Predictive Beamforming in Integrated Sensing and Communication-Enabled Vehicular Networks

Wei Liang[1,2] *Member, IEEE*, Yujie Wang[1] *Student Member, IEEE*, Jiankang Zhang[3], *Senior Member, IEEE*, Lixin Li[1,2] *Member, IEEE*, and Zhu Han[4], *Fellow, IEEE*

[1]School of Electronic and Information, Northwestern Polytechnical University, Xi' an, China.

[2]Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen, China.

[3]Department of Computing and Informatics, Bournemouth University, Bournemouth, U.K.

[4]Department of Electrical and Computer Engineering, University of Houston, TX, USA.

***Abstract*— Integrated sensing and communication (ISAC) has recently attracted significant research attention. This paper develops the deep learning-based predictive beamforming method for the ISAC-enabled vehicular networks. Traditional deep learning (DL) is a data-driven approach, which means that numerous training samples are required to improve system performance. In addition, embedded devices are not able to provide sufficient computing power, which hinders the application of DL solutions. Motivated by this, the dynamic self-attention mechanism is proposed to reduce the dependence of DL on training samples. Aiming for the optimal trade-off between sensing performance and computational complexity, the efficient model design, Self-Attention Channel Shuffle Mobile Network (SACSMN), is formulated. Experimental results demonstrate that SACSMN achieves similar sensing performance to that based on the full training set under the condition of few samples, the dependence of SACSMN on training samples is significantly reduced. Furthermore, SACSMN significantly reduces the computational complexity while achieving the same level of sensing performance as the benchmarks, realizing the optimal trade-off between system sensing performance and computational complexity. Benefiting from the robust sensing performance of SACSMN, the system achieves the same level of communication performance as that based on full training samples in the case of few samples.**

I. Introduction

Radar sensing and wireless communication, as two main applications of electromagnetic radiation, are implemented in numerous scenarios. Over the past few decades, due to the separate operating frequency bands and different performance specifications, these two types of applications have been studied and developed as separate research entities [1]. Nevertheless, increasing sensing resolution and communication rate requirements have exacerbated the problem of spectrum resource shortage. In particular, part of the frequency band suitable for sensing has been used for communication to satisfy the ever-increasing quality of services (QoS) for mobile communication services. Moreover, with the development of sensing and wireless communication, the two types of systems share many commonalities [1]. Based on this, researchers have intended to combine sensing and communication in one unified framework, which has inspired the emerging research theme, namely integrated sensing and communication (ISAC) [1].

The work was supported by National Nature Science Foundation of China under Gants 62101450, in part by Key R&D Plan of Shaan Xi Province Grants 2023YBGY037, and in part by the Shenzhen Science and Technology program under Grant GJHZ20220913143203006.

ISAC technology has the potential for numerous improvements. On the hardware side, hardware costs and complexities will be reduced as a result of sharing a single device for sensing and communication [2]. In terms of resource consumption, two functionalities are performed by ISAC systems with single transmissions and spectrum resources, which reduces energy expenditure and improve the spectral efficiency [3]. Moreover, the mutual benefits of collaborative communication and sensing cooperation are able to enhance the performances of both functionalities [4]. Given the above advantages, ISAC will play an important role in the future development of wireless networks, especially vehicle-to-infrustructure (V2I) [2]. In traditional V2I scenarios, vehicles and infrastructures are equipped with numerous jointly deployed communication and sensing devices to track/interact with system key information. With the inherent benefits of ISAC-V2I systems, there will be significant efficiency gains in terms of hardware, energy and spectrum. In addition, the terrain and meteorological information obtained through communication enables high-resolution sensing, and the communication parameters obtained through sensing allow for high-data rate transmission [5].

Contributions to the ISAC system initially focused on waveform design [6]. In [6], the authors utilized chirp signals to perform sensing and information embedding. Specifically, the down-chirp and up-chirp waveforms were employed to distinguish between binary information "0" and "1". However, the communication rate of this scheme was suboptimal. The introduction of orthogonal frequency division multiplexing (OFDM) provided a new approach for implementing ISAC [7]. In terms of communications, OFDM offers significant advantages in multipath interference, channel adaptation, and frequency selective fading. Additionally, with OFDM waveforms, distance and Doppler estimators are decoupled, resulting in higher resolution positioning services [8]. Considering the vulnerability of OFDM to time-varying channels, the ISAC-assisted orthogonal time frequency space transmission scheme for V2I was developed, where the data symbols are modulated into the delay-Doppler domain instead of the conventional time-frequency domain. Moreover, to meet the system requirements for sum-rate performance, multiple-input multiple-output (MIMO) technology has been introduced to facilitate ISAC execution [9]. In [9], the mainlobes of the spatial beams formulated by MIMO radar were dedicated to object detection, while the sidelobes were used for communication.

Gbps-level communication rates and centimetre-level sensing capabilities are essential to enable V2I systems. In order to meet these requirements, the millimeter wave (mmWave) band and the massive multi-input-multi-output (mMIMO) antenna array offer a promising solution. The large bandwidth available in the mmWave band satisfies the need for high data rates, and the shorter wavelength provides higher sensing resolution. On the other hand, the directional "pencil-like" beams generated by the mMIMO compensate for the high path loss of mmWave signals, with the added benefit that the spatial domain may be exploited to increase the degree of freedom in system design. The authors in [10] considered a mmWave-based ISAC scenario and derived system performance bounds. However, this work is unsuitable for V2I communication and does not consider beam tracking in high-speed motion scenarios, which is critical to the implementation of V2I systems.

In V2I systems, vehicles travelling at high speeds result in a highly time-varying nature of the system, so efficient beam alignment between the vehicles and the road-side unit (RSU) is essential to ensure QoS [11]. Traditional mmWave beam alignment mainly relies on communication protocols. Specifically, the RSU sends communication signals containing pilots to the vehicles, which perform decoding to estimate the channel information and feeds it back to the RSU [12]. Fast beam tracking has been proposed in several works to achieve low-latency beam alignment, where the RSU is capable of performing beam prediction. In [13], one predictive beamforming method based on the extended Kalman filtering scheme was proposed. In the considered scenario, ISAC signals were employed in the downlink transmission, and the RSU performed beam tracking and prediction based on the echo signals reflected by the vehicles. This implies that all downlink signals are employed for communication, without any dedicated pilot signals. The author [14] developed a factor graph and message passing-based algorithm for the RSU to estimate and predict the vehicles' motion parameters.

In recent years, deep learning (DL) has gained significant attention for their remarkable achievements. These breakthroughs have provided a novel approach for beam alignment, and [15] utilized a fully-connected (FC) neural network to estimate the angular parameters of the vehicles for predictive beamforming in ISAC-V2I systems. The deep neural network (DNN) training is described as data-driven, which means that numerous training samples are essential to attain optimal sensing performance [16]. However, datasets generated from a single device or system contain fewer samples. To this end, the concept of few-shot learning [17] was proposed, which aims to reduce the dependence of DNN training on training samples. Furthermore, the self-attention (SA) mechanism [18], which calculates the correlations between all features to compensate for the lack of long-range dependence in DNNs, was proposed offering a promising approach to obtain valid information based on few samples. It should be highlighted that the remarkable achievements of DNNs have come at the cost of enormous computational complexity that is beyond the capacity of many mobile and embedded systems. Accordingly, the depthwise convolution (DWConv) was developed to reduce computational overhead [19].

In this paper, we present a novel design for predictive beamforming in the ISAC-V2I system to achieve high communication and sensing performance, while minimizing signaling overhead. Specifically, during transmission periods, RSU transmits ISAC signals containing information which are decoded and reflected by vehicles. Based on these echoes, the RSU performs tracking and localization to predict beam directions. The entire downlink block is exclusively utilized for communication without allocating dedicated downlink pilots and uplink feedback for sensing, thereby reducing beam tracking overhead and enhancing system resource utilization. To address the strong nonlinearity of the sensing and enhance beamforming performance, we propose the predictive beamforming scheme based on DNN. Under the condition of few echo samples, low-complexity-DNN solutions are investigated. Additionally, we propose the self-attention channel-shuffle mobile network (SACSMN) to further improve beam alignment accuracy, enabling attention-driven modeling for sensing tasks. The primary contributions of this paper are summarized as follows:

- To minimize the overhead and latency of beam tracking, we propose a novel predictive beamforming methodology in ISAC-V2I system. In the proposed system, there is no need for dedicated downlink pilots or uplink feedback. Moreover, the angular parameters are predicted at the RSU for beam prediction to reduce latency.
- To reduce the dependence of DNNs on training samples, the dynamic SA is proposed to improve the feature extraction capability of the DNN model. Furthermore, one trade-off problem between sensing performance and model computational complexity in the ISAC-V2I system is formulated, which aims to achieve stable sensing performance while significantly reducing model computational overhead. In addition, the influence of the size of training samples on the sensing and communication performance of the system is investigated.
- We propose the novel DNN framework, SACSMN, for the considered ISAC-V2I system to perform sensing. SACSMN comprises SA encoder / decoder layer and feature extraction layer. The proposed SA encoder / decoder layer aims to capture the correlation and importance between the elements in the feature map, which improves the feature extraction capability of the model. The feature extraction layer adopts DWConv as the backbone to significantly reduce the computational overhead.
- Extensive simulations are conducted to verify the effectiveness of the proposed scheme with respect to sensing and communication performance. Specifically, satisfactory sensing performance is achieved by SACSMN under few-shot conditions. What's more, SACSMN enables similar sensing performance to the benchmarks while significantly reducing computational complexity. In addition to the sensing ability, SACSMN shows great advantages in communication performance under few-shot conditions. The results demonstrate the tremendous potential of the proposed SACSMN.

The structure of this paper is organized as follows. Some concepts related to DNNs are introduced in Section II. Section

TABLE I: Summary of the abbreviations.

Abbreviation	Meaning
DL	Deep learning
DNN	Deep neural network
FC	Fully-connected
GConv	Group convolution
SA	Self-attention
DWConv	Depthwise convolutions
MACC	Multiply-accumulate operation

III describes the considered ISAC-V2I system, and a trade-off problem for predictive beamforming design in considered system is formulated in Section IV. Then Section V illustrates the proposed predictive beamforming approach, and the simulation results are provided in Section VI. Finally, Section VII concludes this paper. For convenience, abbreviations that appear frequently in the paper are summarised in Table. I.

II. Related Work

A. SA Mechanism

The conventional convolutional layers in DNNs rely on local convolution operations for feature extraction within a local neighborhood, making it challenging for the layers to learn global information. In order to address this limitation, the SA mechanism was introduced. As illustrated in Fig. 1, the SA mechanism calculates the correlation between all features on the input feature map, and then multiplies the attention map with the input feature map to obtain the global information of the features, thus overcoming the limitations of local receptive fields in convolutional layers.

