Concert Hall Monitoring System WIth AI and mmWave Radar

Harshith Nagubandi
Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA
hnagubandi@ucsd.edu

Rahul Polisetti
Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA
rpolisetti@ucsd.edu

***Abstract*—We propose a novel way to solve the problem of people monitoring in concert halls. Instead of the traditional ways to count people i.e., with weight based sensors, camera systems etc., we are proposing a new way to count people and making sure the privacy is preserved. We use traditional CNNs to count the people based on the input range-AoA transforms. We classified a classroom data with 0 to 4 people with an accuracy of 95% on test data. We also propose a CNN based model to detect if any person is present or not in the hall, which performs with an accuracy of 100%. The results show a promising way to use CNNs to detect and monitor people using mmWave radar data and preserving the user privacy.**

***Index Terms*—ML, radar, NN, AI**

I. Introduction

People counting is one of the most useful applications of sensing in daily life and finds applications in building automation control, public safety, and intelligent transportation. Counting people in concert halls and theaters can be used to inform people of empty sections and providing statistics to management about viewership. This information can be displayed outside the concert hall to help guide people to empty sections. The statistics of people counting can be used to set dynamic pricing in the concert halls and also can be utilized in improving the acoustics to better match where people are most likely to be seated, for better user experience. Counting people zone wise also finds its application in regulating the hall management to follow distancing protocols, or also find out if people have correctly evacuated in case of emergency. In this project we take a look at using mmWave radar to detect people in a concert hall or theater. We will further use Machine Learning techniques to provide an accurate prediction of how person count.

Current methods of people detection include using mechanical sensors per seat. Although this method is very accurate, it involves in huge circuitry to connect up every seat and can be very expensive to install and also maintain. Another way is to manually monitor people by either counting the number of people entering the hall or by looking at the overall hall and guessing the density. Another way is to use cameras to record the hall and then use computer vision techniques to track and estimate the number of people. This is effective but one will have to consider the loss of privacy when using such devices. The Wifi CSI activity can also be exploited to detect the density of people in an area but it can be faulty when people use multiple devices at the same time. All these methods prove to be cost prohibitive or require huge infrastructure upgrades and might break privacy. The radar can overcome these challenges and drawbacks since for one we are not recording people. The radar device just has to be installed at the front of the hall, and can also utilize just one device, therefore making it cheap and easy to install.

II. Related Work

Radar systems use reflected signals that are affected by human bodies to detect and classify human activity without the need for devices (such as counting the number cellphones being used on the concert halls network). [4] They analyze the received signals to gather information about the presence and actions of people. Our approach would be to obtain radar data using a mmWave testbed and train a machine learning classifier. We will use various machine learning classifiers and compare them based on accuracy and computational complexity.

In order to estimate distances in the order of meters using the radio frequency is extremely difficult as they travel with the speed of light. One way to solve this is by using Frequency Modulated Carrier Wave to estimate the time of flight and hence the location of the objects. The work on 3D tracking of humans [5] talks about solving this and overcoming few other problems involved in sensing using the radar data. The static object reflections are not in the point of interest and have to be eliminated. Different techniques are used for the stationary clutter removal and we chose the method from [4] that removes the range FFT averages across few frames from the current frame. The other method includes considering the difference across adjacent frames to remove the stationary clutter.

The next step is to get the range azimuth map using which we estimate the location of the people and also use it to count them. To have an accurate estimate on the desired metrics, we need a very good mm-Wave radar. Using low cost mm-Wave radar, machine learning models can be employed to achieve

comparable accuracies. The above discussed methods show how ML models being applied to the range-azimuth maps to locate or count the people.

A. Range-AoA Fourier Transform

The authors of this paper propose a novel method [1] for estimating the range and angle of multiple targets using a frequency-modulated continuous-wave (FMCW) radar system. The authors first analyze the system model of the FMCW MIMO radar and derive the signal model of the received signals. They then propose a joint range-angle estimation algorithm based on the matrix completion technique, which can effectively estimate the range and angle of multiple targets using the received signals. The detailed formulation is discussed below.

