OTFS Signal Detection Method by Fusing Multi-Feature and Optimization Strategies

Xinlin Wang, Liejuan Yin, *Member, IAENG*

Abstract—**Orthogonal time frequency space (OTFS) modulation has garnered a lot of interest in the context of high-speed wireless communication systems because of its distinct delay-Doppler signal processing technique. However, existing signal detection approaches suffer from feature coupling in the delay-Doppler domain and high computational complexity. To overcome these challenges, we propose a CNN-Attention-LSTM network optimized via the nutcracker optimization algorithm (NOA), referred to as NOA-CALNet, for efficient signal detection. This network effectively captures the sparse characteristics of the channel along the delay dimension, dynamically weights key channel response regions, enhances the suppression of inter-symbol interference, and decodes time-varying correlations along the Doppler dimension. By leveraging NOA to automatically optimize hyperparameters, the model achieves an optimal configuration for high-precision signal detection. Simulation results show that, compared with traditional methods such as MMSE and other deep learning-based detectors, the proposed model not only accurately extracts signal features and reduces the symbol error rate but also improves computational efficiency and accelerates convergence, demonstrating strong applicability in high-mobility communication environments.**

Index Terms—**deep learning; orthogonal time frequency space; nutcracker optimization algorithm; signal detection**

I. Introduction

WITH the fast development of wireless communication technology, high-mobility scenarios such as high-speed rail networks, intelligent connected vehicles, and aviation communication systems impose increasingly stringent requirements on stable and efficient data transmission [1]. Orthogonal frequency division multiplexing (OFDM) is widely recognized as one of the most commonly used modulation techniques in modern wireless communication networks. By utilizing the inverse fast Fourier transform (IFFT), OFDM decomposes high-rate data streams into multiple mutually orthogonal subcarriers, thereby enabling parallel transmission and effectively reducing inter-symbol interference (ISI). However, Doppler-induced frequency shifts break subcarrier orthogonality in high-mobility situations with fast channel fluctuations, resulting in significant performance deterioration and severe inter-carrier interference (ICI) [2]. Furthermore, the time-varying nature of such channels necessitates frequent channel estimation and computationally intensive equalization processes. These operations not only increase system complexity but also make it challenging to maintain reliable communication quality under dynamic conditions [3].

Orthogonal time frequency space (OTFS) modulation is increasingly regarded as an effective alternative for addressing these challenges. Unlike other modulation schemes that process signals in the time-frequency (TF) domain signal processing, OTFS converts time-varying channels into a quasi-static representation by mapping signals onto the delay-Doppler (DD) domain. This transformation inherently mitigates ISI and ICI while leveraging channel diversity, thereby improving spectral efficiency and communication robustness[4]. Nevertheless, OTFS signal detection remains challenging due to the interplay of multipath effects and non-stationary noise. Traditional methods like zero-forcing (ZF) and minimum mean squared error (MMSE) detection are highly sensitive to noise under complex channel conditions, which restricts their detection accuracy [5]. Recently, message passing (MP)-based detection [6] and unitary approximate message passing (UAMP)-based methods [7] have reduced computational complexity but demonstrate performance degradation in highly scattered environments. Complementary approaches based on variational Bayes (VB) [8] and expectation propagation (EP) [9] offer partial improvements in global convergence and efficiency, yet fail to fully resolve these limitations.

In order to investigate more potent feature-learning capabilities, neural network techniques have been progressively included into OTFS system signal detection due to the quick development of deep learning technology. To enhance the precision of estimating the channel, a dedicated neural network named CENet [10] was proposed. Subsequently, CCRNet, a recovery network conditioned on channel information, was introduced to further boost detection capability. Reference [11] introduced an LSTM-based approach for OFDM detection, effectively enhancing the extraction capability for temporal features. Additionally, a Bayesian-based deep learning detection architecture, termed Bayesian parallel interference cancellation network (BPICNet), was proposed in [12]. This architecture integrates neural networks with parallel interference cancellation strategies, demonstrating strong

Manuscript received March 12, 2025; revised May 11, 2025.

Xinlin Wang is a lecturer at the Software College, Liaoning Technical University, Huludao 125105, China (e-mail: wangxinlin1107@foxmail.com).

Liejuan Yin is a postgraduate student at the Software College, Liaoning Technical University, Huludao 125105, China (Corresponding author to provide e-mail: yinlielie@163.com).

performance in environments with complex interference. In [13], an OTFS signal identification framework utilizing a two-dimensional convolutional neural network (2D-CNN) was proposed. By applying data augmentation strategies to expand the training data volume, the system achieved improved detection accuracy.

Existing signal detection algorithms still face challenges in effectively modeling the intrinsic relationship between signal features and detection objectives, which limits further improvements in detection accuracy and results in high computational complexity. To overcome these limitations, we propose NOA-CALNet, a deep learning-based model for OTFS signal detection, incorporating the following key enhancements. First, a CNN module with multi-scale convolutional kernels is utilized to extract localized structural features from the delay-Doppler grid, enabling effective capture of multipath delay distributions. Second, a spatial attention mechanism adaptively adjusts the weights of time-frequency features, suppressing non-diagonal interference in the channel matrix. Third, an LSTM module is employed to model temporal dependencies and dynamically track Doppler-induced frequency shifts in high-mobility scenarios, which enhances the adaptability of the model to time-varying channels. Finally, NOA-based hyperparameter optimization is introduced to mitigate local minima issues associated with gradient descent, significantly accelerating the training process.

