IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE POWER & ENERGY SOCIETY SECTION

Received 19 April 2024, accepted 21 June 2024, date of publication 25 June 2024, date of current version 3 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418883

RESEARCH ARTICLE

False Data Injection Attack Detection Method Based on Deep Learning With Multi-Scale Feature Fusion

JINPENG JI[1], YANG LIU[1], (Member, IEEE), JIAN CHEN[2], ZHIWEI YAO[1], MENGDI ZHANG[1], AND YANYONG GONG[1]

[1]School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China
[2]Zibo Metrology Technology Research Institute, Zibo 255025, China

Corresponding author: Yang Liu (bqxyl@sdut.edu.cn)

This work was supported in part by the State Grid Corporation of China Technology Project 52094017003D.

ABSTRACT Cyber-attacks, especially the false data injection attack (FDIA), are gradually becoming a common way to threaten the regular operation of power grid. However, the FDIA is challenging to detect because it prevents the bad data detection mechanism in the energy management system from destroying the integrity of measurement information. Aiming at the problem of the FDIA detection in smart grids, this paper presents a FDIA detection method based on deep learning with multi-scale feature fusion. First, the improved convolution neural network (ICNN) is used to predict measurement data by combining convolution neural network with the Inception v1 module. Then, the attention mechanism is introduced into the ICNN to extract and fuse full and partial features of measurement data. By fitting the function between measurement and state vectors, the state data are generated with predicted measurement data. Eventually, the threshold of divergence is obtained to determine whether the FDIA occurs or not by the difference in probability distribution between predicted and actual state vectors. The performance of the proposed method is evaluated in the IEEE 14-node and 39-node test systems. The results show that the proposed method can accurately detect the existence of FDIA in time. This method has definite robustness to noise and distributed generation switching.

INDEX TERMS False data injection attack, convolution neural network, feature fusion, multi-scale convolution, dynamic state estimation.

I. INTRODUCTION

The enhanced interaction between information flow and power flow makes vulnerable nodes in power system increasingly vulnerable to attacks [1]. As a cyber-attack with higher threat to power system, the false data injection attack (FDIA) can cause system stability problems and system breakdown in the power system by tampering with crucial control parameters, time-scale information, or measurement data [2], [3]. Before the occurrence of the FDIA, the attacker needs to know all or part of the system topology and line parameters information to construct specific attack vectors [4], [5]. In addition, attackers can also generate an approximate system Jacobian matrix to construct specific attack vectors by mining hidden system information in historical operation data of power system [6]. The residual between measured data is still less than the set threshold during the FDIA because of the specific attack vector constructed, making it difficult for bad data detection schemes in energy management systems (EMS) to detect the FDIA [4]. Therefore, how to reliably and effectively detect the FDIA has essential significance for ensuring the stable operation of the power system.

A. RELATED WORK

Depending on whether information on system topology and line parameters is required, the FDIA detection methods are divided into the model-based methods and the data-based methods [7]. Most model-based FDIA detection methods use the state estimation methods to estimate the current state of the system by the information on system topology and

The associate editor coordinating the review of this manuscript and approving it for publication was Sarasij Das.

line parameters [8], such as weighted least squares (WLS) estimation [9], median filtering (MF) [10], and Kalman filter (KF) [11], [12]. On this basis, the FDIA is detected by comparing the difference between actual state data and estimated state data. However, the model-based FDIA method requires prior knowledge of the system topology and line parameters, and its detection reliability is easily affected by uncertainty of the system parameters [13].

Since the measurement data contains system physical structure information, and real-time data analysis can effectively solve the detection performance problems caused by system model uncertainty to a certain extent, data-based methods have been used to detect the FDIA in recent years. The data-based FDIA detection methods do not depend on the information of system topology and line parameters. Dou et al. [14] proposed a hybrid detection mechanism combined with variational mode decomposition (VMD) and an extreme learning machine (ELM). The measurement data decomposed by VMD is input into the ELM to learn data features and realize the FDIA detection. Zhang and Yang [15] proposed a deep neural network-based approach, which transformed the attack detection problem into a supervised learning problem to detect the FDIA. Bhattacharjee et al. [16] proposed a deep latent space clustering algorithm in which the trainable clustering heads are stacked on top of the finetuned autoencoder for model training and detecting the existence of FDIA. The above methods use normal and attack sample data to train and learn the model. However, the difficulty in obtaining actual attack sample data may limit the effectiveness of these methods. In addition, data-based prediction methods have been implemented for the FDIA detection. Lei et al. [17] used a gated recurrent neural network to predict measurement data, which are compared with the observed measurement data to realize the FDIA detection by convolution neural network (CNN). Wang et al. [18] proposed an integrated two-level learner to detect the FDIA by the distribution of the sum of square errors between the observed and the predicted. Considering the inherent temporal correlation between consecutive measurement data in the system, Reda et al. [19] proposed a deep recurrent neural network method and introduced Kullback-Leibler (KL) divergence into the binary classifier to detect the FDIA. However, this method did not consider the potential impact of the spatial features of the data on the detection results. In addition, these data-based prediction methods only consider the full-scale features in measurement data but ignore the partial-scale features in measurement data that can better reflect the change details of the system operation state.

B. CONTRIBUTIONS

This paper introduces the improved Inception v1 module into CNN to construct the improved convolution neural network (ICNN), combining with gated recurrent unit (GRU), attention mechanism, and Wasserstein divergence (WD) to propose a FDIA detection method based on deep learning with multi-scale feature fusion. The contributions of this paper can be summarized as follows:

1) Constructing the measurement matrix in the input layer can enable the model to extract the temporal and spatial features of data. Then, the attention mechanism is used to extract partial-scale features, which makes the model pay more attention to critical partial-scale features, thereby improving the overall performance of the proposed model.

2) A model with system measurement data prediction and state data generation capabilities is built to predict system states, which, together with actual system states, are used for the FDIA detection by WD. This model only requires system average sample data for model training and does not need to consider the difficulty of obtaining actual attack sample data.

3) The FDIA can be detected without relying on information on system topology and line parameters, and the effectiveness of the proposed method has been verified on an open real-time dataset and two open test systems.

The rest of this paper is organized as follows. Section II presents the basic principles of the FDIA and two deep learning methods. Section III focuses on the proposed FDIA detection method, which includes the process of data prediction, state generation, and attack detection. Section IV presents the validation and result analysis of detecting the FDIA for the IEEE 14 nodes and 39 nodes test systems by using the proposed method. Finally, Section V concludes this paper.

II. PRELIMINARIES

This section introduces the principles of the FDIA, CNN method, and GRU method.

A. FALSE DATA INJECTION ATTACK PRINCIPLE

The operating state of the power system is determined by proper measurement data. The measurement data typically include voltage magnitude, power injection, and line power flow. The voltage magnitudes and phase angles can completely describe the operating state of a power system on all buses. The relationship between the measurement data and the states can be formulated as follows in the power system:

$$\boldsymbol{z} = h(\boldsymbol{x}) + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^N$ is the state vector at a time step; $\boldsymbol{z} \in \mathbb{R}^M$ is the measurement vectors at a time step; $h(\cdot)$ is a measurement function between the measurement vectors and the state vectors; $\boldsymbol{e} \in \mathbb{R}^M$ is a measurement error vector. Assuming that each noise $e_i \in \boldsymbol{e}$ is mutually independent and follows a zero Gaussian distribution with a zero mean.

