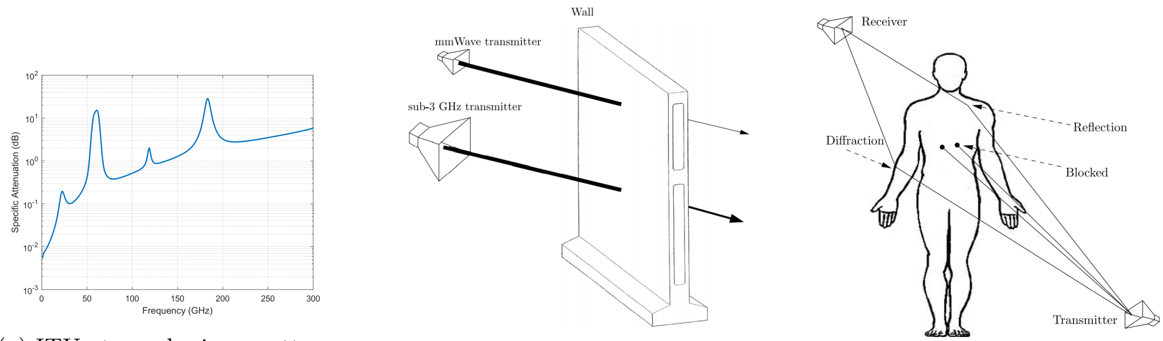
ITU-Challenge-ML5G-PHY

Zecchin Matteo
Team : `BEAMSOUP`
`zecchin@eurecom.fr`

December 11, 2020

1 mmWave communication and the beam selection problem

mmWave communications (30-300GHz) constitute a promising solution to overcome communication bottlenecks of the over-exploited sub-6GHz spectrum. By harnessing higher frequency spectrum, mmWaves unlock a wide and unused frequency band that seems to bring 5G key performance indicators (KPIs) within reach. On the other hand, mmWave communications give rise to many technical challenges due to hostile physical phenomena at high frequencies, such as: higher atmospheric attenuation, penetration losses, severe blockage and many others.



(a) ITU atmospheric gas attenuation model [1]

(b) Penetration loss is proportional to the wavelenght. Image from [2]

(c) Humans, tree and other small object can block the mmWave beam. Image from [2]

On the upside, smaller wavelengths allow to design tinier antennas and, in turn, pack many more at both transmitter and receiver ends, shifting communication towards to the massive MIMO regime. This technical advantage translates into higher beamforming gains that eventually make up for the propagation losses and therefore render mmWaves a viable solution for future generation communications. Nonetheless, there still exists many open challenges to face in order to unleash the full power of mmWave communications, a prominent one consists in reducing the overhead introduced by the beam selection (or alignment) procedures.

Consider a transmitter and receiver pair, each equipped with an analog antenna arrays of N_t and N_r antennas elements respectively. Furthermore, assume a fixed and pre-established pair of DFT codebooks at both ends, denoted with $\mathcal{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_{|\mathcal{H}|} : \mathbf{h}_i \in \mathbb{C}^{N_t \times 1}\}$ for the transmitter and with $\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_{|\mathcal{D}|} : \mathbf{d}_i \in \mathbb{C}^{N_r \times 1}\}$ for the receiver. For a given channel $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, which is typically estimated or unknown at both ends, the beam selection problem consists in solving

$$\underset{(i,j)\in|\mathcal{H}|\times|\mathcal{D}|}{\text{maximize}}\; g(i,j) = \left|\mathbf{d_j}^T \mathbf{H} \mathbf{h_i}\right|$$

Namely, choosing the pair of transmitter and receiver beams that maximizes the beamformed channel gain.
Due to the imperfect knowledge of the quantity to maximize (precisely estimating $\mathbf{H}$ is not affordable in the massive MIMO regime), the beam selection is typically solved by sequential procedures [3, 6] and it is frequently mapped to a multi-armed bandit problem [4, 5]. Many efforts have been made in order to reduce the overhead associated to the beam alignment procedure. In this regard, it is

well-known that leveraging out of band information coming from sensors, not directly involved in the communication process, can reduced the degree of uncertainty about the beam-formed channels and it can greatly boost the performance of sequential search [7]. Vehicular type of communication represents one of the scenarios that is envisioned to benefit the most from this strategy; in fact the abundance of sensory data that is generated by the vehicular infrastructure (Lidar, dashcameras, GPS signals, roadside cameras, etc.) represents a great opportunity to ease the cost of antenna configuration procedures [8–10] in this context; which is infamous due to high mobility, frequent blockage and short beam coherence times. However, it is not clear how to effectively exploit combination of high dimensional sensory inputs to reduce the communication overhead, therefore the question that we address with the proposed solution is:

"Do the data-driven approaches provide a tool to extract useful patterns from high-dimensional sensory data and are they able to produce informative priors to be exploited during the beam alignment procedures?"

The answer is **affirmative**, at least from an empirical perspective, and it is given in the form of a deep learning model that makes the beam alignment overhead 5 times smaller compared to a beam selection procedure that exploits only the empirical distribution of the beams in the training dataset.

2 Proposed solution

In order to investigate the usefulness of deep learning techniques, a synthetic dataset simulating a mmWave vehicle to infrastructure communication was given. Along with ray-tracing data, LIDAR, road side cameras and GPS sensor information were provided. The latters constituted the covariates (or features) available to predict the beamformed channels gains used in the beam alignment phase. Out of the three data modalities, the proposed solution leverages mainly LIDAR maps and receivers coordinates to rank the mmWave beams according to their expected channel gains. LIDAR data becomes a very relevant piece of information in the beam selection process as it contains locations of potential reflectors or blockages in the proximity of the receiver, that are not captured by the low-dimensional GPS information and are crucial in NLoS conditions. The importance of LIDAR data motivates the development of a *customized* pre-processing front-end that better retains information for the NLoS scenarios. This pre-processing step is then combined with a carefully designed feature extraction block based on Convolution Neural Networks (CNN) that produces a lower dimensional representation of the vehicle surroundings. This vector is then concatenated with the output of a shallow fully connected neural network processing GPS coordinates and it is jointly processed by 3 fully connected layers that output the predicted channel gains. The model is trained end-to-end (except the pre-processing pipeline that is hand-crafted) using a loss function *which is inspired from knowledge distillation techniques (KD).* The proposed loss function not only yields better performance but also produces model with much more informative outputs. In the following we introduce the data modalities, the model architecture, the training phase and display the final performance along with the beneficial effects of the proposed loss function.

2.1 GPS data

Despite its simplicity and low dimensional nature, GPS data represents a valuable piece of information that can be used to devise robust alignment procedure [11] and to train simple, yet effective machine learning models [12]. Nonetheless, it comes with some limitations since it does not bear information on the presence of LoS, reflectors, obstacles and, in general, about the geometry of propagation environment. This missing information clearly pose some fundamentals limitation on the predictive capabilities of model that exclusively use GPS coordinates. This aspect is displayed in Fig.2 where we overlay the performance of a 1-nearest neighbor classifier to the road in which communication takes place. In some regions of the street the 1-NN classifier performs poorly and we argue that those regions corresponds to regions where NLoS conditions are more likely and information about surroundings is more important. For this reason LIDAR information is an excellent candidate to complement GPS information.

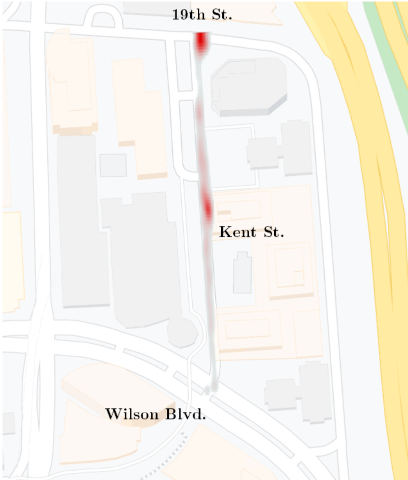
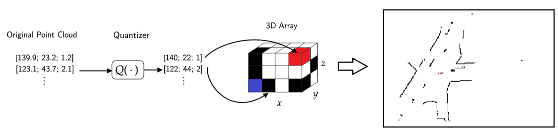


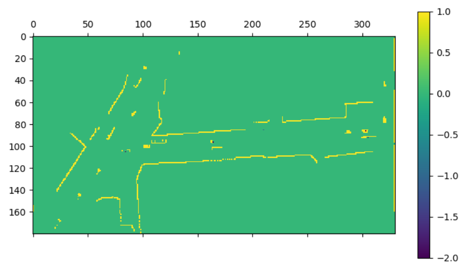
Figure 2: Color-coded magnitude of error of a 1-nearest neighbor classifier.

