Computer Vision Aided Blockage Prediction in Real-World Millimeter Wave Deployments

Gouranga Charan and Ahmed Alkhateeb
School of Electrical, Computer and Energy Engineering - Arizona State University
Emails: {gcharan, alkhateeb}@asu.edu

***Abstract*—This paper provides the first real-world evaluation of using visual (RGB camera) data and machine learning for proactively predicting millimeter wave (mmWave) dynamic link blockages before they happen. Proactively predicting line-of-sight (LOS) link blockages enables mmWave/sub-THz networks to make proactive network management decisions, such as proactive beam switching and hand-off) before a link failure happens. This can significantly enhance the network reliability and latency while efficiently utilizing the wireless resources. To evaluate this gain in reality, this paper (i) develops a computer vision based solution that processes the visual data captured by a camera installed at the infrastructure node and (ii) studies the feasibility of the proposed solution based on the large-scale real-world dataset, DeepSense 6G, that comprises multi-modal sensing and communication data. Based on the adopted real-world dataset, the developed solution achieves $\approx 90\%$ accuracy in predicting blockages happening within the future 0.1s and $\approx 80\%$ for blockages happening within 1s, which highlights a promising solution for mmWave/sub-THz communication networks.**

***Index Terms*—computer vision, deep learning, blockage prediction, mmWave, terahertz.**

I. Introduction

Millimeter wave (mmWave) and sub-terahertz communication systems rely on line-of-sight (LOS) links to achieve sufficient receive signal power. Blocking these LOS links by the moving objects in the environment may disconnect the communication session or cause sudden and significant degradation in the link quality. This is due to the high penetration loss of the mmWave/sub-terahertz signals and the much less receive power of the NLOS links compared to the LOS ones [1], [2]. All that highly challenges the reliability and latency of the mmWave/sub-terahertz communication networks. Initial approaches for overcoming these blockage challenges relied mainly on multi-connectivity [3], [4]. These solutions, however, generally keep the user connected to multiple infrastructure nodes which under utilizes the wireless network resources. This motivated the research for more efficient blockage avoidance approaches.

Leveraging machine learning (ML) to address the blockage challenges has gained increasing interest in the last few years [5]–[7]. In [5], the authors proposed to leverage recurrent neural networks to process the sequence of beams serving a mobile user and to predict whether or not a future blockage will happen. Relying only on beam sequences, however, limits the applications to stationary blockage prediction. Predicting dynamic blockages require more information about these moving blockages in the environment. In [6], [7], in-band mmWave and sub-6GHz based wireless scattering signatures were used to identify/predict the incoming mmWave link blockages. These solutions, however, are mainly capable of predicting immediate blockages and are hard to scale to complex/crowded scenarios. To enable predicting blockages early enough before they block the links, solutions based on radar and LiDAR sensory data were proposed in [8], [9]. Despite their promising results, each sensing modality has its advantages and drawbacks. For example, radar data is mainly suitable for uncrowded scenarios and LiDAR sensors are expensive and have relatively short range.

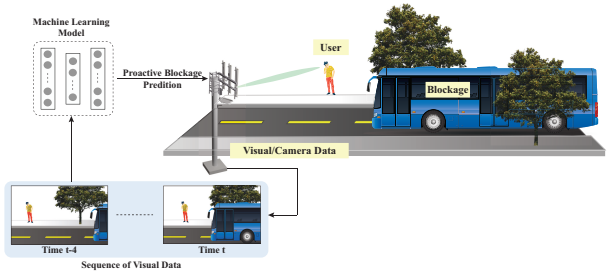


Fig. 1. This figure illustrates the overall system model where a mmWave/THz basestation utilizes the captured RGB images to predict the future link blockage status.

