LSTM-Based Beam Tracking for mmWave Vehicular Networks

Chen Wang, XXX, XXX, XXX
Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications
Beijing, China
wangchen@bupt.edu.cn

***Abstract*—The use of millimeter wave (mmWave) frequency bands for transmission can improve data transmission rate with the help of beamforming technique to overcome the high path and penetration losses. However, utilizing mmWave technology in vehicular networks is challenging, as the high mobility of vehicles results in extremely frequent beam alignment and significant overhead. In this paper, a long short-term memory (LSTM)-based beam tracking scheme was proposed for reducing overhead brought by beam alignment in mmWave Vehicular Networks, by predicting beam angles at next time step through known beam angles at a certain number of consecutive time steps as features. To train this network, a time series array antenna channel data was set up by statistical channel model using time series vehicle information generated from road traffic simulation software named "Simulation of Urban MObility (SUMO)". Simulation results show that the proposed LSTM-based scheme can reduce a significant overhead with an acceptable loss in spectral efficiency compared with exhaustive search scheme.**

***Index Terms*—beam tracking, mmWave vehicular networks, LSTM**

I. Introduction

MmWave is one of the important technology for the future development of communications [1], relies on abundant spectrum resources. In order to apply mmWave better, the high path and penetration loss caused by high frequency of mmWave need to be overcome [2]. Fortunately, it is possible to integrate mmWave phased array antennas into small chip with the help of short wavelength of mmWave, and use beamforming technique to deliver signal power in specified direction for high gain [3]. Analog beamforming, digital beamforming and hybrid beamforming are the most commonly used techniques for beamforming. Considering the digital beamforming is prohibitively complex and costly due to the requirement of a dedicated radio frequency (RF) chain for each antenna element and higher dynamic range for analog front end RF circuitry, it has not been largely adopted by both indoor and outdoor mmWave communications [4].

Unlike conventional omnidirectional antennas, array antennas are directional [5], which means that if high antenna gain needs to be achieved, it is important to ensure that the beams generated by array antennas at the receiver and transmitter are aligned well with each other. Otherwise, misalignment of the beam pairs will result in serious gain reduction and degrading performance. A hierarchical codebook design was proposed in [6], which essentially exploits binary search algorithm to reduce overhead by using different beam widths at different stages to perform beam search. This algorithm requires a high quantization level phase shifter and potentially multiple RF chains to implement, which leads to high cost. In [7], Zhang et al. proposed a Kalman filter based scheme for beam tracking using exhaustive search, whose measurements and protocol overhead increase with the number of elements in the antenna. Based on prior work, [8] exploits extended Kalman filter and requires only a single measurement, making it more suitable for beam tracking in fast-changing environments.

Providing reliable beam alignment for high speed vehicles is more challenging, due to vehicles move out of the beam coverage faster. Frequent beam alignment is required for ensuring the vehicles are always covered by high gain beams, and it will further increase system overhead, especially when high resolution codebooks are used. Fortunately, due to the sparsity of mmWave channel [9] and the correlation between the angles of arrival (AoA) and departure (AoD) of a mmWave channel and the positions of the receiver and transmitter [10], the beam angles of a vehicle traveling along a road under LOS condition can be well predicted. Wang et al. proposed a situational awareness-aided beam training solution using machine learning in mmWave Vehicle to Infrastructure (V2I) communication, which is dependent on availability of accurate localization of vehicles [11]. In practice, however, the inaccurate or incomplete situational awareness may be introduced by localization error, limited location updating frequencies, and data transmission rate of vehicles.

In this paper, a LSTM-based beam tracking scheme is proposed. By using LSTM network to learn the historical channel state information (CSI) which was obtained from exhaustive beam search, AoA and AoD of current channel between vehicle and BS can be predicted to reduce measurement overhead. Meanwhile, a road traffic simulation software is used to generate time series vehicle information (e.g. position, speed, etc.) to set up time series CSI with a statistical channel model for training the network. Compared with prior work, this paper differs in the following ways: 1) More realistic channel model; 2) Simultaneous prediction of AoA and AoD; 3) The outage judgment is based on the SNR of the receiver instead of the angle difference; 4) Fewer beam measurements required on average in a tracking round. 5) In addition, beam search is performed more realistic by using a quantized phase

