A CNN-LSTM based Blockage Prediction in Millimeter Wave Communications

Ayush Patel
Ahmedabad University
Ahmedabad, India
ayush.p3@ahduni.edu.in

Rutul Patel
Ahmedabad University
Ahmedabad, India
rutul.p@ahduni.edu.in

***Abstract*—This paper provides an innovative approach to using visual data and machine learning techniques for proactively predicting the millimeter wave (mmWave) dynamic link blockages before their occurrence. Proactive prediction helps the mmWave/sub-THz networks manage the wireless resources proactively and with enhanced reliability and latency. Using the latest Computer Vision techniques and state-of-the-art methods can improve the performance metric of blockage prediction. To evaluate this, the paper provides a CNN-LSTM network for real-time image feature extraction and time series prediction on DeepSense 6G, a real world data that contains sensing and communication data from the environment. Based on this dataset, the proposed solution is able to near the performance of the state-of-the-art methods by achieving $\approx$ 90% accuracy in predicting blockages within 0.1 seconds and $\approx$ 80% accuracy in predicting blockages within 1 second, depicting the ability of the novel approach over the methods used by base paper researchers.**

***Index Terms*—Blockage Prediction, CNN-LSTM, mmWave Communication, Computer Vision**

I. Introduction

A. Background

The landscape of wireless communication has evolved significantly, transitioning from 2G networks to the recent 5G networks, and now, setting its sights on the promising era of 6G communication systems. In this progression, the millimeter wave spectrum, operating within the high-frequency range of 30 GHz to 300 GHz, has emerged as a key player, offering unparalleled data transfer rates, reduced latency, and increased network capacity—ideal characteristics for the 6th Generation of wireless networks. However, a critical challenge in millimeter wave communication, particularly in line-of-sight (LOS) scenarios, stems from the susceptibility of signals to environmental conditions. The mmWave communication systems are pretty fragile and their usage in non-stationary users could result in very noisy communication as small objects like human beings, cars, trees, and buildings can block the mmWaves. The previous work done in this domain uses RGB camera data to obtain the bounding box coordinates using the YOLO framework and applying machine learning techniques like GRU for time series prediction, but their performance takes a heavy dip when the time frame for future instances is increased from 0.1 seconds to 1 second. Its ability is further minimized when we consider unfavourable weather conditions that work against the smooth workflow of YOLO. Addressing all these limitations, this work showcases a novel approach to creating a robust mechanism, that makes mmWave communication much reliable and latency-free.

B. Motivation

In the upcoming era of 6G networks, we expect data rates up to 1000 gigabits per second and latency to be less than 100 microseconds. The mmWaves play a pivotal role in achieving the envisioned performance. Predicting the blockage scenarios for the LOS connections is crucial for determining the success of 6G networks. In this context, the base paper methodology is extensively dependent upon the performance of the object detection tool - YOLO, and the GRU architectures for their time series prediction. This research is motivated by exploring the possibility of using the image data from RGB images and letting the Convolutional Network learn the features from the images. These features will be precise and provide more details than the YOLO architecture as in the latter case, the output is limited to the precision of bounding box coordinates whereas image features learned by CNN can be of any length, providing much more detailed information about each image of the time frame. By combining the abilities of CNN and LSTM, this CNN-LSTM-based architecture can overcome the challenges faced by the object detection approach. The strategic way of using the feature extraction prowess of CNNs and temporal memory retention of LSTM, we can now significantly enhance the accuracy of predicting the LOS blockages before they can happen in real time, not only ensuring the reliability of the system but also reducing the latency, which is a prominent step towards implementation of 6G networking in real life. In the case of non-stationary conditions, in scenarios like autonomous vehicles and high-speed transportation, the proposed research work aims to assist in the implementation of these technologies safely and efficiently. Most of the technological advancements nowadays are reliant on computer vision-based techniques like CNN, further providing support to the motivation of creating a CNN-LSTM-based approach to this problem. This method not only aims to propel theoretically but also makes the usage of mmWave practical and easy.