The FDIA can circumvent the residual-based bad data detection mechanism of modern energy management system. When the FDIA occurs, the residual value of the bad data detection mechanism during the attack is as follows:

$$r_a = \left\| z_a - h(\hat{\boldsymbol{x}}_c) \right\|_2 = \left\| \boldsymbol{z} + \boldsymbol{a} - h(\hat{\boldsymbol{x}} + \boldsymbol{c}) \right\|_2 \tag{2}$$

where $\boldsymbol{r}_a$ is the residual value of the measurement vectors during the attack; $\boldsymbol{z}_a \in \mathbb{R}^M$ is the measurement vectors

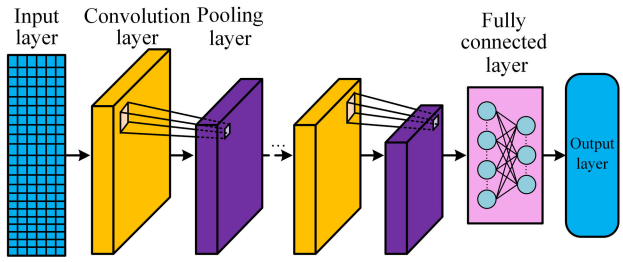


FIGURE 1. Structure diagram of the CNN model.

during the attack; $\hat{\boldsymbol{x}}_c \in \mathbb{R}^N$ is the estimated state vector; $\boldsymbol{a} \in \mathbb{R}^M$ is the attack vector; $\boldsymbol{c} \in \mathbb{R}^N$ is the estimated deviation state vector. If $\boldsymbol{a} = h(\hat{\boldsymbol{x}} + \boldsymbol{c}) - h(\hat{\boldsymbol{x}}_c)$ is satisfied, thus residual value is $r_a = r$, where r is the residual value of the measurement vector under normal condition. This means that attackers can circumvent bad data detection mechanism by constructing specific attack vectors to influence residual value of the measurement vector. The specific principle of the FDIA can be found in [4].

B. CONVOLUTIONAL NEURAL NETWORK

As a feedforward neural network, CNN can effectively extract the relevant features of input data [20]. A typical CNN is mainly comprised of convolution layers and pooling layers, as shown in Figure 1.

The convolution layer, as the core component of CNN, is used to extract the features of the measurement data in the input layer, and convolution operation as follows:

$$c_{i,j,v} = f\left(\sum_{e=1}^{m}\sum_{p=0}^{d-1}\sum_{q=0}^{d-1}\rho_{p,q,e,v}x_{i+p,j+q,e} + b_v\right) \tag{3}$$

where $x_{i,j,e}$ is every entry of the input measurement data in the e^{th} input channel; $c_{i,j,v}$ is every entry of the extracted spatial and temporal feature in the v^{th} output channel; $\rho_{p,q,e,v}$ is each entry of the kernel square; b_v is the bias of kernel in the v^{th} output channel; d is size of kernel; m is the total number of input channel; $f(\cdot)$ is the activation function and the ReLU function is selected as the activation function in this paper.

There are two main pooling layers: the max pooling layer and the average pooling layer. The pooling layer is added after the convolution layer to downsample the feature, which further reduces the location sensitivity of extracted features and improves computational efficiency.

C. GATED RECURRENT UNIT

GRU, a variant of the long short-term memory neural network (LSTM), combines the input gate and the forget gate in the LSTM into a single update gate and performs training learning it together with the reset gate [21]. GRU has the characteristics of a less gating structure with fewer learning parameters. The structure of the GRU model is shown in Figure 2.

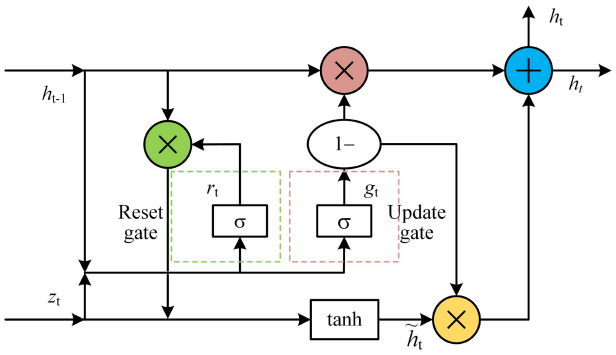


FIGURE 2. Structure diagram of the GRU model.

The GRU cell is the core of the GRU model, the expressions for each variable in the GRU cell are as follows:

$$\begin{aligned} z_t &= \sigma(\boldsymbol{W}_z x_t + U_z h_{t-1}) \\ r_t &= \sigma(\boldsymbol{W}_r x_t + U_r h_{t-1}) \\ \tilde{h}_t &= \tanh(r_t \times U h_{t-1} + \boldsymbol{W} x_t) \\ h_t &= (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \end{aligned} \tag{4}$$

where z_t and r_t, respectively, are the state of update gate and reset gate; x_t is the input measurement data; h_t is the hidden state; $\tilde{h}_t$ is the updated hidden state; $\boldsymbol{U}_z$ and $\boldsymbol{W}_z$, respectively, are the weight of the hidden state and the input measurement data in the update gate; $\boldsymbol{U}_r$ and $\boldsymbol{W}_r$, respectively, are the weight of the hidden state and the input measurement data in the reset gate; $\boldsymbol{U}$ and $\boldsymbol{W}$ are the training parameter.

III. METHODOLOGY

A. MEASUREMENT MATRIX CONSTRUCTION

If the measurement data between different buses has the spatiotemporal correlation in the power system, the temporal and spatial feature of the measurement data can be extracted to analyze the difference between the normal data and the attack data, and then the FDIA can be detected. In order to conveniently extract the temporal and spatial features of the measurement data, the measurement input matrix of the detection model is constructed. The measurement matrix is as follows:

$$\boldsymbol{Z} = \begin{bmatrix} z_{1,k-N'} & \cdots & z_{1,k-1} & z_{1,k} \\ z_{2,k-N'} & \cdots & z_{2.k-1} & z_{2,k} \\ \vdots & \ddots & \vdots & \vdots \\ z_{M,k-N'} & \cdots & z_{M,k-1} & z_{M,k} \end{bmatrix} \tag{5}$$

where $z_{M,k}$ is the measurement data of the M^{th} measurement device at the sample time k; N' is sample time of maximum time correlation. The value of N' can be determined by $N' = \max\{i \,||R(i)| \geq \sigma\}$, where σ is the given confidence level and $R(i)$ is the autocorrelation function value of the time series at different delayed sample time i. In the larger and more complex test systems, test system is divided into several small

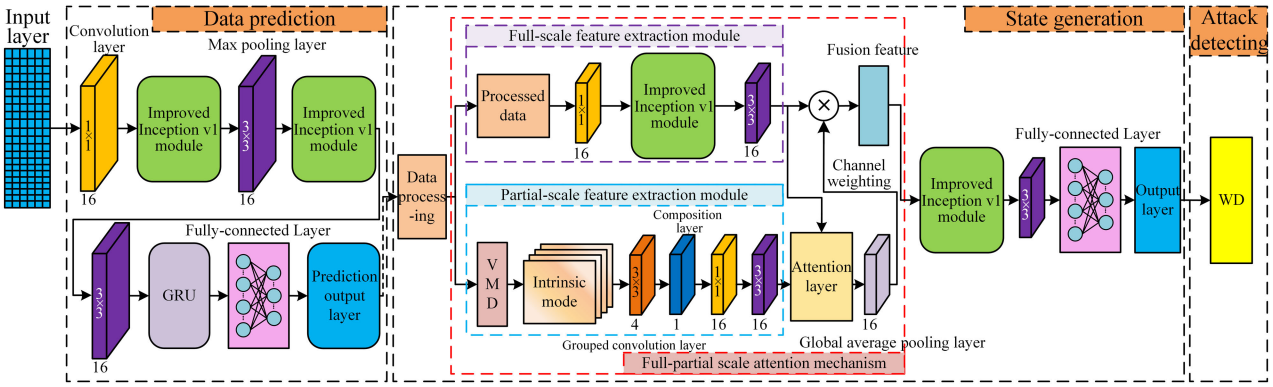


FIGURE 3. Framework of the FDIA detection method.

and simple areas by graph theory knowledge [22], and attacks can be detected by constructing measurement matrices with the node data of this regional system in differently regional systems.

B. DETECTION MODEL OF FDIA

1) FRAMEWORK OF THE FDIA DETECTION METHOD

The framework of the proposed FDIA detection method is comprised of data prediction, state generation, and attack detection in this paper. The detection model of proposed method is shown in Figure 3.