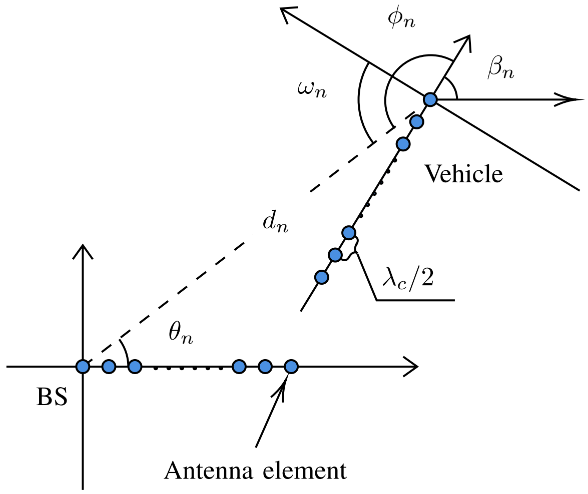


Fig. 1. Channel geometry

shifter based analog beamforming.

The following notation will be used in this paper. Matrices, vectors and scalars are denoted by bold uppercase letters (e.g. $\mathbf{A}$), bold lowercase letters (e.g. $\mathbf{a}$) and lowercase letters (e.g. a), respectively. $(\cdot)^T$ denote transpose and $(\cdot)^H$ denote conjugate transpose (Hermitian). $[\mathbf{A}]_{m,:}$, $[\mathbf{A}]_{:,n}$ and $[\mathbf{A}_{m,n}]$ denote the mth row, nth column and the mth row nth column entry of $\mathbf{A}$, respectively. $[\mathbf{a}]_n$ denote the nth entry of $\mathbf{a}$. Besides, $\|\cdot\|_2$ denote ℓ_2-norm of a vector. $\mathbb{C}$ denote the set of complex number. Gaussian, complex Gaussian, wrapped Gaussian and exponential distribution are denoted by $\mathcal{N}$, $\mathcal{CN}$, $\mathcal{WN}$ and $\mathcal{E}$, respectively.

II. System Model

A mmWave vehicular network scenario including a vehicle with M_r antenna elements as receiver and a base station (BS) with M_t antenna elements as transmitter are considered. Both of them equip with uniform linear array (ULA) of half-wave interval antenna as shown in Fig. 1, and adopted analog beamforming used quantized phase shifter which connect with single analog RF chain. The array response vector of a ULA with M half-wave interval antenna elements is given by:

$$\mathbf{a}(M,\varphi) = \frac{1}{\sqrt{M}}\left[1, e^{j\pi cos(\varphi)}, ..., e^{j\pi(M-1)cos(\varphi)}\right]^T \tag{1}$$

where φ is the arrival angle of the signal. When AoA of the vehicle is ϕ and AoD of the BS is θ, the array response vectors of both are $\mathbf{a}_r(\phi) = \mathbf{a}(M_r, \phi)$ and $\mathbf{a}_t(\theta) = \mathbf{a}(M_t, \theta)$, respectively.

The statistical 28 GHz mmWave channel model in [12] was used, which is modeled by real experimental data collected in New York City. The narrowband time-varying L subpaths channel matrix between the vehicle and BS at nth time step is $\mathbf{H}_n \in \mathbb{C}^{M_r \times M_t}$, which is given as:

$$\mathbf{H}_n = \frac{1}{\sqrt{L}}\sum_{l=1}^{L} g_{ln}\mathbf{a}_r(\phi_{ln})\mathbf{a}_t^H(\theta_{ln}) \tag{2}$$

where $g_{ln} \in \mathbb{C}$ is complex small-scale fading gain on subpath l at nth time step, ϕ_{ln} and θ_{ln} are the AoA and AoD of the lth subpath at nth time step, respectively. $\phi_{ln} = \phi_n + \Delta\phi_{ln}$ and $\theta_{ln} = \theta_n + \Delta\theta_{ln}$, where ϕ_n and θ_n is the AoA and AoD of the cluster center at nth time step, $\Delta\phi_{ln} \sim \mathcal{WN}(\delta_{rn}^2)$, $\Delta\theta_{ln} \sim \mathcal{WN}(\delta_{tn}^2)$, $\delta_{tn} \sim \mathcal{E}(\lambda_t)$ and $\delta_{rn} \sim \mathcal{E}(\lambda_r)$. λ_t and λ_r are root-mean-square (RMS) angular spread of transmitter and receiver, respectively. The complex small-scale fading gain is given by:

