Federated mmWave Beam Selection Utilizing LIDAR Data

Mahdi Boloursaz Mashhadi*, Mikolaj Jankowski*, Tze-Yang Tung, Szymon Kobus, Deniz Gündüz
Dept. of Electrical and Electronic Eng., Imperial College London, UK

***Abstract*—Efficient link configuration in millimeter wave (mmWave) communication systems is a crucial yet challenging task due to the overhead imposed by beam selection on the network performance. For vehicle-to-infrastructure (V2I) networks, side information from LIDAR sensors mounted on the vehicles has been leveraged to reduce the beam search overhead. In this letter, we propose distributed LIDAR aided beam selection for V2I mmWave communication systems utilizing federated training. In the proposed scheme, connected vehicles collaborate to train a shared neural network (NN) on their locally available LIDAR data during normal operation of the system. We also propose an alternative reduced-complexity convolutional NN (CNN) architecture and LIDAR preprocessing, which significantly outperforms previous works in terms of both the performance and the complexity.**

***Index terms*— Federated learning, mmWave beam selection, LIDAR.**

I. Introduction

Millimeter wave (mmWave) is a promising technology for high data rate vehicular communications. However, efficient beam selection in mmWave vehicle-to-infrastructure (V2I) networks is challenging due to the overhead imposed by the beam search process. Recently it was shown that side information from sensors mounted on the connected vehicles can be exploited to reduce the beam-selection overhead for mmWave links. The position information from vehicles is used in [1]–[4], while out-of-band measurements are used in [5], [6] for efficient mmWave beam selection. Information from a radar located in the infrastructure is shown to be beneficial for mmWave link establishment in [7].

The use of light detection and ranging (LIDAR) technology is considered in [8], [9]. LIDAR uses a laser to scan the environment and generates a three-dimensional (3D) image with pixels indicating relative positions from the sensor [10]. Data from the LIDAR sensors mounted on vehicles can be exploited for improved beam-selection in mmWave V2I communications. On the other hand, the lack of analytical models that can relate LIDAR outputs to mmWave channels motivates employing a neural network (NN)-based approach to this problem. In [8], [9], a NN architecture is trained over simultaneous LIDAR and ray-tracing channel datasets with a top-K classification accuracy metric to identify K beam directions that include the beam pair with the best channel condition between the vehicle and the base station (BS) with the highest probability.

The approach in [8], [9] is distributed, in the sense that, each vehicle uses the trained NN on the measurements from its own LIDAR sensor to infer its top-K beam directions. It was shown in [9] that such a distributed approach outperforms centralized beam selection, where a NN at the BS infers the best beams for all the vehicles in its coverage area either by combining LIDAR data from all the vehicles or using a single LIDAR sensor mounted at the BS. Although the NN performs beam selection inference in a distributed fashion in [8], [9], it is trained offline on LIDAR and channel measurements from all the vehicles gathered in a centralized dataset. However, in a practical scenario, gathering a large centralized dataset of individual LIDAR measurements from connected vehicles is challenging as it requires communicating a large amount of LIDAR point cloud data over the uplink channel. Note also that a separate NN needs to be trained for the coverage area of each BS as the trained NN is site-specific, and will not perform well even for the same site if significant changes occur in the scattering environment. Therefore, continuous recollection of up-to-date LIDAR data and retraining or fine-tuning the NN weights is necessary during normal operation of the system. This means that a centralized approach imposes a continuous overhead on the system for transmitting up-to-date LIDAR measurements to the BS.

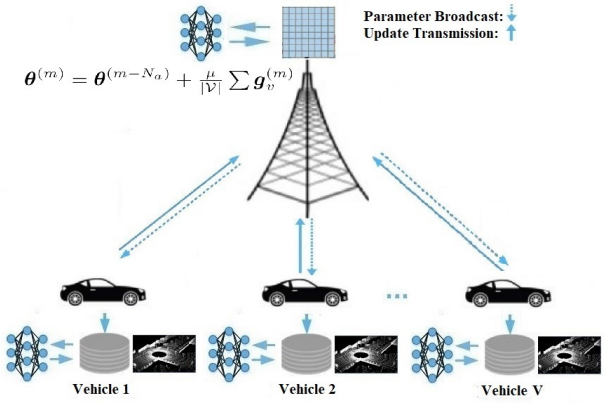

Fig. 1: The proposed federated LIDAR-based beam selection scheme.

