Multimodal Machine Learning Based Equivalent Modeling Method For Active Distribution Networks

Wenhao Wang, Jiehui Zheng, *Member, IEEE*, Zhaoxi Liu, *Senior Member, IEEE*, Wei Yao, *Senior Member, IEEE*, Yuekuan Zhou, Zhigang Li, *Senior Member, IEEE*, Qinhua Wu, *Life Fellow, IEEE*

***Abstract*—With high integration of distributed generators (DGs) and diverse topology reconfigurations, the dynamic states of active distribution networks (ADNs) become more complex, which poses great challenges to the dynamic equivalence of ADNs. In this paper, motivated by the similarities between Newton Raphson power flow calculation (NRPFC) and computation of convolution network (CNN), and based on the capability of recurrent neural network (RNN) to represent the differential algebraic equations (DAEs) of loads, we propose a multimodal machine learning based equivalent modeling method in order to track the dynamic behaviors of ADNs. First, the multimodal machine learning is divided into two modules, one of which is gated recurrent unit (GRU) + fully connected (FC) module to represent the DAEs for load modeling, and the other is CNN + attention module to simulate the NRPFC to extract spatial feature changes of power system caused by disturbances or topology reconfigurations. Then, the robustness and generalization abilities of the proposed method are evaluated by different test systems, where the IEEE 14-bus transmission network is connected to distribution networks with different scales (i.e., IEEE 33-bus distribution network and IEEE 57-bus distribution network). The simulation results reveal that the proposed method can accurately capture the dynamic behaviors of ADNs under different operation conditions. In addition, the precision and generalization of the proposed method is tested in comparison with other deep learning (DL)-based equivalent methods.**

***Index Terms*—Multimodal machine learning, recurrent neural network, load modeling, active distribution network, topology reconfiguration, dynamic equivalence, gated recurrent unit.**

Manuscript received: March 29, 2025; revised: June 5, 2025; accepted: August 11, 2025. Date of CrossCheck: August 11, 2025. Date of online publication: September 18, 2025.

This work was supported by State Key Laboratory of HVDC (No. SKLHVDC-2024-KF-11).

W. Wang, J. Zheng (corresponding author), Z. Liu, and Q. Wu are with the School of Electric power Engineering, South China University of Technology, Guangzhou, China (e-mail: 202410183143@mail.scut.edu.cn; zhengjh@scut.edu.cn; liuzhaoxi@scut.edu.cn; wuqh@scut.edu.cn).

Z. Li is with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China (e-mail: lizg@cuhk.edu.cn).

W. Yao is with the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China (e-mail: w.yao@hust.edu.cn).

Y. Zhou is with the Sustainable Energy and Environment Thrust, Function Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China (e-mail: yuekuanzhou@ust.hk).

DOI: 10.35833/MPCE.2025.000281

I. INTRODUCTION

THE dynamic equivalence of active distribution networks (ADNs) has noticeable effects on simulating and analyzing the dynamics of power system [1], [2]. Moreover, due to the integration of large-scale renewable energy and distributed generators (DGs) in ADNs, the traditional distribution network has changed from a passive one to active one. As a consequence, the power interaction between the ADNs and backbone network becomes more complicated, which poses great threats to the security and operation of ADNs [3].

The dynamic equivalent task of ADNs first determines the equivalent structure, and then identifies the parameters of the constructed equivalent structure. For the equivalent structure of ADNs, the load modeling is the most important part, because it represents the dynamic characteristics of loads and renewable energy generation systems. Load modeling is principally used by backbone networks for their planning, analysis, and control [4], [5]. If the load model does not accurately reflect the dynamic characteristics of ADNs, it will draw incorrect conclusions in dynamic security analysis of power system. Therefore, how to develop an accurate, adaptive, and robust load modeling method for ADNs is an urgent issue, which is needed to be addressed for power system studies.

Nowadays, there are two traditional load modeling methods, which are component-based methods and measurement-based methods. The basic idea of component-based methods is to regard the load as a collection of individual users, then count the proportion of various electrical appliances, and finally synthesize the overall load model [6], [7]. The component-based methods are not only time-consuming and laborious, but also subject to time-varying loads. Therefore, it is impossible to perform these methods at any time. Different from component-based methods, the measurement-based methods regard the load as a whole: they first choose the structure of load model, and then identify the model parameters based on the data collected from nodes. Therefore, the measurement-based methods are widely applied in power system for practical applications because of their better performance in presentation of dynamic characteristics [8]-[13].

Once the equivalent structure of ADNs is determined, the equivalent problem is transformed into a parameter identification problem. Traditional parameter identification methods include mathematical optimization methods and modern evolutionary algorithms. Mathematical optimization methods

[10]-[12] such as least squares and steepest descent have the advantages of fast solving speed and reliable convergence, but they are prone to getting stuck in local optima, especially when the initial value is far from the actual value. Modern evolutionary algorithms [13] have poor robustness, slow convergence speed, and difficulty in setting learning parameters. In addition, they contain many parameters, so their solving process often takes a long time, making it difficult to meet the requirements of online parameter identification for time-varying loads. Mathematical optimization methods generally use simplified models to meet real-time requirements [2], [14], but the accuracy of parameter identification might be affected. The deep learning (DL)-based equivalent methods use typical simplified models with stronger ability to describe dynamic characteristics compared with other methods. After training process, the parameter identification (i.e., testing process) can be performed in a very short time, but the training time is relatively long. Obviously, compared with mathematical optimization methods, the DL-based equivalent methods could better balance the requirements of accuracy and online equivalence.

Meanwhile, in recent years, DL, as an innovative technology, has been widely developed in various fields for feature extraction or data prediction [15], [16]. In the area of power systems, many studies in the application of DL to wind speed forecasting [17], [18], renewable energy prediction [19], fault diagnosis [20], energy management [21], and dynamic security assessment [22] have been carried out. Accordingly, some researchers have suggested employing neural networks (NNs) to model the dynamic characteristics of ADNs in recent years. They regard voltages or frequencies as inputs and find the mapping relationship through NNs to output the active and reactive power [23]-[28]. Reference [23] employs the feed-forward artificial NN (ANN) as the prescribed function for load modeling, and the real load data are used to assess the performance of the constructed load model. Reference [24] applies the NN-based dynamic load identification technique, which utilizes delayed system inputs and outputs to represent the dynamic loads in power system. The use of ANNs with feedback loops for load modeling is investigated in [25], and the simulation results show that the ANN can accurately capture the active and reactive power responses. Reference [26] points out that the radial basis function NN combined with the lookup table can represent the dynamic responses of electrical arc furnace. However, the above simple NN-based methods without memory unit may not accurately capture the long-term impact of disturbance, that is, the disturbance occurring at time t will affect the power response at time $t+\tau$. In addition, the cases with large number of DGs and electric vehicles (EVs) in ADNs have not been taken into consideration. To this end, the recurrent NNs (RNNs) are employed for representing dynamic behaviors of ADNs.

Reference [27] proposes a measurement-based method using RNN to construct an accurate equivalent model for ADNs, and the testing results show that this model can be employed for representing the whole ADN. However, RNN is prone to gradient explosion and gradient disappearance during training, because the gradient cannot be transmitted in a long sequence during training. So RNN cannot capture the correlation of long time-series data well. To solve these issues, the derived algorithms, i.e., long short-term memory (LSTM) [29] and gated recurrent unit (GRU) [30], have been proposed. Based on the similarities between differential algebraic equations (DAEs) of power system and the forward calculation flows of RNNs, [28] proposes an LSTM-RNN-based load modeling method. The generalization and accuracy of this method are certified by using simulations. Nevertheless, compared with GRU, LSTM has a more complex structure and contains more parameters so that it may require larger datasets and more computation resources for training. Although GRU simplifies its model, resulting in slightly inferior performance compared with LSTM in handling tasks that require high precision to maintain long-term memory, it can achieve a comparable performance with LSTM in many other tasks. In addition, due to its simple structure, it can be easily extended to other learning methods without increasing the complexity of the network significantly.

All of the above literature [23]-[28] about DL-based equivalent methods does not consider the time-variant characteristic of ADNs caused by the disturbances (i.e., disturbance locations and types) or topology reconfigurations (i.e., line outrage and line reorganization). Taking an IEEE 33-bus distribution network as example, when the line reorganization occurs, the system topology will change greatly, leading to changes in the dynamic response of ADNs. Moreover, when the three-phase (ABC) short circuit occurs, the load connected to fault line will be cut off and the network topology will change greatly, which also lead to great shifts in dynamic responses. However, these physical changes, as the source of response changes of ADNs after disturbance, cannot be tracked well using existing DL-based equivalent methods, which only consider simple input characteristics to find the mapping relationship between voltage/frequency and active/reactive power, and overlook the spatial change, leading to poor generalization. Meanwhile, these equivalent methods cannot consider these dynamic topology changes due to their single module representation, which further affects their robustness and generalization ability.