$$g_{ln} = \overline{g}_{ln}e^{j2\pi t_n f_{Dn,max}cos(\omega_{ln})}, \overline{g}_{ln} \sim \mathcal{CN}\left(0, 10^{-0.1PL_n}\right) \tag{3}$$

where t_n is the signal transmission time at nth time step, $f_{Dn,max} = v_n/\lambda_c$ is maximum Doppler shift at nth time step, v_n is the speed of vehicle at nth time step, λ_c is the carrier wavelength, $\omega_{ln} = \phi_{ln} - \pi/2$ is the AoA of subpath l relative to the direction of vehicle at nth time step, and PL_n is omnidirectional path loss at nth time step, which is defined as:

$$PL_n = 61.4 + 20\log(d_n) + \xi \ [dB] \tag{4}$$

where d_n is distance between the vehicle and the BS at nth time step, and $\xi \sim \mathcal{N}(0, \sigma_p^2)$. Note that only the case of single cluster under line-of-sight (LOS) condition is considered in this paper, and the cluster angle depends on the geometric position of the vehicle and the BS.

For a transmitted signal x_n from the BS, received signal y_n in the vehicle at nth time step is:

$$y_n = \mathbf{w}_n^H\mathbf{H}_n\mathbf{f}_n x_n + \mathbf{w}_n^H\mathbf{v}_n \tag{5}$$

where $\mathbf{w}_n$ and $\mathbf{f}_n$ are combining vector and beamforming vector at nth time step, respectively. $\mathbf{v}_n$ is Gaussian noise at time step n and $\mathbf{v}_n \sim \mathcal{CN}(0, \sigma_v^2\mathbf{I}_{M_r})$.

III. LSTM-based Beam Tracking Solution

A. The LSTM Network

The purpose of beam tracking is to obtain beam angles (ϕ_{ln} and θ_{ln}) with as few beam measurements as possible for reducing the overhead brought by beam search. Since the vehicle must travel in a particular direction on the road and the mmWave beams are correlated with the position of vehicles, the problem of beam tracking can be formulated as a time series regression problem after the historical CSI of vehicles is collected.

An artificial neural network named LSTM network was used to solve this regression problem. It has been implemented successfully in several fields such as machine translation, image captioning, speech recognition, and even stock prices prediction in economy. As one of the recurrent neural networks (RNNs), LSTM is generally considered to be more robust to long time series than simpler vanilla RNN implementations [13]. It is powerful for handling time series data and learning long-term relations between features, relying on three gates which are forget, input and output gate composed of neural networks (NNs) in LSTM cells.

Fig. 2 shows the structure of the network, which is a sequential model comprises two LSTM layers with 50 hidden

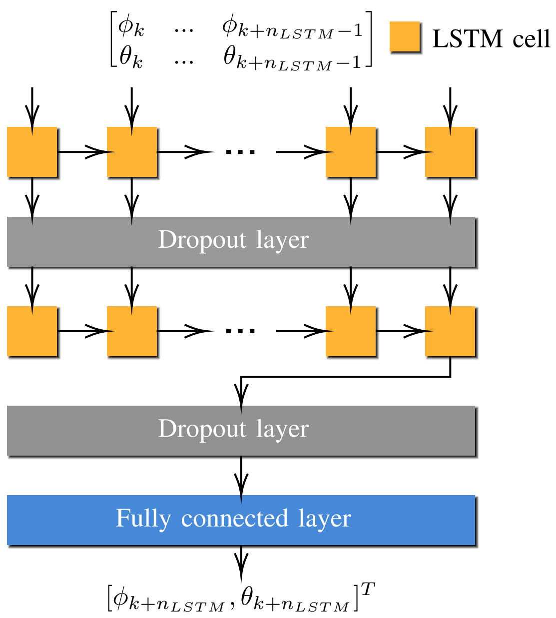


Fig. 2. Structure of network

units, two dropout layers with 0.2 probability and a fully connected layer with 2 nodes. The output of the first LSTM layer is a sequence of length n_{LSTM}, which is used as the input of the second LSTM layer. And two dropout layers are set to prevent overfitting.

