WALDO Finds You using Machine Learning: Wireless Adaptive Location and Detection of Objects

Aditya Singh*†, Pratyush Kumar†, Vedansh Priyadarshi†, Yash More†, Aishwarya Das†, Debayan Gupta†,
* Carnegie Mellon University, Pittsburg, Pennsylvania
†Ashoka University, Sonipat, Haryana
Telephone: +91 9909869367

***Abstract*—We present a novel radar-based system for real-time indoor positioning and detection of objects and human-bodies with low-quality, inexpensive sensors. Using modern deep learning methods, we avoid the use of expensive hardware and computationally-expensive signal processing methods for object detection.**

We train our model on mini–Doppler maps, collected via software defined radios. Crucially, our system is different from existing RF-based detection systems as it operates in a less crowded frequency range of 433 MHz, allowing us to use inexpensive off-the shelf hardware. Our system, based on the VGG-16 model, reports high-accuracy results on: (1) classification of different objects/materials (plastic, glass, metal); (2) detection and classification of multiple visually and materially similar objects and the human-body; and (3) Simple object detection at different distances between the transmitter T_x and the receiver R_x.

WALDO, using low frequency radio waves, is able to handle occlusions and bad lighting environments. Our results demonstrate that Deep Learning methods can be combined with inexpensive, low-frequency radars to achieve high accuracy in real-time on various useful tasks.

I. Introduction

Radio-based detection and location systems have seen huge advances in the past decade. Beyond the usual industrial and military uses, radar has been used in areas as diverse as unsupervised monitoring of adults [1], medical diagnosis, autonomous vehicles [2], [3], load-balancing energy in smart homes [4], improved virtual reality experiences, and even for ensuring social distancing during a pandemic[1].

Radar systems overcome limitations posed by vision and laser based approaches which are often limited by difficult light and weather conditions, presence of occlusions, and high deployability costs. RF-based systems are often better for privacy since they do not involve visual surveillance of the surroundings; they do not involve wearables, so they don't require user compliance and can operate in the background. Micro-doppler signatures, which are rotational and translational components of a body useful for identifying unique dynamic objects and movements [5] can be analysed using RF-waves, enabling indoor radars using portable, readily available, low-cost software defined radios.

[1]https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

Fig. 1: System setup for detection of standing human, distance between Transceiver (T_x) and Receiver (R_x) is 4m

Deep learning has proved to be a powerful approach for solving pattern-recognition problems [6]. Using the immense computational power of modern graphics processing units, it can achieve human level (or beyond) performance in several image and audio classification tasks. Deep neural networks (NNs) with multi-layer non-linear structures allow us to extract useful representations of raw input in the deeper layers of the NN, which capture features important for success on down-stream tasks. Critically, these features often exhibit invariance to viewpoints, lighting conditions, noise, and other transformations, providing powerful generalizations over unseen data.

Although the typical approach of extracting radar reflections and identifying objects from the shape of point-cloud of reflections has been successful, the task becomes increasingly difficult as multiple objects of various sizes are introduced [3]. Further, as the number of, and similarity between, classes of objects increases, the performance of classifiers using pre-defined features worsens significantly [7]. So, Radar Spectra generated by multi-dimensional Fast Fourier Transform (FFT) has emerged as an effective representation – it is not only capable of preserving all the information in the raw signal, but also serves as a data representation upon which NNs can operate (in a similar fashion as in vision). The automation of the feature extraction process in NNs, combined with the spectrogram representations of the signal, results in a powerful combination: it ensures that (important but complex) parts of the signal which might be ignored by manual feature selection can be correctly cast as significant by NNs.

978-1-7281-7609-3/21/$31.00 ©2021 IEEE

Deep Learning methods which work on the radar spectrum after multi-dimensional FFT have been successfully applied in tasks such as human fall detection [8], human pose estimation [9], [10], and human-robot classification [11]. It has also yielded successful results for classification of objects and other traffic participants for scene understanding, which is critical in automated driving [12]. Notably, this approach has been able to extract micro-doppler signatures from the full radar spectrum. However, most of the work done in this area relies on the use of multiple expensive and specialised hardware components operating in the over-crowded high frequency WiFi channel ($f \geq 2.0$ GHz). In general, the use of low-frequency radio-waves for enabling object (or human) detection is not an area which has seen much exploration.