Specifically, as illustrated in Fig. 1, the input feature maps $\boldsymbol{X_{SA}}$ from the preceding hidden layer are initially converted to $\boldsymbol{Q}_{SA}, \boldsymbol{K}_{SA}, \boldsymbol{V}_{SA}$ through a 1×1 convolution kernel. These entities are respectively referred to as queries, keys, and values. Furthermore, it is obtained that

$$\begin{cases} \boldsymbol{Q}_{SA} = \boldsymbol{W}_q \boldsymbol{X}_{SA}, \\ \boldsymbol{K}_{SA} = \boldsymbol{W}_k \boldsymbol{X}_{SA}, \\ \boldsymbol{V}_{SA} = \boldsymbol{W}_v \boldsymbol{X}_{SA}, \\ \boldsymbol{Y}_{SA} = softmax(\frac{\boldsymbol{Q}_{SA} \odot \boldsymbol{K}_{SA}^{\mathrm{T}}}{\sqrt{d_{k'}}}), \\ \boldsymbol{Z}_{SA} = \text{attention}(\boldsymbol{Q}_{SA}, \boldsymbol{K}_{SA}, \boldsymbol{V}_{SA}) = \boldsymbol{Y}_{SA} \odot \boldsymbol{V}_{SA}, \end{cases} \tag{1}$$

where $\boldsymbol{W}_q$, $\boldsymbol{W}_k$, and $\boldsymbol{W}_v$ are the weight matrices learned through training. The $\odot$ symbol denotes the matrix dot product operation. $\boldsymbol{Y}_{SA}$ represents the attention map, which reflects the degree of attention. The dimension of $\boldsymbol{K}_{SA}$ is $d_{k'}$, and the scaling factor of $\frac{1}{\sqrt{d_{k'}}}$ prevents the gradient from vanishing in the $softmax(\cdot)$ function. $\boldsymbol{Z}_{SA}$ is the output of the SA module. Compared to pure convolutional layers, the SA mechanism is effective in capturing both local and long-range dependencies while requiring less computation due to the utilization of dot product operations.

B. Group Convolution Operation

The standard convolution incurs a high computational cost, which may limit the number of channels and deteriorate the learning abilities of DNNs. To this end, group convolution (GConv) was developed. GConv operation partitions the input feature map $\boldsymbol{X}_{gc}$ into g groups of equal size before performing separate convolution operations on each group. A particular instance of GConv is DWConv, where the number of groups equals the number of channels in the input feature map.

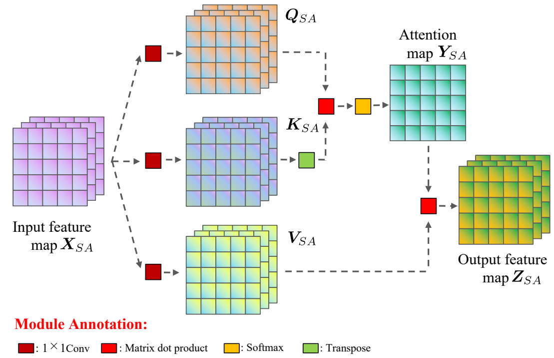


Fig. 1: The processing procedure of the SA mechanism

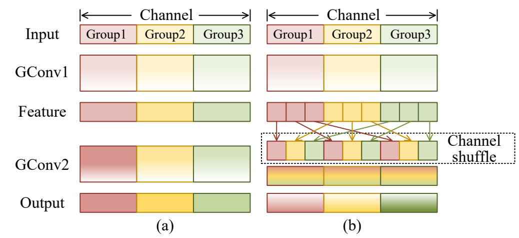


Fig. 2: Traditional group convolution and channel shuffle operations.

Generally, the computational complexity is measured by multiply-accumulate operation (MACC) in DNNs. For a standard convolutional operation, the input feature map $\boldsymbol{X}_{sc} \in \mathbb{R}^{hin \times w_{in} \times c_{in}}$ is filtered with c_{out} convolutional kernels, each of size $k_{sc} \times k_{sc}$, to produce the output feature map $\boldsymbol{Y}_{sc} \in \mathbb{R}^{hout \times w_{out} \times c_{out}}$. The computational cost of the standard convolution operation, denoted as MACC_{sc}, is expressed as

$$\mathrm{MACC}_{sc} = h_{in} \times w_{in} \times c_{in} \times k_{sc} \times k_{sc} \times c_{out}. \tag{2}$$

If the number of groups is denoted by g and the convolutional kernels remain constant, the cost of the GConv operation is expressed as $\mathrm{MACC}_{gc} = h_{in} \times w_{in} \times c_{in} \times k_{sc} \times k_{sc} \times c_{out}/g$. Therefore, the reduction in computional cost is represented by

$$\frac{h_{in} \times w_{in} \times c_{in} \times k_{sc} \times k_{sc} \times c_{out}/g}{h_{in} \times w_{in} \times c_{in} \times k_{sc} \times k_{sc} \times c_{out}} = 1/g. \tag{3}$$

The aforementioned analysis reveals that the computational burden of DNNs is considerably reduced by the GConv operations. This reduction in complexity facilitates the deployment of lightweight DNNs.

C. Channel Shuffle Operation

Fig. 2(a) illustrates an instance of two regular GConv operations, denoted by GConv1 and GConv2. In this case, the input feature map channels are partitioned into three distinct groups, namely Group1, Group2, and Group3. Obviously, there is a drawback when stacking multiple GConv layers: the output of each GConv operation depends only on a specific input channels, which prevents the interaction of information between different groups of channels, leading to a weakened DNN representation.

The channel shuffle operation [20] enables the GConv operation to obtain input features from multiple groups, as

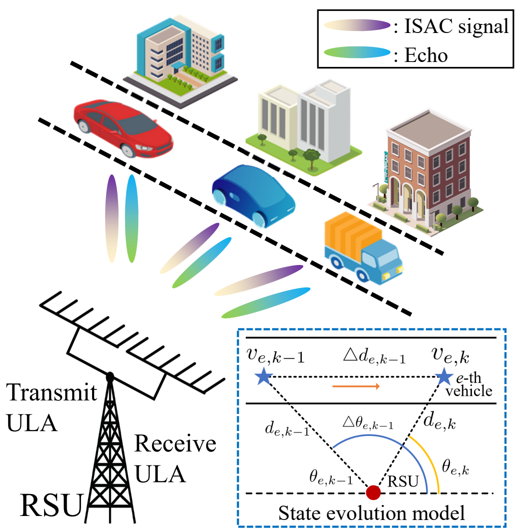


Fig. 3: System model for the considered ISAC-V2I scenario.

illustrated in Fig. 2(b). Specifically, if the output feature map of GConv1 is partitioned into three sub-groups, denoted as (i, j) where j denotes the j-th sub-group in group i, then the sub-group (i, j) from GConv1's output is rearranged as the sub-group (j, i) in GConv2's input feature map, resulting in a novel grouping scheme. When this is completed, the standard GConv operation could be applied to the new feature map.

The channel shuffle operation offers the benefit of fully connecting the channels of input and output feature maps, resulting in a more potent structure for DNNs.

III. System Model

The proposed ISAC-V2I network is developed as depicted in Fig. 3. Specifically, there is a RSU serves E vehicles working in the mmWave frequency band. Based on mmWave and mMIMO technology, the RSU is equipped with a Uniform Line Array (ULA) consisting of M transmit antennas and N receive antennas, which is able to transmit downlink signals and receive vehicle echoes simultaneously. In general, it is assumed that the vehicles travel along the single lane road, and the ULA is deployed parallel to the road [14]. Without loss of generality, there are no obstacles between the vehicles and the RSU. Therefore, the signals transmit through the line-of-sight (LoS) channels [13]. The researches on planar array and non-line-of-sight channels will be designated as future work, and the relevant derivation will be extended.

In this section, the general framework of the ISAC-V2I system is described in detail. We first elaborate the sensing signal model, the state evolution model and the angle prediction model. Subsequently, the communication model is analyzed.

A. General Framework

The system operations are executed according to the following steps:

Step 1 State sensing: At time slot k, $k \in \{1, 2, \cdots, K\}$, the RSU transmits ISAC signals to the vehicles, which are then reflected by the vehicles as echoes. Based on the echoes, the motion parameters of the vehicles, including distances, azimuth angles, and velocities, are sensed.

Step 2 State evolution model construction and state prediction: The state evolution model will be established based on the geometric relationships between adjacent time slots. The RSU will utilize the motion parameters at time slot k to predict the locations and angles of the vehicles at the next time slot.

Step 3 Downlink communication: The transmit beamformer at the RSU will be designed to direct beams containing communication information toward corresponding vehicles based on the predicted angles in step 2. The embedded information on the vehicles' side will be decoded.

B. Sensing Model

1) Sensing Signal Model: In ISAC-V2I systems, the transmitted ISAC signals are simultaneously used for sensing and communication. At the k-th slot, in order to provide service for all vehicles, a E-dimensional multi-beam ISAC signal $\boldsymbol{s}_k(t) = [s_{1,k}(t), s_{2,k}(t), \ldots, s_{E,k}(t)]^{\mathrm{T}} \in \mathbb{C}^{E\times 1}$ is formulated at the RSU, where $s_{e,k}(t)$, $\mathbb{E}\left\{|s_{e,k}(t)|^2\right\} = 1$, represents the downlink signal for the e-th, $e \in \{1, 2, \cdots, E\}$, vehicle at time instant t of the k-th slot. Over the transmit ULA, the transmitted signal is denoted by

$$\hat{\boldsymbol{s}}_k(t) = \boldsymbol{F}_k \boldsymbol{s}_k(t) \in \mathbb{C}^{M\times 1}, \tag{4}$$

where $\boldsymbol{F}_k = [\boldsymbol{f}_{1,k}, \boldsymbol{f}_{2,k}, \ldots, \boldsymbol{f}_{E,k}] \in \mathbb{C}^{M\times E}$ is the downlink beamforming matrix. Transmit beamforming vector $\boldsymbol{f}_{e,k} \in \mathbb{C}^{M\times 1}$ is adopted to steer the corresponding beam towards the e-th vehicle at the k-th slot, which is given by

$$\begin{aligned}\boldsymbol{f}_{e,k} &= \mathfrak{F}\left(\tilde{\theta}_{e,k}\right)\\ &= \sqrt{\frac{1}{M}}\left[e^{-\mathrm{j}\pi(1-1)\cos\tilde{\theta}_{e,k}}, \cdots, e^{-\mathrm{j}\pi(M-1)\cos\tilde{\theta}_{e,k}}\right]^{\mathrm{T}}\\ &= \sqrt{\frac{1}{M}}\left[1, \cdots, e^{-\mathrm{j}\pi(M-1)\cos\tilde{\theta}_{e,k}}\right]^{\mathrm{T}}\end{aligned} \tag{5}$$

where $\mathfrak{F}(\cdot)$ denotes the beamforming function. This paper focuses on exploring the trade-off between sensing performance and the computational complexity of the network model. With the constant transmit power, the influence of other factors will be reduced. Therefore, it is assumed that the transmitted power for each beam at each time instant is unit power [13]. $\tilde{\theta}_{e,k}$ is the angle prediction of vehicle e relative to the RSU at the k-th time slot.