B. Loss Function

Mean Square Error (MSE) is the most commonly used regression loss function for quantifying the error of prediction. In [14], considering the problem of angle ambiguity due to the periodicity of the beam angle results in unnecessary loss for large difference of beam angle when used MSE, D. Burghal et al. proposed to use a shifted cosine function as the loss function. However, when using this loss function, the training process may be unstable due to the multiple global minima, and the responses may not converge in a reasonable interval.

In addition, another problem is that the codewords whose beam angle near 0 or π have close gains at two different angles, resulting in unnecessary loss in some cases, as shown in Fig. 3. For instance, in the case of the codeword whose beam angle is π, there are same gains at 0 and π. The loss should be small, even zero, if estimated angle is 0 and predicted angle is π, although the absolute difference of them is π. A straightforward idea is taking $1-|\mathbf{a}^H(M,\theta^e)\mathbf{a}(M,\theta^p)|$ as the loss. Calculating this loss, however, requires array response vectors of estimated angle and predicted angle at the corresponding time step, which will increase the computational cost when training network. On the other hand, considering it is a complex function, it will be difficult to perform backpropagation (BP).

Hence, to avoid the problem of angle ambiguity, the AoAs and AoDs are limited in $(0,\pi]$ by simple conversion before training. And considering complexity of computation, a modified MSE is used as loss function, which is given as:

$$Loss = \frac{1}{N}\sum_{i=1}^{N}\left[(\phi_i^p-\phi_i^e)^2+(\theta_i^p-\theta_i^e)^2\right] \tag{6}$$

where N is the number of training samples, ϕ_i^e, θ_i^e, ϕ_i^p, θ_i^p, are the estimated AoA, estimated AoD, predicted AoA, and predicted AoD, respectively, for training sample i.

C. Procedure of Beam Tracking

Before exploiting the LSTM network for beam tracking, the BS demands to collect historical CSI, which are estimated by the vehicles using exhaustive beam search, for training the LSTM network. At time step k, the network takes the AoAs and AoDs from time step $(k-n_{LSTM})$ to $(k-1)$ as features and the AoA and AoD at time step k as responses. The dataset is consisted of time series AoAs and AoDs of the vehicles from initial access to disconnection.

After the LSTM network has been trained, the vehicle takes n_{LSTM} consecutive exhaustive beam search to obtain initial estimated AoAs and AoDs at the beginning of a round of beam tracking, and the network will take the estimated AoAs and AoDs to predict AoA and AoD at next time step. Then, the predicted AoA and AoD at kth time steps will be used to compose new feature which from $(k-n_{LSTM}+1)$th to kth time steps to predict the AoA and AoD at $(k+1)$ time step. For every time step, SNR under current predicted AoA and AoD will be calculated and compares with a threshold SNR_T. The network will repeat to use predicted result to compose new features to predict AoA and AoD, when the calculated SNR is greater than SNR_T. Otherwise, a new round of estimation and prediction will be repeated.

It makes more sense to use the SNR of received signal rather than the absolute difference of the angles as a threshold, because there is no way to obtain an accurate beam angle for comparing with the threshold until it is accurately measured or predicted.

