Lightweight Deep Learning-Based Receiver Design for Coded OTFS Modulation in Vehicular Networks

Haohai Huang*, Jianzhe Xue*, Jinshan Su†, Jiaxin Li*, Tingting Zhang‡, and Haibo Zhou*
*School of Electronic Science and Engineering, Nanjing University, Nanjing, China, 210023.
†Key Laboratory of Vibration Signal Capture and Intelligent Processing, Yili Normal University, Yining, China, 835000.
‡College of Mechanical and Automotive Engineering, Chuzhou Polytechnic, Chuzhou, China, 239000.
Email: {haohaihuang, jianzhexue}@smail.nju.edu.cn, sqsjs1968@aliyun.com, jiaxinli@smail.nju.edu.cn, zhangtingting406@chzc.edu.cn, and haibozhou@nju.edu.cn.

2024 IEEE/CIC International Conference on Communications in China (ICCC) | 979-8-3503-7841-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCC62479.2024.10681737

***Abstract*—Vehicular networks have strong requirements for reliable communication in high-mobility environments. Orthogonal time frequency space (OTFS) modulation addresses the challenges of vehicular double-selective channels by mapping information symbols into a two-dimensional delay-Doppler (DD) domain, allowing for enhanced utilization of channel diversity across time and frequency. Traditionally, signal detection in OTFS receivers usually requires accurate channel state information (CSI) which is difficult to obtain. In this paper, we design an efficient intelligent receiver for coded OTFS in vehicular networks based on deep learning (DL), utilizing its power ability in two-dimensional data processing. Specifically, without perfect CSI, we use neural networks to directly extract features from signal data in the DD domain for OTFS signal detection. Moreover, in order to solve the problem of the limited reception field of convolution modules, we design a shift padding (SP) method to enhance data at the edge of DD domain data frames. Extensive simulation results validate the feasibility of the SP method and the effectiveness of the proposed intelligent receiver of coded OTFS, whose performance is considerably ahead of classic signal detection methods in low signal-to-noise ratios (SNR).**

***Index Terms*—OTFS, receiver, deep learning, shift padding.**

I. Introduction

As sixth-generation (6G) networks continue to develop, the demand for ubiquitous mobile terminal connections poses huge challenges for the wireless communication networks from vehicular networks [1]. Due to high-speed movement, the channel demonstrates characteristics of double-selective fading, which include time-selective fading induced by Doppler shift and frequency-selective fading caused by multipath effects [2]. However, the conventional orthogonal frequency division multiplexing (OFDM) modulation in the time-frequency (TF) domain suffers from severe inter-carrier interference (ICI) when applied to the double-selective channel.

Different from the OFDM modulation, the orthogonal time frequency space (OTFS) modulation is a novel two-dimensional (2D) modulation scheme that maps data symbols into delay-Doppler (DD) domain, which is emerging as a promising solution for vehicular networks communication. Each symbol in the DD domain can be thoroughly expanded in the TF domain to explore the channel diversity, ensuring that the channel has a nearly uniform influence on each transmitted symbol [3]. In other words, OTFS converts the TF domain channel, which experiences both frequency-selective fading and time-selective fading, into the DD domain channel, where the channel characteristics remain relatively stable. In the DD domain, the channel effect is a 2D convolution of the data symbols and the channel response. This transformation aids in overcoming the inherent limitations of the TF domain processing framework used in OFDM [4]. Hence, significant research efforts have been devoted to developing signal detection algorithms for OTFS [5], [6]. However, all of these signal processing methods depend on having explicit and accurate channel state information (CSI), which is difficult to estimate due to pilot overhead and dynamic communication environment. In the presence of channel estimation errors, the signal detection results of the traditional algorithm will have a large deviation from the original signal [7].

Recently, the application of deep learning (DL) techniques in the OTFS has garnered increasing attention, which is widely employed for optimizing different modules of OTFS receivers [8], [9]. DL can offer a refined understanding of the multifaceted relationships between transmitted signals and received observations, thus enabling more sophisticated and effective signal processing techniques [10]. The vehicular networks channel conditions and interference patterns are often highly dynamic and complex, which have posed substantial challenges for signal detection through conventional techniques. Fortunately, the DD domain signals can be easily converted into these data formats that are compatible with various DL models. For instance, the signals in the DD domain are inherently 2D and can be interpreted as grid data, allowing the utilization of convolutional neural network (CNN). Conventional OTFS receivers require perfect CSI for signal recovery, whereas the DL-based intelligent receiver can autonomously extract features from extensive datasets to recover information without dependence on perfect CSI, which shows better performance than traditional receivers.

