Human Activities Classification in a Complex Space Using Raw Radar Data

Shufan Yang[1], Julien Le Kernec[1], Francesco Fioranelli[1], Olivier Romain[2]
[1]*School of Engineering, University of Glasgow ,* Glasgow, UK
[2]*ETIS-ASTRE, Université Cergy-Pontoise, Cergy-Pontoise, France*
Email:shufan.yang@glasgow,ac.uk ;

Abstract— **The classification of human activities through the utilisation of radar mainly focuses on the analysis of the time-frequency domain, typically through spectrograms obtained with Short-Time Fourier Transform. Commonly the original information format in the raw complex radar data is being ignored. In this work, we propose and evaluate a new recurrent neural network architecture to decode the time sequence of raw radar data over a longer time than previously attempted, through training in a complex space of in-phase and in-quadrature data. We explore a solution for sequence analysis problems when adapting a recurrent neural network; our best network architecture exhibits a significant accuracy in performance for over seven participants with six activities, collected over 60 seconds.**

Keywords— radar classification, machine learning, LSTM networks, human activity recognition.

I. INTRODUCTION

In the context of assisted living, automatic classification of various human indoor activities is important to monitor the wellbeing of vulnerable people, as this can detect possible anomalies in activity patterns that may be linked to worsening health conditions. The techniques of automatic human monitoring based on radar give the advantage of contactless and non-intrusive monitoring. Compared to optical imaginary, privacy concerns may be less relevant as plain optical images of people and environments are not recorded; compared to wearable sensors, end-users do not need to wear, carry, or interact with devices that may be perceived as invasive [1-2].

Typically, human micro-Doppler information has been exploited for assisted living using radar, for applications ranging from the classification of different activities of daily living, detection of critical events such as falls, characterisation of gait, and monitoring of vital signs parameters [3-9]. The classification of different signatures in the micro-Doppler domain was typically performed extracting handcrafted features from spectrograms, for example empirical features (e.g. bandwidth, periodicity, RCS ratios), features based on Singular Value Decomposition (SVD) and Cadence Velocity Diagram (CVD), Cepstral features, features based on Wavelet transformations [4-5, 10-13]. However, this approach can induce false positive detection for various experimental radar data. Indeed, the traditional classification methods are based on a number of parameters in the feature extraction algorithm that have been optimised based on the radar operator's experience, with the risk of overfitting them to a specific set of data.

To address this problem, the usage of neural networks to process radar signatures for classification has gained significant interest in the past few years. Convolutional neural networks have been proven highly successful at human activities classification using micro-Doppler signatures [3, 14]. A limitation of this approach is that radar data for training and testing are typically captured as an image corresponding to a fixed time duration, a sort of "snapshot" spectrogram where only one activity is performed. This might be unsuitable for analysing long sequences of realistic data, where a continuous flow of activities is performed by people, with more or less significant transitions between them.

Therefore, we hypothesise that learning the temporal relationship in a sequence of physical actions is important for human activity classification. The idea of interpreting radar data as temporal sequences rather than images for classification is not totally new, and the recurrent networks have been applied to spectrograms [15-16]. However, this approach is still seldom explored in the open literature, and even more so for raw radar data. A further computational challenge is that deep neural networks can be hard to train in long sequences and struggle to model significant features out of the data.

In this paper, we propose using Long Short Term Memory (LSTM) units in a recurrent neural network to classify six different activities, expanding from the preliminary results of simpler binary classification in our previous work [17]. As shown in Figure 1, 60s of radar data for six activities performed by six participants have been segmented as 0.5-second subsamples. These real and imaginary data samples are then used as inputs in the form of complex-valued vectors. Two LSTM layer1 and layer 8 are input and output layer. Due to the size of the tensor, the data has been reshaped into the array from [batch_size, seq_len, n_channels] to [seq_len, batch_size, n_channels]. So that the network would properly split the data into a list of [batch_size, lstm_size] arrays. New sequences can be predicted, and be labelled using the "one hot encoding" approach. In this way, we present a scalable framework by sharing parameters through time. In our approach, the number of hyper-parameters remains constant even with larger inputs (more activities).