C. Contribution

In the pursuit of overcoming the challenges faced by the existing research on Line-Of-Sight blockage prediction, this research aims to produce better accuracy and F-1 score, more importantly in the future time instances of 0.495 seconds and 1 second, which are the cases where the existing research work underperformed.

It is important to integrate the spatial as well as temporal elements as they both provide accurate patterns and discrepancies in the image data. The feature extraction ability of CNN can only be excelled upon if the temporal features are extensively learned by an advanced time series prediction model – LSTM. This hybrid architecture capitalizes on the strengths of both CNNs and LSTMs, combining spatial and temporal learning capabilities for a more comprehensive understanding of the dynamic mmWave environment.

The proposed architecture aims to provide a comprehensive solution capable of addressing the challenges associated with extended prediction intervals in millimeter wave communication systems.

II. System Model

The work considers the scenario where both the receiver as well as transmitter are placed in a line with several obstacles coming in the LOS, leading to the disruption of the mmWave/sub-THz network connection. The mmWave base station employed in the system is mounted with an RGB camera that captures the sensing data from the environment. The base station is transmitting to a single user which is considered stationary for simplicity.

The code book used by the base station can be modelled as:

$$V = \{v_n\}_{n=1}^{N}$$

where $v_n \in \mathbb{C}^{K\times 1}$ and $N =$ No. of beamforming vectors in the codebook.OFDM transmission has been considered with K subcarriers and D length cyclic prefix. Let the vector being used be $v_n \in V$ to serve the user. The downlink signal recieved at pth subcarrier is :

$$y_p[t] = h_p^T[t] v_n x[t] + n_p[t],$$

where $h_p[t] \in \mathbb{C}^{K\times 1}$ is the channel between the basestation and the user at the k_pth subcarrier, $x[t]$ is a transmitted signal, $E\{|x[t]|\}^2 = P$, where the average transmitting power is P, and n_p is the noise received, $n_p \sim \mathcal{NC}(0, \sigma_n^2)$.

At any time instant t', the channel model for LOS Blockage can be expressed as :

$$H_p[t'] = (1 - B[t'])H_{\text{LOS},p}[t'] + H_{\text{NLOS},p}[t'],$$

where $H_{\text{LOS},p}$ and $H_{\text{NLOS},p}$ are the corresponding channel components. $B[t'] \in [0, 1]$ is used to represent whether the link is blocked or not. If $B[t'] = 1$, we have blocked link and if $B[t'] = 0$, we do not.

Note: The channel gain for LOS communication is much greater than NLOS communication in an mmWave/sub-THz network.

III. Methodology

A. Problem Formulation

The primary objective of the work is to work with the sequential data using concepts of Convolutional Neural Network and LSTM. A sequence of images is captured in the RGB camera mounted at the base station. This data is then used to predict whether the object will obstruct the LOS connection or not, which is the primary goal of the work. Let $i[t] \in \mathbb{R}^{X\times Y\times C}$ express a single image of the RGB image at time instant t, here X, Y, and C are the width, height, and the number of color channels for the captured image. We can define the sequence of length r (ie. r no.of images) at any time instant $\tau \in \mathbb{Z}$ as :

$$\text{Seq}[\tau] = \{i[t]\}_{t=\tau-r+1}^{\tau},$$

Our work aims to process this sequence of captured data $\text{Seq}[\tau]$ at a time instant τ and predict the probability of its LOS blockage within a window of r_0 future instances. For any sequence $\text{Seq}[\tau]$ and future window r_0, the state of blockage can be represented as

$$s[\tau] = \begin{cases} 1, & B[t] = 1, t \in \{\tau + 1, \ldots, \tau + r_0\}, \\ 0, & \text{else}, \end{cases}$$

Here, 0 represents that there is no blockage in the r_0 future instances and 1 represents that the LOS connection is going to be disrupted and a proactive response is required.