2.2 LIDAR

LIDAR data has the role of enhancing the prediction capabilities of GPS data providing information about position of obstacles and reflector, for this reason the processing pipeline for LIDAR data was carefully designed to retain this information. The LIDAR raw data comes in the format of a list with each entry associated to a point in 3D space. The pre-processing consists in representing the data in an image-like format to be used as input for the CNN. This is done by quantizating the coordinates of the original point cloud and filling 3D matrix where each entry is set to -1 in case of transmitter, -2 for the receiver and 1 for obstacles. This procedure is similar to the one that was used to build the baseline features [10]; however, the quantization granularity and the range of coordinates are changed in order to capture also the extrema of the road, where NLoS is more likely and beam selection is more challenging. The quantization step is 1 in all the 3 dimensions and the spanned intervals are X_{max} : 841, Y_{max} : 680, Z_{max} : 10, X_{min} : 661, Y_{min} : 350 and Z_{min} : 0. Figure 1 is an example of the output for episode 0 of s008 dataset; note that, differently from the baseline features, the vehicles at the end of the street are also captured in this representation.



(a) The Lidar point cloud is first quantized and then used to fill a 3D array that represent a discrete version of the propagation environment.



(b) Flattend version (in the Z-coordinate) of the 3D matrix resulting from the pre-processing of LIDAR data.

2.3 Image data

There is a plethora of literature about successful use cases of deep learning to classify and in general process images for different purposes. Despite this, for the specific task, image data proved to be useless, at least when processed by CNNs. Our guess is that, contrarily to standard classification tasks, we are not requiring the model to craft high level image features and use them to perform classification. We are rather implicitly asking CNNs to capitalize on a 3D geometry reconstructed from a 2D image to infer how a mmWave beam will propagate in space. This task is challenging for humans too, which can be tricked by images that leverage on perspective to induce on false

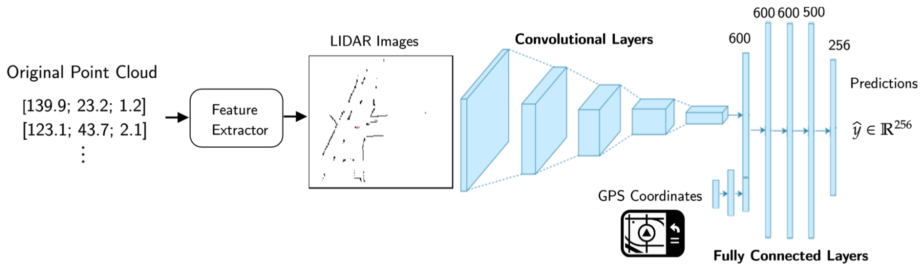


Figure 4: The proposed network architecture

assumption on what is portrayed in the picture. We argue that CNNs alone may not be the proper tool for this challenging task and using them to process images did not provide significant gains for the problem at hand. For this reason they were not used in the final solution.

2.4 Model

We now turn in describing the proposed model architecture. The deep neural network is made of 3 building blocks:

- 1 fully connected layer (FC) to create a 128 dimensional embeeding of the $[x, y, z]$ coordinates of the receiver.
- A convolutional neural network with 5 conv layers each followed by a max pooling operation and one FC layer after flattening, this block is used to process lidar data and outputs a vector of 400 entries.
- A 3 FC layers classifier taking as input the concatenated processed features by the two previous blocks. It outputs a 256 dimensional vector with the predicted normalized gains. The layers of this block have 600, 600 and 500 neurons respectively.

The neural network is depicted in Fig.4. It contains $\approx 2.5M$ parameters and therefore needs to be regularized during the training process as explained in the next subsection.