Newton-Raphson power flow calculation (NRPFC) calculates the power correction value through the multiplication of Jacobian matrix and the residual vector of set state value, and then continuously corrects the power flow through iteration. When the corrected state value is less than the set threshold, the convergence result is achieved. This iterative method is consistent with the NN in reducing the loss value and propagating the loss value backward in the training process. At the same time, the convolution calculation of two-dimensional convolution neural network (CNN) can also be viewed as the multiplication of standard matrix and vector, because the convolution calculation is essentially a special matrix multiplication calculation, and the convolution of CNN can be calculated in the form of vectors. Therefore, based on the similarity between NRPFC and convolution computation of CNN in calculation and iteration form, we utilize CNN to simulate the NRPFC to capture the dynamic topology changes caused by topology reconfigurations or disturbances, and apply stacked attention block to extract important features. Thus, the CNN + attention module can perceive the dynamic behaviors of ADNs from the prospective of power flow calculation.

Multimodal machine learning is a subfield of artificial intelligence, which attempts to improve the learning ability by employing a large number of data in different forms of expression (homogeneous or heterogeneous) in training. This enables the optimization model to learn new patterns and correlations between different forms of data, so as to improve the accuracy and interpretation ability. Recently, multimodal machine learning has been applied in several fields such as image and natural language processing [31], [32], biomedicine [33], and sentiment analysis [34], and has achieved great success. However, in the field of power system, the application of multimodal machine learning is rarely studied [35].

The dynamic behaviors of ADNs are influenced by both disturbances and topology reconfigurations. If we only consider the relationships between the voltage of boundary bus and dynamic response of ADNs, the online application capability of DL-based equivalent method will be greatly reduced. Therefore, we need to track the impact of topology reconfiguration on the dynamic behaviors of ADNs. Because the dynamic behaviors can be obtained by load modeling and the dynamic topology changes can be captured by CNN, we merge the load modeling module (i.e., GRU + FC module) and feature extraction module (i. e., CNN + attention module) as multimodal machine learning to track the dynamic behaviors of ADNs under different disturbances and topology reconfigurations. Due to the merits of multimodal machine learning, the multimodal machine learning based equivalent modeling method can consider both of the dynamic and static characteristics of the load and the topology changes of power grid caused by disturbances, thereby better tracking the dynamic behaviors of ADNs in different situations.

Within this context, the key contributions of this paper are listed as follows.

1) This paper proposes a CNN to simulate NRPFC process and applies an attention module to extract important features caused by disturbances or topology reconfigurations.

2) The feature extraction module and load modeling module are fused into a multi-modal machine learning based equivalent modeling method to track the dynamic behaviors of ADNs. It is capable of tracking time-variant characteristics of ADNs under different disturbances and topology reconfigurations, and the case studies indicate that the proposed method outperforms other DL-based equivalent methods.

3) To better extract the power grid characteristics, based on the similarity between CNN and NRPFC in calculation form and iteration form, this paper employs the Jacobian matrix of power grid to form the spatial information of power grid, which serves as inputs to the CNN to help feature extraction module better perceive changes in network topology. Then, the feature extraction strategy is proposed for extracting the dynamic responses of ADNs caused by disturbances and topology reconfigurations.

The remainder of this paper is organized as follows. Section II discusses some important concepts and basic design principles of the proposed method. Then, the numerical case studies for evaluating the abilities of the generalization and robustness of the proposed method are presented in Section III. Finally, conclusions are drawn in Section IV.

II. Some Important Concepts and Basic Design Principles of Proposed Method

In this section, first, we describe some important concepts of CNN, multi-head attention (MHA), and GRU. Then, the motivation of adopting multi-modal machine learning for load modeling is explained and discussed in detail. Finally, the basic design principles using CNN + attention module as feature extraction module and GRU + FC module as load modeling module are introduced.

A. CNN Layer

As a strong feature extraction structure, CNN has drawn great attention of researchers in recent years [36]. The main structure of CNN is shown in Fig. 1. The core of the CNN is the convolutional layer with multiple filters and feature maps, which reduces the network complexity and number of parameters. In this layer, the characteristics of the input data are revealed, which can be expressed as:

$$h_{ij}^{l}=f\left(\left(\boldsymbol{W}^{l}*\boldsymbol{x}\right)_{ij}+\boldsymbol{b}^{l}\right) \tag{1}$$

where $f(\cdot)$ is the activation function; $\boldsymbol{x}$ is the input vector; l is the index of feature map; h_{ij}^{l} is the output of convolutional layer; $*$ is the convolution operation; and $\boldsymbol{W}^{l}$ and $\boldsymbol{b}^{l}$ are the weight matrix and bias vector of the kernel to feature map, respectively.

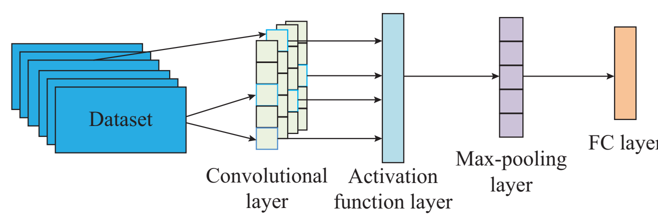

Fig. 1. Main structure of CNN.

B. MHA Layer

MHA extends the attention mechanism to multiple heads, thus to enhance the attention of model to different features [37]. The input of MHA includes three vectors: query vector ($\boldsymbol{Q}$), key vector ($\boldsymbol{K}$), and value vector ($\boldsymbol{V}$). For a given query vector, MHA computes the weighted sum of the key vector, where the weight is calculated using the similarity between the query vector and key vector, and then the weighted sum is multiplied by the value vector to produce the output [38]. The calculation of MHA can be expressed as:

$$MHA(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Concat\left(h_1,h_2,\dots,h_M\right)\cdot\boldsymbol{W}^{O} \tag{2}$$

where M is the number of headers; $Concat(\cdot)$ is the vector concatenation operation; $\boldsymbol{W}^{O}$ is the output transformation matrix; and h_i is the output of the i^{th} header, which can be expressed as:

$$h_i=\tilde{A}\left(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},\boldsymbol{V}\boldsymbol{W}_i^{V}\right) \tag{3}$$

where $\boldsymbol{W}_i^Q$, $\boldsymbol{W}_i^K$, and $\boldsymbol{W}_i^V$ are the query, key, and value transformation matrices of the i^{th} header, respectively; and $\tilde{A}(\cdot)$ is the attention calculation function. Generally, the self-attention mechanism is applied to attention calculation function:

$$\tilde{A}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{4}$$

where d_k is the dimension of attention head; and $softmax(\cdot)$ is the softmax activation function.

MHA enables the model to learn more diverse and complex features, effectively improving its expressive power.

C. GRU Layer

The LSTM stores useful information in memory cells and eliminates unnecessary information to achieve better performance than classical RNNs in many fields such as machine translation and load forecast. However, due to its complex structure, LSTM may lead to long training time and overfitting problem when it is overlapped with other modules. Therefore, with the similar structure to LSTM, GRU is proposed to overcome the above issues of LSTM [30]. The GRU network contains two gates, i.e., update gate and reset gate, the outputs of which are defined as (5) and (6), respectively.

$$\boldsymbol{z}_t=\sigma\left(\boldsymbol{W}^{(z)}\boldsymbol{x}_t+\boldsymbol{U}^{(z)}\boldsymbol{h}_{t-1}\right) \tag{5}$$

$$\boldsymbol{r}_t=\sigma\left(\boldsymbol{W}^{(r)}\boldsymbol{x}_t+\boldsymbol{U}^{(r)}\boldsymbol{h}_{t-1}\right) \tag{6}$$

where t is the index of time step; $\sigma(\cdot)$ is the activation function; $\boldsymbol{x}_t$ and $\boldsymbol{h}_{t-1}$ are the vectors of inputs and hidden states, respectively; and $\boldsymbol{W}^{(z)}$, $\boldsymbol{U}^{(z)}$ and $\boldsymbol{W}^{(r)}$, $\boldsymbol{U}^{(r)}$ are the weight matrices of reset gate and update gate, respectively.

