frequency distributions, and the manual extraction of handcrafted features from the spectrograms. This can be illustrated in Fig. 1, which indicate the presence of our work with red circle among the existing ones in the literature. CNN & LSTM refer to [13], where hand gesture recognition is performed using range-Doppler inputs to CNN and LSTM networks. Additionally, Stacked Auto Encoders (SAE) were utilized with spectrograms as mentioned in [7]. Recurrent Neural Networks (RNNs) are

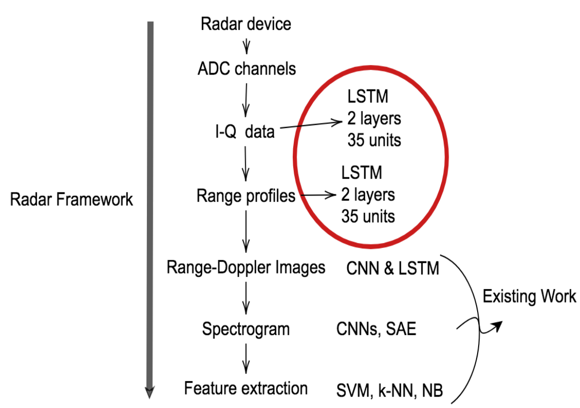

Fig. 1. Radar framework process for classification of human movements.

suitable for this task, treating the series of received radar pulses as a time sequence, a time series of samples to process, rather than an image. RNNs have the ability to maintain memory while iterating through the sequence elements, whereas CNNs and the other densely connected networks treat each input independently. However, RNNs can struggle to learn long sequences, hence their LSTM variant is used in this work to address this issue [9]. LSTM have been briefly explored to classify different radar classes through spectrograms [10], but, to the best of our knowledge, the use of range profiles and direct, raw I & Q data presented here is an innovative approach. Furthermore, the significance of using directly the I & Q data of the backscattered radar signal is revealed, as the proposed network scheme can provide a class prediction every 0.5 s of received data once the network has been properly trained. This is a promising result for future real-time implementations of the proposed method.

Using the aforementioned two representations of sequential data, range profiles and raw I & Q data, different network topologies are investigated and their prediction time required is considered. This research is opening the path for real-time classification, avoiding any cumbersome pre-processing for time-frequency transforms and feature extraction.

II. Data and Representations

The dataset containing two movements 'walking' and 'sitting & standing' was constructed using an off-the-shelf Frequency Modulated Continuous Wave (FMCW) radar. Its carrier frequency was set at 5.8GHz, bandwidth at 400MHz and sweep time at 1ms resulting in 128 samples per sweep. Five subjects contributed, where each action was recorded continuously for 60 s. In total 19 recordings were collected, 10 of them refer to 'walking' and the rest to 'sitting & standing'.

For the first type of representation, backscattered signals processed with an FFT to obtain the range profiles and each profile was limited to 35 range bins to keep in track the actual information of the performed action. Each input sample to LSTM layer is instructed to have a specific format which is characterized by the batch processed size, the time steps and the number of features. Our aim was to predict a class of action every 0.5 s, thus each time-sequence sample consisted of 500 range profiles. In that way 2,280 samples were obtained and each one represent 0.5 s of recorded action, with 500 time steps (range profiles) and 35 features (range bins).

Second type of representation is the I & Q data for the same recordings. Each time-sequence sample of 0.5 s amounts 64,000 time steps and only two features, the quadrature components from the two radar ADC (Analogue to Digital Converter) channels. Therefore the second representation is characterized by 2,280 samples.

In Fig. 2 the range profiles for two successive movements (walking and sitting & standing) as well the raw I & Q representation of the same movements are illustrated.

