



Guo, J., Shu, C., Zhou, Y., Wang, K., Fioranelli, F., Romain, O. and Le Kernec, J. (2021) Complex Field-based Fusion Network for Human Activities Classification With Radar. In: IET International Radar Conference 2020, Chongqing City, China, 4-6 Nov 2020, pp. 68-73. ISBN 9781839535406.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/223726/>

Deposited on: 1 October 2020

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Complex Field-based Fusion Network for Human Activities Classification With Radar

Jiaqi Guo¹, Chang Shu¹, Yiyi Zhou¹, Kun Wang¹, Francesco Fioranelli², Olivier Romain³, and Julien Le Kerneec^{1,3,4*}

¹James Watt School of Engineering, University of Glasgow, Glasgow, UK

²MS3-Microwave Sensing Signals and Systems, TU Delft, Delft, The Netherlands

³ETIS-Signal and Information Processing lab, University Cergy-Pontoise, Cergy, France

⁴School of Information and Communication, University of Electronic Science and Technology of China, Chengdu, China

*Email: *Julien.lekerneec@glasgow.ac.uk

Abstract

In the context of assisted living, human activity recognition (HAR) is of increased importance to maintain people at home living independently longer. Compared with directly using spectrograms (μ D) or range profiles (RP) as inputs for classification discarding either range or explicit Doppler information, range-Doppler-surface (RDS) can more fully represent the information contained in the observed activities. However, because the data are collected from different activities, people, and locations, the RDS has a requires non-trivial adjustments for pre-processing as discussed in this paper to maintain the number of points in the RDS to fixed integer. Although it has improved performance (92%) compared to CNN (90%), this algorithm also discards phase information as most algorithms do in HAR with radar. In contrast, the phase information of the range-Doppler domain, although its performance was not the best (88%), it had no obvious weakness in the recognition of all the movements. Our proposed complex field-based fusion network (CFFN) combines the amplitude and phase which improves both the accuracy of classification (94%) as well as accelerating training time by 12.5%.

keyword: Human activities classification, Convolution neural network, Doppler Radar, Adaptive thresholding, Deep fusion scheme, Phase information

1 Introduction

With the increasing number of elderly people, assisted living has attracted a lot of recent attention as of late in the research community. For example, many older adults are facing risks of falling as well as managing multiple chronic diseases, and their carers may not be on site to monitor them 24/7. The research in human activity recognition (HAR) can help to solve this problem with remote sensing. There are few technologies used in HAR like cameras, various wearable sensors, ambient sensors, and radar. Compared with the camera

and some other types of sensors, radar systems are versatile, contactless, and are perceived as non-intrusive [1-3].

In radar based HAR, spectrograms are widely used for the relative simplicity to interpret the data compared to other data domains and because continuous wave (CW) radar was cheaper historically. Handcrafted features can be extracted from these spectrograms and extensive literature exist covering this data domain [1,4]. Various classification techniques are applied in HAR both from supervised or unsupervised learning techniques [5]. The development of deep learning techniques like convolutional neural network (CNN) has been used with great success to recognize HAR. CNNs have been used successfully in radar applications on spectrograms to detect humans and classify human activities in [6,7] yielding high accuracy.

With the advent of automotive radar, the radar community suddenly had increasingly access to more data domains providing supplemental information (Raw data, range-time, spectrograms, cadence velocity diagrams, cepstrograms, and composite representations) for human activity classification. Using data from only one date domain is reductive. However, this challenges the radar community as there is now a plethora of data domains and features to choose from to classify data effectively. In terms of composite representations, B. Erol et al. [8] combined both range and Doppler features in HAR improving accuracy. Multiple joint-variable domains processed together showed to achieve an improvement HAR [9], and others applied the joint time-frequency analysis as well as the dynamic range-Doppler trajectory method to continuous HAR [10]. Range-doppler surface (RDS) was first introduced in [11] as a tool in HAR. In [12], the authors exploit RDS, sample the surface into a point cloud and use the PointNet [13] to classify the activities which allow leveraging information from both range and Doppler domains as they evolve in time.

