RF-Net: Radio Frequency based Objection Detection

Vedansh Priyadarshi
Department of Computer Science
Ashoka University
vedansh.priyadarshi_ug21@ashoka.edu.in

Abstract

Object detection is one of the most important application of deep learning, which is widely applied in various tasks like detecting cracks in the metal, class attendance system, autonomous vehicle and so on, with the purpose of locating object of a certain class. In recent years, a new object detection approach, which takes advantages of radio frequency signal, has been proposed. In this paper, we talk about how various features of radio signal changes when an object is kept between transceiver and a receiver. We propose a deep learning model trained on dataset of spectral density of radio frequency, with and without object. Our proposed model learns the features of radio signal which change when we place an object between them. At last, we point out various future research direction of this new topic.

Keywords: RF, CNN, RTL-SDR

1 INTRODUCTION

Object detection system aims to accurately predict the class of object based on a predefined model. It is a hot research topic in the field of computer vision and widely used in human and industry centred applications such as self-driving cars (Fujiyoshi et al., 2019), steel crack detection (Lee, 2011), human activity detection (Jobanputra et al., 2019), etc.

To recognise objects, physical sensor like cameras, gyroscope, radar, etc. are often deployed in the environment to continuously collect these sensor readings. Then, based on the model trained on the data collected by these sensors, object recognition is done. The traditional method of collecting data for object detection is to use a camera. Camera sensor based methods take advantage of camera to capture a particular event, or record the sequence of events. Depending on the camera, an image or video might be of RGB, depth, or RGB-D type. Using other physical sensors like pressure, proximity, RFID can be used to infer the shape or size of the objects or activity of humans. Although traditional object recognition methods are widely accepted, they require expensive sensing module and raise some concerns, such as energy, privacy, feasibility, and deployment costs.

The importance of this problem and the limitations of existing methods have motivated researchers to find the feasible solution of detecting objects which takes privacy, energy, and costs into the consideration. Radio based methods haven't been implemented by many researchers to detect the physical objects. Most works have been done on the systems that transmit a low power RF-signal and analyze its reflection off *human*'s bodies to analyze their emotion, movements, etc.

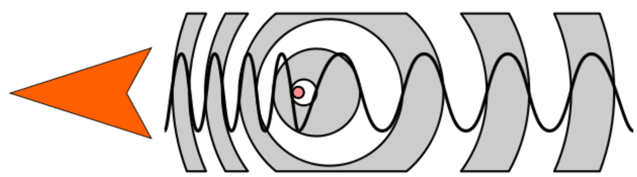

Figure 1: Doppler effect

State-of-the-art RF based experiments on detecting various human activities or traits can be classified into two categories. The first is Doppler's effect based radar (Gadde et al., 2014). This method is useful in human body fall detection. Fast motion in human body causes spike in the Doppler's frequency. The second method is using WiFi's Channel State Informations (CSIs) to detect various human activities. This uses Fourier's transform to compute changes in the WiFi channel. It is worth noting that both of these methods have been applied on detecting the moving humans.

In this paper, our main aim is to address the issue of detecting the objects at rest which generalizes across various environments like electrically noisy room, lots of people moving in the room, etc.

This paper is organised as follows. Section 1 introduces this area. Section 2 talks about different techniques of sensing objects in an wireless channel. The previous works that have been done is discussed in the section 3. Section 4 introduces to the preliminaries required to understand this paper. Section 5 talks about the approach that we are taking. Section 6 concludes the paper with future works that could be done.

2 TECHNIQUES FOR SENSING

There are many physical layer properties that can be extracted over an wireless channel to facilitate activity sensing and object detection. The four common sensing techniques based on different layer properties are Received signal strength indicator(RSSI), channel state information(CSI), frequency shift for frequency modulated carrier wave(FMCW), and Doppler shift.

2.1 Received Signal Strength Indicator (RSSI)

RSSI is available in most of the modern routers, which indicates the path loss of signals with respect to distance, and can be derived using Log-normal distance Path Loss(LDPL) (Seidel and Rappaport, 1992) model:

$$P(d) = P(d_0) + 10\gamma \log \frac{d}{d_0} + X_\delta$$

where $P(d)$ denotes the RSSI measurement at distance d in Decibel(dB), $P(d_0)$ is the path loss at distance d_0. γ is the path loss exponent, and X_δ is the zero-mean normal noise.

