IET Radar, Sonar & Navigation

Research Article

IET Journals
The Institution of Engineering and Technology

Practical classification of different moving targets using automotive radar and deep neural networks

ISSN 1751-8784
Received on 8th March 2018
Accepted on 5th April 2018
doi: 10.1049/iet-rsn.2018.0103
www.ietdl.org

Aleksandar Angelov[1], Andrew Robertson[2], Roderick Murray-Smith[3], Francesco Fioranelli[1] ✉

[1]*School of Engineering, University of Glasgow, Glasgow, UK*
[2]*NXP Semiconductors, Glasgow, UK*
[3]*School of Computing Science, University of Glasgow, Glasgow, UK*
✉ *E-mail: francesco.fioranelli@glasgow.ac.uk*

Abstract: In this work, the authors present results for classification of different classes of targets (car, single and multiple people, bicycle) using automotive radar data and different neural networks. A fast implementation of radar algorithms for detection, tracking, and micro-Doppler extraction is proposed in conjunction with the automotive radar transceiver TEF810X and microcontroller unit SR32R274 manufactured by NXP Semiconductors. Three different types of neural networks are considered, namely a classic convolutional network, a residual network, and a combination of convolutional and recurrent network, for different classification problems across the four classes of targets recorded. Considerable accuracy (close to 100% in some cases) and low latency of the radar pre-processing prior to classification (~0.55 s to produce a 0.5 s long spectrogram) are demonstrated in this study, and possible shortcomings and outstanding issues are discussed.

1 Introduction

Autonomous vehicles have been gaining significant interest in the past few years, with considerable attention and investments from technology-intensive companies (such as data management and algorithms developers, vehicles and electronic sensors and systems manufacturers), governments and academic research community, and media and the general public [1–3]. As research in this vast field grows, an attempt of standardising the different levels of autonomy that advanced driver assistance systems can enable in ground vehicles has been made, specifying six levels ranging from 0 to 5, from rather standard car accessories such as antilock braking system, to fully autonomous dynamic driving with little to no inputs from the human driver [2].

To achieve complete driving autonomy, the capability of sensing the surrounding environment and other moving entities, other vehicles or humans, and animals, is paramount. Different sensing technologies have been proposed [4]. Cameras are suited for objects classification exploiting colour and texture data, and can be relatively cheap compared with the other types of sensors, but may suffer from the limited depth of view and adverse weather and light conditions, as well as requiring high data processing power, depending on the image classification algorithm.

LiDAR uses rotating laser arrays to generate an accurate 3D map of the surrounding environment around the autonomous vehicle, but this type of sensors are still rather expensive and may require significant computational power to address the adverse effect of light and weather (rainy, foggy, snowy conditions).

Radar sensors provide the advantage of not being affected by light and weather conditions, as well as exploiting mature range-Doppler and classification processing developed for different end applications over the years [5]. However, the applicability and adaptation of these techniques to the specific automotive context, and the development of the most suitable processing to fuse information from different radar channels and heterogeneous sensors are still open research questions. In particular, significant research in the context of automotive radar has been devoted to the issue of detecting and classifying accurately vulnerable road users, such as pedestrians, to preserve their safety.

One of the earliest classification studies on automotive radar reported over 90% accuracy when distinguishing vehicles and pedestrians [6], as well as other objects [7], by extracting features from micro-Doppler (MD) signatures combined with joint probability data association tracking, in order to account for discrepancies in amplitude and shape due to the aspect angle changes. Although the use of trackers worn by vulnerable road users would help their detection and classification [8], the reliability of the whole system would be poor, as relying uniquely upon compliance of them wearing the devices.

Other studies looked at using range-Doppler maps as the domain to perform classification. Object tracking through clustering algorithms and a linear classifier was used to distinguish vehicles and scenarios of walking pedestrians in [9], and in [10] features related to the size, orientation, and frequency of the pedestrians' step were used in conjunction with ordered statistics-constant false alarm rate (OS-CFAR) and density-based cluster algorithm. Further works focused on using different domains of information to achieve vehicles–pedestrians classification, such as [11] through the phase characteristics (coherent/non-coherent) of the object signature, and [12] through features related to the differences in radar cross section (RCS) between the different classes of targets, used together with a support vector machine classifier. As systems working at a higher frequency, tens but also hundreds of GHz, become available, work has been carried out to characterise the radar signatures of pedestrians in the automotive context, such as in [13–15] which considers the frequency ranges around 300 GHz. Another group of studies looked at characterising the radar signatures of cyclists, to highlight differences and similarities with those of pedestrians and vehicles that can be useful to improve their detection and classification [16–18]. Bicycles can travel at significantly higher speed than pedestrians and present high manoeuvrability on the road, as well as at the same time exhibiting low RCS compared with vehicles; they are therefore a challenging class of targets for automotive radar applications.

Many of the classification studies considered some form of 'handcrafted' extraction process on the radar data in order to obtain the most suitable combination of features to maximise classification accuracy [19, 20]; this often requires significant expertise and inputs from the human radar operator/engineer, thus not lending too well to achieving reliable automatic classification in the large diversity of situations and scenarios expected for automotive radar. To address this issue, an emerging stream of work in the literature has been looking at neural networks as a

Table 1 Radar parameters for the data analysed in this paper

number of samples per chirp	512
number of chirps per frame	256
chirp bandwidth	1.0 GHz
chirp duration	25.6 μs
carrier frequency	76.5 GHz
analogue to digital converter (ADC) sampling frequency	20 MHz
transmitter-receiver (TX/RX) channels	1/4
radar field of view (azimuth and elevation)	±35° at 50 m/7.5°

processing tool to bypass the feature extraction step and enable automatic selection of the most suitable features and meaningful information for classification within the network itself. One of the first work in this aspect was [21], in which deep convolutional neural networks (DCNNs) were given spectrograms directly as input data to distinguish four classes of targets (humans, dogs, horses and cars signatures), and seven different human activities. The DCNN was a scaled-down model of the famous VGG16 (Visual Geometry Group) network that won the ImageNet classification challenge in 2014, and accuracy in the region of 91% was achieved for target identification. Further work on the use of convolutional neural networks (CNNs) in the context of human activity recognition for assisted living has been presented [22], focusing on aspects such as most suitable pre-processing and time-frequency distribution for the MD signatures [23], combination of information from different radar domains including range-Doppler and range-time to enhance performance [24], different architectures mixing auto-encoders with CNNs [25, 26], and challenges and strategies to train deep networks effectively with limited experimental radar data available [27]. Other works have looked at classifying different human gaits in the context of area surveillance using a ground-based radar, in particular identifying individual pedestrians as opposed to group of multiple people, either using CNNs or recurrent neural networks (RNNs) on the spectrograms [28, 29], and at classification of armed/unarmed personnel using a multi-static radar [30].