In [10], we proposed to leverage visual data (captured by cameras) to predict future dynamic blockages. These solutions and analysis in [10], though, were based only on synthetic datasets, and an important question that arises is whether the promising results in [10] can be achieved in reality? In this paper, we attempt to answer this question. In particular, the main contributions of the paper can be summarized as follows:

- Formulating the vision-aided blockage prediction problem in mmWave/THz wireless networks considering practical visual and communication models.
- Developing a machine learning approach that is capable of (i) pre-processing the real-world visual data to enhance the blockage prediction performance, (ii) extracting the relevant features about the scatterers/environment, and (iii) efficiently predicting future dynamic link blockages.
- Providing the first real-world evaluation of vision-aided blockage prediction based on our large-scale dataset, DeepSense 6G [11], that consists of co-existing multi-modal sensing and wireless communication data.

Based on the adopted real-world dataset, the developed so-

978-1-6654-5975-4/22 © 2022 IEEE

lution achieves $\approx 90\%$ accuracy in predicting blockages happening within a future prediction interval of 0.1s and $\approx 80\%$ for a prediction interval of 1s. This highlights the potential of leveraging machine learning and visual data in addressing the critical LOS link blockage challenges.

II. System Model

This work considers a communication scenario where a mmWave basestation is serving a stationary user located in a busy environment with multiple moving objects, such as vehicles, pedestrians, etc., as shown in Fig. 1. The mmWave basestation is equipped with an RGB camera to monitor and gather sensing data about the surrounding environment. This information could potentially be leveraged to proactively predict future link blockages caused by the moving objects.

The adopted system model consists of a mmWave basestation equipped with an N-element antenna array and a standard-resolution RGB camera. The basestation is serving a stationary user that is, for simplicity, considered to have a single antenna. The basestation uses a pre-defined beam codebook $\boldsymbol{\mathcal{F}} = \{\mathbf{f}_m\}_{m=1}^{M}$ to serve the user, where $\mathbf{f}_m \in \mathbb{C}^{N\times 1}$ and M is the total number of beamforming vectors in the codebook. As will be described in Section V, the beamforming codebook adopted by the hardware prototype has 64 beamforming vectors (i.e., $M = 64$) with the azimuth angles uniformly quantized between $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$. The communication system further adopts OFDM transmission with K subcarriers and cyclic prefix of length D. At any time instant t, if the basestation uses the beamforming vector $\mathbf{f}_m \in \boldsymbol{\mathcal{F}}$ to serve the user, then the downlink received signal at the user at the kth subcarrier can be expressed as

$$y_k[t] = \mathbf{h}_k^T[t]\mathbf{f}_m x[t] + n_k[t], \tag{1}$$

where $\mathbf{h}_k[t] \in \mathbb{C}^{N\times 1}$ is the channel between the basestation and the user at the kth subcarrier, $x[t]$ is a transmitted data symbol, $\mathbb{E}|x[t]|^2 = P$, with the average transmit power P, and n_k is a receive noise sample, $n_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2)$.

LOS Blockage: The channel model $\mathbf{h}_k$, defined in (1), is generic and can be expressed as follows at time instant t

$$\mathbf{h}_k[t] = (1 - b[t])\, \mathbf{h}_k^{\text{LOS}}[t] + \mathbf{h}_k^{\text{NLOS}}[t], \tag{2}$$

where $\mathbf{h}_k^{\text{LOS}}$ and $\mathbf{h}_k^{\text{NLOS}}$ are the LOS and NLOS channel components. The binary variable $b[t] \in [0, 1]$ represents the link status at time instant t, with $b[t] = 1$ indicating that the LOS path is blocked and $b[t] = 0$ otherwise.

It is important to note here that for mmWave and sub-THz communication systems, the LOS channel gain is much greater than the NLOS channel gain [1], [2]. Therefore, LOS link blockages challenge the reliability of these networks. Next, we provide a formal definition of the proactive vision-aided blockage prediction problem which is the focus of this work.