In Figure 3, the data prediction part consists of the ICNN, GRU, fully-connected layer, and predictive output layer, where the ICNN includes a convolution layer, two improved Inception v1 modules, and two max pooling layers. The measurement data in the power system has a specific temporal and spatial correlation. Therefore, the spatial and temporal features of the measurement data are extracted by the multi-scale convolution layer of the ICNN. Then, the implicit correlation between the continuous time sequence of output data from ICNN is learned by GRU, so the active power data of the power system at the future time are effectively predicted. Eventually, the predicted active power data of the predictive output layer is entered into the state generation part.

Considering that some fine-grained details of measurement data contain important information. This paper introduces the attention mechanism into ICNN to constitute the state generation part. The attention mechanism can extract and fuse full and partial scale features of the measurement data after data processing, and the attention layer weights and sums the extracted full and partial scale features, which are further fused into fusion features. The spatial and temporal features of the fusion features are further extracted by the improved Inception v1 module in ICNN. Then, the full-connected layer connects the features extracted by ICNN with the attention mechanism with the preloaded system state data to mine the intrinsic relationship between extracted features and state data in the training process, and fit the measurement

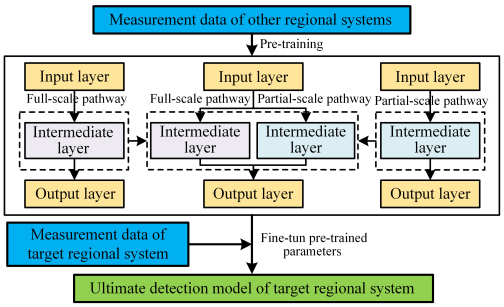


FIGURE 4. Transfer learning process of the proposed detection model.

function $h(\cdot)$ between the measurement vectors $\boldsymbol{z}$ and the state vectors $\boldsymbol{x}$, thereby realizing the purpose of the state data generation based on the measurement data. Finally, the distance of probability distributions between the actual state vectors and the predicted state vectors is calculated to detect the FDIA by the WD method.

2) TRANSFER LEARNING

In the process of detection model training, it is more difficult to directly train the whole model, while the alone training of each module in the whole model is relatively fast and effective. Therefore, this paper introduces the idea of model-based transfer learning [23] into the training process to reduce the training time of the detection model by transferring pre-trained parameters of the model to the objective network model. The transfer learning process of the model is shown in Figure 4.

The state generation part in the proposed detection model mainly includes the full and partial scale feature extraction module. As shown in Figure 4, during the training process of the proposed detection model, the full and partial scale feature extraction modules are trained, respectively, and the trained parameters are used as pre-trained parameters in the overall detection model. Then, the pre-trained parameters, such as the dropout rate, the learning rate, step sizes of convolution and max pooling layers, and pooling size, need to be further fine-tuned by prior knowledge until the training of the ultimate

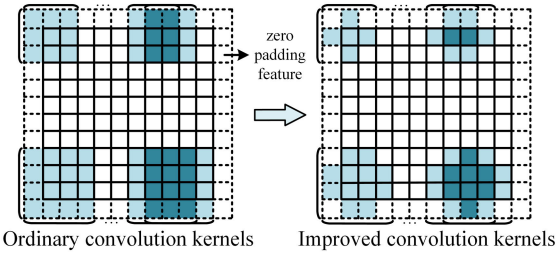


FIGURE 5. Improved processing of convolution kernels.

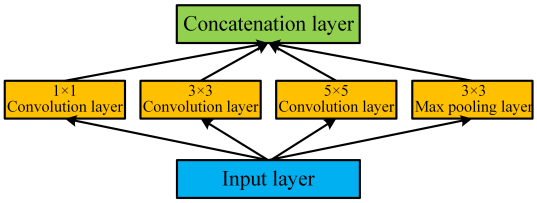


FIGURE 6. Structure of inception v1 module.

detection model is completed. If the system model is large, the system model can be divided into different small regional systems. During the training process of the detection model for the target regional system, the trained parameters can be introduced as pre-trained parameters into the detection model of the target regional system, and fine-tuned the pre-trained parameters by prior knowledge until the training of the ultimate detection model in the target regional system is completed.

C. IMPROVED CONVOLUTION NEURAL NETWORK

In ICNN, this paper improves the structures of the convolution kernels and convolution layers to enhance the effectiveness of the detection method. This section includes an improved convolution kernel and an improved Inception v1 module.

1) IMPROVED CONVOLUTION KERNEL

As the core of the convolution layer, the convolution kernel is used to extract data features. In order to avoid the edge features of the feature map being ignored, it is usually necessary to perform operations of zero padding before the convolution operation. At the same time, the operations of zero padding increase calculations of feature extraction, especially zero padding features. This paper improves the convolution kernel. The improved processing of the convolution kernel is shown in Figure 5. All 3×3 and 4×4 convolution kernels are improved convolution kernels in the proposed model.

2) IMPROVED INCEPTION V1 MODULE

The Inception v1 module has excellent capability in data feature extraction [24]. The Inception v1 module is shown in the Figure 6.

The computation of feature extraction operation in the Inception v1 module is as follows:

$$C_v = \sum_{d=1,3,5} D_{H,d}^2 mn D_F^2 + C_p \tag{6}$$

where $D_{H,d}^2$ is the size of the $d \times d$ convolution kernel; m and n, respectively, are the total channel number of feature map before and after convolution; D_F^2 is the size of the feature map; $C_p = D_{H,3}^2 m D_F^2$ is the computation of max pooling layer. The computation of ordinary convolution layer is $C_o = D_{H,3}^2 mn D_F^2$, it can be seen that the computation of the Inception v1 module is more than three times that of the ordinary convolution layer. In addition, the concatenation layer increases the channel number of the feature map, which further increases the subsequent computation of the module.

To improve the feature extraction capability of the ordinary convolution layer in CNN, the ordinary convolution layer is replaced with the Inception v1 module in this paper. However, introducing the Inception v1 module can increase the computation of the model, so this section improved the Inception v1 module. The improvement processes of the improved Inception v1 module are as follows:

Step 1: The Improved Inception v1 module removes the 3×3 max pooling layer of ordinary Inception v1 module, and replaces the 1×1, 3×3, and 5×5 convolution layers in Inception v1 modules with the multi-size grouped convolution layer containing 2×2, 3×3, and 4×4 grouped convolution layers, which could produce three $a \times b \times m$ feature maps by grouped convolution operation. The convolution operation is formulated follows:

$$c(d)_{i,j,e} = f\left(\sum_{p=0}^{d-1}\sum_{q=0}^{d-1} \rho(d)_{p,q,e} x_{i+p,j+q,e} + b(d)_{i,j,e}\right) \tag{7}$$

where is the extracted feature of 2×2 convolution kernel; $d \in \{2, 3, 4\}$ is the size of convolution kernel.

Step 2: To reduce the number of output channels of the terminal feature map, the concatenation layer is replaced by the composition layer and the convolution layer. The three different output feature maps in the multi-size grouped convolution layer are combined into a $a \times b \times m$ feature map by the composition layer. The operation of the composition layer is formulated as follows:

$$c_{i,j,e} = c(2)_{i,j,e} + c(3)_{i,j,e} + c(4)_{i,j,e} \tag{8}$$

Then, the corresponding position feature of different channels are convoluted to a $a \times b \times n$ feature map by the convolution layer. The convolution operation is as follows:

$$c_{i,j,v} = f\left(\sum_{e=1}^{m} \rho_{i,j,e,v} x_{i,j,e} + b_{i,j,v}\right) \tag{9}$$

Through the above operations of feature extraction, the structure of the improved Inception v1 module is shown in the Figure 7.

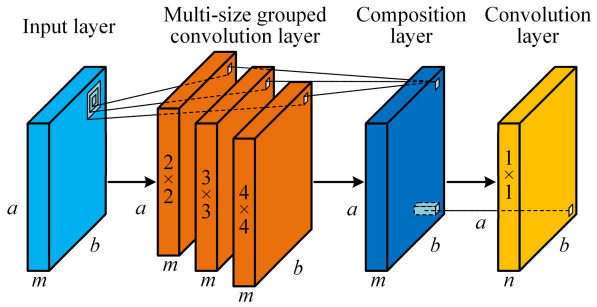


FIGURE 7. Structure of improved inception v1 module.