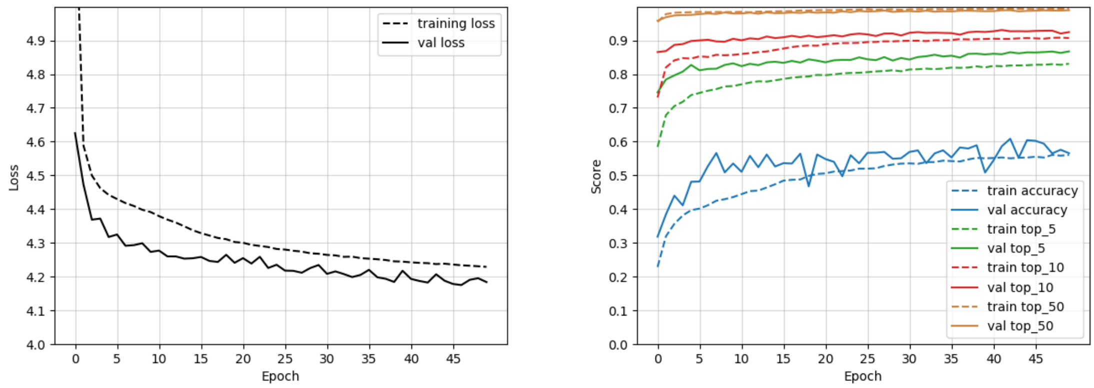
2.5 Training

The model is trained taking batches of 32 samples from the s008 dataset that contains 11194 samples. The model is trained end-to-end and for this purpose we propose a loss inspired by Knowledge Distillation techniques. Define as $\vec{y} \in [0,1]^{256}$ the vector of the normalized channel gains after beamforming and as $\vec{y}_H \in \{0,1\}^{256}$ the same vector where only the largest component is set to 1, the loss that is used to train has the following form

$$\mathcal{L}_{KD}(\theta) = (1-\beta)\sum_{\mathcal{D}^n} H(\vec{y}, f_\theta(x)) + \beta \sum_{\mathcal{D}^n} H(\vec{y}_H, f_\theta(x))$$

where H is the cross-entropy loss, $f_\theta(x)$ is the model output and $\beta \in [0,1]$. This loss is used to promote learning of soft and hard labels, so that the model not only learns to correctly classify the first beam but also to predict the other channel gains, the value β is used to balance this two driving forces during learning and is set to 0.8.

Given the size of the dataset and the number of model parameters, regularization techniques were used to prevent overfitting. In particular a $L2$ penalty term is used to prevent the configuration with large network weights. Additionally, white noise was added to the input layer in order to smooth out the prediction rule. This last regularization technique is well also justified in this setting, as it represents sensor noise and possibly induce robust predictors. We train the model for 50 epochs using Adam optimizer and we tracking top-K accuracy for $K = \{1, 5, 10, 50\}$, it tooks roughly one hour on a GTX1080Ti . We save the model with the best top-10 accuracy. As we can see from the training curves the model does not over-fit and attains good validation performance.



(a) Loss vs. Epoch

(b) Top-K accurcay vs. Epoch

2.6 Testing and the effect of the loss

The model top-K accuracy is evaluated over a set of 7k unseen samples and it compared against predictors that uses an increasing level of contextual information:

- **Random Search**: a random predictor that represent the full uninformed setting without any degree of information.
- **Statistical**: a predictor based on the empirical distribution of the beams in the training set that works by ranking beams according to their relative frequency.
- **GPS**: a 2 layer fully connected neural network that uses only GPS coordinates.
- **GPS+LIDAR**: the proposed neural network that combines the GPS and LIDAR information.

As can we see an increased amount of contextual information results in increased predictive performance and higher top-K accuracy (a steeper curve is better). To reach a 90% confidence of outputting the best beam from a sequential search we need 8 beams for GPS+LIDAR, 13 for GPS only , 43 using statistical information and 230 for a random search. Therefore the overall beam alignment overhead can be greatly reduced exploiting contextual information and leveraging high dimensional data, such as LIDAR, can enhance performance of the performing GPS predictor by 30%.