III. Vision-Aided Blockage Prediction: Key Idea and Problem Formulation

One of the major challenges in the high-frequency wireless communication networks is the LOS link blockages; the mmWave/THz communication systems suffer from link disconnection and significant dips in the received SNR when an object/blockage intersects the LOS path between the basestation and the user. Re-establishing a LOS connection is usually done in a reactive way, which incurs critical latency and impacts the reliability of such systems. The presence of dynamic moving objects in the environment further increases these reliability/latency challenges. These challenges could potentially be addressed if these blockages can be proactively predicted [8]–[10]. In order to develop an efficient solution that can *proactively* predict the occurrence of such future blockages, it is essential to equip the wireless network with a sense of its surroundings. This work attempts to do so by utilizing machine learning and visual data captured by cameras placed at the basestation to proactively predict future blockages before they happen. In this section, we will first present the key idea in Section III-A and then formulate the vision-aided blockage prediction problem in Section III-B.

A. The Key Idea

In a wireless network, the link blockages are often caused by moving objects, such as cars, trucks, buses, and humans, present in the wireless environment. Given the dynamic nature of multiple moving objects in a real-wireless scenario, the task of future blockage prediction becomes extremely challenging. While the detection of the different objects in the environment, such as cars and humans, can be well achieved using a single image, the success of the dynamic blockage prediction task also relies on characterizing the mobility patterns and geometric features of these objects. For example, for a vehicle-to-infrastructure use case, the visual data at the infrastructure needs to characterize the speed/direction of travel and the size of the various objects in the environment. In order to capture these additional indicators, which are normally obtained by analyzing a *sequence* of images, **our proposed solution observes a sequence of r image samples instead of making the predictions based on just one sample and attempts to predict future LOS link blockages before they happen**. Building upon the key idea presented here, in the next subsection, we provide the formal definitions for the proactive vision-aided blockage prediction problem.

B. Problem Formulation

The main objective of this work is to observe a sequence of camera image samples captured at the basestation and utilize the sensing data to predict whether or not the stationary user will be blocked within a window of future instances. Let $\mathbf{X}[t] \in \mathbb{R}^{W\times H\times C}$ denote a single RGB image of the environment captured at the basestation at time instant t, where W, H, and C are the width, height, and the number of color channels for the image. At any time instant $\tau \in \mathbb{Z}$, the basestation uses a sequence of RGB images, $\mathbf{S}[\tau]$, defined as

$$\mathbf{S}[\tau] = \{\mathbf{X}[t]\}_{t=\tau-r+1}^{\tau}, \tag{3}$$

where $r \in \mathbb{Z}$ is the length of the input sequence or the observation window to predict future link blockages. In particular,

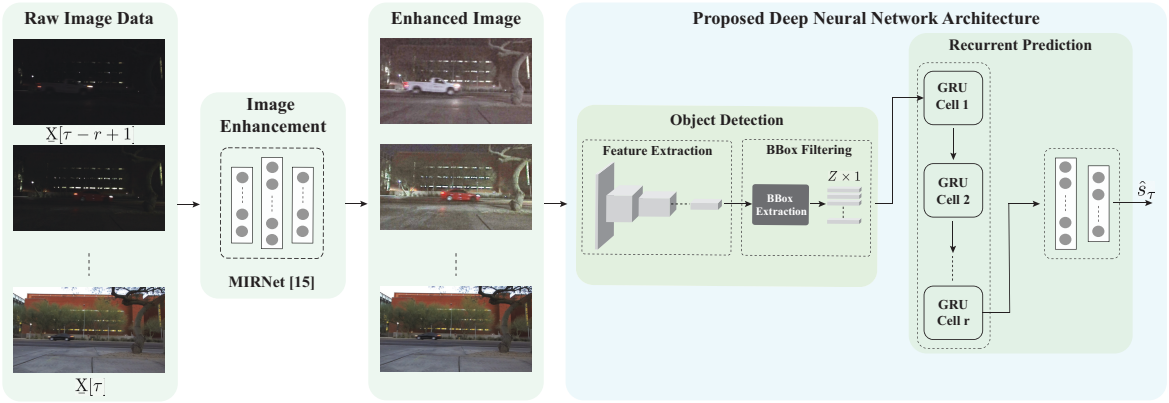


Fig. 2. This figure presents the proposed architecture for vision-aided future blockage prediction. The raw visual data is first processed to enhance the images captured during night or low-light scenario. The processed sequence of images are then passed through the proposed deep neural network architecture consisting of a object detection model and recurrent neural network to predict the future blockage.