The computation of improved Inception v1 module is formulated follows:

$$C = \sum_{d=2}^{4} D_{H,d}^2 m D_F^2 + 3D_F^2 + mnD_F^2 \tag{10}$$

The computation ratio of improved Inception v1 module to ordinary convolution layer is formulated follows:

$$K = (\sum_{d=2}^{4} D_{H,d}^2 m + 3 + mn)/(D_{H,3}^2 mn) \tag{11}$$

During process of training and test, the convolution layers in the proposed detection model have 16 channels. Therefore, when $m = n = 16$, the computation ratio K is approximately 2/5. According to the above, the improved Inception v1 module requires lesser computation than the ordinary Inception v1 module and the ordinary convolution layer under the same simulation conditions.

This paper replaces the ordinary convolution layer in CNN with the improved Inception v1 module to construct the ICNN, and combines the ICNN with GRU to realize system measurement data prediction.

D. FULL-PARTIAL SCALE ATTENTION MECHANISM

The changes of measurement data in the partial nodes can affect influence the system state in the normal operation of system. Most neural networks only extracted the full-scale information from the input measurement data, ignoring the partial fine-grained details in measurement data, but the partial scale features in measurement data were often most important. The full-scale features refer to the data features under power frequency, and the data can be directly measured by system measurement equipment. The full-scale features in this paper are the amplitude features of active power data under power frequency. The partial-scale features refer to the data features at a specific frequency scale. According to signal processing theory, the partial features of the measurement data signal are clearer at a specific frequency scale, so this paper uses the data features of the intrinsic modes obtained from active power data are processed by VMD as the specific scale data.

The attention mechanism [25] can extract global information from the whole important image and local fine-grained

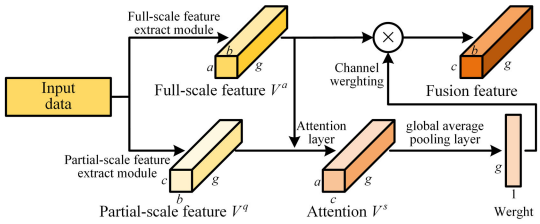


FIGURE 8. Structure of the full-partial scale attention mechanism.

details from local patches, which highlighted import information with high weight by giving features different weights. Therefore, this paper introduces the attention mechanism into the proposed model to extract full and partial scale features in system data. Considering that there are differences in acquisition ways between local patches in [25] and the specific scale data in this paper, the VMD processing part is added in the partial-scale feature extraction module. Similar to [28], to avoid unimportant features in the full-scale features having larger weighting, this paper adds a global average pooling layer after the attention layer to realize the channel weighting, which can more effectively highlight important detailed features in the full-scale features and improve effectiveness of fusion. The improved attention mechanism is named the full-partial scale attention mechanism, which is shown in Figure 8.

As shown in Figure 8, the attention layer can weight and sum the extracted full and partial scale features. The attention is formulated as follows:

$$\boldsymbol{V}^s = \text{softmax}(\boldsymbol{V}^a(\boldsymbol{V}^q)^T/\sqrt{g}) \tag{12}$$

where $\boldsymbol{V}^s$ is the attention; $\boldsymbol{V}^a$ and $\boldsymbol{V}^q$, respectively, are the full and partial scale feature; g is the amount of channels; softmax$(\cdot)$ is the normalized exponential function.

This paper introduces full-partial scale attention mechanism into ICNN to extract full and partial scale features of predicted measurement data, which makes ICNN pay more attention to the key partial-scale features and improves the overall performance of proposed method. Before the predicted measurement data are input into the attention mechanism, this paper uses the principal component analysis method [26] to deal with the predicted measurement data and reduce data dimensionality. The attention mechanism uses improved Inception v1 module with excellent capabilities of feature extraction to extract the full-scale features. For the partial-scale features, this paper uses VMD to obtain the partial-scale data of the predicted measurement data at the specific frequency scale. The VMD adaptively realizes the frequency domain decomposition of the signal and the effective separation of each component by iteratively searching for the optimal solution of the variational mode [27]. The VMD can decompose the processed data into several intrinsic modes, i.e., variational modes, which contain essential the partial-scale features of the system. The intrinsic modes are input into the partial-scale feature extraction module to

extract the partial-scale features. Then, the attention layer can weight and sum the extracted full and partial scale features, which are further fused into fusion features with the full-scale feature by global average pooling layer.

E. TRAINING STRATEGY AND HYPERPARAMETER TUNING

Essentially, the proposed detection model is a regression task. The mean-square error function is the loss function of the proposed detection model in this paper, and the loss function is formulated as follows:

$$L = \sum_{i=1}^{H} \sum_{j=1}^{R} |t_{i,j} - y_{i,j}|/H \tag{13}$$

where $t_{i,j}$ the actual value; $y_{i,j}$ is output value of network; R is the amount of output; H is the length of sequence.

For the detection model to have better computational efficiency, the adaptive moment estimation (Adam) algorithm is designated as the optimizer. The two exponential decay rates of the Adam optimizer, respectively, are set to 0.9 and 0.999, and its learning rate is set to 0.001. In addition, this paper uses the 10-fold cross-validation method [29] to tune the hyperparameters of the proposed method model, such as size and stride of the convolution kernel and pooling size, the dropout rate and the neuron number of the GRU module, and the total epochs of the training strategy. For the convolution layer and max pooling layer in the proposed method model, Figure 3 and Figure 7 showed the size and channel number of the convolution layer and max pooling layer, and the strides of the convolution kernel and pooling are both set to 2. For the GRU module, the proposed method model sets the neuron number and dropout rate of GRU module are 200 and 40%, respectively. In addition, the epoch of the proposed method model is 100.

F. FDIA DETCETION

In a normal system state, the predicted state vectors have almost the same probability distribution as the actual state vectors. However, the FDIA tamper with the measurement data of partial nodes, resulting in changes in the probability distribution of actual state vectors. Therefore, FDIA can be detected by the distance between probability distributions of the actual and predicted state data. WD method originated from optimal transport theory, which can measure difference of probability distributions between two data sets [30]. Compared with the KL and Jensen-Shannon divergence, WD can calculate the difference between the two probability distributions without overlap [31], so this paper used WD to catch the differences between the probability distribution of actual and predicted state vector for detecting the FDIA. The WD is formulated as follows:

$$W_d(P_{\hat{X}}, P_{\hat{Y}}) = \inf_{\gamma \sim \prod(P_{\hat{X}}, P_{\hat{Y}})} E_{(x,y)\sim\gamma}[||x - y||] \tag{14}$$

where $\hat{\boldsymbol{X}} \in \boldsymbol{X}$ is the actual state vector; $\hat{\boldsymbol{Y}} \in \boldsymbol{Y}$ is the predicted state vector; $P_{\hat{X}}$ and $P_{\hat{Y}}$ are the probability distributions of $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$ respectively; $\prod(P_{\hat{X}}, P_{\hat{Y}})$ is the set of all joint distributions between $P_{\hat{X}}$ and $P_{\hat{Y}}$; x and y are the state sample data of $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$ respectively. $||x - y||$ is the distance of state data x and y. The WD between actual state vector $\boldsymbol{X}$ and predicted state vector $\boldsymbol{Y}$ is formulated follows:

$$W_d(P_X, P_Y) = [W_d^1, W_d^2, \cdots, W_d^{N-b+1}] \tag{15}$$

where b is the window size of state vectors $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$; P_X and P_Y are probability distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. In the simulation, W_d is the WD between actual state vector and predicted state vector, which is used to detect existence of the FDIA.

