Intelligent Predictive Beamforming for Integrated Sensing and Communication Based Vehicular-to-Infrastructure Systems

Yujie Wang*, Wei Liang*, Lixin Li*, Jiankang Zhang†, Constantinos Marios Angelopoulos†
* School of Electronic and Information, Northwestern Polytechnical University, China
† Department of Computing and Informatics, Bournemouth University, U.K.
Emails: yjiewang@mail.nwpu.edu.cn, {liangwei,lilixin}@nwpu.edu.cn,
{jzhang3, mangelopoulos}@bournemouth.ac.uk

***Abstract*—Integrated Sensing and Communication (ISAC) has become a promising paradigm for next-generation wireless communications, which are capable of jointly performing sensing and communication operations. In ISAC systems, sensing accuracy and transmission rate are two major metrics to be targeted. In this paper, we propose a predictive beamforming approach based on the multi-dimensional feature extraction network (MDFEN) for vehicle-to-infrastructure (V2I) systems. In particular, in order to achieve high precision and low latency beamforming, the roadside unit (RSU) will perform angle parameter estimation and prediction based on the ISAC signal echoes. Furthermore, our predictive beamforming approach based on the multi-dimensional feature extraction network (MDFEN) is capable of improving the efficient beam alignment by exploiting the joint spatio-temporal characteristics of the received signals at the RSU side. Simulation results demonstrate that the proposed approach achieves a higher accuracy in angle tracking compared to convolutional neural network and long short-term memory models. At the same time, the system is capable of obtaining a higher transmission rate.**

***Index Terms*—Integrated sensing and communication, vehicle-to-infrastructure, beam alignment.**

I. Introduction

VEHICLE-TO-INFRASTRUCTURE (V2I) has become a promising technology for future intelligent vehicles transportation, which enables vehicles and roadside-units (RSUs) to interact in real-time with their respective traffic information. With the support of V2I, the vehicular network systems are capable of improving traffic efficiency, realizing autonomous driving, and providing other information services [1]. However, in order to enable V2I systems in practical use, Gbps-level data transmission, millimeter-scale sensing performance and dedicated operating frequency bands are indispensable [2].

Communication and sensing are the two core functionalities of a V2I system. But studying communication and sensing performance separately does not allow to fully exploit the available energy and spectrum resources. To eventually achieve the efficient use of energy and spectrum resources, integrated sensing and communicaiton (ISAC) has attracted significant attention [3] [4]. In particular, ISAC-V2I system adopts a single device to simultaneously perform communication and sensing, which can also reduce the complexity and cost of the vehicles. To meet the transmission rate demands, millimeter wave (mmWave) with a wider available bandwidth has been considered as a key factor, there is also additional benefit of improving the sensing resolution due to its shorter wavelength [5]. At the same time, by leveraging the massive multi-input-multi-output (mMIMO) array technology, beamforming is designed to generate pencil-like spatial beams, which is capable of compensating the high path-loss of mmWave signals. In addition, the system is expected to achieve a low latency and a high accuracy beamforming for stable communication links as well as high quality of service (QoS). The traditional beamforming is typically based on communication protocols [6], which might hinder the efficiency of the whole system. To this end, the ISAC signals were adopted in downlink transmission in the V2I systems [7], and the authors designed an extended Kalman filter to predict the angles of the beams, which is capable of reducing the delay of beam alignment. In order to improve estimation performance, a factor graph and message passing based beamforming method was proposed in [8].

In order to further improve the estimation performance, deep learning (DL) has attracted extensive attention, which provides a new paradigm, especially for parameters estimation of nonlinear models [9]. To solve beam alignment problem, we propose a multi-dimensional feature extraction network (MDFEN) for predictive beamforming in ISAC systems. Specifically, the RSUs transmit ISAC signals to vehicles in downlink communications. When encountering vehicles, the transmitted signals will be reflected, and the received echo signals at the RSUs are exploited for beam tracking. It is obvious that there are no specific pilots for sensing, with the whole downlink block for communication. In order to estimate the angle parameters of vehicles based on the echo signals, we invoke the MDFEN composed of convolutional neural networks (CNN) and long short-term memory (LSTM)

The work was supported by the Shenzhen Science and Technology program under Grant JCYJ20210324121006017 and in part by National Nature Science Foundation of China under Gants 62101450 and in part by Key R&D Plan of Shaan Xi Province Grants 2023YBGY037.