This paper will explore the use of phase information, which is discarded in the vast majority of algorithms for HAR. This paper proposed PhaseNet that uses the phase of the RDS, and a novel the classification algorithm complex field-based fusion network (CFFN) that fuses the RDS amplitude and phase information from PhaseNet for improved performance, the discussion of pre-processing adjustments to optimise the

performance of the CFFN.

This paper is organised as follows. Section 2 will describe the methodology followed from this work introducing the database used for training and testing and the classification algorithms’ development. Section 3 will discuss the results obtained using our different approaches. Section 4 provides insight on the pitfalls and advantages of our approach over spectrogram only classification as well as lessons learned. Finally, Section 5 will provide the conclusions and further development directions.

2 Methodology

2.1 Introduction to our database

We use the University of Glasgow (UoG) Radar Signature of Human Activities dataset [14] to validate our proposed network. The data was collected from 23 female and 49 male participants aged 25-98 in 4 different locations, the laboratory and a common room of the University of Glasgow, rooms of Glasgow NG Homes, and Age Uk West Cumbria.

Each participant was asked to perform 5 daily activities repeatedly, namely walking back and forth (A01), sitting down (A02), standing up (A03), picking up an object (A04) and drinking water (A05) within a certain area in front of the radar. Only under laboratory-controlled conditions, was the sixth action (A06), simulated frontal fall, collected. The snapshots of all activities mentioned above are illustrated in Figure 1.



Figure 1: Sketch of six daily human activities

The off-the-shelf FMCW radar (by Ancortek) operating at 5.8 GHz with a bandwidth of 400 MHz and 1 ms chirp duration was utilized for data collection. Micro-Doppler signatures received by the radar are recorded by Yagi antennas with a gain of about +17 dB. The overall duration of a single piece of the data sample (walking excepted) is 5 s, whereas the recording of walking is a bit longer, lasting for 10 s.

2.2 Range-Doppler-Time point clouds construction

In [12], the authors proposed the range-Doppler-time (RDT) point clouds for human activity recognition. The data pre-processing work starts from the time domain complex radar data (in-quadrature and in-Phase). Then they are processed through the 2D Fourier transform with a sliding window to obtain a series of range-Doppler (RD) images. The sequence of RD images also captures the time domain information with the sliding window. Since these images describe the evolution of energy distribution in the range-time domain, the interval between two successive RD images is the time step of the sliding window.

However, as the background noise in RD images will influence the performance of classification model, a 2D constant

Constant False Alarm Rate (CFAR) detector is employed to detect the signal of human activities. The next step is iso-surface extraction. Depending on the energy distribution, an iso-surface mesh is generated by fine-tuning the iso-threshold value to find a surface that contains the same energy intensity. After that, the meshes are saved in the polygon file format, which can preserve the geometry feature (faces information) of the iso-surface. The vertex information in the polygon file is the discrete representation (down-sampling) of the iso-surface.

A key requirement for the classification algorithm is a uniform point cloud size. Since batch size is 32 in network training, the size of point clouds we choose are multiples of 32, i.e. 512,1024,2048. In reality, many point clouds, especially those constructed from data of elderly subjects, have a much smaller size than required. In this case, the geometric center point of each face is used to up-sample the point clouds and maintain the shape information of the point clouds at the same time. The up-sampled point clouds contain much more details than the original ones. However, an increased point cloud size not only brings about useful information, but may amplify the noise as well. Besides, a larger size of point clouds makes it harder for the classification model to converge. In order to select a satisfactory size, several experiments were conducted. The results are shown in Table 1. A point cloud size of 1024 did have higher accuracy in motion recognition than that of 512. However, when we used the up-sampled point clouds with a size of 2048, there was no significant improvement in accuracy. Therefore, we choose 1024 as the default size of our point clouds.