Whether the FDIA is determined to exist depends on whether the distance of probability distributions exceeds the set threshold after calculating the result of WD. Assuming that the measurement noise follows a Gaussian distribution, the actual and predicted state vectors conform to the normal distribution. Therefore, the mean and standard deviation of calculated divergences are used to realize the FDIA detection based on 3 times standard deviation principle and reference [32]. Considering that data fluctuations can affect the calculation divergence in the system normal operation, similar to [33], this paper adds a correction factor τ to the standard deviation of calculation divergence in the 3 times standard deviation principle, which can reduce the false alarm rate as much as possible when the detection probability meets the requirement. The set divergence threshold is as follows:

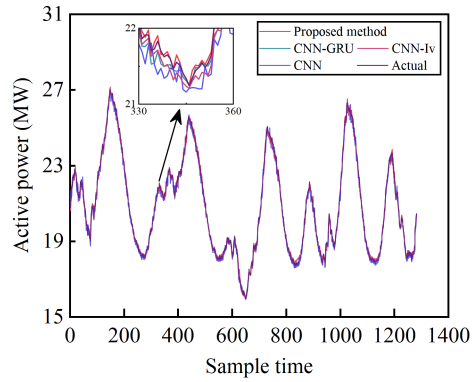
$$\psi = \varsigma + \tau S \tag{16}$$

where ψ is the threshold of divergence; ς is the average of divergence; S is the standard deviation of divergence. The expressions of the variance and mean in divergence are as follows:

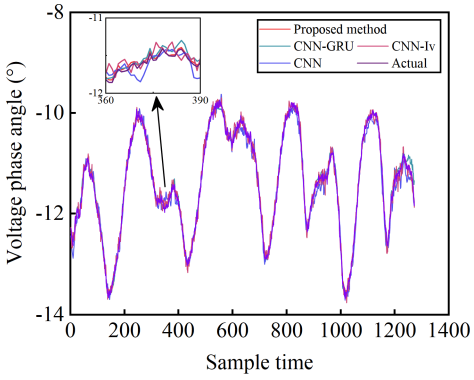
$$\begin{aligned} \varsigma &= \sum_{i=1}^{N-b-1} W_{d_i}/(N-b+1) \\ S &= \sqrt{\sum_{i=1}^{N-b-1} (W_{d_i} - \varsigma)^2/(N-b-1)} \end{aligned} \tag{17}$$

IV. SIMULATION AND EXPERIMENTAL RESULTS

The simulation test environment in this paper is deployed on a computer with an NVIDIA GeForce MX250 GPU (having 12 GB GDDR5 VRAM), Intel(R) Core i5 2.5GHz CPU, and 8GB RAM. And MATLAB R2020b is adopted as the software platform. The abnormal datasets in this paper are structured by the attack model [4]. In order to simulate the more realistic running state of the power system, this paper uses the load data of 5-minute interval provided by New York Independent System Operator (NYISO) [34] during March 1, 2023 and March 31, 2023 and the test systems similar to [14] to test the proposed method. The load data of NYISO are used to produce measurement data and state data of system containing 4286 sample times by Matpower toolbox [35] as



(a) Comparison of predicted active power



(b) Comparison of generated voltage phase angle

FIGURE 9. Result comparison of different methods.

sample data, where the active power data in measurement data as input measurements of the proposed method model. The ratio of training and testing samples in detection model of the proposed method is 7:3, that is, the first 3000 sample data are used for training and the last 1286 sample data are used for testing.

A. SIMULATION ANALYSIS OF DATA PREDICTION AND STATE GENERATION

The proposed detection method uses the data prediction part to extract the temporal and spatial features to predict the active power data of the nodes. Then, the state generation part uses the predicted measurement data to generate the state data of nodes by extracting the temporal and spatial features of the data and fitting the measurement function. For comparing the model performance of the proposed method with CNN, the method combining CNN with ordinary Inception v1 module (CNN-Iv), and the method combining CNN with GRU (CNN-GRU) in data prediction and data generation, taking the 11^{th} node as an example, the measurement data prediction and state data generation results of 11^{th} node are shown in Figure 9.

As shown in Figure 9, the difference between the actual data in the proposed method and the output data is smaller than CNN, CNN-Iv, and CNN-GRU. In order to future evaluate the prediction and fitting results of different detection methods, this paper uses mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) as the evaluation indexes. The prediction and fitting results of different detection methods are shown in Table 1.

TABLE 1. Performance comparison of different methods.

Detection methods	Data prediction stage			State generation stage		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
Proposed method	0.052	0.068	0.250	0.018	0.022	0.153
CNN-GRU	0.078	0.101	0.597	0.050	0.068	0.440
CNN-Iv	0.082	0.103	0.643	0.059	0.080	0.496
CNN	0.136	0.136	1.242	0.095	0.119	0.846

TABLE 2. Model decision-making speed of different methods.

Detection methods		Proposed method	CNN-GRU	CNN-Iv	CNN
Data prediction stage	Total time/s	468.943	515.943	623.315	258.580
	Training time/s	468.177	515.073	622.430	258.010
	Testing time/s	0.766	0.870	0.885	0.570
State generation stage	Total time/s	725.237	1024.614	1265.715	604.476
	Training time/s	724.067	1022.631	1263.523	603.574
	Testing time/s	1.170	1.983	2.192	0.902

As shown in Table 1, the model evaluation indexes of the proposed method are lower than CNN-GRU, CNN-Iv, and CNN, that is, the proposed method has higher effectiveness in extracting temporal and spatial features of measurement data and has batter capabilities of data prediction and state generation. For example, the MAPE value of the proposed method in state generation stage is about 1/5, 1/3, and 1/3 of the CNN, CNN-Iv, and CNN-GRU, respectively.

In order to further compare the simulation performance of above methods, this section compares model decision-making speed of different methods, which are shown in Table 2.

As shown in Table 2, the training and testing time of the proposed method is more than the CNN but lesser than the CNN-GRU and CNN-Iv, that is, the decision-making speed of the proposed method is between CNN and CNN-GRU. On the basis of the same epoch, training samples, and testing samples, Although the training time of the proposed method is far lower than the CNN method, its testing time is slightly lower than the CNN method, and its results of prediction and fitting is better than the CNN method. Therefore, the proposed method is higher than CNN, CNN-Iv, and CNN-GRU in efficiency, and the model of proposed method has batter performance.

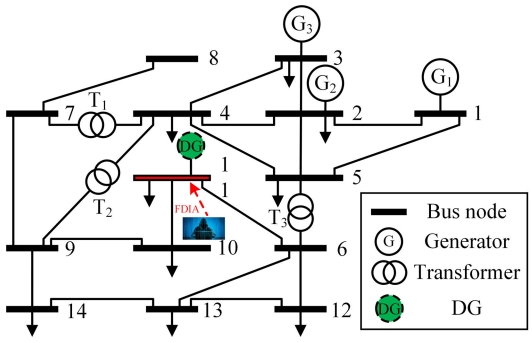


FIGURE 10. The IEEE 14 nodes test system.

B. SIMULATION AND RESULTS ANALYSIS FOR IEEE 14 NODES SYSTEM

The proposed detection method is tested by the IEEE 14 nodes test system [36], and the IEEE 14 nodes test system model is shown in Figure 10.

The FDIA occurred at the 487th sample moment in the 11th node of the test system and lasted 682 sample moments. Similar to [14], the attack intensity of the FDIA is the ratio between the deviation of the estimated state variable amplitude after injecting false data and the actual state data amplitude in the target attack node.

1) EVALUATION METRICS

The common metrics, precision (P), recall (R), accuracy (A), F_1-score (F_1), and false alarm rate (F) are utilized to evaluate performance of FDIA detection methods [18], which are as follows:

$$\begin{aligned} P &= N_{\mathrm{TP}}/(N_{\mathrm{TP}} + N_{\mathrm{FP}}) \times 100\% \\ R &= N_{\mathrm{TP}}/(N_{\mathrm{TP}} + N_{\mathrm{FN}}) \times 100\% \\ A &= (N_{\mathrm{TP}} + N_{\mathrm{TN}})/(N_{\mathrm{TP}} + N_{\mathrm{FP}} + N_{\mathrm{FN}} + N_{\mathrm{TN}}) \times 100\% \\ F &= 2 \times (P \times R)/(P + R) \times 100\% \\ F &= N_{\mathrm{FP}}/(N_{\mathrm{FP}} + N_{\mathrm{TN}}) \times 100\% \end{aligned} \tag{18}$$

where true positive (N_{TP}) and false negative (N_{FN}) denote the number of true FDIA samples predicted as FDIA data and normal data, respectively; false positive (N_{FP}) and true negative (N_{TN}) denote the number of true normal samples predicted as FDIA data and normal data, respectively.

