DeepActivity: a micro-Doppler spectrogram-based net for human behaviour recognition in bio-radar

Hao Du¹, Tian Jin¹, Yongping Song¹, Yongpeng Dai¹
¹College of Electronic Science, National University of Defense Technology, Changsha, People’s Republic of China
E-mail: tianjin@nudt.edu.cn

Abstract: The movements of the human body and limbs result in unique micro-Doppler signatures, which can be exploited for classifying human activities. In this work, the authors propose a Convolutional Gated Recurrent Units Neural Network (CNN-GRU) to classify human activities of varying duration based on micro-Doppler spectrogram. Unlike conventional deep learning approaches which often treat the micro-Doppler spectrogram the same way as natural image, the authors extract local feature of micro-Doppler signatures via convolutional layer and encode temporal information with gated recurrent units. Through this unified framework, the temporal evolution of body motions within a short time can be better utilised. It avoids the resolution limitation caused by the fixed-size time window of input data and identifies human activity of duration shorter than the time window length. The experiment shows that CNN-GRU model is capable of recognising and temporally localising activity sequence contained in the spectrogram.

1 Introduction

Human activity classification is a field of particular interest to researchers varying from physical security to intelligent interface. Visual perception of the human body motion can be affected by distance, variations in lighting, deformations of clothing, and occlusions on the appearance of human body segments [1]. Owing to excellent day and night performance and ability to penetrate obstacles, radar system is an appropriate alternative to analyse human behaviour. The radar echoes from non-rigid-body contain valuable information related to human behaviours, which is known as micro-Doppler signatures [2]. Taking advantage of this distinctive effect, various human activities are promising to be classified.

Motivated by the successful application of deep neural networks in various fields, deep convolutional neural network architectures have been proposed for activity classification based on micro-Doppler spectrogram [3–6] and significantly outperform the previous state-of-the-art schemes that mainly relied on domain knowledge-based features. In addition to convolutional gated recurrent (CNN) trained from scratch, transfer-learned neural networks are also used for this task [7, 8]. Based on the pre-trained knowledge-based features. In addition to convolutional gated recurrent unit (GRU) to classify human activities of varying duration based on micro-Doppler spectrogram. Unlike conventional deep learning approaches which often treat the micro-Doppler spectrogram the same way as natural image, the authors extract local feature of micro-Doppler signatures via convolutional layer and encode temporal information with gated recurrent units. Through this unified framework, the temporal evolution of body motions within a short time can be better utilised. It avoids the resolution limitation caused by the fixed-size time window of input data and identifies human activity of duration shorter than the time window length. The experiment shows that CNN-GRU model is capable of recognising and temporally localising activity sequence contained in the spectrogram.

2 Models

A micro-Doppler spectrogram-based human activity classification framework embodying the CNN-GRU model is shown in Fig. 1. Raw human backscattering radar echo is transformed into micro-Doppler spectrogram at first. After applying the short-time Fourier transforms (STFTs), the Doppler spectrogram is processed with convolution and pooling operations to extract local features. Then, two-layer GRU encodes the temporal patterns. The output of GRUs flows into a fully-connect layer with a softmax activation function to classify the identity of the micro-Doppler spectrogram.

2.1 Motivation

The core idea behind the designed model is that we treat the micro-Doppler spectrogram as time-sequence data rather than general RGB image. The micro-Doppler spectrogram, which is the STFT of the returned radar echo signal, represents the time-varying velocity of the separate body parts exactly. As human activity is a complex combination of the time-varying motion of the whole-body segments, it is possible to analyse human behaviour based on backscattering phenomenon of human body, the time-varying velocity of separate body parts may overlap to some extent in time–frequency plane (which can be seen from Fig. 2), which makes it difficult to analyse the temporal information conveyed in spectrogram directly. To address the problem, we combine...
convolutional neural network and GRUs in one model, convolutional and pooling layer is used to extract local features, GRUs are introduced to analyse feature maps along time axis. Compared with stacked convolutional layers or auto-encoders which are generally used for this task, this unified model fits the time-varying nature of micro-Doppler spectrogram better in theory.

### 2.2 Local feature learning with convolutional neural network

The convolutional neural network used for local feature learning is two-layer architecture. We use $5 \times 5$ frequency–time filter for the first convolutional layer, followed by the same size filter for the second convolutional layer. These filters slide to every position of the spectrogram and compute a new element as a weighted sum of the elements it floats over. The aim of the convolution operation is to extract robust feature from the blurred curves in joint velocity–time plane. Unlike natural images, each ‘pixel’ in one column of spectrogram corresponds to the echo intensity and radial velocity of the illuminated parts of human body in one-time step. Each column records the time-varying characteristics of human motion in a time interval. Max-pooling function replaces the output of the convolutional layer at a certain location with the maximum output within neighbourhood. It can ensure the representation become approximately invariant to small translations of the input. This operation is useful, because motion speed of different human is usually not the same and this difference can be seen as the translation of the corresponding spectrogram. Through max-pooling operation, we can focus more on whether some feature is present than exactly where it is and learn more robust and informative representations of the spectrogram.

### 2.3 Temporal encoding with GRUs

The proposed pattern of temporal encoder is GRUs network with recurrent connections between hidden units. It reads an entire pooled sequence and then produces output to make classification. Built on a simpler architecture compared with long short-term memory network (LSTM), the GRU alleviates the computational burden to some extent and is promising for embedded platform. What is more, in emotion classification from noisy speech tasks, it is found that LSTM performs better than GRU in the case of continuous noise, while GRU performs better for the noise which is not usually continuous [9]. Since the background noise in radar backscattering echo is often non-periodic, we decide the GRU block to encode temporal information.

