Multimodality in mmWave MIMO Beam Selection Using Deep Learning: Datasets and Challenges

Jerry Gu, Batool Salehi, Debashri Roy, and Kaushik R. Chowdhury

The authors present an experimentally obtained dataset, composed of 23 GB of data, which aids in beam selection in vehicle to everything (V2X) mmWave bands, with the goal of facilitating machine learning (ML) in the wireless communication required for autonomous driving.

Abstract

The increasing availability of multimodal data holds many promises for developments in millimeter-wave (mmWave) multiple-antenna systems by harnessing the potential for enhanced situational awareness. Specifically, inclusion of non-RF modalities to complement RF-only data in communications-related decisions like beam selection may speed up decision making in situations where an exhaustive search, spanning all candidate options, is required by the standard. However, to accelerate research in this topic, there is a need to collect real-world datasets in a principled manner. This article presents an experimentally obtained dataset, composed of 23 GB of data, which aids in beam selection in vehicle-to-everything mmWave bands, with the goal of facilitating machine learning (ML) in the wireless communication required for autonomous driving. Beyond this specific example, the article describes methodologies of creating such datasets that use time synchronized and heterogeneous types of LiDAR, GPS, and camera images, paired with the RF ground truth data of selected beams in the mmWave band. While we use beam selection as the primary demonstrator, we also discuss how multimodal datasets may be used in other ML-based PHY-layer optimization areas, such as beamforming and localization.

Introduction

Interest in using large-scale multimodal datasets has surged within the past few years given the widespread availability of small form-factor sensors and decreasing data storage costs. Such datasets may include historical and real-time streaming samples of camera images, GPS coordinates, light detection and ranging (LiDAR) pointclouds, radar signals, infrared signals, and acoustic signals, to name a few. These sensors may have a wide range of applications that include biomedical sensing, autonomous vehicles, and the ubiquitous Internet of Things. However, effective utilization of the collected information to extract meaningful context from the operational environment remains an open challenge, particularly when real-time execution is needed.

Diversity of data sources for wireless PHY optimization use cases. Classical wireless datasets use radio frequency (RF) data for applications such as channel estimation and interference detection, generally with multiple- input multiple-output (MIMO) systems that utilize antenna arrays. While rapid advances in machine learning (ML) have demonstrated benefits of using RF data, recent results indicate the exciting possibility of leveraging non-RF data as well to improve system performance and increase solution robustness [1]. Non-RF modalities can complement tasks where rich environmental representations circumvent the shortcomings of RF data, albeit at increasing data collection costs and management overhead.

Challenges in creating multimodal datasets. Naturally, the creation of multimodal datasets comes with its own set of challenges. Depending on the application of a multimodal dataset, data samples may need to be synchronized, particularly for use with ML techniques that require a uniform input size. This requires meticulous consideration of the timescale characteristics of the different sensor types, including device startup delay and sampling frequency. These samples should have accurate labels and timestamps based on a common timescale, and in the likely scenario that different device capture frequencies do not align, sensor-specific "gaps" in the dataset should immediately be addressed for synchronization. Finally, any collected raw data needs to be checked for compatibility with a given task. For example, LiDAR pointclouds vary in size, which is undesirable for deep learning (DL)-based approaches that require input tensors of fixed size. Hence, preprocessing steps may be necessary to compress and standardize data representations. Our contributions are as follows:

- We discuss how multimodal data can be used for beam selection for wireless links, specifically concerning beamforming in the millimeter-wave (mmWave) band.
- We survey different existing multimodal datasets that have the potential to contribute to emerging research directions.
- We describe a novel generalized process for collecting multimodal datasets using both our dataset, e-FLASH, and the collection process behind e-FLASH to illustrate this process. e-FLASH, which consists of heterogeneous LiDAR, camera, and GPS sensor-generated real-world data, was collected for the purpose of beam selection in mmWave bands. We make this dataset publicly available [4] to accelerate community contributions in the

Digital Object Identifier: 10.1109/MCOM.002.2200028

The authors are with Northeastern University, USA.

exciting research area of mmWave communication.
- We conduct preliminary experiments that utilize multimodal data for beam selection, showing the potential to achieve low-latency high-reliability communication within mmWave systems.