2) PERFORMANCE TESTING OF DETECTION METHODS

This section uses the FDIA with 10% attack intensity as an example to test the performance of the proposed detection method. The detection results are shown in Figure 11.

As shown in Figure 11, the test results of the proposed method did not exceed the set threshold before the FDIA started. The attacks are launched at the 487th sample time. After the FDIA occur, the proposed method can immediately exceed the set threshold at the 488th sample time. Compared with the normal scenario, only after a sample time delay, the proposed method detects the existence of attacks in noise scenario. Therefore, the FDIA can be effectively detected by the proposed detection method.

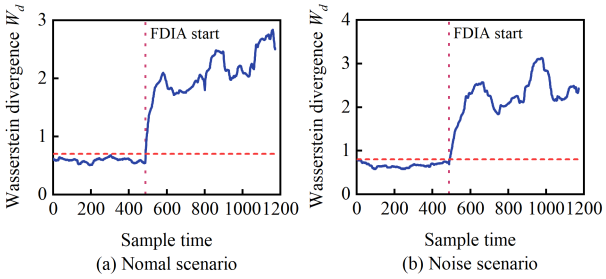


FIGURE 11. Detection results of proposed method.

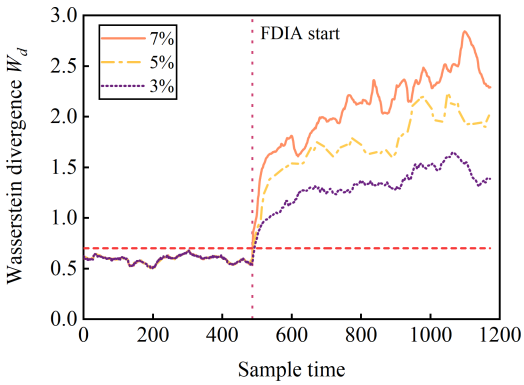


FIGURE 12. Detection results in different attack intensity.

3) INFLUENCE OF ATTACK INTENSITY

There is a negative correlation between attack intensity and detection difficulty, that is, the weaker the attack intensity, the higher the detection difficulty. In order to test detection performance of the proposed method under different attack intensities, the proposed method is tested at attack intensity of 3%, 5% and 7%. The detection results are shown in Figure 12.

As shown in Figure 12, after the FDIA occur, the FDIA of 7%, 5%, and 3% attack intensity, respectively, are detected at 489th, 490th, and 492th sample time by the proposed detection method. It can be seen that as the attack intensity of the FDIA decreases, the sensitivity of the detection method has decreased.

4) INFLUENCE OF NOISES

In order to further test the influence of noise, the Gaussian white noises are added to the measurement data with the signal to noise ratio (SNR) of 50 dB, 40 dB, 30 dB, and 25 dB. In this section, the proposed detection method is tested at the FDIA of 3% attack intensity. The detection results are shown in Figure 13.

As shown in Figure 13, after the FDIA occur, the proposed detection method detects the FDIA in different noises of SNRs. For example, the proposed method exceeds the set threshold at 494th sample time in the noise scenario with

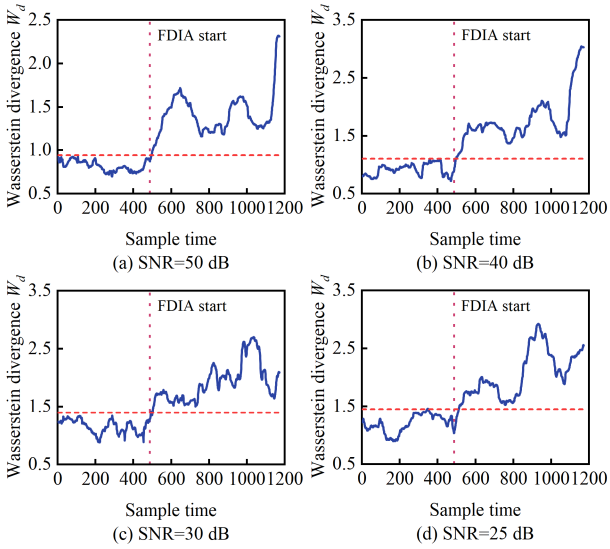


FIGURE 13. Detection results in different SNRs.

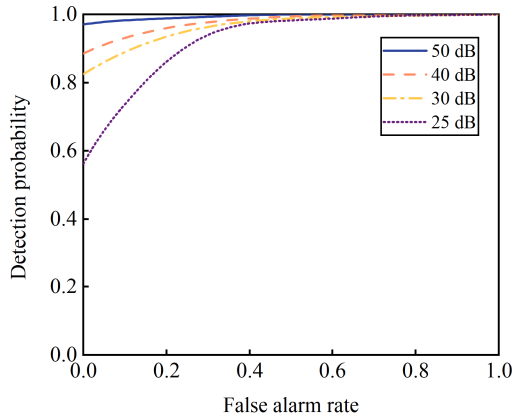


FIGURE 14. Detection performance in different SNRs.

SNR of 50 dB. As the SNR of noise decreases, the FDIA are detected at the 497th, 502th, and 510th sample time, respectively. To further compare the performance of detection methods, the test system suffered 100 simulated attacks of 3% attack intensity in four noise scenarios with SNR of 50 dB, 40 dB, 30 dB, and 25 dB. The detection performance of the proposed detection method is shown by the receiver operating characteristic (ROC) graph, which is shown in Figure 14.

As shown in Figure 14, the change of SNR can affect the detection performance of proposed method. For example, when the false alarm rate is 1%, the detection probability of the proposed method is 97.26%, 89.06%, 83.24%, and 58.38% in the noise scenarios with SNR of 50 dB, 40 dB, 30 dB, and 25 dB. As shown in Figure 13 and Figure 14, as the SNR of noise decreases, the proposed method has a delay in detecting the FDIA, and detection probability of the proposed method is decreased.

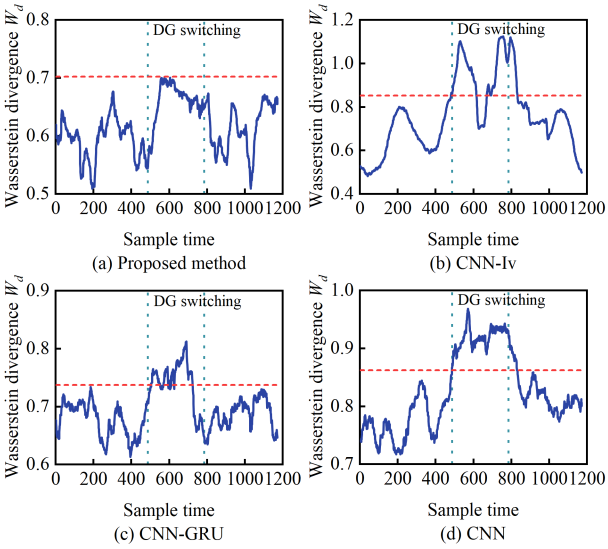


FIGURE 15. Influence of distributed generation switching.

5) INFLUENCE OF SWITCHING EVENTS

The switching of DG can cause fluctuations in system power flow, voltage, and current, which can affect attack detection methods. To test the effects of DG switching, the DG with 3MW capacity is put into the test system at the 11th node between the 487th to the 786th sample time. The test results are shown in Figure 15.

As shown in Figure 15, during DG input and remove test system, the test results of the proposed method did not exceed the set threshold, but the CNN, CNN-Iv, and CNN-GRU have exceed the set threshold at multiple persistent sample time. Therefore, compared with CNN, CNN-Iv, and CNN-GRU methods, the detection performance of proposed method is not affected by DG switching and determine no attack existence.