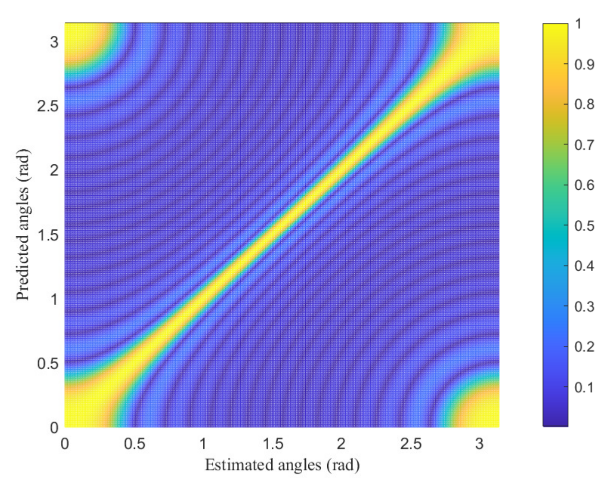


Fig. 3. Heatmap of $|\mathbf{a}^H(16,\theta^e)\mathbf{a}(16,\theta^p)|$

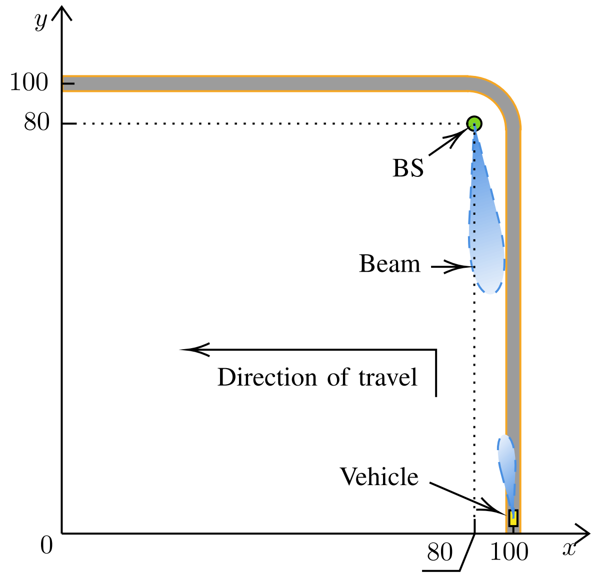


Fig. 4. Scene of mmWave vehicular networks for data generation

IV. Data Generation

A. Channel Matrix

A road traffic simulation software named "SUMO" was used to generate time series information of the vehicle in the scenario mentioned in Section II, which including a BS and a vehicle driving on curve road as shown in Fig. 4. The vehicle departs from $(0, 100)$ at first time step, and arrives $(100, 0)$ at final time step. Note that the speed of vehicle is non-linear and depends on the road conditions and behavior of the driver which defined by SUMO.

The generated vehicle time series information is composed of position, v_n and directional angle β_n of the vehicle at each time step. For nth time step, the position of vehicle is given by (x_{rn}, y_{rn}), v_n is used to determine maximum Doppler shift in (3), and β_n is used to determine ϕ_{ln} and ω_{ln}.

Once the position of the vehicle and the BS are obtained, θ_n and ϕ_n can be calculated by simple geometric methods as following:

$$\begin{cases} \theta_n = \arg(\Gamma_{xn} + j\Gamma_{yn}) \\ \phi_n = \arg(\Gamma_{xn} + j\Gamma_{yn}) - \beta_n \end{cases} \tag{7}$$

where $\Gamma_{xn} = x_{rn} - x_t$, $\Gamma_{yn} = y_{rn} - y_t$, (x_t, y_t) are coordinates of the BS, β_n is directional angle of the vehicle. Furthermore, the signal transmission time $t_n = d_n/c$ can be calculated, where the distance between the BS and the vehicle is $d_n = \sqrt{\Gamma_{xn}^2 + \Gamma_{yn}^2}$, and c is light speed. Now the channel matrix at each time step can be obtained from (2), (3), (4) and (7).

B. Beam Search

In this paper, downlink (DL) configuration scheme is considered for beam management. The beam management procedure is as follows. The reference signals are sent to the vehicle from the BS, the vehicle measures and determines the optimal beam for communication, then transmits beam reporting back to the BS.

AoAs and AoDs measured from exhaustive beam search scheme is used to train the network, rather than using the generated CSI directly. Although AoAs and AoDs are in range $[0, 2\pi]$, only the angles in range $[0, \pi]$ are considered, since the cosine function of the angles is symmetric around π. To reduce the overhead of beamforming, a discrete resolution b-bit codebook is adopted to perform beam search, where b is the number of bits for quantized phase shifters. For a b-bit codebook, it is consisted of 2^b beam patterns. In this paper, b is assumed to be $2log_2M$ for a M-element antenna terminal. For a terminal with M-element antenna, the candidate beam angles vector $\mathbf{b}_M$ is given by:

$$\mathbf{b}_M = \left[\frac{1}{2M}\pi, \frac{2}{2M}\pi, ..., \pi\right] \tag{8}$$