Use Cases of Multimodal Datasets

Multimodal data may be leveraged to speed up the utilization of heavily congested RF bands (Fig. 1). As an example, modalities such as camera images and LiDAR are leveraged to provide contextual information and/or guide existing RF-only approaches. Earlier works such as [2, 3] have demonstrated that it is possible to increase accuracy of beam selection by using multimodal datasets. The contextual awareness drastically reduces inference time for optimal beam selection. These initial studies motivate us to identify different PHY-layer topics in mmWave and MIMO systems, describing how multimodal data in general can address research challenges in these topics. An overview of some state-of-the-art (SOTA) techniques is presented in Table 1.

Beam Selection

Propagation in the mmWave spectrum is severely limited due to the increased path loss of RF signals in high carrier frequencies. To mitigate this propagation characteristic, directional links that offer the strongest signal strength in the dominant signal lobe between the transmitter (Tx) and receiver (Rx) pair are used, but they must be carefully aligned. This alignment process, known as beam selection, is essential and must be repeated frequently in mobile environments to ensure continuous connectivity. Several prior efforts use a combination of RF and non-RF data, such as light-based sensors [5, 6], to perform beam selection. Such techniques exploit the benefit of variety of light-based modalities (e.g., LiDAR and images) to either offset the sole reliance on RF data or increase beam selection efficiency.

Beamforming

Beamforming focuses on the generation of multiple elements in a Tx antenna array with the goal of creating constructive interference at the Rx while producing destructive interference elsewhere, which is then paired with beam steering and selection steps in a system for optimal performance. While there is myriad research available on various beamforming techniques, they are either:
- Solely validated on simulation studies [8]
- Have the potential to be improved using DL-based data-driven approaches [7]

In either case, SOTA techniques may benefit if their outcomes are demonstrated on a real-world dataset.

Localization

Localization is the prediction of a device's placement in space through the use of one or more wireless transmitters. Specific to scenarios involving mmWave links, [9] uses pre-existing automotive radars and Van Atta arrays to build a low-power tag that may be localized with high accuracy over an extended range. In [10], the authors use signal-to-noise ratio (SNR) measurements to construct an ML-based system that performs only location-based classification, location and orientation classification, and coordinate estimation. In [11], channel state information (CSI) from commercial off-the-shelf mmWave routers is used for localization and then to optimize network operations. We believe that all of these approaches stand to benefit from the inclusion of LiDAR and GPS data, given the significant environmental context awareness achievable beyond what has been demonstrated with only the classical parameters of CSI and SNR.