In this work, we propose fully distributed LIDAR-aided beam selection in V2I mmWave communication systems. In the proposed approach, both the inference and training of the NN is performed in a distributed fashion at the vehicles in the coverage area of the BS. We propose a three-phase procedure, which enables the vehicles to periodically collect up-to-date data and retrain or fine-tune the NN in a federated manner during normal operation of the system. Federated training helps avoid the large communication overhead that would be imposed by the transmission of LIDAR measurements from the connected vehicles to the BS to gather a centralized dataset for offline training. After the training phase, each vehicle leverages the trained NN and its locally available LIDAR data to infer a subset of beams that are most likely to contain the best transmitter/receiver beam pair. We also propose an alternative reduced-complexity convolutional NN (CNN) architecture along with LIDAR preprocessing which significantly outperforms previous works. The proposed architecture achieves a top-10 classification accuracy of 91.17% on the benchmark Raymobtime dataset [11], which is a significant improvement over the previous works in [8], [9], while reducing the number of floating point operations (FLOPs) and parameter complexity of the NN by factors of 100 and 55, respectively. The reduction in the number of trainable NN parameters facilitates efficient federated training of the proposed architecture during normal operation of the system with reduced communication overhead.

The rest of the paper is organized as follows: Section II presents the system model. Section III presents our proposed federated LIDAR-aided beam selection scheme. Simulation results are presented in Section IV. Finally, Section V concludes the paper. For further reproduction of the reported results, our codes are available at: https://github.com/galidor/ITU_Beam_Selection_TF

* Indicates equal contribution.

This work was supported by the European Research Council (ERC) through project BEACON (grant no 677854).

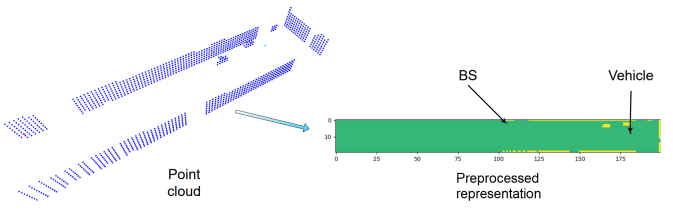

Fig. 2: Preprocessing of the LIDAR point cloud.

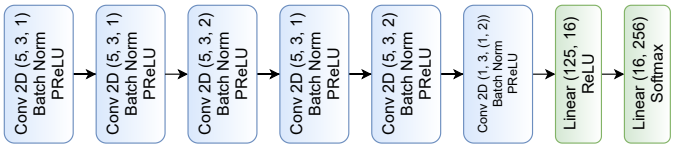

Fig. 3: The proposed model architecture.

II. System Model

We consider a downlink orthogonal frequency division multiplexing (OFDM) mmWave system, where a BS located on the street curb serves connected vehicles in its coverage area over N_c subcarriers. The BS and the vehicles are equipped with N_t and N_r antennas, respectively. Denote by $\mathbf{H}_n$ the downlink channel matrix from the BS to a vehicles over the n'th subcarrier. We assume that both the BS and the vehicles have antenna arrays with only one radio frequency (RF) chain and fixed beam codebooks and apply analog beamforming. We assume beam codebooks $\mathcal{C}_t = \{\mathbf{f}_i\}_{i=1}^{C_t}$ and $\mathcal{C}_r = \{\mathbf{w}_j\}_{j=1}^{C_r}$ at the transmitter and the receiver sides, respectively.