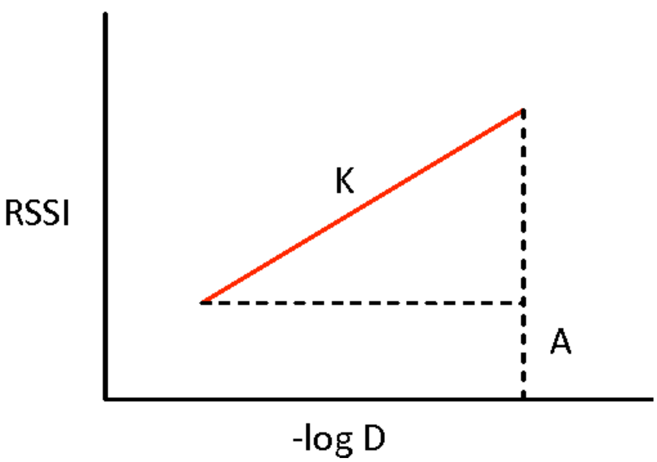

Figure 2: Expected relationship between RSSI and [-log (distance)]

The RSSI has been used for a very long time in position estimation (Bahl and Padmanabhan, 2000). It has also been seen that the presence of human body within wireless range causes signal attenuation, leading to the spike in RSSI measurement. Thus, RSSI has been widely used in various human-based activity recognition, for instance, device-free indoor localisation (Zhang et al., 2011), density estimation of crowd in a room (Xu et al., 2013), breathing rate monitoring (Kaltiokallio et al., 2014).

The limitations of RSSI is that it can only detect limited types of human activities due to single path loss. Also, It has already been shown that RSSI is not effective in the indoor environment (Parameswaran et al., 2009).

2.2 Channel State Information(CSI)

To achieve reliable human sensing, it becomes very important that we capture fine-grained CSI information, which represents the combined effect of scattering, fading, and power decay with distance.

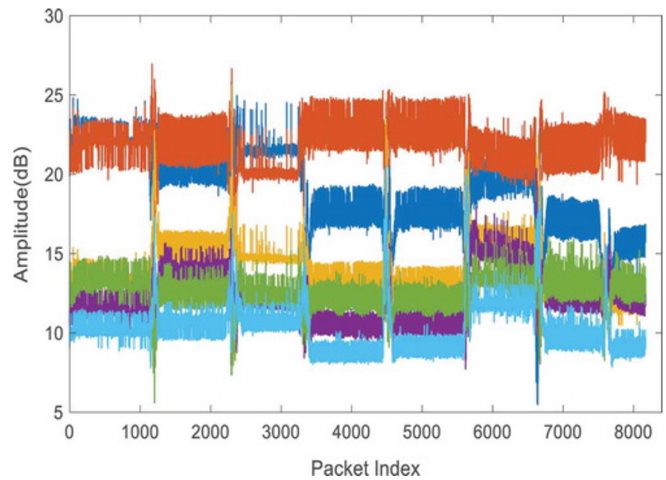

Figure 3: Raw CSI data with noise

Since wireless signal can travel through any indoor corner, the presence of any object would alter the propagation of the wireless signal, resulting in small changes as shown in Figure 4. All these multi-path rays contribute to the measurable CSI values, and can be used to detect and track human body movements.

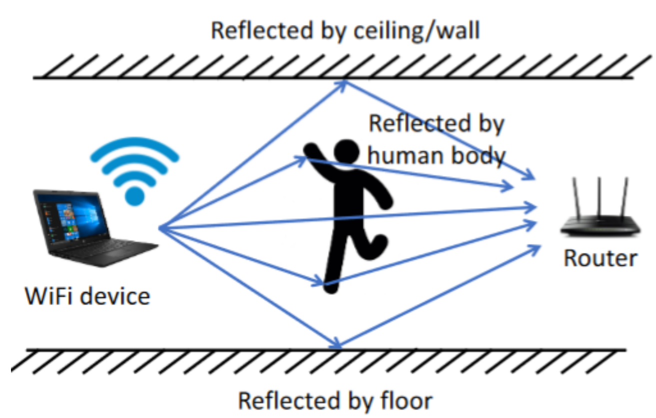

Figure 4: Illustration of the multi path effect of wireless signals.

In contrast to RSSI, CSI consists of sets of complex values containing both amplitude and phase information for multiple orthogonal frequency-division multiplexing(OFDM) subcarriers. Each carrier has a different center frequency. This causes different multi-path fading effects. IEEE 802.11n standard can render the CSI measurements for 52 and 128 subcarriers with 20 MHz and 40MHz bandwidth for each subcarrier respectively.

CSI allows fine-grained channel estimation, which can be expressed as:

$$H = [H_1, H_2, ..., H_i, ..., H_{N-1}, H_N]^T, \text{where } i \in [1, N]$$

where N is the number of subcarriers, and H_i can be represented as:

$$H_i = |H_i| e^{j \sin H_i}$$

Similar to RSSI, CSI measurements can be obtained in any off-the-shelf WiFi devices by tinkering with WiFi drivers. It has already been used by many researchers for human intrusion detection, walking direction, and activity recognition (Wang et al., 2014) (Wang et al., 2015).