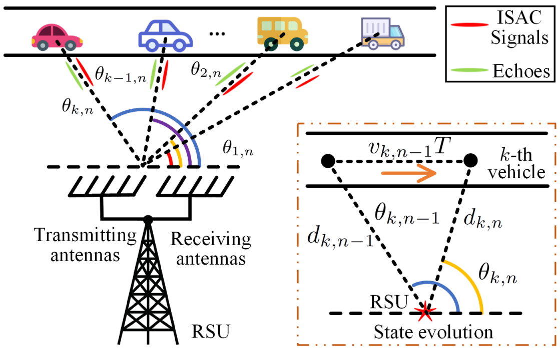

Fig. 1. System model of ISAC-V2I.

for angle tracking. In our proposed scheme, CNN components extract the spatial features between different antenna signals, and the temporal features are captured by LSTM. By extracting features in different dimensions, the joint spatio-temporal information enables the approach to achieve superior sensing performance. Meanwhile, based on the state evolution model of the vehicle, the RSUs are able to perform angle prediction, which has been embedded in the ISAC signals. Therefore, the system latency will be reduced by the prediction.

II. SYSTEM MODEL

In this paper, we consider a system where one RSU provides service for K vehicles, as depicted in Fig. 1. The RSU and all vehicles operate in the mmWave band. In order to ensure that the reflected echoes by the vehicles can be received while transmitting the ISAC signals, two mMIMO uniform linear arrays (ULAs) with M transmitting antennas and N receiving antennas are deployed at the RSU. In addition, the line-of-sight (LoS) channel model is adopted in the considered mmWave system, and the extension to non LoS channels will be considered in our future work. For brevity, we denote the angle, speed and distance of the k-th vehicle relative to the RSU at the n-th time slot by $\theta_{k,n}$, $v_{k,n}$ and $d_{k,n}$, respectively. According to geometric principle, the angle of the RSU relative to the k-th vehicle can also be expressed as $\theta_{k,n}$. The duration of each time slot is T, and according to the standard assumption [7], vehicle k travels at a constant speed within the time slot T.

A. State Evolution Declaration

Initially, the RSU works as a pure radar to perform sensing, and the initial vehicle parameters $\theta_{k,0}$, $v_{k,0}$ and $d_{k,0}$ will be inferred from the reflected echoes. To achieve high-precision beam alignment while the vehicles are in motion, the system should track the angular changes of the vehicles related to the RSU in real time. As shown in Fig. 1, according to the geometric relations between the positions of vehicle in two adjacent time slots, the state evolution model can be formulated as

$$\begin{cases} d_{k,n} = \frac{1}{2} c\tau^s_{k,n}, \\ d_{k,n-1} = \frac{1}{2} c\tau^s_{k,n-1}, \\ \sin(\theta_{k,n-1} - \theta_{k,n}) d_{k,n} = T v_{k,n-1} \sin(\theta_{k,n-1}), \end{cases} \tag{1}$$

where c is the signal propagation speed and $\tau^s_{k,n}$ denotes the sensing delay of k-th vehicle at the n-th time slot.

B. Sensing Model

At the n-th time slot, the K-dimensional downlink ISAC streams sent from the RSU to the vehicles can be denoted by $\mathbf{s}_n(t) = [s_{1,n}(t), s_{2,n}(t), \ldots, s_{K,n}(t)]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ with the k-th element $s_{k,n}(t)$ for k-th vehicle. In general, the signal $s_{k,n}(t)$ is assumed to have one unit power. The transmitted signal by the ULA of the RSU can be represented as

$$\widetilde{\mathbf{s}}_n(t) = \mathbf{F}_n \mathbf{s}_n(t) \in \mathbb{C}^{M\times 1}, \tag{2}$$

