



Shrestha, A., Murphy, C., Johnson, I., Anbulselvam, A., Fioranelli, F., Le Kerneec, J. and Gurbuz, S. Z. (2019) Cross-Frequency Classification of Indoor Activities with DNN Transfer Learning. In: IEEE Radar Conference, Boston, MA, USA, 22-26 Apr 2019, ISBN 9781728116792 (doi:[10.1109/RADAR.2019.8835844](https://doi.org/10.1109/RADAR.2019.8835844))

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/178053/>

Deposited on: 17 January 2019

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

Cross-Frequency Classification of Indoor Activities with DNN Transfer Learning

Aman Shrestha*, Corben Murphy[†], Ivan Johnson[†], Ajaymehul Anbulselvam[†],
Francesco Fioranelli*, Julien Le Kernec*[‡] and Sevgi Zubeyde Gurbuz[†]

*School of Engineering, University of Glasgow, Glasgow, United Kingdom

[†]Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, Alabama, USA

[‡]School of Information and Communication, University of Electronic Science and Technology of China, Chengdu, China

Abstract—Remote, non-contact recognition of human motion and activities is central to health monitoring in assisted living facilities, but current systems face the problems of training compatibility, minimal training data sets and a lack of interoperability between radar sensors at different frequencies. This paper represents a first work to consider the efficacy of deep neural networks (DNNs) and transfer learning to bridge the gap in phenomenology that results when multiple types of radars simultaneously observe human activity. Six different human activities are recorded indoors simultaneously with 5.8 GHz and 25 GHz radars. Firstly, the bottleneck feature performance of the DNNs show that a baseline of 76% is achieved. On models trained only with 25 GHz data when 5.8 GHz data is used for testing 81% accuracy is achieved. In absence of a large dataset for radar at a certain frequency, we demonstrate information from a different frequency radar is better suited for generating the classification models than optical images and by using time-velocity diagrams (TVD), a degree of interoperability can be achieved.

Keywords: *Micro-Doppler, Assisted living, Transfer learning*

I. INTRODUCTION

Indoor monitoring of motions and activities through a non-contact paradigm is becoming necessary to face the need for ambient care of the increasingly elderly population of the world. As we are getting older and living longer, the need for quality care and quick response times in cases of accidents has increased importance in this environment where multi-morbidity conditions are present. The key aim is to maintain user autonomy while keeping their independence. Typical technologies to address this gap have been sensors, including but not limited to wearable inertial measuring units [1], RGB and stereoscopic camera systems, passive acoustic sensors and radar [2].

Micro-Doppler(mD) signatures, which are rotational, translational or vibrational components of a body moving in the line of sight of an electromagnetic detector, can be utilised to identify dynamic objects or movements [3]. As the sensor is capable of distinguishing changes in the environment in addition to the composing entities, there is an expansive array of applications which mD use is suggested for, including but not limited to: detecting armed people [4]; differentiating between pedestrian, cyclists and automotive cars [5]; human activity classification [6] and general in-home monitoring of elderly persons with focus on fall detection [7]. Radar is

suitable for indoor monitoring due to its contactless nature since there is no need for the users to wear or carry devices or change their behaviour. It is more suitable than other technologies which may be either forgotten by the user or unused due to the negative connotations associated with the devices. It is insensitive to luminescence levels which means it can record throughout the day and because it does not capture digital images, it is more privacy oriented than other technologies. The radiated power level from a radar sensor is comparable or less than that of other radio sources such as WiFi or smartphones meaning it is safe for prolonged exposure [2].

Deep learning [8] has brought a new revolution in cognitive computing and attempts are being made to utilise these technologies with radar [5] [9] [10]. Previously the focus has been on generating handcrafted features which exploit spectral and temporal properties which are then classified with the use of support vector machines [7] [6] [11]. A further improvement on this method relies on extracting salient features to improve accuracy [11]. Multistatic and multimodal sensor fusion have also been shown to improve the recognition of activities. To better generalise the extraction of the features and increase the automation of the classification process, neural networks are being used increasingly. In this paper we continue this trend where we use a selection of convolutional neural network (CNN) architectures, specifically: AlexNet [12], GoogleNet [13] and a simple 3 unit CNN with emphasis on batch normalisation.