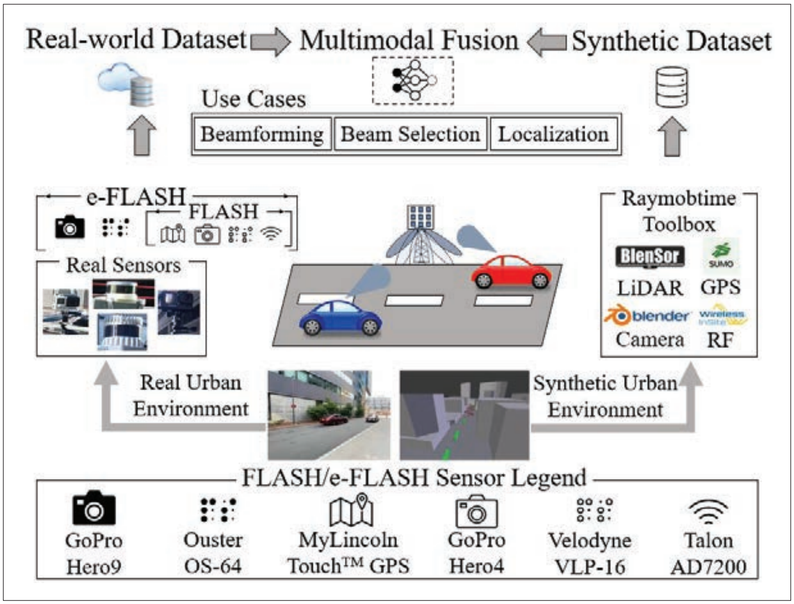

FIGURE 1. Use cases and overview of real-world FLASH [2], e-FLASH, and synthetic Raymobtime [3] datasets.

Application	Modalities	Methods	Paper
Beam selection	RF, image	Computer vision, deep learning	Alrabeiah *et al.* [5]
Beam selection	RF, LiDAR	Deep learning, fusion	Dias *et al.* [6]
Beamforming	RF	Channel measurements, hybrid RF/baseband system	Roh *et al.* [7]
Beamforming	RF, radar	Hybrid beamforming, sub-arrayed radar techniques, minimization	Liu *et al.* [8]
Localization	RF, radar	Reflector array	Soltanaghaei *et al.* [9]
Localization	RF	Deep learning	Koike-Akino *et al.* [10]
Localization	RF	Channel measurements	Palacios *et al.* [11]

TABLE 1. Survey of literature leveraging the use of different modalities for solving mmWave challenges.

Classification of Existing Multimodal Datasets for Wireless PHY

We survey available simulation and real-world multimodal datasets consisting of primarily non-RF data that address different problems in the mmWave band.

Simulation Datasets

Raymobtime: The Raymobtime dataset [3] consists of a high-fidelity virtual vehicle-to-infrastructure (V2I) deployment for modeling 5G and mmWave MIMO channel propagation. The dataset is obtained from a simulated urban canyon region, where buildings flank the two sides

The first step in collecting a multimodal dataset is to choose the right data collection environment. In our case, e-FLASH addresses the sector selection problem in the mmWave band, which will support low-latency communication via short-range links within high-traffic regions.

FIGURE 2. The experimental setup of e-FLASH data collection.

of a road. Using the Simulator for Urban Mobility (SUMO) software, a variety of moving vehicles are generated to form traffic patterns near a roadside Tx. Image, LiDAR, and ray tracing (RT) data are collected by Blender, Blender Sensor Simulation (BlenSor), and Remcom Wireless Insite, respectively. Synchronized samples of LiDAR pointclouds, GPS coordinates, and camera images are collected for each scene, in which an active Rx is designated to one of the three vehicle types. Overall, 256 beam configurations, consisting of 32 codebook elements at the Rx and eight elements at the Tx, are collected, cumulatively giving over 6000 scenes (55 GB data).

ViWi: The Vision-Wireless (ViWi) framework utilizes 3D modeling and RT in combination with wireless data for the purpose of facilitating vision-aided mmWave wireless communication [12]. Simulated outdoor environments containing objects such as buildings, cars, and people are generated with Blender, while RT is done with Remcom Wireless Insite. The initial version of ViWi features four sets of raw data, consisting of visual, LiDAR, and wireless data (e.g., depth maps, signal angles of departure, and channel gains), and a data-generating package. An extended version of this dataset, named ViWi Vision-Aided mmWave Beam Tracking (ViWi-BT), contains an additional 13 pairs of consecutive beam indices with corresponding street view images in a new heavy-traffic downtown environment. ViWi-BT contains 281,000 training samples, 120,468 validation samples, and 10,000 test samples for ML-based beam tracking.