Therefore, the final memory of current time step is:

$$\boldsymbol{h}_t'=\tanh\left(\boldsymbol{W}\boldsymbol{x}_t+\boldsymbol{r}_t\odot\boldsymbol{U}\boldsymbol{h}_{t-1}\right) \tag{7}$$

$$\boldsymbol{h}_t=\boldsymbol{z}_t\odot\boldsymbol{h}_{t-1}+\left(1-\boldsymbol{z}_t\right)\odot\boldsymbol{h}_t' \tag{8}$$

where $\odot$ is the vector element multiplication operation; $\boldsymbol{W}$ and $\boldsymbol{U}$ are the weight matrices; and $\boldsymbol{h}_t'$ is the candidate state.

The basic architecture of GRU block is shown in Fig. 2. The detailed explanation can be seen in [30].

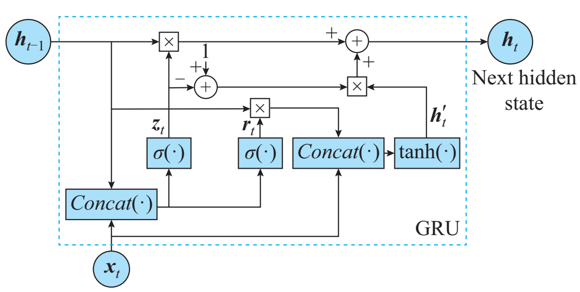

Fig. 2. Basic architecture of GRU block.

D. Basic Design Principles of Proposed Method

The basic elements of the proposed method have been introducted in Section II-A-C and the feasibility of the proposed method is explained in Supplementary Material A. All of these have made sufficient preparations for designing a suitable NN for constructing the equivalent model of ADNs under different conditions.

Due to the fact that differential equations (DEs) and algebraic equations (AEs) could represent the dynamic loads (motor machine and wind turbine (WT)) and static loads (lights and air-conditioning), respectively, it is necessary to increase the number of GRU layers when there are a large number of dynamic elements, and the same applies to FC layers. Therefore, based on [28] and the proportion of dynamic load and static load in the simulation system presented in this paper, one GRU layer, two FC layers, and one CNN + attention layer are employed for constructing an equivalent model for ADN, as illustrated in Fig. 3, where the GRU layer contain 64 blocks, and two FC layers both contain 64 neurons.

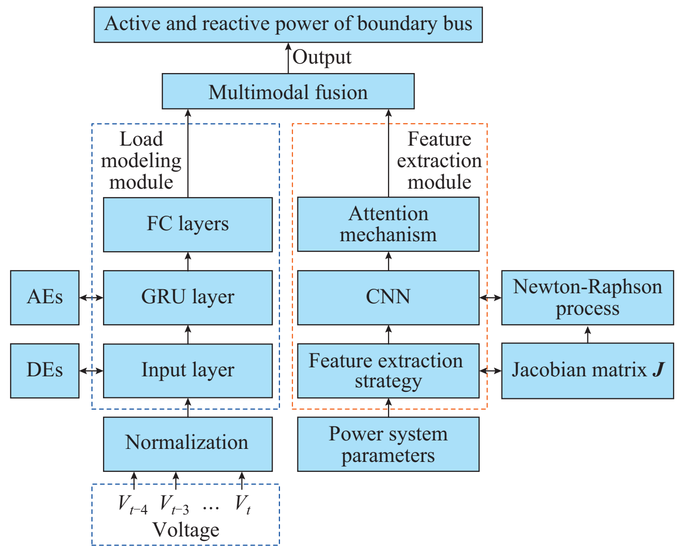

Fig. 3. Framework of equivalent model for ADN.

1) Multimodal Fusion Method

The modal fusion method is very important for multimodal learning. If the modal fusion method is relatively simple and focuses on fusing at the global feature level through cascading operations, it cannot fully learn the connections and representation methods between various modules. Therefore, we adopt a multimodal fusion method called cross-attention (co-attention) fusion method to improve the modal perception ability of equivalent model, thereby enhancing the robustness of equivalent model. The structure of the co-attention fusion method is shown in Fig. 4.

The co-attention fusion method contains two multi-head co-attention blocks, the detail information of which can be seen in [39]. The output $\boldsymbol{F}_{\mathrm{L}}$ of the load modeling module and the output $\boldsymbol{F}_{\mathrm{P}}$ of the feature extraction module are embedded and fed into co-attention blocks for complementary information fusion, and then a mixed modal representation $\boldsymbol{F}_{\mathrm{C}}$ is generated by the feed-forward layer.

The co-attention fusion method realizes the interaction between the load modeling module and feature extraction module, establishes the temporal and spatial relationships between the modeling part and power system network features, and utilizes the complementary advantages of the two data modes to achieve a more powerful and comprehensive modal representation to track dynamic behaviors of ADNs.

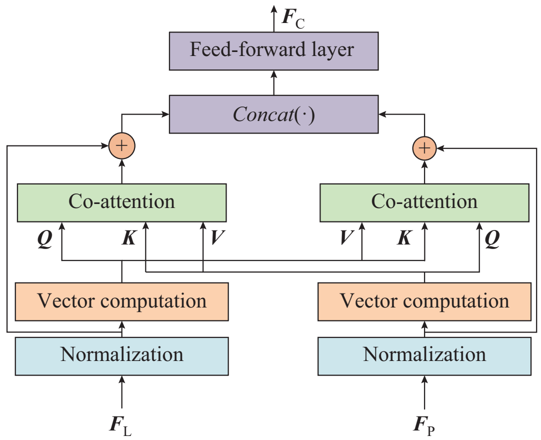

Fig. 4. Structure of co-attention fusion method.

2) Inputs of Equivalent Model

According to the definition of non-mechanism load modeling in power system, the voltage of boundary bus in the T-step time window $\mathcal{X}=\left[\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_T\right]\in\mathbb{R}^{T\times 2}$ is selected as the input of GRU + FC module, where $\boldsymbol{X}_t=[V_t,\delta_t]$, and V_t and δ_t are the voltage amplitude and phase angle at time t, respectively. In our experimental test, the bus voltage vector within the five-step time window is selected as the input of GRU + FC module. The parameters of power system at time t, such as line conductance and susceptance, are selected as the input of CNN + attention module.

Meanwhile, the convolutional computation of CNN is similar to that of NRPFC. Therefore, the Jacobian matrix of power system is employed as inputs to CNN + attention module. Specifically, for a power system with n nodes, assuming the network contains $m-1$ PQ nodes and $n-m$ PV nodes, the Jacobian matrices corresponding to which, i. e., $\boldsymbol{J}_{PQ}\in\mathbb{R}^{2(m-1)\times 2(n-1)}$ and $\boldsymbol{J}_{PV}\in\mathbb{R}^{2(n-m)\times 2(n-1)}$, respectively, are divided into blocks as:

$$\begin{cases}\boldsymbol{J}_{PQ}=\begin{bmatrix}\boldsymbol{H}_{ij} & \boldsymbol{N}_{ij}\\ \boldsymbol{J}_{ij} & \boldsymbol{L}_{ij}\end{bmatrix}\\ \boldsymbol{J}_{PV}=\begin{bmatrix}\boldsymbol{H}_{ij} & \boldsymbol{N}_{ij}\\ \boldsymbol{R}_{ij} & \boldsymbol{S}_{ij}\end{bmatrix}\end{cases} \tag{9}$$

where $\boldsymbol{H}_{ij}$, $\boldsymbol{N}_{ij}$, $\boldsymbol{J}_{ij}$, $\boldsymbol{L}_{ij}$, $\boldsymbol{R}_{ij}$, and $\boldsymbol{S}_{ij}$ are the partial derivatives of voltage for each variable to be solved. Combining $\boldsymbol{J}_{PQ}$ and $\boldsymbol{J}_{PV}$ yields the Jacobian matrix of power system $\boldsymbol{J}\in\mathbb{R}^{2(n-1)\times 2(n-1)}$.

The framework of feature extraction process is shown in Fig. 5. First, the measurement vector of power system can be gathered through limited supervisory control and data acquisition (SCADA) systems, phasor measurement units (PMUs), and smart meters. Then, the power system state (i. e., voltage amplitude and phase angle) can be calculated via state estimation. Next, the Jacobian matrix of power system can be calculated according to the estimated state, topology structure, and related physical parameters. Finally, the Jacobian matrix is input into the CNN for feature extraction. It is worth noting that after the system undergoes topology reconfiguration or disturbance, the elements of relevant Jacobian matrix will also change. Therefore, when the topology reconfiguration or disturbance is detected, the Jacobian matrix is also modified, thus achieving the ability of feature tracking. In addition, it should be noted that when the equivalence starts, we will detect whether the network topology has changed, and when a disturbance occurs, we will locate its location, so as to update the input of CNN. After the disturbance is removed, the network topology will return to the original state before the failure.