The rest of this paper is structured as follows: Section II outlines the fundamental principles of the OTFS system. Section III presents the proposed NOA-CALNet-based OTFS signal detection method. Section IV discusses the experimental results, and finally, Section V summarizes the findings of this study.

II. System Model

A. OTFS Transmitter

The modulation process of the OTFS system is illustrated in Fig. 1. The transmission procedure follows a structured sequence: first, the bit information undergoes constellation mapping, converting it into modulated symbols. A structured collection of modulated symbols of length $N \times M$ is created by mapping these symbols onto the two-dimensional DD domain. These symbols are then grouped into a 2D matrix $X_{\text{DD}} \in \mathbb{C}^{N\times M}$. Where $x_{\text{DD}}[k,l] \in X_{\text{DD}}, k = 0,\ldots,N-1, l = 0,\ldots, M-1$ represents the transmitted symbol located at the k^{th} Doppler index and the l^{th} delay index. Subsequently, the inverse symplectic finite Fourier transform (ISFFT) is applied to map the information symbols from the DD domain to the TF domain, producing the transformed matrix $\boldsymbol{X}_{\text{TF}}[n,m], n = 0,\ldots N-1, m = 0,\ldots M-1$, represented by the following expression:

$$\boldsymbol{X}_{\text{TF}}[n,m] = \frac{1}{\sqrt{MN}}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1} x_{\text{DD}}[k,l]\exp\left[j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)\right] \quad (1)$$

Next, the Heisenberg transform is applied to project the information symbol $\boldsymbol{X}_{\text{TF}}[n,m]$, defined in the TF domain, into the time-domain signal $s(t)$ [14], where $s(t)$ is given by:

$$s(t) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\boldsymbol{X}_{\text{TF}}[n,m] f(t-nT)\exp\left[j2\pi m\Delta f(t-nT)\right] \quad (2)$$

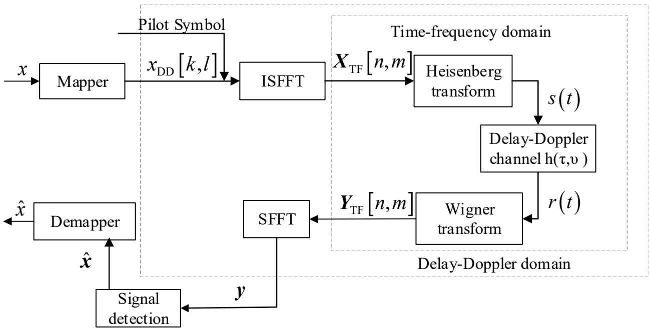

Fig. 1. OTFS system modulation block diagram

where $f(t)$ is the pulse waveform. At this point, the modulation and signal conversion processes at the transmitter are completed, and the modulated information $s(t)$ is obtained.The channel response characterized in the DD domain within OTFS is formulated as [15]:

$$m(\tau,\nu) = \sum_{i=1}^{P} m_i\delta(\tau-\tau_i)\delta(\nu-\nu_i) \quad (3)$$

where P is the number of paths, m_i denotes the gain of the i^{th} path, and $\delta(\bullet)$ is the Dirac function. The corresponding delay and Doppler shift for the i^{th} path are denoted by τ_i and ν_i, respectively, and are defined as:

$$\tau_i = \frac{l_{\tau_i}}{M\Delta f} = \frac{l_{\tau_i}T}{M}, \nu_i = \frac{k_{\nu_i}}{NT} = \frac{k_{\nu_i}\Delta f}{N} \quad (4)$$

where T represents the symbol period, $\Delta f = \frac{1}{T}$, l_{τ_i} and k_{ν_i} denote the delay index and Doppler shift index, respectively, both being integers.

B. OTFS Receiver

After $s(t)$ propagating through the delay-Doppler channel, the received signal $z(t)$ at the receiver is expressed as:

$$z(t) = \iint m(\tau,\nu)s(t-\tau)\exp\left[j2\pi\nu(t-\tau)\right]\mathrm{d}\nu\mathrm{d}\tau + w(t) \quad (5)$$

where $w(t)$ denotes additive Gaussian white noise (AWGN) with $\mathcal{N}(0,\sigma^2)$. Subsequently, the received signal $z(t)$ is processed by applying the Wigner transform, through which the information symbols $\boldsymbol{Y}_{\text{TF}}[n,m]$ in the TF domain can be obtained. This relationship can be expressed as:

$$\begin{aligned}&\boldsymbol{Y}_{\text{TF}}[n,m] = A_{f,z}(\tau,\nu)|_{\tau=nT,\nu=m\Delta f}\\ &A_{f,z}(\tau,\nu) \triangleq \int \overset{*}{f}(t-\tau)z(t)\exp\left[-j2\pi\nu(t-\tau)\right]\mathrm{d}t\end{aligned} \quad (6)$$

where $A_{f,z}(\tau,\nu)$ denotes the cross-ambiguity function and $f(t)$ represents the waveform pulse at the receiver. The information symbol $y_{\text{DD}}[k,l]$ is then obtained by applying the symplectic finite Fourier transform (SFFT) to the signal, which converts it from the TF domain to the DD domain. This may be represented as:

$$y_{\text{DD}}[k,l] = \frac{1}{\sqrt{NM}}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\boldsymbol{Y}_{\text{TF}}[n,m]\exp\left[-j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)\right] \quad (7)$$