Cross-frequency compatibility is a key aspect sensing that has seen limited discussion in the literature, as most works focus on single frequency systems. In preliminary work, Vishwakarma [14] proposed a dictionary learning approach for classification across multiple frequency, motivated by cognitive radar applications. However, only four easily discernable classes were considered for transmit frequencies varying by just 2.5 GHz; therefore, the methods ability to deal with the interoperability of transmitters with large difference in frequency, in realistic environments, to classify numerous signatures for similar activities is an outstanding research challenge. In this work, we consider the ability of deep neural networks and transfer learning to aide in the classification of 6 human activities using radar systems whose transmit

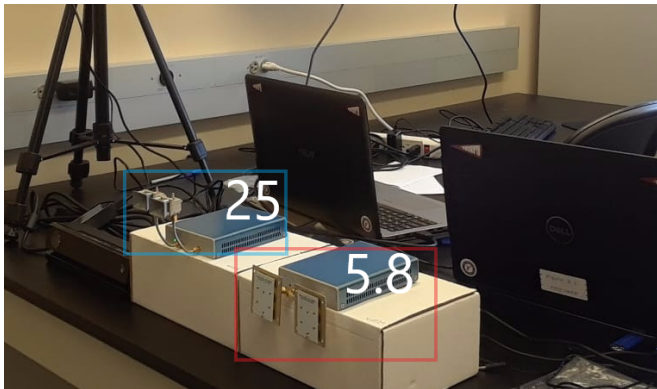


Fig. 1. Radar and antenna setup. The 25 GHz and 5.8 GHz radars are connected to two different laptops, which are connected via a local network.

frequencies differ by 20 GHz.

The paper is organised as follows: in Section II, the hardware setup and the experimental design is discussed, this is followed by the data processing. Then, in Section III results from the use of bottlenecked features from the classifiers transfer learning on our subset of activities are presented in Section IV and finally, the use of transfer learning and TVD from the secondary radar is shown to establish compatibility.

II. EXPERIMENTAL DESIGN

A. Hardware Setup

Simultaneous measurements from frequency modulated continuous wave (FMCW) radars at 5.8 GHz and 25 GHz were taken in a 4 m by 4 m square space in the Radar Signal Processing Laboratory (RSPL) of the University of Alabama, Tuscaloosa. The bandwidth of the two systems shown in Fig.1 were set at 400 mHz and 750 mHz, while the pulse repetition frequency of both radars was set at 1 kHz. The systems were aligned at waist height (1 m) to ensure the torso returns were focused upon as the rest of the body has narrow radar cross section. Physically, the radars were separated by 0.2m for the beam to effectively have the same area however, as they were transmitting in two different frequency bands, no cross interference was expected to occur.

The 5.8 GHz C-band system used patched antennas and the 25 GHz K-band system utilised horn antennas. The transmitted power settings were 16 dBm and 19 dBm respectively.

B. Data collection

The dataset in this experiment was collected from seven participants with variations in build, age, height and weight. The range of motions performed were typical of an assisted living scenario where gait movements and the bending of the torso is common. These activities are useful for monitoring purposes as they require macro movements of large segments of the body where degradation of mobility becomes more easily identifiable. Six activities were performed by the participants of which five seconds were recorded from, with at least ten repetitions of each activity. The aspect angle was

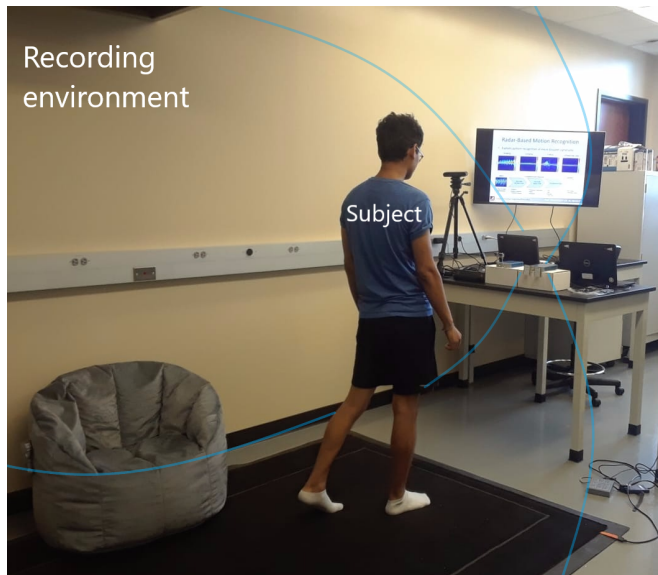


Fig. 2. The recording environment with a participant. Aspect angle was varied therefore the participant was not directly in line of sight.

varied from 0 to 15 onto 30 degrees as to simulate an indoor environment, as shown in Fig 2. and also increase the variation and the classification challenge.