The transmitted signal of the FMCW MIMO radar can be expressed as:

$$s(t) = \sum_{m=1}^{M} \sum_{n=1}^{N} w_m(t) w_n\left(t - \tau_{m,n}\right) e^{j(2\pi f_{\mathrm{IF}} t + \phi_{m,n})} \tag{1}$$

where $w_m(t)$ is the transmitted waveform of the m^{th} antenna, f_{IF} is the intermediate frequency, $\tau_{m,n}$ is the time delay between the m^{th} and n^{th} antennas, and $\phi_{m,n}$ is the phase difference between the signals received by the two antennas.

The received signal can be expressed as:

$$\begin{aligned} r(t,\theta) = \sum_{i=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} & A_i G\left(\theta_i\right) w_m\left(t - \tau_{m,i}\right) w_n(t - \tau_{n,i}) \\ & \times e^{j(2\pi f_{\mathrm{IF}} \tau_{m,i} + \phi_{m,i} - 2\pi f_{\mathrm{FF}} \tau_{n,i} - \phi_{n,i})} + w(t,\theta) \end{aligned} \tag{2}$$

where L is the number of targets, A_i is the amplitude of the i^{th} target, $G(\theta_i)$ is the antenna gain in the direction of the i^{th} target, $w(t,\theta)$ is the noise, and θ is the angle of arrival.

The range and angle of arrival of the targets can be estimated using the Range-AoA Fourier Transform (RAFT):

$$\begin{aligned} X(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} & r(mT, nT) w_m\left(nT \right. \\ & \left. - \tau_{m,l}\right) e^{-j2\pi kn/N} e^{j2\pi lm/M} \end{aligned} \tag{3}$$

where k and l are the frequency and angle indices, T is the sampling interval, and N and M are the numbers of samples and antennas, respectively.

The range and angle estimates can be obtained by finding the peaks of the 2D FFT of the RAFT output:

$$\hat{r}_i = \frac{cT_s}{2\left(B + f_c\right)}\left(k_i - 1\right) \tag{4}$$

$$\hat{\theta}_i = \arcsin\left(\frac{\lambda}{2d}\left(l_i - \frac{L-1}{2}\right)\right) \tag{5}$$

where c is the speed of light, T_s is the sampling period, B is the bandwidth of the transmitted signal, f_c is the carrier frequency. λ is the wavelength, d is the distance between the consecutive antennas.

B. mmWave Radar for People Counting

The authors first introduce the system model of the millimeter wave radar [2] and derive the received signal model. They then propose two feature extraction techniques, namely, the spectrogram-based method and the time-frequency analysis-based method, which can effectively extract features from the received signals. The authors also consider three classification techniques, namely, the support vector machine (SVM), the k-nearest neighbor (k-NN), and the random forest (RF), which can classify the extracted features into different classes.

The authors evaluate the performance of their proposed techniques through experiments conducted in a laboratory and in a real-world scenario. The experimental results show that the spectrogram-based feature extraction method combined with the SVM classification technique achieves the best performance in terms of counting accuracy, with an average error rate of less than 3%. The time-frequency analysis-based feature extraction method combined with the k-NN and RF classification techniques also perform well, but not as well as the spectrogram-based method.

The paper also provides detailed equations for the received signal model, the spectrogram-based feature extraction method, and the classification techniques. The paper is significant because it demonstrates the potential of millimeter wave radar for people counting applications, and compares the performance of different feature extraction and classification techniques, providing insights for future research in this area.

C. Neural Networks for People Counting

The authors first introduce the system model [3] of the millimeter-wave radar and derive the received signal model. They then propose a two-stage neural network architecture for people counting. In the first stage, a convolutional neural network (CNN) is used to extract features from the spectrogram of the received signal. In the second stage, a recurrent neural network (RNN) is used to classify the features and count the number of people.