where $\mathbf{F}_n \in \mathbb{C}^{M\times K}$ is the transmitting beamforming matrix, the k-th column of $\mathbf{F}_n$ denotes the beamformer corresponding to the k-th vehicle. The beamforming matrix $\mathbf{F}_n$, designed based on the predicted angles of target vehicles, is applied to steer each beam in the direction of the target. Assuming that the predicted angle $\overline{\theta}_{k,n}$ for served vehicles $\forall k \in [1, K]$ has been obtained, the k-th column of $\mathbf{F}_n$ can be expressed as [8]

$$\mathbf{f}_{k,n} = \sqrt{p_{k,n}}\,\mathbf{a}(\overline{\theta}_{k,n}), \tag{3}$$

where $p_{k,n}$ is the transmit power. The term $\mathbf{a}(\overline{\theta}_{k,n})$ is the steering vectors of the transmit antennas array at the RSU, which is given by

$$\mathbf{a}(\overline{\theta}_{k,n}) = \sqrt{\frac{1}{M}} \left[1, e^{-\mathrm{j}\pi \cos\overline{\theta}_{k,n}}, \cdots, e^{-\mathrm{j}\pi(M-1)\cos\overline{\theta}_{k,n}}\right]^{\mathrm{T}}. \tag{4}$$

When encountering the vehicles, the transmitted signal $\widetilde{\mathbf{s}}_n(t)$ will be reflected. Therefore, the reflected echo $\mathbf{r}_n(t)$ received at the RSU can be formulated as

$$\mathbf{r}_n(t) = \varrho \sum_{k=1}^{K} \beta_{k,n} e^{\mathrm{j}2\pi\alpha^s_{k,n} t} \mathbf{G} \widetilde{\mathbf{s}}_n(t - \tau^s_{k,n}) + \mathbf{n}_{\mathrm{r}}(t), \tag{5}$$

where $\varrho = \sqrt{MN}$ represents the antenna array gain, $\beta_{k,n}$ and $\alpha^s_{k,n}$ denote the reflection coefficient and the sensing Doppler for the k-th vehicle at n-th time slot, respectively. $\mathbf{G} = \mathbf{b}(\theta_{k,n})\mathbf{a}^{\mathrm{H}}(\theta_{k,n})$, $\mathbf{b}(\theta_{k,n})$ is the receiving steering vectors of antenna array at the RSU, which has the same form as in (4) with N antennas, and $\mathbf{n}_{\mathrm{r}}(t) \in \mathbb{C}^{N\times 1}$ represents the complex additive white Gaussian noise with zero mean. Referring to [8], the ULAs of the mMIMO system are considered to have the following property:

$$\left|\mathbf{a}^{\mathrm{H}}(\theta_{k,n})\mathbf{a}(\theta_{k',n})\right| \approx 0, \forall k \neq k'. \tag{6}$$

This implies that the steering vectors are asymptotically orthogonal to each other for different vehicles [10], thus there is negligible interference between the reflected echoes [7]. Based on this, the RSU can differentiate the echoes by the corresponding steering vectors. Therefore, at n-th time slot, the reflected echo of the k-th vehicle is given by

$$\mathbf{r}_{k,n}(t) = \varrho \beta_{k,n} e^{\mathrm{j}2\pi\alpha^s_{k,n} t} \mathbf{G} \mathbf{f}_{k,n} s_{k,n}(t - \tau^s_{k,n}) + \mathbf{n}_{k,n}(t). \tag{7}$$

After matched filtering, the received signal samples of $\mathbf{r}_{k,n}(t)$ can be expressed as

$$\mathbf{y}_{k,n} = \zeta_{k,n} G \begin{bmatrix} \sum_{i=1}^{M} e^{\mathrm{j}\pi(i-1)\cos\bar{\theta}_{k,n}} e^{-\mathrm{j}\pi(i-1)\cos\theta_{k,n}} \\ \sum_{i=1}^{M} e^{\mathrm{j}\pi(i-1)\cos\bar{\theta}_{k,n}} e^{-\mathrm{j}\pi(i-2)\cos\theta_{k,n}} \\ \cdots \\ \sum_{i=1}^{M} e^{\mathrm{j}\pi(i-1)\cos\bar{\theta}_{k,n}} e^{-\mathrm{j}\pi(i-N)\cos\theta_{k,n}} \end{bmatrix} + \mathbf{n}_{\mathrm{y}}, \quad (8)$$