In summary, the input-output connection of the OTFS system can be expressed as follows [16]:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \quad (8)$$

where $x_{\text{DD}}[k,l]$ denotes the transmitted symbols arranged into a vector $\boldsymbol{x}$, and $y_{\text{DD}}[k,l]$ represents the received symbols similarly organized into a vector $\boldsymbol{y}$, both having dimensions of $NM \times 1$. $\boldsymbol{H}$ Is the complex 2D channel description matrix, and $\boldsymbol{w}$ indicates the noise vector. Finally, de-mapping is applied to the received symbol $y_{\text{DD}}[k,l]$ to recover the transmitted bit information.

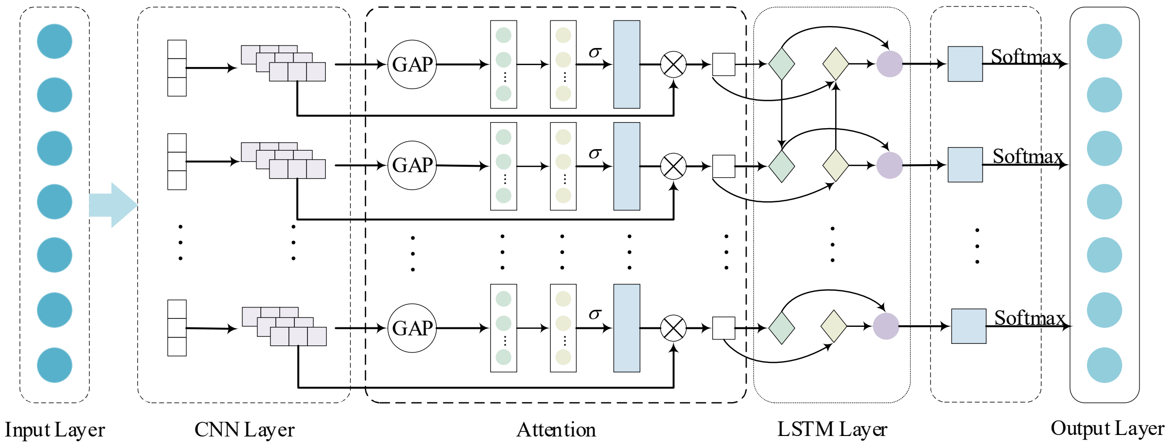

Fig. 2. CALNet model structure

III. The Proposed NOA-CALNet Model

We propose an NOA-CALNet model for OTFS system signal detection. When applying this model for detection, a CNN first extracts local spatial features of the delay-Doppler domain signals. Specifically, convolutional kernels are utilized to capture signal intensity variations and characteristic patterns across different delay-Doppler regions, thus obtaining detailed local feature information. Subsequently, an attention mechanism is introduced, enabling the model to automatically identify and enhance key channel features while effectively suppressing irrelevant interference. This significantly improves feature-learning accuracy and helps the network focus more precisely on regions containing richer information. Next, an LSTM network is employed to model the temporal characteristics of the signals, leveraging its gating mechanisms to capture long-term dependencies and thus extract global correlation information within the time sequences. Finally, the nutcracker optimization algorithm is used to dynamically optimize multiple hyperparameters, such as learning rate and regularization coefficients, achieving the optimal configuration for the model. The architecture of the proposed model is illustrated in Fig. 2.

A. *CNN layer architecture design*

In the OTFS signal detection task, local spatial features are extracted from received signals using CNNs [17], thereby enhancing the representational capability for signals in the DD domain. In this paper, a complex signal decomposition strategy is adopted, where the received signal $\boldsymbol{y}$ of the OTFS system is decomposed into its real and imaginary parts, expressed respectively as:

$$\boldsymbol{y}_{\text{real}} = \Re(\boldsymbol{y}), \boldsymbol{y}_{\text{imag}} = \Im(\boldsymbol{y}) \tag{9}$$

Then the CNN receives the signal as:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_{\text{real}} \\ \boldsymbol{y}_{\text{imag}} \end{bmatrix} \in \mathbb{R}^{N\times C_0} \tag{10}$$

where N denotes the number of time steps in the feature sequence, and the initial number of channels C_0 is set to 1. In this study, two convolutional layers are utilized to extract features from the input signals. The first convolutional layer primarily captures local low-level features, such as channel impulse response characteristics induced by multipath effects, enhancing the fundamental morphological information of the signal to serve as a basis for deeper feature extraction in subsequent layers. The second convolutional layer further extracts more abstract time-frequency features, which enhances the robustness of the model against complex channel fading and noisy environments.