The authors evaluate the performance of their proposed two-stage neural network through experiments conducted in a laboratory and in a real-world scenario. The experimental results show that their proposed method achieves high accuracy in people counting, with an average error rate of

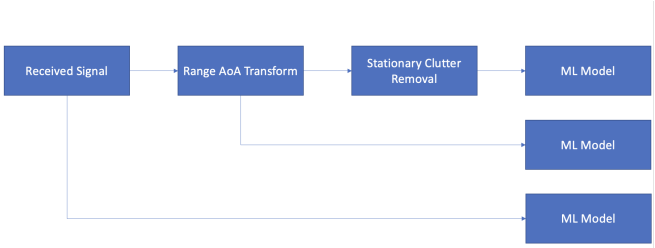
less than 2%. The proposed method also has fast processing time, making it suitable for real-time applications.

The paper provides detailed equations for the received signal model, the spectrogram-based feature extraction method, and the two-stage neural network architecture. The paper is significant because it proposes a novel two-stage neural network architecture for real-time people counting using millimeter-wave radar, which achieves high accuracy and fast processing time. The proposed architecture can be applied to various applications, such as smart homes, surveillance systems, and public spaces.

III. Design

The design of people monitor is based on utilizing the Range-AoA fourier transform with a machine learning model. There are traditional algorithms using peak detection to locate the person and also count them. Solutions for locating people using mmWave radar data already exist in literature but counting people using these techniques can be process intensive and are also prone to errors in cases of counting more number of people. So making use of a machine learning model can be of help here. The base architecture of our design involves capturing the received radar signal data, then computing the Range-AoA fourier Transform, stationary clutter removal and then ML algorithm in the final stage.

We experimented with different architectures given below feeding the received signal directly to the machine learning model, then feeding the Range-AoA transform to the ML model and feeding the Range-AoA transform after stationary clutter removal. The experiments show that only the last architecture is easier for the weights of the ML model to converge and also shown promising results.



A. Convolutional Neural Networks

For each of the below tasks we experimented with multiple architectures and found that the model which takes in the range-AoA fourier transform works better. The convolutional neural network we employed has three sets of convolutional layers that contain convolutional step, Max pooling and Batch Normalization.

The primary function of the convolutional layer is to extract features from the input image. It is a mathematical operation that involves multiplying each element of the input image with a corresponding element of a filter or a kernel. The result of this operation is a feature map, which is a representation of the input image that highlights certain features, such as edges, corners, and textures.

The convolutional layer consists of multiple kernals that are applied to the input image to give multiple kernal maps. Each filter extracts a different feature from the input image to give rise to this kernal maps. The output of the convolutional layer basically is a set of feature maps that represent the input image in a more abstract and compact form.

In addition to feature extraction, the next part of convolutional layer which is the pooling layer, performs downsampling also called as pooling. Pooling reduces the spatial size of the feature maps and helps to reduce overfitting of the traning model. We used Max pooling which is a commonly used pooling technique, that involves selecting the maximum value in each sub-region of the feature maps.

The next part of the layer involves batch normalization. The goal of batch normalization is to normalize the input to each layer of the network, ensuring that the mean and variance of the inputs are roughly the same, regardless of the input data. By reducing the internal covariate shift, which is the change in the distribution of the inputs to a layer, batch normalization helps to stabilize the training process. Batch normalization helps to reduce the vanishing and exploding gradient problems, which can slow down the convergence of the network.

Other advantages of batch normalization include it acting as a form of regularization, reducing the need for other ways of regularization such as dropout, it helps reducing the chance of overfitting the model, and also batch normalization can improve the generalization performance of the network on unseen data.