Utilizing a pair $(i,j) \in \mathcal{C}_t \times \mathcal{C}_r$ of precoder and combiner vectors, the resulting channel gain at subcarrier n is $\mathbf{w}_j^H \mathbf{H}_n \mathbf{f}_i$, where $(\cdot)^H$ denotes the conjugate transpose. For the (i,j) pair, the normalized signal power over all subcarriers is given by

$$y_{ij} = \sum_{n=1}^{N_c} |\mathbf{w}_j^H \mathbf{H}_n \mathbf{f}_i|^2. \tag{1}$$

Hence, the optimum beam label is $b^* = (i^*, j^*) = \underset{(i,j)}{\text{argmax}}\ y_{ij}$. Without any side information, the transmitter and reciever would search through all $C_t C_r$ beam pairs to identify b^*. Our goal is to infer a small subset of K beam pairs $\mathcal{S} = \{(i_k, j_k)\}_{k=1}^{K} \subset \mathcal{C}_t \times \mathcal{C}_r$ using the available position and LIDAR data, such that $b^* \in \mathcal{S}$. This results in a reduction of $\frac{K}{C_t \times C_r}$ in the search space for beam selection. In the next section, we propose a novel NN architecture as well as a federated training approach for top-K beam classification from simultaneous position and LIDAR data.

III. Federated Beam Selection Utilizing LIDAR Data

We propose a novel data-driven beam selection scheme, where connected vehicles in the coverage area of a BS collaborate to train a shared NN for top-K beam classification using their position and LIDAR data in a distributed manner. Collaborative training is orchestrated by the BS, and takes place during normal operation of the network as depicted in Fig. 1.

A. Three-Phase Network Operation

Our proposed solution consists of three network operation phases: (i) data collection phase, (ii) federated training phase, and (iii) distributed inference phase.

During phase (i), a subset of connected vehicles in the coverage area of the BS, denoted by $\mathcal{V} = \{v\}_{v=1}^{V}$, each acquires a local dataset $\mathcal{D}_v = \{(\mathcal{P}_v, \mathcal{B}_v)\}_{v \in \mathcal{V}}$, where $\mathcal{P}_v = \{P_i\}_{i=1}^{|\mathcal{D}_v|}$ contains instances of the point cloud P_i recorded by the LIDAR sensor and $\mathcal{B}_v = \{b_i^*\}_{i=1}^{|\mathcal{D}_v|}$ contains the corresponding best beam pair labels $b_i^* \in \mathcal{C}_t \times \mathcal{C}_r$, i.e., index of the best beam pair. Note that during this phase the connected vehicles employ exhaustive beam search to identify the optimal beam pair. Although brute-force beam search in this phase imposes an overhead on the network, it provides accurate beam labels required for training of the NN.

Algorithm 1: FedAvg for LIDAR-assisted beam selection

Init: Initial parameters $\boldsymbol{\theta}_v^{(0)} = \boldsymbol{\theta}^{(0)}$, $\forall v \in \mathcal{V}$.
for *each* $m = 1, 2, \ldots$ **do**
  Each vehicle performs a local epoch using mini-batch gradient decent iterations according to (3) ;
  if m *is an integer multiple of* N_v **then**
    Each vehicle v sends $\boldsymbol{g}_v^{(m)} = \boldsymbol{\theta}_v^{(m)} - \boldsymbol{\theta}_v^{(m-N_v)}$ to BS;
    BS computes $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m-N_v)} + \frac{\mu}{|\mathcal{V}|} \sum \boldsymbol{g}_v^{(m)}$;
    BS distributes $\boldsymbol{\theta}^{(m)}$ such that $\boldsymbol{\theta}_v^{(m)} = \boldsymbol{\theta}^{(m)}$, $\forall v \in \mathcal{V}$;
  end
end
Output: Trained $\boldsymbol{\theta}^{(m)}$ shared among all vehicles.

During phase (ii), the vehicles collaborate in a federated learning scheme to train or fine-tune a shared NN for top-K beam classification as depicted in Fig. 1. In particular, the vehicles employ federated averaging (FedAvg) [12], where a global model is sent to the vehicles by the BS at each round, and the vehicles perform mini-batch stochastic gradient descent (SGD) updates based on their local datasets. The local updates are aggregated by the BS, and used to update the global model for the next round. The duration of this phase is propotional to the number of global aggregation rounds required to train the model, denoted by N_a. Note that, the vehicles train a single site-specific NN, which learns the statistical characteristics of the coverage area of the BS for efficient beam selection.