2.3 Frequency Modulated Carrier Wave(FMCW)

The activities can also be detected by capturing the radio reflections off body, by estimating the time it takes the wireless signal to travel from the transmitter, bouncing off the body, and then coming back to the receiver. The main issue here is measuring this time because wave travels at the speed of light. FMCW maps differences in time to the shifts of carrier frequency which is deployed to measure the time of flight of radio signals.

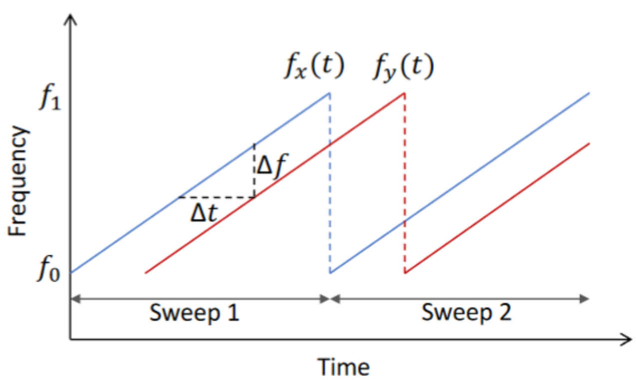

Figure 5: Illustration of FMCW operation

As shown in Figure 4, the carrier frequency of the transmitting wireless signal is repeatedly swept across a specific bandwidth. After bouncing off the human body, it will introduce the frequency shift and the time shift can easily be calculated with respect to the transmitting signal. It should be noted that FMCW method relies on special devices like USRP to generate signals that sweeps the frequency across the time. FMCW techniques have already been used in various research works like capturing human figure through a wall (Adib et al., 2015a), track user's 3D motion using WiTrack (Adib et al., 2014), monitor breathing and heart rate (Adib et al., 2015b), etc.

2.4 Doppler Shift

Doppler shift is also one the property of wireless signals which can be used to perform objection detection and activity sensing. It measures the frequency change as transmitter and receiver move to each other. When signals transmitted bounces back from human body; if we think this bounced back signal as signal emitted by a transmitter, any movements in human body will result in a Doppler shift.

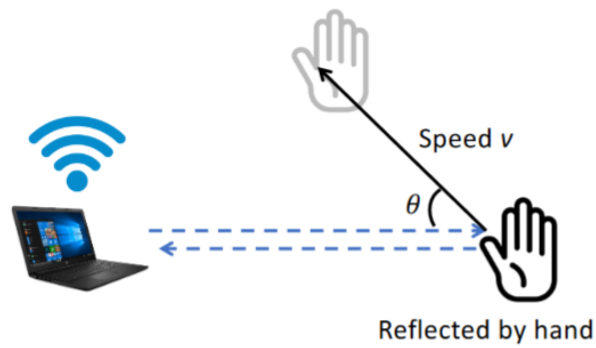

Figure 6: Illustration of Doppler Shift

If an object or a person moves toward the receiver, positive Doppler shift is inferred. If an object or a person moves way from the receiver, negative Doppler shift is inferred. When an object moves at speed v along the direction θ with respect to the receiver, the Doppler shift is:

$$\Delta f = \frac{2v \cos\theta}{c} f$$

where f is the center frequency of the wireless signal and c is the speed of light.

This technique has already been used with Software Defined Radios(SDRs) to detect walking (Adib and Katabi, 2013), running (Chetty et al., 2012), and human gestures (Pu et al., 2013a) (Tan et al., 2014).

3 PREVIOUS WORKS

Most of the works done before are related to reflecting radio frequency from human body and using deep neural network methods to learn complex features which cannot be easily extracted.

(Al-qaness, 2019) proposed a device-free CSI-based human activity recognition system which exploits CSI of an indoor ubiquitous wireless device. They described recent research in device-free human sensing technologies and proposed an effective method of CSI filtering, pattern segmentation using machine learning methods, and micro-activity classification. They evaluated the proposed method in an indoor environment and took useless noise into consideration. Their model was able to classify nine human micro-activities. They evaluated their model in various scenarios with ten people and their model reached high accuracy. There are some limitation to their methods as well. It can't track movements of two or more people. Also, they didn't talk about the problem we're trying to solve.