B. Analysis

The proposed scheme is a hybrid approach to of a CNN and LSTM[2]. In which CNN extracts the features of the input image sequence and the sequence is fed to the LSTM which predicts the blockage status. This hybrid model was trained on the loss provided by the LSTM.

The combination of CNN and LSTM offers distinct advantages, especially when dealing with issues such as obstructions and dynamic shifts in a scene. The Long Short-Term Memory (LSTM) model's intrinsic capacity to learn and forget information over time helps it navigate through obscured frames, which improves the model's understanding of scene dynamics.

YOLO is mainly intended for detection of objects and might not be able to capture subtle semantic information. In contrast, the CNN-LSTM hybrid model is very good at using rich spatial information from convolutional layers. As a result, the model can make more precise forecasts.

Moreover, the weights of CNN architecture are being updated along with the LSTM architecture based on the LSTM output, which enables the CNN to learn according to the temporal prediction as well. The bi-directional nature of information traansfer is one of the main advantages of this netowrk over existing work. This makes model much more flexible to learn as compared to the object detection approach using YOLO. Even the long-term dependencies can now be retained by using LSTM over GRU making the prediction process more robust and reliable.

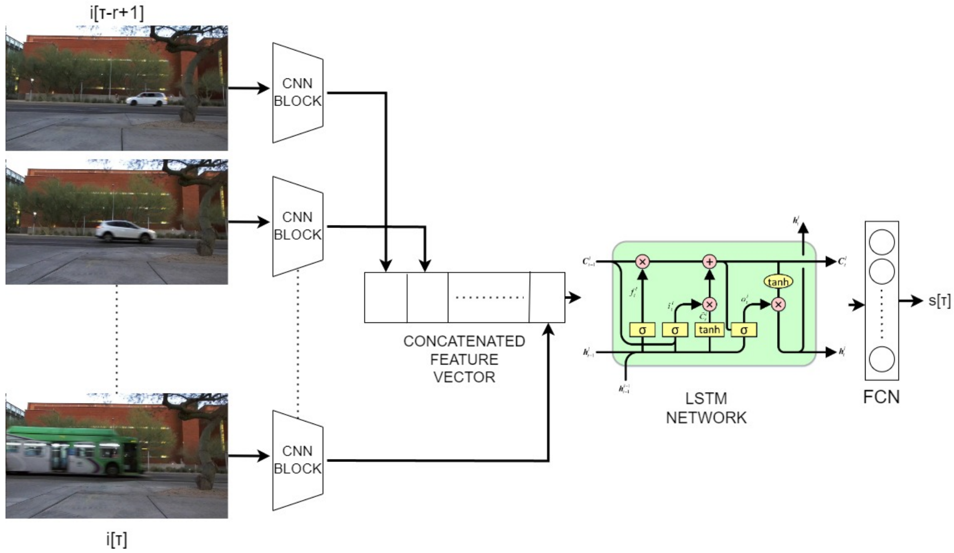


Fig. 1: Proposed Architecture - A hybrid CNN-LSTM architecture that is efficient to capture spatio-temporal features of the images

IV. Experiments

A. Dataset

The multi-modal DeepSense 6G dataset was created with the objective of researching and assessing various aspects of sensing and communication in a dynamic setting. The dataset is multi-modal, including RGB images, millimeter wave (mmWave) sensor data, and GPS information. The data collection setup used two stationary units positioned on opposite sides of a two-way street.

Unit-1, located on one side of the street, is equipped with a mmWave receiver, an RGB camera, and a GPS receiver. Unit-2, positioned on the opposite side of the street, is equipped with a mmWave transmitter and a GPS receiver.

For the purpose of blockage prediction, the dataset is organized into temporal sequences so that each sample contains a series of 8 RGB images (temporally dependent), and the ground truth blockage after t time instances (each instance on average 95 ms) is represented by 0 or 1. To facilitate research and evaluation, the dataset is partitioned into three distinct sets: 70% for training, 20% for validation, and 10% for testing.