at any given time instant τ, the goal in this work is for the basestation to observe $\mathbf{S}[\tau]$ and predict whether or not the stationary user is going to be get blocked within a window of r' future instances. It is important to note here that we do not focus on the exact future instance but consider the entire future window sequence for denoting the future blockage status. Given $\mathbf{S}[\tau]$ and the future window r', the future blockage status at time instant τ can then be expressed as

$$s[\tau] = \begin{cases} 1, & b[t] = 1,\ t \in \{\tau + 1, \ldots, \tau + r'\} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where 0 indicates that the user remains LOS within the next r' future instances and 1 points towards the occurrence of blockage within the same window.

In order to predict the future blockage status, we define a function f_Θ that maps the observed sequence of images, $\mathbf{S}[\tau]$ to a prediction (estimate) of the future blockage status, $\hat{s}_\tau$. The function f_Θ can be formally expressed as

$$f_\Theta : \mathbf{S}[\tau] \rightarrow \hat{s}[\tau]. \tag{5}$$

In this work, we adopt a machine learning model to learn this prediction function f_Θ, that takes in the observed image sequence and predicts the future blockage status, $\hat{s}[\tau] \in \{0, 1\}$. Here, Θ represents the parameters of the machine learning model and is learned from a dataset of labeled sequences. For this, a dataset of independent sample pairs $\mathcal{D} = \{(\mathbf{S}_v, s_v)\}_{v=1}^{V}$ is collected, where s_v is the ground-truth future blockage label for the observed sequence $\mathbf{S}_v$, and V is the total number of *sequence-label* pairs in the dataset. The labeled dataset $\mathcal{D}$ is then used to optimize the prediction function f_Θ such that it maintains high fidelity for any samples drawn from this dataset. The optimization problem can be written as

$$f^\star_{\Theta^\star} = \underset{f_\Theta(.)}{\text{argmax}} \quad \prod_{v=1}^{V} \mathbb{P}(\hat{s}_v = s_v | \mathbf{S}_v), \tag{6}$$

where the joint probability in (6) is factored out to convey the identical and independent (i.i.d.) nature of the samples in dataset $\mathcal{D}$. In the next section, we present the proposed deep learning-based solution for the vision-aided future blockage prediction task.

IV. Vision-Aided Blockage Prediction: A Deep Learning Solution

Guided by the principles mentioned in Section III-A, the blockage prediction task is divided into two sub-tasks: (i) object detection and (ii) recurrent prediction. The first sub-task deals with detecting the relevant objects of interest in the FoV of the basestation. The objective of the second stage needs is to predict the future blockages based on the features extracted from the first stage. In Fig. 2, we illustrate the proposed deep learning-based blockage prediction solution. In this section, we first present the details of the image enhancement pre-processing stage adopted to deal with the low-light/dark images. Then, we take a deeper dive into the developed two-stage vision-aided blockage prediction solution.

A. Data Processing (Image Enhancement)

Compared to LiDAR, radars and other sensing modalities, RGB cameras provide a low-cost, high-resolution, and low-footprint alternative, making it one of the preferred choices for wireless sensing applications. However, there is major bottleneck associated with the visual images captured using an RGB camera. Under low-light conditions, the visual data turns out to be noisy and dark, making it unsuitable for further computer vision tasks. Fig. 2 shows an image captured under such low light conditions. The white truck in the first image and the red sedan in the second are hardly visible highlighting the challenges associated with such images. In order to develop a robust and reliable solution that can work in most of the natural lighting conditions, it is essential to perform some sort of image enhancement to extract the hidden details and make the low-light images more usable. For the post-processing stage, we adopt the state-of-the-art MIRNet [12] model developed for low-light image enhancement. It is a fully-convolutional architecture that learns an enriched set of features by combining contextual information from multiple scales, while simultaneously preserving the high-resolution spatial details. As shown in Fig. 2, that the objects in the low-light images are clearly visible after the image enhancement