Real-World Datasets

Image-Based mmWave Beamforming: The dataset collected by Salehi *et al.* [1] uses camera images to complement beamforming optimization between two mmWave antenna arrays. The Tx and Rx antenna arrays are placed on mobile sliders and are set to move to five distinct positions with an obstacle in between the arrays to represent the radiation pattern under both line-of-sight (LoS) and non-LoS (NLoS) conditions. Two GoPro Hero4 cameras synchronized with mmWave radios are used to observe movement in the environment. The corresponding raw images are fed through a binary classifier to detect the locations of the Tx and Rx in the image and are later passed to a second classifier to direct one of 13 beam directions at both the Tx and Rx. The dataset contains over 3750 samples (2.85 GB), which are used to predict the optimal beam pairing using raw input images.

FLASH: The Federated Learning for Automated Selection of High-band mmWave Sectors (FLASH) dataset is a real-world multimodal dataset wherein sensory data from LiDAR, camera, and GPS sensors participate in federated learning to speed up beam selection in mmWave vehicular networks [2]. Overall, this multimodal dataset is collected in real time with sensors mounted on an autonomous vehicle in an urban canyon environment in a variety of LoS and NLoS scenarios. Using the locally collected data, obstacles are detected within the immediate surroundings, and their locations are used to make guided inferences for selecting the best mmWave sector. The sensor suites in FLASH contain three sensors:
- 16-channel LiDAR
- A side-facing camera
- GPS

The labels in FLASH are time synchronized RF ground truth of the received signal strength at each mmWave beam. Overall, the dataset contains 31,923 samples (∼20 GB processed data), covering a wide variety of real-world vehicle-to-base station (BS) LoS and NLoS scenarios.

Multimodal Dataset Creation Framework

In this section, we present a step-by-step methodology of collecting a multimodal dataset, using the extended FLASH (e-FLASH) dataset as a guiding example. e-FLASH is an extension of the original FLASH dataset with the addition of two new sensors that provide richer environmental representation, which we detail in the Sensors subsection. Additionally, e-FLASH provides a structured way of representing the multimodal data in the portable Hierarchical Data Format 5 (HDF5) format with self-explanatory metadata. The e-FLASH dataset is available for community use [4].

Data Collection Environment

The first step in collecting a multimodal dataset is to choose the right data collection environment. In our case, e-FLASH addresses the sector selection problem in the mmWave band, which will support low-latency communication via short-range links within high-traffic regions. Thus, we set up the data collection in an environment containing a two-lane 6-m-wide alleyway between high-rise buildings located 4 m from the road and multiple building material types, vegetation, and weather conditions.

Sensors

The selection of the right number of proper sensors is crucial to defining a well-structured problem using multimodal data. Using a number of modalities allows for both a more complete representation of the environment, which may improve inference performance, and greater flexibility in the selective contribution of the situationally favored modality for overall performance optimization. Most modern cars come pre-equipped with at least GPS tracking systems, which perform well in LoS environments. Although vehicular LiDAR systems are less common, they are more dependable in NLoS scenarios. However, low-light conditions and reflections from strong sunlight can hamper the performance of LiDAR systems [13].

For e-FLASH, we consider *five* different sensors which capture data via three different modalities e-FLASH Dataset

Category 1: LOS		Category 2: NLOS Pedestrian		Category 3: NLOS Static Car		Category 4: NLOS Moving Car		
Car Speed	10 mph	Car Speed	15 mph	Car Speed	10 mph	Car Speed	15 mph	20 mph
BS Location	Same lane	BS Location	Opposite lane	BS Location	Opposite lane	BS Location	Opposite lane	
# Samples	Static (1568)	# Samples	Walking R-L (1593) Walking L-R (1603) Static (1595)	# Samples	In front of the BS (1506)	# Samples	15 mph same dir. (1557)	20 mph opp. dir. (1431)

FIGURE 3. Summary of different categories in the e-FLASH dataset.