B. Simulations

Simulations were carried out to evaluate the performance of the proposed CNN-LSTM architecture for predicting blockages in mmWave communication systems. The simulations were implemented using a deep learning framework (PyTorch).

1) Simulation Parameters: The CNN-LSTM architecture was configured with the following parameters:

- Number of CNN layers: $n_{\text{CNN}} = 7$
- Number of LSTM layers: $n_{\text{LSTM}} = 1$
- CNN filter size: 3×3
- LSTM input sequence size: 8

2) Training Details: The model was trained using the DeepSense 6G dataset, consisting of multi-modal sensing and communication data. The training process involved the following details:

- **Number of Epochs:** 25
- **Batch Size:** 64
- **Optimizer:** Adam
- **Learning Rate:** 0.001

C. Performance Evaluation

In the conducted experimental setup, we focus on parameters set for the neural network and the evaluation metrics used to assess the efficacy of the proposed CNN-LSTM model for LOS blockage prediction. As elaborated in the code documentation, the model architecture integrates eight parallel CNN layers, each comprising convolutional and batch normalization operations, culminating in a fully connected layer. Subsequently, the CNN features are concatenated and given as an input into an LSTM layer with hidden states, followed by a final linear layer for prediction.

The training process involves utilizing an Adam optimizer with a specified learning rate and a maximum gradient norm to mitigate the risk of exploding gradients. The binary cross-entropy loss function is employed during the training of the model, aiming to optimize its parameters. The training loop iterates through multiple epochs, and the model's performance is evaluated using metrics such as accuracy and F1 score.

The CNN-LSTM model is implemented in PyTorch and trained on a computing environment featuring a GPU P100 provided on kaggle platform. The training data, derived from a labeled development dataset, consists of inputs representing eight parallel streams of data, mirroring the eight CNN layers in the architecture. The chosen batch size is 64, providing a

Fig. 2: Examples of the RGB images captured by the base station camera - Each image is originally of size 960x540 px

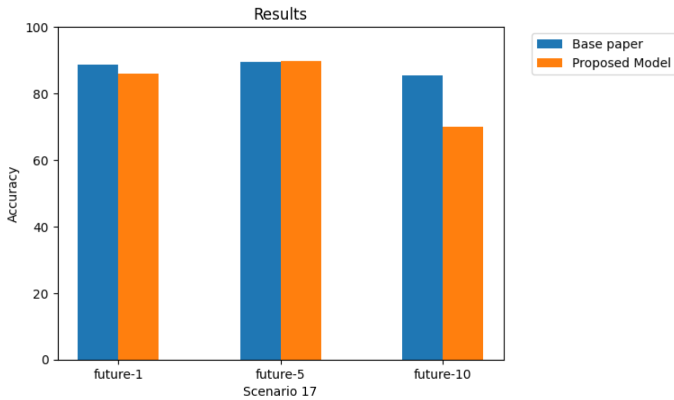


Fig. 3: Comparison of Prediction Accuracy

The graph compares the prediction accuracy of our proposed model and the base paper across future-1, future-5, and future-10 prediction windows.

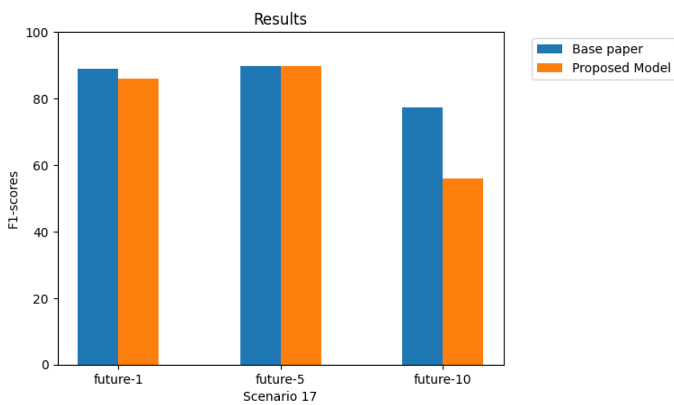


Fig. 4: Comparison of F-1 Scores

This graph illustrates a comparative analysis of F1 scores between our proposed model and the baseline model across future-1, future-5, and future-10 prediction windows.

balance between computational efficiency and model update frequency.