Finally, in phase (iii), any vehicle in the coverage area of the BS utilizes the trained NN on its local LIDAR data to infer the K beams and reduce the beam search overhead. Note that, in this phase, the BS can use a low frequency control channel to transmit the trained NN model to any new vehicle entering its coverage area.

Each BS in a large network can orchestrate training of a site-specific NN for its own coverage area following the above three phases. This three-phase process continues periodically to enable updating the NN parameters to adapt to the changes in the environment.

B. LIDAR and Location Preprocessing

For each scene, the LIDAR sensor mounted on each vehicle outputs a point cloud $P = \{(x_p, y_p, z_p)\}_{p=1}^{|P|}$ representing obstacles measured by the LIDAR sensor. Each vehicle v also has its own location information (x_v, y_v, z_v), and the BS location (x_{BS}, y_{BS}, z_{BS}), which is broadcast to all the vehicles. We preprocess this data to obtain a tensor of fixed size, which contains both the location and LIDAR data and is input to the NN for each scene.

To reduce both the NN dimension and the computation load, we propose a two-dimensional (2D) representation of the LIDAR measurements, where we partition the coverage area of the BS into a grid of equal-size square cells from the top view. We define the corresponding 2D tensor L, where the cells containing the vehicle and the BS are set to -1 and -2, respectively, while each of the remaining cells is populated with a 1 if it accommodates at least one of the cloud points, and with a 0 otherwise. We remark that this 2D representation discards the height data along the z-axis, resulting in a significant reduction in the input size, and hence, the complexity of the NN, which in turn reduces the communication overhead for federated training. Moreover, we observed through experiments that the proposed 2D representation even improves the

TABLE I: Comparison between the proposed NN architecture and the baseline in [8], [9], both trained in a centralized manner.

Model	Top-10 accuracy	Top-10 throughput ratio	FLOPs	# of NN parameters, $\lvert\boldsymbol{\theta}\rvert$
Baseline [8], [9]	$83.92 \pm 0.93\%$	$86.15 \pm 0.82\%$	179.01×10^6	403677
Proposed centralized	$91.17 \pm 0.28\%$	$94.78 \pm 0.61\%$	1.72×10^6	7462

performance in comparison with a 3D representation. Fig. 2 illustrates this preprocessing scheme.

C. NN Architecture

Our NN architecture consists of 6 convolutional layers followed by 2 linear layers as depicted in Fig. 3. In the convolutional layers, we vary the value of stride between 1 and 2, depending on whether we intend to downscale the intermediate features, or not. We apply batch normalization and parametric rectified linear unit (PReLU) activation after each convolutional layer. The first linear layer is followed by rectified linear unit (ReLU) activation, and softmax is used at the output to obtain the predictions. To achieve better generalization, convolutional layers downscale the features and ensure that only essential information is preserved. This helps avoid overfitting to the training data. Note that, to reduce the communication overhead for federated training, we have minimized the trainable model parameters utilizing a convolutional structure with limited kernel sizes. We denote the NN model function by $\pi(L;\boldsymbol{\theta})$, which is a vector of length C_tC_r at the softmax output. L is the preprocessed LIDAR and location input while $\boldsymbol{\theta}$ denotes the trainable NN parameters. The best beam is predicted as $\hat{b^*} = \underset{b\in\mathcal{S}}{\arg\max}\, y_b$, where the prediction set $\mathcal{S}$ is given by the top-K softmax outputs.

D. Federated Training

Due to the individual characteristics of a specific vehicle (e.g., its trajectory, dimensions, speed, etc.), its local dataset may not capture all the subtleties of the coverage area. In such cases, the NN trained solely on a local dataset $\mathcal{D}_v$ is highly biased and may not operate reliably for other vehicles entering the coverage area of the BS. We exploit the fact that, while each vehicle may capture a limited amount of training data that is biased towards its own specific circumstances, the overall dataset captured by several vehicles (i.e., $\mathcal{D}_U = \{\mathcal{D}_v\}_{v=1}^{V}$) within the coverage area of the BS is more diverse to allow training a generalizable NN model for the coverage area in consideration. On the other hand, gathering a large dataset of LIDAR measurements from various connected vehicles for centralized training at the BS increases the communication overhead, particularly due to the large size of LIDAR point cloud measurements.