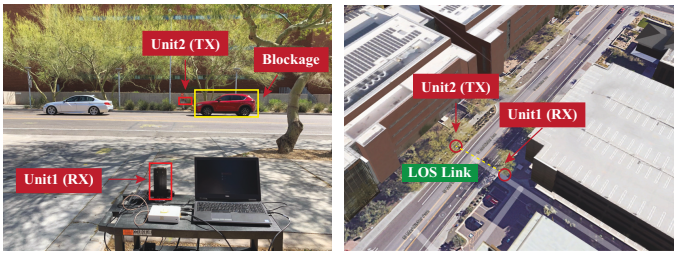


Fig. 3. This figure shows the data collection setup used for DeepSense 6G Scenarios 17-22. The figure on the left depicts the street view from Unit 1 perspective. It also highlights the location of the transmitter (Unit 2 TX) and the receiver (Unit 1 RX) during the data collection process. The figure on the right shows the exact location of Unit 1 and Unit 2 highlighted on the Google Earth 3D view.

post-processing step, which is important for the performance of the proposed blockage prediction solution.

B. A Two-Stage Deep Learning Model

Here, we present the details of the proposed blockage prediction architecture, which consists of two key functions, namely object detection and recurrent prediction.

Object Detection: The first stage of the proposed solution is the object detection deep learning model. There are two primary goals of this stage: (i) Perform accurate and quick detection of the objects of interest in the FoV of the basestation and (ii) extract the coordinates of the bounding boxes placed around the relevant objects. For this, in our proposed solution, we adopt the state-of-the-art COCO pre-trained YOLO object detection model. For each image sample, the pre-trained YOLOv3 is used to detect the relevant objects and extract the bounding box coordinates of the detected objects. In particular, for each detected object in the image, we extract a 4-dimensional vector consisting of the bottom-left coordinates $[x_1, y_1]$ and the top-right coordinates $[x_2, y_2]$. These coordinates are normalized to be between $[0, 1]$. In order to account for multiple detected objects in the FoV of the basestation, the extracted bounding boxes are concatenated to form one dimensional vector $\mathbf{d} \in \mathbb{R}^{4Y\times 1}$, where Y is the number of objects detected by the YOLOv3 model. It is important to highlight here that the number of detected objects might not be the same in each data sample, which results in a variable length vector $\mathbf{d}$. This will lead to inconsistency in the size of the extracted features and create unnecessary complications for the next stage of the proposed solution pipeline, i.e., the recurrent predictions. In order to avoid this inconsistency, the extracted bounding box vector $\mathbf{d}$ is further padded with $Z - Y$ zeros to obtain a fixed size vector $\tilde{\mathbf{d}} \in \mathbb{R}^{Z\times 1}$. The fixed size bounding box feature vector $\tilde{\mathbf{d}}$ is then provided as an input to the recurrent network to predict the future link blockage status.

Recurrent Prediction: Convolutional neural networks inherently fail in capturing sequential dependencies in input data. In order to learn the inherent relation in a sequence of input data, the final stage of the proposed solution utilizes recurrent neural network (RNN) to make the final prediction. In this work, we consider a two-stage Gated Recurrent Unit (GRU) separated by a dropout layer. These two layers are followed by a fully-connected layer that acts as a classifier. More specifically, the model receives a sequence of r extracted bounding box feature vectors, $\{\tilde{\mathbf{d}}[\tau-r+1, \ldots, \tilde{\mathbf{d}}[\tau]]\}$, as input and predicts the future link blockage status over a window of r' time instance.

V. Testbed Description and Development Dataset

In order to evaluate the performance of the proposed vision-aided blockage prediction solution, we adopt multiple scenarios from the DeepSense 6G [11] dataset. DeepSense 6G is a real-world multi-modal dataset enabling sensing-aided wireless communication applications. It contains co-existing multi-modal data such as vision, mmWave wireless communication, GPS data, LiDAR, and Radar, collected in realistic wireless environments. In this section, we present a brief overview of the scenarios adopted from the DeepSense 6G dataset followed by the analysis of the final development dataset utilized in the blockage prediction task.