To delve into the model's architectural details, the CNN layers seek to capture spatial features, while the subsequent LSTM layer processes these features sequentially to capture temporal dependencies. This hybrid approach is particularly

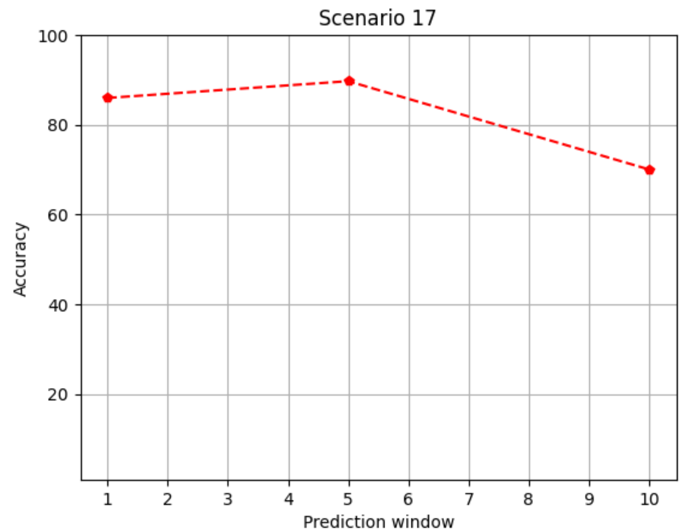


Fig. 5: Nature of Accuracy Values Over Increasing Future Instances

The figure demonstrates the accuracy comparison in different prediction windows. Accuracy increases from future-1 to future-5 but decreases from 5 to future-10. Predicting LOS blockages before a time window of 1 second is challenging, leading to a significant drop in accuracy for future-10.

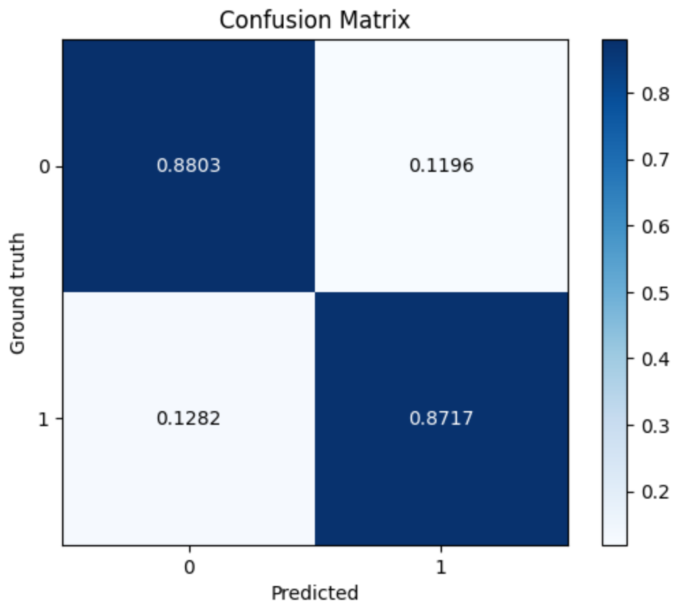
precise for predicting LOS blockage in scenarios where both spatial and temporal information play crucial roles.

In terms of evaluation, the primary metric is the accuracy, offering a fundamental measure of the model's correctness in predicting LOS blockage. Additionally, the F1 score is employed to gauge the model's robustness, providing insights into its precision and recall performance.

