CS230

Multi-Modal Sensing-Aided Wireless Prediction Models

Durga P Malladi
Stanford University
dmalladi@stanford.edu

Faith T Adegbenro
Stanford University
faithade@stanford.edu

Abstract

Traditional wireless communication beamforming techniques rely upon the reception of transmitted reference signals to process wireless channel information (amplitude, phase, delay), prior to data communication. As the usage of higher frequencies (60 GHz above) and near line-of-sight (LOS) operation becomes more prevalent, we investigate augmentation of traditional wireless signal processing techniques with the usage of additional sensor (Camera, Radar) input to improve the performance of beam prediction algorithms. In this paper, we propose and evaluate the performance of novel AI based beam prediction algorithms in 60 GHz wireless communication systems, using wireless reference signals, RGB camera and Radar. Based on the analysis conducted, we conclude that joint AI based processing of wireless signal and RGB Camera / Radar input significantly improves the accuracy, precision, recall and F1-score of beam prediction.

1 Introduction

In the past two decades, local and wide area wireless networks have become increasingly complex, as they offer high data rates, low latency and higher degree of reliability to users.

Over this period, multiple important trendlines have emerged: (1) Usage of higher frequencies (e.g. 28 GHz, 60 GHz) where plenty of bandwidth is available for data communication (2) Nonlinear signal processing in baseband modems and RF front end components (3) Large amount of data generated and available from networks devices for offline processing (4) Increasing computational power (CPU, GPU, NPU) embedded in the devices, complemented by edge and cloud based processing (5) Availability of additional sensors (Camera, LiDAR, Radar, Gyro, Accelerometer) in wireless devices such as smartphones and vehicles.

Given the trendlines on availability of modem baseband and RF data, processing power and additional complementary sensory input, a question arises as to whether one can incorporate all available wireless and sensory input into a single unified input, and use AI algorithms to process that unified input to achieve overall better performance.

Case in point, high frequency wireless systems (e.g. 60 GHz) frequently rely upon line-of-sight communication and any awareness of the channel geometry (buildings, vehicles, obstacles) via cameras, LiDAR and Radar should be considered as additional signals available at the device receiver, to complement wireless signals being processed.

In this paper, we focus on high frequency millimeter-wave (60 GHz) wireless communication systems, where large antenna arrays are used to generate a large number of narrow (in azimuth and elevation) analog plus digital beams to transmit and receive signals. The data rate at a device is maximized by picking the best possible beam from all the candidate beams transmitted by the network.

CS230: Deep Learning, Spring 2022, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

However, finding the best beam can be a sub-optimal and inefficient process, if one only relies on a subset of input signals (wireless signals only) using classical signal processing techniques. Conversely, using classical signal processing on the entire input signal (including sensors) can be computationally expensive, leading to prohibitive time for optimal beam search and/or higher power consumption.

To tackle this problem, in this paper we propose five distinct AI algorithms for beam prediction in 60 GHz wireless communication systems using (a) mmWave wireless signals only (b) RGB Camera based input only (c) Wireless signals plus Camera input (d) 3D complex raw Frequency Modulated Continuous Wave (FMCW) Radar Sensor data only (e) Wireless signals plus FMCW Radar input.

2 Related Work

The state of the art on AI research and applications in wireless networks was comprehensively covered in (3), with applications in five areas (a) Sensing AI (b) Network AI (c) Device AI (d) Medium Access AI (e) Data Provenance AI. Our focus is on AI assisted beam prediction in 60 GHz Wireless Communications, which falls under the broad category of Medium Access AI.

Most recently, CNNs have been used for channel state information and mmWave beamforming processing (6)(8)(1), while DNNs have been used for channel precoding and allocation techniques (4)(9). Additional work on AI assisted mmWave beamforming techniques can be found in (10)(7)(11).

However there is sparse published research on joint AI processing of blended data from wireless signals and sensors (cameras, Radar) for Medium Access AI. Our project attempts to tackle this problem in a holistic manner, combining all input data from wireless baseband processing and RGB cameras.