So, there are $2M_r$ beam patterns for the vehicle, and $2M_t$ beam patterns for the BS. Note that another codebook which achieves uniform maximum gain in all direction is not used for this solution, because its low spatial resolution near 0 and π resulting in network underfitting. The beam codebooks of the vehicle and the BS are:

$$\begin{aligned} \mathbf{W} &= [\mathbf{a}_r([\mathbf{b}_{M_r}]_1), \mathbf{a}_r([\mathbf{b}_{M_r}]_2), ..., \mathbf{a}_r([\mathbf{b}_{M_r}]_{2M_r})] \\ \mathbf{F} &= [\mathbf{a}_t([\mathbf{b}_{M_t}]_1), \mathbf{a}_t([\mathbf{b}_{M_t}]_2), ..., \mathbf{a}_t([\mathbf{b}_{M_t}]_{2M_t})] \end{aligned} \tag{9}$$

where $\mathbf{W} \in \mathbb{C}^{M_r \times 2M_r}$ is the codebook matrix of vehicle, and $\mathbf{F} \in \mathbb{C}^{M_t \times 2M_t}$ is the codebook matrix of BS. Each column of the codebook matrix represents a beam pattern, each entry in the column is phase rotation for corresponding antenna element to generate directional beam.

Assume $x_n = 1$ is sent as reference signal, the observation matrix comprises all measurement of exhaustive search now can be given by:

$$\mathbf{Y}_n = \mathbf{W}^H \mathbf{H}_n \mathbf{F} + \mathbf{V}_n \tag{10}$$

where $\mathbf{V}_n \in \mathbb{C}^{2M_r \times 2M_t}$ is Gaussian noise matrix composed of independent identically distribution (i.i.d.) elements, which are the noise part of (5). $\mathbf{Y}_n \in \mathbb{C}^{2M_r \times 2M_t}$, $[\mathbf{Y}_n]_{i,j}$ is the received signal of using $[\mathbf{W}]_{:,i}$ as combining vector and $[\mathbf{F}]_{:,j}$ as beamforming vector.

For each time step, both of the vehicle and the BS use their beam codebook to perform exhaustive beam search to find the optimal beam pair. The indexes of optimal beam pair can be obtained by solving an optimization problems as following:

$$\begin{aligned} (i_n^*, j_n^*) = \ &\underset{i_n, j_n}{\text{argmax}} \|[\mathbf{Y}_n]_{i_n, j_n}\|_2 \\ &s.t. \ i_n \in [1, 2M_r], \\ &\quad\ \ j_n \in [1, 2M_t] \end{aligned} \tag{11}$$

A simple nested loop algorithm can be exploited to solve this problem. Then, at nth time step, the optimal combining vector $\mathbf{w}_n^*$ and the optimal beamforming vector $\mathbf{f}_n^*$ are $\mathbf{w}_n^* = [\mathbf{W}]_{:,i_n^*}$

and $\mathbf{f}_n^* = [\mathbf{F}]_{:,j_n^*}$, respectively. And the estimated AoA ϕ_n^e and AoD θ_n^e used to training network are given by:

$$\begin{cases} \phi_n^e = \frac{i_n^* \pi}{2M_r} \\ \theta_n^e = \frac{j_n^* \pi}{2M_t} \end{cases} \tag{12}$$

The advantage of training with radian instead of the codeword indices is that it allows the proposed beam tracking solution to be generalized under terminals with different antenna configuration and beamforming techniques, without additional computation and transmitted data.

Note that due to the limited resolution of the beam codebook and the noise received at the vehicle, there will be some errors for the channel estimation, which is reflected in fluctuations in the estimated time series AoAs and AoDs. To reduce the fluctuations, a first order Savitzky-Golay filter [15] with frame length 11 was used to smooth the estimated AoAs and AoDs for training the LSTM network better.

V. Simulation Results

In this section, the performances of proposed LSTM-based beam tracking solution is presented by analyzing the SNR of received signal, number of measurements and outage probability. The outage probability for a simulation is given by:

$$P_o = \frac{\sum_{i=1}^{N} \mathbf{I}(SNR_i \leqslant SNR_T)}{N} \tag{13}$$

where SNR_i is the SNR of received signal for ith sample and $\mathbf{I}$ is an indicator function.

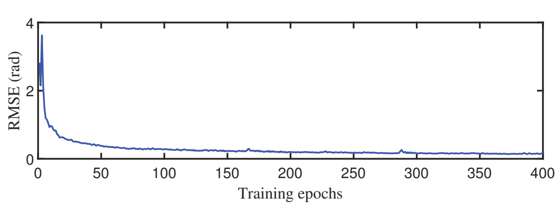
In this paper, the sample rate of series channel data is 20 Hz (50 ms interval) and there are total 613 samples for training and 401 samples for testing. Table I shows the common simulation parameters. For each simulation, parameters default to the values in Table I, except for those mentioned specifically.