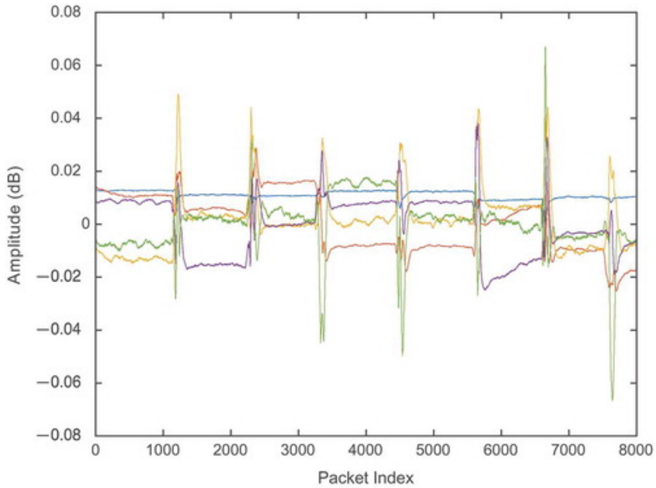

Figure 7: CSI data after PCA

(Zhao et al., 2018) proposed a wireless system called EQ-Radio that infers people emotions from the radio signals that bounce off their bodies. EQ-Radio tried to segment these RF reflection to the individual heartbeats. It had no idea how the heartbeats look like in a reflected signals. There are two additional issues that they faced while classifying various human emotions. First, the reflected signals from human bodies were noisy. Second, they were operating in a low Signal-to-Interference-and-Noise Ratio (SINR) where there were too much interference due to breathing. EQ-Radio gave results very close to the ECG-based models.

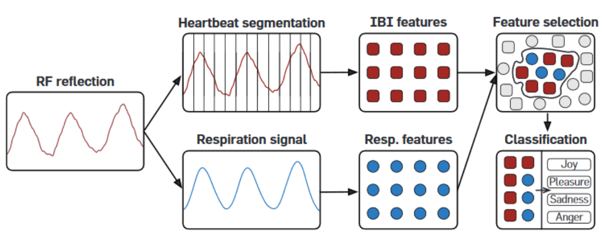

Figure 8: EQ-Radio Architecture

CSI based methods have currently been popular and powerful tool for intrusion detection system. This is due to fine-grained wireless channel measurement property of CSI. (Nishimori et al., 2011) measure the influences of antenna arrangement on the radio signal propagation and then use the channel matrix data to detect intrusion. (HONG and OHTSUKI, 2012) proposed a method that uses eigenvector and eigenvalue spanning the signal subspace as features, obtained from the matrix channel data. The aforementioned studies mainly rely on the characteristic from matrix of CSI amplitude, phase or phase difference to detect the sudden changes associated with human intrusion.

(Pu et al., 2013b) proposed a Doppler shift based gesture recognition system, WiSee which is able to recognise 9 hand/leg gestures based on the unique Doppler shift profiles extracted from wireless signals. Due to different body movements with respect to the SDR and transmitter, there is an unique positive and negative Doppler shift pattern for different activities. WiSee has achieved the highest accuracy of 94%.

Due to low granularity, RSSI-based methods depend on multiple wireless links from various devices to capture minute movements related to vital signs like heartbeats and breathing rates. Many studies turn to CSI signals for detecting such vital sign without using any complex devices. (Liu et al., 2015) proposed to reuse the wireless device like WiFi router to detect the vital signs without requiring any additional wearable and hardware. CSI amplitude received from various wireless devices over time when person is asleep in Figure 9.

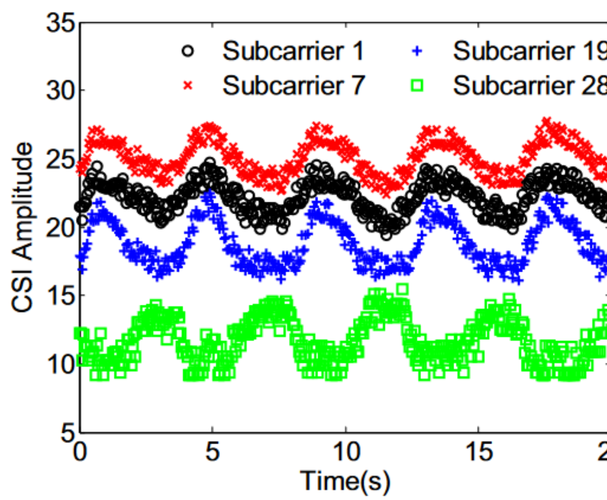

Figure 9: CSI amplitude over time when person is asleep

4 Prelimiaries

In this section, we talk about the preliminaries required to understand this paper. The preliminaries include basics of Power spectral density, I/Q components of the wave, and CNNs.