In this paper, we design a DL-based receiver for coded OTFS modulation in vehicular networks. Specifically, the intelligent receiver uses the CNN model and other lightweight neural network models, including MobileNet and MobileViT. However, when dealing with 2D symbols in the DD domain, convolution windows cannot handle some edge symbols whose information is located on the other pair of edges due to their limited reception field. To address the issue, we designed a

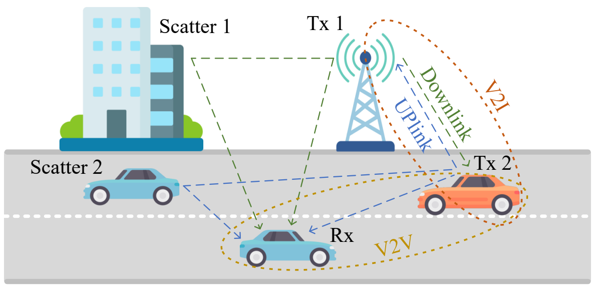


Fig. 1: OTFS application in vehicular networks.

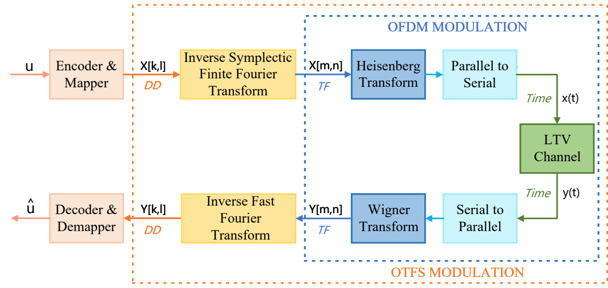


Fig. 2: Architecture of OTFS modulation.

padding method to process symbol data in the DD domain, which is called shift padding (SP). This paper makes three primary contributions, outlined as the following:

- We leverage neural networks to directly extract features from signal data in the DD domain, which can perform better signal recovery than traditional algorithms.
- We propose a novel padding method called SP for neural networks to address the limitation where the convolutional window cannot handle some edge symbols whose information is located on the other pair of edges due to their limited reception field.
- We carry out extensive simulation experiments in the vehicular networks environment, and verify the advantages of the proposed intelligent receiver and the effectiveness of the SP method.

The rest of this paper is organized as follows. Section II elaborates the OTFS communication system model. Section III describes SP method and the proposed DL-based receiver. Simulation results are presented in Section IV, followed by the conclusion of this work in Section V.

II. System Model

The OTFS application in vehicular networks can be shown in Fig. 1, in which each vehicle travels at a certain speed. In vehicular networks, vehicles communicate with other vehicles and roadway infrastructures. When the vehicle is on the move, the communication environment of the vehicle networks will become dynamic and complex, which will introduce double-selective fading. In the high-speed moving environment, the parameters of the communication channel are uncertain and time-varying. Meantime, the Doppler shift which can be regarded as a fixed value in a short time will destroy the orthogonality between subcarriers and cause OFDM to suffer from severe ICI. Symbol modulation techniques in the TF domain no longer suffice to meet the demands of stable and reliable communication under mobility. Therefore, OTFS modulation has emerged as an efficient solution for vehicular network communication.

Fig. 2 shows the specific architecture of of OTFS modulation, which can be implemented by using straightforward pre-processing and post-processing steps applied to traditional multicarrier modulation. The internal box is similar to the OFDM modulation and the external box implements the OTFS modulation with pre-processor and a post-processor blocks. OTFS converts the modulation domain on the basis of OFDM, and converts the one-dimensional sequence signal to DD domain, TF domain and time domain successively for signal transmission, which can effectively resist the fading caused by Doppler shift.