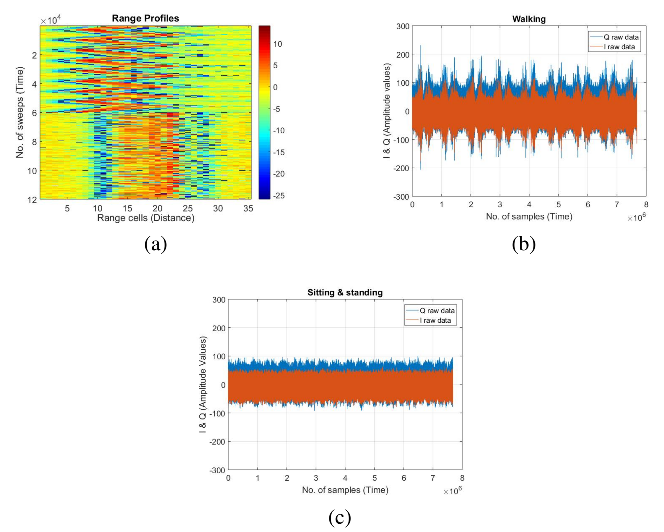

Fig. 2. (a) Range profiles processed for walking and sitting & standing movements for 60 s each. (b), (c) illustrate I & Q data extracted from the two ADC channels for walking and sitting & standing movements respectively. The patterns that exist in (a), (b) and (c) are exploited through the LSTM layers.

The network investigation was performed using as backend tensorflow gpu, version 1.4 in a single NVIDIA Pascal 1080.

III. LSTM NETWORKS & INVESTIGATION

The first type of experiment that was performed is common to both representations, setting a common ground of comparisons between the classification accuracy and the selected type of data. A stratified 5-fold method was used, mainly due to low number of samples and the possible expected variance in the classification accuracy result on the selected test set. For both data types, the samples were shuffled in a stratified manner and 80% was used for training and the rest 20% for testing, under the 5-fold scheme. The neural network model for the range profiles had two layers of 35 LSTM units each, using the RMSprop optimizer with learning rate of 0.001. The batch size equals to one, stating that each movement is independent from the other and the number of epochs was set to 50. On the other hand, for the I & Q data, two layers of four LSTM units were used, with the same optimizer, learning rate and batch size whereas the number of epochs was set to 10 this time due to the high number of time steps. Note that for the I & Q data, fewer LSTM units were tested manually, however they resulted in under-fitting during the training procedure. In both network models, the final layer has a sigmoid activation function. The mean test classification accuracy and the standard deviation of the results for both representations in the same activities is summarized in Table I.

As can be seen from Table I, this method indicates the potential for using directly the I & Q data for predicting every 0.5 s, which scored higher accuracy than the range profiles. In addition, the standard deviation of the five test accuracies quantify: about 1-2%, indicating small variance between random choices of test sets. For sake of clarity, the 20% of data that was used for testing was not used during the training phase (comes from unseen movements from the 5 subjects which represents 456 samples).

TABLE I
STRATIFIED 5-FOLD RESULTS FOR TWO-LAYER LSTM NETWORKS:

Metrics	Range profiles	I & Q data
LSTM units	35	4
Mean Test Accuracy	94.16%	**97.67%**
Standard Deviation	1.14%	2.02%

The next series of experiments are related with optimized neural network architectures for both representations and investigation of the prediction time per sample, considering a real-time implementation.

Regarding the range profiles data, hyperparameter optimization was performed with a two-layer architecture, splitting the dataset to 64% training, 16% validation and 20% testing. Dropout [11] between layers was added to increase the generalization capability of the model. This optimization was achieved using the Hyperopt library in Python and the algorithm of Tree Parzen Estimators [12]. The characteristics of the optimized network architecture with two layers were 37 LSTM units in each layer, Dropout 0.42359 and 0.24449, the learning rate 0.00043 and the total number of parameters

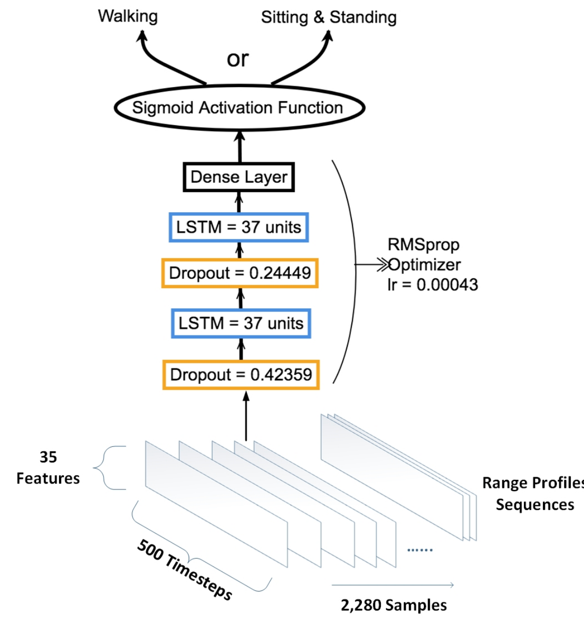

Fig. 3. Optimized LSTM network architecture for range profiles sequences.

equals to $22,238$ for the model. The optimized network architecture for the range profiles is shown in Fig. 3.