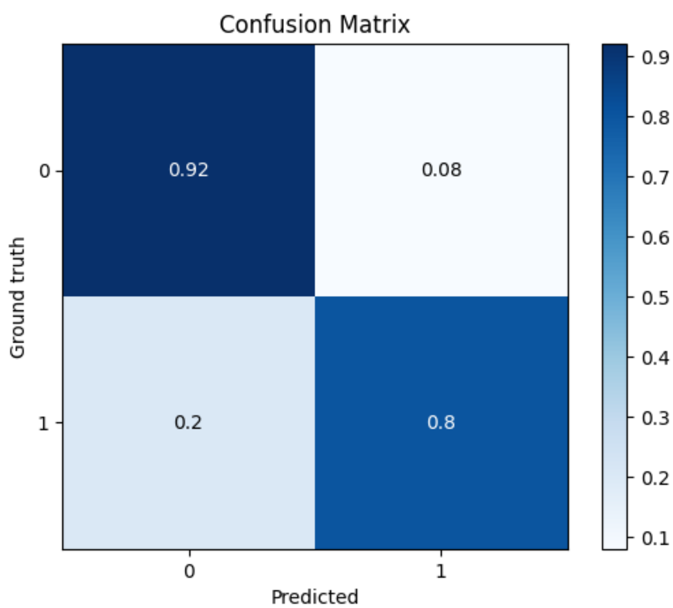
The proposed methadology although does not beat the outcomes provided in the base paper, but further improvements mentioned in the forecoming section can out perform the evaluation metric for the proposed architecture.

V. Conclusion

This paper proposes a novel approach to address the problem of blockage prediction in millimeter wave deployments using computer vision techniques. We determined the system model, formulated the problem definition, and proposed a



(a) Future-1



(b) Future-5

Fig. 6: This figure shows the confusion matrices for future-1 and future-10 blockage prediction intervals.

computer vision-based machine-learning approach to predict the Line-of-Sight (LOS) blockages before 0.1, 0.495, and 1 second. For this, we used a real-world, multi-modal dataset, DeepSense 6G.

The proposed scheme, though could not outperform the evaluation metric, did display its potential by nearing the accuracies for the three future instances. It was able to achieve an accuracy of $\approx 90\%$ for the shorter prediction time frame and $\approx 80\%$ for the longer prediction time frame (up to 1 second). Thus, the efficient usage of visual data does contain a decent scope of real-world implementation for the mmWave/sub-THz communication systems.

VI. Future Work

Since the proposed architecture was able to achieve similar outcomes with a different approach, further specifications in this method can make this work outperform the existing research work on this dataset. Further usage of self-attention-based Convolutional Neural Networks can be used to target specific portions of the image data that provide crucial information about spatiotemporal features, leading to enhanced blockage prediction accuracy as well as F-1 score. Further, since the DeepSense 6G dataset is a multi-modal dataset, providing Lidar as well as mmWave power data can be used simultaneously to create a more robust, multi-modal prediction mechanism. This method can also be applied to address the scenario of a non-stationary user and base station or a mobile receiver and transmitter.

References

[1] G. Charan and A. Alkhateeb, ”Computer Vision Aided Blockage Prediction in Real-World Millimeter Wave Deployments,” 2022 IEEE Globecom Workshops (GC Wkshps), Rio de Janeiro, Brazil, 2022, pp. 1711-1716, doi: 10.1109/GCWkshps56602.2022.10008524.

[2] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 15, 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[3] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, and N. Srinivas, “DeepSense 6G: A large-scale real-world multi-modal sensing and communication dataset,” available on arXiv, 2022. [Online]. Available: https://www.DeepSense6G.net

[4] S. Wu, M. Alrabeiah, C. Chakrabarti, and A. Alkhateeb, “Blockage prediction using wireless signatures: Deep learning enables real-world demonstration,” arXiv preprint, arXiv:2111.08242, 2021.