Denoting M as the number of delay bins/number of subcarriers, N as the number of Doppler bins/number of time slots, T as the time slot duration, and $\triangle f$ as the subcarrier spacing. At the transmitter, the input bit stream u is encoded by an encoder and subsequently modulated into 2D symbols $\mathbf{X_{DD}}$ through a mapper, representing the symbols in the DD domain. With the use of 2D inverse symplectic finite Fourier transform (ISFFT), the transition from $\mathbf{X_{DD}} \in \mathbb{C}^{M\times N}$ in the DD domain to the TF domain symbols $\mathbf{X_{TF}} \in \mathbb{C}^{M\times N}$ can be derived as,

$$\mathbf{X_{TF}}[m,n] = \frac{1}{\sqrt{\mathrm{MN}}} \sum_{l=0}^{N-1} \sum_{k=0}^{M-1} \mathbf{X_{DD}}[k,l] e^{j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}. \quad (1)$$

After adding the transmit window function $W_{tx}[m,n]$, Eq. (1) can be expressed as,

$$\mathbf{X_{TF}}[m,n] = \frac{W_{tx}[m,n]}{\sqrt{\mathrm{MN}}} \sum_{l=0}^{N-1} \sum_{k=0}^{M-1} \mathbf{X_{DD}}[k,l] e^{j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}. \quad (2)$$

Next, the Heisenberg transform with the transmit waveform $g_{tx}(t)$ is then applied to $\mathbf{X_{TF}}$ to obtain the discrete time domain signal $x(t)$. Specifically, the Heisenberg transform corresponds to the inverse fast Fourier transform (IFFT) that is implemented in the OFDM modulation. Hence, the signal $x(t)$ in the time domain can be formulated as,

$$x(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \mathbf{X_{TF}}[m,n] g_{\mathrm{tx}}(t-nT) e^{j2\pi m\Delta f(t-nT)}, \quad (3)$$

where $g_{\mathrm{tx}}(t-nT)$ denotes the receive pulses.

Moreover, for a linear time-varying (LTV) channel comprising P resolvable paths, the channel response $\mathbf{H_{DD}}(\tau,\nu)$

in the DD domain characterizes an impulse with delay τ and Doppler ν , which can be expressed as,

$$\mathbf{H_{DD}}(\tau,\nu)=\sum_{p=1}^{P}h_p\delta\left(\tau-\tau_p\right)\delta\left(\nu-\nu_p\right), \tag{4}$$

where h_p, τ_p and ν_p are the complex fading coefficient, time delay, and Doppler frequency associated with the p-th resolvable path, respectively, whose values are determined by the selected channel model.

The signal $x(t)$ is transmitted over a double-selective fading channel with the complex DD domain channel response $\mathbf{H_{DD}}(\tau,\nu)$. Thereby, the received time domain signal $y(t)$ can be formulated as,

$$y(t)=\iint \mathbf{H_{DD}}(\tau,\nu)x(t-\tau)e^{j2\pi\nu(t-\tau)}d\tau d\nu+\omega(t), \tag{5}$$

where $\omega(t)$ is the additive white Gaussian noise with mean 0 and variance σ^2. The received time domain signal $y(t)$ is firstly transformed to the TF domain through the Wigner transform with the matched windowing $g_{rx}(t)$, and the cross-ambiguity function $\mathbf{Y}(t,f)$ between $g_{rx}(t)$ and $y(t)$ can be formulated as,

$$\mathbf{Y}(t,f)=\int g_{\rm rx}^{*}(t^{'}-t)y(t^{'})e^{-j2\pi f(t-t^{'})}dt^{'}. \tag{6}$$

Then, the matched filter output is obtained by adding the received window function $W_{rx}[m,n]$ and sampling $\mathbf{Y}(t,f)$ at intervals $t=nT$, $f=m\Delta f$ as,

$$\mathbf{Y_{TF}}[m,n]=W_{rx}[m,n]\mathbf{Y}(t,f)\mid_{t=nT,f=m\Delta f}. \tag{7}$$

In the TF domain, the input-output relation between $\mathbf{X_{TF}}$ and $\mathbf{Y_{TF}}$ by adding terms $\mathbf{H_{m,n}}[m^{'},n^{'}]$ can be derived as,

$$\mathbf{Y_{TF}}[m,n]=\sum_{n^{'}=0}^{N-1}\sum_{m^{'}=0}^{M-1}\mathbf{H_{m,n}}[m^{'},n^{'}]\mathbf{X_{TF}}[m^{'},n^{'}], \tag{8}$$

where $\mathbf{H_{m,n}}[m^{'},n^{'}]$ is the combined effects of $g_{tx}(t)$, $\mathbf{H_{DD}}(\tau,\nu)$, and $g_{rx}(t)$. Let $C_{g_{rx},g_{tx}}$ denote the cross-ambiguity function between $g_{tx}(t)$ and $g_{rx}(t)$ as,

$$C_{g_{rx},g_{tx}}(t,f)\triangleq\int g_{\rm rx}^{*}(t^{'}-t)g_{tx}(t^{'})e^{-j2\pi f(t-t^{'})}dt^{'}. \tag{9}$$