Afterwards, the optimized architecture of two layers was used to train a model on 80% of data and tested with the remaining 20%; the same samples were used for testing in the 5-fold experiment. The result of testing in 20% of unseen movements after 100 epochs of training was **97.58%** and the confusion matrix is illustrated in Fig. 4.

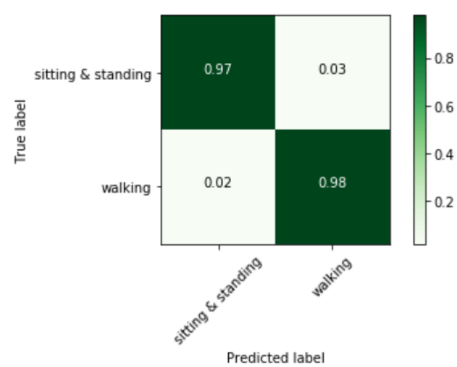

Fig. 4. Normalized confusion matrix for optimized range profiles network.

It is important to mention that the prediction time per sample of unseen movement with that network architecture was 2 ms.

The same process of optimization was followed also for a three-layer network, an optimization procedure which lasted 28.8 hours but resulted in a classification accuracy of **93.85%**. Hence, the two-layer network, for this binary problem using range profiles was found to give the highest accuracy and any more complex network approach, anticipating to learn more abstract geometric presentations of inputs, resulted in overfitting.

Considering the raw I & Q data, the dataset of 2,280 samples was split to 64% training, 16% validation and to the remainder 20% (used also in the 5-fold experiment for testing). The architecture of the network with same characteristics as in the first experiment was used to train the model for 10 epochs and is described in Fig. 5.

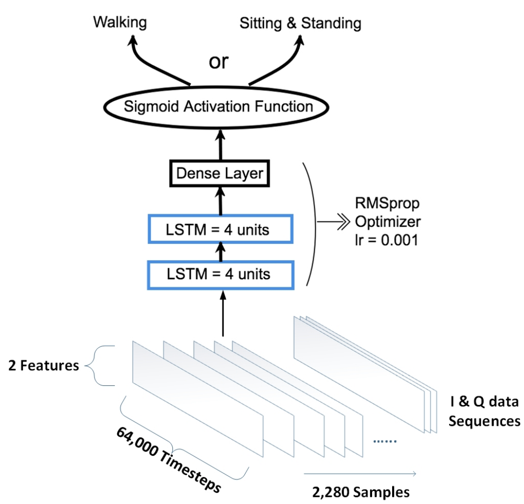

Fig. 5. LSTM network architecture for I & Q data sequences.

At the end of the training procedure, the weights of the model with best validation accuracy were saved and compiled to predict the class of 20% of unseen movements. Due to the high number of time steps (64,000) for each time-sequence sample, the training time was 16.41 hours. The classification accuracy in the test set using the weights of the model that achieved the highest validation accuracy was found to be **99.56%** and the confusion matrix is presented in Fig. 6.

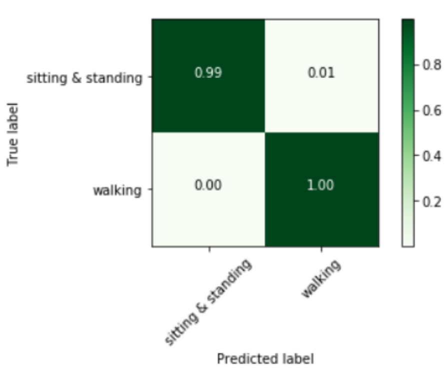

Fig. 6. Normalized confusion matrix for I & Q data network.

The LSTM layers of the proposed network architecture achieved a correct generalization for all the 456 unseen movements of 'walking' and 'sitting & standing' except from two time-sequence samples. This result demonstrated that this type of Deep Learning is able to train backscattered radar data as time sequences. In that way further pre-processing is not required to first visualize the micro-Doppler signatures to then manually or automatically extract features.