Fig. 5 demonstrates the convergence of the LSTM network. The root-mean-square error (RMSE) of the LSTM network is shown in Fig. 5(a). It is observed that the RMSE rapidly converges to about 0.4 around 50th epoch, and finally converges to about 0.2 after 400 epochs. And the loss of the LSTM network is shown in Fig. 5(b). It shows that the loss decreases sharply around 20 epochs, and converges to a small value about 0.01 after 400 epochs.

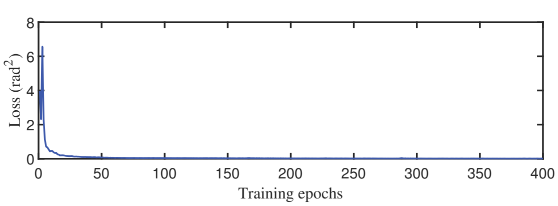
TABLE I
Simulation Parameters

Parameters	Values
Carrier frequency f_c	28×10^9 Hz
Coordinates of BS (x_t, y_t)	$(80, 80)$
Number of antenna elements for BS M_t	16
Number of antenna elements for Vehicle M_r	4
Number of subpaths L	20
Expectation of BS RMS angular spread λ_t	10.2°
Expectation of Vehicle RMS angular spread λ_r	15.5°
Path loss standard deviation σ_p	5.8 dB
Noise variance of received signal σ_v^2	-140 dBW
Threshold of SNR SNR_T	5 dB
Lookback of LSTM n_{LSTM}	5
Optimizer of LSTM	Adam
Training Epochs	400
Learning rate	0.05

Fig. 6 shows the performances of schemes versus noise variance of received signal. The vehicle using exhaustive search scheme uses exhaustive search method to find the optimal beam angles at each time step. It illustrates that for both exhaustive search and LSTM-based scheme, the average SNR of received signal is decreasing and the probability of outages is increasing with the increasing of σ_v^2. The difference in performance between the LSTM-based scheme and the exhaustive search scheme is small before $\sigma_v^2 = -110$, up to about 6 dB. When $\sigma_v^2 = -140$ dBW, the average SNR of received signal in exhaustive search is 55.62 dB and in LSTM-based scheme is 51.93 dB, there is only 3.69 dB loss in SNR, and about 6.7% loss in spectral efficiency.



(a) RMSE of the LSTM Network



(b) Loss of the LSTM Network

Fig. 5. Convergence of the LSTM network

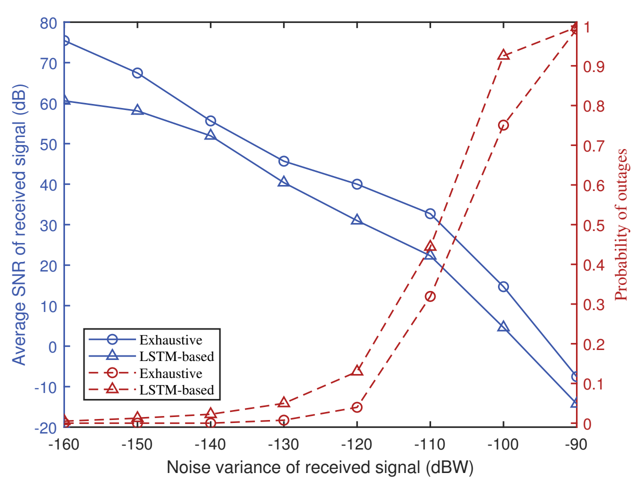


Fig. 6. Average SNR and outages probability versus σ_v^2

TABLE II
COMPARISONS OF THE NUMBER OF MEASUREMENTS

Schemes	Complexity	Number of Measurements
LSTM-based	$N_s + N_f M_t M_r$[1]	721
Exhaustive search	$N M_t M_r$	25984

[1] N_s is the number of successful prediction samples and $N_f = N - N_s$.