The remainder of this paper is organised as follows. Section II provides more details on the methodology used, and section III presents initial results. Finally, section IV concludes this paper and outlines some possible future work.

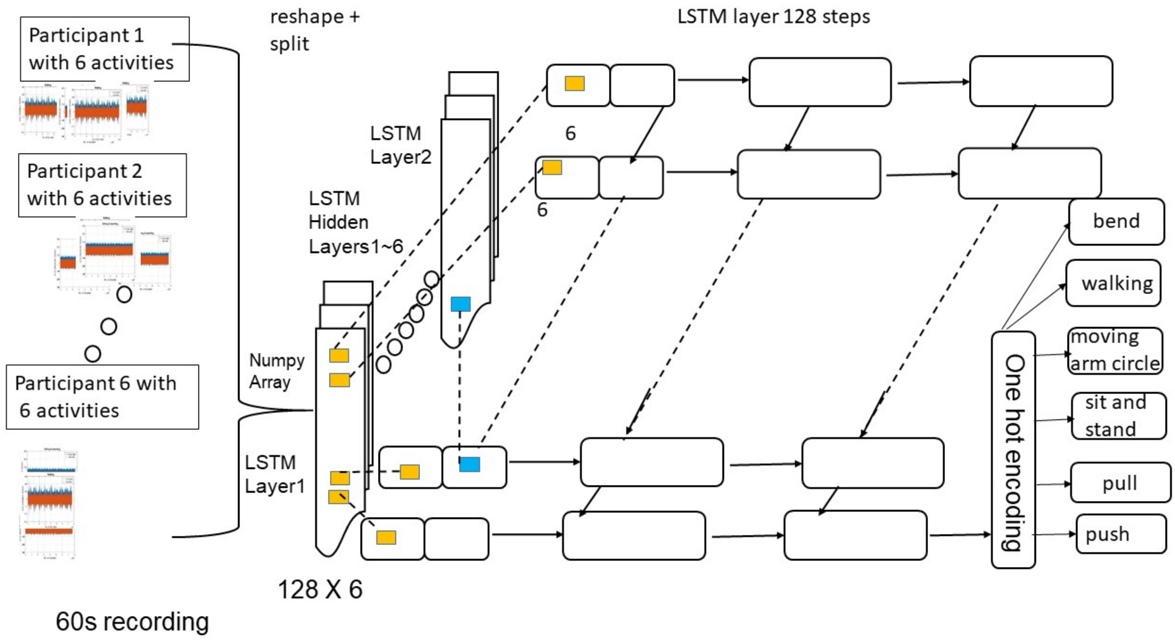

Fig. 1. Overview of the proposed method with LSTM architecture, input radar data (left-hand side) and output labels (right-hand side)

II. Methodology

A. Data preparation

The dataset contains six activities recorded for classification using a C-band Frequency Modulated Continuous Wave (FMCW) radar at the University of Glasgow. These include walking back and forth in front of the radar; standing and sitting repeatedly from a chair; bending down to pick up an object from the floor and standing back up; moving one arm in circle while standing; pushing (moving one hand quickly towards the radar while standing and pulling it slowly backwards); and pulling (moving one hand slowly towards the radar while standing and pulling it quickly backwards). The walking activities were performed in the line of sight of the radar from ~6 to 0.5m from the antennas, whereas the other activities were performed on the spot at a distance of ~5 meters from the radar system. Two 60 seconds-long recordings were collected for each subject performing each type of activities. During data collection, the participants were facing the line of sight of the radar. The radar had a carrier frequency set at 5.8GHz, bandwidth at 400MHz and sweep time of 1ms, resulting in 128 samples per sweep and a Doppler ambiguity of ±500 Hz. For a 60s recording, the number of sweeps were 60,000, and each had 128 complex samples.