In this work, we present and discuss a modular pipelined approach to achieve near real-time radar data processing and multiple moving object tracking and to subsequently classify these objects. Three different neural network architectures have been explored – a downscaled version of the network VGG16, utilising the same block structure; the very deep ResNET-50 [31], which uses shortcuts between network blocks to avoid over-fitting and achieve better generalisation; and an innovative CNN + long short-term memory (LSTM) architecture, which is able to extract features from MD spectrogram segments, and learn their representation as time series (sequences of data). This is an innovative approach, as the radar data will be considered by the LSTM network part not as snapshot spectrogram images (as currently done in many works in the literature [22–28]), but as temporal data sequences. Although demonstrated on preliminary results on a small experimental dataset, this classification approach may prove well suited to radar data, exploiting the inherent information from a sequence of radar waveforms, rather than casting the problem as the classification of images.

Although the dataset of experimental samples is small, the work presented here aims to demonstrate the potential of this approach. It provides a proof of concept evaluation of the lean implementation of radar signal processing necessary for radar MD-based classification, and of different architectures of neural networks that do not require manual fine-tuning of parameters of external inputs to guide the feature extraction process.

These processing steps have been implemented with the following objectives:

- To use real experimentally-gathered data for training and testing the neural networks, in order to investigate the generalisation capabilities of the network architectures beyond the ideal cases of using simulated data. This includes the implementation of radar signal processing for detection and tracking of multiple targets, which can provide good performance even in the presence of significant noise generated within the radar system.
- To have a significantly low classification latency – below 0.5 s, since studies have shown that the average driver reaction time is around 0.7 s [32].
- To use the MD spectrograms directly as input to the classifier and network, avoiding handcrafted features (e.g. MD bandwidth and frequency, Cepstral coefficients, moments of vectors extracted by singular value decomposition, and many others proposed in the literature [20]). This allows avoiding possible loss of relevant information and fine-tuning of the many parameters involved when defining the feature extraction algorithms.

The remainder of this paper is organised as follow. Section 2 describes the experimental setup, the radar kit used, and the data collection protocol. Section 3 introduces the implementation of the radar signal processing developed, and the structure of the neural networks used in this study. Section 4 presents comments on the experimental results. Finally, Section 5 draws conclusions and discusses some possible future work.

2 Experimental setup and data collection

All data have been collected using the TEF810X fully integrated automotive radar transceiver manufactured by NXP Semiconductors and S32R274 radar micro-controller unit. The radar operation mode was configured as frequency modulated continuous wave, with linear chirp modulation, and the parameters, shown in Table 1. These parameters were empirically found to provide the clearest MD signatures at visual inspection, as well as providing a reasonable compromise in terms of range resolution, Doppler unambiguous range, and data throughput for fast transferring and processing. The system had one transmitter and four receiver channels, and digitised data were transferred from the micro-controller unit to a computer via User Datagram Protocol (UDP) packets. These packets were then decoded to form ‘frames’, matrices with 512 rows and 256 columns, which essentially correspond to range-time matrices with 256 radar chirp and 256 [after removing fast Fourier transform (FFT) mirroring] range bins for each chirp. The time for one frame to be transmitted and received for processing (for all four receiver channels) is set internally in the micro-controller unit (MCU) as 50 ms, and this is a firmware parameter that cannot be modified in this version of the system.

Three different types of movements and targets were recorded, namely a single person walking at an average speed of 4–5 km/h (type 1), a car accelerating and decelerating (type 2), a bicyclist following the trajectory of an eight-figure (type 3), and finally two people walking side by side (type 4). All activity types were performed with objects moving towards and away from the radar covering a distance of around 0–17 m, at 0 degree aspect angle (radial trajectory with respect to the line-of-sight of the radar), with some little variability for the bicyclist to turn when cycling towards and away from the radar.

The radar was positioned ∼0.7 m above the ground, to correspond to the height at which the automotive radar is usually mounted on a car. This also allows to capture the micro-motions contributing most to the MD effect, such as hands, torso and the upper leg parts from walking people; the body of a moving vehicle; and bicycle frame/pedalling legs. Around 30 min of data were collected for the single person walking and the car, and ∼15 min each for the bicycle and the multiple people class. The raw digitised data were then divided into blocks, which are the starting point of the processing steps described in the next section.

3 Data processing and neural networks architecture

As described in Section 1, a lot of research has been conducted on target classification using the MD signature of objects. When the target signature is spread across many different range bins, the

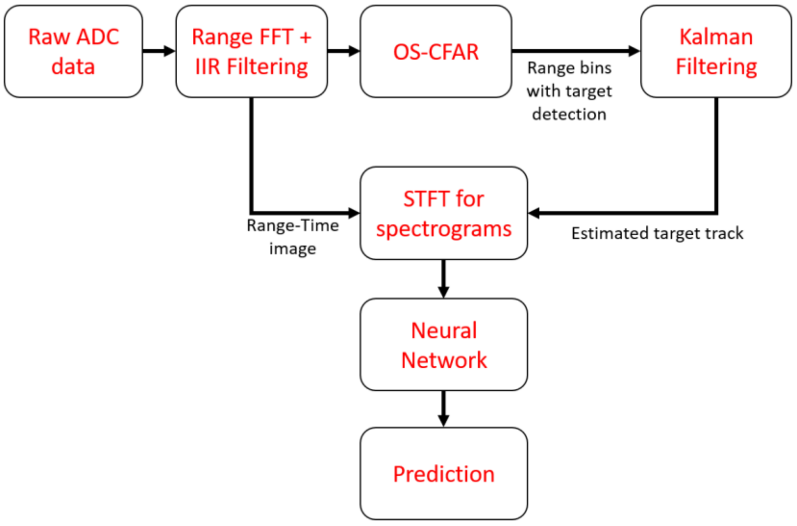

Fig. 1 *Block diagram of the multi-target classification system*

different target contributions need to be aggregated prior to performing short time Fourier transform (STFT), or an alternative time-frequency distribution, and this is even more important in case of multiple targets crossing their trajectories. To address this issue and easily track multiple moving targets, we have implemented the following processing on the raw data obtained from the NXP radar. The different processing steps have been summarised in Fig. 1.