Based on the above insight, in our proposed scheme, the connected vehicles collaborate to train a single NN architecture on the overall dataset captured by all the vehicles withing the cell area via the FedAvg algorithm [12].

To train our NN classifier we use the empirical cross entropy loss, hence the local loss calculated at vehicle v is given by

$$\psi_v(\boldsymbol{\theta},\mathcal{D}_v) = -\frac{1}{|\mathcal{D}_v|}\sum_{i=1}^{|\mathcal{D}_v|}\log[\pi\left(L_i;\boldsymbol{\theta}\right)]_{b_i^*}, \tag{2}$$

where $[\pi]_b$ denotes the b'th element of the model's softmax output. Each connected vehicle performs mini-batch SGD iterations to update its local vector of model parameters, denoted by $\boldsymbol{\theta}_v$, via

$$\boldsymbol{\theta}_v^{(l)} = \boldsymbol{\theta}_v^{(l-1)} - \rho_l \nabla\psi_v(\boldsymbol{\theta}_v^{(l-1)}, \{(b_{i_l}, L_{i_l})\}_{i_l\in 1,\dots,|\mathcal{D}_v|}), \tag{3}$$

where l is the local iteration index, $\rho_l > 0$ is the local step-size, and the set $\{(b_{i_l}, L_{i_l})\}_{i_l\in 1,\dots,|\mathcal{D}_v|}$ is a mini-batch of the local dataset with $i_l \in 1,\dots,|\mathcal{D}_v|$. The training consists of N_v local epochs at each vehicle (i.e. N_v cycles of training on the vehicle's local dataset) and N_a aggregation rounds at the BS as summarized in Algorithm 1.

Such distributed learning orchestrated by the BS during phase (ii) requires the vehicles to periodically exchange and synchronize their local model parameters $\boldsymbol{\theta}_v$ through reliable low-rate communications with the BS. This imposes an overhead of communicating $O_{UL} = V \times N_a \times |\boldsymbol{\theta}|$ float32 variables in the uplink and $O_{DL} = N_a \times |\boldsymbol{\theta}|$ in the downlink channel. Minimizing the number of trainable parameters $|\boldsymbol{\theta}|$ is hence critical to reduce the communication overhead during phase (ii) of the network operation.

IV. Numerical Evaluations

We provide numerical evaluations on the benchmark Raymobtime datasets [11], where we train the models on samples from dataset s008 and test on those from s009 (refer to [11], [13] for details on these datasets e.g. locations, frequencies, etc.). For performance comparison, we use the top-K classification accuracy defined as the probability of correctly identifying the optimal beam pair within the top-K output of the network, and the top-K throughput ratio, R, defined as $R \triangleq (\sum_{t=1}^{T}\log_2(1+y_{\tilde{i}\tilde{j}}))/(\sum_{t=1}^{T}\log_2(1+y_{i^*j^*}))$, where T is the number of test samples, and (i^*, j^*) and $(\tilde{i},\tilde{i})$ denote the optimum beam pair index and the best beam pair within the top-K prediction set $\mathcal{S}$, respectively.

In Table I, we compare the performance of the NN architecture presented in Subsection III-C with the baseline architecture proposed in [8], [9], both trained in a centralized manner. In this experiment, we trained our model using the Adam optimizer [14] with an initial learning rate of 10^{-3} and batch size of 16, and train the models for 20 epochs. Besides the learning rate adjustment imposed by the Adam optimizer, we further reduce the learning rate by a factor of 10 after the 10th epoch. In Table I, we present 95% confidence intervals for Top-10 accuracy and throughput ratio of the models calculated from 10 Monte Carlo simulations.