3 Dataset

The dataset from DeepSense 6G (2) (5) consists of 22,823 wireless 60GHz mmWave signals collected across 9 different scenarios. These scenarios were collected from 9 different indoor and outdoor locations across varying weather conditions and times of day, focusing on 5G vehicle-to-infrastructure communications.

The mmWave receiver, RGB camera and Radar sensor capture data at different rates. The mmWave receiver runs at 8Hz, RGB camera captures data at 30 frames/second (fps) while the radar sensor captures data at 10 fps with a frequency range of 76-81 GHZ.

Given that not all scenarios have camera images and radar sensor signals, we specifically focus on Scenarios 5/8/9 for mmWave signals plus Camera based processing, and Scenario 9 for mmWave signals plus Radar based processing.

4 Model Architecture and Approach

4.1 Wireless Signals Only

The received wireless mmWave signal is a 64x1 real-valued vector and is used directly as the input signal, while the labeled output is a 64x1 vector comprising of the correct beam indices.

Given that both the input and desired output can be modeled as a Nx1 vector, we pick a Deep Neural Network (DNN) to process the data.

The mmWave signal 64x1 input vector is fed to a DNN with 4 hidden FC layers involving (1024,512,256,128) activations, with a Softmax 64x1 output layer that indicates the probability of each beam index. This DNN network has 763,840 parameters as shown in Table 1 and is trained using Adam Optimization with a categorical cross-entropy loss function.

4.2 RGB Camera Only

The RGB camera input signal is a 960x540x3 volume, while the labeled output is a 64x1 vector comprising of the correct beam indices.

Table 1: DNN Architecture - Wireless mmWave Signal Processing

Network	Input Volume	Final Output	Architecture	Number of Parameters
DNN-mmW	64x1	64x1	FC-FC-FC-FC-Softmax	763,840

One approach would be to view this as a classical image processing problem and solve using a ConvNet architecture (e.g. VGG-16) comprising of several CONV-POOL-ReLU layers and FC layers towards the end. However note that the label is not an object embedded in the image, rather it's the index of the correct beam.

So we use a Transfer Learning technique to pre-process the RGB camera images using VGG-16 and cascading the VGG-16 FC2 output to a new fully connected DNN, which is then trained with the labels comprising of correct beam indices.

Concretely, we implemented a network architecture wherein the RGB camera 960x540x3 volume is re-sized to 224x224x3 and run through VGG-16 trained on ImageNet, and extracted the 4096x1 FC2 layer prior to the output Softmax layer. Note that there are 134,260,544 parameters in VGG-16 up to FC2 layer, as shown in 2.

This 4096x1 FC2 layer output constitutes the input to a new fully connected DNN with 4 hidden FC layers involving (512,384,256,128) activations, with a Softmax 64x1 output layer that indicates the probability of each beam index. This DNN network has 2,434,368 parameters as shown in Table 2 and is trained using Adam Optimization with a categorical cross-entropy loss function.

Table 2: VGG-16 plus DNN Architecture - RGB Camera Image Processing

Network	Input Volume	Final Output	Architecture	Number of Parameters
Original VGG-16	224x224x3	1000x1	See literature	138,357,544
Modified VGG-16	224x224x3	4096x1	VGG-16 upto FC2	134,260,544
DNN-RGB	4096x1	64x1	FC-FC-FC-FC-Softmax	2,434,368

4.3 Wireless Signals plus RGB Camera

We note that this is a heterogeneous input data problem, with a 64x1 real-valued received mmWave signal coupled with a 960x540x3 RGB camera signal. The labeled output remains a 64x1 vector comprising of the correct beam indices.

Instead of unifying the input signal up-front, we first process RGB camera input signal as described in previous section to get a 64x1 Softmax output, that is used as a image feature vector. This image feature vector is further augmented by appending the 64x1 real-valued received mmWave signal, to create a new 128x1 input vector.