The synchronization process is crucial when creating a multimodal dataset for DL-based applications that expect constant input sizes. In e-FLASH, individual modalities use the same reference clock to record real-time observations with different sampling frequencies.

that are representative of sensor types found in modern vehicular systems to record various representations of the environment for mmWave beam selection. The sensor suite consists of one side-facing GoPro Hero4 Silver camera, one front-facing GoPro Hero9 Black camera, two Velodyne VLP-16 LiDAR sensors, one Ouster OS164 LiDAR sensor, and a GPS system onboard a 2017 Lincoln MKZ Hybrid autonomous car. The sampling frequencies are 30 fps for the cameras, 10 Hz for the LiDARs, and 1 Hz for the GPS. The GPS and LiDAR sensors are connected to an onboard computer in the vehicle using low-latency Ethernet cables and assigned unique IP addresses. The computer runs the Robot Operating System (ROS) suite, which reads measurements instantaneously. The GoPro cameras are also connected using Wi-Fi and USB cables for Hero4 and Hero9, respectively, for simultaneous operation. We use a customized Python code employing OpenCV to control the camera recordings, labeling them with the appropriate timestamps. All sensors are connected to the same computer for common reference timestamping.

In e-FLASH, we use mmWave radios to collect the RF ground truth corresponding to each multimodal sensor input. Two TP-Link Talon AD7200 tri-band routers, which use Qualcomm QCA9500 IEEE 802.11ad Wi-Fi chips with an antenna array consisting of 32 elements to communicate at the 60 GHz band, are set to function as the BS, with a coverage angle and height of 168.57° and 1.5 m, respectively, and the Rx. We use a default codebook of sector IDs 1–31 and 61–63, with IDs 32–60 undefined, and the open source Linux Embedded Development Environment (LEDE) and Nexmon firmware patching released by [14] to access the PHY-layer characteristics of the routers, using the following process. First, propriety IP addresses are assigned to each router. Then the open source iPerf3 tool is used to set the Tx-Rx pair as "client" and "server," respectively, and generate the stream to be transmitted. We use patched LEDE firmware to send full-buffer TCP traffic, recording the time-synchronized RF ground truth data at 1–1.5 Hz via data transmission rate and received signal strength indication (RSSI) for each mmWave beam sampled. The throughput is approximately ~1.5 Gb/s with a 2 GHz channel at the 60 GHz center frequency.

Data Categories

The inclusion of diverse scenarios in the collected datasets is necessary to create a comprehensive solution that addresses every aspect of the problem. We diversify possible scenarios in *categories*, as shown in Fig. 3. For each category, we define specific variations, or *scenarios*, and collect 10 *episodes*, or trials, of each scenario, with each episode lasting approximately 15 s. To start each collection, we arrange the testbed according to the signal propagation conditions (LoS or NLoS) and obstacles present (none, pedestrian, or car). We set up the server/client and ensure that the transmission traffic is observed at the Rx, which is set up in the road lane specified by the "BS Location" with respect to the collection vehicle. Then we start recording the non-RF data on all sensors while driving the collection vehicle at the designated "Car Speed" and assign unique names that describe the ongoing category. In total, we collect 70 episodes over four categories.

Synchronization

The synchronization process is crucial when creating a multimodal dataset for DL-based applications that expect constant input sizes. In e-FLASH, individual modalities use the same reference clock to record real-time observations with different sampling frequencies. On the other hand, the beam sweeping mechanism used in the Talon AD7200 routers is triggered whenever a drop in RSSI is observed. Given the information above, we design a synchronization procedure with four steps as follows:

1. We identify the image and LiDAR sensor information between two consecutive RF measurement timestamps.
2. We pair the LiDAR sample with the closest image sample.
3. We estimate missing GPS sensor samples by interpolation.
4. We assign the same RSSI measurement to all sensor modality samples in the range.

With this scheme, we extract a total of 10,853 samples (~23 GB raw data) from the raw data.

Post-Processing

The use of sensors that generate continuous representations of the environment inevitably generates a large amount data in raw formats which contain some information that may not be used. Therefore, it is essential to perform post-processing steps on the raw data to convert the data to more portable formats with high data efficacy.