Thereby, $\mathbf{H_{m,n}}[m^{'},n^{'}]$ can be formulated as,

$$\mathbf{H_{m,n}}[m^{'},n^{'}]=\iint \mathbf{H_{DD}}(\tau,\nu)C_{g_{rx},g_{tx}}((n-n^{'})T-\tau,$$
$$(m-m^{'})\Delta f-\nu)e^{j2\pi(\nu+m^{'}\Delta f)((n-n^{'})T-\tau)}e^{j2\pi\nu n^{'}T}d\tau d\nu. \tag{10}$$

Correspondingly, the Wigner transform is the fast Fourier transform (FFT) that is implemented by the conventional OFDM demodulation. Then, the TF domain symbols $\mathbf{Y_{TF}}\in\mathbb{C}^{M\times N}$ can be formulated as,

$$\mathbf{Y_{TF}}[m,n]=W_{rx}[m,n]\int g_{\rm rx}^{*}(t-nT)y(t)e^{-j2\pi m\Delta f(t-nT)}dt. \tag{11}$$

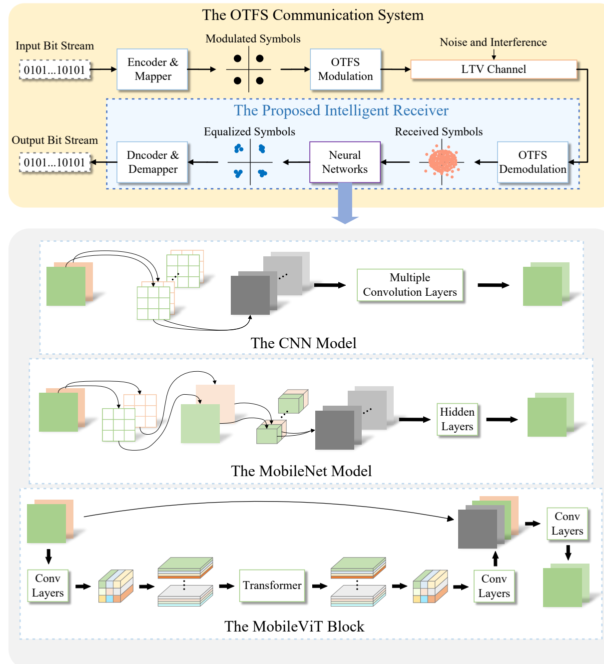


Fig. 3: Architecture of the proposed DL-based receiver.

Then, $\mathbf{Y_{TF}}$ is transformed to DD domain using the 2D symplectic finite Fourier transform (SFFT) formulated as,

$$\mathbf{Y_{DD}}[k,l]=\frac{1}{\sqrt{NM}}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\mathbf{Y_{TF}}[m,n]e^{-j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}. \tag{12}$$

The data in the DD domain symbols $\mathbf{Y_{DD}}$ is distributed along the delay axis and Doppler axis, representing the modulated constellation points. Finally, the data symbols $\mathbf{Y_{DD}}$ in the DD domain at the receiver are de-mapped to the one-dimensional sequence, and decoded by a decoder to obtain the output bit stream $\hat{u}$.

III. Intelligent Receiver for OTFS

The structure of our DL-based receiver for coded OTFS in vehicular networks is shown in Fig. 3, which achieves dependable data restoration from distorted received symbols to equalized symbols and then to the original information bit stream. The neural networks in our receiver, aim to recover chaotic received symbols to equalized symbols, which is similar to original modulated symbols.