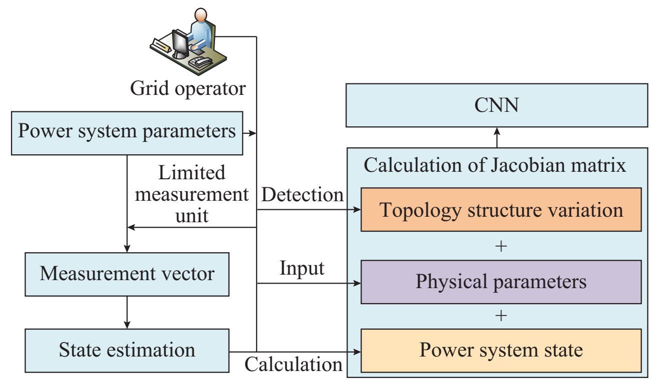

Fig. 5. Framework of feature extraction process.

3) Outputs of Equivalent Model

The active power P_t and reactive power Q_t of the boundary bus are selected as the output of equivalent model, which is denoted as $\mathcal{Y}=\left[\boldsymbol{Y}_1,\boldsymbol{Y}_2,\ldots,\boldsymbol{Y}_T\right]\in\mathbb{R}^{T\times 2}$, and $\boldsymbol{Y}_t=[P_t,Q_t]$.

Furthermore, in order to overcome the shortcomings of gradient disappearance or gradient explosion caused by the input values beyond the limits of nonlinear activation function employed by GRU + FC module, we design all the bus voltage inputs within the range of $(0,1)$ before they are fed to the input layer of GRU + FC module. And the parameters of power system also must be normalized for the same reason, while there is no limitation on the output of equivalent model.

III. Numerical Case Studies

In this section, the performance of the proposed method will be comprehensively evaluated by employing various hypothetical measurement data such as PMU data collected by time-domain simulations of the test system. The proposed method is compared with other DL-based equivalent methods. The framework of the proposed method is built by utilizing Tensorflow 2.2.0 and Keras 2.3.1 in Python 3.9 with a personal computer equipped with 1.80 GHz Intel Core i7-8565U CPU and 8 GB RAM.

A. Introduction for Test Systems

The topology of test system 1 is shown in Fig. 6, which integrates a modified IEEE 33-bus distribution network to IEEE 14-bus transmission network, through node 15 in distribution network and node 16 in transmission network. The frequency of test system 1 is 50 Hz, and the time domain step is 50 μs. The electrical parameters of distribution network are available in MATPOWER 6.0. Furthermore, the entire test system is simulated in Simulink. For example, the DGs are modeled by the specifically designed photovoltaic (PV), WT, and gas turbine (GT) blocks, and static loads are modeled by three-phase series RLC loads, and RLC loads can be designed as R, L, C, RL, RC, and LC loads based on the real demand.

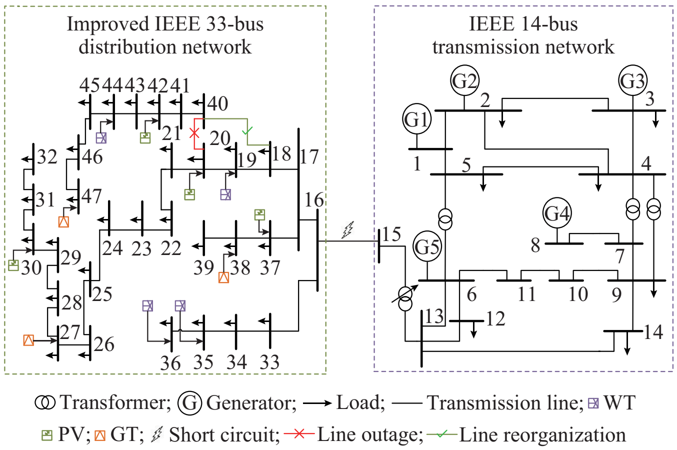

Fig. 6. Topology of test system 1.

To verify the robustness and generalization of the proposed method, a test system with more buses, DGs, and renewable energy sources is needed. To this end, the test system 2 is introduced, which integrates an IEEE 57-bus distribution network with more DGs and renewable energy sources to IEEE 14-bus transmission network, through node 13 in transmission network and node 1 in distribution network. The topology can be found in MATPOWER 6.0.

B. Performance Evaluation of Proposed Method at Different Disturbance Locations or Under Topology Reconfigurations

In this subsection, we select the ABC short circuit as the disturbance type, due to its serious impact on power system, to evaluated the performance of the proposed method. The security and stability analysis of power system is also essential to track the dynamic behaviors of ADN.

The simulation data include 24 pairs of active and reactive power responses. The 24 pairs of power responses are from the disturbances occuring on lines between different nodes within the distribution network in the test system 1: 8 pairs do not have line outages or topology reconfigurations (e.g., Do-14-15), while other 16 pairs do (e.g., Do-15-16-LO-33-34 and Do-16-17-LR-17-33), as shown in Table I.

Based on basic rule for dividing training set and testing set of DL, the proportion of training set and testing set is 3:1, with 18 pairs of simulation data for training set and 6 pairs for testing set.

The root mean square percentage error (RMSPE) is adopted to calculate the error between the calculated value of the proposed method and measurement value of simulation results. In addition, the relative absolute error per point (RAEP) is applied as an index to indicate the severity of the deviation of the calculated value from the measurement value at each snapshot. Their mathematical expressions are:

TABLE I
DETAILS OF DISTURBANCES OCCURING ON LINES BETWEEN DIFFERENT NODES WITHIN DISTRIBUTION NETWORK IN TEST SYSTEM 1

No.	Disturbance	No.	Disturbance
1	Do-14-15	13	Do-16-33-LO-17-37
2	Do-15-16	14	Do-16-33-LO-20-40
3	Do-16-17	15	Do-17-18-LO-19-20
4	Do-17-18	16	Do-18-19-LR-16-33
5	Do-16-33	17	Do-33-34-LR-16-33
6	Do-17-37	18	Do-33-34-LO-17-37
7	Do-33-34	19	Do-18-19
8	Do-15-16-LO-33-34	20	Do-16-17-LR-17-21
9	Do-16-17-LR-17-33	21	Do-16-17-LO-37-38
10	Do-16-17-LR-17-33	22	Do-16-33-LR-22-38
11	Do-16-17-LO-20-21	23	Do-33-34-LO-19-20
12	Do-16-33-LR-19-37	24	Do-17-33-LO-16-33

Note: Do, LO, and LR denote disturbance occurrence, line outage, and line reorganization, respectively. For example, Do-14-15 indicates that the disturbance occurs on line 14-15; Do-16-17-LO-20-21 indicates that the disturbance occurs on line 16-17 with the line outage on line 20-21; and Do-16-33-LR-19-37 indicates that the disturbance occurs on line 16-33 with the line reorganization on line 19-37.

$$RMSPE=\frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y_{km}-y_{kc}\right)^2}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}y_{km}^2}}\times 100\% \tag{10}$$

$$RAEP=\frac{\left|y_{km}-y_{kc}\right|}{y_{km}}\times 100\% \tag{11}$$

where y_{km} and y_{kc} are the measurement value of simulation results and the calculated value of the proposed method at sampling point k, respectively; and N is the total number of sampling points during entire time-domain period.

The testing set is applied to evaluate the generalization ability of the trained model, which is crucial for practical applications. Therefore, the RMSPEs of active and reactive power in training set and testing set are illustrated Fig. 7, and we can draw the following conclusions.

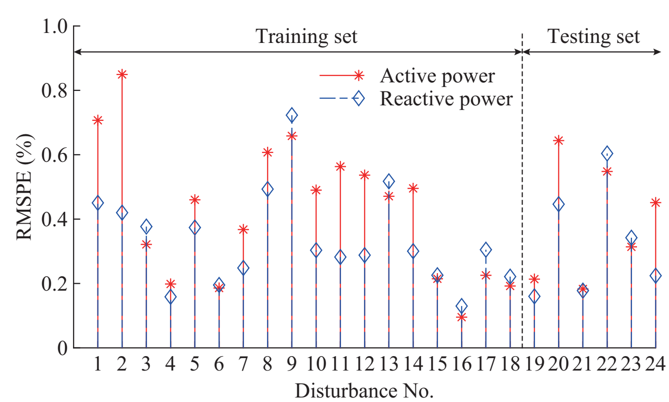

Fig. 7. RMSPEs of active and reactive power in training set and testing set.