Table 1: The influence of point cloud size on accuracy

Size of point cloud	prediction accuracy (CFFN+hybrid data)
512	88 %
1024	94.27 %
2048	92.3 %

The architecture of our network is shown in Figure 2. It consists of two parts, a feature extraction network and a fusion network. The feature extraction network is composed of the PointNet [13], a deep learning framework on point clouds for 3D classification, and the PhaseNet, which captures the phase information of RD. The inputs of PointNet are RDT point clouds, which are firstly aligned by an input transform matrix (annotated as T-Net) and then put into multi-layer perceptrons (MLPs) to obtain RD magnitude features. Then these features are aligned by a feature transform matrix and processed through another series of MLPs to get local point features. In the final stage, a symmetric function, max pooling, is employed to aggregate the learned features into a global one. The PhaseNet takes RD phase grey images as its inputs. Through a series of 2D convolution and max-pooling layers, the network can output global phase features. In the fusion network, we adopt the deep fusion scheme instead early fusion or late fusion to enable more interactions among features [15]. For the intermediate layers, common operations includes concatenation and element-wise mean. As concatenation leads to high dimension intermediate layers, we choose the latter for the sake of computation complexity. During this stage, the m-

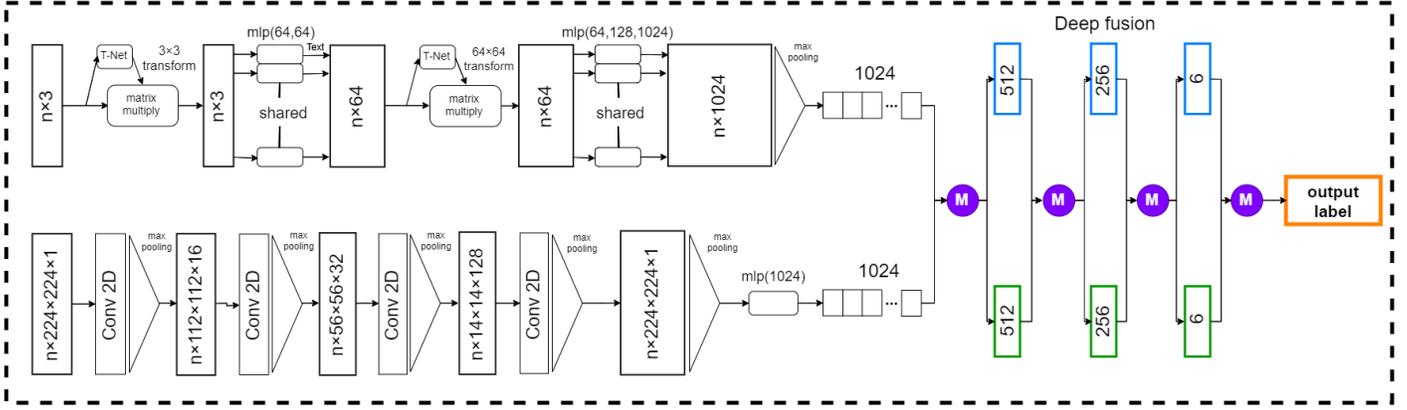


Figure 2: Basic structure of the proposed CFFN

-agnitude and phase features learned by PointNet and PhaseNet interact with each other deeply. This is realized by feeding back the average error in the fusion network to both the PointNet and PhaseNet.

2.3 Discussion on thresholding

Data preprocessing is an important step before the network training. According to the 3D outlier detection module [12], we need to determine several parameters to alleviate the effect of anomalies and extract the energy layer. Because the CFAR is used for noise detection and elimination, its iso-thresholding value constrains our scope to a certain energy level. As the magnitude of every pixel in the RD images can somehow represent the instantaneous energy at a point in time, the essence of iso-value selection is to pick up an energy layer that contains the most information. In other words, only the most representative point cloud should be inputted to our neural network. In this paper, we classify the six activities into two cases and propose a hybrid iso-value method to extract the RD iso-surface.

The intensity of the noise in a specific place is unpredictable but relatively stable, therefore, two cases were considered. Case one concerns vigorous activities (like walking and falling) as their SNR is relatively higher compared to the more subtle activities. This way, we need to extract the RD surface with a smaller iso-value. By contrast in the second case, for more subtle activities (e.g., stand up and down), their energy levels are relatively lower than for vigorous activities, the iso-value needs to be set higher (focus on a relative higher energy iso-surface) to neglect the unwanted noise. In this way, the salient energy layer can be extracted to improve the training accuracy of the proposed neural network. Given the energy of RD sequence E_{RD} , the relationship between the iso-value A_{iso} and target iso-thresholding value E_{th} is:

$$A_{iso} = 20lg \left(\frac{E_{th}}{\max(E_{RD})} \right) \quad (1)$$

3 Results

In this section, the experimental results of the proposed network are presented. The UoG research dataset contains over 1700 radar signatures of the six daily human activities,

among which 1472 signatures were selected to form two training sets and 128 samples into three validation sets. The participants are divided into two age groups. Those aged over 65 are considered the elderly whereas the others under are the younger group. Our training sets are composed of participants from both age groups. However, the validation sets are formed of young, old, and mix-aged participants separately to enhance the generalization of our trained model. We use the data of 100 radar signatures of all the activities provided for the Radar Challenge as the test set.