A. Shift Padding Method

Neural networks are used to extract data features through filters including the convolutional window, which has a limited receptive field and can only aggregate local information. When applying neural networks at the receiver to process OTFS frames, this feature leads to the inability of the convolutional window to efficiently process edge symbols whose information probably lies on the opposite edge far across the entire frame. For instance, in Fig. 4, information in Region 6 is diffused into Region 4 due to the Doppler shift, resulting in insufficient

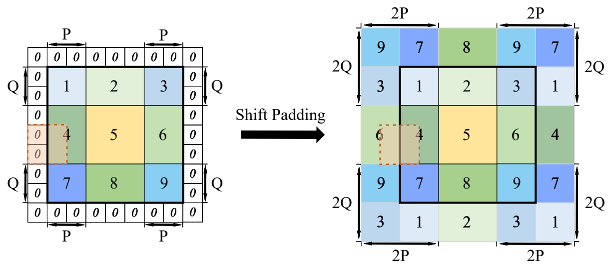


Fig. 4: The SP method.

size of the convolution window to capture this changing relationship. To this end, we design the SP method to solve the conflict of the limited reception field of neural networls and the signal characteristics of OTFS.

As shown in Fig. 4, the proposed SP method firstly cuts the received OTFS frame into 9 pieces, and then copies 8 of them and pads the original frame perimeter according to the characteristics of the 2D circular convolution. In the extension of SP, Q represents the maximum time delay tap and P represents the maximum Doppler frequency tap. The extension length of SP is the length extending the maximum time delay tap at both ends of the delay axis and the length extending the maximum Doppler frequency tap at both ends of the Doppler axis, which ensure the information affected by the channel can be included in the extended range. Denoting x_d as the initial input of neural networks, y_d as the label of neural networks, $\tilde{y}_d$ as the output of neural networks, and $E\left(\cdot;(Q,P)\right)$ as the SP function with extended lengths Q and P. Then, the input data x_{sp} after adding SP can be formulated as,

$$x_{sp} = E\left(x_d;(Q,P)\right), \tag{13}$$

where x_{sp} acts as the final input of the neural network. Denoting $\tilde{E}\left(\cdot;(Q,P)\right)$ as the inverse process of the SP method, which means that after neural networks get the output $\tilde{y}_d$, the SP extension is removed to get the final output. The final output $\hat{y}_d$ can be formulated as,

$$\hat{y_d} = \tilde{E}\left(\tilde{y}_d;(Q,P)\right). \tag{14}$$

Through the SP method, the information composing edge symbols is moved to their neighboring locations, where this information can be well aggregated by neural networks. In this way, when neural networks use the convolutional window for feature extraction, the information that would be lost at the edge will be captured, which can significantly improve the performance of channel equalization.

B. Intelligent Receiver Framework

In the case of input data, the DL model typically processes three-dimensional images, and the third dimension is usually the number of channels 2 or 3, representing black-and-white images and color images respectively. However, different from traditional image data, the data in the DD domain symbols $\mathbf{Y_{DD}} \in \mathbb{C}^{M\times N}$ consists entirely of complex numbers and the number of channels of $\mathbf{Y_{DD}}$ is 1. To match the input data format of neural networks, the complex numbers in received symbols can be split into their real and imaginary parts. Therefore, denoting $\mathbf{r}$ as the 2D received symbols, $Re(\mathbf{r})$ as the real part and $Im(\mathbf{r})$ as the imaginary part of $\mathbf{r}$, the input data x_d can be formulated as,

$$x_d = \left[Re(\mathbf{r}); Im(\mathbf{r})\right], \tag{15}$$

where $x_d \in \mathbb{C}^{M\times N\times 2}$.

The training dataset of neural networks consists of received symbols at the receiver and modulated symbols at the transmitter. Moreover, the received symbols x_d and the modulated symbols y_d serve as the input and labels of neural networks, respectively. The dataset Ω with N_B samples can be expressed as,

$$\Omega = \left\{\left(x_d^{(i)}, y_d^{(i)}\right)\right\}_{i=1}^{N_B}, \tag{16}$$

where i is the index value of the set, and $i = 1,\ldots,N_B$.