B. Training method

Unlike standard recurrent neural networks (RNN), the Long Short Term Memory (LSTM) architecture [18] uses memory cells to store and output information, allowing it to better discover long-range temporal relationships. The value stored in the LSTM cell c is maintained unless it is added to the input gate i or diminished by the forget gate f. The output gate o controls the emission of the memory value from the LSTM cell. At each time step, the network is present with an input x_t and updates states to: as shown in the following equations:

$$\hat{i_t},\ \hat{o_t}, \hat{f_t}, \hat{u} \ = W_{xh}\, x_t + W_{hh}\, h_{t-1} + b_h \quad (1)$$

$$i_t = \sigma(\hat{i_t}) \quad (2)$$
$$o_t = \sigma(\hat{o_t}) \quad (3)$$
$$f_t = \sigma(\hat{f_t}) \quad (4)$$
$$u = \tanh(\hat{u}) \quad (5)$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot u \quad (6)$$
$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

Where $\sigma(x)$ is the logistic sigmoid function; The u vector is a proposed update to the cell state c. W_{xh} and W_{hh} are weight matrices, and b_h is a bias vector. · denotes element-wise multiplication of two vectors.

We followed the same procedure as training CNNs with two modifications:

First modification is explicitly considering sequences of activations. Since subsamples contain dynamic content, the variations between frames may encode additional information, which could be useful in making more accurate predictions.

Secondly subsamples is generated using a fixed window to meet batch training method. Since the idea of RNN decoder is to use the previous output as next input. The final output will then become a list of outputs at different time stamp.

Our dataset contains 184,320,000 rows (7,680,000 * 24 recording). There are no missing values in all data. The complex vector for one 60s recording has total 7,680,000 samples (60,000 sweeps with 128 bins per sweep). Only 1 second of data corresponds to 128,000 complex samples, with possible 3D format [bins, sweeps, real&imaginary data].

For one activities, with 7,680,000 time steps divided with fix window 240 (a free parameter). The shape of data for each sequence of real part is (1, 32,000, 6) and the sequence of imaginary part is (1, 32,000, 6). In total there are 240 *24 = 5,760 samples. The final array has shape to (5760, 64,000, 6). When 25% is used for validation, the final training data shape is (4320, 64,000, 6) and the validation data shape is (1,440, 6,400, 6). For 6 dense layer used in LSTM the batch input shape is (720, 6,400, 6) with a stateful prediction (means to remember the entire sequence for classification task). Each of the time steps will be trained, with a target of "000001", “0000010”, “000100”, “001000”, “010000” or "100000" depending of where this sequence belongs.

III. RESULTS

Practically, we demonstrate that the learning with random permutations are effective at the depth of layer in the neural network, when we vectorise total weights of each layer. We see that weights in layer 2, 4 and 6 are less disorganised by noise as the number of copies grows. While the layer2 network is forming a pattern, we expect that the network figures out the features are drastically different in frequency between walking and the rest of classes. With the layer increase, more subtle patterns are being learnt gradually. The initial results of the proposed approach are promising, as shown in Figure 3 with accuracy and loss of the model. The evaluation results show faster learning and high prediction accuracy. As shown in Table I, although dropout increases from 0.5 to 0.7, the classification performance increases but at the same time, the chance of network overfitting is increased. It is interesting to find out that the data for activity “push slow, pull fast”, are the most frequently misclassified for all three values of dropout rate. Upon further inspection, it was found out that there were hardware issues with the recorded data for that specific activity. Hence their reliability is questionable. Nevertheless, the proposed network responds well to the training data, despite those for one class/activity being partially corrupted.