Let $\boldsymbol{Y}^{(l)} \in \mathbb{R}^{N\times C^{(l)}}$ represent the input of the l^{th} layer, $\boldsymbol{W}^{(l)} \in \mathbb{R}^{K\times C^{(l-1)}\times C^{(l)}}$ denote the convolution kernel, K denote the kernel size, $C^{(l-1)}$ represent the number of input channels from the previous layer, $C^{(l)}$ represent the number of output channels in the current layer, and $\boldsymbol{b}^{(l)}$ denote the bias term. The computation performed by the CNN layer can be mathematically described as follows:

$$\boldsymbol{Y}^{(l+1)} = f\left(\boldsymbol{W}^{(l)} * \boldsymbol{Y}^{(l)} + \boldsymbol{b}^{(l)}\right) \tag{11}$$

where $*$ denotes the convolution operation, and $f(\cdot)$ represents the activation function.

B. *Attention mechanism*

A squeeze-and-excitation (SE) attention module is introduced after the CNN to adaptively reweight the extracted channel features. The main process consists of three stages: squeeze, adaptive excitation, and channel recalibration[18]. For the OTFS signal detection task, considering the varying importance of different delay-Doppler regions, global average pooling is applied to each channel to compress its spatial information into a single scalar value. The global representation of the c^{th} channel is defined as:

$$z(c) = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{Y}(i,j,c) \tag{12}$$

where M and N are the spatial dimensions along the delay and Doppler axes, respectively, and the resulting vector z is the global statistical representation used to generate channel attention, capturing the overall response of each feature channel across the OTFS grid. A fully connected network is employed to perform a nonlinear transformation on the aggregated vector, generating activation weights for each channel. To reduce the number of parameters and prevent overfitting, a bottleneck structure is typically adopted. Specifically, the vector z is first projected into a lower-dimensional space of size $\frac{c}{r}$ through a fully connected layer, where r is the reduction ratio (set to 4 in this work), and is subsequently passed through a ReLU activation to obtain $\boldsymbol{s}$. Then, a second fully connected layer is applied to project $\boldsymbol{s}$ back to a c-dimensional vector, and a Sigmoid function is used to normalize the output to the range $[0,1]$, where the resulting vector $\boldsymbol{w}$ is the set of attention weights assigned to each channel.

Finally, the activation weights $\boldsymbol{w}$ are used to scale each channel of the original feature map $\boldsymbol{Y}$, resulting in the reweighted output feature map, where each channel is adaptively emphasized according to its learned importance:

$$\tilde{\boldsymbol{Y}} = \boldsymbol{w} \odot \boldsymbol{Y} \tag{13}$$

This channel-wise multiplication operation enhances the feature channels that are more important within the OTFS grid, thereby improving the ability of the subsequent LSTM to model the correlation in the symbol sequence.

C. LTSM layer model design

Due to multipath interference and channel time variability, there exists a temporal correlation between adjacent symbols. The extracted feature sequences are processed using an LSTM network [19] in order to model the underlying temporal dependency. LSTM utilizes a gating mechanism to maintain long-term states and prevent gradient vanishing, enabling the accurate modeling of temporal relationships across extended sequences. To exploit this capability for modeling inter-symbol dependencies, the feature maps are reshaped into one-dimensional sequences and fed into the LSTM module in order. The network sequentially processes the input and outputs the detection result for each symbol step-by-step. At each step, contextual information from previously detected symbols is leveraged to enhance the reliability of subsequent decisions. The architecture of the LSTM cell is illustrated in Fig. 3.

The forget gate f_t in the LSTM unit is used to determine whether the historical state information in the current symbol sequence should be retained, thereby enabling adaptation to dynamically changing channel conditions. The input gate i_t, on the other hand, controls the extent to which the extracted feature information from the current received symbol should be written into the cell state $\tilde{C}_t$. The cell state is then updated by combining the retained past information and the newly selected input features, where the updated cell state is computed as:

$$C_t = f_t \odot C_{t-1} + i_t \cdot \tilde{C}_t \tag{14}$$

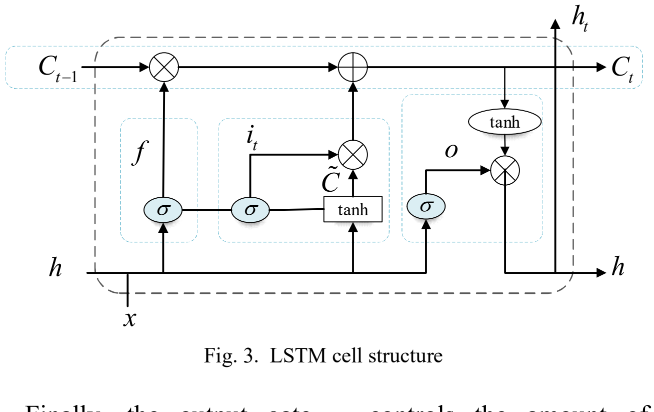

Fig. 3. LSTM cell structure

Finally, the output gate o_t controls the amount of information released from the current cell state to generate the hidden state h_t, which integrates both historical and current features and serves as the output for the final symbol decision.

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{15}$$

$$h_t = o_t \odot \tanh\left(C_t\right) \tag{16}$$

where x_t is the feature vector of the reconstructed symbol at time step t, h_{t-1} is the signal hidden state from the previous time step, W_o and U_o are the time weight matrices, b_o is the time bias term.