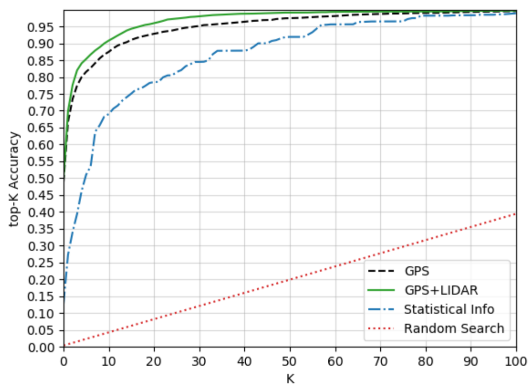
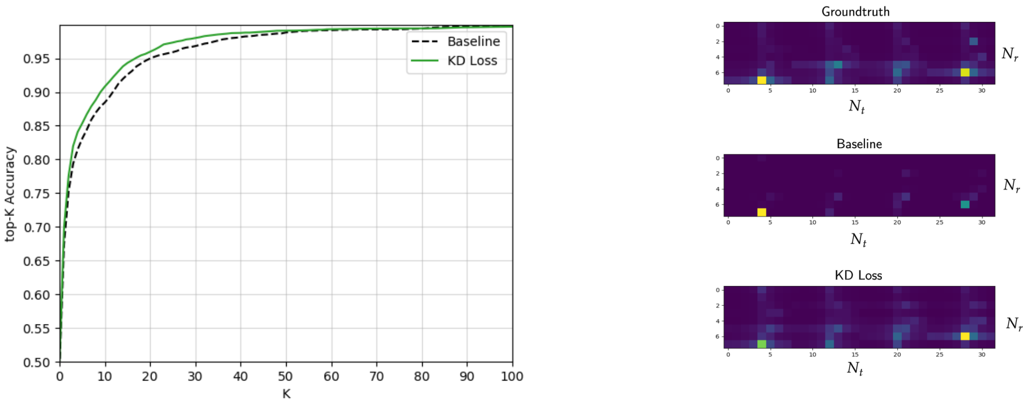


Figure 6: Top-K accuracy vs K

In Fig. 7a we report the final performance of the same model trained with the baseline loss and the proposed one, inspired by knowledge distillation technique. The use of this *costum* training objective not only determines a better final top-K accuracy but also yields model with more informative outputs that better predict the beam gains matrix. In Fig. 7b, we consider an instance from the validation set and compare the prediction of the model trained with the baseline loss and KD loss against the ground-truth. The KD loss yields prediction that better capture the true channels gains and are therefore more informative for the beam selection procedure.



(a) Final top-K accuracy

(b) Predictions

Figure 7: Loss benefits

References

[1] Radiocommunication Sector of International Telecommunication Union. Recommendation ITU-R P.676-10: Attenuation by atmospheric gases 2013.

[2] Hemadeh et al., "mm-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions, and Link-Budget"

[3] Lee, Junho, et al. "High-Resolution Hierarchical Beam Alignment with Segmented Beams." 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019.

[4] Booth, Matthew B., et al. "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications." IEEE Communications Letters 23.7 (2019): 1244-1248.

[5] Hashemi, Morteza, et al. "Efficient beam alignment in millimeter wave systems using contextual bandits." IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018.

[6] Noh, Song, Michael D. Zoltowski, and David J. Love. "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment." IEEE Transactions on Wireless Communications 16.9 (2017): 5689-5701.

[7] González-Prelcic, Nuria, et al. "Millimeter-wave communication with out-of-band information." IEEE Communications Magazine 55.12 (2017): 140-146.

[8] Wang, Yuyang, Murali Narasimha, and Robert W. Heath. "MmWave beam prediction with situational awareness: A machine learning approach." 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018.

[9] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang and R. W. Heath Jr., "5G MIMO Data for Machine Learning: Application to Beam-Selection using Deep Learning" in Information Theory and Applications Workshop (ITA)

[10] A. Klautau, N. González-Prelcic and R. W. Heath, "LIDAR Data for Deep Learning-Based mmWave Beam-Selection," in IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 909-912, June 2019.

[11] Maschietti, Flavio, et al. "Robust location-aided beam alignment in millimeter wave massive MIMO." GLOBECOM 2017-2017 IEEE Global Communications Conference. IEEE, 2017.

[12] Heng, Yuqiang, and Jeffrey G. Andrews. "Machine Learning-Assisted Beam Alignment for mmWave Systems." 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019.