In e-FLASH, we incorporate post-processing steps that extract the relevant features from raw data to generate a processed dataset. Note that

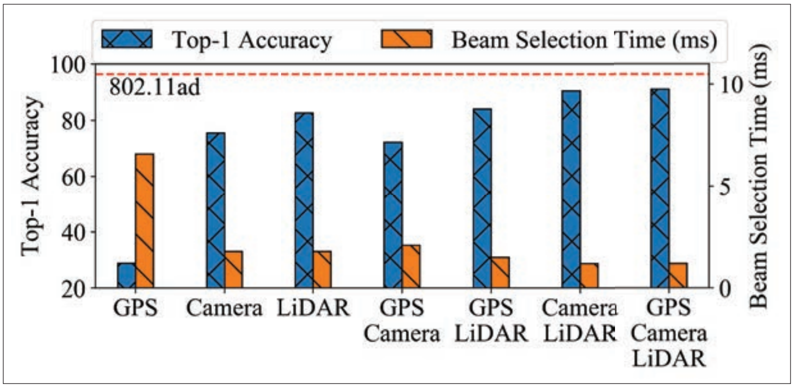

FIGURE 4. Comparing the top-1 accuracy and beam selection time of single and multiple modalities. The red horizontal line denotes the beam selection time of the IEEE 802.11ad standard.

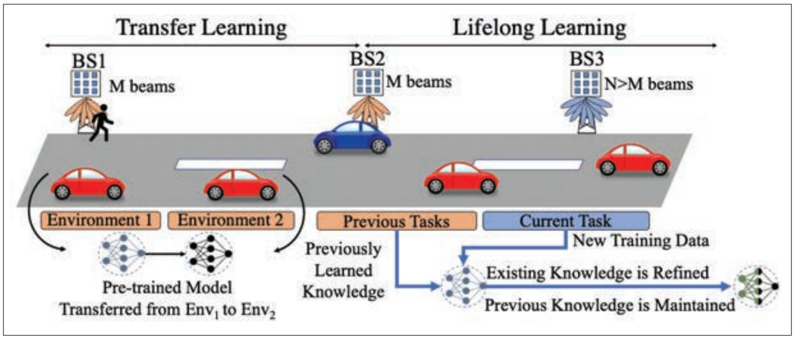

FIGURE 5. Conceptual overview of transferring the learning from different environments (similar to the categories in e-FLASH dataset) and lifelong learning to serve users on infrastructures with different proprietary service providers.

the raw GPS measurements are in decimal degrees with 13-digit precision, while the LiDAR data range is set to ±80 m. In order to resolve the measurement scale conflicts, we define a fixed origin and calculate the distances of GPS positioning recordings to arrange a Cartesian coordinate system.

Meanwhile, LiDAR pointclouds are a collection of (x, y, z) points that indicate the location of detected objects in a coordinate system with the LiDAR sensor location being the origin. As a result, the size of pointcloud files scales with the number of objects in the environment. Therefore, we apply a post-processing step on the LiDAR data as follows [3]: We quantize recorded space such that pointclouds are mapped to a rigidly defined representation where dimensionality is fixed, and the role of each object is identified with unique numerical indicators.

Deep-Learning-Based Fusion for Fast Beam Selection

In this section, we present a feasibility study of using a DL-based framework on e-FLASH, where we leverage all five sensor inputs and evaluate the end-to-end latency improvement over the IEEE 802.11ad standard from the beam selection performance in vehicle-to-everything (V2X) communication. We assume that most modern vehicles, particularly ones with onboard LiDAR systems for autonomous driving, already have the required computational power to process this specific task and other similar tasks.

Collaborative Learning on e-FLASH

As part of a supervised learning scheme, it is assumed that the training and test sets follow the same distribution. In order to achieve a generalized model, the training set must have a rich representation of different scenarios, such as the four categories described earlier. We demonstrate this by considering a collaborative learning scheme, where multiple vehicles contribute to the common training set by sharing data from local repositories to the BS.