The output of the LSTM module is passed through a fully connected layer followed by a softmax classification layer to produce a probability distribution over modulation classes for each symbol. Symbol detection is then performed by selecting the constellation point corresponding to the highest probability, thereby reconstructing the transmitted symbol sequence $\hat{x}$. This end-to-end detection process directly optimizes the symbol mapping accuracy by minimizing the symbol error rate (SER), achieving a nonlinear inverse transformation from received signal $\boldsymbol{y}$ to the original transmitted data $\hat{x}$.

D. Nutcracker optimization algorithm design

Important factors, such as the regularization factor and learning rate, influence the performance of the network in the proposed model architecture. Traditional optimization methods exhibit two main issues: first, gradient descent algorithms are prone to becoming trapped in local optima; second, manually tuning parameters is both time-consuming and unable to effectively adapt to dynamic channel environments. To address these issues, we adopt the nutcracker optimization algorithm [20], which automatically identifies the optimal parameter combinations through an intelligent search strategy. This scheme specifically focuses on optimizing three core parameters: the learning rate η, which controls the granularity of gradient updates to prevent oscillation or stagnation during training; the L2 regularization coefficient λ, which balances model complexity and generalization capability; and the learning rate decay factor γ, which dynamically adjusts the learning step size to ensure stable convergence in the later stages of training.

The fundamental idea behind the nutcracker optimization algorithm [21] is derived from nutcrackers' foraging and food-caching habits. The algorithm consists of two key phases. First, the foraging and storage strategy initializes multiple candidate solutions within the search space and employs heuristic methods for random exploration to identify optimal hyperparameter combinations. Second, the cache-search and recovery strategy adaptively adjusts the search scope during the optimization process by integrating global exploration with local exploitation mechanisms, thereby guiding hyperparameters progressively toward the global optimum and accelerating convergence. Set the maximum number of iterations (M_{N}) and population size (P_{N}) first, then initialize the hyperparameter set at random as follows:

$$\boldsymbol{\theta}^{(0)} = \left\{\eta^{(0)}, \lambda^{(0)}, \gamma^{(0)}\right\} \tag{17}$$

where $\eta^{(0)} \in [\eta_{\min}, \eta_{\max}]$, $\lambda^{(0)} \in [\lambda_{\min}, \lambda_{\max}]$, $\gamma^{(0)} \in [\gamma_{\min}, \gamma_{\max}]$ .Second, define the fitness function $f(\boldsymbol{\theta})$ to measure the symbol error rate of the model on the validation set, thereby providing a criterion for evaluating the quality of hyperparameters during the optimization process:

$$f(\boldsymbol{\theta}) = \mathrm{SER}(\boldsymbol{\theta}) \tag{18}$$

Determine which candidate hyperparameter combination

$\theta^{(i)}$ has the highest fitness value, then note the current best option.

During the foraging and storage phase, the NOA searches the hyperparameter space by employing random perturbations, enabling individuals to perform global exploration:

$$\theta^{(t+1)} = \theta^{(t)} + \alpha \cdot \mathcal{N}(0,1) \tag{19}$$

where α is the adaptive step size and $\mathcal{N}(0,1)$ is a random variable drawn from the standard normal distribution, used to simulate the nature-inspired search strategy.

In the cache-search and recovery phase, NOA stores historically superior hyperparameters and uses them as a basis for local optimization:

$$\theta^{(t+1)} = \theta^{(t)} + \beta \cdot \left(\theta_{\text{best}} - \theta^{(t)}\right) \tag{20}$$

where β is the adaptive adjustment factor and θ_{best} is the optimal hyperparameter found during the current search. Finally, The optimization process terminates when the number of iterations reaches the predefined upper limit M_{N} or when the fitness function converges. The optimal hyperparameter θ_{best} is then output and applied to model training.

E. Loss function

Signal detection is essentially a multi-class classification problem, with the primary objective of accurately assigning received signals to their corresponding modulation types. The loss function should be designed to establish a dynamic balance among model convergence speed, classification accuracy, and the ability to distinguish challenging samples. Therefore, we take the categorical cross-entropy loss function as the optimization benchmark. The L2 coefficients are dynamically optimized by combining the adaptive adjustment mechanism of the NOA. This approach effectively balances model complexity and feature discrimination capability, significantly enhancing the robustness and generalization performance of the detection system. The loss function is defined as:

$$L_{CL} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) + \lambda \|W\|_2^2 \tag{21}$$

where C denotes the total number of categories, y_i is the true category label for symbol i, $\hat{y}_i$ is the probability distribution predicted by the model, λ is the regularization coefficient, and W is the weight parameter for each layer of the model. This loss function measures classification errors while simultaneously imposing constraints on the model parameters, effectively reducing the risk of overfitting.