- Perform FFT on raw digitised data to convert them into the range-time domain, and apply a fourth-order Butterworth infinite impulse response high-pass filter with 0.04 Hz cut-off to remove stationary objects (i.e. objects with Doppler signature at 0 Hz or close to that value).
- Apply OS-CFAR algorithm [33] to perform target detection and reduce the undesired contribution from noise and clutter.
- Detect the position of the targets (i.e. the range bins they occupy) for a given frame and store these coordinates in a detection matrix.
- Input the detection matrix frame-wise in an algorithm, which combines constant acceleration Kalman filtering and the Hungarian algorithm [34]. The former would produce a better estimation of the target position, as well as continue to output predictions, even if frames are temporarily lost or corrupted. The latter would constantly assign identities to the object detections, based on the estimates from the Kalman filter. The algorithm can also take into consideration new objects entering the radar field of view, or those leaving it, using markers for each track.
- Concatenate several range-time frames and generate segments of MD signatures using the object track position estimates, i.e. the range bins where the target signature is located. The duration of the overall MD signature can be varied depending on the classification algorithm just by concatenating more or less frames together.
- Use the generated MD spectrograms to train and test classifiers based on neural networks.

Using the aforementioned approach, samples of MD signatures have been generated by concatenating eight 0.25 s segments to provide spectrograms that are 2 s long. Examples of MD spectrograms plotted using the method described above for the different cases are shown in Fig. 2, with one spectrogram for each class of targets considered in this work. Even through visual inspection, it is possible to see some discriminant features of the different classes. For example, the single human (Fig. 2*a*) appears to present some peaks around the main Doppler component, as expected for the swinging of limbs. This effect becomes more blurred for multiple people (Fig. 2*b*) because their movements are not synchronised. For the car class (Fig. 2*d*), we can see a clear main Doppler shift with no major additional components, whereas the bicycle (Fig. 2*c*) presents an intermediate situation with a clear main Doppler component, plus some additional effects due to the movement of the legs while cycling. The STFT window size was 512 points (equal to two concatenated range-time frames), with 95% overlap. Although segmentation is present as an artefact of the concatenation process, it does not seem to affect the learning capabilities of the neural network classifiers, as will be further demonstrated in the next section.

After removing the unsuitable datasets where there was false target detection and hence no clear MD signature, we generated 60 samples for movement types 1 and 2 each (single person walking and car), 22 samples for type 3 (bicycle), and 44 samples for type 4 (two people walking together). The samples for each class are created using data collected at different time instances rather than continuously and this helps reduce the intra-class correlation between the samples. The data were partitioned into training and testing subsets to validate the neural network performance with an 80/20% proportion, and this partition was performed randomly. The networks used the training data for learning and the test data for validation. Furthermore, all evaluations were performed using the same number of samples for each class, to avoid class imbalance, with the final number of samples governed by the class with the least datasets. Four types of evaluations were performed, in particular:

- Binary classification of type 1 versus type 2, a single person walking versus car.
- Three-class problem with the single car, a single person, and single bicycle as classes of interest.
- Three-class problem with the single car, a single person, and two people.
- Four-class problem with all the available data.

Three different network architectures were used for the classification of experimental data, detailed as follows. A pictorial representation of the different layers in each architecture is shown in Fig. 3, where different functionalities of the layers have been highlighted in different colours. The input to the networks is a 3D structure containing the 2 s long spectrogram samples for each of the four receiver channels of the radar, so that the overall dimensions of each input samples are 4 (number of channels) ×512 (number of Doppler bins for each spectrogram) × 120 (number of time bins for each spectrogram, for eight segments). Each spectrogram is normalised between 0 and 1 and centred around the mean value.

Network I: This is a VGG-like convolutional neural network, as in Fig. 3*a*. Each ‘block’ consists of a convolutional layer, with a different number of filters of the same size, and a pooling layer, which reduces the dimensionality of the block output by a factor of 4, selecting the maximum value in the kernel. The convolutional filters would learn features from the datasets, specific for each class. The addition of a dropout layer (20%) has been proven in the