In this paper, for the first time, we demonstrate radar-sensing with deep learning using inexpensive, off-the-shelf software radios (utilizing low-frequency radio waves) in an indoor environment to:

1) Classify different objects/materials (polyethylene, glass, metal) with high accuracy
2) Detect human body, multiple visually and materially similar objects with high accuracy.
3) Successfully perform simple object detection up to a range of 8m.

Our experiments demonstrate that deep neural networks can be applied on full radar spectra to carry out (the otherwise intractable task of) object detection and localisation even with low-frequency radio waves. Our approach was inspired by the use of low-frequency waves (around 20 kHz) by certain owls. These animals are able to localise their prey during night-time and through occlusions due to their heightened sensitivity to low-frequency waves [13]. Due to its portability and cheap deployment costs, our system has the potential to be easily scaled to larger areas and developed to perform well in different environments.

II. Background

We review recent literature in object-detection, material-classification and grid-based localisation in radar-based systems. We specifically look at how deep-learning and classical machine learning methods have been utilised.

Deep neural networks have played a versatile role across different object detection tasks and radar-based methods. Yeo et al. developed a radar based system for material and object detection [14]. They experimented with 26 materials which included complex composite objects, transparent materials, and human body-parts. Their model uses a random forest classifier coupled with 10 fold cross-validation technique and achieves high accuracy on both material and human body-part detection. However, such extensive classification was only possible by using a Google Soli chip operating at 60Ghz.

On similiar grounds of human-detection, Jin et al. developed a mmWave radar capable of real-time tracking and detecting patients, with the help of DNNs [15]. The mmWave radar is used to collect doppler patterns of the subject in motion, which eventually serve as inputs to the CNN-based classifier. Their model is able to generalize well on test data, and serves the potential to be scaled to different environments, while taking into account more varied motions. Amin et al. have also utilized DNNs in the domain of human-motion detection. DNNs are used to determine the relative significance of different regions of spectrograms, generated from data collected via a radar, which can then be applied to detect different regions responsible for human motion [16]. Radar coupled with deep learning has also been used in detecting different objects in a scene for self driving cars. Patel et al. [2] apply deep convolutional networks directly to regions of interest (ROI) in the radar spectrum and achieve accurate classification of different objects in the scene.

Hicham et.al [17] present a smart monitoring system based on 433 MHz using an arduinonano, arduinouno, a GSM module and a buzzer module called SMIS (Smart monitoring information system). The Arduino uno acts as the data gathering unit and the GSM acts as the transmission module. The RF emitter and transmitter both operate in the 433 MHz band. SMIS operates in real time by sending a message every 500ms to the RF receiver. If the object is not in the remotely monitored area, the receiver doesn't get any message and SMIS sends an SMS to the object's owner using the GSM module. During this process, the buzzer is also activated to let the user know that the object is out of the monitored area. The use of the 433 MHz is preferred for SMIS as it can transmit over very long distances without requiring high power input. This wireless frequency band is also an open source alternative and available worldwide, making it an apt choice of use for our proposed experiments.

Using techniques inspired by recent works, we propose an alternative system which aims to achieve considerably high accuracy in object-detection, localisation, material classification, and human detection with the help of a low-cost radar setup operating on low-frequency radio waves.

III. Convolutional Neural Network

Convolutional Neural Network (CNN) has been the main workhorse of recent breakthroughs in understanding images [18], videos [19] and audios [20]. Below we describe the basic building blocks of a CNN that are relevant to this paper.

Deep Neural Network: A deep neural network contains multiple layers of neurons that extract information from the input signal. Each neuron receives input from the neurons in the previous layer and combines them through weights and a nonlinearity (e.g., softmax). Mathematically, the value of a neuron a_i^n at the n-th layer is $\sigma\left(\sum_j w_{ij}^n a_j^{n-1}\right)$, where a_j^{n-1} are the neurons from the previous layer, w_{ij}^n are the weights and $\sigma(\cdot)$ is a nonlinearity.

CNN: In CNN, neurons in a layer are locally connected to a few neurons in the previous layer, contrary to the ordinary neural networks where each neuron in a layer is connected to all the neurons in the previous layer. This local connectivity allows CNNs to even learn sparse features present in data. All the neurons in the same layer of a CNN share the weights, and the value of neurons in a CNN can be computed as the convolution of a weight kernel with the neurons in the previous layer, that is $a^l = \sigma\left(f^l * a^{l-1}\right)$, where $*$ is the convolution operator, f^l refers to the weight kernel at layer l, and a^l is the neurons in layer l.