After feeding the training dataset into the proposed intelligent DL-based receiver, the convolutional layers will extract features from input data using their kernels. In other word, neural networks aim to learn the function $\mathscr{F}: x_{sp} \to y_d$ with the trainable weights θ, and $\tilde{y}_d = \mathscr{F}\left(x_{sp};\theta\right)$. So the output data $\hat{y_d}$, which is the equalized symbols, can be formulated as,

$$\hat{y_d} = \tilde{E}\left(\mathscr{F}\left(x_d;\theta\right);(Q,P)\right). \tag{17}$$

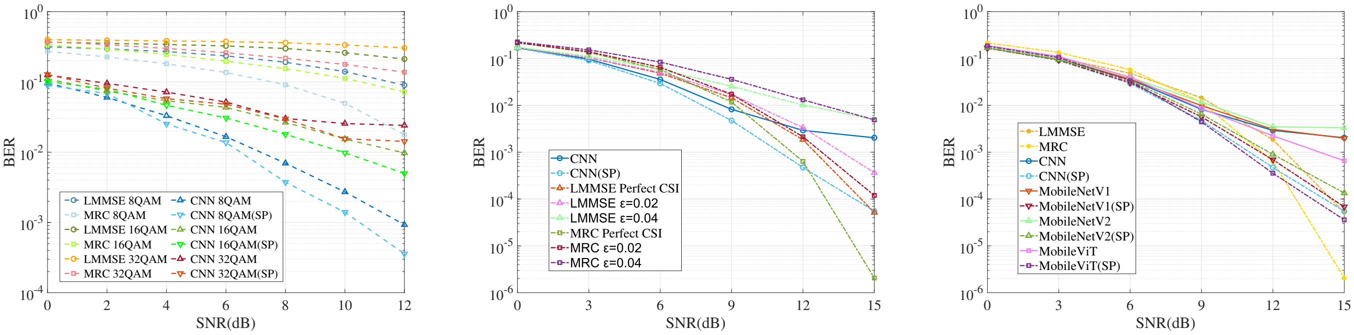
Bit error rate (BER) between the output bit stream and the input bit stream is a primary indicator that reflects the performance of the intelligent receiver. Under the same conditions, the lower the bit error rate, the better the performance of the receiver. Therefore, the goal of the intelligent receiver is to minimize the error between y_d and $\hat{y_d}$, which can minimize BER as much as possible. The loss function L mean square error (MSE) can be expressed as,

$$L = MSE(y_d,\hat{y_d}) = \frac{1}{N_D}\sum_{d=1}^{N_D}\left(y_d - \hat{y_d}\right)^2, \tag{18}$$

where N_D represents the data length and the goal is to minimize L.

In order to adapt to the high-speed vehicular networks, it is of great significance to reduce computational costs of neural networks. Lightweight neural networks employ a flexible and unique structural design that maintains accuracy while enabling easier deployment on intelligent receivers requiring mobility and flexibility [11].

Besides the traditional neural network such as CNN, we use MobileNetV1 presented in [12], which proposes depthwise separable convolution to make the network lightweight and ensure accuracy, and MobileNetV2 presented in [13], which proposes inverted residual with the linear bottleneck to improve performance of information recover. Furthermore, we use MobileViT presented in [14], which combines the characteristics of CNN's spatial inductive bias and correlation between local pixels with the transformer's self-attention function to capture global information.



(a) BER performance of different modulation. (b) BER performance with imperfect CSI. (c) BER performance without channel coding.

Fig. 5: BER performance evaluation.

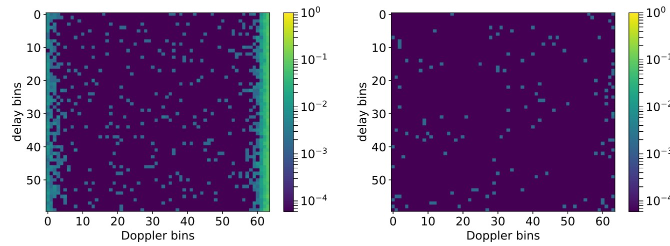
IV. Simulation and Results

A. Simulation Parameter Setting

The EVA channel model is considered for vehicular networks, where the vehicle speed is 150 kmph and the channel stays constant within a data frame. Some null symbols in the DD domain are placed as interleaved zero-padding (ZP) guard bands in the time domain [15]. The channel coding adopts (10,9) Reed-Solomon (RS) code and Bose-Chaudhuri-Hocquenghem (BCH) code. The signal detection algorithm adopts linear minimum mean square error (LMMSE) and maximal ratio combining (MRC). At each signal-to-noise ratio (SNR), the training set consists of 70,00 samples, while the validation set contains 3,000 samples. Other simulation parameters are detailed in Table I.