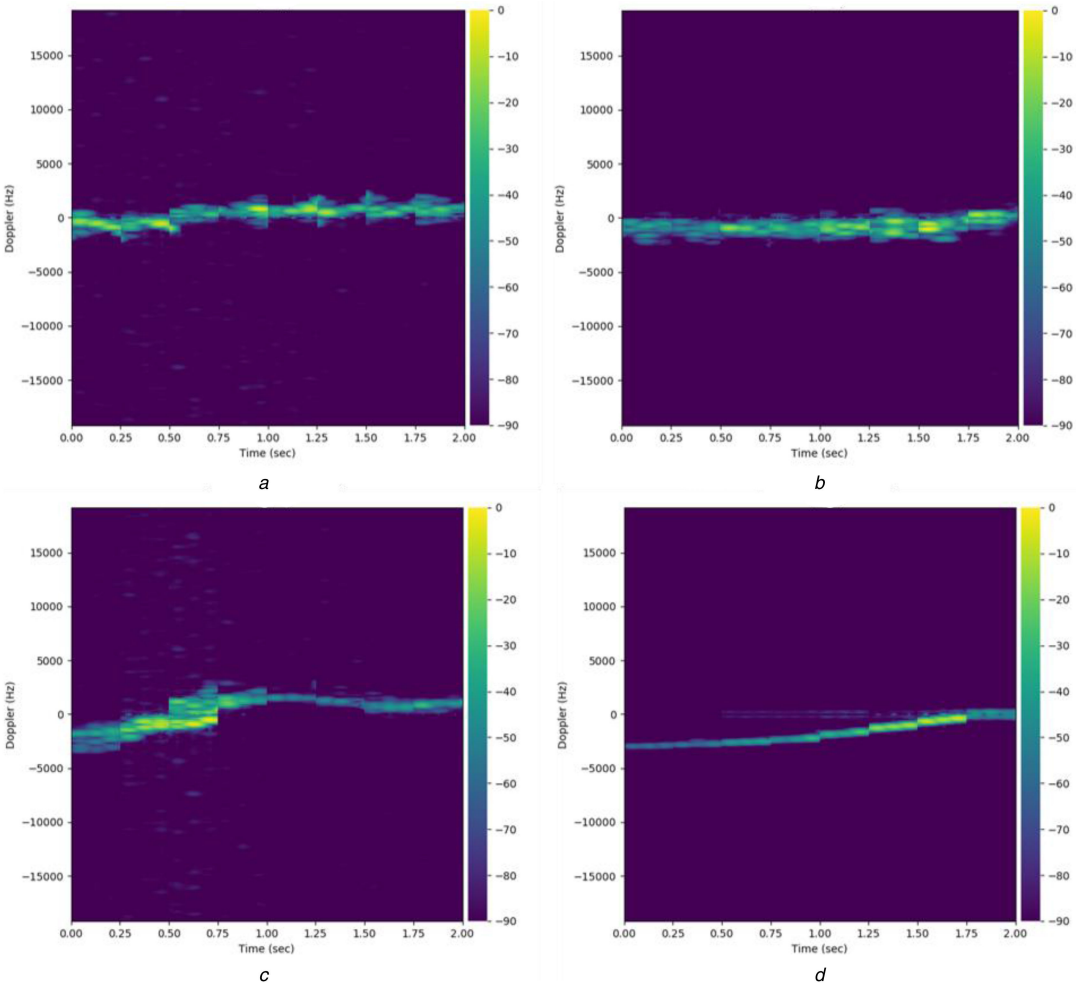

Fig. 2 *Examples of spectrograms for different targets:*
(a) Single person walking, ***(b)*** Two people walking together, ***(c)*** Bicycle, ***(d)*** Car

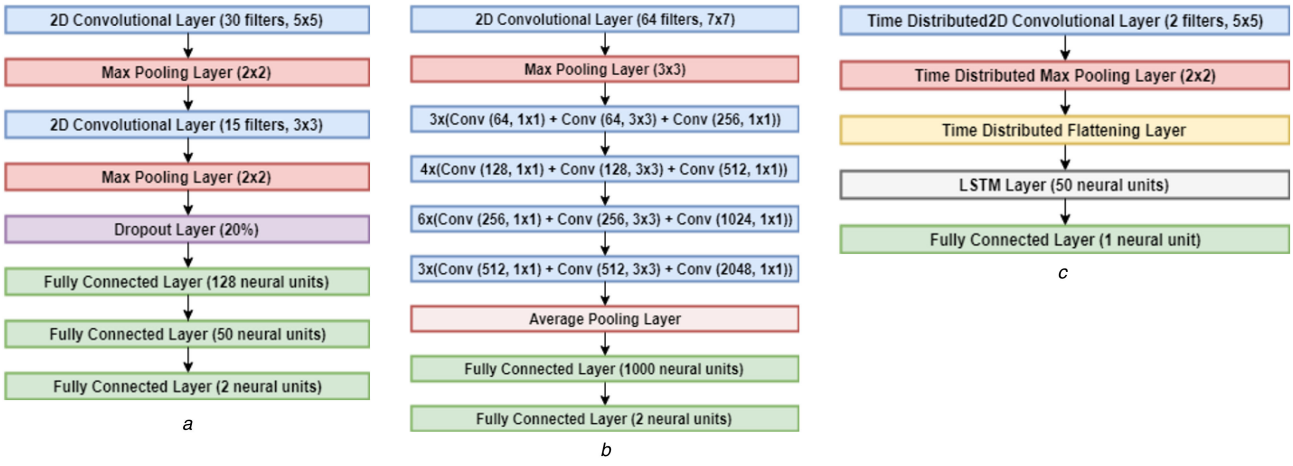

Fig. 3 *Representation of the different network architectures*
(a) Convolutional neural network similar to VGG type, ***(b)*** Convolutional residual network, ***(c)*** Combination of convolutional and recurrent LSTM network

literature to improve learning regularisation [35], which is paramount for a small amount of data like in this case. Finally, three fully connected layers are used, where each neural unit in the layer is connected to the rest. Rectified linear unit activation function has been used in all but the last layer, where the function used is Softmax. In this and in all subsequent models, Adam optimizer algorithm was implemented, due to its very fast convergence rate and reliability.

Network II: This architecture is shown in Fig. 3*b* and is based on a residual network, in which the input and output of a convolutional block are connected via a shortcut. In very deep networks of the VGG type, the back-propagation gradient tends to diminish as it propagates through the network layers, hence having little effect on the initial ones. This is partly because, in a VGG type architecture, subsequent blocks have to learn data features anew, from the output of the preceding block. However, due to the shortcuts in a ResNet architecture, the blocks only have to learn the residual of the output from the preceding one. This largely improves the representation capability, allowing for correct classification of data with very similar features, as it is the case with radar spectrograms. In this work, the ResNet-50 architecture has been used [31].

Network III: This architecture is based on RNNs and is shown in Fig. 3*b*. RNNs have been used for years to analyse time series

Table 2 Five-fold evaluation test accuracy when using a different number of radar channels for binary classification car versus person walking

Number of channels	1	2	3	4
test accuracy /standard deviation	98.33/2.04%	98.33/2.04%	97.50/2.04%	98.33/2.04%

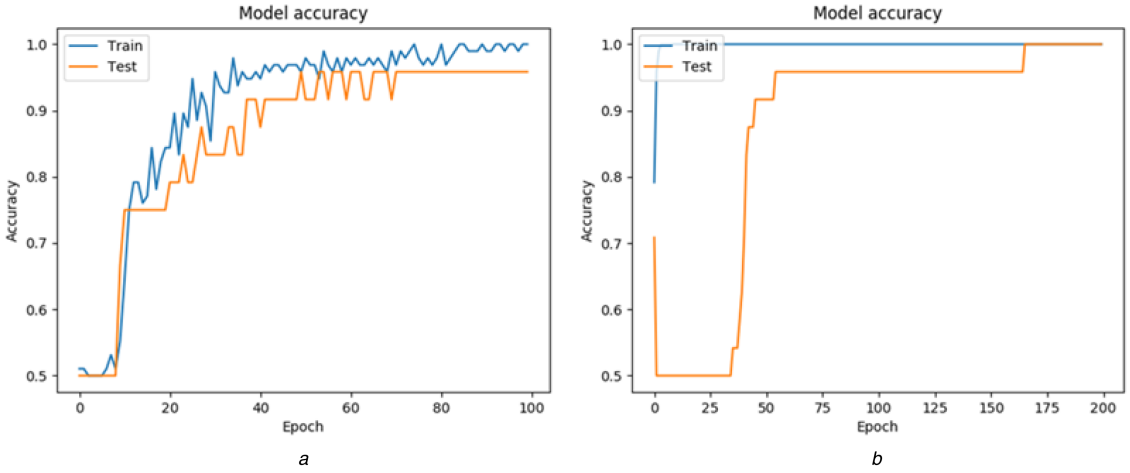

Fig. 4 *Neural network performance (accuracy) for* ***(a)*** VGG-like network, ***(b)*** Residual network

data, for example in speech processing and acoustics, and have been demonstrated to work very well to predict and classify sequences of data. Out of different types of RNNs, LSTM networks are mainly used in practice, because they can overcome the issue of vanishing/exploding back-propagation gradients [36] and are able to learn the representation of longer sequences (around 1000 instances) compared with other architectures of recurrent networks. In this work, we have modelled each 0.25 s-long segment in a 2 s spectrogram sample as instances from a data sequence, with variable length, depending on the requirements. The convolutional part of the network would extract features from single segments, which would then serve as input to the LSTM part, analysing their progression and evolution with time.