According to Table I, our proposed architecture not only outperforms those in [8] and [9] in terms of both the top-10 accuracy and the throughput ratio, but also significantly reduces the complexity of the model. Our architecture reduces the FLOPs and the number of trainable parameters roughly by factors of 100 and 55, respectively. Such a significant reduction in the number of trainable model parameters is specifically desirable in federated training as it leads to a significant reduction of the communication overhead.

Remember that the beam search complexity of these schemes depends on K, the size of the prediction set. Figure 4 plots the top-K accuracy and throughput ratio for the proposed and baseline architectures as a function of K, when trained in a centralized fashion. It is observed that, our proposed model architecture significantly outperforms [8], [9], e.g., to achieve a throughput ratio $R \geq 90\%$, our proposed model architecture requires $K \geq 3$ while the baseline needs $K \geq 16$. This is more than 5 times reduction in the required search space for beam selection. Also, the proposed architecture can achieve close to 80% of the optimal throughput with $K = 1$; that is, with no beam search at all.

We next evaluate the performance of our proposed federated beam selection scheme. To generate the local dataset at each connected vehicle v, we choose $|\mathcal{D}_v| = 11000/V$ samples from the training set s008 uniformly at random, where 11000 is the total number of samples in s008. We use mini-batch SGD with an initial learning rate

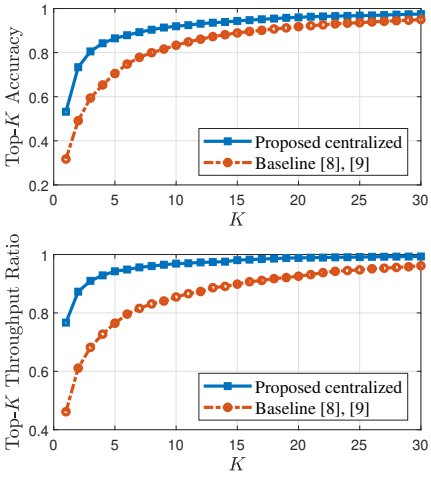

Fig. 4: Top-K accuracy and throughput ratio as a function of K.

of 0.2 and exponential rate decay of 0.001 with a batch size of 16 for local optimization at the vehicles. We set the learning rate $\mu = 0.2$ for aggregation at the BS.

We provide the performance tradeoffs for our proposed federated beam selection scheme in Table II. The notation $(N_a)^{0.88}$ in this table represents the number of global aggregation rounds required for the training to achieve a top-10 accuracy larger than 88%. This is an important measure as it determines the communication overhead required to train the model to the specified accuracy. Notations $(O_{DL})^{0.88}$ and $(O_{UL})^{0.88}$ used in this table represent this overhead in terms of the number of float32 variables needed to be communicated over the downlink and uplink channels, respectively.

According to Table II, the number of aggregation rounds required to achieve top-10 accuracy larger than 88% increases when more vehicles take part in federated training. A larger $(N_a)^{0.88}$ increases the communication overhead and duration of the training phase. On the other hand, the duration of the data collection phase decreases with V. This is because we keep the total number of samples the same across all cases, and hence, increasing V means that each vehicle needs to collect fewer samples. Here, we assume that there are always V vehicles in the cell that can participate in the training process. This leads to a tradeoff between the duration of the data collection and training phases, and the communication overhead. For practical deployment, the number of vehicles participating in federated training should be decided according to the requirements of the system and the amount of communication overhead that can be afforded. We note here that, thanks to our simple NN architecture, which only has $|\boldsymbol{\theta}| = 7462$ trainable parameters, the maximum communication overhead required for federated training (i.e., $1620 \times 7462 \sim 1.2 \times 10^7$ float32 communications for $V = 20$, $N_v = 1$) is orders of magnitude smaller than the overhead that would be imposed by offloading the LIDAR point clouds to the BS for centralized training (i.e., $\sim 4 \times 10^9$ float32 communications for samples in s008).