The transmitted signal $\hat{\boldsymbol{s}}_k(t)$ will be reflected by all vehicles, and the received echoes at the receive ULA is formulated as

$$\begin{aligned}\boldsymbol{r}_k(t) =& \alpha \sum_{e=1}^{E} \beta_{e,k} e^{\mathrm{j}2\pi t\delta_{e,k}} \boldsymbol{b}(\theta_{e,k}) \boldsymbol{a}^{\mathrm{H}}(\theta_{e,k}) \hat{\boldsymbol{s}}_k(t - \tau_{e,k})\\ &+ \boldsymbol{z}_s(t),\end{aligned} \tag{6}$$

where $\alpha = \sqrt{MN}$ is the sensing array gain coefficient, $\beta_{e,k} = \frac{\gamma}{2d_{e,k}}$ denotes the reflection factor with γ being the complex radar cross-section and $d_{e,k}$ being the distance between vehicle e and the RSU at k-th slot. $\delta_{e,k}$ and $\tau_{e,k}$ represent the Doppler and the delay of the e-th vehicle for sensing, respectively. $\boldsymbol{z}_s(t) \in \mathbb{C}^{N\times 1}$ is a complex additive white Gaussian noise with zero mean. Transmit steering vector $\boldsymbol{a}(\theta_{e,k})$ and receive steering vector $\boldsymbol{b}(\theta_{e,k})$ at the RSU are expressed as

$$\boldsymbol{a}(\theta_{e,k}) = \sqrt{\frac{1}{M}}\left[1, \cdots, e^{-\mathrm{j}\pi(M-1)\cos\theta_{e,k}}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1} \quad (7)$$

and

$$\boldsymbol{b}(\theta_{e,k}) = \sqrt{\frac{1}{N}}\left[1, \cdots, e^{-\mathrm{j}\pi(N-1)\cos\theta_{e,k}}\right]^{\mathrm{T}} \in \mathbb{C}^{N\times 1}, \quad (8)$$

respectively. And $\theta_{e,k}$ is the actual angle of vehicle e relative to the RSU at the k-th time slot.

For the ULA in mMIMO antenna arrays, the steering vectors are asymptotically orthogonal to each other [21], i.e.,

$$\left|\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{a}(\theta')\right| \Rightarrow 0, \forall\theta \neq \theta', M \gg 1, \quad (9)$$

which means that there is negligible inter-beam interference between the echoes from different vehicles. Based on this, the reflected echo of the corresponding vehicle will be identified by the RSU at each time slot. Therefore, the received echo from e-th vehicle could be modeled as

$$\begin{aligned}\boldsymbol{r}_{e,k}(t) =& \alpha\beta_{e,k}e^{\mathrm{j}2\pi t\delta_{e,k}}\boldsymbol{b}(\theta_{e,k})\boldsymbol{a}^{\mathrm{H}}(\theta_{e,k})\boldsymbol{f}_{e,k}s_{e,k}(t-\tau_{e,k}) \\ &+ \boldsymbol{z}_s(t),\end{aligned} \quad (10)$$

where $\boldsymbol{r}_{e,k}(t) = \left[r^1_{e,k}(t), r^2_{e,k}(t), \ldots, r^N_{e,k}(t)\right]^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ with $r^n_{e,k}(t)$ representing the echo received by the n-th antenna of receive ULA from vehicle e at time slot k. Adopting a series of $s_{e,k}(t)$ with different time delays and frequency shifts, the radar's matched filtering will be performed on $\boldsymbol{r}_{e,k}(t)$ and the estimates of the delay and the Doppler shift are formulated as

$$\left\{\bar{\tau}_{e,k}, \bar{\delta}_{e,k}\right\} = \underset{\tau,\delta}{\operatorname{argmax}}\left|\int_0^{\triangle T} r_{e,k}(t)s^*_{e,k}(t-\tau)e^{-j2\pi t\delta}dt\right|. \quad (11)$$

Compensating $\boldsymbol{r}_{e,k}(t)$ with $\bar{\tau}_{e,k}$, $\bar{\delta}_{e,k}$, the received signal model is defined as

$$\widehat{\boldsymbol{r}}_{e,k} = \alpha\beta_{e,k}\sqrt{G}\boldsymbol{b}(\theta_{e,k})\boldsymbol{a}^{\mathrm{H}}(\theta_{e,k})\boldsymbol{f}_{e,k} + \widehat{\boldsymbol{z}}_{e,k}, \quad (12)$$

where G is the compensation gain, $\widehat{\boldsymbol{r}}_{e,k} \in \mathbb{C}^{N\times 1}$.

2) Sensing Measurement Model: After the radar's matched filtering, the delay and Doppler are measured as [13]

$$\begin{cases}\tau_{e,k} = \frac{2d_{e,k}}{c} + z_\tau, \\ \delta_{e,k} = \frac{2v_{e,k}\cos\theta_{e,k}f_c}{c} + z_\delta,\end{cases} \quad (13)$$

where c and f_c denote signal propagation speed and the carrier frequency, respectively. Terms z_τ and z_δ represent the measurement noises obeying Gaussian distribution, i.e., $z_\tau \sim \mathcal{N}(0, \sigma^2_\tau)$ and $z_\delta \sim \mathcal{N}(0, \sigma^2_\delta)$. Based on the sensing model, the distance $d_{e,k}$ and the volecity $v_{e,k}$ of the e-th vehicle relative to the RSU at the k-th time slot are obtained.

3) State Evolution Model: Establishing a state evolution model for parameter prediction is crucial. The kinematic equations between two adjacent time slots for each vehicle is modeled based on the geometric relationships and motion parameters. Fig. 3 illustrates the analysis of the e-th vehicle in the considered system, with the vehicle traveling away from the RSU. The state evolution model is expressed as

$$\begin{cases}\triangle d_{e,k-1} = \triangle T v_{e,k-1}, \\ \triangle\theta_{e,k-1} = \theta_{e,k-1} - \theta_{e,k}, \\ d_{e,k}\sin\triangle\theta_{e,k-1} = \triangle d_{e,k-1}\sin\theta_{e,k-1},\end{cases} \quad (14)$$

where $\triangle T$ denotes the time duration of each time slot, and it is assumed that the vehicle is traveling at a constant speed with a short slot $\triangle T$ [13].

4) Angle Prediction Model: It is not sufficient to simply observe and track vehicles. The system should possess predictive capabilities to reduce delays [13]. Generally, the angle variation $\triangle\theta$ is small during $\triangle T$. Adopting the approximation $\triangle\theta \approx \sin\triangle\theta$ [13], (14) is rewritten as

$$\triangle\theta_{e,k-1} \approx \sin\triangle\theta_{e,k-1} = \frac{\triangle d_{e,k-1}\sin\theta_{e,k-1}}{d_{e,k}}. \quad (15)$$

In addition, the approximation $d_{e,k-1} \approx d_{e,k}$ generally holds in two consecutive time slots [13]. Substituting this to (15), the prediction model for $\theta_{e,k}$ is expressed as

$$\tilde{\theta}_{e,k} = \theta_{e,k-1} - \frac{\triangle T v_{e,k-1}\sin\theta_{e,k-1}}{d_{e,k-1}} + z_\theta, \quad (16)$$

where $z_\theta \sim \mathcal{N}(0, \sigma^2_\theta)$ represents the noise generated by the distance approximation.

C. Communication Model

At time slot k, the received downlonk signal at the e-th vehicle is formulated as

$$\begin{aligned}c_{e,k}(t) =& \bar{\alpha}\epsilon_{e,k}e^{\mathrm{j}2\pi t\vartheta_{e,k}}\mathbf{a}^{\mathrm{H}}(\theta_{e,k})\boldsymbol{f}_{e,k}s_{e,k}(t-\iota_{e,k}) \\ &+ z_{\mathrm{c}}(t),\end{aligned} \quad (17)$$

where $\vartheta_{e,k}$ is the Doppler frequency and $\iota_{e,k}$ represent the delay for communication. $z_c(t) \sim \mathcal{CN}(0, \sigma^2_c)$ is the Gaussian noise, $\bar{\alpha} = \sqrt{M}$ denotes the communication array gain factor between the RSU and each vehicle, $\epsilon_{e,k} = \epsilon_0 {d_{e,k}}^{-1}$, where ϵ_0 is the path-loss at reference distance $d_0 = 1$ m. The the receive signal-to-noise ratio (SNR) for the e-th vehicle at the k-th slot is expressed as

$$\mathrm{SNR}_{e,k} = \frac{\left|\bar{\alpha}\epsilon_{e,k}\boldsymbol{a}^{\mathrm{H}}(\theta_{e,k})\boldsymbol{f}_{e,k}\right|^2}{\sigma^2_c} = \frac{\bar{\alpha}^2\epsilon^2_{e,k}\varepsilon_{e,k}}{\sigma^2_c}, \quad (18)$$

where

$$\varepsilon_{e,k} = \left|\boldsymbol{a}^{\mathrm{H}}(\theta_{e,k})\boldsymbol{a}(\tilde{\theta}_{e,k})\right|^2 \leq 1. \quad (19)$$

From (19), it is observed that there will be a peak for $\mathrm{SNR}_{e,k}$ when $\tilde{\theta}_{e,k} = \theta_{e,k}$. In particular, the achievable sum-rate R_k for all vehicles in the system at the k-th time slot is given by

$$R_k = \sum_{e=1}^{E}\log_2\left(1 + \mathrm{SNR}_{e,k}\right). \quad (20)$$

The sensing and communication models indicate that the design of the transmit beamformer, which relies on sensing results, serves as the foundation for beam alignment. This is a critical factor in enhancing the communication rates of the system.

IV. Problem Formulation

The focus of this section is to analyze the optimal balance between the sensing capabilities and computational complexity in ISAC-V2I systems. Specifically, the evaluation metrics of DNN are introduced, followed by the formulation of the trade-off problem.

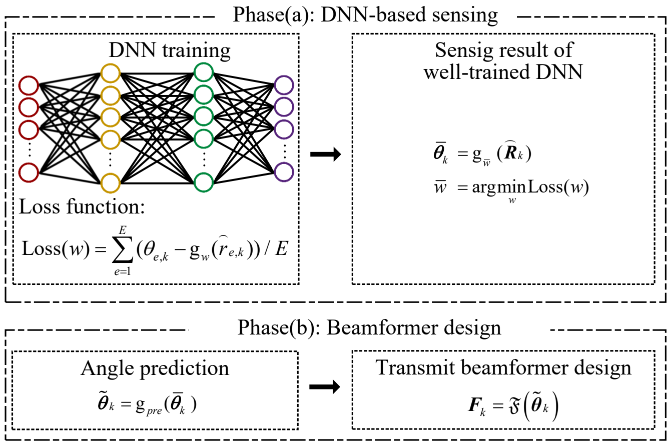


Fig. 4: The predictive bamforming framework for V2I system.