4 Classification results

Initially, the effect on the classification performance of using data from a subset of the available four receiver channels is evaluated using 5-fold validation with the VGG-like convolutional network (see Fig. 3*a*). This was done on the binary classification problem of distinguishing a single person and a car. The results in terms of classification accuracy and standard deviation across the 5-fold tests are shown in Table 2, where the number of channels used increased from one to all four We can see that the results are very similar with little or no difference with the number of channels. This may be because the receiver antennas are mounted very close to each other in this version of the radar kit, hence the aspect angles on the target at ranges of a few metres are practically the same so that the different channels do not seem to provide additional information. Nevertheless, all further evaluations have been performed using samples containing all four channels, due to the expected increase in the number of hardware channels in a near future as technology improves. This would provide bigger data discrepancy, hence better network generalisation for objects moving at different aspect angles, especially for less favourable trajectories for MD-based classification (i.e. trajectories which are tangential or close to tangential to the radar field of view).

When evaluating the performance of a neural network, two main indicators are generally used, namely the accuracy, which shows the percentage of correctly classified samples, and the logarithmic loss measure, which is the negative logarithm of the network-predicted probability for a dataset to belong to a certain class, taking into account the true class label. Back-propagation algorithms strive to minimise this loss and forcing this to zero by adjusting the weights of the network layers at the training stage. By analysing how the loss gradient changes over time, one can judge for the generalisation capabilities of a network, i.e. whether it overfits on the training data, compromising its ability to classify correctly new test/validation data.

An initial test compared the VGG-like network and the residual network for the binary classification problem of moving car versus single person walking. The validation accuracy of the residual network achieved 100% (Fig. 4) after only 200 epochs with a batch size of eight datasets (i.e. the weights of the network have been updated every eight input samples). The validation accuracy of the VGG-like network was in the range of 98%, as shown in Fig. 4 and Table 2. In terms of loss function for training and validation, Fig. 5 shows these over different epochs.

When using the VGG-like CNN (Fig. 5*a*), both training and test losses fluctuate heavily, and although the test loss continues to decrease, its value at epoch 100 is significantly above the training loss, which tends to zero, and this may be a sign of overfitting as the network has nearly exhausted its capability to learn from the available data. In contrast, the residual network losses exhibit an almost non-existent fluctuation, even using a very small number of datasets as in this case. Both training and validation loss continue to decrease, and their values at epoch 200 may be an indication of a significant potential for further learning, as the training loss has not reached values close to zero. This, combined with the very high accuracy score close to 100%, seems to confirm the assumption that the use of residual networks for this classification task would yield better, more generalised results.

The same binary classification problem has been evaluated on the CNN-LSTM network architecture and the results are presented in Fig. 6 in terms of accuracy and loss function for training and testing. In this case, we have considered two different temporal durations of the input samples, namely 2 s (equal to eight MD segments) as done previously for the other networks, and 0.5 s (equal to just two segments) in order to reduce the latency required to provide a classification result.

Comparing the performance of this CNN-LSTM network on 2 s long samples (Fig. 5*a*) with the previous architectures (Fig. 4), we can see that the overall validation accuracy is reduced (~92%) and there is very significant overfitting, as while the training loss has reached values close to zero, the validation loss remains stationary at a non-zero value. Looking at the results for 0.5 s long samples, despite the decreased latency, the overall accuracy appears to be very significant, in the range of 99%. This increased accuracy could be related to the combined effect of having a larger dataset of samples for training and testing (as each 2 s spectrogram was divided into 0.5 s segments, with a 4-fold increase in the dataset size), but also to the fact that the tested LSTM architecture might

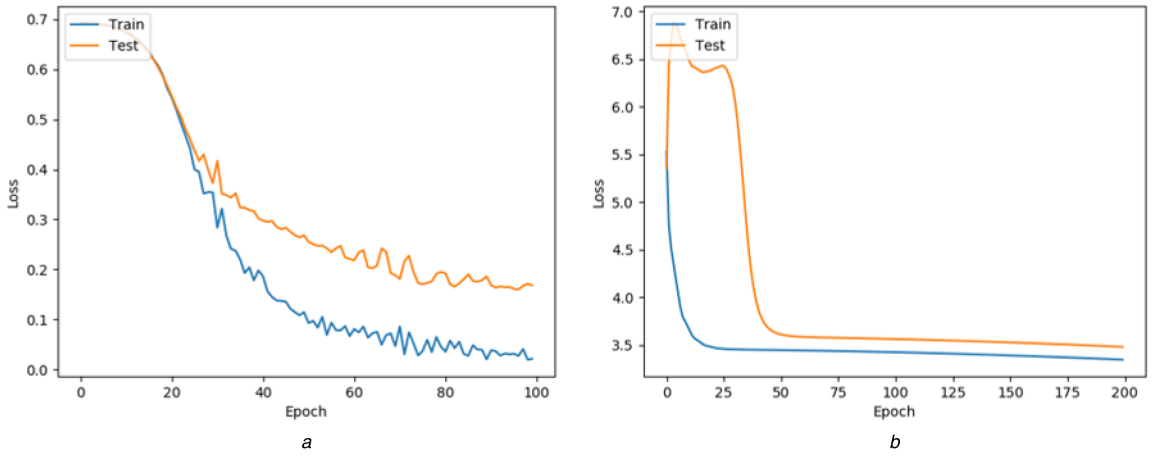

Fig. 5 *Neural network performance (loss function) for*
(a) VGG-like network, ***(b)*** Residual network