B. Current Model

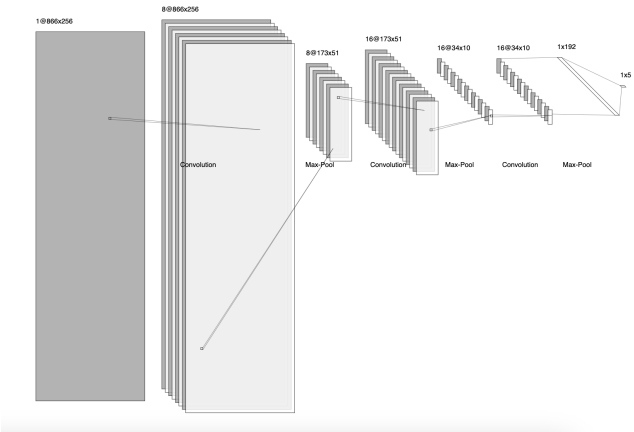


Fig. 1. Train and Test Loss over Epochs

All the models we employed are CNNs with different number of features in each layer. The first problem we aim to solve is detecting a person in the concert hall. The goal of this neural network is to provide with a probability of finding a person in the hall. For this purpose we found a

convolutional neural network more suitable based on the literature. We approached to solve this problem with multiple architectures as discussed above. The final model we propose takes in Range-AoA Fourier transform to solve this coupled with a Convolutional Neural Network.

The second model involves counting the number of people in the concert hall. For this we can have a classification or linear regression based models. The classification limits the number of people the model can count where as the linear regression based model will be more generic for counting purposes. For classification problem we use cross entropy based loss and for linear regression model we use mean square error loss.

Cross-entropy loss is more informative than MSE loss. Cross-entropy loss is specifically designed for classification tasks where the goal is to estimate the probability distribution over the classes. In contrast, MSE loss is designed for regression tasks where the goal is to minimize the distance between the predicted and true values. The cross-entropy loss provides more information about the probability distribution over the classes, making it easier for the optimization algorithm to converge faster.

Cross-entropy loss has a logarithmic form: Cross-entropy loss has a logarithmic form, which means that it can converge faster than MSE loss, which has a quadratic form. This is because the derivative of the logarithmic function is steeper near its minimum than the derivative of the quadratic function.

IV. Implementation

A. Hardware Experimental Setup

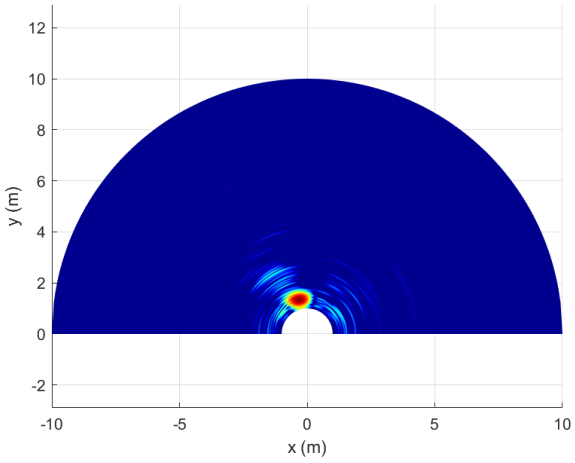


Fig. 2. one person in the room

For the physical experiment aspect of the project we first had to decide which location to use for data collection and testing. We obviously didn't have access to a concert hall and also the required amount of people to train and test the models, so we decided to start with a scaled down version of the problem by finding a conference room/classroom to test our setup in.