A. Evaluation Metrics of DNN

Generally, there are three primary evaluation metrics for DNN, namely performance metrics, total parameter, and computational complexity. The details are as follows.

1) Performance Metrics: Evaluation criteria based on performance metrics are widely applied and are typically derived from the output results of DNN. Performance metrics are classified into several types, such as classification accuracy, estimation error, and detection accuracy.

2) Total parameters: The total number of parameters is determined by the intrinsic property of DNN model. To a certain degree, the potent learning ability of one DNN model is granted by its learnable parameters. The learning ability of the DNN is enhanced when the number of parameters in the DNN is increased , which improves the target performance of the model. However, the increase in the number of parameters results in a larger memory footprint and longer training cycles, consequently increasing the expense of the DNN.

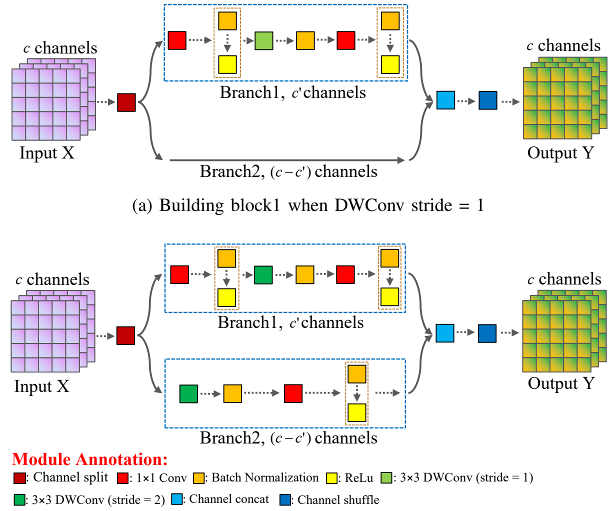
3) Computational Complexity: The computational complexity of DNN is closely related to its characteristics. Specifically, it quantifies the amount of computation required for a DNN to process the given input. The computational complexity increases when the DNN performs tasks such as feature extraction, error backpropagation, and nonlinear mapping. The most commonly applied measure of computational complexity is the MACC. larger MACC means higher power consumption during the execution of the corresponding DNN model.

B. Problem Formulation

ϱ is adopted to indicate the sensing performance of the system, which is the root mean squared error (RMSE) on angles and is calculated by averaging the entire observation time and the all test samples. The mathematical expressions for ϱ will be shown in Section V-D. Computational complexity, on the other hand, is evaluated by the MACC of the entire DNN. Therefore, the trade-off performance is formulated as

$$\eta = \mathrm{MACC} \times \varrho, \tag{21}$$

where η is utilized to quantify the efficiency of the DNN. Specifically, the trade-off performance of the DNN is considered superior while η is smaller.



(a) Building block1 when DWConv stride = 1

(b) Building block1 when DWConv stride = 2

Fig. 5: Building blocks of SACSMN.

V. The Proposed SACSMN-Based Predictive Beamforming Approach

In this section, we initially present the DNN-based predictive beamforming framework in the ISAC-V2I system, followed by the introduction of building block (BB) in the proposed SACSMN. Thereafter, the SACSMN model is designed and finally the SACSMN-based predictive beamforming algorithm is proposed.

A. DNN Based Predictive Beamforming Framework in V2I system

As depicted in Fig. 4, the predictive beamforming framework consists of two phases, i.e., (a) DNN-based sensing and (b) transmit beamformer design. In phase (a), the DNN is adopted to perform sensing, i.e.,

$$\boldsymbol{\theta}_k^* = g_{w^*}(\widehat{\boldsymbol{R}}_k), \tag{22}$$

where $\boldsymbol{\theta}_k^* = [\theta_{1,k}^*, \theta_{2,k}^*, \cdots, \theta_{E,k}^*] \in \mathbb{R}^{1\times E}$ denotes the sensing result during the network model training, $\widehat{\boldsymbol{R}}_k = [\widehat{\boldsymbol{r}}_{1,k}, \widehat{\boldsymbol{r}}_{2,k}, \cdots, \widehat{\boldsymbol{r}}_{E,k}] \in \mathbb{C}^{N\times E}$. $g_{w^*}(\cdot)$ represents the mapping function from the echo signals at the receive ULA to the sensing result $\boldsymbol{\theta}_k^*$, and w^* denotes the model weight of the DNN model in training. In additon, the loss function of DNN is defined by

$$\mathrm{Loss}(w^*) = \sum_{e=1}^{E}(\theta_{e,k} - g_{w^*}(\widehat{\boldsymbol{r}}_{e,k}))/E. \tag{23}$$

To minimise (23), the model weight w^* is updated after each training epoch. Following the whole training, the optimal sensing result output by the well-trained model is expressed as

$$\bar{\boldsymbol{\theta}}_k = g_{\bar{w}}(\widehat{\boldsymbol{R}}_k), \tag{24}$$

where $\bar{\boldsymbol{\theta}}_k$ represents the sensing result of well-trained DNN, and $\bar{w} = \arg\min_{w} \mathrm{Loss}(w)$ denotes the trained optimal model weight. In phase (b), after obtaining the sensing result, the angle prediction, denoted by $g_{pre}(\cdot)$, is performed based on

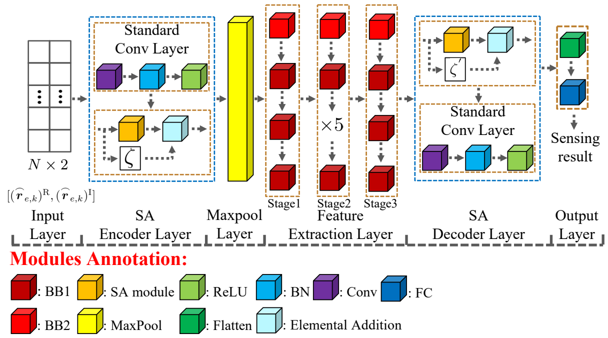


Fig. 6: The overall structure of SACSMN.

(16), and the transmit beamformer $\boldsymbol{F}_k$ is designed based on (5) at RSU.

B. Building blocks of SACSMN

The backbone of SACSMN is constructed from a series of stacked BBs, and the BB is illustrated in Fig. 5.

Based on the strides of the DWConv operation, two types of BBs, namely BB1 and BB2, are identified. At the beginning of each BB, the c channels of the input feature map $\boldsymbol{X}$ are divided into two branches, i.e., c' and $c - c'$, via channel splitting operation. BB1 executes a sequence of operations in branch 1, namely 1×1 convolution operation, 3×3 DWConv with stride = 1, and 1×1 convolution operation, which perform feature aggregation. Batch normalization (BN) operation is incorporated to prevent gradient vanishing. The rectified linear unit (ReLu) function performs a sequence of linear mappings, resulting in a transformation of low-level features into high-level ones. As for branch2 in BB1, it does not modify the input. The two branches are then concatenated in the channel dimension, and a channel shuffle operation follows. In BB2, only branch2 is modified compared to BB1, and it consists of a 3×3 DWConv with stride = 1, a BN operation, a 1×1 convolution operation, a BN operation, and a ReLu function, as shown in Fig. 5 (b).

C. The Proposed SACSMN Model

The SACSMN is proposed in ISAC-V2I system, and the comprehensive model architecture is depicted in Fig. 6, while the full specifications are summarized in Table II. It is observed from Fig. 6 that SACSMN is composed of six primary components, including the input layer, SA encoder layer, maxpool layer, feature extraction layer, SA decoder layer, and output layer. The entire SACSMN is detailed as follows.

1) Input Layer: Based on (12), the received echo signal after filtering is a complex vector. However, the DNN cannot directly process the complex vector. As a result, the following transformation is performed first:

$$\widehat{\boldsymbol{r}}_{e,k} \Rightarrow (\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{R}} + \mathrm{j}(\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{I}}, \tag{25}$$

where $(\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{R}} \in \mathbb{R}^{N \times 1}$ and $(\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{I}} \in \mathbb{R}^{N \times 1}$ represent the real part and imaginary part of $\widehat{\boldsymbol{r}}_{e,k}$, respectively. By omitting the imaginary units, a new form of expression is obtained, i.e., $[(\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{R}}, (\widehat{\boldsymbol{r}}_{e,k})^{\mathrm{I}}]$, and the input feature map is change from a complex vector to a real matrix of size $(N \times 2)$. The input

TABLE II: The Full Specification of SACSMN

Layer	Output size	Kernel size	Stride	Repeat	Output channels
Input layer	$N \times 2$	-	-	-	1
SA encoder layer	$N \times 2$	1×1	1	1	24
Maxpool	$(N/2) \times 2$	3×3	2	1	24
Feature extraction 1	$(N/4) \times 2$ $(N/4) \times 2$	- -	2 1	1 3	48 48
Feature extraction 2	$(N/8) \times 2$ $(N/8) \times 2$	- -	2 1	1 7	96 96
Feature extraction 3	$(N/16) \times 2$ $(N/16) \times 2$	- -	2 1	1 3	192 192
SA encoder layer	$(N/16) \times 2$	1×1	1	1	192
Flatten layer	$N/8$	-	-	-	192
Output layer	1	-	-	-	-

layer is only responsible for feeding the echo signal values into the DNN and does not do any other processing.

2) SA Encoder Layer: This layer includes two distinct sub-layers, namely the standard convolutional layer and the SA layer. Initially, the standard convolution operation is executed on the input feature map in the first sub-layer, with subsequent application of BN and ReLU nonlinearity. In the second sub-layer, the SA mechanism is implemented to apply distinct weighting to the output of the first sub-layer. Additionally, the secondary weighting and shortcut operations are utilized in this layer, which is expressed as

$$\boldsymbol{Z}_{en} = \zeta_{en} \boldsymbol{Y}_{en} + \boldsymbol{X}_{en}, \tag{26}$$

where $\boldsymbol{X}_{en}$, $\boldsymbol{Y}_{en}$, and $\boldsymbol{Z}_{en}$ correspond to the output of the first sub-layer, the SA layer output, and the output of the SA encoder layer, respectively. ζ_{en} represents a learnable scalar that performs element-wise multiplication with $\boldsymbol{Y}_{en}$. Furthermore, during the SACSMN training process, ζ_{en} will be updated. (26) is the mathematical representation of the proposed dynamic SA, which assigns more reasonable weights to the features by the learnable ζ_{en} and mitigates the gradient vanishing problem of the model by the residual structure.

3) Maxpool Layer: In this layer, spatial down-sampling is accomplished through the selection of the maximum value within the area encompassed by the convolution kernel. This approach has the benefit of reducing computational costs.