The raw radar data were processed through the 2D Fourier transform to obtain the complex-field RD. It was swept by a sliding window with a length of 200 ms and a padding factor 95% to generate the RD magnitude image sequences. Unlike their magnitude counterparts, the RD phase images were produced by setting the time window-length equal to the whole sample duration. From the RD magnitude images, we extracted the RDT point clouds according to the pre-processing data pipeline described above. The false alarm rate is set as 0.36. Figure 3 is an illustration of the data pre-processing of walking. After that, the outputs are trained by the network.

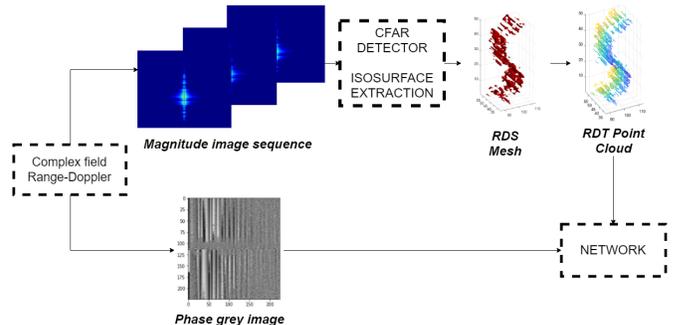


Figure 3: The construction of RDT point clouds and phase extraction

To validate the feasibility of using the phase of RD to do human activity classification, the PhaseNet was implemented as a standalone classifier and compared with the PointNet and CFFN. In the experiment, the CFFN uses the same input point clouds as the PointNet and the same input phase images as the PhaseNet. The evaluation results are illustrated in Table 2. In the hybrid dataset, A02, A03, and A04 are ex-

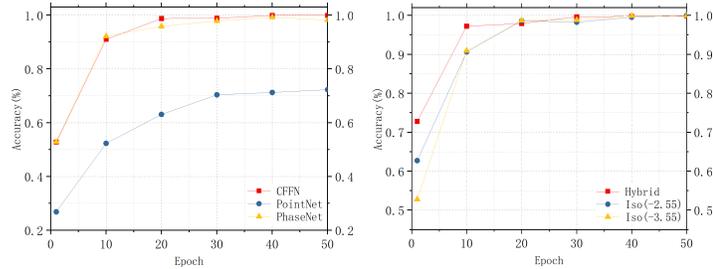
Table 2: Human activity classification accuracy on the UoG dataset

	avg acc	avg class acc	A01	A02	A03	A04	A05	A06
CNN	90.62 %	91.22 %	100 %	100 %	100 %	78.8 %	68.6 %	100 %
PhaseNet	86.45 %	86.95 %	97.2 %	94.3 %	83.3 %	72.7 %	80 %	94.1 %
PointNet+Hybrid	92.18 %	92.75 %	96.6 %	96.7 %	86.7 %	83.3 %	83.3 %	100 %
CFFN+Iso(-3.55 dB)	88.02 %	87.39 %	94.8 %	100 %	96.7 %	76.7 %	63.3 %	92.9 %
CFFN+Iso(-2.55 dB)	86.45 %	83.89 %	97.2 %	94.3 %	88.9 %	51.5 %	71.4 %	100 %
CFFN+Hybrid	94.27 %	94.73 %	100 %	97.1 %	88.9 %	90.9 %	91.4 %	100 %

-tracted with iso-value of -3.55 dB whereas A01, A05, and A06 are extracted with iso-value of -2.55 dB.