4.1 Power Spectral Density

The power spectrum of any time series data describes the distribution of power in frequency component that composes the signal. For continuous signals, there is a need to define the Fourier transforms of the signals, which is called Power spectral density(PSD). The average power P of the signal $x(t)$ is given by the following squared of signal averaged over time:

$$P = \lim_{T \to \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt$$

4.2 I/Q components

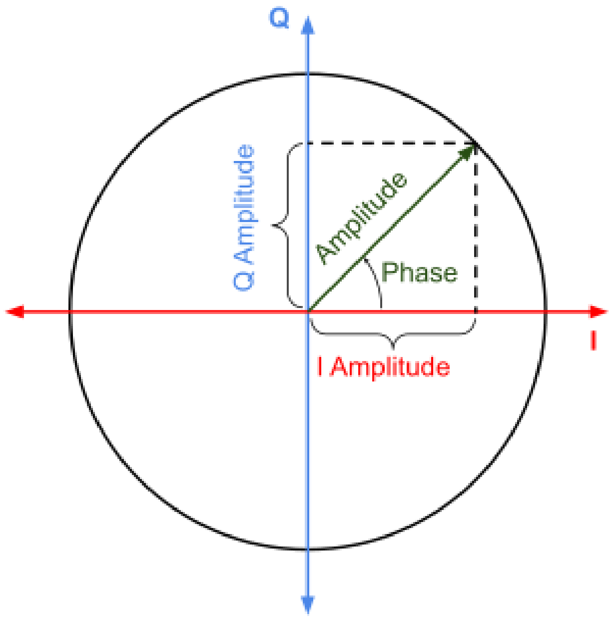

Figure 10: IQ Phasor diagram

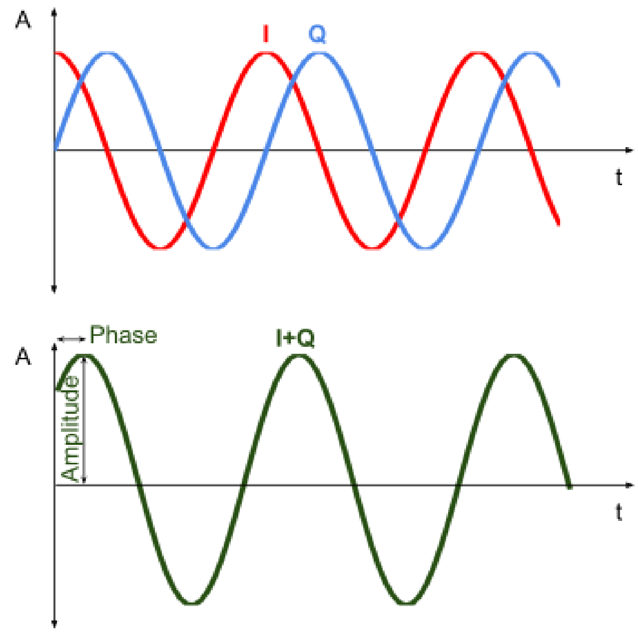

Figure 11: Composition of I and Q components

A sine wave with angle modulation can be decomposed into two amplitude modulated sine waves that are off-set in-phase by $\frac{\pi}{2}$ radians. The original wave functions and the decomposed wave functions, all three, have same center frequency. These two amplitude modulated components are called in-phase(I) and quadrature(Q) components of the wave.

In amplitude modulation with carrier frequency f and ϕ as a time variant function, when multiplied by amplitude gives I and Q components of the wave as given below:

$$\begin{aligned} A(t) \cdot \sin[2\pi ft + \phi(t)] = A(t) \cdot \sin(2\pi ft) \cdot \cos[\phi(t)] \\ + A(t) \cdot \cos(2\pi ft) \cdot \sin[\phi(t)] \end{aligned}$$

$A(t) \cdot \sin(2\pi ft) \cdot \cos[\phi(t)]$ is the I component of the sine wave and $A(t) \cdot \cos(2\pi ft) \cdot \sin[\phi(t)]$ is the Q component of the sine wave.

4.3 Convolutional Neural Network

Convolutional Neural Networks(CNNs) is the widely used deep learning framework inspired by the visual cortex of the animals. LeNet was the pioneering work in Convolutional Neural Networks by (Lecun et al., 1998). ConvNets are very similar to the normal neural networks which can be visualized as a collection of neurons arranged as an acyclic graph. The main difference from a neural network is that neurons are sparsely connected to the previous layer. This sparse connection of neurons make CNN capable of learning features implicitly. We're using CovNets because we want the network to learn features automatically due to complex nature of the data. Classical machine learning methods require extracted feature as an input.

5 APPROACH

In this section, we talk about the experiments which includes setup, training of our model, and our observations.

5.1 Setup

We make our own transmitter using Raspberry Pi to transmit the data. We use official RTL-SDR as the receiver to demodulate the signal. The transmitter and SDR is placed at 90 cm apart. We use `pyrtlsdr` Python library to demodulate and preprocess data from SDR. The data is collected in the raw I/Q format over time. We also use SDRSharp to visualize data as an spectrogram over time by placing different bottles between transmitter and receiver. The data collection is done in an empty room to reduce interference due to other structures.