4) Feature Extraction Layer: The feature extraction layer is constructed based on BBs depicted in Fig. 5. This layer is comprised of three stages, with each stage consisting of a stack of BBs. With stacked BBs, the ability of the model to extract different features will be enhanced to improve the sensing performance. As for the internal relationship between neighboring stacked BBs, at the data processing level, each BB performs data processing independently to maintain the flexibility and trainability of the model. Each BB obtains input data from the previous BB when processing data and transmits the processed output data to the subsequent BB. During each stage, BB2 serves as the initial component, while BB1 is repeated a designated number of times for subsequent operations. Additionally, within each BB, the channel split operation maintains an equal number of channels in both branch1 and branch2 of the input feature matrix.

5) SA Decoder Layer: This layer is similar to the SA encoder layer, where the SA layer is performed initially,

Algorithm 1: SACSMN-based Predictive Beamforming Algorithm

Offline Training:
Initialization:
  $epoch = 0$
  $Epoch = Epoch_{max}$
  φ^* with random weights
  one training set $\boldsymbol{\Omega}^{tr}$
Input: Training set $\boldsymbol{\Omega}_{tr}$
for *epoch* = 1 : *Epoch* **do**
  Calculate $\mathrm{L}(\varphi^*)$ as defined in (28)
  Update φ^* by BP to minimize $\mathrm{L}(\varphi^*)$
  epoch ++
end
Output: Well-trained $\mathrm{g}_{\bar{\varphi}}(\cdot)$ and sensing result $\boldsymbol{\theta}_k^*$ as defined in (29)
Online Testing:
Input: Randomly selected testing sample $\boldsymbol{X}_k^{nte}$ in $\boldsymbol{\Omega}^{te}$
Output: Sensing result based on well-trained SACSMN
do:
  Angle prediction based on state evolution model
  Transmit beamformer design

followed by the standard convolutional layer. The output of SA layer in this part is expressed as

$$\boldsymbol{Z}_{de} = \zeta_{de}' \boldsymbol{Y}_{de} + \boldsymbol{X}_{de}, \tag{27}$$

where $\boldsymbol{X}_{de}$, $\boldsymbol{Y}_{de}$ and $\boldsymbol{Z}_{de}$ represent the outputs of the feature extraction layer, the direct output of the SA layer, and the final output of the SA module, respectively. The scalars ζ_{en} and ζ_{de}' are similar, but they are independent of each other.

6) Output Layer: This layer is composed of the flatten sublayer and FC sublayer. The former serves to expand the last two dimensions of the input, such that when a feature map with dimensions $(batchsize, channels, height, weight)$ is used as input, the dimension of the output of the flatten sublayer is $(batchsize, channels, height \times weight)$. Meanwhile, the FC sublayer executes a linear weighted summation of the output from the flattened sublayer to produce the sensing result.

D. SACSMN-Based Predictive Beamforming Algorithm

The SACSMN-based predictive beamforming algorithm is proposed in this section, which adopts the proposed SACSMN as the core and consists of offline training and online beamforming design. The details of the algorithm will be illustrated in the following.

1) Offline Training: The training set with N_{tr} system samples, i.e., $\boldsymbol{\Omega}^{tr} = [\boldsymbol{\chi}^1, \boldsymbol{\chi}^2, \cdots, \boldsymbol{\chi}^{N_{tr}}]$, is considered in this part. $\boldsymbol{\chi}^{ntr} = [(\boldsymbol{R}_1^{ntr}, \boldsymbol{\theta}_1^{ntr}), (\boldsymbol{R}_2^{ntr}, \boldsymbol{\theta}_2^{ntr}), \cdots, (\boldsymbol{R}_K^{ntr}, \boldsymbol{\theta}_K^{ntr})]$, $ntr \in \{1, 2, \cdots, N_{tr}\}$, denotes the signal samples for the entire observation period in the ntr-th system sample. $(\boldsymbol{R}_k^{ntr}, \boldsymbol{\theta}_k^{ntr})$ represents the signal sample at the k-th slot in the ntr-th system sample, where $\boldsymbol{R}_k^{ntr} = [\widehat{\boldsymbol{r}}^{ntr}_{1,k}, \widehat{\boldsymbol{r}}^{ntr}_{2,k}, \cdots, \widehat{\boldsymbol{r}}^{ntr}_{E,k}]$ denotes the echoes for the vehicles at the k-th time slot, $\boldsymbol{\theta}_k^{ntr} = [\theta_{1,k}^{ntr}, \theta_{2,k}^{ntr}, \cdots, \theta_{E,k}^{ntr}]$ denotes the actual angle parameters for the vehicles. Based on the proposed network model and the predictive beamforming framework, the loss function for SACSMN is expressed as

$$\mathrm{L}(\varphi^*) = \frac{1}{N_{tr}} \frac{1}{K} \frac{1}{E} \sum_{ntr=1}^{N_{tr}} \sum_{k=1}^{K} \sum_{e=1}^{E} \left| \theta_{e,k}^{ntr} - \mathrm{g}_{\varphi^*}(\widehat{\boldsymbol{r}}^{ntr}_{e,k}) \right|, \tag{28}$$

where φ^* represents the model weight of SACSMN. In the training process of SACSMN, back-propagation (BP) algorithm is adopted to optimize φ^* according to (28) in each training epoch, which in turn minimizes $\mathrm{L}(\varphi^*)$ to achieve convergence. After the DNN training, SACSMN that learns the mapping relationship between echo signals and sensing result is expressed mathematically as

$$\mathrm{g}_{\bar{\varphi}}(\widehat{\boldsymbol{R}}_k) = \boldsymbol{\theta}_k^*, \tag{29}$$

where $\mathrm{g}_{\bar{\varphi}}(\cdot)$ represents the well-trained SACSMN with optimal model weight $\bar{\varphi}$, and $\boldsymbol{\theta}_k^*$ denotes the final sensing result based on the well-trained SACSMN.

2) Online beamforming design: The testing set $\boldsymbol{\Omega}_{te}$ with N_{te} samples, i.e., $\boldsymbol{\Omega}^{te} = [\boldsymbol{\Phi}^1, \boldsymbol{\Phi}^2, \cdots, \boldsymbol{\Phi}^{N_{te}}]$, is of the same form as the training set, and both sets are independent of each other. $\boldsymbol{\Phi}^{nte} = [(\boldsymbol{X}_1^{nte}, \boldsymbol{y}_1^{nte}), (\boldsymbol{X}_2^{nte}, \boldsymbol{y}_2^{nte}), \cdots, (\boldsymbol{X}_K^{nte}, \boldsymbol{y}_K^{nte})]$, $nte \in \{1, 2, \cdots, N_{te}\}$. During the online testing process, the testing angle parameter, i.e., $\boldsymbol{y}_k^{nte}$, $\forall nte$, $\forall k$ will not be fed into the SACSMN. Therefore, one testing sample is randomly selected from $\boldsymbol{\Omega}^{te}$, i.e., $\boldsymbol{X}_k^{nte}$, and fed into the well-trained SACSMN. After that, a series of steps will be performed according to the framework in Fig. 4 to design the transmit beamformer

$$\boldsymbol{F}_k^* = \mathfrak{F}(\mathrm{g}_{pre}(\mathrm{g}_{\bar{\varphi}}(\boldsymbol{X}_k^{nte}))), \tag{30}$$

and the sensing performance ϱ is formulated as [13]

$$\varrho = \frac{1}{N_{te}} \frac{1}{K} \frac{1}{E} \sum_{nte=1}^{N_{te}} \sum_{k=1}^{K} \left\| \boldsymbol{y}_k^{nte} - \mathrm{g}_{\bar{\varphi}}(\boldsymbol{X}_k^{nte}) \right\|_1. \tag{31}$$

3) Algorithm Summary: The proposed predictive beamforming algorithm is summarized in Algorithm 1, where *epoch* denotes the training epoch index and $Epoch_{max}$ represents the total amount of training epoch.

VI. Simulation Results

The simulation results are presented to validate the effectiveness of the proposed SACSMN for predictive beamforming in this section. We consider the ISAC-V2I system with one RSU and three vehicles, and the system operates in the 30 GHz band. Without loss of generality, the Cartesian coordinate system is applied in the simulation environment. All vehicles depart from the RSU in a direction parallel to the road, with the direction considered as the positive half-axis of the x-axis. The RSU is located at [0m, 0m], and the number of antennas for both transmit ULA and receive ULA is adjustable. The vehicles are initialized at [25m, 20m], [30m, 20m] and [35m, 20m], respectively. The duration for each time slot is $\triangle T = 0.02$s, and the velocity of each vehicle follow a uniform distribution of $[10, 15]$m/s during the considered time slots.

A. Dataset Description

Since the studied probelm and the proposed method are novel, there are no available open datasets. Therefore, we generated the corresponding dataset based on our proposed signal model. The duration of each time slot in the data set is $\triangle T = 0.02$s. The system operates in the 30GHz frequency band, with the vehicles traveling away from the RSU. The sensing compensation gain G and complex radar cross-section γ are set to 10 and $10 + 10$j, respectively. The dataset is comprised of training as well as the test sets. The training set consists of one RSU located at [0m, 0m] and three vehicles positioned at [20m, 20m], [25m, 20m], and [30m, 20m], respectively. Furthermore, the training set is divided into three parts based on the RSU configuration, namely $M = N = 16$, $M = N = 32$, and $M = N = 64$. Additionally, in our assumpations, $N_{tr} = 2000$ is performed for each type of antenna configuration, and time slots $K = 512$ is considered for each simulation. This leads to 1024×10^6 echo samples for each RSU configuration in the training set. To ensure the consistency of the data distribution, the system parameters of the test set are the same as the training set, except while $N_{te} = 3000$. Then, all simulation results are averaged from the test set, i.e., 3000 independent simulations.

B. The Influence of Channel Shuffle Parameter g

The influence of channel shuffle parameter g of SACSMN on the sensing performance is the focus of this part.

Since the channel shuffling operation has no influence on the number of parameters and the MACC, the results are evaluated based on the sensing error ϱ. In addition, there is no inter-channel information exchange during convolution operations when $g = 1$. Table III presents the experimental datas, which clearly indicate the efficiency of the channel shuffle operation in improving the sensing performance of SACSMN. The results show that the sensing errors generally decrease and then increase as the g grows. Specifically, the proposed model usually achieves the best sensing performance when $g = 2$. This is because moderate value of g facilitates the interaction of information between channels and contributes to improved feature extraction ability of the model. However, the data stream containing valuable information is segmented when the g is large, which leads to loss of information and disruption of the original information flow, thus degrading the sensing performance. Therefore, SACSMN with $g = 2$ will be employed for further performance comparison in subsequent experiments.

C. The Necessity of SA Encoder & Decoder Layers

In this subsection, we investigate the necessity of the SA encoder and decoder layers, which are fundamental modules in SACSMN. The experimental results are presented in Tables IV. In the benchmark SACSMN (no SA), the dynamic SA module is replaced by a standard convolutional layer comprising convolutional operation, BN, and ReLU. It is obvious that the proposed SACSMN achieves superior sensing performance than the comparator under the same experimental settings, which implies that proposed dynamic SA significantly improves the model sensing capability. This is because the proposed dynamic SA module is able to pay more attention on the important feature information in the feature map and ignore the trivial features, which reduces the interference in the feature extraction process and improves the feature extraction capability of the model, allowing the model to extract more general and effective features.