1) The accuracy of the proposed method varies with differ-

ent disturbance locations. For example, the RMSPE of Do-14-15 is high, at 0.707% for active power and 0.4504% for reactive power, but that of Do-17-18 is low, at 0.1984% for active power and 0.1586% for reactive power. This may be because as the disturbance location moves away from the boundary bus, the fluctuations are weakened during the propagation to boundary bus, allowing the equivalent model to capture the weakened dynamic response.

2) The accuracy of the proposed method after topology reconfiguration is different from that before topology reconfiguration. For example, the sum RMSPE of the active and reactive power is 0.6882% for Do-16-17, but 0.7933% for Do-16-17-LR-17-33 and 0.8454% for Do-16-17-LO-20-21. This indicates that the dynamic response has changed after topology reconfiguration.

3) The proposed method can maintain dynamic response tracking under the topology reconfiguration after training, e. g., Do-16-17-LO-37-38 in testing set. This indicates that the feature extraction module (i.e., CNN + attention module) can effectively track the dynamic changes of this network topology, thereby enabling the proposed method to better track the dynamic behaviors of ADNs.

Overall, the above conclusions reveal that the proposed method can accurately represent the dynamic response of ADNs, even though under the topology reconfigurations. Moreover, in all datasets, the RMSPE of active and reactive power is very small (below 1%), especially that in testing set, indicating that the proposed method has great generalization capability.

In order to better demonstrate the training effectiveness and the tracking performance of the proposed method under the topology reconfiguration, we take two representative examples, i.e., Do-16-17 in training set and Do-16-17-LO-37-38 in testing set, by visualizing the active and reactive power responses, as shown in Figs. 8 and 9, respectively. $P_{measured}$ and $Q_{measured}$ are the measurement active and reactive power of simulation results, respectively; and $P_{proposed}$ and $Q_{proposed}$ are the calculated active and reactive power of the proposed method, respectively. Note that the disturbance type here is the ABC short circuit with 0.01 Ω resistance.

As can be observed from Figs. 8 and 9, the dynamic behaviors (i.e., active and reactive power responses) of ADNs change greatly after the topology reconfiguration, especially in reactive power. Specifically, after the line outage occurs on line 37-38, the active power is generally shifted downwards (i.e., 108 kW of active power for Do-16-17 and 100 kW of active power for Do-16-17-LO-37-38 in steady stage), and there are significant differences during the oscillation phases (i.e., the dynamic reactive power response within the time interval of 0.5-2.5 s).

Moreover, Figs. 8 and 9 indicate that the outputs of the proposed method well fit the original simulation results during the disturbance, recovery, and steady-state periods. It is worth mentioning that Fig. 9 shows the results obtained in the testing set, and the characteristics of interference are different from those in the training set. This also indicates that the feature extraction module can effectively track this topology reconfiguration, thereby better assisting load modeling module in tracking the dynamic behaviors of ADNs.

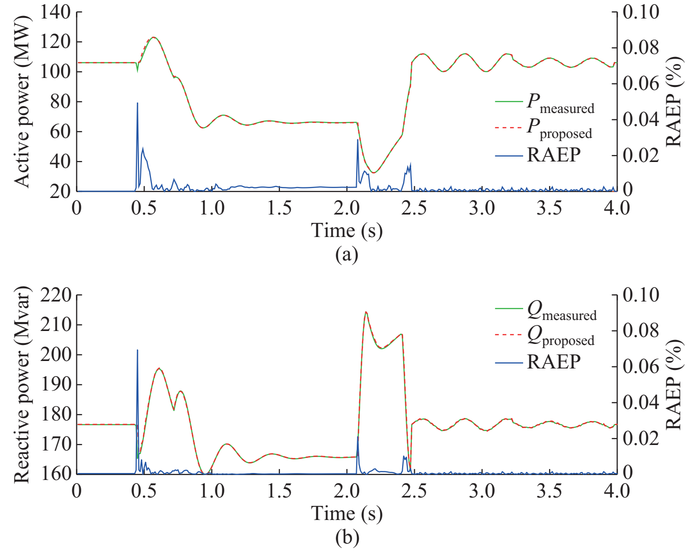

Fig. 8. Active and reactive power responses under Do-16-17 in training set. (a) Active power. (b) Reactive power.

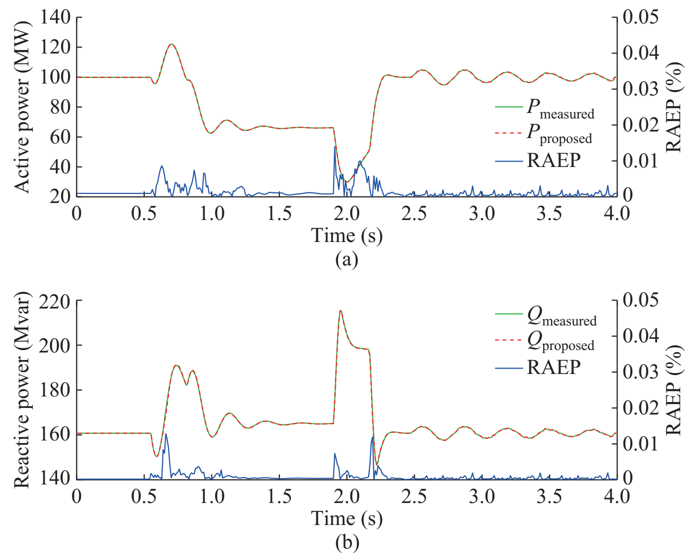

Fig. 9. Active and reactive power responses under Do-16-17-LO-37-38 in testing set. (a) Active power. (b) Reactive power.

It can be clearly observed from Figs. 8 and 9 that the dynamic behaviors of ADNs can be divided into three stages after the disturbance. The first stage is the overshoot after the disturbance, followed by oscillations in the second stage, and finally the steady state in the third stage after the disturbance is eliminated. Overall, the above results suggest that the proposed method can accurately track the dynamic active and reactive power responses under topology reconfigurations.

C. Performance Evaluation of Proposed Method Under Different Disturbance Types

In this subsection, the performance of the proposed method under different disturbance types are evaluated, including three-phase ground (ABCG) short circuit, ABC short circuit, two-phase ground (ABG) short circuit, two-phase (AB) short circuit, and single-phase ground (AG) short circuit conducted on line between node 15 in the transmission network and node 16 in the distribution network (i.e., line 15-16) .

Table II shows the RMSPE of the proposed method under five different disturbance types, and the results reveal that the proposed method can also accurately track the dynamic behaviors of ADNs under different disturbance types, especially under ABG and AB short circuits.

TABLE II
RMSPE OF PROPOSED METHOD UNDER FIVE DIFFERENT DISTURBANCE TYPES

Disturbance type	RMSPE (%)	
	Active power	Reactive power
ABCG short circuit	0.8510	0.3776
ABC short circuit	0.8956	0.4200
ABG short circuit	0.7559	0.3648
AB short circuit	0.5408	0.3758
AG short circuit	0.6291	0.4158

In order to better demonstrate the dynamic tracking capability of the proposed method, we visualize two cases of the above examples in Table II, i.e., the active and reactive power responses under AG and ABG short circuit on line 15-16, as shown in Figs. 10 and 11, respectively.

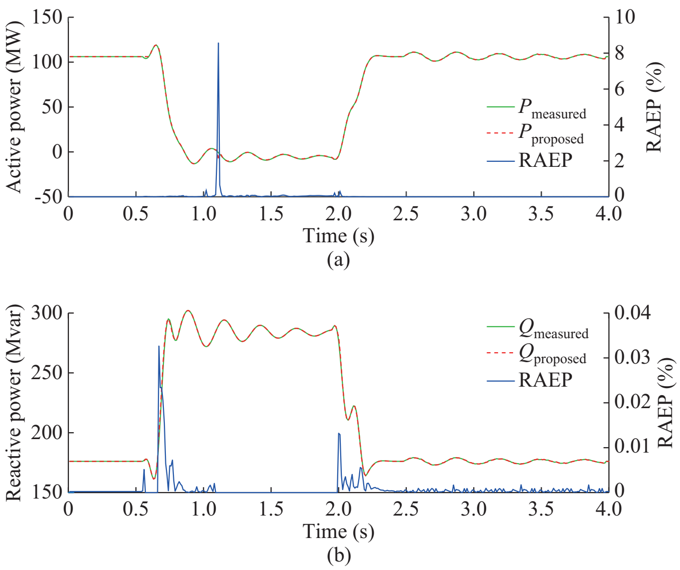

Fig. 10. Active and reactive power responses under AG short circuit on line 15-16. (a) Active power. (b) Reactive power.