where $\zeta_{k,n} = \beta_{k,n}\sqrt{\frac{p_{k,n}}{N}}$, and G is the signal-to-noise ratio (SNR) gain obtained by matched filter gain. $\mathbf{n}_{\mathrm{y}} = [n_{\mathrm{y}}^1, n_{\mathrm{y}}^2, \ldots, n_{\mathrm{y}}^N]^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ is the filter output noise sample at different receive antennas, and each element in $\mathbf{n}_{\mathrm{y}}$ obeys the same distribution, i.e., $n_{\mathrm{y}}^i \sim \mathcal{CN}(n_{\mathrm{y}}; 0, \sigma_{\mathrm{y}}^2), \forall i \in [1, N]$.

C. Communication Model

At n-th time slot, the k-th vehicle adopts the receiving beamformer $\mathbf{w}_{k,n}$ to receive the ISAC signal from the RSU. Therefore, we have the received communication signal [7] as

$$\begin{aligned} c_{k,n}(t) =& \overline{\varrho}\gamma_{k,n} e^{j2\pi t\alpha_{k.n}^c} \mathbf{w}_{k,n}^{\mathrm{H}} \mathbf{u}(\theta_{k,n}) \mathbf{a}^{\mathrm{H}}(\theta_{k,n}) \cdot \\ & \mathbf{f}_{k,n} s_{k,n}(t - \tau_{m,n}^c) + n_{\mathrm{c}}(t), \end{aligned} \quad (9)$$

where $\overline{\varrho} = \sqrt{ML}$ is array gain factor between the RSU and the vehicle, $\gamma_{k,n}$ denotes the communication channel pathloss coefficient, it can be easily estimated based on the range $d_{k,n}$ relative to the RSU in a LoS dominated channel model. $n_{\mathrm{c}}(t)$ is a zero-mean AWGN with variance σ_{c}^2. $\alpha_{k,n}^c$ and $\tau_{k,n}^c$ represent the Doppler and time delay for communication, respectively. The term $\mathbf{u}(\theta_{k,n})$, the steering vector of the vehicle's ULA antenna array, has the same form as $a(\overline{\theta}_{k,n})$ with L antennas. The beamformer $\mathbf{w}_{k,n}$ is designed based on the predicted angle $\overline{\theta}_{k,n}$, i.e., $\mathbf{w}_{k,n} = \mathbf{u}(\overline{\theta}_{k,n})$. Assuming that the transmitted ISAC signals $s_{k,n}(t)$ from the RSU have a unit power, then for k-th vehicle, the SNR of receiving signal is obtained by

$$\mathrm{SNR}_{k,n} = \frac{\left|\overline{\varrho}\gamma_{k,n} \mathbf{w}_{k,n}^{\mathrm{H}} \mathbf{u}(\theta_{k,n}) \mathbf{a}^{\mathrm{H}}(\theta_{k,n}) \mathbf{f}_{k,n}\right|^2}{\sigma_{\mathrm{c}}^2}. \quad (10)$$

According to the k-th vehicle's SNR, the achievable sum-rate of all vehicles at n-th time slot can be formulated as

$$R_n = \sum_{k=1}^{K} \log_2 (1 + \mathrm{SNR}_{k,n}). \quad (11)$$

With the aforementioned analysis, it can be found that beam alignment based on angle estimation and prediction in ISAC-V2I is crucial to establish a high-quality communication link. Recent previous works [7] [8] on angle estimation employ approximations; an approach which may reduce the precision.

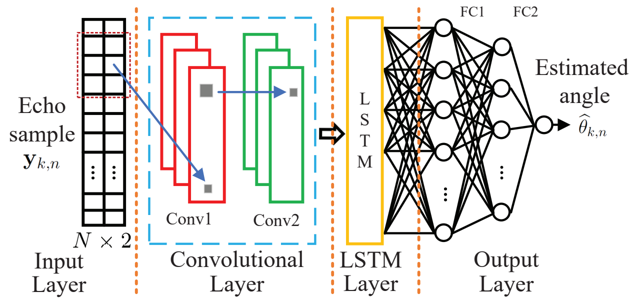

Fig. 2. The structure of MDFEN.