The six activities were:

- 1 : Walking forwards
- 2 : Walking away from the sensor
- 3 : Bowing forwards
- 4 : Picking up an object
- 5 : Foot drag
- 6 : Walking with short steps

1009 independent observations were taken with each radar, in total 2018 files were generated through this experiment.

III. DATA PROCESSING

Using the radar returns from the system, a Short Time Fourier Transform (STFT) was performed on each recording to generate the spectrogram. STFT visualises the intrinsic instantaneous frequency components of the signal spectrum which allows the micro-Doppler signal to be visible. With it, the movement of limbs and rotation of torso can be seen.

The generated spectrograms had a Hamming window with an duration of 0.2 seconds for both systems. The overlap between the windows was set at 95%.

As a large proportion of data has been taken at an off aspect angle the returns would be attenuated by the cosine relationship. Contrast enhancing methods are used therefore to improve classification accuracy and generalise performance [12]. Methods like Naka-Rushton contrast enhancement and principle component based methods have been suggested, but in this paper we use a simplified z-score based threshold where the mean amplitude of the spectrogram is subtracted from each bin then normalised by the standard deviation of the bins.

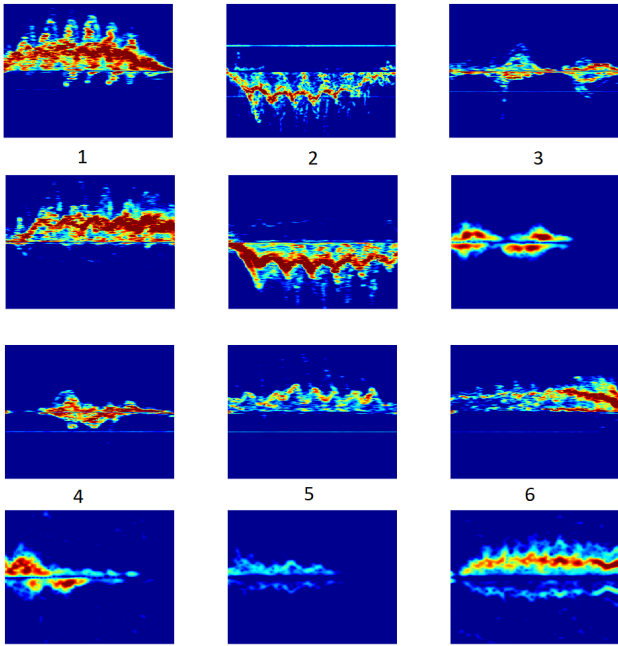


Fig. 3. Image inputs to the neural networks. The higher resolution images above the Acitivity number are from 25 GHz radar in rows 1 and 3. The 5.8 GHz inputs are below the number in rows 2 and 4.

After generation of the spectrograms, the time velocity diagrams are taken using the Doppler relationship:

$$v = \frac{cf_d}{2f_0} \quad (1)$$

where f_d is the Doppler frequency and f_0 is the carrier frequency of the radar. The time velocity diagrams generated with limits at 5 m/s are passed through the contrast threshold then converted to RGB images. These images, as shown in Fig 3., are then used as inputs to the neural networks.

IV. TRANSFER LEARNING

Transfer learning (TL) utilises the weights generated by a deep neural network trained on a large dataset [15], which allows the detection of edges, curves and other image patterns/properties. This then is ported to a application specific problem set with a smaller set of labelled data and identifiable classes. This approach mitigates two general issues with Deep Convolutional Neural Networks: the size of the training dataset and the training time required to train the classifier. Fig 4. shows the layer decomposition of one of the architectures used in the design and shows the layers removed in the transfer learning process.