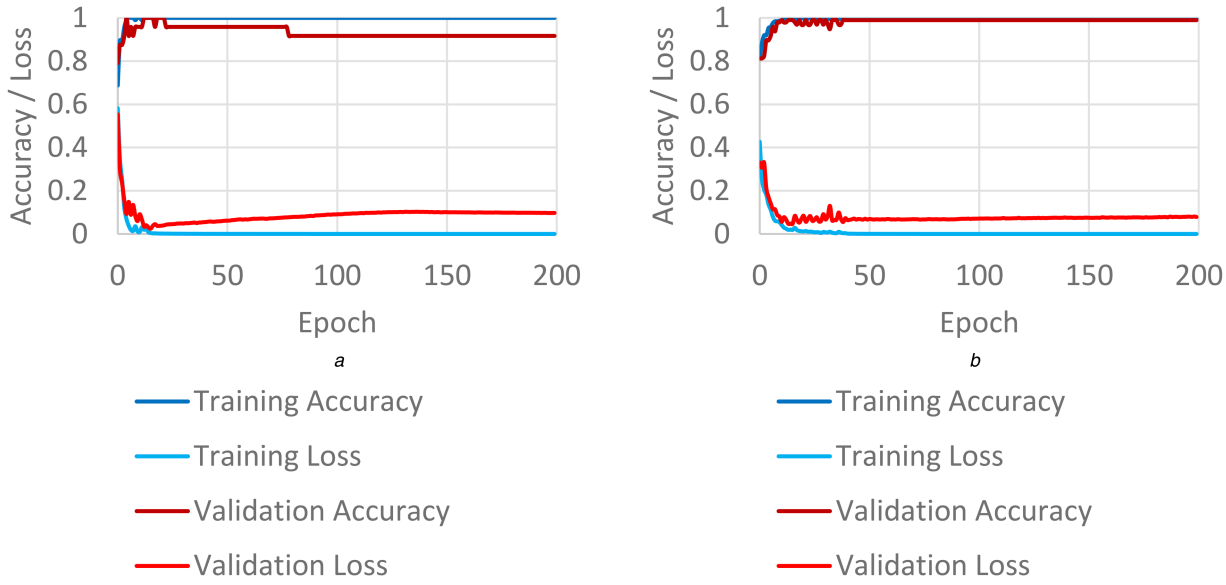

Fig. 6 *CNN-LSTM network performance (accuracy and loss function for both training and validation) when using*
(a) 2 s long inputs, ***(b)*** 0.5 s long inputs

Table 3 Test accuracy for two network architectures evaluated on three class problems

Evaluation/ network type	VGG-like CNN (2 s long datasets)	VGG-like CNN (0.5 s long datasets)	CNN-LSTM (2 s long datasets)	CNN-LSTM (0.5 s long datasets)
car-person-bicycle classification	79%	83%	93%	83%
car-person-2 people classification	81%	78%	80%	84%

be more capable to infer relevant features from shorter sequences. There is some residual overfitting (validation loss stationary with training loss already close to zero), and this can be caused by the use of a relatively shallow convolutional layer before the LSTM layer in this architecture (see Fig. 3*c*), as this may not be able to learn relevant features from the input data.

Subsequently, we have analysed three-class problems by adding to the binary dataset with moving car data and a single person walking data, either bicycle data or data for two people walking together. These three-classes problems have been tested with different network architectures and some results are shown in Table 3.

The results in Table 3 suggest that adding a different class of targets can have a very significant impact on the results, and in general the accuracy is reduced compared with the binary class scenario analysed before. The CNN-LSTM case shows increased accuracy when using shorter sequences (from 2 to 0.5 s) for the classification of the car and single or multiple people, as observed in Fig. 5. This is not true for the classification scenario involving the bicycle, where the accuracy degrades from ~93 to 83%, and this could be due to the fact the bicycle and car signatures are similar in such a short period of time (especially as at times the cyclist was not pedalling but just coasting with the bicycle). The VGG-like network presents results around 80% for both three-class problems, which appears to suggest that extending the dwell time on target for extraction of MD signatures does not provide a significant classification benefit. This may be due to the specific settings of the radiofrequency radar parameters and spectrogram extraction algorithm, which could not capture enough details to differentiate the spectrograms belonging to each target class. On average, the CNN-LSTM architecture appears to provide higher accuracy and therefore better capability to generalise on additional target classes, making it a promising approach.

In terms of loss functions (not shown here for conciseness), the CNN-LSTM architecture suffers significantly from overfitting problem as already noted when commenting Fig. 6 for the binary classification problem. This poor performance can be linked to the very shallow convolutional part (two filters, with 5×5 kernel size), which is not able to learn the discriminating details between one person and two people walking. To evaluate this hypothesis, we

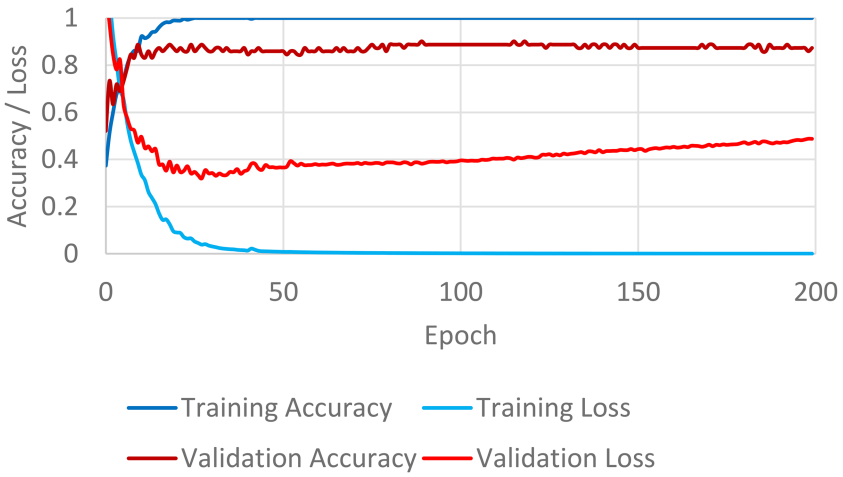

Fig. 7 *Alternative CNN-LSTM network performance (accuracy and loss function for both training and validation) for the four-class classification problem*

Table 4 Test accuracy for three types of networks (VGG-like, CNN-LSTM, and VGG-LSTM) on all considered problems, with regularisation and batch normalisation

Evaluation/ network type	VGG-like CNN (2 s long datasets)	VGG-like CNN (0.5 s long datasets)	CNN-LSTM (2 s long datasets)	CNN-LSTM (0.5 s long datasets)
car-person-bicycle classification	78.6%	81.1%	50%	73.5%
car-person-2 people classification	77.8%	88.6%	44.4%	78.3%
all-4-classes-classification (VGG LSTM)	—	—	—	70%

have run further tests by substituting the convolutional layer of the CNN-LSTM in Fig. 3*c* with the VGG-like model in Fig. 3*a* (excluding the dropout and fully connected layers). This creates an alternative CNN-LSTM architecture, where the initial convolutional part is much deeper than the initial choice with just one layer. With this alternative architecture, we managed to achieve 87.3% test accuracy after only 200 epochs (an increase of about 3–4% with respect to the results in Table 3). This was achieved on a more challenging classification scenario, which includes all four classes of interest (moving car, a single person walking, bicycle, and two people walking) and low latency with 0.5 s long spectrograms. The results in terms of accuracy and loss function for training and testing are shown in Fig. 7. Although overfitting is still present, the validation loss is expected to reduce by using a residual network approach for the convolutional part.