This 128x1 input vector is fed to a new fully connected DNN with 4 hidden FC layers involving (1024,512,256,128) activations, with a Softmax 64x1 output layer that indicates the probability of each beam index. This DNN network has 763,840 parameters as shown in Table 3 and is trained using Adam Optimization with a categorical cross-entropy loss function.

The new 128x1 input vector is to be processed by a DNN, consisting of several hidden layers followed by a 64x1 softmax output layer that indicates the probability of best beam index, and is trained using Adam Optimization with a cross-entropy loss function.

Table 3: VGG-16 plus New DNN Architecture - RGB Image Processing

Network	Input Volume	Final Output	Architecture	Number of Parameters
DNN-RGB-mmW	128x1	64x1	FC-FC-FC-FC-Softmax	829,376

4.4 Radar Sensors Only

The FMCW (frequency-modulated continuous-wave) radar outputs 3D complex I/Q radar measurements of dimension 4 (number of Rx antennas) x 256 (samples per chirp) x 128 (chirps per frame) at 10 fps. This output can be pre-processed to produce 3 distinct data mappings: radar cube, range velocity mapping, and range angle mapping.

The range angle mapping outputs the propagation range and direction (azimuth and elevation angles) of a signal from a source point. Each data sample of the range angle mapping outputs a vector of size 1x256x64. This is fed into a CNN that outputs a 64x1 softmax prediction of 64x1 as is shown in Table 5.

Table 4: CNN Architecture - FMCW Radar Processing

Network	Input Volume	Final Output	Architecture	Number of Parameters
CNN-Radar	1x256x64	64x1	PL-5(Conv2D)-Flatten-FC-FC-Softmax	2,141,334

4.5 Wireless Signal plus Radar Sensors

The wireless signal and radar sensors were initially concatenated in a fashion analogous to that of wireless signal plus RGB camera data and it was then passed into 4 fully connected ReLU layers and a softmax output. However this approach produced validation and test metrics that were 35-40%.

The next approach tried was to concatenate the pre-processed radar angle data with the power vectors, but this approach put too much weight on the radar data and a lot of irrelevant details were included in the model input. The current hypothesis is to try extracting the relevant features using convolutional layers and combining this, as opposed to the initial softmax output or the pre-processed radar data, might improve the model performance.

Table 5: CNN plus DNN Architecture - FMCW Radar Processing

Network	Input Volume	Final Output	Architecture	Number of Parameters
DNN-Radar-mmW	128x1	64x1	FC-FC-FC-FC-Softmax	837,312
CNN-Radar-mmW	1x257x64	64x1	PL-5(Conv2D)-Flatten-FC-FC-Softmax	2,141,334

5 Dataset and Evaluation Metrics

Based on the different Model architectures presented in the previous section and amount of data available, we split the data into Train/Val/Test sets as shown in Table 7.

We used batch processing (batch size = 32) during training over 200 epochs for wireless signal processing, RGB camera processing and wireless + RGB camera processing. For radar signal processing and wireless + radar processing, we also used batch sizes of 32 during training but it was over 20 epochs.

Each scenario method is evaluated with a combination of metrics, namely Accuracy, Precision, Recall and F1 Score, to take a comprehensive view of the efficacy of algorithms.

6 Results and Analysis

The results for consolidated and individual Scenarios 5, 8 & 9 are listed in Tables 8 910 11 6 respectively.