DeepSense 6G: [Scenarios 17 - 22] We adopt Scenarios 17-22 of the DeepSense 6G dataset for evaluating the efficacy of our proposed solution. The hardware testbed and the exact location used for collecting these data is shown in Fig. 3. The DeepSense testbed 3 is utilized for this data collection and is placed on the opposite sides of a 2-way street with a passing-lane in-between. The primary components of the adopted testbed are: (i) A stationary 60 GHz omni-directional mmWave transmitter (unit2), (ii) a directional mmWave receiver (unit1), (iii) an RGB camera. The receiver employs a 16-element ($N = 16$) 60 GHz phased array and it receives the transmitted signal using an over-sampled beam codebook of 64 pre-defined beams ($M = 64$). Unit 1 is also equipped with a camera and it captures RGB images for the wireless environment in the FoV of the receiver. The data capture rate varies from 12 samples/sec for scenarios $17 - 19$ to 6.5 samples/sec for scenarios $20 - 22$. The weighted average for all the scenarios combined is 10 samples/sec. Each data sample consists of an RGB image of the environment and a 64-element mmWave receive power vector. For more information regarding the data collected testbed and setup, please refer to [6], [11].

DeepSense 6G: [Development Dataset] The adopted DeepSense scenarios include diverse data collected during different times of the day (day and night). Each row in the dataset scenarios consists of a tuple of an RGB image, $\mathbf{X}[\tau]$, and the corresponding receive power vector and the ground-truth link blockage status, $s[\tau]$ (manually labeled). To form the development dataset of the blockage prediction task described in Section III-B, the offered DeepSense data is further processed using a sliding window to generate a time-series dataset consisting of 8 input image samples ($r = 8$) and the corresponding future blockage status in a future window of r' samples (we generate 10 such time-series datasets at $r' = 1, 2, ..., 10$. In order to perform an in-depth study, the development dataset per scenario was further processed to generate datasets for different future prediction window size. For example, *future-1* dataset, consists of data sequences

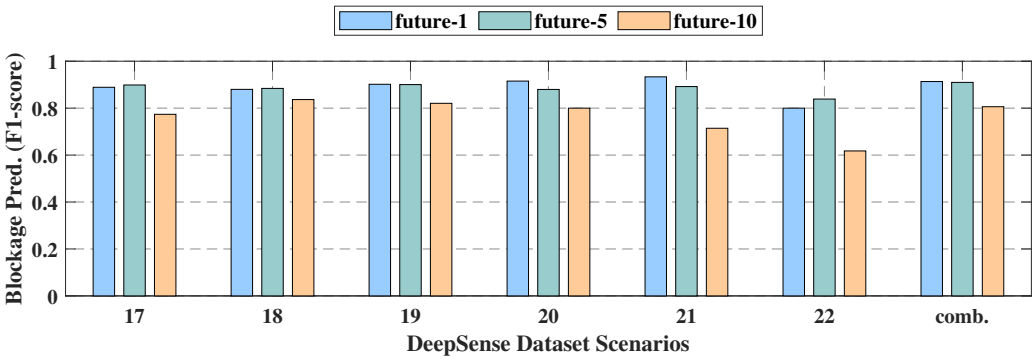


Fig. 4. This figure plots the future-1, future-5, and future-10 future blockage prediction scores (f1-score) for scenarios 17-22 and the combined scenarios. The combined scenario achieves comparable or better prediction accuracy highlighting the gain of having sufficient diversity in the dataset.

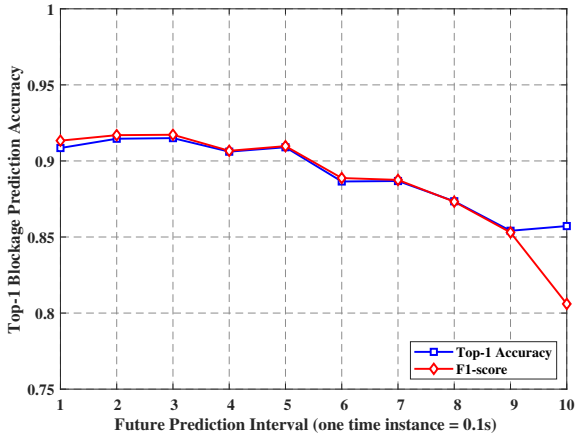


Fig. 5. Top-1 blockage prediction accuracy and F1-score for different future prediction intervals observed in the combined case. It is observed that both the top-1 accuracy and f1-score decreases as we predict further into the future.