For completeness, all the above-mentioned tests have been repeated by training using a kernel (weight) and activity (activation function) regularising approaches applied on the last fully connected output layer, with a penalty of 0.01 and 0.001 for the VGG-like and CNN LSTM networks, respectively. Furthermore, batch normalisation layers have been used after each activation function. Both these are common strategies in the literature to help improve the performance, as they should, in theory, improve the generalisation capabilities of the network in particular [37]. However, in our case these results appear to show that the classification accuracy has degraded, as per summary provided in Table 4 (with 200 epochs training).

It can be seen that the regularisation and batch normalisation can at times improve the performance for CNNs (for example, for the car-person-2 people problem with 0.5 s long datasets, the accuracy increased ∼10% compared with Table 3), but this is not always consistent (for example, with the other three-class problem involving the bicycle the accuracy degraded from 83 to 81%). Furthermore, results appear to become worse for the CNN-LSTM network cases. However, the training history over epoch (not shown here for conciseness) shows a very large variability of the accuracy, possibly meaning that the CNN-LSTMs need more time and longer training to converge and exploit effectively regularisation and batch normalisation (as happened for the VGG-like network in some cases). This will be considered in future work, as well as investigating the most suitable hyper-parameter values (for example the penalty ratio of the regularisation process) for these specific classification problems, with a small amount of data available for training effectively.

5 Conclusions and future work

This study has presented results for classification problems in the automotive radar context using different neural network architectures. Although validated on a small set of experimental data, these proof-of-concept results demonstrated benefits (classification close to 100% in some cases) and potential shortcomings (overfitting and non-robust generalisation) of different networks, as well as the importance of choosing suitable radar parameters and radar signal processing (proper target detection and tracking) to provide the best input data as possible to the networks.

Residual networks appear to provide improved performance compared with simpler convolutional networks when the radar classification is cast as an image recognition problem among different spectrograms. Combinations of convolutional and recurrent networks have also been proposed. One potential problem with these networks is the overfitting for scenarios with a low amount of data available, as in this paper, especially if the initial convolutional part is not deep enough to capture the subtle differences between spectrograms of different classes of targets. Further work is needed to characterise how classification performance could be improved by adding a robust residual network as convolutional part and multiple LSTM layers in a mixed CNN-LSTM architecture explored in this study. Furthermore, one could consider purely recurrent network architectures without the convolutional part, so that the radar classification problem is cast as a data sequence classification (sequence of radar pulses), rather than reducing this to an image discrimination problem. This would allow exploring the information in different radar domains other than Doppler-time (MD) patterns, such as sequences of range profiles or even raw complex data, which would be an interesting innovative approach. In any case, priority for further work should aim at collecting a larger experimental dataset for the training and validation of the chosen neural networks, especially for the very deep ones where many parameters need to be tuned. This availability of radar data for deep learning is a known issue, for both collecting and properly labelling the data, and strategies such as transfer learning and pre-training are being explored for its mitigation [27].

In terms of radar architecture, the availability of additional channels in multiple-input multiple-output, spatially distributed

architectures would benefit the classification performance if data from additional aspect angles to the targets of interest can be captured. In terms of radar signal processing, the detection, tracking, and MD extraction presented in this study have been achieved in ∼0.55 s computational time for 0.5 s long MD (on a Python based implementation on a desktop machine). This shows that the overhead latency of the radar processing is not very significant, with respect to the amount of dwell time on the target to collect data (0.5 s is fairly close to an average gait cycle of a human walking). Moving away from MD-based classification, perhaps exploiting other sequential radar domain with LSTMs as mentioned before, could enable to avoid this minimal dwell time requirement. Implementations in C++ or other languages more suitable for low level programming in micro-controller units could also allow for faster classification time and reduced latency, and firmware improvements could speed up the data transfer from the radar chip to the processing unit (50 ms for a single frame in this work).

Additional further work could look at making the clutter cancellation filter adaptive, taking into account the velocity and the orientation of the vehicle carrying the radar, which for simplicity has been considered stationary in this work. Fusion of data from heterogeneous sensors (be it cameras, Lidar or other sensors) is also an interesting area for further work to improve classification performance and mutual learning of the classifiers.

Finally, research on different architectures of networks should focus on evolution and predictability of their learning capability and performance, making sure that this adheres to the relevant regulations in the automotive sector for standardisation and safety issues.

6 Acknowledgments

This research is supported in part by the EPSRC UK Quantum Technology Programme (Grant No. EP/ M01326X/1) and EU Horizon 2020 project MoreGrasp, award number 643955.

7 References

[1] Jenn, U.: 'The road to driverless cars: 1925–2025'. Available at http://www.engineering.com/DesignerEdge/DesignerEdgeArticles/ArticleID/12665/The-Road-to-Driverless-Cars-1925-2025.aspx, 2016, accessed February 2018

[2] SAE International: 'Taxonomy and definitions for terms related to driving automation systems for On-road motor vehicles'. Available at http://standards.sae.org/j3016_201609/, 2016, accessed February 2018

[3] '*Expect the unexpected – an IET transport sector report on the unintended consequences of connected and autonomous vehicles*' (Institution of Engineering and Technology, 2017) accessed February 2018

[4] 'Radar, camera, lidar, and V2X for autonomous cars'. Available at https://blog.nxp.com/automotive/radar-camera-and-lidar-for-autonomous-cars, 2017, accessed February 2018

[5] Hasch, J.: 'Driving towards 2020: automotive radar technology trends'. 2015 IEEE MTT-S Int. Conf. on Microwaves for Intelligent Mobility (ICMIM), Heidelberg, 2015, pp. 1–4

[6] Heuel, S., Rohling, H.: 'Two-stage pedestrian classification in automotive radar systems'. 2011 12th Int. Radar Symp. (IRS), 2011, pp. 477–484

[7] Heuel, S., Rohling, H.: 'Two-stage pedestrian classification in automotive radar systems'. 2012 13th Int. Radar Symp. (IRS), 2012, pp. 39–44