III. MDFEN-BASED PREDICTIVE BEAMFORMING METHOD

At the RSU side, the design of the transmitting beamformer can be divided into two stages. At the first stage, the RSU estimates the angle parameter $\widehat{\theta}_{k,n}$ of k-th vehicle based on the received echo signal $\mathbf{y}_{k,n}$. After obtaining $\widehat{\theta}_{k,n}$, the system is capable of predicting the angle $\overline{\theta}_{k,n+1}$ from (1) at the second stage. The prediction can be formulated as

$$\overline{\theta}_{k,n+1} = \widehat{\theta}_{k,n} - \arcsin\left(\frac{T v_{k,n} \sin\widehat{\theta}_{k,n}}{d_{k,n}}\right). \quad (12)$$

The RSU adopts $\overline{\theta}_{k,n+1}$ to design the transmitting beamformer $\mathbf{f}_{k,n+1}$, and the transmitted ISAC signals will contain the predicted angle. At the vehicle side, the k-th vehicle obtains $\overline{\theta}_{k,n+1}$ by decoding the information to design the receiving beamformer. Therefore, our goal is to achieve real-time accurate estimation of $\widehat{\theta}_{k,n}$, and to design the beamformer based on the predicted value $\overline{\theta}_{k,n+1}$ derived from $\widehat{\theta}_{k,n}$.

In the considered ISAC-V2I system, the k-th vehicle is analyzed in this part. For the RSU, the corresponding sample value of the echo signal at the n-th time slot can be expressed as $\mathbf{y}_{k,n} = \left[y_{k,n}^1, y_{k,n}^2, \cdots, y_{k,n}^N\right]^{\mathrm{T}}$, where $y_{k,n}^i = (y_{k,n}^{\mathrm{R}})_i + \mathrm{j}(y_{k,n}^{\mathrm{I}})_i, \forall i \in [1, N]$. This indicates that the received signal sample from each antenna is related to the relative positions between the antennas, and the received signals contain spatial information. By deforming (8), we have that

$$\begin{aligned} \mathbf{y}_{k,n} =& \psi_{k,n} \sum_{i=1}^{M} e^{\mathrm{j}\pi(i-1)\cos\bar{\theta}_{k,n}} e^{-\mathrm{j}\pi i\cos\theta_{k,n}} \begin{bmatrix} e^{\mathrm{j}\pi 1\cos\theta_{k,n}} \\ e^{\mathrm{j}\pi 2\cos\theta_{k,n}} \\ \cdots \\ e^{\mathrm{j}\pi N\cos\theta_{k,n}} \end{bmatrix} \\ & + \mathbf{n}_{\mathrm{y}}, \end{aligned} \quad (13)$$

where $\psi_{k,n} = \beta_{k,n} G\sqrt{\frac{p_{k,n}}{N}}$. The element $e^{\mathrm{j}\pi i\cos\theta_n}$ in the above column vector can be transformed as

$$\begin{aligned} e^{\mathrm{j}\pi i\cos\theta_{k,n}} &= \cos\left(2\pi i\frac{\cos\theta_{k,n}}{2}\right) + \mathrm{j}\sin\left(2\pi i\frac{\cos\theta_{k,n}}{2}\right), \\ &= \cos(2\pi t f) + \mathrm{j}\sin(2\pi t f), \end{aligned} \quad (14)$$

where we define $t = i$ to regard the serial number of the antennas as the time factor, and $f = \frac{\cos\theta_{k,n}}{2}$ represents the frequency factor. Derived from the (14), we propose the MDFEN-based beam tracking approach. The proposed neural network model mainly consists of CNN and LSTM. Via convolutional operations, CNN exploits the knowledge of the receptive field to extract local feature information from the input features. Although CNN specializes in exploring spatial features of the signal, the temporal features could be better extracted by LSTM. To this end, LSTM is added to our model. LSTM records information for a long time and provides better performance compared to ordinary recurrent neural networks (RNN) in many tasks. Therefore, the MDFEN complement the spatial and temporal features of the received signal effectively, thus being capable of improving the performance of angle tracking.

The overall architecture of our model is shown in Fig. 2. The model is composed of four parts, including input layer, convolutional layer, LSTM layer and output layer. The above mentioned layers are detailed below.