where the input sequence length is still 8, but the future prediction window length is 1. During these process of generating datasets with different future prediction window, we ensure that the dataset is balanced {Number of LOS data sequences $\approx$ Number of NLOS sequences.}. Each of the development dataset is further divided into training, validation and test sets following a split of $70-20-10\%$.

VI. Performance Evaluation

In this section, we first discuss the neural network training parameters and the adopted evaluation metrics. Next, we present the numerical evaluation of the proposed solution.

Experimental Setup: As described in Section IV, in this work, the extracted bounding boxes(output of the YOLOv3) are provided as an input to the recurrent neural network proposed earlier in Section IV. The designed RNN has an embedding dimension (Z) of 30 and the hidden state dimension of 128. The GRU model is trained using the labeled development dataset discussed in Section V using a cross-entropy loss function. The model is trained with an ADAM optimizer with an initial learning rate of $1 \times 10^{-}3$ and a batch size of 128. All the simulations were performed on a single NVIDIA Quadro 6000 GPU using the PyTorch deep learning framework. We utilize the top-1 accuracy metric as the primary method of evaluating the proposed solution. In order to study the robustness of the proposed solution, we also utilize the F1-score metric.

TABLE I
Impact of speed on blockage prediction accuracy

scenario	vehicle speed		num of samples	
	slow-moving	fast-moving	slow-moving	fast-moving
17	**0.8863**	0.8571	88	21
18	0.8230	**1.0**	130	21
19	0.8983	**1.0**	295	52
20	0.7959	**0.9655**	147	29

Can visual data predict LOS blockages? For each of the scenarios 17-22, we evaluate the proposed solution for various future prediction window lengths. In Fig. 4, we show the F1-score of the future-1, future-5 and future-10 blockage prediction. Note that 'future-5' here represents the development dataset that considers a future blockage prediction window of length 5 time instances (i.e., predicting a blockage that will happen in the future 450ms in scenarios 17-19 and 750ms in scenarios 20-22). In Fig. 4, we observe that for all the scenarios, the proposed two-staged solution achieves an accuracy of $\approx 0.88-0.90$ future-1 and future-5 blockage prediction F1-score, highlighting the high efficiency of the proposed vision-aided blockage prediction approach. It is observed that there is a slight degradation in the model's performance for scenario 22, which could be attributed to the lower number of samples in the development dataset of this scenario; the lower number of training samples can often lead to under-fitting and impede the model's capability to learn efficiently.

What is the impact of vehicle speed? Given the dynamic nature of the dataset, the vehicles travel at different speeds. Therefore, it is important to consider the impact of vehicle speed on blockage prediction accuracy. To estimate the vehicle speed, we compute the difference between the initial and final position in the input sequence and then divide it by the number of data samples the vehicle is present. We further divide the sequences into two buckets based on the mean and standard deviation of the speeds: (i) slow-moving vehicles and (ii) fast-moving vehicles. In Table I, we present the future-5 blockage prediction accuracy versus the vehicle speed for

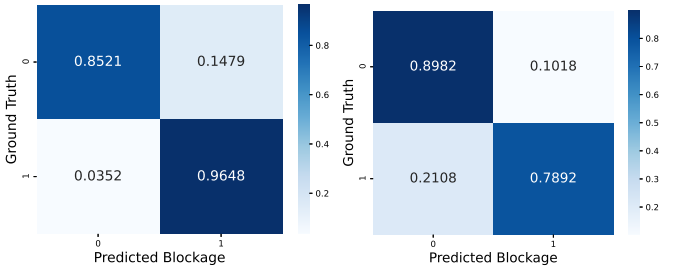
scenarios $17 - 20$. It is observed that for these scenarios, the prediction is more accurate for faster-moving vehicles. This higher prediction performance can be partially attributed to the imbalance in the dataset, highlighting the need for a balanced dataset. However, the performance achieved on the available dataset indicates the model's ability to predict future blockages for dynamic users with very high confidence.