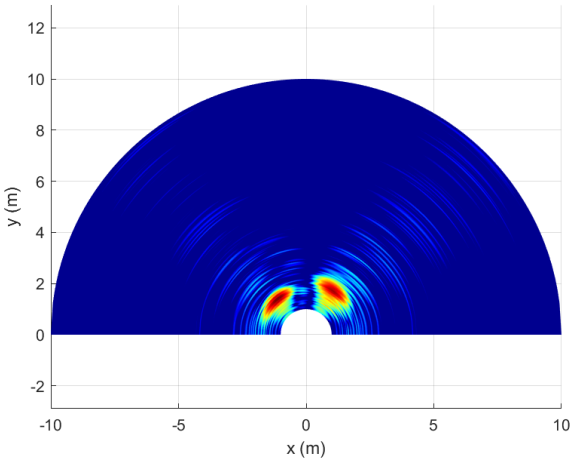


Fig. 3. two people in the room

We placed the RadarBook in the front of the room and collected data for 0, 1, 2, 3, and 4 people in the room. For example when we were collecting data for 2 people in the room, we asked each person to sit in a random location for a few minutes and then change locations and collect data again for another few minutes. This was repeated over and over again till we got 1000 ticks for each amount of people in the room.

In order to provide better results, we asked the participants to fidget a bit in their seats, which allowed us to get better radar data due to our background suppression algorithm we used.

B. Software Setup

All our machine learning models were experimented on in a Jupyter Notebook. Artifacts have been posted on github. Data loading requires the installation of scipy, plotting requires matplotlib and seaborn, and pytorch is used for creating the neural network.

V. Results

A. Radar Output

As can be seen in fig 2 we can see if there is no noise or unwanted reflections such as in indoor scenarios we should get one peak. This can be seen as a red blob in a background of blue. Similar results can be seen in fig 3 as well when there are two people in a clean environment. But as soon as we consider a realistic indoor environment like a classroom with multiple reflections we see a lot of noise in the radar output. This can be seen in fig 4