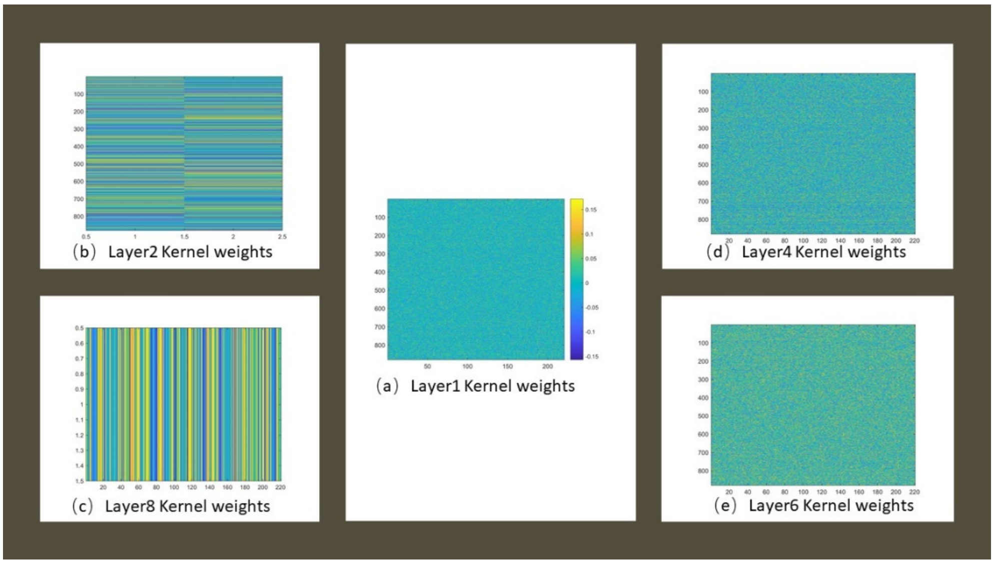

Fig. 2. The representation of weights from the sample layer1, layer2, layer 4, layer6 and layer8

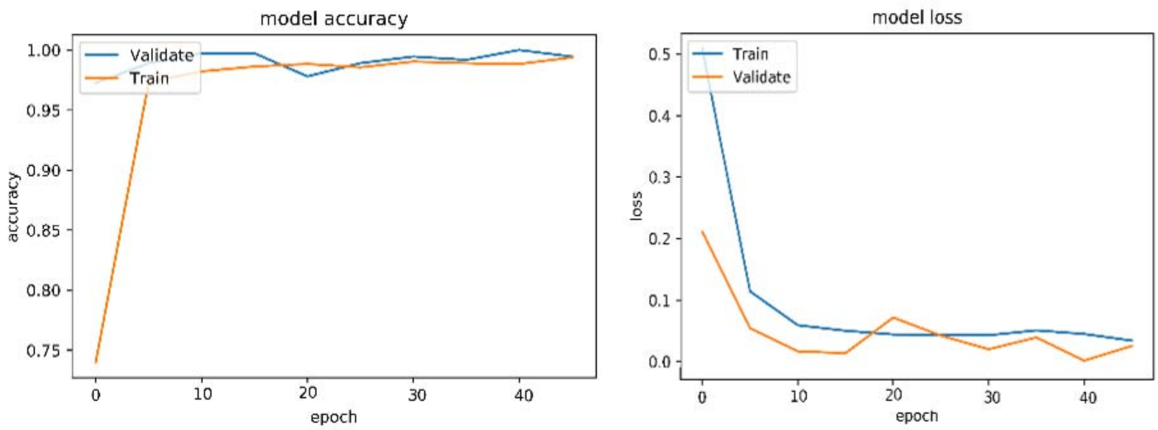

Fig. 3. The training performance of recurrent neural network: evaluation of average accuracy (left), and evaluation of average loss (right)

TABLE I. THE NORMALISED CONFUSION MATRIX RESULTS (CLASSIFICATION ACCURACY) OF RNN WITH VARIOUS DROPOUT

ACTIVITIES	Bending	Circling arms	Push	Pull	Sitting and stand	Walking	Average across activities
Dropout = 0.5	0.97	0.85	0.75	0.67	0.74	0.73	0.79
Dropout = 0.6	0.99	0.85	0.85	0.80	0.94	0.89	0.89
Dropout = 0.7	0.98	0.96	0.95	0.89	0.91	0.89	0.93

IV. CONCLUSIONS AND DISCUSSION

This work demonstrates the proof-of-concept of training the recurrent neural network using complex vectors made of raw radar data, without pre-processing through STFT or similar time-frequency transformations. We believe recurrent neural networks have the inherent capability to reconstruct and interpret the temporal information within the radar data. The results in this work on the classification of six human activities show average accuracy up to 93% when the network is trained on raw complex radar data. Initial attempts using sequences of range profiles obtained after initial FFT of the received data (FMCW data) show a reduction in performance, which is currently investigated. However, the high accuracy of recurrent neural networks comes in contrast with high computational costs. For example, we had to train the network using four NVidia GeForce GTX 1080 GPUs for 240 hours. To some extent this is expected, as the processing load in this approach is somewhat transferred to network complexity, removing some layers of more conventional radar signal processing (STFT for example).