C. SIMULATION AND RESULTS ANALYSIS FOR IEEE 39 NODES SYSTEM

In order to test the applicability of the proposed detection method in different system models, this section tested the performance of the proposed method in the IEEE 39 nodes test system [37], where the node, the sample time, and the continuous sample time of attacks are same as IEEE 14 nodes test system. The IEEE 39 nodes test system model is shown in Figure 16.

The detection performance of the proposed method at different attack intensity is shown in Figure 17.

As shown in Figure 17, after the FDIA occur, the FDIA of 10%, 7%,5%, and 3% attack intensity are detected at 489th, 490th, 493th, and 494th sample times, respectively. In order to test the performance of the proposed detection method in the IEEE 39 nodes test system, the test system suffered 100 simulated attacks of 3% attack intensity in four noise scenarios with SNR of 50 dB, 40 dB, 30 dB, and 25 dB. The detection results are shown in Figure 18.

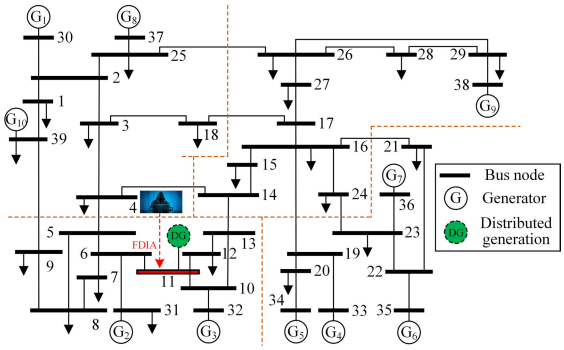


FIGURE 16. The IEEE 39 nodes test system.

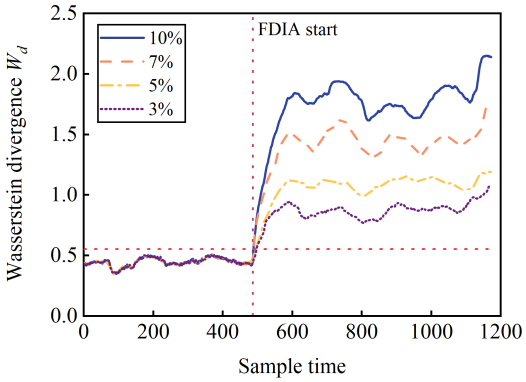


FIGURE 17. Detection results in different attack intensity.

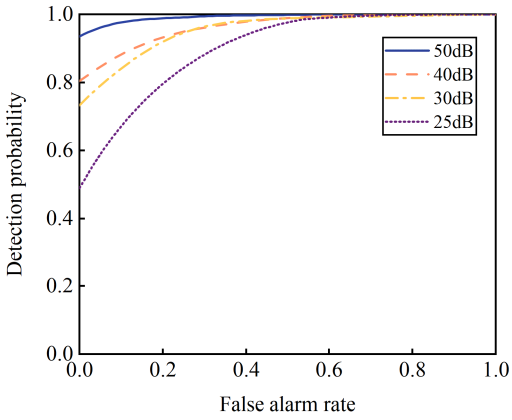


FIGURE 18. Detection results in different SNRs.

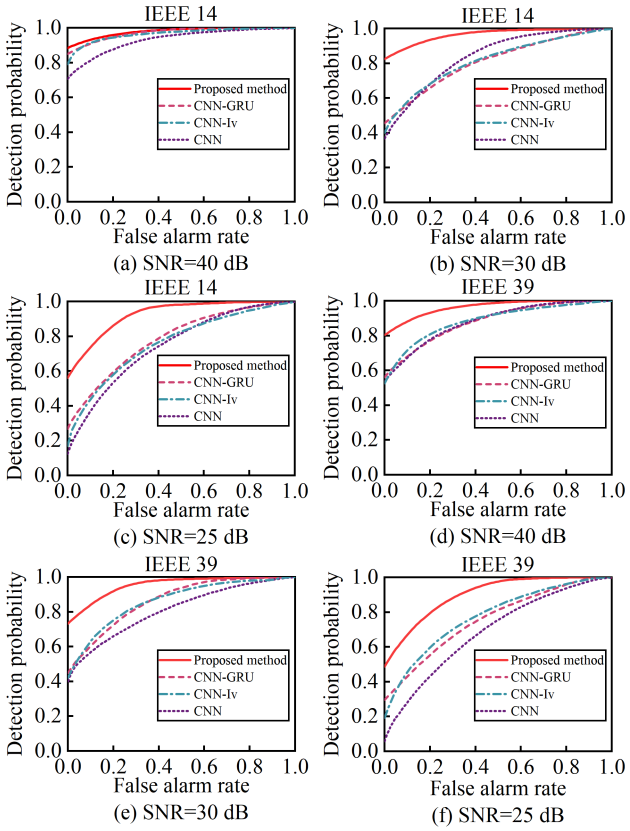


FIGURE 19. ROC comparison of different detection methods.

As shown in Figure 18, the proposed method can effectively detect the existence of FDIA. For example, when the false alarm rate is 1%, the detection probability of the proposed method is 94.16%, 81.29%, 74.52%, and 51.03% in the noise scenarios with SNR of 50 dB, 40 dB, 30 dB, and 25 dB. As shown in Figure 14 and Figure 18, the increase in model complexity can reduce the sensitivity of the proposed method to detect the FDIA.

D. COMPARATIVE ANALYSIS OF DETECTION RESULTS

In noise scenarios with different SNRs, the different detection methods are tested by the 100 simulated attacks of 3% attack intensity. The results of different detection methods are shown in Figure 19.

As shown in Figure 19, the detection probability of the proposed method is higher than the detection probability of CNN-GRU, CNN-Iv, and CNN under the same false alarm rate in different systems. For example, it can be seen from Figure 19 (a) that the detection probability of the proposed method, CNN-GRU, CNN-Iv, and CNN, respectively, is 89.06%, 85.78%, 81.82%, and 72.32% in noise scenario with SNR of 40 dB when the false positive rate is 1%. When the SNR is 30 dB, the detection probability of the proposed method is 83.24% in Figure 19 (b). Compared with CNN-GRU, CNN-Iv, and CNN, the detection probability of the proposed method was improved by 44.09%, 43.17%, and 36.45% at the same false alarm rate, respectively. The proposed method is affected by high noise, and its detection probability is 58.38% in noise scenario with an SNR of 25 dB. However, the detection probability of CNN-GRU, CNN-Iv, and CNN are only 31.02%, 22.08%, and 16.26% at the same false alarm rate, respectively. Therefore, the proposed method has better detection performance in noisy scenarios compared with other methods.

For comparing the performance of different detection methods in the IEEE 14 nodes test system and the NYISO

TABLE 3. Detection performance of different detection methods.

Detection methods	P (%)	R (%)	A (%)	F_1 (%)	F (%)
Proposed method	99.26	97.26	98.38	98.25	1.00
VMD-ELM [14]	–	–	90.00	–	–
MTMCN [38]	99.80	86.40	98.30	92.60	–

dataset, the proposed method, the method combining VMD with ELM (VMD-ELM) [14], and modified temporal multi-graph convolutional network (MTMCN) [38] are compared in this section. The comparison results are shown in Table 3.

As shown in Table 3, the proposed method is better than other detection methods at common metrics of the precision, recall, accuracy, and F_1 score for the detection results of different methods. For example, the accuracy of the proposed method improves by 8.38 and 0.08%, respectively, compared with VMD-ELM and MTMCN. Although the proposed method is slightly lower than MTMCN in precision, the recall, accuracy, and F_1 score of the proposed method is higher than MTMCN. Therefore, the proposed method is better than other methods regarding detection performance.