Although other disturbance types do not cause as much impact to the power grid as ABC short circuit, their impacts on the oscillation of power grid cannot be ignored. From Figs. 9 and 11, although the fluctuation in the oscillation stage under AB short circuit is not as high as that under ABC short circuit, its waveform (i.e., the dynamic reactive power response within the time interval of 0.5-2.5 s) is more complex. However, the proposed method can also track this dynamic changes well. This also proves that the feature extraction module can track the network physical changes caused by different disturbance types.

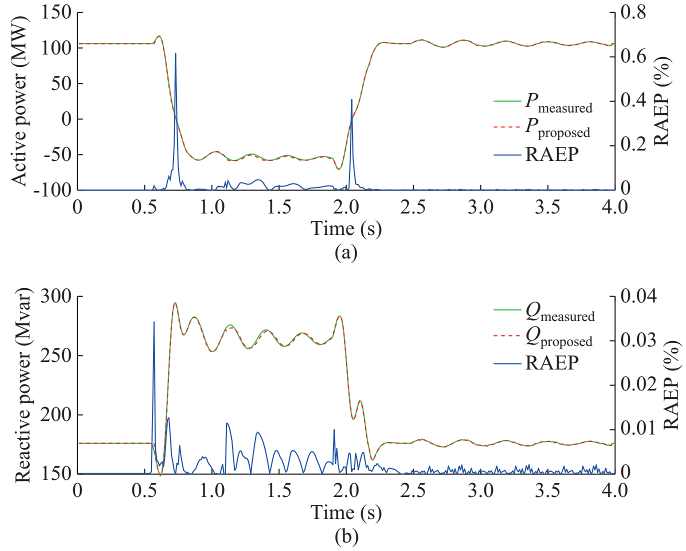

Fig. 11. Active and reactive power responses under ABG short circuit on line 15-16. (a) Active power. (b) Reactive power.

D. Performance Evaluation of Proposed Method in Test System 2

Five representative disturbances in test system 2 consisting of ABC short circuit at different locations with topology reconfigurations are conducted. The results are shown in Table III.

TABLE III
RMSPE OF PROPOSED METHOD UNDER FIVE REPRESENTATIVE DISTURBANCES IN TEST SYSTEM 2

No.	Disturbance	RMSPE (%)	
		Active power	Reactive power
1	Do-T13-D1-LO-D5-D6	0.6910	0.5464
2	Do-D1-D2-LO-D23-D24	0.8231	0.4548
3	Do-D2-D3-LR-D2-D14	1.1716	0.6814
4	Do-D25-D30-LR-D24-D32	0.4818	0.5934
5	Do-D1-D15	0.7543	0.6329

Note: D*i* and T*j* represent the node Nos. in the distribution network and transmission network, respectively. For example, Do-T13-D1-LO-D5-D6 indicates that the ABC short circuit occurs on line T13-D1 with line outage on line D5-D6.

Among these results, the proposed method performs best under Do-D25-D30-LR-D24-D32 (with an RMSPE of 0.4818% for active power and 0.5934% for reactive power), while performs worst under Do-D2-D3-LR-D2-D14 (with an RMSPE of 1.1716% for active power and 0.6814% for reactive power). However, these results are all within an acceptable and optimistic range.

Moreover, as shown in Table IV, different disturbance types are also conducted on the connection line between the transmission network and distribution network, i.e., line T13-D1, to evaluate the performance of the proposed method in a larger system. In Table IV, for active power, the maximum RMSPE is 1.1981% under the AB short circuit, and the minimum RMSPE is 0.4850% under the AG short circuit. For reactive power, the maximum RMSPE is 0.7820% under the ABCG short circuit, and the minimum RMSPE is 0.2085% under the ABG short circuit.

TABLE IV
RMSPE OF PROPOSED METHOD UNDER DIFFERENT DISTURBANCE TYPES ON LINE T13-D1

Disturbance type	RMSPE (%)	
	Active power	Reactive power
ABCG short circuit	0.7802	0.7820
ABC short circuit	0.6910	0.5464
ABG short circuit	0.9064	0.2085
AB short circuit	1.1981	0.6455
AG short circuit	0.4850	0.6452

The active and reactive power responses of a representative case, i.e., AG short circuit on line T13-D1, is visualized in Fig. 12. Overall, Table IV proves that the proposed method can track the dynamic behaviors of ADNs under various disturbances, even though in large systems with high integration of DGs.

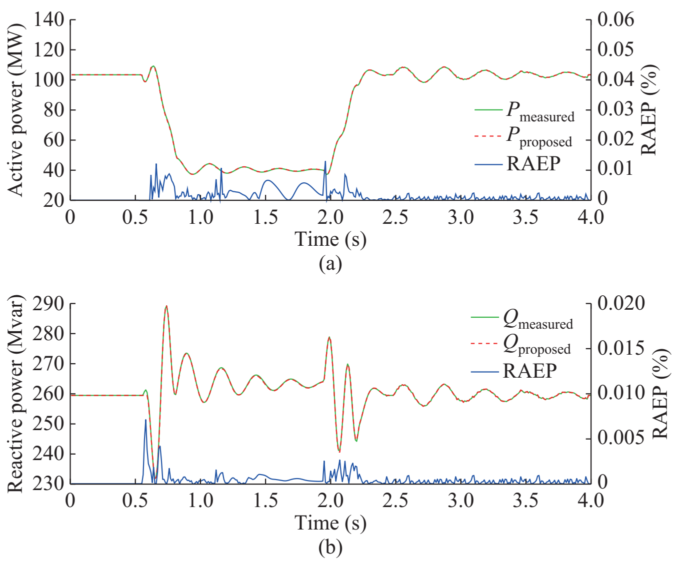

Fig. 12. Active and reactive power responses under AG short circuit on line T13-D1. (a) Active power. (b) Reactive power.

Based on the results above, it is concluded that the proposed method is robust enough for tracking the dynamic behaviors of ADNs under different disturbance locations and types, even with more buses and loads.

E. Sensitivity Analysis of Dataset Size on Accuracy of Proposed Method

This subsection compares the training time and testing accuracy of the proposed method under different dataset sizes. The ratio of training set and test setting is still set to be 3:1, and the dataset size is set to be 12-40 with a step of 4. As visualized in Fig. 13, the RMSPE of active and reactive power decreases with the increase of dataset size, and then has a slight upward trend. This indicates that when the dataset is small, the equivalent model is in the state of under-fitting, so the accuracy of equivalent model is non-ideal. When the dataset size increases to a certain extent, the equivalent model is in the state of over-fitting, so the RMSPE will have an upward trend. In addition, the training time increases with the increase of dataset size. Accordingly, in order to give consideration to the efficiency and accuracy of equivalent model, it is a good choice to choose the dataset with the size of 24 or 28, as shown in Fig. 13.

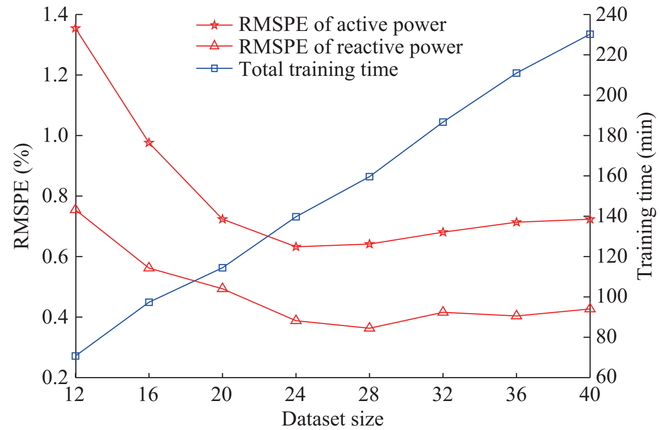

Fig. 13. Sensitivity results of dataset size.