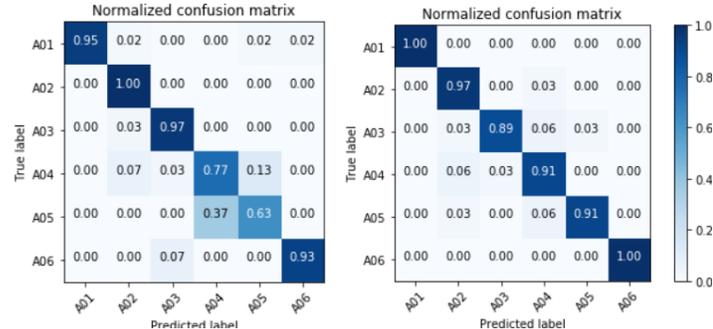
Additionally, we set up a control group that finishes the classification through a simple CNN. This CNN has almost the same network structure as our PhaseNet, but it takes a traditional Doppler-Time spectrogram as input. It has superior prediction accuracy over activities like walking, standing up, sitting down and falling down, whereas there is a serious confusion between picking up an object and drinking water.

The CFFN takes full advantage of the RD phase features and has the highest average accuracy. The training process of three networks is illustrated in Figure 4(a). It takes 200 epochs for the PointNet to converge. With the fusion of PhaseNet and PointNet, our CFFN converges within 50 epochs, which is only a quarter of the time to train the PointNet.



(a) Comparison among CFFN, (b) Comparison among different PointNet and PhaseNet iso-value settings

Figure 4: The training accuracy of different networks and different iso-value settings



(a) The confusion matrix of CFFN with mono iso-value (b) The confusion matrix of CFFN with hybrid iso-values

Figure 5: The confusion matrices of CFFN with mono iso-value and hybrid iso-values

Figure 5 shows the confusion matrices of our CFFN. According to Figure 5(a), the CFFN with mono iso-value has difficulty in distinguishing between A04: picking up an object and A05: drinking water. After applying the hybrid iso-

values, the accuracy for those activities increases by 14 and 28 %, respectively. Figure 4(b) reveals that during the training process, the model with hybrid iso-values outperforms that with mono iso-value.

As is shown in Table 2, the combination of CFFN+Hybrid iso-values has the best overall performance in training, the combination of CFFN+Mono iso-value also performs well in the recognition of some actions, for example, the combination of CFFN+Iso (-3.55 dB) achieves the highest accuracy in the recognition of A02 and A03. Hence, when applying the trained model to the test set of radar challenge, we introduced the idea of voting in order to achieve higher accuracy. Based on the evaluation results, we selected the trained models in the CFFN experimental group, they are CFFN pretrained with Hybrid data, Iso (-3.55 dB) and Iso (-2.55 dB). And the voting procedure is shown in Figure 6.

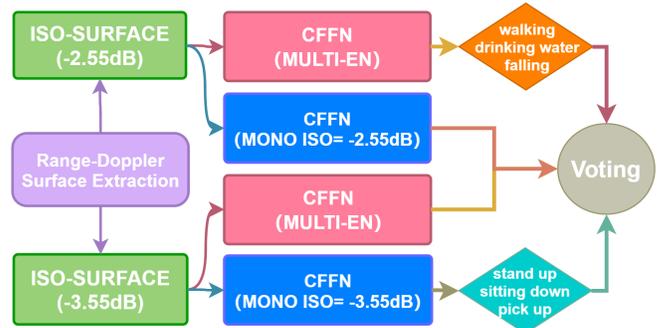


Figure 6: The voting procedure for predictions on test sets

The provided test data were processed with two iso-values and then evaluated by three pre-trained models. For the prediction results of CFFN+Hybrid, we only selected the results corresponding to iso-value to participate in the final voting. For example, since A02 (standing up), A03 (sitting down), A04 (picking up objects) were extracted with iso-value of -3.55 dB, so, according to Figure 6, only data judged as A02, A03, and A04 (activities in the cyan rhombus) would be selected to participate in the final voting. In other words, only these three activities' voting weight will be set to 1, whereas the rest of them will be set to 0.