We try to differentiate by using bottle made up of plastic and metal, filled with and without water. The Raspberry Pi emits a low power signal of 433 MHz with a bandwidth of 8MHz. The signal is received and sampled by SDR at 2.6 ms/s with each recorded sample being 0.5 seconds long. We create dataset by placing object between and no object between the transmitter.

5.2 Training

For training we use VGGNet based architecture to train our model (Liu and Deng, 2015). Then, we do lots of experiments with different model and change the hyper-parameters of the model as suggested by (Soon et al.,

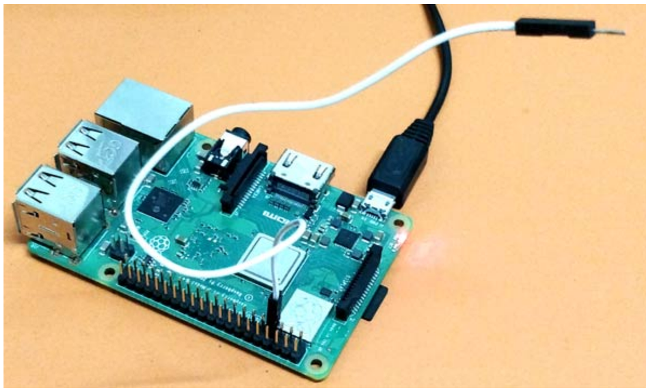

Figure 12: Raspberry Pi based transmitter

2018). We also try to use more deeper Convolutional Neural Networks like AlexNet, but our results are not getting better.

Finally, we stay with architecture given in the Figure 17. We use RELU activation to learn non-linearity in the data. We also use Batch after convolutional layer. We also add pooling layer alternatively. We use dropout at the final layer to prevent model from overfitting.

We train our model on various data as mentioned below.

5.2.1 I/Q data

We input raw I/Q data by treating the resizing the data as two dimensional input vectors to the model for learning complex features. It is the most common used method in the signal processing. We generate 10,000 samples of length 1024 with and without bottle.

5.2.2 Power Spectral Density

We generate the power spectral density function for each 0.5 second sample of I/Q data and plot the power values at different frequencies with respect to time. We obtain a 20×100 dimension vector capturing the variation of PSD values at 433 MHz. This method has also been used in (Ma et al., 2020). We obtain 1,000 samples for both with and without bottle class which are used to train the CNN. Power spectral Density plots for each case are shown in Figure 13.

5.2.3 Spectrograms

We apply Fast Fourier Transform to each I/Q data points and convert all into a spectrogram representation in the frequency-time domain. We obtain 1,000 samples for each bottles and without bottles class. A sample of spectrogram is shown in the Figure 14. The image is cropped and resized to 224, and then the image is converted as a vector representation and normalised between 0 and 1, the the resulting tensor of size $224 \times 224 \times 3$ are fed into the network for training.

5.3 Results

Our model overfitted on the RAW I/Q and Power Spectral Density data.