What is the gain of combining the datasets? To evaluate that, we constructed a *combined* dataset by combining the training, validation, and test sequences of the individual scenarios. This increases the size and diversity of the dataset. As shown in Fig. 4, the model that is trained based on this combined dataset is generally achieving better than the models trained on the individual scenario datasets.

How early can a blockage be predicted? To answer this question, we evaluated the top1-accuracies and F1-scores for different future prediction window lengths (based on the combined dataset) in Fig. 5. As shown in this figure, the proposed approach achieves more than 90% prediction accuracy till the future-5 prediction interval (an average of 500 ms before the blockage happens). Even though the prediction accuracy and F1-score starts degrading after the future-5 instance, we observe that the model achieves almost 80% accuracy for predicting up to the 10th future instance. Accuracy being a holistic metric may not reflect the intricacies of the blockage prediction task. To develop a deeper insight into the model's performance, we plot the confusion matrices in Fig. 6(a) and Fig. 6(b), for future-1 and future-10 combined predictions, respectively. The high precision of 96% and 78% for both cases further highlights the high efficiency of the proposed architecture in the future blockage prediction task. The final adoption of any solution depends on achieving both high accuracy and extremely low latency. All three different stages of the proposed solution, i.e., (i) image enhancement, (ii) object detection, and (iii) recurrent prediction, contributes toward the final inference latency. For this, we compute the prediction latency associated with each of the three stages of the solution. The analysis is performed on a single NVIDIA Quadro 6000 GPU. The total inference latency is $\approx$ 45 ms, where the contribution of the three stages are ≈ 10ms, 30ms, and ≈ 4ms. The solution is designed to predict the future blockages, the earliest being 100ms in the future. The computed inference latency lies well within this prediction window, highlighting the computational efficacy of the solution.

VII. Conclusion

This paper explores the potential of leveraging visual sensory data for proactive blockage prediction in a mmWave communication system. We formulate the vision-aided blockage prediction problem and develop an efficient machine learning-based solution to predict future blockages. The key takeaways of evaluating our vision-aided blockage prediction solution based on the large-scale real-world dataset, DeepSense, can be summarized as follows: (i) the vision-aided solution achieves high blockage prediction accuracy of more than 90% for a shorter prediction window, i.e., for predicting future moving blockages that are within 500 ms. (ii) For predicting further into the future (within one second), the proposed solution achieves an average prediction accuracy of more than 80%. These results highlight the potential gains of leveraging visual data in predicting future link blockages and enable proactive network management decisions.



(a) Future-1 (b) Future-10

Fig. 6. This figure shows the confusion matrices for future-1 and future-10 blockage prediction interval (based on the combined dataset).

VIII. Acknowledgment

This work is supported in part by the National Science Foundation under Grant No. 2048021.

References

[1] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[2] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, 2017.

[3] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5g mmwave mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069–2084, 2017.

[4] V. Petrov and et. al, "Dynamic multi-connectivity performance in ultra-dense urban mmwave deployments," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2038–2055, 2017.

[5] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in *in Proc. of IEEE GlobalSIP*, Nov 2018, pp. 1055–1059.

[6] 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.

[7] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmwave beam and blockage prediction using sub-6ghz channels," *IEEE Transactions on Communications*, pp. 1–1, 2020.

[8] U. Demirhan and A. Alkhateeb, "Radar aided proactive blockage prediction in real-world millimeter wave systems," in *Proc. of IEEE ICC, arXiv preprint, arXiv:2111.14805*, 2021.

[9] S. Wu, C. Chakrabarti, and A. Alkhateeb, "Lidar-aided mobile blockage prediction in real-world millimeter wave systems," in *Proc. of IEEE WCNC, arXiv preprint, arXiv:2111.09581*, 2021.

[10] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 193–10 208, 2021.

[11] 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

[12] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *ECCV*, 2020.