1) Input Layer: The samples of each echo signal are composed of N complex numbers. Neural networks cannot directly deal with complex numbers, so we represent each sample in a matrix form. More specifically, as shown in the input layer of Fig. 2, the i-th row represents the sample value of the i-th receiving antenna, the matrix's first column is the real part of the samples, the second column is the imaginary part, and the size of neural network input data is $N \times 2$.

2) Convolutional Layer: The convolutional layer contains two CNNs. Each CNN consists of two parts; one for the convolution operation and the other for the nonlinear mapping. For the purpose of extracting local feature information, the former employs convolution kernels that sequentially scan the data from the previous layer for convolution operations, and the latter selects a suitable activation function to achieve a non-linear mapping from low-level features to high-level features. To effectively prevent gradient disappearance or explosion, both CNNs adopt the rectified linear unit as the activation function. Considering the limited size of the input data, 3 $\times$ 3 and 2 $\times$ 2 convolution kernels are used to perform convolution operations respectively.

3) LSTM Layer: LSTM is a unique kind of RNN and it specializes in time modeling, which is mainly designed to resolve the gradient disappearance problems during the long sequences training. In order to attain a higher angle tracking performance, we exploit the complementary nature of CNN and LSTM and combine them into one coherent model.

4) Output Layer: After the LSTM layer, the joint time-frequency features of the echo signal will be extracted. The output layer performs a weighted summation of the joint features. To be specific, the output layer contains two Fully-Connected (FC) layers, i.e. FC1 and FC2. The number of neurons in the FC layers is adjustable, and the final output is the estimation of the angle. In addition, batch normalization process is performed at the FC1 layer to prevent overfitting.

After several training epochs, the MDFEN model learns the joint spatio-temporal relationship of the echo signals, which allows for angle estimation. Furthermore, the predicted angle with respect to the next timeslot will be obtained by the RSU based on Eq. (12).

According to [11], the complexity of convolutional layer is given by $\mathcal{O}\left(\sum_{l=1}^{N_c} s_l^2 c_l^{in} c_l^{out} h_l^f w_l^f\right)$, where N_c is the number of CNN, subscript l represents the index of CNN, and s_l is the spatial size of convolution kernel. c_l^{in} and c_l^{out} denote the number of channels for input and output feature map of l-th CNN, respectively. Moreover, h_l^f and w_l^f are the height and width of input feature map. The complexity of LSTM layer is $\mathcal{O}\left(4\tau\left(d_h\left(d_h+d_x\right)+d_h\right)\right)$, where τ is time step, and d_x and d_h represent the input and the output dimension of the LSTM, respectively. As for the complexity of output layer, it can be expressed as $\mathcal{O}\left(\sum_{m=1}^{N_{fc}} 2a_m b_m\right)$, where N_{fc} is the number of FC layer, subscript m represents the index of FC layer. a_m and b_m denotes the number of neurons in the input and output of the m-th FC layer, respectively. In summary, the complexity of the proposed approach is

$$\begin{aligned}&\mathcal{O}(I_t N_e(\textstyle\sum_{l=1}^{N_c} s_l^2 c_l^{in} c_l^{out} h_l^f w_l^f + \sum_{m=1}^{N_{fc}} 2a_m b_m \\ &\quad + 4\tau\left(d_h\left(d_h+d_x\right)+d_h\right))),\end{aligned} \tag{15}$$

where I_t and N_e denote the maximum epoch number and the total amount of training examples, respectively.

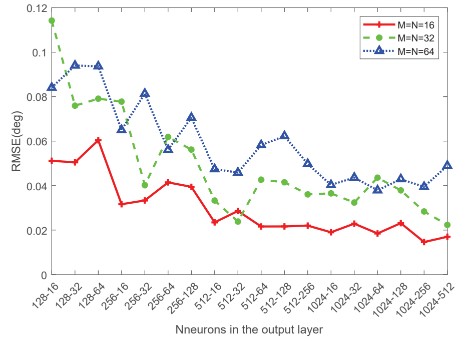

Fig. 3. Influence of the number of neurons in the output layer on the sensing performance.