[8] Saebboe, J., Viikari, V., Varpula, T., *et al.*: 'Harmonic automotive radar for VRU classification'. 2009 Int. Radar Conf. on 'Surveillance for a Safer World' (RADAR 2009), Bordeaux, 2009, pp. 1–5

[9] Sorowka, P., Rohling, H.: 'Pedestrian classification with 24 GHz chirp sequence radar'. 2015 16th Int. Radar Symp. (IRS), 2015, pp. 167–173

[10] Schubert, E., Meinl, F., Kunert, M., *et al.*: 'Clustering of high-resolution automotive radar detections and subsequent feature extraction for classification of road users'. 2015 16th Int. Radar Symp. (IRS), 2015, pp. 174–179

[11] Lee, J., Kim, D., Jeong, S., *et al.*: 'Target classification scheme using phase characteristics for automotive FMCW radar'. *IET Electron. Lett.*, 2016, **52**, (25), pp. 2061–2063

[12] Lee, S., Yoon, Y.J., Lee, J.E., *et al.*: 'Human–vehicle classification using feature-based SVM in 77-GHz automotive FMCW radar', *IET Radar Sonar Navig.*, 2017, **11**, (10), pp. 1589–1596

[13] Marchetti, E., Du, R., Noruzian, F., *et al.*: 'Radar reflectivity and motion characteristics of pedestrians at 300 GHz'. 2017 European Radar Conf. (EURAD), Nuremberg, 2017, pp. 57–60

[14] Marchetti, E., Du, R., Noruzian, F., *et al.*: 'Comparison of pedestrian reflectivities at 24 and 300 GHz'. 2017 18th Int. Radar Symp. (IRS), Prague, 2017, pp. 1–7

[15] Gashinova, M., Hoare, E., Stove, A.: 'Predicted sensitivity of a 300 GHz FMCW radar to pedestrians'. 2016 European Radar Conf. (EuRAD), London, 2016, pp. 350–353

[16] Belgiovane, D., Chen, C.C.: 'Micro-Doppler characteristics of pedestrians and bicycles for automotive radar sensors at 77 GHz'. 2017 11th European Conf. on Antennas and Propagation (EUCAP), Paris, 2017, pp. 2912–2916

[17] Belgiovane, D., Chen, C.C.: 'Bicycles and human riders backscattering at 77 GHz for automotive radar'. 2016 10th European Conf. on Antennas and Propagation (EuCAP), Davos, 2016, pp. 1–5

[18] Stolz, M., Schubert, E., Meinl, F., *et al.*: 'Multi-target reflection point model of cyclists for automotive radar'. 2017 European Radar Conf. (EURAD), Nuremberg, 2017, pp. 94–97

[19] Tahmoush, D.: 'Review of micro-Doppler signatures', *IET Radar Sonar Navig.*, 2015, **9**, (9), pp. 1140–1146

[20] Fioranelli, F., Ritchie, M., Gürbüz, S., *et al.*: 'Feature diversity for optimized human micro-Doppler classification using multistatic radar', *IEEE Trans. Aerosp. Electron. Syst.*, 2017, **53**, (2), pp. 640–654

[21] 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, **13**, (1), pp. 8–12

[22] Jokanović, B., Amin, M.: 'Fall detection using deep learning in range-Doppler radars', *IEEE Trans. Aerosp. Electron. Syst., PP, (99)*, 2018, **52**, (1), pp. 180–189

[23] Jokanovic, B., Amin, M.G., Ahmad, F.: 'Effect of data representations on deep learning in fall detection'. 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Rio de Janerio, 2016, pp. 1–5

[24] Jokanovic, B., Amin, M., Erol, B.: 'Multiple joint-variable domains recognition of human motion'. 2017 IEEE Radar Conf. (RadarConf), Seattle, WA, 2017, pp. 0948–0952

[25] Seyfioğlu, M.S., Gürbüz, S.Z., Özbayoğlu, A.M., *et al.*: 'Deep learning of micro-Doppler features for aided and unaided gait recognition'. 2017 IEEE Radar Conf. (RadarConf), Seattle, WA, 2017, pp. 1125–1130

[26] Parashar, K.N., Oveneke, M.C., Rykunov, M., *et al.*: 'Micro-Doppler feature extraction using convolutional auto-encoders for low latency target classification'. 2017 IEEE Radar Conf. (RadarConf), Seattle, WA, 2017, pp. 1739–1744

[27] Seyfioğlu, M.S., Gürbüz, S.Z.: 'Deep neural network initialization methods for micro-Doppler classification with low training sample support', *IEEE Geosci. Remote Sens. Lett.*, 2017, **14**, (12), pp. 2462–2466

[28] Trommel, R.P., Harmanny, R.I.A., Cifola, L., *et al.*: 'Multi-target human gait classification using deep convolutional neural networks on micro-Doppler spectrograms'. Radar Conf. (EuRAD), 2016 European, 2016, pp. 81–84

[29] Klarenbeek, G., Harmanny, R.I.A., Cifola, L.: 'Multi-target human gait classification using LSTM recurrent neural networks applied to micro-Doppler'. 2017 European Radar Conf. (EURAD), Nuremberg, 2017, pp. 167–170

[30] Patel, J.S., Fioranelli, F., Ritchie, M., *et al.*: 'Multistatic radar classification of armed vs unarmed personnel using neural networks', *Evol. Syst.*, 2017, pp. 1–10

[31] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition', arXiv:1512.03385, 2015

[32] Khashbat, J., Tsevegjav, T., Myagmarjav, J., *et al.*: 'Determining the driver's reaction time in the stationary and real-life environments (comparative study)'. 2012 7th Int. Forum on Strategic Technology (IFOST), 2012

[33] Rohling, H.: 'Radar CFAR thresholding in clutter and multiple target situations', *IEEE Trans. Aerosp. Electron. Syst.*, 1983, **AES-19**, (4), pp. 608–621

[34] Munkres, J.: 'Algorithms for the assignment and transportation problems', *J. Soc. Ind. Appl. Math.*, 1957, **5**, (1), pp. 32–38

[35] Srivastava, N., Hinton, G., Krizhevsky, A., *et al.*: 'Dropout: a simple way to prevent neural networks from overfitting'. *J. Mach. Learn. Res.*, 2014, **15**, pp. 1929–1958

[36] Hochreiter, S., Schmidhuber, J.: 'Long short-term memory', *Neural Comput.*, 1997, **9**, pp. 1735–1780

[37] Ioffe, S., Szegedy, C.: 'Batch normalization: accelerating deep network training by reducing internal covariate shift', arXiv:1502.03167v3