IV. Experimental Analysis

A. Simulation experiment

The experimental setup of the OTFS system parameters is as: $N = M = 16$, $P = 4$, the maximum delay index $l_{\max} = 6$, the maximum Doppler shift index $k_{\max} = 4$, the carrier frequency is 4GHz, and the subcarrier spacing is 15kHz. We use a QPSK modulation scheme to construct a constellation diagram for symbol mapping. We consider the 300Hz Doppler expansion. Under randomly varying delay-Doppler

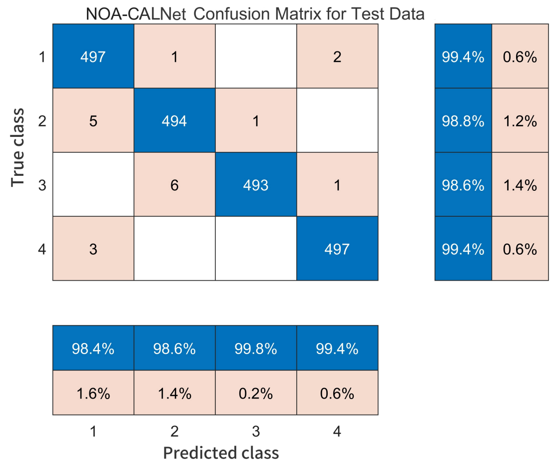

Fig. 4. NOA-CALNet confusion matrix for test data

channel conditions, we generated 10000 sets of training samples for each OTFS frame size to ensure that the model could fully learn across diverse channel environments. The training dataset and validation dataset are divided according to 8:2.

For the proposed model the parameters are set as: Each of the two stacked convolutional layers that make up the CNN module has a stride of one and a kernel size of 3×1. The number of input channels for these two layers is 32 and 64, respectively. The LSTM module sets the number of hidden cells to 40. The attention mechanism uses the squeeze-and-excitation with the compression ratio set to 4. With a population size of 20 and a maximum iteration count of 50, the NOA is implemented. The initial learning rate, L2 regularization coefficient, and learning rate decay factor are randomly selected as $[0.0001, 0.2]$, $[0.001, 0.2]$, and $[0.1, 0.9]$, respectively.

B. Simulation results analysis

Fig. 4 illustrates the confusion matrix of the NOA-CALNet model on the test set, which displays a clear diagonal pattern, reflecting high signal recognition accuracy. Specifically, the diagonal elements are significantly higher than the off-diagonal elements across all classes, indicating that the model demonstrates strong capability in distinguishing between different symbols. A comparison between the vertical axis (true classes) and the horizontal axis (predicted classes) reveals occasional minor misclassifications, primarily between the first and third classes. Additionally, a small degree of confusion is observed between the second and first classes. These misclassifications account for less than 1% of the total samples, suggesting that the model can reliably recognize signals across all categories, with errors occurring only under conditions of strong noise or closely spaced symbol boundaries.

The NOA-CALNet model achieves an overall recognition accuracy of 99.05% on the test set. This high level of accuracy results from the effective integration of feature extraction and temporal sequence modeling within the network architecture. Moreover, the attention mechanism effectively focuses on key feature regions, further enhancing detection performance. Under highly dynamic environments, the NOA-based adaptive optimization strategy enables efficient exploration of the parameter space and convergence toward optimal solutions, thereby mitigating the effects of overfitting and gradient vanishing.

C. Multi-scenario model performance analysis

The proposed model in our study is compared with commonly used models, including CNN, LSTM, CNN-LSTM, CNN-LSTM-Attention with the Adam optimizer, and the classical signal detection algorithm MMSE. Where the deep learning models share the same parameter settings as the proposed model in this study. This section analyzes the SER of the proposed model and various comparison models from two perspectives: the number of pilots and different noise environments.

As illustrated in Fig. 5, when the number of pilot symbols increases linearly, the symbol error rate (SER) of each algorithm exhibits a logarithmic decay trend, which aligns with theoretical expectations. The traditional MMSE detector exhibits noticeable accuracy loss within the 0–20 pilot range, as a result of instability in the channel matrix caused by its ill-conditioned characteristics. Deep learning-based methods exhibit relatively high SER due to their limited generalization capability in capturing channel characteristics. In contrast, NOA-CALNet maintains a low SER at a conductivity number of 5. As the number of pilots increases further, the SER of NOA-CALNet decreases at a faster rate, clearly outperforming the comparison models. This indicates that under low-pilot conditions, NOA-CALNet can more effectively maintain detection performance, allowing the system to allocate more resources to actual data transmission and thereby significantly improving overall communication efficiency.

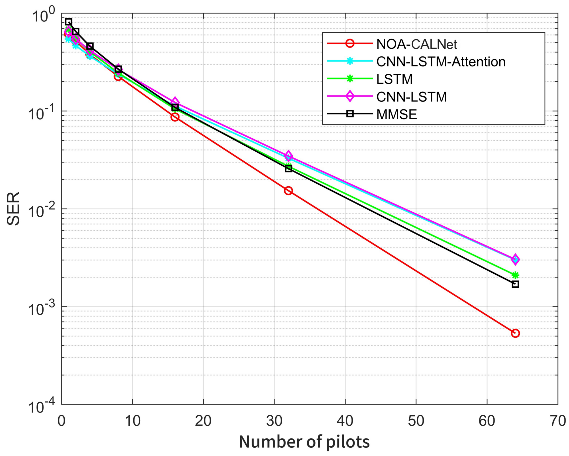

Fig. 5. Variation of SER performance with different numbers of pilots

An analysis of SER curves under Gaussian, Gamma, and Rayleigh conditions is conducted to assess the robustness of the proposed NOA-CALNet model across varying channel environments. In the standard Gaussian noise environment (Fig. 6), the SER of all detection methods decreases significantly with increasing signal-to-noise ratio (SNR). It is worth noting that when the SNR exceeds 8 dB, the proposed model exhibits a more pronounced reduction in SER. To achieve the same error performance, our model provides an SNR gain of nearly 4 dB compared with the conventional MMSE algorithm. This demonstrates that, in the presence of additive white Gaussian noise (AWGN), the suggested model