From the collected dataset of the K-band radar, 50% of the data was used for training the classifiers, 25% for validation and 25% for testing. The segmentation of the dataset was stratified therefore class ratios were preserved despite the random selection of the data points. The pretrained models used for transfer learning are AlexNet and GoogLeNet, both trained with the ImageNet database prior to their modification.

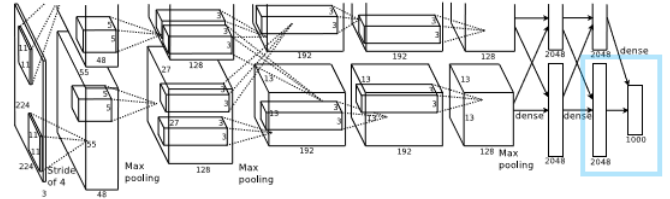


Fig. 4. Alexnet network architecture [12]. For transfer learning, after training on the ImageNet database, the end fully connected, softmax and output layer are replaced and the network is retrained with the target dataset to make it specific to our application.

A convolutional neural network has been included for comparison between the intensive deep learning methods and simpler classifiers. These three networks differentiate in three general properties: depth, presence of a batch normalisation layer and network architecture.

All neural networks are tuned with an initial learning rate of 0.0001 and Stochastic Gradient Descent method. The software package used for this work was MATLAB with graphics processing unit (GPU) acceleration utilised for a faster training process. Three Nvidia GTX 1080ti are utilised here and the learning rate has been scaled according to the number of GPUs.

AlexNet was famously used in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. It consists of three convolutional units (5 convolutional layers and 3 max pooling layers). For better generalisation and to prevent overfitting, it contains dropout layers towards the top of the structure [12].

Two years later, GoogLeNet won ILSVRC [16] due to its new inception module and hierarchical decision design. The module performs dimensionality reduction by performing 1x1, 3x3 and 5x5 convolutions implicitly and filter concatenating. This means it has increased depth compared to AlexNet.

The third neural network is a simple design which relies on a number of batch normalisation layers as this improves generalisation by reducing the covariant shifts between the layers [17]. This, in practice, would benefit training with two radar systems as the compatibility of the model is reliant on the fitting of the spectrograms. The entire design consists of three convolutional units, each with a convolutional layer; a batch normalisation layer; a rectified linear unit layer and a max pooling layer, in this specific sequence. Three units are cascaded with the relu layer in the third convolutional unit connected to a fully connected and a softmax layer for classification.

A. DNN Initialization with Optical Imagery

To establish a baseline for the results, within the challenging classification environment where the age, body type and aspect angles varied, prior to removal of the fully connected layers, bottleneck features from the pre-trained networks were generated and classified with a multi-class support vector machine. We compare bottleneck features as these reflect the efficacy of

different source domains in initializing the network weights. Bottlenecked features are generated by using the weights of the DNNs' end connected layers then using data unknown to the classifier to generate features which give indications of how suited the current weights are for mD classification. In this section, we present the bottleneck feature performance when the ImageNet database is used as the source domain.

Table I shows the bottlenecked feature performance of the deeper network is significantly improved. Features from AlexNet find it difficult to correctly identify activities 3,4 and 6 while maintaining a 80-91% classification rate for the remaining activities.

TABLE I
CLASSIFICATION RESULTS OF TRANSFER LEARNING ON MODELS TRAINED ON IMAGENET WITHOUT ANY RADAR DATA

Classifier\Class	1	2	3	4	5	6	Mean
AlexNet	80.6	96.3	55	36.6	90	29.2	64.7
GoogLeNet	87.8	92.6	94.3	44.7	86.1	55.1	76.8

GoogLeNet derived features performs better with the exception of 2 and 5 where the accuracy is 5% lower for these two activities compared to AlexNet. In Fig 5. the main incidents of confusion are between the similar classes as expected. As activity 6 is a variant of activity 1 it is normal that a degree of confusion occurs between these classes. The central component of activities 3 and 4 also have the same motion where the forward bend of the torso is detected by the radar, since radial movements are detected more easily than traversal perpendicular to the radar.