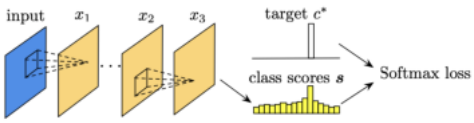

Fig. 2: CNN architecture for classification task, where x_1, x_2 and x_3 are feature maps, s shows the class scores, and c^* refers to the label. The Softmax loss function compares the score vector with the label.

Besides the typical convolutional layer, the following building blocks are relevant to this paper:

1) *Batch Normalization (BN) Layer* [21]: BN normalises activation value and can be represented mathematically as,
$$BN_{\gamma^k \beta^k}(\boldsymbol{x}^k) = \gamma^k \boldsymbol{x}^k + \beta^k,$$
where $\boldsymbol{x}^k$ is some input tensor, and γ^k and β^k are learnt during optimization process.
2) *Dropout Layer*: Dropout is a technique used to ignore a percentage of randomly selected neurons during training to improve generalisation performance of the network. It is an approach to regularisation which helps reduce interdependent learning amongst neurons by adding noise to the training process [22].
3) *Softmax Layer*: In classification tasks, the final layer of the network outputs a score for each class (i.e., a vector of scores). The Softmax loss is used to measure the difference between the class scores $s = \{s_C\}_{C=1}^k$ and the target label C^* as follow:
$$L_{\text{Softmax}}(s, C^*) = -\log \frac{e^{s_{C^*}}}{\sum_C e^{s_C}}$$
where s_{C^*} is the score prediction of the target class. Finally, the CNN computes class scores for new input and predicts it as the class with the highest score, as shown in Figure 2.

IV. Experimental Design

In previous experiments, a CNN was used with off-the-shelf Software-defined Radios (SDRs) operating at 433MHz to simply detect the presence and absence of an object in an enclosed space [23]. Building on those initial results, in this paper we use CNNs for three more complex tasks: (1) Object detection/material classification (polyethylene terephthalate, glass, metal), (2) Detection of multiple visually and materially similar objects and the human-body, and (3) Object detection with increasing distance between T_x and R_x. As in the general schema, the method for each experiment follows four steps, namely data collection, preprocessing, feature extraction, and classification. The preprocessed representation is fed as an input to CNN which performs the feature extraction process.

Fig. 3: The 3 different objects and materials we classify in the first set of experiments

A. Hardware and Software Setup

We use a hardware platform to send and receive signals and software to control the transmission and run the classification algorithms. The transmitter (T_x) comprises of a Raspberry-Pi 3 and the receiver (R_x) is made of a RealTek Software Defined Radio (RTL-SDR) dongle and an off-the-shelf dipole antenna. Across all of our experiments, the dipole antenna is directed towards the receiver. We use Python libraries `pyrtlsdr` and `scipy` for interfacing and transmitting a continuous narrowband FM wave centred at 433 MHz (with a bandwidth of 8 MHz). The signal is sampled at 2.048 MS/s in raw I/Q form with each recorded data point/sample being $0.5s$ long.

B. Data Collection & Labelling

We first perform an initial experiment by training a CNN to detect the presence and absence of an object in a closed room. The CNN is able to achieve a near perfect accuracy on this simple task. Following this, we systematically explore the potential of low frequency radio-based object detection systems for everyday surveillance tasks. For this, we gradually increase the complexity of our experiments.

Three datasets, namely D_A, D_B, and D_C are collected in a closed room of dimensions : $8.5\text{m} \times 5.5\text{m}$

1) D_A: Building on system used by [23] for detecting objects, we conduct experiments to classify three different types of objects and materials. For this, we use three bottles of roughly the same dimensions, each made of different materials – polyethylene terephthalate (PET), steel, glass. The three bottles (polyethylene terephthalate, steel, glass) have height and diameter of $(24\text{ cm}, 7\text{ cm})$, $(25\text{ cm}, 6.5\text{ cm})$, and $(26\text{ cm}, 7\text{ cm})$ respectively. The bottles are placed exactly in the middle of the T_x and R_x placed 2.1 m apart. We collect 2000 samples for each bottle. Figure 3 displays the three bottles used for this experiment.
2) D_B: Through our second experiment, we want to test if our system can learn to differentiate between multiple objects of the same class and humans simultaneously. To test this, we train a network on spectrograms of single PET bottle, two identical PET bottles placed 30 cm apart

in line-of-sight of the antenna, human in the middle, and no objects placed between T_x and R_x. For each of these cases, 2000 samples are collected with T_x and R_x placed 4 m apart.