The information modulation process inside the GRUs seems similar to LSTM except that the GRUs do not have a separate memory cell. The GRU block diagram is illustrated in Fig. 3. The hidden state $h_t$, which is also the activation of the GRU at time $t$, is a linear interpolation between the previous activation $h_{t-1}$ and the candidate activation $\tilde{h}_t$. The update gate $z_t$ decides how much the unit updates its activation. The candidate activation $\tilde{h}_t$ is computed similarly to that of the recurrent unit. The value of reset gate $r_t$ decides whether to remember or forget the previously computed state. Equations (1)–(4) describe the mathematical model of the GRUs. Here, $z_t$ and $r_t$ is the update gate and reset gate separately. $h_t$ is the hidden state, $W$ terms denote different weight matrices, and $\odot$ denotes an element-wise multiplication.

$$h_t = (1 - z_t) \odot h_{t-1} + z \odot \tilde{h}_t.$$  \hfill (1)

$$z_t = \sigma(W_z h_{t-1} + W_x x_t + b_z).$$  \hfill (2)

$$\tilde{h}_t = \tanh(W_h (r_t \odot h_{t-1} + W_x x_t + b)).$$  \hfill (3)

$$r_t = \sigma(W_r h_{t-1} + W_x x_t + b_r).$$  \hfill (4)

### 2.4 Training for implemented model

The model is trained in a fully-supervised way, calculating error between the predicted outputs and ground-truth labels with cross-entropy loss function then backpropagating the gradients from the softmax layer to the convolutional layer. We choose the Adaptive Moment Estimation (Adam) as the gradient descent optimisation method. The size of mini-batches data is 32. Model is trained with a learning rate of 0.001. Drop-out function is introduced during the training phase. As a form of regularisation, it connects the inputs of every dense layer, setting the activation of randomly-selected units during training to zero with probability of 0.5. An open source toolkit Pytorch is used for training the network. The training process is accelerated by NVIDIA GTX 1080 GPU and CUDA library (cuDNN).

### 3 Results

In this section, we evaluate the proposed CNN-GRU model on MOCAP-based non-parametric simulation.
In activity classification, the input data is trimmed to a fixed size. The fine temporal resolution of human activity sequence is composed of walking, running, boxing and kicking as shown in Fig. 5.

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
Conclusion

In this work, we have developed a CNN-GRU neural network that combines convolution operation and temporal encoding to classify human activities based on micro-Doppler spectrogram. The MOCAP database-based experiments show that the proposed model avoids the resolution limit caused by the fixed-size training samples and yields a class probability distribution for every single time step. In contrast to previous works, it is able to recognise and temporally localise activities of varying duration, which shows finer temporal resolution and better generalisation. Featured with fine temporal resolution and generalisation ability, this approach is not only suitable for classification tasks but also promising towards capturing context and reasoning about human pose.

References

[1] Chen, V.C., et al.: ‘Micro-Doppler effect in radar: phenomenon, model, and simulation study’, IEEE Trans. Aerosp. Electron. Syst., 2006, 42, pp. 2–21
[2] Chen, V.C.: ‘The micro-Doppler effect in radar’ (Artech House, London, UK, 2011).

[3] Yin, W., et al.: ‘ECG monitoring system integrated with IR-UWB radar based on CNN’, IEEE Access., 2016, 4, pp. 6344–6351

[4] Juan, Z., Yong, L., Junping, Y.: ‘A novel modulation classification method for FM signals based on the time-frequency distribution and CNN’, IET Radar Sonar Navig., 2017

[5] Lang, Y., Hou, C., et al.: ‘Convolutional neural network for human micro-Doppler classification’. European Microwave Conf., Nuremberg, Germany, 2017.

[6] Kim, Y., Moon, T.: ‘Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks’, IEEE Geosci. Remote Sens. Lett., 2016, 1, pp. 8–12

[7] Park, J., et al.: ‘Micro-Doppler based classification of human aquatic activities via transfer learning of convolutional neural networks’, Sensors, 2016, 16, pp. 1990

[8] Hao, D., Yuan, H., Tian, J.: ‘Transfer learning for human activities classification using micro-Doppler spectrograms’. IEEE ICCEM, Chengdu, China, 2018

[9] Rajib, R.: ‘Gated recurrent unit (GRU) for emotion classification from noisy speech’, Eprint Arxiv, 2016

[10] He, Y., et al.: ‘Range-Doppler surface: a tool to analyse human target in ultra-wideband radar’, IET Radar Sonar Navig., 2015, 9, pp. 1240–1250

[11] Ram, S.S., Ling, H.: ‘Simulation of human micro-Dopplers using computer animation data’. Radar Conf., IEEE, Rome, Italy, 2008, pp. 1–6

[12] Ram, S.S., et al.: ‘Simulation and analysis of human micro-Dopplers in through-wall environments’, IEEE Trans. Geosci. Remote Sens., 2010, 48, pp. 2015–2023

[13] Drillis, R., Contini, R., Bluestein, M.: ‘Body segment parameters. Research division’, S. Artif. Limbs, 1964, 1, pp. 44–66

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)