1	88			3	4	5
2	1	93	2		4	1
3	4		94	1	1	1
4	6	2	42	45	5	
5	4		2	3	86	5
6	41		1	1	1	55
	1	2	3	4	5	6

Fig. 5. Confusion matrix for the SVM with bottlenecked features from GoogLeNet

Bottleneck accuracy of approximately 64-76% means that the models have room for improvement with fine tuning. At this point we removed the classification layers from these classifiers for fine tuning and tested with the C-band radar data, to assess if the models can be generalised to radars operating at significantly different frequencies and if training with a set of spectrograms results in improvements compared to having a training set of optical images.

V. CROSS FREQUENCY TRAINING RESULTS

This model was then tested with independent data from the C-band only. There is no fine tuning of the model with C-band data in this stage to simulate a scenario where a large number of training data for the radar is not available allowing an assessment of compatibility to be made. Key image attributes such as sharpness and pixel intensity can change markedly when carrier frequencies which are 20 GHz wide apart are considered. The degree to which the classification accuracy attenuates as a result of this and use of machine learning to approach this problem has not been fully assessed in literature.

TABLE II
CLASSIFICATION RESULTS OF CROSS-FREQUENCY TESTING WITH 5.8 GHZ DATA TUNED ON 25 GHZ DATA

Classifier\Class	1	2	3	4	5	6	Mean
AlexNet	99.4	99.4	87.9	16.7	100	59.86	77.3
GoogLeNet	100	100	80.3	31.8	100	57.2	78.2
3- CNN	92.8	99.4	66.5	52.8	96.2	82.9	81.7

Table II shows the parity in performance of the pre-trained classifier is present. Both networks identify the classes at 77-78% accuracy which is an improvement from the case in Section IV by 12% for AlexNet but a similar measure for GoogLeNet. Classification rates of some activities are reduced and new minimum rates of 16% for AlexNet and 31% for GoogLeNet for activity 4 and performance of the 3 unit CNN is at 53% which is an improvement on the other architectures which is 20% more than the nearest pre-trained classifier .

1	93		1	1	2	3
2		99		1		
3			66	30	1	3
4		10	22	53	13	2
5			1	1	96	2
6	14			2	1	83
	1	2	3	4	5	6

Fig. 6. Confusion matrix for the best performing learning method for cross frequency data: 3-CNN

The best performing architecture here is the 3 unit CNN which maintains an accuracy of approximately 81%, an improvement from the 76% shown in Table I. The confusion for class 4 is due to the smaller Doppler shift which occurs when 'bending' relative to the mass movement of the whole body.

Fig. 6 shows the distribution of the misclassification is different when radar data is introduced into the models. There is less widespread misclassification as a number of activities,

including 1: walking, is recognised as the other activities at low incident rates. Significant confusion, however occurs between the classes identified earlier which are expected to be confused (i.e. 3,4 and 1,5,6) but activity 4 has alarming rates of misclassification as it is identified as a gait at times. Activity 4 and 3 are very similar in terms of bowing and bowing with picking up an object. In other words, it is a challenging task to classify.

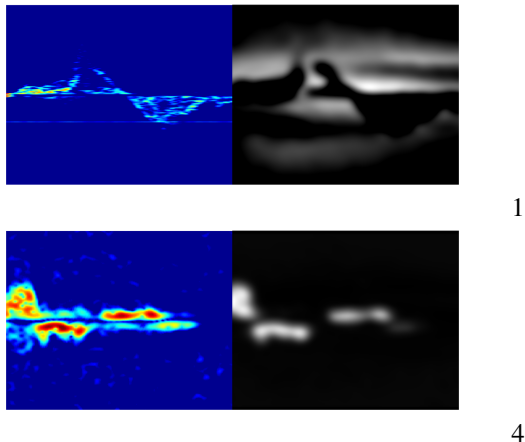


Fig. 7. Activation map for the third convolutional layer for activity 4 for X-band radar and C-band radar.

Towards the top layers of the CNN, the information from the micro-motions appear to be ignored and only the strong Doppler signals remain with spurious, weaker components being filtered out. The weights are reliant on edges as shown in Fig. 7 and it explains the challenging results for activity 4, as the inferential information required to accurately classify the activity is different for the two systems.