TABLE I: Simulation Parameters.

Parameter	Value
Number of time slots N	64
Number of subcarriers M	64
ZP length	4
Carrier frequency	5.9 GHz
Subcarrier spacing (Δf)	15 kHz
Initial learning rate	0.001
Optimizer	Adam
Mini-batch size	64
Loss function	MSE



(a) Average BER without SP. (b) Average BER with SP.

Fig. 6: The average BER of DD domain symbols.

B. Results and Evaluation

In this section, we evaluate the DL-based receiver performance via empirical results with the corresponding BER performance. We demonstrate the effectiveness of the proposed DL-based receiver in the EVA channel of vehicular networks, verify the feasibility of the SP method, and finally test the effectiveness of the DL-based receiver for coded OTFS.

The BER performance evaluation of the proposed DL-based receiver in 4-quadrature amplitude modulation (QAM) mode is shown in Fig. 5. Fig. 5a illustrated the BER performance evaluation of the conventional OTFS signal detection algorithms and DL-based intelligent receiver with adopting RS code. As the modulation order increases from 8 to 16 and then to 32, the BER performance of the traditional signal detection algorithms significantly decreases, while the proposed DL-based receiver maintains relatively stable reception performance in high-order modulation modes. Fig. 5b shows that as the error coefficient ε increases, the BER performance of traditional receivers degrades significantly. Different from traditional signal detection algorithms with perfect CSI, the proposed DL-based intelligent receiver can achieve reliable information recovery even without channel estimation. Fig. 5c demonstrates that the SP-aided neural networks perform better than neural networks without SP, which proves that our SP method has significantly improved the effect of intelligent receivers. Moreover, the SP-aided neural networks perform best when SNR$\leq$12dB, outperforming traditional signal detection algorithms with perfect CSI. However, when SNR$\geq$12dB, because the traditional algorithm has perfect CSI, it can perform excellent signal recovery under relatively high SNR, which exceeds the signal detection method based on deep learning.

When SNR is 15dB, the average BER of DD domain symbols is shown in Fig. 6. Fig. 6a demonstrates that in the whole DD domain, the average BER is significantly lower in the central portion of data bits, whereas it is higher at the edges since the information in the edge part is not well extracted by the convolutional window. On the contrary, Fig. 6b demonstrates that because of the SP expansion of the data, the convolutional window can extract the information lost at the edge to decrease the average BER in the whole DD domain, which reaches the order of 10^{-4}. By comparing

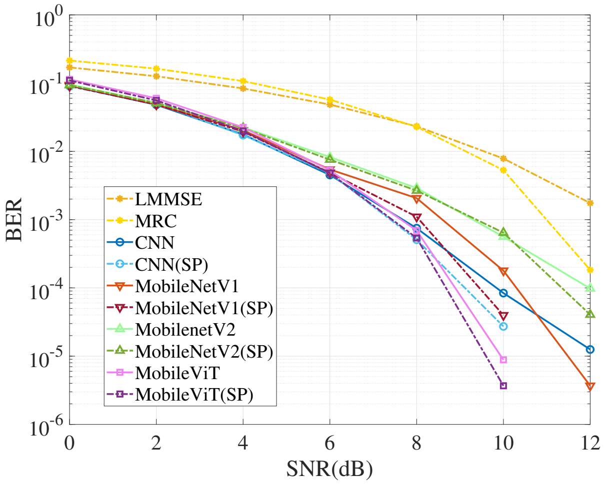


Fig. 7: BER performance with RS code.

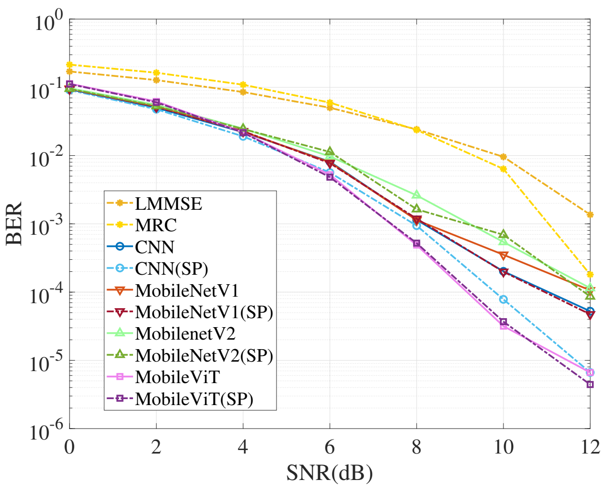


Fig. 8: BER performance with BCH code.

these two figures, the effectiveness of the SP method proposed in this paper in processing edge symbols and improving the overall performance is well demonstrated.