3) D_C: Following the above experiment, we try to find out the impact of range of our setup on its detection capabilities. We collect datasets by gradually increasing the difference between T_x and R_x. A total of three datasets are collected at with distance between T_x and R_x at 4m, 6m and 8m; for each of the mentioned cases, 3000 samples each were collected, each, with the bottle placed in line-of-sight and with no bottle in the middle. We train three separate networks to examine how the accuracy of simple object detection is affected as the range is increased.

(a) (b)

Fig. 4: Spectrograms for a) presence of human-body and b) absence of object/human in the room for D_B. Indistinguishable for humans, but CNNs can find useful patterns to classify.

C. Preprocessing

Using the radar output from the system setup, Short Time Fourier Transform (STFT) was performed on each recorded I/Q sample to generate the spectrogram in the frequency-time domain. STFT visualises the intrinsic instantaneous frequency components of the signal spectrum which allows the micro-Doppler signal to be visible. STFT is a Fourier related transform in which we divide a longer time signal into shorter segments of equal length and then compute the Fourier transform of each of the shorter segments separately. For an arbitrary signal $x(t)$ in the time domain, STFT is defined as:

$$\mathrm{STFT}_x[n,k] = X[n,k] = \sum_{m=n}^{n+(N-1)} x[m]e^{-j\frac{2\pi k}{N}m},$$

where $k = 0, 1, \ldots, N-1$, and the discrete variables n and k represent time and frequency respectively. The STFT at a certain time n corresponds to the FFT of the sequence from samples $x[n]$ to $x[n+(N-1)]$ [24].

We plot the changing spectra of the input signal, resulting in a spectrogram [25]. As no discernible differences are observable in spectrograms of different classes, we retain their original size of 288×432. The RGB values are converted to floating point integers and normalised (divided by 255.0), and the resulting floating point matrices of size $288 \times 432 \times 3$ are fed into the CNN. Our results show that while the human eye cannot find anything to classify here, CNNs can still identify useful patterns.

V. Experiments and Results

A. CNN Architecture

We start with a baseline architecture consisting of two convolution layers and two dense layers, then progressively vary the hyperparameters to analyze their effect on performance and arrive at the final model(s). The selected CNN model is based on the VGG-16 [26] architecture which has shown remarkable success in image recognition. In our final model(s), two or more CNN blocks are followed by one Dense block. A *CNN Block* is defined as:

$$\mathrm{CL} \rightarrow \mathrm{BN} \rightarrow \mathrm{CL} \rightarrow \mathrm{BN} \rightarrow \mathrm{MP} \rightarrow \mathrm{DL},$$

where CL is a Convolutional Layer with k neurons and kernel size 3×3 with SAME padding. ReLU activation functions are used in the convolutional layers to introduce non-linearity. BN is Batch Normalization Layer, MP is Max Pooling Layer with pool size (2×2), and DL is the Dropout Layer. We define a *Dense Block* as:

$$\mathrm{DeL} \rightarrow \mathrm{BN} \rightarrow \mathrm{DL} \rightarrow \mathrm{DeL} \rightarrow \mathrm{SL},$$

where DeL is the Dense Layer and SL is the Softmax Layer. For classification, the output of the final CNN block is flattened, and a dense layer with ReLu activations is added, followed by BN and the final dense layer with softmax outputs for classification. The number of neurons in the final Dense Layer depends on the number of classes in our data.

B. Training and Inference

We use `keras`' [27] application program interface (API) built on top of TensorFlow [28] to create our models. We use a system with 32 GB RAM and RTX 2070 GPU for training our models.

We one-hot encode all labels belonging to different classes. For multiple labels, we generate multi-hot encoded vectors. We keep 20% of each dataset for testing and use 80% for training and validation. We use *Categorical-cross-entropy loss* for multi-class classification and *Binary-cross-entropy loss* for the multi-label case, along with Adam optimizer and learning rate of value α. A mini-batch size of 8 is used for training the models, keeping 10% of the training data for validation

1) Object/Material Classification: The CNN model uses two CNN blocks (32 and 64 neurons), one Dense block (with 512 and 3 neurons), $\alpha = 0.001$ and dropout rate of 0.5 in both the CNN and Dense blocks.