The proposed system consists of vehicles in an urban canyon region, where buildings flank the sides of the road with roadside BSs in an mmWave V2X setup. We use a back-channel-based framework [2] to collect the data from all vehicles and train on the collective data in a collaborative fashion. We then use a DL-based fusion network with stochastic gradient descent-based training and a learning rate of 0.0001 over the cloud, where multimodal sensor data with RF ground truth is transmitted collectively from the vehicle to the cloud. In brief, the designed neural network consists of three unimodal architectures (each with 4–9 convolutional and 2–4 fully connected layers) and a fusion network of 5 more fully connected layers with up to ∼105 million parameters [2]. The trained model is later transmitted from the cloud to each individual vehicle, which selects the best beam for the mmWave radios in inference time. For the most complex case utilizing fusion of all three modalities, the time complexity of this framework with 13 layers using 80 percent of data, N_{data}, samples for training is on the order of $0.8 \times N_{data}^{2^{13}}$ [15].

Experimental Evaluation

The preliminary results of predicting best beams on the e-FLASH dataset are presented in Fig. 4. We use three out of the five available modalities present in the e-FLASH dataset and consider different combinations where only one, two, or three modalities are present. Using time-synchronized samples of three sensor modalities, we train a fusion architecture that employs concatenation at the feature level to predict the optimum sector [2] and report the prediction accuracy. We show that LiDAR performs better than the other two sensors by 7.18–53.86 percent in top-1 accuracy. Our initial results show the effectiveness of a DL-based fusion over the individual modalities with 8.17 percent improvement with respect to best single modality (i.e., LiDAR). Moreover, we observe that unimodal scenarios achieve 62.32 percent average top-1 accuracy, whereas the fusion of two and three modalities improves the performance by up to 82.14 and 90.84 percent, respectively. We also analyze the end-to-end latency of the proposed beam selection scheme over the IEEE 802.11ad standard, showing that the sensor-based beam selection achieves 54.28–85.71 percent lower latency than the current IEEE 802.11ad standard while targeting >99 percent accuracy.

Potential Research Directions

We present some specific scenarios with open research problems that can benefit from the data in e-FLASH.

Adaptive Beam Selection Using Transfer Learning

The general idea of transfer learning is to propagate the knowledge from one learning task to another for an efficient outcome. The proposed beam selection task can be formulated to adapt to new environments within different datasets by consolidating the knowledge from previously seen environments. For example, an architecture previously trained on the simulation-based Raymobtime dataset can be adapted for the real-world e-FLASH dataset. The pre-trained model weights of one dataset or category can be loaded as the starting point for training the other datasets or categories. We show an overview of the proposed idea for adaptive beam selection using transfer learning in Fig. 5.

Lifelong Learning for Adaptive Beam Selection of Variant Codebooks

The goal of lifelong learning is to develop a core feature extractor that can be used for a set of different tasks by only branching the input and output layers. Consider a scenario, such as that in Fig. 5, where single/multiple vehicles are using an ML module to speed up the beam selection. The road contains three BSs with proprietary service providers. As a result, each of these BSs have unique mmWave codebooks with different numbers of beams and radiation patterns. Using lifelong learning, we may derive a model that is able to adapt to all three tasks without changing the core ML model architecture.

Conclusion

This article introduces a step-by-step multimodal dataset collection process through the e-FLASH dataset, a large multimodal dataset consisting of camera image, GPS, and LiDAR data supported with RF ground truth, for the purpose of mmWave band communication. The dataset contains 10,853 samples, consisting of seven unique real-world LoS and NLoS vehicular network scenarios that are structured, synchronized, labeled, and ready for use in ML-aided applications. While we describe two use cases of this dataset in detail for beam selection in mmWave bands, we also outline other topics where this dataset can accelerate research, including beamforming and localization. Given the number of potential applications, we believe e-FLASH will become a benchmark dataset for the emerging area of multimodal data fusion for wireless communication.