Fig. 7 and Fig. 8 illustrate the BER performance of conventional OTFS signal detection algorithms and the DL-based receiver proposed in this paper in 4-QAM mode with adopting RS and BCH code, respectively. When the channel coding in vehicular networks EVA channel reflects a certain stability, the DL-based receiver achieves better BER performance than conventional OTFS receivers. Specifically, in Fig. 7, the BER performance of MobileViT, SP-aided MobileNetV1, and CNN can reach 0 in SNR 12dB. Meantime, when SNR is 12dB, the BER of the DL-based receiver can reach 10^{-6} with adopting BCH code in Fig8. In the RS and BCH coded OTFS system in the EVA channel, the DL-based receiver consistently demonstrates superior performance compared to conventional signal detection algorithms. Moreover, the SP-aided intelligent receiver also shows better performance than the intelligent receiver without the SP method, which proves the effectiveness of the SP method. Due to its complete reliance on convolutional windows, the effectiveness of the SP method is most evident on CNN.

V. Conclusions

In this paper, we have proposed a DL-based receiver with lightweight neural networks and SP method for vehicular networks with OTFS modulation. Extensive simulation results have demonstrated the effectiveness of the proposed DL-based receiver, which provides an efficient solution for achieving reliable communication in high-mobility vehicular networks.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant 623B2052, 62271244, the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067.

References

[1] J. Xue, K. Yu, T. Zhang, H. Zhou, L. Zhao, and X. Shen, “Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks,” *IEEE Transactions on Mobile Computing*, vol. 23, no. 8, pp. 8143–8157, 2024.

[2] J. Chen, X. Liang, J. Xue, Y. Sun, H. Zhou, and X. Shen, “Evolution of RAN architectures towards 6G: Motivation, development, and enabling technologies,” *IEEE Communications Surveys & Tutorials*, in press.

[3] W. Yuan, S. Li, Z. Wei, Y. Cui, J. Jiang, H. Zhang, and P. Fan, “New delay doppler communication paradigm in 6G era: A survey of orthogonal time frequency space (OTFS),” *China Communications*, vol. 20, no. 6, pp. 1–25, 2023.

[4] L. Gaudio, G. Colavolpe, and G. Caire, “OTFS vs. OFDM in the presence of sparsity: A fair comparison,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4410–4423, 2022.

[5] S. Tiwari, S. S. Das, and V. Rangamgari, “Low complexity LMMSE receiver for OTFS,” *IEEE Communications Letters*, vol. 23, no. 12, pp. 2205–2209, 2019.

[6] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, “Interference cancellation and iterative detection for orthogonal time frequency space modulation,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6501–6515, 2018.

[7] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, “Receiver design for OTFS with a fractionally spaced sampling approach,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4072–4086, 2021.

[8] T. Sun, J. Lv, and T. Zhou, “A transformer-based channel estimation method for OTFS systems,” *Entropy*, vol. 25, no. 10, 2023.

[9] Y. Gong, Q. Li, F. Meng, X. Li, and Z. Xu, “Data-driven deep learning for OTFS detection,” *China Communications*, vol. 20, no. 1, pp. 88–101, 2023.

[10] J. Xue, Y. Xu, W. Wu, T. Zhang, Q. Shen, H. Zhou, and W. Zhuang, “Sparse mobile crowdsensing for cost-effective traffic state estimation with spatiotemporal transformer graph neural network,” *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 16 227–16 242, 2024.

[11] B. Wang, Z. Yuan, S. Zheng, and Y. Liu, “Data-driven intelligent receiver for OTFS communication in internet of vehicles,” *IEEE Transactions on Vehicular Technology*, pp. 1–12, 2023.

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[14] S. Mehta and M. Rastegari, “MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer,” *arXiv preprint arXiv:2110.02178*, 2021.

[15] T. Thaj and E. Viterbo, “Low complexity iterative rake decision feedback equalizer for zero-padded OTFS systems,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 606–15 622, 2020.