Classified/Actual Class	Glass	Metal	Plastic
Glass	100%	4%	–
Metal	–	95%	–
Plastic	–	1%	100%

TABLE I: Confusion matrix for object and material classification using spectrograms. The success rate is 98.0%

We formulate material/object classification as a 3-label classification problem (results in Table I). We observe that each class achieved a high success rate, resulting in an overall accuracy of 98%. The result shows promise in the use of NNs on spectrograms for detecting different materials.

2) Human-body/Multi-identical Object Detection: Here, our model uses three CNN blocks (32, 64, and 128 neurons) and a Dense block (512 and 4 neurons). In the first block, Dropout is set to 0.7, and 0.5 is used for the other two blocks. We set the learning rate to $\alpha = 0.0001$ as a higher learning rate in this case was not conducive for a higher accuracy. Furthermore, while performing hyperparameter tuning we remove Dropout and Batch Normalization layers as it yields best results.

It's advantageous for us to use the 433 MHz sensor for the detection of multiple objects, as it has a better penetration when compared to the WiFi-spectrum. Penetration of RF signals is inversely proportional to its frequency so the range and performance of a 433 MHz sensor is significantly better than a 2.4 GHz sensor especially in environments where there are absorbing materials like walls, metals and human objects [29]. At lower energies in free space, the incident wave interacts with the material in various different ways which might lead to it getting reflected, re-emitted and absorbed uniquely depending on the nature of the material.

The task of detecting multiple identical objects and human-body is formulated as a 4-label classification problem. The results can be seen in Table II. The size of the objects under detection are also comparable to the wavelength of the emitted signal which makes it theoretically possible to detect them. We observe a high accuracy for each class. This suggests that the CNN is able to extract relevant features to distinguish between a single instance of an object and two of the same objects present together, implying that such systems have potential to monitor multiple similar objects in an environment. Moreover, this property combined with the high accuracy achieved in human-body detection suggests that novel, competent and inexpensive systems can be developed with further research in low-frequency radar.

Classified/Actual Class	1bottle	2bottles	Human-body	Air
1bottle	100%	1%	–	–
2bottles 30 cm apart	–	99%	–	–
Human-body	–	–	97%	3%
Air	–	–	3%	100%

TABLE II: Confusion matrix for multi-object and human-body detection using spectrograms. The success rate is 99.3%

3) Range Analysis: For studying the range of our setup, we use a similar model architecture as used in 2) with three CNN blocks (32,64 and 128 neurons) and a Dense block (512 and 2 neurons). Learning rate is set to $\alpha = 0.0001$ and a dropout rate of 0.5 is used in both the CNN and Dense block.

Detecting the presence and absence of an object is posited as a simple binary classification task. In accord with the results observed by Singh et al. [23], we obtain a high accuracy for the detection task as the range of the setup increases from 4m to 8m. It is interesting to note that even after the distance between T_x and R_x is doubled, no detrimental effect on the accuracy is observed. By using a dipole antenna of 13 cm long, and increasing its end to end distance, we optimize it to detect emitted waves up to 8 metres. The objects under detection are also kept in the line of sight of the transmitter and receiver, which makes it easy for our system to detect the emitted waves. Moreover, our setup is ideal for indoor detection as it can transmit/receive over long distances without requiring high power consumption on a battery [17]. While this result is promising, it would be interesting to find out how our system would perform if an multiple objects are occluded behind one another.

Dataset	**Validation Accuracy**	**Testing Accuracy**
$4m$	0.98	0.99
$6m$	0.98	0.99
$8m$	0.98	0.99

TABLE III: Comparison of performance of CNN as the distance between T_x and R_x is increased from 4m to 8m

VI. Conclusion and Future Work

We show the capability of low-cost radars, solely backed by deep neural networks, to perform well in different types of object detection and classification tasks. In comparison to traditional radar-systems, our approach distinguishes itself, as it relies on low-frequency radio waves of the magnitude one-twentieth of its contemporaries. This also means that the radiated power level from our radar sensor is much lesser than that of other radio sources such as WiFi or smartphones, making it safe for prolonged exposure [30]. We demonstrate that our system not only achieves high accuracy in multi-class object/material detection with almost no explicit signal processing, but it is also scalable to larger distances. Our approach isn't strictly limited to object-based detection, and shows immense potential for detection of humans as well. This opens up an opportunity for countless applications in our day-to-day lives: Ambient-assisted living (AAL) services for the elderly; human activity recognition to detect anomalies in health; Multi-modal Internet of things (IoT) based systems for patient monitoring in hospitals. As our paper primarily relies on portable and inexpensive software defined radios, it can be easily deployed within indoor environments.