F. Performance Comparison Between Proposed Method and Other DL-based Equivalent Methods

In this subsection, the effectiveness of the proposed method is compared with that of other DL-based equivalent methods (i. e., LSTM-based method, RNN-based method, ANN-based method). To fairly compare the performance, these benchmark methods are constructed with three hidden layers, which are the same as the expression module for DAEs in the proposed method. In addition, in order to prove that GRU can better describe DAEs than LSTM, we also employ the LSTM-based multimodal structure, i. e., LSTM-CNN-based method, to conduct comparative experiments. The feature extraction module of LSTM-CNN-based method is consistent with that of the proposed method. The difference is that LSTM-CNN-based method uses LSTM to represent DAEs rather than GRU.

To present the comparison results, a representative case, i.e., AB short circuit on line 15-16 in test system 1, is visualized and the results are shown in Fig. 14. The conclusion are drawn as follows.

1) The zoomed-in parts in Fig. 14 show that the results of ANN-, RNN-, and LSTM-based methods have greatly deviated from the simulation values, indicating that they are no longer suitable under different disturbances and topology reconfigurations.

2) In contrast, the results of the proposed method fit the true value well, indicating that the feature extraction module can extract the topology reconfigurations, thereby enabling the proposed method to better perceive the dynamic responses of ADNs.

3) The accuracy of LSTM-CNN-based method is higher than that of other benchmark methods but inferior to that of the proposed method, which shows the scalability of GRU and the effectiveness of the feature extraction module.

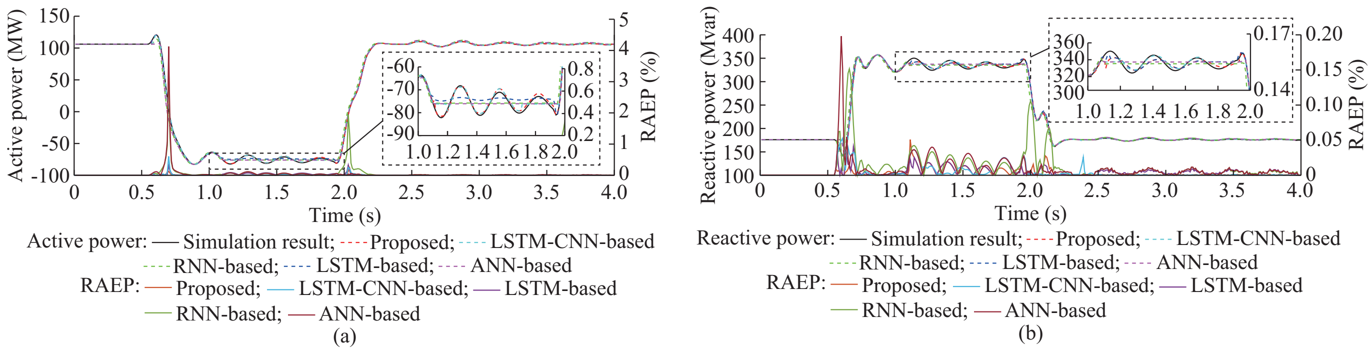

Fig. 14. Comparison results of active and reactive power responses using proposed method and other benchmark methods. (a) Active power. (b) Reactive power.

The RMSPEs with the proposed method and benchmark methods, i.e., LSTM-CNN-, LSTM-, RNN-, and ANN-based methods, are listed in Table V, where different disturbance locations and types and topology reconfigurations are considered.

TABLE V
RMSPEs WITH PROPOSED METHOD AND BENCHMARK METHODS

Disturbance	Short circuit type	RMSPE of active power (%)					RMSPE of reactive power (%)				
		Proposed	LSTM-CNN-based	LSTM-based	RNN-based	ANN-based	Proposed	LSTM-CNN-based	LSTM-based	RNN-based	ANN-based
Do-16-17	ABC	0.3214	0.5440	0.9286	1.5356	0.5821	0.3768	0.6271	1.0045	1.6161	0.6016
Do-16-33	ABC	0.4599	0.5334	0.8491	1.8865	1.3842	0.3738	0.7929	0.7341	1.1487	0.7390
Do-33-34	ABC	0.3675	0.6347	0.4682	1.1271	0.5593	0.2482	0.3129	0.5980	1.2631	0.4924
Do-15-16	ABC	1.0494	1.2767	1.2609	2.3227	1.8099	0.4200	0.4563	0.4688	1.1865	0.7871
Do-15-16	ABCG	0.8570	0.9359	1.5221	2.5588	1.6795	0.3821	0.5321	0.7753	1.2066	1.0997
Do-15-16	ABG	0.7529	1.1705	2.9188	3.2296	2.1113	0.3637	0.5633	0.5109	0.9356	1.4513
Do-15-16-LO-33-34	ABC	0.6073	0.8316	1.0183	1.5839	1.3207	0.4931	0.6047	0.9042	1.1967	0.9668
Do-16-17-LR-17-21	ABC	0.6440	0.8904	0.9718	1.7031	1.5875	0.4463	0.5829	0.8783	1.3091	1.1256

The average training time and RMSPEs with these methods are shown in Table VI. The conclusions can be drawn as follows.

TABLE VI
AVERAGE TRAINING TIME AND RMSPEs WITH PROPOSED METHOD AND BENCHMARK METHODS

Method	Average training time (s)	Average RMSPE (%)	
		Active power	Reactive power
Proposed	465.800	0.6324	0.3880
LSTM-CNN-based	646.300	0.8522	0.5559
LSTM-based	471.200	1.2422	0.7343
RNN-based	252.500	1.9934	1.2328
ANN-based	102.625	1.3793	0.9079

1) Compared with benchmark methods, the RMSPEs with the proposed method are lower.

2) Except for RNN-based method, the RMSPEs with other methods are all small.

3) The average training time of the proposed method and LSTM-based method is almost the same, but the proposed method outperforms the LSTM-based method in terms of equivalent precision.

4) The ANN-based method has the highest training efficiency, but it performs worst in the complex cases (i.e., line 15-16 and line 16-33), which is not suitable for serious scenarios.

5) Compared with LSTM-CNN-based method, the proposed method has shorter training time and better fitting efficiency. Therefore, it can be concluded that the GRU performs better than LSTM in representing the DAEs of loads.

Overall, it is validated that the proposed method performs better than other benchmark methods for tracking dynamic behaviors of ADNs even under different disturbances and topology reconfigurations.

IV. CONCLUSION

In this paper, we propose a multimodal machine learning based equivalent modeling method for tracking dynamic behaviors of ADNs. The proposed method consists of two modules, one of which is GRU + FC module used to simulate the DAEs of loads, and the other one is CNN + attention module to simulate the NRPFC to achieve the function of feature extraction. The Jacobian matrix of power grid is the core part of NRPFC and can be applied to form the spatial information, therefore the dynamic spatial changes of power grid can be represented and input to the CNN for feature extraction. Therefore, the proposed method can effectively track the dynamic behaviors of ADNs through the multimodal perception of load modeling module and feature extraction module. Then, two test systems of different scales are applied to evaluate the performance of the proposed method. The simulation results show that the proposed method can accurately track the dynamic behaviors of ADNs under different disturbances and topology reconfigurations. In addition, compared with the results of LSTM-CNN-, LSTM-, RNN-, and ANN-based methods, the proposed method performs best among all the methods. Therefore, the proposed method can be a novel and effective way for capturing the dynamic behaviors of ADNs in both DAE and NRPFC aspect.

References

[1] J. Ruan, G. Liang, J. Zhao *et al*., "Deep learning for cybersecurity in smart grids: review and perspectives," *Energy Conversion and Economics*, vol. 4, no. 4, pp. 233-251, Aug. 2023.

[2] A. Rouhani and A. Abur, "Real-time dynamic parameter estimation for an exponential dynamic load model," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1530-1536, May 2016.

[3] W. Price, C. Taylor, and G. Rogers, "Standard load models for power flow and dynamic performance simulation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1302-1313, Aug. 1995.

[4] A. Arif, Z. Wang, J. Wang *et al*., "Load modeling: a review," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5986-5999, Nov. 2018.

[5] J. Zheng, W. Wang, Z. Li *et al*., "Multi-layer double deep *Q* network for active distribution network equivalent modeling with internal identification for EV loads," *Applied Soft Computing*, vol. 147, p. 110834, Nov. 2023.

[6] I. Dzafic, M. Glavic, and S. Tesnjak, "A component-based power system model-driven architecture," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 2109-2110, Nov. 2004.

[7] D. Karlsson and D. J. Hill, "Modelling and identification of nonlinear dynamic loads in power systems," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 157-166, Feb. 1994.

[8] B. K. Choi, H. D. Chiang, Y. Li *et al*., "Measurement-based dynamic load models: derivation, comparison, and validation," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1276-1283, Aug. 2006.