4 Discussions

Pitfalls: To enhance the performance of our network, we introduced the multi-energy threshold. However, when we applied this method to the test set in the radar challenge, we found that we needed to implement two threshold parameter settings to extract the point cloud data for optimal performances. Although our prediction results were very satisfactory, we spent much time in data pre-processing. In real

conditions, we would need to develop an adaptive method to determine the threshold to account for variations in aspect angle, signal-to-noise ratio, and interference.

Advantages over spectrograms: All of our training is done on a GPU NVIDIA GTX 2080ti. The training time is counted and illustrated in Table 3. By incorporating PhaseNet into our CFFN, we improved the network’s predictive accuracy and improved the training time by 12.5 % compared to PointNet.

Table 3: Time counting of the total training time

Networks	Time Consumption (second)
CFFN	1400
PointNet	1600
PhaseNet	256

Lessons learned/take away message: Picking up objects and drinking water has a high degree of similarity. Without adopting the hybrid iso-value method, the accuracy will be low.

5 Conclusion

Summarising results: The predicted results on the ‘Radar Challenge’ test set has demonstrated that our CFFN can realize an accuracy of 100 % in radar-based human activities classification. Our model achieves a satisfactory accuracy of 94.7 % even on a mixed data set in which both older and younger samples exist. On the one hand, it shows that PhaseNet combined with PointNet can correct the abnormal prediction and significantly improve the convergence rate. On the other hand, it also suggests that for every action, its features may be more prominent in a particular energy layer. So we need to devise an adaptive method to predict the salient energy level for real-time processing.

Opening research directions: Several things that need to be mentioned are the iso-value we adopted is only a relatively optimal value, and the data pre-processing procedure should be simplified. So, future work will involve designing an adaptive algorithm to find the optimal iso-value and find a possible way to integrate the energy layers with different energy (multiple energy layers to single data) together. Apart from this, since our data samples are all discrete activities, future research can extend the scope which enables our network to predict a series of continuous human activities.

6 Acknowledgments

The authors thank Glasgow College - UESTC and the British Council 515095884 and Campus France 44764WK-PHC Alliance France-UK for their financial support.

References

- [1] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, J. Lorandel, and O. Romain, “Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 29–41, 2019.
- [2] X. Li, Y. He, and X. Jing, “A survey of deep learning-based human activity recognition in radar,” *Remote Sensing*, vol. 11, no. 9, p. 1068, 2019.
- [3] A. Shrestha, J. Le Kernec, F. Fioranelli, Y. Lin, Q. He, J. Lorandel, and O. Romain, “Elderly care: activities of daily living classification with an S band radar,” *The Journal of Engineering*, vol. 2019, no. 21, pp. 7601–7606, 2019.
- [4] S. Z. Gurbuz and M. G. Amin, “Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.
- [5] J. Le Kernec, F. Fioranelli, S. Yang, J. Lorandel, and O. Romain, “Radar for assisted living in the context of internet of things for health and beyond,” in *2018 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC)*, pp. 163–167, IEEE, 2018.
- [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, 2015.
- [7] R. Zhu, Z. Xiao, Y. Li, M. Yang, Y. Tan, L. Zhou, S. Lin, and H. Wen, “Efficient human activity recognition solving the confusing activities via deep ensemble learning,” *IEEE Access*, vol. 7, pp. 75490–75499, 2019.
- [8] B. Erol and M. G. Amin, “Fall motion detection using combined range and doppler features,” in *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 2075–2080, IEEE, 2016.
- [9] B. Jokanovic, M. Amin, and B. Erol, “Multiple joint-variable domains recognition of human motion,” in *2017 IEEE Radar Conference (RadarConf)*, pp. 0948–0952, 2017.
- [10] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li, “Continuous human motion recognition with a dynamic range-doppler trajectory method based on fmcw radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6821–6831, 2019.
- [11] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. Le Chevalier, and A. Yarovoy, “Range-doppler surface: a tool to analyse human target in ultra-wideband radar,” *IET Radar, Sonar Navigation*, vol. 9, no. 9, pp. 1240–1250, 2015.
- [12] H. Du, T. Jin, Y. Song, Y. Dai, and M. Li, “A three-dimensional deep learning framework for human behavior analysis using range-doppler time points,” *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [13] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

- [14] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kerneç, “Radar signatures of human activities,” 2019.
- [15] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3d object detection network for autonomous driving,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1907–1915, 2017.