We make the following conclusions from the results:

- Using Wireless Signals alone leads to low F1 scores (0.64-0.70) from val/test set results across all scenarios

Table 6: Training, Validation & Test Set Results - Scenario 9

Method	Train/Val/Test	Accuracy	Precision	Recall	F1 score
Wireless Signals	Train	0.9000	0.9144	0.8868	0.9004
	Val	0.6879	0.7080	0.6795	0.6935
	Test	0.6851	0.7063	0.6767	0.6912
RGB Camera	Train	1.0000	1.0000	1.0000	1.0000
	Val	0.5084	0.5084	0.5050	0.5067
	Test	0.5159	0.5176	0.5159	0.5168
Wireless Signals plus RGB Camera	Train	0.9593	0.9603	0.9579	0.9591
	Val	0.9077	0.9091	0.9060	0.9076
	Test	0.9012	0.9027	0.9012	0.9019
Radar Data	Train	0.9917	0.9919	0.9910	0.9914
	Val	0.3951	0.4009	0.3865	0.3936
	Test	0.3879	0.3962	0.3862	0.3911
Wireless Signals plus Radar Data(DNN)	Train	0.9976	0.9976	0.9974	0.9975
	Val	0.3831	0.3883	0.3797	0.3840
	Test	0.3929	0.3951	0.3845	0.3897
Wireless Signals plus Radar Data(CNN)	Train	0.9564	0.9810	0.9228	0.9510
	Val	0.3959	0.4322	0.3729	0.4004
	Test	0.3963	0.0.4271	0.0.3609	0.3912

- The RGB camera network has a large variance (over-fitting), as seen from the delta between train and val/test set accuracy, precision, recall and F1 scores. The variance is very pronounced in Scenario 5 and remains large for Scenarios 8 and 9 as well.
- Using a combination of Wireless Signals plus RGB Camera, provides the best results, as seen from the consistent accuracy, precision, recall and F1 scores (0.86 or above) from val/test across all scenarios
- The Radar network has a large variance (over-fitting) as seen from the delta between train and val/test set accuracy, precision, recall and F1 scores
- Combining wireless signals with Radar input does not improve overall performance or lower the variance

7 Potential Future Work

Based on the analysis, we recommend the following next steps as a part of future enhancements:

- Reduce variance in RGB camera processing methods (smaller network, larger dataset)
- Try different ConvNet models (e.g. ResNet50) for RGB camera image processing
- Investigate different techniques to improve Wireless signal processing in Scenario 8
- Analyze other representations of radar data (radar cube range velocity) for beam prediction
- Analyze feature extraction for radar data plus wireless signals
- Explore the combination of Camera plus Radar plus Wireless signal processing in a DNN

8 Contributions

Durga worked on creating models for the wireless signals, RGB camera data and the combination of both signals, while Faith worked on creating models for the wireless signals, Radar input and combination of wireless signal and Radar data.

References

[1] A.ELBIR, AND A.PAPAZAFEIROPOULOS. Hybrid precoding for multiuser millimeter wave massive mimo systems. *IEEE Transactions on Vehicular Technology* (January 2020), 552–563.

[2] ALKHATEEB, A., CHARAN, G., OSMAN, T., HREDZAK, A., AND SRINIVAS, N. DeepSense 6G: Large-scale real-world multi-modal sensing and communication datasets. *to be available on arXiv* (2022).

[3] D. NGUYEN, P. CHENG, M. D. D. L.-P. P. P. J. L. A. S. Y. L., AND POOR, V. Enabling ai in future wireless networks: A data life cycle perspective. *IEEE Communications Surveys and Tutorials 23* (March 2021), 553–595.

[4] DE KERRET, P., AND D.GESBERT. Robust decentralized joint precoding using team deep neural network. *Proceedings 15th International Symposium on Wireless Communication Systems (ISWCS)* (2018), 1–5.

[5] DEMIRHAN, U., AND ALKHATEEB, A. Radar aided 6g beam prediction: Deep learning algorithms and real-world demonstration. 2655–2660.

[6] J.YUAN, H., AND M.MATTHAIOU. Machine learning based channel estimation in massive mimo with channel aging. *Proceedings 20th International Workshop on Signal Processing and Advanced Wireless Communications (SPAWC)* (2019), 1–5.

[7] N. YE, X. LI, J. W. L., AND HOU, X. Beam aggregation based mmwave mimo-noma: An ai enhanced approach. *IEEE Transactions on Vehicular Technology* (March 2021), 2337–2348.