IV. SIMULATION RESULTS

Simulation results are presented to evaluate and validate the efficiency of the proposed system. We consider one RSU providing service to three vehicles. The vehicles travel in the direction away from the RSU. The parameters of the system are configured as follows: the RSU is located at $[0,0]^{\mathrm{T}}$, initial locations of the vehicles are set as $[25,20]^{\mathrm{T}}$, $[30,20]^{\mathrm{T}}$, $[35,20]^{\mathrm{T}}$ with 5 m spacing. The speed of signal propagation is approximated as $c=3\times10^8$ m/s, and the average velocity of the k-th vehicle during the n-th slot is set to obey the uniform distribution, i.e., $v_{k,n} \sim \mathcal{U}\left[10,15\right]$ m/s . Both the RSU and the vehicles operate at a frequency of $f_c = 30$ GHz, and the number of antennas for all vehicles is $L=16$. The original transmitted signal for each vehicle is assumed to have unit power, then $G=1$, and $\sigma_{\mathrm{y}} = \sigma_{\mathrm{c}} = 1$ is used as the noise variances for sensing and communication. Unless otherwise stated, all results are averaged from 3,000 independent simulations.

The output layer plays a crucial role in combining and transforming spatio-temporal joint features extracted from earlier layers, ultimately producing sensing results. Therefore, the number of neurons in the output layer will affect the sensing performance. To this end, 18 neuronal configuration

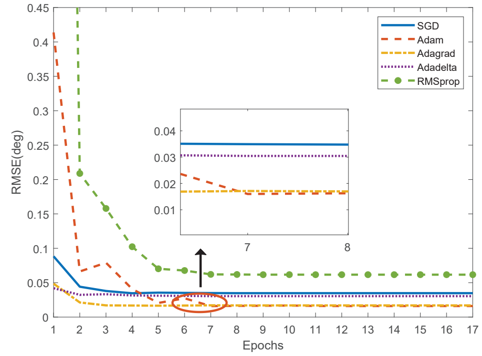

Fig. 4. Influence of optimizer on the sensing performance when M=N=16.

experiments were performed, and the sensing performance is demonstrated in terms of root mean squared error (RMSE) of angle tracking in Fig. 3, where the RSU is equipped with ULAs of different sizes. In Fig. 3, the horizontal coordinate indicates the set of neuron numbers, such as “128-32” meaning the number of neurons in two FC layers are 128 and 32, respectively. It is evident that with an increase in the size of the output layer, there is an initial decline in sensing error followed by an increase. This is because, if the number of neurons is too small, the model may underfit and fail to extract sufficient feature information, leading to deterioration of the sensing performance. On the contrary, if the number of neurons is too high, it can cause the model to be overfitted and over-fit the training samples, resulting in poor generalization on the test set. In order to balance the complexity and the accuracy, the configuration of “1024-256” is selected as the size of the output layer. Interestingly, the sensing performance deteriorates as the size of the antenna array increases. This could be attributed to the fact that with larger input feature matrix dimensions, the convolution operation generates more invalid feature information, thereby reducing the model’s learning capacity.

The parameter optimization process of neural networks is significantly impacted by the optimizer. In order to determine the most appropriate optimizer for optimizing MDFEN, a study was conducted to evaluate the effects of Stochastic Gradient Descent, Adam, Adagrad, Adadelta and RMSprop on sensing performance under three different antenna configurations. Using $M = N = 16$ as an example, the experimental results are illustrated in Fig. 4. Upon the number of training iterations increased, the perceptual error gradually decreased until it converged. Compared to other optimizers, networks optimized by Adam exhibited superior performance. Therefore, Adam was selected for the subsequent experiments in this paper.

Fig. 5 displays the instantaneous tracking results regarding $\theta_{2,n}$ when $M = N = 16$. The proposed approach demonstrates closer approximation to the true values in its sensing results, while exhibiting smaller sensing errors in comparison to schemes based on pure CNN and LSTM. Additionally, the sensing error fluctuation of the proposed approach is reduced, indicating a higher level of robustness.