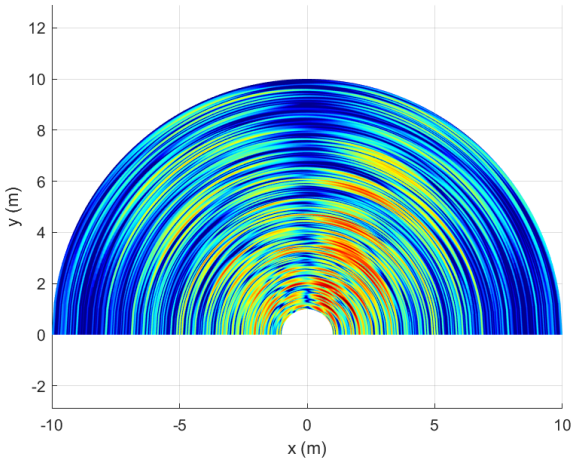


Fig. 4. Two people in an indoor environment

B. Detecting the Person

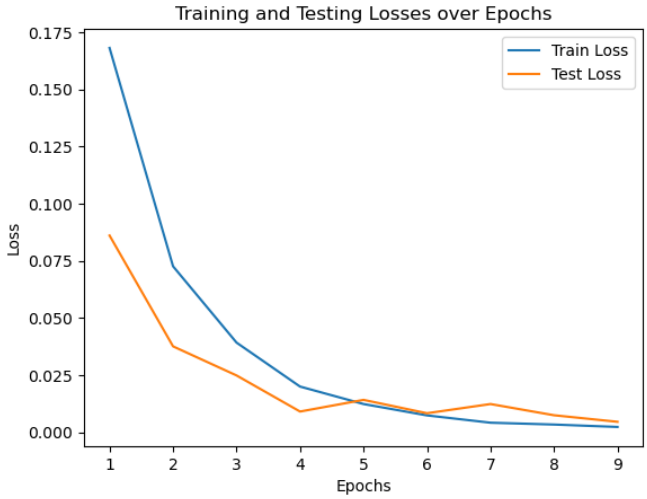


Fig. 5. Train and Test Loss over Epochs

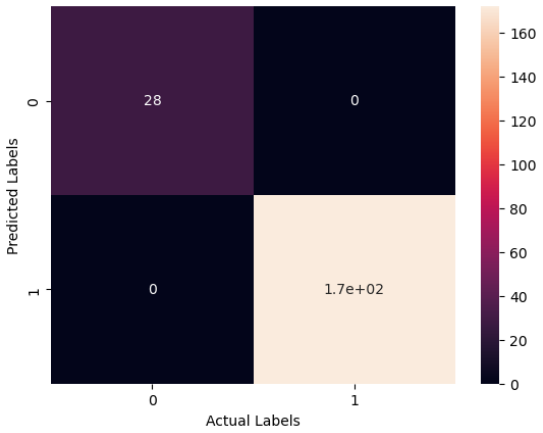


Fig. 6. Confusion Matrix For Detection Model

The CNN mentioned in the design section is used to train for the task of detecting a person. The train and test losses kept decreasing and the training was stopped at the stage where the train loss starts increasing because of overfitting. The model performed very good on the test data giving 100% accuracy. This shows CNNs are very good to solve this task.

The performance of the above model can also be visualized using a confusion matrix below.

C. Counting Number of People

The CNN with more number of parameters in each level is used to train for this task. The training and test loss curves show that there is no over fitting for this task. The confusion matrix can be visualized to see the actual performance. The true positives contribute for the correct predictions and the rest are false predictions. We see that the false predictions are most probable with the near by classes.

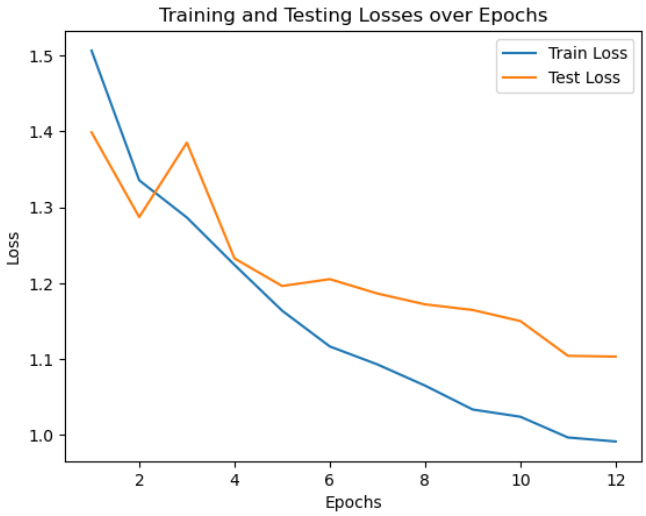


Fig. 7. Train and Test Loss over Epochs

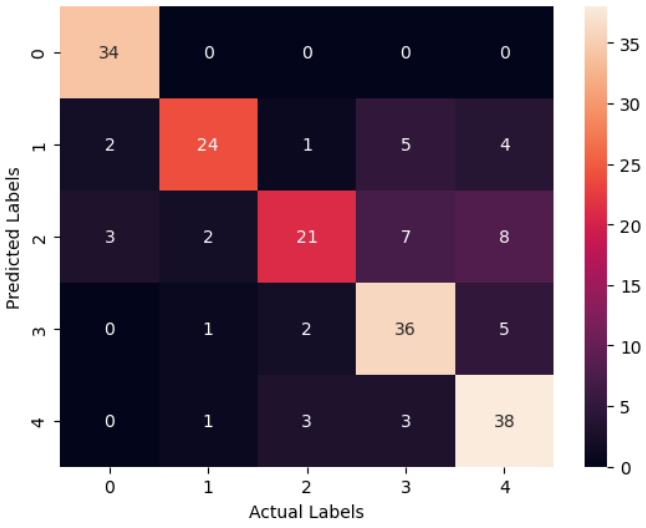


Fig. 8. Confusion Matrix For Classification Model

We also experimented with counting people as a linear regression task. Though the losses are decreasing, the model is not able to generalize the task and not performing to a good extent. All the predicted labels seems to be falling under 0-1 and hence needs to be debugged further.