However, considering a real-time implementation of human monitoring, the prediction time per sample using the raw I & Q data was 2 s. This comes from the fact that each movement is presented with 64,000 time steps and no undersampling was performed during the pre-processing phase. As a result, the network is predicting every 0.5 s of movement with a delay of 2 s. If we suppose that a carer is monitoring an elderly patient for a potential fall, a delay of 2 s is not crucial, acknowledging that the system is predicting the class of the action every 0.5 s.

Nonetheless, this type of network can be used in many other applications that up to now target micro-Doppler signatures had to first be extracted in order to perform classification. Looking ahead, the delay of 2 s may be too important for some applications and monitoring requirements and will need to be decreased.

IV. Conclusion

The experimental results introduced in this paper are believed to present an innovative approach for radar-based classification in the context of assisted living, using LSTM layers as an alternative to image processing via CNNs, and using range profiles and raw I & Q radar data as inputs to the classification stage. This enables to bypass cumbersome pre-processing with the possible trade-offs and loss of information they may cause such as FFT and time-frequency transforms. These preliminary results show a great potential of using directly the raw I & Q data, as the classification accuracy achieved a maximum of **99.56%**, using 64% of data for training, whereas for the same activities, range profiles attained **97.58%**, with 80% of samples for training. This proves that LSTM layers, integrated into a neural network, can learn good representations of the radar data treating them as sequences, rather than images as it is currently explored with CNNs. In this way, the need for pre-processing of radar data is reduced, and so it's the computational delay related to that. Additionally, with the proposed scheme, prediction of the class is performed every 0.5 s giving the advantage for continuous monitoring as the time of radar observation grows.

The problem of a long prediction time per movement is reported, when using raw radar data. In future work, it is anticipated this delay will be decreased under 1 s, with ad-hoc network architectures optimized for the specific classification scenario. Furthermore, we will focus on how to detect multiple activities while exploring the balance between long prediction time and classification accuracy. The trade-off between testing prediction time and classification accuracy will also be explored, comparing the proposed approach with more conventional supervised learning techniques and the use of CNNs with image-processing inspired approaches.

.

ACKNOWLEDGMENT

We would like to thank A. Angelov for his work on extracting the experimental movements and the subjects taking part.

REFERENCES

[1] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante,“Radar and RGB-Depth Sensors for Fall Detection: A Review,” IEEE Sensors Journal, vol. 17, no. 12, pp. 3585-3604, Jun. 2017.

[2] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho,“Radar Signal Processing for Elderly Fall Detection: The future for in-home monitoring,” IEEE Signal Processing Magazine, vol. 33, no. 2, pp. 71-80, Mar. 2016.

[3] Y. Kim and H. Ling,“Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine,” IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 5, pp. 1328-1337, May 2009.

[4] F. Fioranelli, M. Ritchie, S. Z. Gürbüz, and H. Griffiths,“Feature Diversity for Optimized Human Micro-Doppler Classification Using Multistatic Radar,” IEEE Transactions on Aerospace and Electronic Systems, vol. 53, no. 2, pp. 640-654, Apr. 2017.

[5] S. Z. Gürbüz, B. Erol, B. Çağlyan, and B. Tekeli,“Operational assessment and adaptive selection of micro-Doppler features,” IET Radar, Sonar & Navigation, vol. 9, no. 9, pp. 1196-1204, 2015.

[6] Y. Kim and T. Moon,“Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks,” IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 8-12, Jan. 2016.

[7] B. Jokanovic and M. Amin, “Fall Detection Using Deep Learning in Range-Doppler Radars,” IEEE Transactions on Aerospace and Electronic Systems, pp. 1-1, 2017.

[8] Y. Shao, S. Guo, L. Sun, and W. Chen, “Human Motion Classification Based on Range Information with Deep Convolutional Neural Network,” in 2017 4th International Conference on Information Science and Control Engineering (ICISCE), 2017, pp. 1519-1523.

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[10] G. Klarenbeek, R. I. A. Harmanny, and L. Cifola, “Multi-target human gait classification using LSTM recurrent neural networks applied to micro-Doppler,” in 2017 European Radar Conference (EURAD), 2017, pp. 167-170.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929-1958, 2014.

[12] J. Bergstra, D. Yamins, and D. D. Cox: “Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures,” Proc. of the 30th International Conference on Machine Learning (ICML), 2013

[13] Z. Zhang, Z. Tian, and M. Zhou, “Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor,” IEEE Sensors Journal, vol. 18, no. 8, pp. 32783289, Apr. 2018.