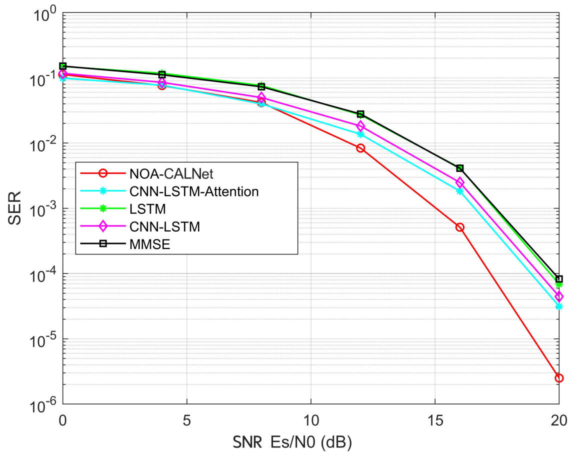

Fig. 6. SER performance variation in Gaussian noise environments

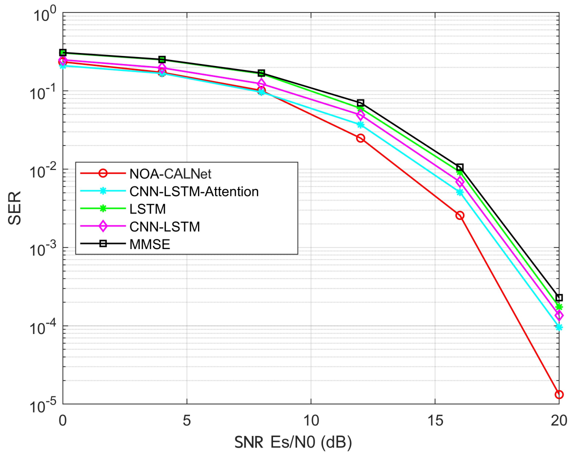

Fig. 7. SER performance variation in gamma noise environments

is accurate and robustness.

In the Gamma noise interference scenario (Fig. 7), the performance of all baseline models degrades, indicating their sensitivity to non-Gaussian heavy-tailed noise. Nevertheless, NOA-CALNet consistently achieves the best SER performance. Its feature selection mechanism effectively suppresses gradient dispersion caused by impulsive noise, demonstrating superior robustness in handling bursty and strong interference. Gamma noise is typically used to simulate non-ideal interference environments, such as areas with high electromagnetic disturbance or densely populated urban settings. The results suggest that the proposed model possesses strong practical applicability and generalization potential in real-world deployments.

Under Rayleigh time-varying fading (Fig. 8), traditional detection methods exhibit relatively limited adaptability to random channel fluctuations. Deep learning algorithms rely on convolutional and recursive structures to mitigate some of the effects of fading. Our model achieves a steeper SER decay curve by attention-gating and LSTM-based weighted diversity strategies. At an SNR of 20 dB, the SER reaches 2×10^{-3}. This validates its adaptability to fast-fading wireless channels and its resilience against multipath interference.

In addition, we compare the SER accuracy of different neural network-based detectors in OTFS signal detection. The detailed results are provided in Table I.

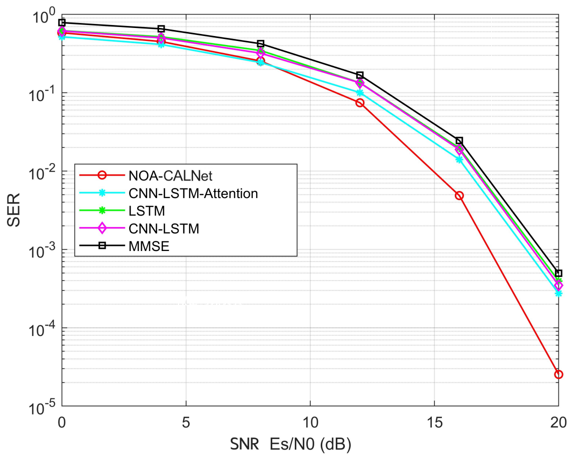

Fig. 8. SER performance variation in Rayleigh noise environment

TABLE I
SER ANALYSIS OF SIGNAL DETECTION ACROSS DIFFERENT MODELS

SNR	BPICNet	GNN	NOA-CALNet
10	0.1306	0.1528	0.1289
15	0.0158	0.0149	0.0132
20	0.0013	0.0012	0.0009

In summary, NOA-CALNet shows more prominent detection performance in multi-scenario experiments. This provides valuable insights for future applications in practical high-mobility OTFS communication scenarios.

D. Computational complex analysis

A comparison is conducted between the computational demands of the NOA-CALNet framework and the MMSE detection method. The overall computational cost of the model consists of the forward propagation process and the hyperparameter tuning overhead introduced by the NOA. However, during the inference phase, only forward propagation is required. Therefore, the overall complexity during inference can be expressed as:

$$T = \mathrm{O}\left(MN \cdot \sum_{l=1}^{L_{cnn}} K^2 \cdot F_l \right) + \mathrm{O}\left(MN \cdot H^2 + (MN)^2 \cdot H \right) \quad (22)$$

where K is the kernel size, L_{cnn} is the number of CNN layers, F_l is the number of convolutional filters in each layer, and H is the number of LSTM hidden units. The SE attention module introduces only a small number of global average pooling and fully connected operations, and its computational overhead is negligible.