TABLE III: The influence of g on sensing error for different training sample sizes.

(a) 1% of the training set

	$g = 1$	$g = 2$	$g = 4$	$g = 8$
$M = N = 16$	2.57E-02	**1.08E-02**	2.57E-02	1.77E-02
$M = N = 32$	2.12E-02	**1.03E-02**	2.12E-02	1.47E-02
$M = N = 64$	2.73E-02	**1.81E-02**	2.73E-02	1.92E-02

(b) 9% of the training set

	$g = 1$	$g = 2$	$g = 4$	$g = 8$
$M = N = 16$	3.17E-03	**1.74E-03**	2.00E-03	2.94E-03
$M = N = 32$	1.68E-03	**1.63E-03**	1.83E-03	1.96E-03
$M = N = 64$	1.82E-03	**8.00E-04**	1.19E-03	1.14E-03

(c) 100% of the training set

	$g = 1$	$g = 2$	$g = 4$	$g = 8$
$M = N = 16$	2.02E-03	**9.92E-04**	1.61E-03	1.56E-03
$M = N = 32$	1.00E-03	**4.67E-04**	5.08E-04	5.15E-04
$M = N = 64$	3.63E-04	**2.28E-04**	2.72E-04	2.63E-04

TABLE IV: The influence of SA on sensing error for different training sample sizes.

(a) 1% of the training set.

	$M = N = 16$	$M = N = 32$	$M = N = 64$
SACSMN	**1.08E-02**	**1.03E-02**	**1.81E-02**
SACSMN(no SA)	3.30E-02	2.64E-02	1.86E-02

(b) 9% of the training set.

	$M = N = 16$	$M = N = 32$	$M = N = 64$
SACSMN	**1.74E-03**	**1.63E-03**	**8.00E-04**
SACSMN(no SA)	2.71E-03	1.88E-03	1.74E-03

(c) 100% of the training set.

	$M = N = 16$	$M = N = 32$	$M = N = 64$
SACSMN	**1.56E-03**	**4.67E-04**	**3.28E-04**
SACSMN(no SA)	2.09E-03	1.28E-03	4.81E-04

D. Sensing performance

The sensing performances under various few-shot training scenarios are examined in this part. Specifically, the SACSMN is trained based on 1%, 5%, 9%, and 100% of the training set samples to investigate the efficiency. We compare the proposed model with several widely used DNN models, including AlexNet [22], CLDNN [23], MobileNet-V3 [19], ResNet18 [24], and ShuffleNet-V2 [20].

The angle sensing performance are presented in Tables V, and the system sensing error is denoted by ϱ_1, ϱ_5, ϱ_9, and ϱ_{100}, which correspond to the DNNs' training with 1%, 5%, 9%, and 100% of the samples in the training set, respectively. Additionally, η defined in (21) is utilized to quantify the performance of the models in terms of the trade-off between computational complexity and sensing performance. The analysis are summarized as follows: First, the effect of the configuration at RSU (i.e., N) on the parameters and MACC is examined. Second, we investigate the impact of the training sample size on sensing performance. Furthermore, the metric η is employed to evaluate the trade-off performance of each network model. Finally, the influences of the initial

TABLE V: The sensing trade-off performance under different conditions

(a) $M = N = 16$

	Params(M)	MACC(M)	ϱ_1	ϱ_5	ϱ_9	ϱ_{100}	η_1(M)	η_5(M)	η_9(M)	η_{100}(M)
AlexNet	8.96	34.74	89.87	89.78	54.03	3.24E-01	3122.22	3118.90	1876.95	11.24
CLDNN	8.64	8.77	3.33E-03	**6.33E-04**	9.21E-03	**1.33E-04**	2.92E-02	5.55E-03	8.08E-02	1.17E-03
MobileNet V3	4.2	14.62	134.60	2.74E-03	5.67E-03	3.50E-04	1.97E+03	4.01E-02	8.30E-02	5.12E-03
ResNet18	11.17	357.77	**4.66E-04**	7.56E-04	**2.61E-04**	1.32E-03	1.67E-01	2.71E-01	9.34E-02	4.72E-01
ShuffleNet V2	1.25	1.7	9.48E-03	2.89E-03	3.48E-03	8.36E-03	1.61E-02	4.91E-03	5.91E-03	1.42E-02
SACSMN	**0.45**	**0.6**	1.08E-02	5.74E-03	1.74E-03	9.92E-04	**6.50E-03**	**3.45E-03**	**1.04E-03**	**5.95E-04**

(b) $M = N = 32$

	Params(M)	MACC(M)	ϱ_1	ϱ_5	ϱ_9	ϱ_{100}	η_1(M)	η_5(M)	η_9(M)	η_{100}(M)
AlexNet	13.16	65.54	90.53	89.60	89.10	46.03	5933.66	5872.53	5905.36	3016.99
CLDNN	33.81	34.1	3.29E-03	7.10E-03	2.10E-03	3.77E-04	1.12E-01	2.42E-01	7.14E-02	1.29E-02
MobileNet V3	4.2	26.49	62.09	2.10E-03	1.31E-03	1.21E-04	1.64E+03	5.57E-02	3.46E-02	3.21E-03
ResNet18	11.17	715.54	**3.48E-04**	**6.89E-05**	**2.08E-04**	**6.06E-05**	2.49E-01	4.93E-02	1.49E-01	4.34E-02
ShuffleNet V2	1.25	3.41	3.99E-03	8.99E-04	1.29E-03	3.54E-04	1.36E-02	3.07E-03	4.41E-03	1.21E-03
SACSMN	**0.45**	**1.22**	1.03E-02	2.46E-03	1.63E-03	4.67E-04	**1.26E-02**	**3.01E-03**	**1.98E-03**	**5.70E-04**

(c) $M = N = 64$

	Params(M)	MACC(M)	ϱ_1	ϱ_5	ϱ_9	ϱ_{100}	η_1(M)	η_5(M)	η_9(M)	η_{100}(M)
AlexNet	21.55	127.14	90.96	90.41	59.21	9.80	11564.45	11622.12	7527.82	1245.64
CLDNN	134.49	135.09	7.39E-03	5.60E-03	4.76E-03	2.09E-03	9.98E-01	7.57E-01	6.44E-01	2.82E-01
MobileNet V3	4.2	50.25	12.37	1.12E-03	6.75E-04	1.40E-04	6.21E+02	5.62E-02	3.39E-02	7.03E-03
ResNet18	11.17	1431.09	**8.68E-04**	**3.66E-04**	**1.82E-04**	**1.36E-05**	1.24E+00	5.23E-01	2.61E-01	1.94E-02
ShuffleNet V2	1.25	6.81	3.97E-03	1.38E-03	8.13E-04	1.22E-04	2.71E-02	9.42E-03	5.54E-03	8.34E-04
SACSMN	**0.45**	**2.43**	1.81E-02	2.37E-03	8.00E-04	3.28E-04	**4.40E-02**	**5.77E-03**	**1.94E-03**	**7.98E-04**

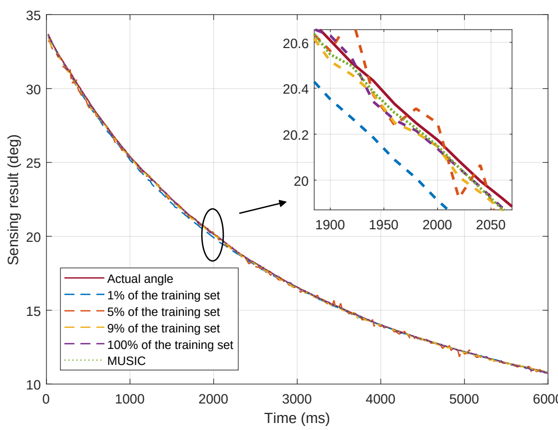


Fig. 7: Sensing performance comparison between SACSMN and MUSIC

vehicle position and direction of vehicle travel on the sensing performance is explored.

1) The Influence of the Configuration N at RSU on the Parameters, MACC and Sensing Performance: It is evident that the parameters of the initial two DNNs increase proportionally with N, since the number of neurons in their FC layers is linked to N. When other conditions remain constant, the total number of parameters in the DNN will be increased when the sum of neurons in the FC layer grows. The structures of the remaining DNN models remains largely unaffected by N, thus keeping the number of parameters constant. In addition, the MACCs of all DNN models exhibit a positive correlation with N, and both parameters and MACCs of the proposed model are minimal for the exact RSU configuration. Furthermore, it is worth noting that when the DNN model and the training sample size remain constant, the sensing errors of most DNNs decrease as N increases. This is because more features of echo sample are provided when N is larger, which improves the discrimination of the corresponding angle.

2) The Influence of Training Sample Size on Sensing Performance: The sensing errors of most DNN models decrease with more training samples when the ULA size is constant. This is because, when all other conditions remain constant, more training samples enable the models to better learn the nonlinear relationships between echo signals and angles. However, the sensing performances of some DNNs are improved and then degraded with the increase of training samples, possibly due to overfitting. Notably, the sensing performance of ResNet18 is nearly optimal under the same conditions. It is evident that the sensing performances of SACSMN are deteriorated as the ULA size increases when training with only 1% of the training set samples. This may be due to the fact that when the training samples are fewer and the model obtains more features of each echo signal, SACSMN suffers from interference leading to a decrease in its feature extraction capability. Moreover, when the training sample size is increased by the same magnitude, i.e., from 1% to 9%, SACSMN with $N = 64$ achieves a greater sensing performance improvement. It is observed that the sensing performance achieved by the 9% training sample is very similar to the 100% performance when $N = 64$, indicating that further increasing the training sample size does not yield significant sensing performance gains.