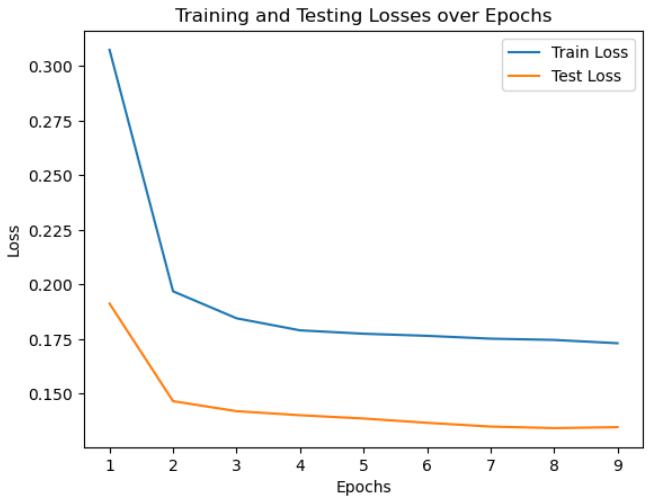


Fig. 9. Train and Test Loss over Epochs

The detection model can be helpful to improve the power efficiency of the system. The main model which can be complex will only run once the detection model triggers any person detection. So a low complex model will run all the time instead of the highly complex people counting algorithm making it less power consuming.

VI. Conclusion

We observed that when trying to use radar to gather information about people in a room, the indoor environment, with the challenges of having multiple reflections make it very hard to visually count the number of people. But using machine learning models help us find patterns in the data to predict this count with a high enough accuracy. The drawback being that we are getting this high accuracy on classification tasks and not regression task. That is to say we are able to accurately predict only the max number of people in our training set, and cannot extrapolate our small training set onto a larger number of people. This may be able to be fixed by simply increasing our training data size.

We also have a set of improvements based on our learning. We can improve the model by forcing the network to look at only certain seats, if we use a fixed seating arrangement. We can also use an LSTM network to consider time series data for more accurate prediction. Knowledge distillation, which is a machine learning technique in which a smaller model is trained to mimic the output of a larger, more complex model, can be used to increase efficiency.

In conclusion, the proposed method of using CNNs with input range-AoA transforms to count and monitor people in concert halls using mmWave radar data is promising, achieving accuracy rates of 98% for counting people and 100% for detecting their presence.

VII. Acknowledgment

- We would like to thank

References

[1] M. Guan, Y. Huang, and Y. Zhang, "Range and angle estimation of multiple targets using FMCW MIMO radar," in IEEE Transactions on Aerospace and Electronic Systems, vol. 53, no. 5, pp. 2329-2346, Oct. 2017, doi: 10.1109/TAES.2017.2693720

[2] J. Chen, J. Huang, and H. Wymeersch, "People counting using millimeter wave radar: A comparison of feature extraction and classification techniques," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, doi: 10.1109/TIM.2020.3004981

[3] J. Lee, J. Lee, and S. Jang, "Real-time people counting using millimeter-wave radar with a two-stage neural network," in IEEE Access, vol. 8, pp. 196574-196585, 2020, doi: 10.1109/ACCESS.2020.3035239

[4] Abedi, H., Luo, S., Mazumdar, V., Riad, M. M., and Shaker, G. (2021). AI-powered in-vehicle passenger monitoring using low-cost mm-wave radar. IEEE Access, 10, 18998-19012

[5] Adib, F., Kabelac, Z., Katabi, D., and Miller, R. C. (2014). 3D tracking via body radio reflections. In 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14) (pp. 317-329).

[6] Hou, Y. L., and Pang, G. K. (2010). People counting and human detection in a challenging situation. IEEE transactions on systems, man, and cybernetics-part a: systems and humans, 41(1), 24-33

[7] Yang, Junjing, Mattheos Santamouris, and Siew Eang Lee. "Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings." Energy and Buildings 121 (2016): 344-349.

[8] Abedi, Hajar, et al. "AI-powered in-vehicle passenger monitoring using low-cost mm-wave radar." IEEE Access 10 (2021): 18998-19012.