Since the network parameters remain fixed during the testing phase, the complexity of the model grows linearly as $\mathrm{O}((MN)^2)$. As shown in Table II, a notable reduction in computational complexity is observed relative to the traditional MMSE detector. In high-dimensional OTFS systems, its advantage becomes even more pronounced.

TABLE II
COMPUTATIONAL COMPLEXITY

OTFS detector	Computation complexity
MMSE	$\mathrm{O}((MN)^3)$
NOA-CALNet	$\mathrm{O}((MN)^2)$

V. CONCLUSION

We propose NOA-CALNet, a model that integrates multi-feature extraction and optimization strategies. It integrates multi-feature extraction and optimization strategies. Experimental results show that NOA-CALNet effectively reduces SER under different noise environments. Compared to other detection methods, it shows superior detection performance across different SNR conditions. In terms of computational cost, NOA-CALNet is more efficient than the MMSE approach. This makes it more advantageous for practical detection applications.

REFERENCES

[1] H. H. H. Mahmoud, A. A. Amer, and T. Ismail, "6G: A comprehensive survey on technologies, applications, challenges, and research problems," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 4, p. e4233, 2021.

[2] M. H. Abid, I. A. Talin, and M. I. Kadir, "Reconfigurable intelligent surface-aided orthogonal time frequency space and its deep learning-based signal detection," *IEEE Access*, vol. 11, pp. 47321-47338, 2023.

[3] W. Kabir, "Orthogonal frequency division multiplexing (OFDM)," in *2008 China-Japan Joint Microwave Conference*, 2008, pp. 178-184: IEEE.

[4] R. Hadani *et al*., "Orthogonal time frequency space modulation," in *2017 IEEE wireless communications and networking conference (WCNC)*, 2017, pp. 1-6: IEEE

[5] P. Singh, A. Gupta, H. B. Mishra, and R. Budhiraja, "Low-complexity ZF/MMSE receivers for MIMO-OTFS systems with imperfect CSI," *arXiv preprint arXiv:2010.04057*, 2020.

[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] Z. Yuan, F. Liu, W. Yuan, Q. Guo, Z. Wang, and J. Yuan, "Iterative detection for orthogonal time frequency space modulation with unitary approximate message passing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 714-725, 2021.

[8] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational Bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7976-7980, 2020.

[9] H. Li, Y. Dong, C. Gong, Z. Zhang, X. Wang, and X. Dai, "Low complexity receiver via expectation propagation for OTFS modulation," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3180-3184, 2021.

[10] X. Yi and C. Zhong, "Deep learning for joint channel estimation and signal detection in OFDM systems," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2780-2784, 2020.

[11] G. Pan, Z. Liu, W. Wang, and M. Li, "A signal detection scheme based on deep learning in OFDM systems," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021, pp. 1-6: IEEE.

[12] A. Kosasih, X. Qu, W. Hardjawana, C. Yue, and B. Vucetic, "Bayesian neural network detector for an orthogonal time frequency space modulation," *IEEE Wireless Communications Letters*, vol. 11, no. 12, pp. 2570-2574, 2022.

[13] Y. K. Enku *et al*., "Two-dimensional convolutional neural network-based signal detection for OTFS systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2514-2518, 2021.

[14] S. Li, J. Yuan, W. Yuan, Z. Wei, B. Bai, and D. W. K. Ng, "Performance analysis of coded OTFS systems over high-mobility channels," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 6033-6048, 2021.

[15] T. Xiaogang, Z. Yiqing *et al*., "Survey of channel estimation method in delay-Doppler domain for OTFS," *Journal on Communications*, vol. 43, no. 12, pp188–201, 2022.

[16] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Effective diversity of OTFS modulation," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 249-253, 2019.

[17] D. Hartama, Mesran, A. P. Windarto, P. Alkhairi, and W. Rosdiana, "Improving stock market forecasting: Comparative insights into CNN architectures with a focus on ReLU activation," *IAENG International Journal of Computer Science*, vol. 52, no. 3, pp862-871, 2025

[18] M. Patacchiola, J. Bronskill, A. Shysheya, K. Hofmann, S. Nowozin, and R. Turner, "Contextual squeeze-and-excitation for efficient few-shot image classification," *Advances in Neural Information*

Processing Systems, vol. 35, pp. 36680-36692, 2022.
[19] S. M. Al-Selwi *et al*., "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *Journal of King Saud University-Computer and Information Sciences*, p. 102068, 2024.
[20] M. Abdel-Basset, R. Mohamed, M. Jameel, and M. Abouhawwash, "Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems," *Knowledge-Based Systems*, vol. 262, p. 110248, 2023.
[21] X. Liu, J. Zhu, J. Wang, S. Zhang, and X. Guan, "Nutcracker optimization algorithm to address power system's optimal power flow issue," *Engineering Letters*, vol. 32, no. 8, pp1644-1658, 2024