[8] Q.YANG, M., AND D.GUNDUZ. Deep convolutional compression for massive mimo csi feedback. *Proceedings 29th International Workshop on Machine Learning and Signal Processing* (2019), 1–6.

[9] S.LEE, H., AND B.SHIM. Pilot assignment and channel estimation via deep neural network. *Proceedings 24th Asia-Pacific Conference on Communications (APCC)* (2018), 454–458.

[10] T. MAKSYMYUK, J. GAZDA, O. Y., AND NEVINSKIY, D. Deep learning based massive mimo beamforming for 5g mobile network. *International Symposium on Wireless Systems* (September 2018), 241–244.

[11] W.C.KAO, S., AND T.S.LEE. Ai-aided 3d beamforming for millimeter wave communications. *IEEE Transactions on Vehicular Technology* (December 2018), 278–283.

A Appendix

Table 7: Dataset Overview

Scenario	Dataset	Train Samples	Val Samples	Test Samples
Scenario 1	2411	1675	451	296
Scenario 2	2974	1914	607	453
Scenario 3	1487	1037	300	150
Scenario 4	1867	1248	401	218
Scenario 5	2300	1840	230	230
Scenario 6	915	653	98	164
Scenario 7	854	568	185	103
Scenario 8	4043	3234	404	405
Scenario 9	5964	4771	596	597

Table 8: Training, Validation & Test Set Results - Consolidated Scenarios 1-9

Method	Train/Val/Test	Accuracy	Precision	Recall	F1 score
Wireless Signals	Train	0.8792	0.8955	0.8644	0.8797
	Val	0.7411	0.7626	0.7254	0.7435
	Test	0.7642	0.7790	0.7473	0.7628

Table 9: Training, Validation & Test Set Results - Consolidated Scenarios 5,8,9

Method	Train/Val/Test	Accuracy	Precision	Recall	F1 score
Wireless Signals	Train	0.8671	0.8827	0.8488	0.8654
	Val	0.7439	0.7640	0.7236	0.7432
	Test	0.7054	0.7224	0.6908	0.7062
RGB Camera	Train	0.9347	0.9384	0.9323	0.9353
	Val	0.4976	0.5012	0.4902	0.4957
	Test	0.5000	0.5037	0.4951	0.4994
Wireless Signals plus RGB Camera	Train	0.9627	0.9660	0.9598	0.9629
	Val	0.8472	0.8567	0.8455	0.8511
	Test	0.8644	0.8674	0.8604	0.8639

Table 10: Training, Validation & Test Set Results - Scenario 5

Method	Train/Val/Test	Accuracy	Precision	Recall	F1 score
Wireless Signals	Train	0.8989	0.9073	0.8886	0.8979
	Val	0.6652	0.6773	0.6478	0.6622
	Test	0.6435	0.6476	0.6391	0.6433
RGB Camera	Train	1.0000	1.0000	1.0000	1.0000
	Val	0.3826	0.3843	0.3826	0.3834
	Test	0.4217	0.4254	0.4217	0.4236
Wireless Signals plus RGB Camera	Train	0.9755	0.9761	0.9755	0.9758
	Val	0.8826	0.8826	0.8826	0.8826
	Test	0.8435	0.8546	0.8435	0.8490

Table 11: Training, Validation & Test Set Results - Scenario 8

Method	Train/Val/Test	Accuracy	Precision	Recall	F1 score
Wireless Signals	Train	0.6871	0.7141	0.6586	0.6852
	Val	0.7054	0.7205	0.6510	0.6840
	Test	0.6642	0.6764	0.6296	0.6522
RGB Camera	Train	1.0000	1.0000	1.0000	1.0000
	Val	0.5767	0.5782	0.5767	0.5774
	Test	0.5877	0.5906	0.5876	0.5891
Wireless Signals plus RGB Camera	Train	0.9542	0.9542	0.9536	0.9539
	Val	0.9332	0.9332	0.9332	0.9332
	Test	0.9210	0.9256	0.9210	0.9233