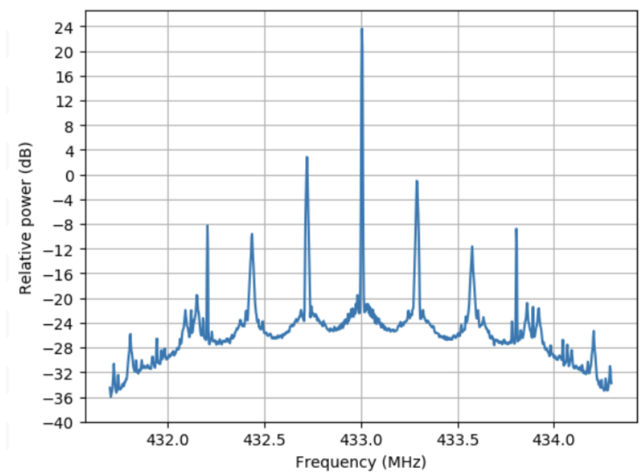

Figure 13: Power Spectral Density

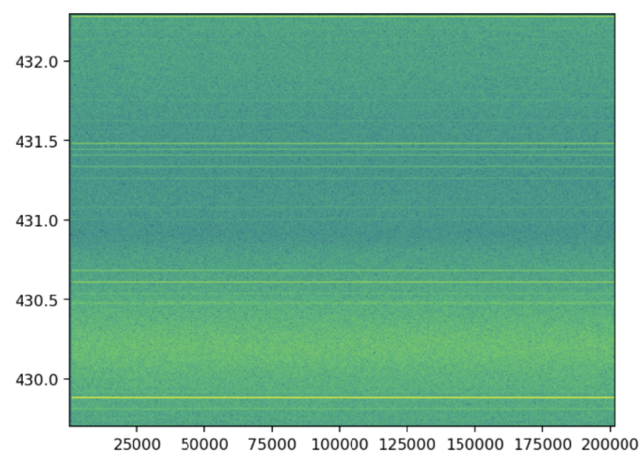

Figure 14: Spectrogram

After not getting better accuracy with raw I/Q and PSD, we used spectrogram to train our model with softmax outputs and binary cross-entropy loss. Softmax outputs are more suitable for this multiclass problem as softmax enforces some dependency upon the predictions. We observe that the network is able to locate a global minima on the loss surface and create a generalised decision boundary that achieves a high accuracy on the test set. The graphs for loss and accuracy are also shown below in Figure 15 and 16. We deduce that spectrogram representation is better representation than other two for detecting the presence of object between the transmitter and receiver. However, this accuracy is not achieved every time we train our model, which further supports our speculation about the extreme noise and very little relevant information in data which makes it harder for the CNN to discover useful features.

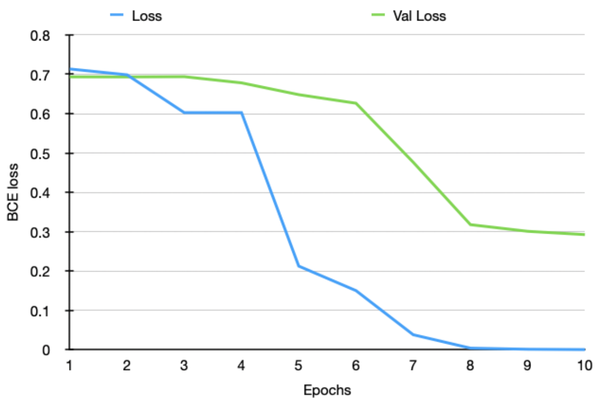

Figure 15: Training and Validation Loss Curve

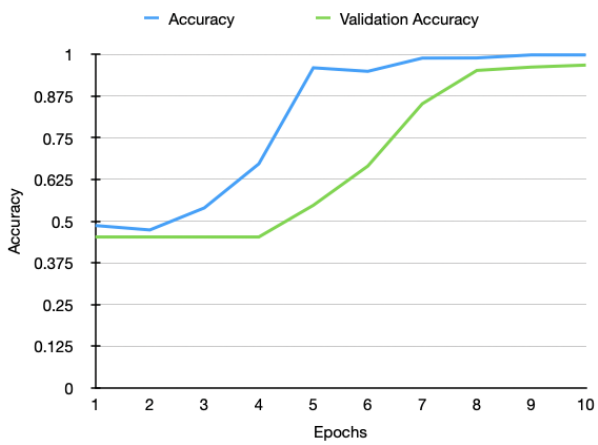

Figure 16: Training and Validation Accuracy Curve

6 CONCLUSION

In this paper, we show that the radio frequency could be exploited to detect object between transmitter and receiver. In particular, our system exploits Spectrograms data to learn the features engrained in the data. Our algorithms grounded on I/Q data in both time and frequency domain, and proved that RF signals have the capability to estimate whether the object is present or absent. Extensive experiments in our lab over a four-month time period confirm that our proposed approach using the existing radio frequency can achieve comparable or even better accuracy as compared to existing dedicated radio based.

We can also make some various changes in the model by using the different architectures like LSTM which works very well on sequential data.

7 ACKNOWLEDGEMENTS

I'd like to thank Professor Debayan Gupta for all his guidance and support.

Data	Accuracy
Raw I/Q	0.44
Power Spectral Density	0.51
Spectrogram	0.96

Table 1: Accuracy on various data

References

Fadel Adib and Dina Katabi. 2013. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.*, 43(4):75–86, August.

Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3d tracking via body radio reflections. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, NSDI'14, page 317–329, USA. USENIX Association.

Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015a. Capturing the human figure through a wall. *ACM Trans. Graph.*, 34(6), October.

Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015b. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 837–846, New York, NY, USA. Association for Computing Machinery.

Mohammed A. A. Al-qaness. 2019. Device-free human micro-activity recognition method using wifi signals. *Geo-spatial Information Science*, 22(2):128–137.

P. Bahl and V. N. Padmanabhan. 2000. Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 2, pages 775–784 vol.2.

K. Chetty, G. E. Smith, and K. Woodbridge. 2012. Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1218–1226.

Hironobu Fujiyoshi, Tsubasa Hirakawa, and Takayoshi Yamashita. 2019. Deep learning-based image recognition for autonomous driving. *IATSS Research*, 43(4):244 – 252.

Ajay Gadde, Moeness G. Amin, Yimin D. Zhang, and Fauzia Ahmad. 2014. Fall detection and classifications based on time-scale radar signal characteristics . In Kenneth I. Ranney and Armin Doerry, editors, *Radar Sensor Technology XVIII*, volume 9077, pages 330 – 338. International Society for Optics and Photonics, SPIE.