Fig. 7 provides a further elaboration of the sensing performance of SACSMN with respect to $\theta_{2,k}$, where multiple signal classification (MUSIC) [25] based scheme is viewed as upper bounds. It is found that the sensing performance of SACSMN based on 1% of the training set is the worst. This is because the sample size is insufficient in this case, and SACSMN cannot effectively learn the general features and patterns between the echoes and the angles, weakening the feature extraction ability in the test set. On the other hand, when the training samples are insufficient, SACSMN overlearns specific features and noises in the training data, resulting in poor sensing performance when dealing with new and unseen data in the test set. The sensing performance of SACSMN is gradually improved with more training samples. This is because more training samples contain a wide variety of scenarios and noises, which helps to train a more robust

TABLE VI: Default settings in the simulation

ULA size	Initial angle parameter	Initial vehicle location
M=N=16	$\theta_0 = 11°$	[104.17m, 20m]
M=N=32	$\theta_0 = 14°$	[81.85m, 20m]
M=N=64	$\theta_0 = 16°$	[71.62m, 20m]

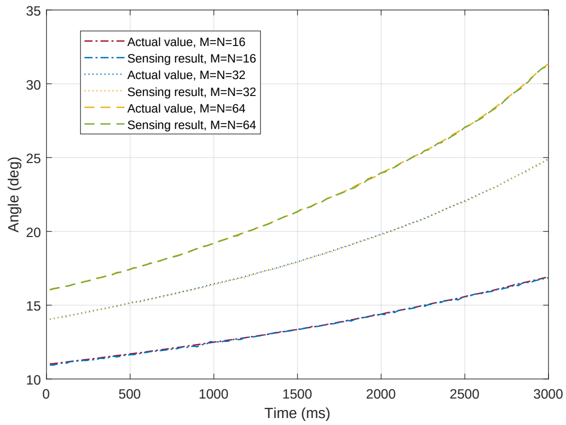


Fig. 8: Sensing results based on the SACSMN trained from 9% of the training set.

SACSMN that achieves stable sensing performance in the test set. On the other hand, more training data means that the model is exposed to more types of echoes, which helps to improve the feature extraction ability of SACSMN in the test set. Since the exhaustive research method is adopted in the MUSIC, the sensing performance of MUSIC is optimal. In addition, it is observed that the sensing performance of SACSMN based on 9% of the training set is very close to the actual angle parameters and the system sensing performance is not significantly improved when more training samples are provided. The above results demonstrate that SACSMN extracts more general and effective features from the limited echoes. The feature extraction ability of SACSMN is improved and satisfactory sensing performance is achieved in the case of few samples.The dependence of SACSMN on training samples is significantly reduced.

3) The Sensing Trade-off Performance of SACSMN under Different Conditions: As previously mentioned, the efficiency of the DNNs is measured by η. It is observed that ResNet18 achieves optimal sensing performance in most cases. However, this comes at a high computational cost. For example, when $M = N = 32$, the MACC of ResNet18 is the highest compared to other models. In contrast to SACSMN, the high computational cost of ResNet18 is insignificant for sensing performance improvement, and this is even more evident when $M = N = 64$. Compared with benchmarks, SACSMN achieves similar sensing performance while significantly reducing the number of model parameters and computational complexity. The experimental results in Tables V fully demonstrate that under the same experimental conditions, the sensing trade-off performance parameter η corresponding to SACSMN is the smallest, and SACSMN realizes the optimal trade-off between the system sensing performance and the model computational complexity.

4) The Influence of Initial Vehicle Position and Direction of Vehicle Travel on the Sensing Performance: In this part, the influence of the initial vehicle positions and directions of vehicle travel on the sensing performance is explored. In the simulation, three ULA sizes are considered, and the initial angle parameter and location of each vehicle are shown in Table VI. In each scenario, the vehicle drives towards the RSU.

The sensing results are depicted in Fig. 8, which are calculated based on the SACSMN trained from 9 % of the training set. It is observed that the angular parameters of the vehicles gradually increase as the vehicles approach the RSU. Even though all the vehicles drive away from the RSU in the training set, high precision sensing is still achieved when all the vehicles drive towards the RSU in the test phase. This is because SACSMN implements sensing based on echo signals, which do not contain information about the vehicle's travel direction. The above experiment demonstrates that the directions of travel of the vehicles and the initial vehicle locations have a very limited impact on the sensing performance of the SACSMN.

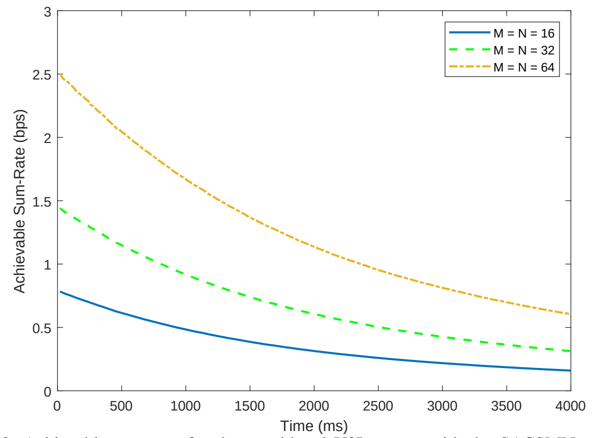
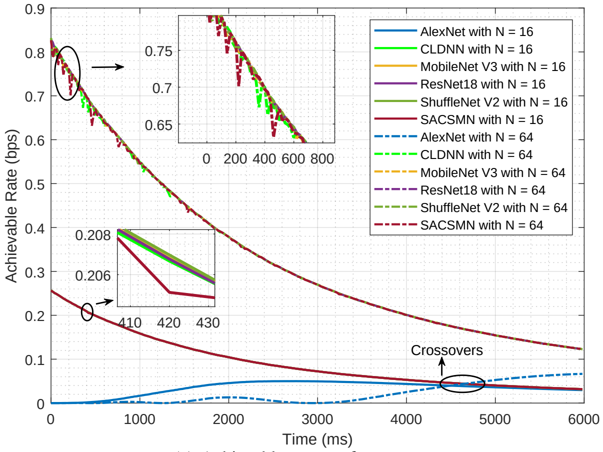


Fig. 9: Achievable sum-rate for the considered V2I system with the SACSMN trained from 100% of the training set.

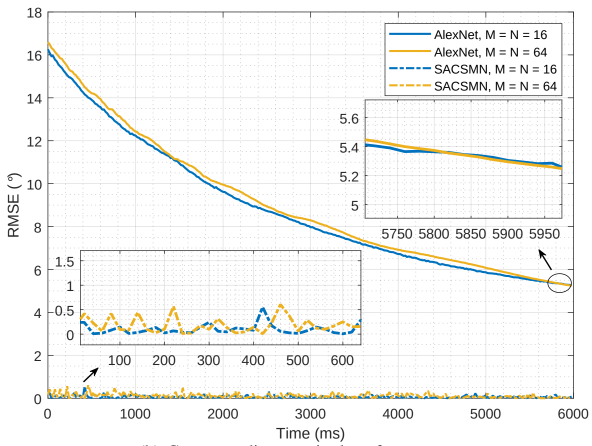
E. Communication Performance

In this subsection, the communication performance are investigated. Initially, the influence of the RSU configuration N on the communication rate is analysed. After that, the comparisons between SACSMN and the benchmarks in the case of few samples are presented. What's more, the effect of training sample size on the communication rates for SACSMN is discussed. Lastly, the influences of the initial vehicle position and direction of vehicle travel on the communication performance is explored.

1) The Influence of RSU Configuration Parameter on Achievable Sum-Rate for SACSMN: The RSU configuration parameters M and N are the fundamental parameters in the considered system, and the simulations are carried out based on SACSMN trained from the complete training set in this experiment. The experimental results are illustrated in Fig. 9. It is apparent that the sum-rates decline over time, owing to the persistent increase of the distance between the vehicles and the RSU, which causes a more significant communication path loss. Furthermore, it is notable that the communication rates are progressively improved as the RSU antenna array size

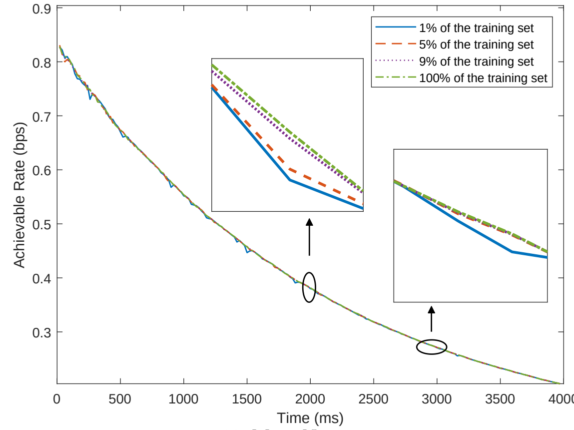


(a) Achievable rate performances.

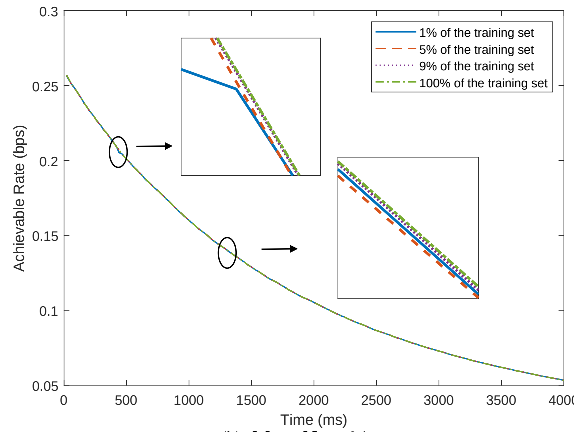


(b) Corresponding sensingl performances.

Fig. 10: Achievable rate performances for a single vehicle with initial state $[30\text{m}, 20\text{m}]$.



(a) $M = N = 16$



(b) $M = N = 64$

Fig. 11: Achievable rate for a single vehicle with initial state $[30\text{m}, 20\text{m}]$.

increases, since a larger ULA leads to greater communication array gain. All of these phenomena are explained by (18).

TABLE VII: Average rates of each model based on 100% of the training set

	$M = N = 16$	$M = N = 64$
AlexNet	3.42E-02 bps	2.03E-02 bps
CLDNN	9.40E-02 bps	3.33E-01 bps
MobileNet V3	9.40E-02 bps	3.34E-01 bps
ResNet18	9.40E-02 bps	**3.35E-01** bps
ShuffleNet V2	9.40E-02 bps	3.34E-01 bps
SACSMN	**9.40E-02** bps	3.32E-01 bps

2) Comparison With Other DNNs in Communication Performancess under Few-shot Scenarios: SACSMN is compared with the benchmarks, such as AlexNet, CLDNN and MobileNet V3, under the few-shot condition in this part. Specifically, 1% of the training set is used to train the models. The initial state of the single vehicle is set to [30m, 20m], and two antenna array sizes are considered. The average communication rates for each model are summarized in Table VII. SACSMN stands out as it achieves comparable communication performance to other models with the least total number of parameters and lowest computational cost for the same RSU configuration. Compared to the benchmark, SACSMN achieves higher communication efficiency.

The Fig. 10(a) shows the achievable rates of all considered DNN models for a specific test sample, while the corresponding sensing performances based on AlexNet and the SACSMN are illustrated in Fig. 10(b). Throughout the simulation, it is observed that the communication rate curves are closely aligned with each other when the ULA size is the same. Therefore, it is approximated that all DNNs, except AlexNet, achieve similar communication performance. Furthermore, the majority of DNNs experience a gradual decline in their communication rates as the distances increase.