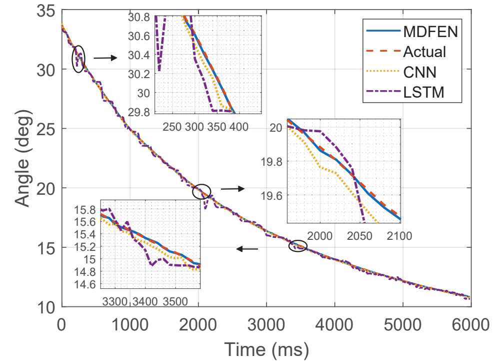

Fig. 5. Instantaneous tracking results regarding $\theta_{2,n}$ when M=N=16.

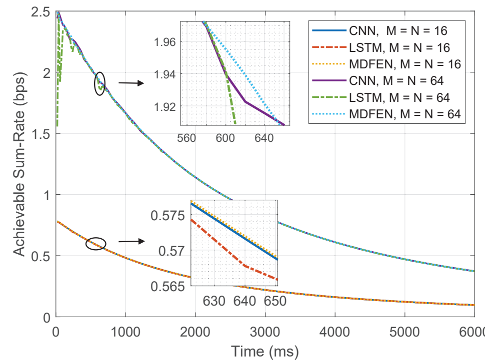

Fig. 6. System’s achieved sum-rates for different schemes.

These collective findings suggest that the proposed approach adequately tracks changes in angle, showcasing the MDFEN’s capacity to extract joint spatio-temporal information from the received signal and effectively learn the nonlinear features of the system.

Two instantaneous achievable sum-rates for the conceived system at the setting of $M = N = 16$ and $M = N = 64$ are illustrated in Fig. 6. Intuitively, the sum-rates at the setting of $M = N = 64$ are always higher than the rates when $M = N = 16$. This is because a higher SNR will be obtained by the larger-scale antenna array according to (10). In addition, since all vehicles drive away from the RSU, the channel pathloss coefficient gets higher. Therefore, the general trend of all rates tends to decrease. Due to the poor angle tracking performance of LSTM, the communication performances are inferior. In particular, the rate of LSTM model experiences a significant deterioration during the initial phase of the simulation at the setting of $M = N = 64$. On the contrary, benefiting from superior sensing performance, the proposed approach achieves a much more accurate beam alignment, which results in higher communication gain thus improving the communication rate.

V. CONCLUSION

In this paper, we proposed a novel ISAC-V2I framework that utilizes the ISAC signals for sensing and communication simultaneously. By jointly exploiting CNN and LSTM, we proposed a MDFEN based predictive beamforming approach and validated the sensing performance. For the considered system, the RSU performs angular tracking based on the echo signals, followed by beam prediction. Simulation results demonstrate that the proposed approach is capable of sufficiently extracting the joint spatio-temporal features from the received signals at different antennas. Based on the extracted spatio-temporal features, the system achieves a higher sensing performance compared to CNN and LSTM models. Furthermore, the RSU is capable of improving the accuracy of beam alignment and consequently improving the overall transmission rate of the system.

REFERENCES

[1] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[2] H. Wymeersch, G. Seco-Granados, G. Destino, D. Dardari, and F. Tufvesson, "5g mmwave positioning for vehicular networks," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 80–86, Dec. 2017.

[3] D. K. Pin Tan, J. He, Y. Li, A. Bayesteh, Y. Chen, P. Zhu, and W. Tong, "Integrated sensing and communication in 6G: Motivations, use cases, requirements, challenges and future directions," in *2021 1st IEEE International Online Symposium on Joint Communications Sensing (JCS)*, Feb. 2021, pp. 1–6.

[4] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6g and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[5] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[6] C. Liu, M. Li, L. Zhao, P. Whiting, S. V. Hanly, I. B. Collings, and M. Zhao, "Robust adaptive beam tracking for mobile millimetre wave communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1918–1934, Mar. 2021.

[7] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7704–7719, Nov. 2020.

[8] W. Yuan, F. Liu, C. Masouros, J. Yuan, D. W. K. Ng, and N. González-Prelcic, "Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1442–1456, Mar. 2021.

[9] L.-H. Shen, T.-W. Chang, K.-T. Feng, and P.-T. Huang, "Design and implementation for deep learning based adjustable beamforming training for millimeter wave communication systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2413–2427, Mar. 2021.

[10] T. L. Marzetta, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[11] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015.