Jihoon HONG and Tomoaki OHTSUKI. 2012. State classification with array sensor using support vector machine for wireless monitoring systems. *IEICE Transactions on Communications*, E95.B(10):3088–3095.

Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. 2019. Human activity recognition: A survey. *Procedia Computer Science*, 155:698 – 703.

Ossi Johannes Kaltiokallio, Hüseyin Yigitler, Riku Jäntti, and Neal Patwari. 2014. Non-invasive respiration rate monitoring using a single cots tx-rx pair. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, IPSN '14, page 59–70. IEEE Press.

Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324.

Sangwook Lee. 2011. Color image-based defect detection method and steel bridge coating.

S. Liu and W. Deng. 2015. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734.

Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '15, page 267–276, New York, NY, USA. Association for Computing Machinery.

Xunguang Ma, Dashuai Wang, Danhua Liu, and Jimin Yang. 2020. DWT and CNN based multi-class motor imagery electroencephalographic signal recognition. *Journal of Neural Engineering*, 17(1):016073, feb.

K. Nishimori, Y. Koide, D. Kuwahara, N. Honmay, H. Yamada, and M. Hideo. 2011. Mimo sensor - evaluation on antenna arrangement-. In *Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP)*, pages 2771–2775.

Ambili Thottam Parameswaran, Mohammad Iftekhar Husain, and Shambhu J. Upadhyaya. 2009. Is rssi a reliable parameter in sensor localization algorithms – an experimental study.

Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013a. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing Networking*, MobiCom '13, page 27–38, New York, NY, USA. Association for Computing Machinery.

Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013b. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing Networking*, MobiCom '13, page 27–38, New York, NY, USA. Association for Computing Machinery.

S. Y. Seidel and T. S. Rappaport. 1992. 914 mhz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40(2):207–217.

F. C. Soon, H. Y. Khaw, J. H. Chuah, and J. Kanesan. 2018. Hyper-parameters optimisation of deep cnn architecture for vehicle logo recognition. *IET Intelligent Transport Systems*, 12(8):939–946.

B. Tan, K. Woodbridge, and K. Chetty. 2014. A real-time high resolution passive wifi doppler-radar and its applications. In *2014 International Radar Conference*, pages 1–6.

Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *MobiCom '14*.

Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 65–76, New York, NY, USA. Association for Computing Machinery.

C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. 2013. Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 79–90.

D. Zhang, Y. Liu, and L. M. Ni. 2011. Rass: A real-time, accurate and scalable system for tracking transceiver-free objects. In *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 197–204.

Mingmin Zhao, Fadel Adib, and Dina Katabi. 2018. Emotion recognition using wireless signals. *Commun. ACM*, 61(9):91–100, August.

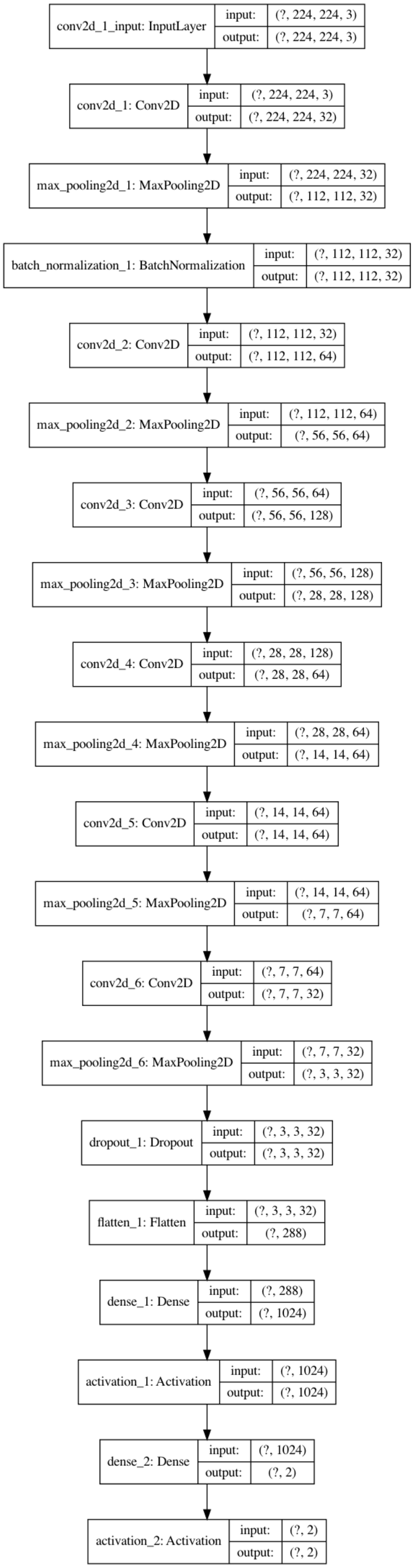

Figure 17: Our model