V. CONCLUSION

In this paper, a FDIA detection method based on deep learning with multi-scale feature fusion is proposed. This method uses the multi-scale convolution kernel of ICNN to extract the temporal and spatial features of measurement data, and the data prediction module predicts the future measurement data of the system. Then, considering the importance of the partial-scale information in measurement data, the attention mechanism is introduced into the ICNN to structure the state generation module, and which is used to extract the full-scale and the partial-scale features and fuse them. State generation module is used to generate the state data by fitting the measurement function between measurement and state vectors. Finally, using WD, the divergences of probability distribution between predicted and actual state vectors are calculated to decide whether the presence of the FDIA.

The performance of the proposed method was tested in the IEEE 14 nodes test system and IEEE 39 nodes test system. The results of simulating showed the proposed method can accurately detect existence of the FDIA in time without knowing the topology information and parameters of the system. The precision, recall, and F_1 score of the proposed method is 99.26%, 97.26%, and 98.25% in noise scenario with SNR of 50 dB, respectively. In addition, this method also has definite robustness for noises and distributed generation switching.

Many technical aspects deserve deepening and expanding in future studies. As the DG continues to be introduced into the power grid, the FDIA detection in large systems and complex working conditions with a high penetration rate of DG should be further researched. Moreover, the problems of the FDIA location and false data removal will also be part of future research.

REFERENCES

[1] Q. Wang, W. Tai, Y. Tang, and M. Ni, "Review of the false data injection attack against the cyber-physical power system," *IET Cyber-Physical Systems: Theory Appl.*, vol. 4, no. 2, pp. 101–107, Jun. 2019.

[2] H. Zhang, B. Liu, and H. Wu, "Smart grid cyber-physical attack and defense: A review," *IEEE Access*, vol. 9, pp. 29641–29659, 2021.

[3] A. Takiddin, R. Atat, M. Ismail, O. Boyaci, K. R. Davis, and E. Serpedin, "Generalized graph neural network-based detection of false data injection attacks in smart grids," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 3, pp. 618–630, Aug. 2023.

[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–33, May 2011.

[5] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2239–2248, Sep. 2017.

[6] X.-J. Ma and H. Wang, "Blind false data injection attacks in smart grids subject to measurement outliers," *J. Control Decis.*, vol. 9, no. 4, pp. 445–454, Oct. 2022.

[7] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2218–2234, May 2020.

[8] A. Parizad and C. J. Hatziadoniu, "A real-time multistage false data detection method based on deep learning and semisupervised scoring algorithms," *IEEE Syst. J.*, vol. 17, no. 2, pp. 1753–1764, Jun. 2023.

[9] B. Xie, C. Peng, M. Yang, X. Kong, and T. Zhang, "A novel trust-based false data detection method for power systems under false data injection attacks," *J. Franklin Inst.*, vol. 358, no. 1, pp. 56–73, Jan. 2021.

[10] I. Lukicheva, D. Pozo, and A. Kulikov, "Cyberattack detection in intelligent grids using non-linear filtering," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur. (ISGT-Europe)*, Oct. 2018, pp. 1–6.

[11] A. S. L. V. Tummala and R. K. Inapakurthi, "A two-stage Kalman filter for cyber-attack detection in automatic generation control system," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 1, pp. 50–59, Jan. 2022.

[12] D. Lin, Q. Zhang, X. Chen, J. Qian, W. Yan, and S. Wang, "Quaternion Kalman filter for false data injection attacks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 71, no. 3, pp. 1501–1505, Mar. 2024.

[13] Y. Zhang, J. Wang, and B. Chen, "Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 623–634, Jan. 2021.

[14] C. Dou, D. Wu, D. Yue, B. Jin, and S. Xu, "A hybrid method for false data injection attack detection in smart grid based on variational mode decomposition and OS-ELM," *CSEE J. Power Energy Syst.*, vol. 8, no. 6, pp. 1697–1707, Nov. 2022.

[15] F. Zhang and Q. Yang, "False data injection attack detection in dynamic power grid: A recurrent neural network-based method," *Frontiers Energy Res.*, vol. 10, Sep. 2022, Art. no. 1005660.

[16] A. Bhattacharjee, A. K. Mondal, A. Verma, S. Mishra, and T. K. Saha, "Deep latent space clustering for detection of stealthy false data injection attacks against AC state estimation in power systems," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2338–2351, May 2023.

[17] W. Lei, Z. Pang, H. Wen, W. Hou, and W. Han, "FDI attack detection at the edge of smart grids based on classification of predicted residuals," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9302–9311, Dec. 2022.

[18] Y. Wang, Z. Zhang, J. Ma, and Q. Jin, "KFRNN: An effective false data injection attack detection in smart grid based on Kalman filter and recurrent neural network," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6893–6904, May 2022.

[19] H. T. Reda, A. Anwar, A. Mahmood, and N. Chilamkurti, "Data-driven approach for state prediction and detection of false data injection attacks in smart grid," *J. Modern Power Syst. Clean Energy*, vol. 11, no. 2, pp. 455–467, Mar. 2023.

[20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[21] C. Kyunghyun, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. EMNLP*, 2014, pp. 1724–1734.
[22] O. Boyaci, M. R. Narimani, K. R. Davis, M. Ismail, T. J. Overbye, and E. Serpedin, "Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 807–819, Jan. 2022.
[23] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[25] S. He, P. E. Grant, and Y. Ou, "Global–local transformer for brain age estimation," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 213–224, Jan. 2022.
[26] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, Aug. 1987.
[27] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[28] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE-ICCV*, 2019, pp. 510–519.
[29] J. G. Moreno-Torres, J. A. Saez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, Aug. 2012.
[30] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Proc. ICCV*, Jan. 1998, pp. 59–66.
[31] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. ICML*, vol. 1, 2017, pp. 298–321.
[32] J. Harmouche, C. Harmouche, and D. Harmouche, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part I," *Singal Process*, vol. 94, pp. 278–287, Jan. 2014.
[33] L. Cui, Y. Liu, L. Wang, J. Chen, and X. Zhang, "High-impedance fault detection method based on sparse data divergence discrimination in distribution networks," *Electric Power Syst. Res.*, vol. 223, Oct. 2023, Art. no. 109514.
[34] (2023). *Real-Time Actual Load Data Reports*. [Online]. Available: https://www.nyiso.com/energy-market-operational-data
[35] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
[36] (2014). *IEEE 14 Bus Power Flow Test Case*. [Online]. Available: https://labs.ece.uw.edu/pstca/pf14/pg_tca14bus.htm
[37] G. W. Bills, "On-line stability analysis study, RP 90–1," North American Rockwell Inf. Syst. Company, Anaheim, CA, USA, Tech. Rep. NP-2901022; ON: DE82901022, 1970. [Online]. Available: https://www.osti.gov/biblio/5984031
[38] Y. Han, H. Feng, K. Li, and Q. Zhao, "False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids," *Comput. Secur.*, vol. 124, Jan. 2023, Art. no. 103016.

JINPENG JI received the B.E. degree in smart grid information engineering from Shandong University of Technology, Zibo, China, in 2021, where he is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering. His research interest includes detection of false data injection attack in smart grid.

YANG LIU (Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2016. He is currently an Associate Professor with the College of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, Shandong. His work has addressed power system protection and control, situational awareness, and the integration of renewable resources.

JIAN CHEN is currently a Senior Engineer with the Institute of Metrology and Technology, Zibo, Shandong. Her research interests include the verification and calibration of measuring instruments, such as DC resistors, AC and DC currents, voltages, and electrical energy.

ZHIWEI YAO received the B.E. degree in smart grid information engineering from Shandong University of Technology, Zibo, China, in 2021, where he is currently pursuing the M.E. degree with the School of Electrical and Electronic Engineering. His research interest includes active fault detection in distribution networks.

MENGDI ZHANG received the B.E. degree in measurement and control technology and instrument from Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, in 2020. He is currently pursuing the M.E. degree with the School of Electrical and Electronic Engineering, Shandong University of Technology. His research interest includes monitoring the condition of energy storage systems.

YANYONG GONG received the B.E. degree in agricultural electrification from Shenyang Agricultural University, Shenyang, China, in 2022. He is currently pursuing the M.E. degree with the School of Electrical and Electronic Engineering, Shandong University of Technology. His research interest includes fault location in distribution networks.