Additional work can look at validating this approach for a wider range of radar data, regarding classes to be classified and number of subjects. Outstanding research challenges from the perspective of network structure and training are related to better capturing the long-term dependencies in the data by optimising the memory/forgetting mechanism in the LSTM units. Furthermore, there is a need to optimise the efficiency of the training when more classes are considered in more realistic applications. In order to understand the behaviour of neural network, t-SNE, introduced by van der Maaten and Hinton in 2008 [19] will be used to explore high dimensional data in each LSTM units.

Finally, we will aim to investigate multi-static radar data for classification purposes to increase performance by embedding additional spatial information and multiple perspectives into the data.

ACKNOWLEDGEMENT

The authors acknowledge support from UK EPSRC grant EP/R041679/1, and the preliminary work from A. Angelov and C. Loukas towards these results.

REFERENCES

[1] 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 Process. Mag.*, vol. 33, no. 2, pp. 71–80, 2016.

[2] 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, 2017.

[3] M. S. Seyfioğlu, A. M. Özbayoğlu and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709-1723, Aug. 2018.

[4] K. Youngwook, L. Hao, 'Human activity classification based on micro-Doppler signatures using a Support Vector Machine', *IEEE Transactions on Geoscience and Remote Sensing,* vol. 47, pp. 1328-1337, 2009.

[5] B. Erol, M. G. Amin and S. Z. Gurbuz, "Automatic Data-Driven Frequency-Warped Cepstral Feature Design for Micro-Doppler Classification," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1724-1738, Aug. 2018.

[6] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multi-sensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, pp. 1-1, 2018.

[7] B. Y. Su, K. C. Ho, M. J. Rantz, and M. Skubic, "Doppler Radar Fall Activity Detection Using the Wavelet Transform," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 3, pp. 865-875, Mar, 2015.

[8] C. W. Ding, L. Zhang, C. Gu, L. Bai, Z. C. Liao, H. Hong, Y. S. Li, and X. H. Zhu, "Non-Contact Human Motion Recognition Based on UWB Radar," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 8, no. 2, pp. 306-315, Jun, 2018.

[9] F. Wang, M. Skubic, M. Rantz and P. E. Cuddihy, "Quantitative Gait Measurement With Pulse-Doppler Radar for Passive In-Home Gait Assessment," in *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2434-2443, Sept. 2014.

[10] F. Fioranelli, M. Ritchie and H. Griffiths, "Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler," in *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1702-1710, 12 2016.

[11] S. Z. Gürbüz, B. Erol, B. Çağlıyan, B. Tekeli, 'Operational assessment and adaptive selection of micro-Doppler features', *IET Radar, Sonar & Navigation*, vol. 9 (9), p. 1196-1204, December 2015.

[12] R. Ricci, A. Balleri, 'Recognition of humans based on radar micro-Doppler shape spectrum features', *IET Radar, Sonar & Navigation*, vol. 9 (9), pp. 1216-1223, December 2015.

[13] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of Unarmed/Armed Personnel Using the NetRAD Multistatic Radar for Micro-Doppler and Singular Value Decomposition Features," *IEEE Geoscience and Remote Sensing Letters,* vol. 12, no. 9, pp. 1933-1937, Sep, 2015.

[14] K. Youngwook, H. Sungjae, K. Jihoon, 'Human detection using Doppler radar based on physical characteristics of targets', *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 289-293, 2015.

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

[16] M. Wang, G. Cui, X. Yang and L. Kong, "Human body and limb motion recognition via stacked gated recurrent units network," in *IET Radar, Sonar & Navigation*, vol. 12, no. 9, pp. 1046-1051, 9 2018.

[17] C. Loukas, F. Fioranelli, J. Le Kernec and S. Yang, "Activity Classification Using Raw Range and I & Q Radar Data with Long Short Term Memory Layers," *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, Athens, 2018, pp. 441-445.

[18] Gers, Felix A, Schmidhuber, J¨urgen, and Cummins, Fred. Learning to forget: Continual prediction with lstm. Neural computation, 12(10):2451–2471, 2000.

[19] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9, no. Nov (2008): 2579-2605.