Acknowledgment

This work was made possible by the U.S. National Science Foundation CCRI Award #2120447 and CNS #1923789.

References

[1] B. Salehi *et al.*, “Machine Learning on Camera Images for Fast mmWave Beamforming,” *IEEE MASS*, 2020, pp. 338–46.

[2] B. Salehi *et al.*, “FLASH: Federated Learning for Automated Selection of High-band mmWave Sectors,” *IEEE INFOCOM*, 2022.

[3] A. Klautau *et al.*, “5G MIMO Data for Machine Learning: Application to Beam-Selection Using Deep Learning,” *2018 Info. Theory and Applications Wksp.*, 2018.

[4] “Multimodal Fusion for NextG V2X Communications”; https://genesys-lab.org/multimodal-fusion-nextg-v2xcommunications.

[5] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, “Millimeter Wave Base Stations with Cameras: Vision-Aided Beam and Blockage Prediction,” *IEEE VTC*, 2020.

[6] M. Dias *et al.*, “Position and LIDAR-Aided mmWave Beam Selection Using Deep Learning,” *IEEE SPAWC*, 2019.

[7] W. Roh *et al.*, “Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results,” *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, p. 106–13.

[8] F. Liu and C. Masouros, “Hybrid Beamforming with Subarrayed MIMO Radar: Enabling Joint Sensing and Communication at mmWave Band,” *IEEE ICASSP*, 2019.

[9] E. Soltanaghaei *et al.*, “Millimetro,” *ACM MOBICOM*, 2021.

[10] T. Koike-Akino *et al.*, “Fingerprinting-Based Indoor Localization with Commercial MMWave WiFi: A Deep Learning Approach,” *IEEE Access*, vol. 8, 2020, pp. 84,879–92.

[11] J. Palacios *et al.*, “LEAP: Location Estimation and Predictive Handover with ConsumerGrade mmWave Devices,” *IEEE INFOCOM*, 2019.

[12] M. Alrabeiah *et al.*, “ViWi: A Deep Learning Dataset Framework for Vision-Aided Wireless Communications,” *IEEE VTC*, 2019.

[13] R. Heinzler *et al.*, “Weather Influence and Classification with Automotive Lidar Sensors,” *IEEE Intelligent Vehicles Symp.*, 2019, pp. 1527–34.

[14] D. Steinmetzer, D. Wegemer, and M. Hollick, “Talon Tools: The Framework for Practical IEEE 802.11ad Research,” 2018; https://seemoo.de/talon-tools/.

[15] R. Livni, S. Shalev-Shwartz, and O. Shamir, “On the Computational Efficiency of Training Neural Networks,” *ACM NIPS*, 2014.

Biographies

Jerry Gu (gu.je@northeastern.edu) is pursuing a Ph.D. degree in computer engineering at Northeastern University, Boston, Massachusetts, where he received his M.S. (2021) in computer engineering. His current research focuses on the use of ML in wireless communications, including multimodal data fusion and RF fingerprinting.

Batool Salehi (salehihikouei.b@northeastern.edu) is pursuing a Ph.D. degree in computer engineering at Northeastern University. She received her M.S. (2019) in electrical engineering from the University of Tehran, Iran. Her current research focuses on mmWave beamforming, IoT, and the application of ML in the domain of wireless communication.

Debashri Roy (d.roy@northeastern.edu) received her Ph.D. degree (2020) in computer science from the University of Central Florida. She is currently a postdoctoral fellow at Northeastern University. Her research interests are in the areas of experiential AI and ML in wireless communication.

Kaushik Roy Chowdhury [M’09, SM’15] (krc@ece.neu.edu) is a professor at Northeastern University. His current research interests involve systems aspects of networked robotics, machine learning for agile spectrum sensing/access, wireless energy transfer, and large-scale experimental deployment of emerging wireless technologies.