In the case of SACSMN, the communication performance curves display more fluctuations during the initial period when $M = N = 64$ compared to $M = N = 16$. As shown in Fig. 10(b), this is due to the larger sensing error of SACSMN at this stage, which leads to the degradation of the beam alignment performance. Furthermore, the larger communication array gain exacerbates the reduction in the rate. When $M = N = 64$, during $t \in [500\text{ms}, 620\text{ms}]$ and $t \in [1600\text{ms}, 6000\text{ms}]$, the sensing performance of SACSMN remains stable, leading to a steady reduction in the communication rate. However, when $M = N = 16$ and $t = 420\text{ms}$, the communication rate decreases dramatically due to the failure to achieve high-precision beam alignment.

It is observed that the beam alignment performance based on AlexNet exhibits a gradual improvement over time. Specifically, when $M = N = 16$ and $t \in [0\text{ms}, 2600\text{ms}]$, the sensing error experiences a sharp decline, and the consequent communication gain offsets the channel fading caused by

TABLE VIII: average achievable rates of SACSMN at different training sample sizes

(bps)	1%	5%	9%	100%
$M = N = 16$	9.397E-02	9.397E-02	9.398E-02	9.398E-02
$M = N = 64$	3.340E-01	3.343E-01	3.343E-01	3.344E-01

longer distances, resulting in an increase in communication rate. When $M = N = 16$ and $t \in [3000\text{ms}, 6000\text{ms}]$, the improvement in beam alignment performance gradually levels off and the effect of distance dominates the achievable downlink rate, resulting in rate degradation. When $M = N = 16$, despite an overall decrease in sensing error, it descends slowly during certain periods, coupled with the impact of distance, resulting in more fluctuations in the achievable rate.

The point of intersection between the 16- and 64-antenna curves at $t \in [4200\text{ms}, 4600\text{ms}]$ is noteworthy for AlexNet. This is because, during this time interval, both scenarios achieve approximate sensing performance gains, resulting in similar improvements in beam alignment performance, while the larger array gain significantly amplifies the contribution of the beam alignment gain to the achievable rate. Notably, the sensing error with $M = N = 64$ for AlexNet is significantly smaller during $t \in [5850\text{ms}, 6000\text{ms}]$, leading to an even greater improvement in the rate. The intersection of the AlexNet with $M = N = 64$ and SACSMN with $M = N = 16$ curves is primarily caused by the differences between communication array gains.

3) The Influence of Training Sample Size on Single Vehicle Achievable Rate for SACSMN: The influence of training sample size on achievable communication rate of a single vehicle based on SACSMN is investigated in this part. The vehicle is placed at [30m, 20m] with $\theta_0 = 33.69°$, and two scenarios, i. e., $M = N = 16$ and $M = N = 64$ are considered. The instantaneous rates are depicted in Fig. 11, and the corresponding average achievable rates at various sizes of training sets are summarized in Table VIII.

As shown in Fig. 11, the communication rate increases with the size of the training set when the time and RSU configuration parameters N are kept constant. This is due to the fact that larger training sets enable the system to achieve smaller sensing errors, which, according to (19), will further enhance the communication rate. Additionally, there are many fluctuations in the rate curve when $M = N = 16$ and only 1% of the data set is utilized for training. This is attributed to the poor learning performance of SACSMN on the non-linear relationship between the echo signals and sensing results, which reduces the beam alignment precision. Furthermore, it is observed from Table VIII that the communication performances achieved based on 1% and 100% of the training set are remarkably similar. This phenomenon indicates that the improvement in communication rate becomes negligible by further increasing the size of the training sample size. Therefore, the above experiment further validates the high efficiency of the proposed SACSMN on communication performance under few-shot learning conditions.

4) The Influence of Initial Vehicle Position and Direction of vehicle travel on the communication performance: In this part, the influences of the initial vehicle position and direction of vehicle travel on the communication performance is explored. The simulation settings in this part is the same as in Table VI. In each scenario, the vehicle drives towards the RSU.

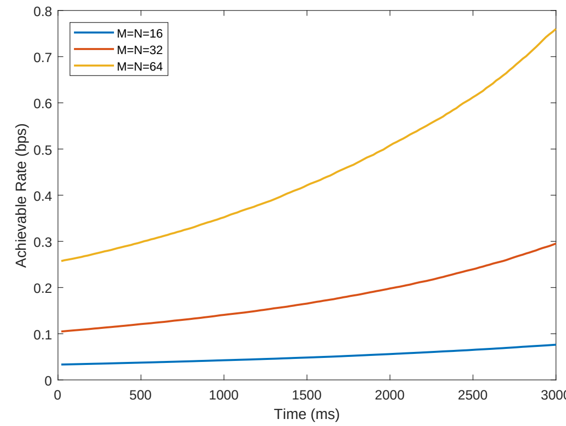


Fig. 12: Achievable communication rate based on the SACSMN trained from 9 % of the training set.

In Fig. 12, we show the achievable rates based on the SACSMN trained from 9% of the training set, where three sizes of ULA are considered. It is observed that as the vehicle approaches the RSU, the communication rate is gradually improved due to the weakened effect of channel fading. Benefiting from the larger communication array gain due to the large-scale ULA, the system achieves the highest communication rate when $M = N = 64$. In the considered scenario, the communication performance is mainly affected by the SACSMN sensing performance and the distance between the vehicle and the RSU. With the benefit of the stable sensing performance achieved by the SACSMN, the rates change gently without drastic fluctuations. According to the above results, it is observed that the initial position and driving direction of the vehicle do not directly affect the communication rate, which influence the communication performance by changing the distance between the vehicle and the RSU.

VII. Conclusion

In this paper, we considered the ISAC-assisted V2I system and investigated the tradeoff problem between system sensing capability and computational complexity as well as the performance of the system in terms of sensing and communication under the condition of few samples. Aiming to reduce the dependence of DNN models on training samples, dynamic SA was proposed to enable the model to extract more general and effective features from few echo signals, improving the feature extraction capability of the model. To achieve the optimal trade-off between system sensing performance and computational complexity, we proposed the groundbreaking network model, SACSMN. In particular, to improve the feature extraction capability, dynamic SA and channel shuffle operation were adopted, and DWConv operation was applied to reduce the computational cost in SACSMN. The simulation results demonstrated that, SACSMN achieves satisfactory sensing performance when based on 9% of the training set samples, and the dependence of the proposed model on training samples is significantly reduced. Compared with the benchmarks, SACSMN achieves the same level of sensing performance with lower computational complexity, realizing the optimal trade-off between system sensing performance and computational cost. With the robust sensing performance of SACSMN in the case of few samples, the system achieves communication performance similar to that of the full training set when based on 1% of the training set samples.

REFERENCES

[1] 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.

[2] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[3] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Radar-communications convergence: Coexistence, cooperation, and co-design," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 1, pp. 1–12, Mar. 2017.

[4] X. Liu, T. Huang, and Y. Liu, "Transmit design for joint mimo radar and multiuser communications with transmit covariance constraint," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1932–1950, Jun. 2022.

[5] Z. Ye, Z. Zhou, P. Fan, Z. Liu, X. Lei, and X. Tang, "Low ambiguity zone: Theoretical bounds and doppler-resilient sequence design in integrated sensing and communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1809–1822, Jun. 2022.

[6] G. N. Saddik, R. S. Singh, and E. R. Brown, "Ultra-wideband multifunctional communications/radar system," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 7, pp. 1431–1437, Jul. 2007.

[7] S. Sen and A. Nehorai, "Adaptive ofdm radar for target detection in multipath scenarios," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 78–90, Jan. 2011.

[8] C. Shi, F. Wang, M. Sellathurai, J. Zhou, and S. Salous, "Power minimization-based robust ofdm radar waveform design for radar and communication systems in coexistence," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1316–1330, Mar. 2018.

[9] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, "Ieee 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018.

[10] S. H. Dokhanchi, B. S. Mysore, K. V. Mishra, and B. Ottersten, "A mmwave automotive joint radar-communications system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 3, pp. 1241–1260, Jun. 2019.

[11] Y.-R. Tsai and W.-H. Chen, "High efficiency beam alignment based on multi-modal beam patterns for massive mimo antenna systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 035–13 046, Dec. 2021.

[12] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[13] 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.

[14] 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.

[15] J. Mu, Y. Gong, F. Zhang, Y. Cui, F. Zheng, and X. Jing, "Integrated sensing and communication-enabled predictive beamforming with deep learning in vehicular networks," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3301–3304, Jul. 2021.

[16] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.

[17] S. Rahman, S. Khan, and F. Porikli, "A unified approach for conventional zero-shot, generalized zero-shot, and few-shot learning," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5652–5667, Nov. 2018.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[19] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[20] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, Sep. 2018, pp. 116–131.

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

[22] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, no. 2, Jun. 2012.

[23] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4580–4584.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[25] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

Wei Liang received her M.Sc. and Ph.D. degree in wireless communication at University of Southampton, Southampton, U.K in 2010 and 2015, respectively. She was a research fellow with Lancaster University, UK, during 2015-2018. She is currently an associate professor in the School of Electronics and Information, Northwestern Polytechnical University. Her research interests include Integrated sensing and communication, cognitive radio network, Non-orthogonal multiple access scheme and Mobile edge computing etc.

Yujie Wang received the MSc. degree in wirelsss communications from the Northwestern Polytechnical University in 2024. His research direction is integrated sensing and communication.

Jiankang Zhang received the B.Sc. degree in mathematics and applied mathematics from Beijing University of Posts and Telecommunications, Beijing, China, in 2006, and the Ph.D. degree in communication and information systems from Zhengzhou University, Zhengzhou, China, in 2012. He is a Senior Lecturer with Bournemouth University, Poole, U.K. Prior to joining in Bournemouth University, he was a Senior Research Fellow with the University of Southampton, Southampton, U.K. He was a Lecturer from 2012 to 2013 and then an Associate Professor from 2013 to 2014 with Zhengzhou University, Zhengzhou, China. His research interests are in the areas of aeronautical communications and networks, evolutionary algorithms, machine learning algorithms, and edge computing. Dr. Zhang serves as an Associate Editor for IEEE ACCESS.

Lixin Li received the B.S. (Hons.), M.S. (Hons.) and the Ph.D. degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2008, respectively. From 2009 to 2011, he was a Postdoctoral Research Fellow with NPU. In 2011, he joined the School of Electronics and Information, NPU, where he is currently a Full Professor and Chair of Department of Communication Engineering with the School. His research interests include 5G/6G wireless networks, federated learning, game theory, and machine learning for wireless communications. He was the recipient of the 2016 NPU Outstanding Young Teacher Award, which is the highest research and education honors for young faculties with NPU. He was an exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2020.

Han Zhu received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively. From 2000 to 2002, he was a R&D Engineer of JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as the Computer Science Department, University of Houston, Houston, TX, USA. His research interests include the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources, wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security and privacy. Dr. Han was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018 and ACM Distinguished Speaker from 2022 to 2025. He has been an AAAS Fellow since 2019 and ACM distinguished Member since 2019. Dr. Han has been a 1% highly cited Researcher since 2017 according to Web of Science.