Our future work aims at exploring the potential of our system to perform motion detection by using low-frequency FMCW waves, and to develop a real-time monitoring system which can be used in several domains such as health monitoring, theft detection, causality detection, pedestrian surveillance, etc. However, the questions of immediate interest about the robustness of low-frequency detection - is it possible to accurately localise objects? what would be the scope of such localisation? it possible to monitor a moving object in 3-D space? When does the low-frequency aspect of our system become a bottleneck? Our contribution indicates that further research into these questions may be a very worthy endeavour for the radar community.

References

[1] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[2] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep learning-based object classification on automotive radar spectra," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[3] E. Schubert, F. Meinl, M. Kunert, and W. Menzel, "Clustering of high resolution automotive radar detections and subsequent feature extraction for classification of road users," in *2015 16th International Radar Symposium (IRS)*. IEEE, 2015, pp. 174–179.

[4] T. Müller, P. Marquardt, and S. Brüggenwirth, "A load balancing surveillance algorithm for multifunctional radar resource management," in *2019 20th International Radar Symposium (IRS)*, 2019, pp. 1–9.

[5] V. C. Chen, F. Li, S. . Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] M. S. Seyfioğlu, S. Z. Gürbüz, A. M. Özbayoğlu, and M. Yüksel, "Deep learning of micro-doppler features for aided and unaided gait recognition," in *2017 IEEE Radar Conference (RadarConf)*. IEEE, 2017, pp. 1125–1130.

[8] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 180–189, 2017.

[9] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[10] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 267–281.

[11] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-doppler based human-robot classification using ensemble and deep learning approaches," in *2018 IEEE Radar Conference (RadarConf18)*. IEEE, 2018, pp. 1043–1048.

[12] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep learning-based object classification on automotive radar spectra," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[13] T. T. Takahashi, "How the owl tracks its prey – ii," *Journal of Experimental Biology*, vol. 213, no. 20, pp. 3399–3408, 2010. [Online]. Available: https://jeb.biologists.org/content/213/20/3399

[14] H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley, "RadarCat," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, Oct. 2016. [Online]. Available: https://doi.org/10.1145/2984511.2984515

[15] F. Jin, R. Zhang, A. Sengupta, S. Cao, S. Hariri, N. K. Agarwal, and S. K. Agarwal, "Multiple patients behavior detection in real-time using mmWave radar and deep CNNs," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, Apr. 2019. [Online]. Available: https://doi.org/10.1109/radar.2019.8835656

[16] M. G. Amin and B. Erol, "Understanding deep neural networks performance for radar-based human motion recognition," in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 1461–1465.

[17] O. Hicham, A. Mouhsen, H. Lagraini, T. Abdelmoumen, and M. Chhiba, "Smart monitoring information system based on rf 433 mhz (smis)," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, p. 5143, 12 2019.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386

[19] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," 2016.

[20] Y. Aytar, C. Vondrick, and A. Torralba, "Soundnet: Learning sound representations from unlabeled video," 2016.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, p. 448–456.

[22] N. Srivastava and N. Srivastava, "Improving neural networks with dropout."

[23] "radio-based object detection using deep learning."

[24] M. Garrido, "The feedforward short-time fourier transform," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 9, pp. 868–872, 2016.

[25] E. Sejdić, I. Djurović, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153 – 183, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105120040800002X

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[27] F. C. et al., "Keras,," 2015. [Online]. Available: https://keras.io

[28] M. A. et al., "Tensorflow: Large-scale machine learning on heterogeneous systems,," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[29] "Choosing an active rfid based rtls platform: 433mhz vs wi-fi," Dec 2020. [Online]. Available: https://www.guardrfid.com/blog/choosing-an-active-rfid-based-rtls-platform-433mhz-vs-wi-fi/

[30] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "Radar and rgb-depth sensors for fall detection: A review," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3585–3604, 2017.