The last column in Table II reports the final top-10 accuracy achieved for each number of vehicles V and local epochs N_v. This column shows a slight performance degradation when more users take part in federated training. This is due to the limited number of training samples available to each vehicle when the same number of training samples (e.g., 11K samples available in s008) are distributed among more vehicles. Increasing N_v in this case tends to overfit to local datasets, which do not efficiently represent the true distribution of the data leading to some performance degradation. This can be mitigated if more data can be collected by each vehicle.

TABLE II: Performance tradeoffs for federated beam selection.

V	N_v	$(N_a)^{0.88}$	$(O_{DL})^{0.88}$	$(O_{UL})^{0.88}$	Top-10 Acc.
5	1	19	$19\|\theta\|$	$95\|\theta\|$	90.12%
	2	13	$13\|\theta\|$	$65\|\theta\|$	90.34%
	5	10	$10\|\theta\|$	$50\|\theta\|$	89.92%
10	1	31	$31\|\theta\|$	$310\|\theta\|$	89.77%
	2	22	$22\|\theta\|$	$220\|\theta\|$	89.16%
	5	15	$15\|\theta\|$	$150\|\theta\|$	88.64%
20	1	81	$81\|\theta\|$	$1620\|\theta\|$	88.81%
	2	48	$48\|\theta\|$	$960\|\theta\|$	88.53%
	5	NA	NA	NA	87.33%

V. Conclusions

We have studied efficient link configuration in mmWave V2I communication networks, and considered exploiting side information in the form of LIDAR and position data in a supervised learning scheme to reduce the beam search overhead. In this letter, we first proposed LIDAR preprocessing and a convolutional NN architecture that improves the state-of-the-art classification accuracy with a significantly reduced model complexity. We have then proposed a federated training scheme that enables connected vehicles to collaboratively train a shared NN on their locally available LIDAR data. Once the NN is collaboratively trained, any vehicle entering the coverage area of the BS can employ it to reduce the beam search overhead.

References

[1] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," *IEEE Trans. on Vehicular Technology*, vol. 67, no. 5, pp. 4042–4058, 2018.

[2] W. B. Abbas and M. Zorzi, "Context information based initial cell search for millimeter wave 5G cellular networks," in *European Conf. on Networks and Comms. (EuCNC)*, 2016, pp. 111–116.

[3] J. C. Aviles and A. Kouki, "Position-aided mm-Wave beam training under NLOS conditions," *IEEE Access*, vol. 4, pp. 8703–8714, 2016.

[4] A. Loch, A. Asadi, G. H. Sim, J. Widmer, and M. Hollick, "mm-Wave on wheels: Practical 60 GHz vehicular communication without beam training," in *Int'l Conf. on Communication Systems and Networks (COMSNETS)*, 2017, pp. 1–8.

[5] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with eyes closed: Mm-Wave beam steering without in-band measurement," in *IEEE Conf. on Computer Comms. (INFOCOM)*, 2015, pp. 2416–2424.

[6] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. on Wireless Comms.*, vol. 17, no. 2, pp. 1038–1052, 2018.

[7] N. González-Prelcic, R. Méndez-Rial, and R. W. Heath, "Radar aided beam alignment in mmWave V2I communications supporting antenna diversity," in *Info. Theory and Apps. Workshop (ITA)*, 2016, pp. 1–7.

[8] A. Klautau, N. González-Prelcic, and R. W. Heath, "LIDAR data for deep learning-based mmWave beam-selection," *IEEE Wireless Comms. Letters*, vol. 8, no. 3, pp. 909–912, 2019.

[9] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, "Position and LIDAR-aided mmWave beam selection using deep learning," in *IEEE 20th Int'l Workshop on Signal Proc. Adv. in Wireless Comms. (SPAWC)*, 2019, pp. 1–5.

[10] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Comms. Magazine*, vol. 54, no. 12, pp. 160–167, 2016.

[11] RAYMOBTIME, accessed November 2020, https://www.lasse.ufpa.br/raymobtime/.

[12] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th Int'l Conf. on Artificial Intelligence and Statistics*, 2017, pp. 1273–1282.

[13] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in *Info. Theory and Apps. Workshop (ITA)*, 2018, pp. 1–9.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.