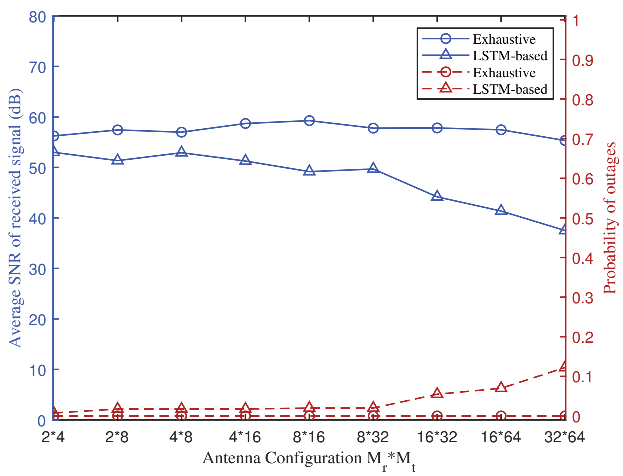


Fig. 7. Average SNR and outages probability versus $M_t \times M_r$

Table II shows the comparison of the number of measurements in different schemes. Compared with the exhaustive search scheme, proposed LSTM-based scheme can reduce about 97.2% measurement overhead. Considering the significant reduction in measurement overhead, the 6.7% loss of spectral efficiency in Fig. 6 seems to be acceptable. Note that the proposed LSTM-based scheme obtain initial beam angles for prediction by exhaustive search method in this paper, and it is obvious that the measurement overhead can be further reduced by utilizing a less complex beam search method (e.g. hierarchical method).

Fig. 7 shows the performances of schemes versus antenna configuration $M_t \times M_r$. It is illustrates that the performances of LSTM-based scheme is decreasing with the increasing of antennas scale. The reason is the beamwidths of both are so small that reducing the tolerance for prediction error and making it difficult for LSTM-based scheme to accurately align with each other. It is observed that there are fluctuations in both SNR and outages probability, due to the randomness of the channel which is caused by the statistical characteristics (e.g. angular spread, noise variance, etc.) of the channel and the vehicle movement.

VI. CONCLUSIONS

This paper studies the beam tracking problem in mmWave vehicular network with ULA. A LSTM-based beam tracking scheme is proposed to reduce the beam alignment overhead in a common vehicular scenario. In order to train and evaluate the proposed scheme, a time series array antenna channel data was set up by statistical channel model using time series vehicle information generated from SUMO. Simulation results show that the proposed LSTM-based scheme can reduce significant measurement overhead compared with the exhaustive search scheme, while ensuring an acceptable loss in SNR.

ACKNOWLEDGMENT

This work was supported by ...

REFERENCES

[1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, “A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges,” *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11276-015-0942-z

[2] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, “28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city,” *IEEE International Conference on Communications*, no. Icc, pp. 5163–5167, 2013.

[3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[4] J. Mercy Sheeba and S. Deepa, “Beamforming Techniques for Millimeter Wave Communications - A Survey,” *Lecture Notes on Data Engineering and Communications Technologies*, vol. 35, no. 2, pp. 1563–1573, 2020.

[5] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A tutorial on beam management for 3GPP NR at mmwave frequencies,” *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 173–196, 2019.

[6] S. Noh, M. D. Zoltowski, and D. J. Love, “Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, 2017.

[7] C. Zhang, D. Guo, and P. Fan, “Tracking angles of departure and arrival in a mobile millimeter wave channel,” *2016 IEEE International Conference on Communications, ICC 2016*, 2016.

[8] V. Va, H. Vikalo, and R. W. Heath, “Beam tracking for mobile millimeter wave communication systems,” *2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2016 - Proceedings*, no. 1, pp. 743–747, 2017.

[9] J. Lee, G. T. Gil, and Y. H. Lee, “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications,” *2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 3326–3331, 2014.

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

[11] Y. Wang, A. Klautau, M. Ribero, M. Narasimha, and R. W. Heath, “MmWave Vehicular Beam Training with Situational Awareness by Machine Learning,” *2018 IEEE Globecom Workshops, GC Wkshps 2018 - Proceedings*, pp. 87 479–87 493, 2019.

[12] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[13] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, 1997.

[14] D. Burghal, N. A. Abbasi, and A. F. Molisch, “A Machine Learning Solution for Beam Tracking in mmWave Systems,” *Conference Record - Asilomar Conference on Signals, Systems and Computers*, vol. 2019-Novem, pp. 173–177, 2019.

[15] A. Savitzky and M. J. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” *Analytical Chemistry*, 1964.