VI. CONCLUSION

In this paper, we compare bottleneck performance for DNNs which are trained with image data as well as performance comparisons for a two-stage training process, where a small amount of measured data is used to fine-tune the weights initialized with data from the other domain (e.g. ImageNet or data at another frequency).

Despite variation due to physical decomposition, gender and age of the participants and variable aspect angles, bottleneck feature performance of 76% is reached when no mD data is made available to the classifiers. When the models generated are trained with 25 GHz data then tested with 5.8 GHz data, accuracy of about 81% is achieved. This indicates that in absence of a large dataset for radar at a certain frequency, data from another frequency is better suited for generating the classification models than general images.

These initial results show that the radar sensors can be compatible with the caveat that signatures with lower intensities can be misclassified to a great degree. To continue on this topic, use of high resolution time frequency techniques such hyperbolically wrapped cepstral coefficients and more intricate

model design as in convolutional auto-encoders [18] will be investigated. A refined fine tuning method where a small amount of measured data is used to tune the weights initialized with data from the other sensor will be evaluated. The different algorithmic pre-processing steps for image enhancement will be explored.

ACKNOWLEDGMENT

A. Shrestha is supported for his PhD by the UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Award to the School of Engineering. The authors acknowledge support from the UK EPSRC through grant EP/R041679/1 INSHEP. The authors would like to thank the participants of the data collection campaign for their help.

REFERENCES

- [1] S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," *IEEE Sensors Journal*, vol. 15, no.3, pp. 1321-1330, 2015.
- [2] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "radar and RGBDepth Sensors for Fall Detection: A Review", *IEEE Sensors Journal*, vol. 17, no. 12., pp. 3585-3604, Jun. 2017.
- [3] V.C. Chen, *The micro-Doppler effect in radar*, Artech House, 2011. ISBN: 978-1-60807-057-2
- [4] J.S.Patel, F. Fioranelli, M. Ritchie, et al.: "Multistatic radar classification of armed vs unarmed personnel using neural networks", *Evol. Syst.*, 2017, pp. 1-10
- [5] A. Angelov, A. Robertson, R.Murray-Smith, and F.Fioranelli,"Practical classification of different moving targets using automotive radar and deep neural networks," *IET radar, Sonar & Navigation*, vol. 12, no. 10, pp. 1082-1089, 10 2018.
- [6] Y. Kim and H. Ling, "Human Activity Classification Based on MicroDoppler Signatures Using a Support Vector Machine" , *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328-1337, May 2009.
- [7] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "radar Signal Processing for Elderly Fall Detection: The future for in-home monitoring" , *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 7180, Mar. 2016.
- [8] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521, pp. 436-444, May 2015.
- [9] M. S. Seyfioglu and S. Z. Gurbuz, "Deep Neural Network Initialization Methods for Micro-Doppler Classification With Low Training Sample Support," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2462-2466, Dec. 2017.
- [10] Y. Kim and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks", *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8-12, Jan. 2016.
- [11] 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.
- [12] Krizhevsky,A., Sutskever,I.,Hinton, G.E.. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." *Neural Information Processing Systems*. 25. 10.1145/3065386.
- [13] C. Szegedy et al., "Going deeper with convolutions," 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9.
- [14] S. Vishwakarma, S. Sundar Ram. "Dictionary Learning with Low Computational Complexity for Classification of Human Micro-Dopplers Across Multiple Carrier Frequencies." *IEEE Access*. PP. 1-1. 10.1109, 2018.
- [15] Y. Bengio, G. Guyon, V. Dror, G. Lemaire, D. Taylor, D. Silver. (2011). "Deep learning of representations for unsupervised and transfer learning." *emphProceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 277*, Washington, pp. 17-37, 2011.

- [16] O. Russakovsky*, J. Deng*, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei. (* = equal contribution) "ImageNet Large Scale Visual Recognition Challenge." *IJCV*, 2015
- [17] S. Ioffe, C. Szegedy "Batch normalization: accelerating deep network training by reducing internal covariate shift", arXiv:1502.03167v3
- [18] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz and M. G. Amin, "Diversified radar micro-Doppler simulations as training data for deep residual neural networks," *2018 IEEE radar Conference (radarConf18)*, Oklahoma City, OK, pp.0612-0617, 2018.