[9] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos *et al*., "Measurement-based dynamic load modeling using the vector fitting technique," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 338-351, Jan. 2018.

[10] J. Zhao, Z. Wang, and J. Wang, "Robust time-varying load modeling for conservation voltage reduction assessment," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3304-3312, Jul. 2018.

[11] F. Hu, K. Sun, A. D. Rosso *et al*., "Measurement-based real-time voltage stability monitoring for load areas," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2787-2798, Jul. 2016.

[12] Z. Wang and J. Wang, "Time-varying stochastic assessment of conservation voltage reduction based on load modeling," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2321-2328, Sept. 2014.

[13] J. Wen, L. Jiang, Q. H. Wu *et al*., "Power system load modeling by learning based on system measurements," *IEEE Transactions on Power Delivery*, vol. 18, no. 2, pp. 364-371, Apr. 2003.

[14] J. K. Kim, K. An, J. Ma *et al*., "Fast and reliable estimation of composite load model parameters using analytical similarity of parameter sensitivity," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 663-671, Jan. 2016.

[15] C. A. C. F. da Silva, D. C. Rosa, P. B. C. Miranda *et al*., "Automated CNN optimization using multi-objective grammatical evolution," *Applied Soft Computing*, vol. 151, p. 111124, Jan. 2024.

[16] J. T. Chien and Y. C. Ku, "Bayesian recurrent neural network for language modeling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 361-374, Feb. 2016.

[17] Y. Zhang and H. Wang, "Multi-head attention-based probabilistic CNN-BiLSTM for day-ahead wind speed forecasting," *Energy*, vol. 278, p. 127865, Sept. 2023.

[18] Y. Li, H. Wu, and H. Liu, "Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction," *Energy Conversion and Management*, vol. 167, pp. 203-219, Jul. 2018.

[19] X. Zheng, F. Bai, Z. Zhuang *et al*., "A new demand response management strategy considering renewable energy prediction and filtering technology," *Renewable Energy*, vol. 211, pp. 656-668, Jul. 2023.

[20] G. Lu, C. W. Tsang, H. N. Yim *et al*., "Interpretable fault diagnosis for overhead lines with covered conductors: a physics-informed deep learning approach," *Protection and Control of Modern Power Systems*, vol. 10, no. 2, pp. 25-39, Mar. 2025.

[21] C. Wu, Y. Zhou, and J. Wu, "Two-layer data-driven robust scheduling for industrial heat loads," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 1, pp. 265-275, Jan. 2024

[22] J. S. Farkhani, Ö. Çelik, K. Ma *et al*., "Fault detection, classification, and location based on empirical wavelet transform-teager energy operator and ANN for hybrid transmission lines in VSC-HVDC systems," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 3, pp. 840-851, May 2025.

[23] B. Y. Ku, R. J. Thomas, C. Y. Chiou *et al*., "Power system dynamic load modeling using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1868-1874, Nov. 1994.

[24] M. Bostanci, J. Koplowitz, and C. W. Taylor, "Identification of power system load dynamics using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1468-1473, Nov. 1997.

[25] T. Hiyama, M. Tokieda, W. Hubbi *et al*., "Artificial neural network based dynamic load modeling," *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1576-1583, Nov. 1997.

[26] G. Chang, C. Chen, and Y. Liu, "A neural-network-based method of modeling electric arc furnace load for power engineering study," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 138-146, Feb. 2010.

[27] C. Zheng, S. Wang, Y. Liu *et al*., "A novel RNN based load modelling method with measurement data in active distribution system," *Electric Power Systems Research*, vol. 166, pp. 112-124, Jan. 2019.

[28] C. Zheng, S. Wang, Y. Liu *et al*., "A novel equivalent model of active distribution networks based on LSTM," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2611-2624, Sept. 2019.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[30] J. Chung, C. Gulcehre, K. Cho *et al*., "Empirical evaluation of gated recurrent neural networks on sequence modeling," *Eprint Arxiv*, 2014. DOI:10.48550/arXiv.1412.3555.

[31] W. Nam and B. Jang, "A survey on multimodal bidirectional machine learning translation of image and natural language processing," *Expert Systems with Applications*, vol. 235, p. 121168, Jan. 2024.

[32] C. Du, K. Fu, J. Li *et al*., "Decoding visual neural representations by multimodal learning of brain-visual-linguistic features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10760-10777, Sept. 2023.

[33] S. Ali, S. El-Sappagh, F. Ali *et al*., "Multitask deep learning for cost-effective prediction of patient's length of stay and readmission state using multimodal physical activity sensory data," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 5793-5804, Dec. 2022.

[34] H. Liu, T. Lou, Y. Zhang *et al*., "EEG-based multimodal emotion recognition: a machine learning perspective," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, p. 4003729, Feb. 2024.

[35] D. Wang, J. Gan, J. Mao *et al*., "Forecasting power demand in China with a CNN-LSTM model including multimodal information," *Energy*, vol. 263, p. 126012, Jan. 2023.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, May 2017.

[37] W. Cui, X. Deng, and Z. Zhang, "Improved convolutional neural network based on multi-head attention mechanism for industrial process fault classification," in *Proceedings of 2020 IEEE 9th Data Driven Control and Learning Systems Conference*, Liuzhou, China, Nov. 2020, pp. 918-922.

[38] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, Mar. 1994.

[39] P. Zhu, Y. Li, Y. Hu *et al*., "MCIGRU: stock prediction model based on multi-head cross-attention and improved GRU," *Neurocomputing*, vol. 638, p. 130168, Mar. 2025.

Wenhao Wang received the B.S. degree from the School of Architecture and Engineering, Northeast Electric Power University, Jilin, China, in 2020, and the M.S. degree in electrical engineering from the School of Electric

Power, South China University of Technology, Guangdong, China, in 2024. He is currently pursuing the Ph.D. degree from the School of Electric Power, South China University of Technology. His research interests include dynamic equivalence of active distribution network and artificial intelligence.

Jiehui Zheng received the B. E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree from South China University of Technology, Guangzhou, China, in 2017, all in electrical engineering. He is currently an Associate Professor with the School of Electric Power Engineering, South China University of Technology. His research interests include optimization algorithm, decision-making method, and their applications to integrated energy system.

Zhaoxi Liu received the B. E. and M. E. degrees from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2006 and 2008, respectively, and the Ph. D. degree in the Centre for Electric Power and Energy, Department of Electrical Engineering, Technical University of Denmark, Copenhagen, Denmark, in 2015. He is currently a Professor with the South China University of Technology, Guangzhou, China. His research interests include power system operation planning, integration of renewable and distributed energy source, and power system security and risk management.

Wei Yao received the B.S. and Ph.D. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2010, respectively. He was a Post-doctoral Researcher with the Department of Power Engineering, Huazhong University of Science and Technology, from 2010 to 2012 and a Postdoctoral Research Associate with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U. K., from 2012 to 2014. Currently, he is a Professor with the School of Electrical and Electronics Engineering, Huazhong University of Science and Technology. His current research interests include stability analysis and control of power system with bulk renewable energy, AC/DC hybrid power system, cyberphysical power system, and artificial intelligence and its application.

Yuekuan Zhou is the Assistant Professor at the Department of Mechanical and Aerospace Engineering, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, majoring in sustainable energy and environment. His research interests include low-carbon and resilient urban energy system, renewable energy, and new energy storage technology.

Zhigang Li received the B. E. and Ph. D. degrees from the Department of Electrical Engineering and Applied Electronic Technology of Tsinghua University, Beijing, China, in 2011 and 2016, respectively. He is currently an Assistant Professor with the Chinese University of Hong Kong (Shenzhen), Shenzhen, China. His research interests include smart grid, integrated energy system, renewable energy, optimization theory, and artificial intelligence.

Qinhua Wu received the Ph. D. degree in electrical engineering from the Queen's University of Belfast (QUB), Belfast, U.K., in 1987. From 1987 to 1991, he was a Research Fellow and a Senior Research Fellow with QUB. In 1991, he was a Lecturer and Senior Lecturer with the Department of Mathematical Sciences, Loughborough University, Loughborough, U. K. In 1995, he joined the University of Liverpool, Liverpool, U.K., to take up his appointment to the Chair of Electrical Engineering with the Department of Electrical Engineering and Electronics. He is a Chartered Engineer with the Institute of Measurement Control (InstMC), London, U.K. He is currently a Distinguished Professor with the School of Electric Power Engineering, South China University of Technology, Guangzhou, China, where he is also the Director of the Energy Research Institute. His research interests include nonlinear adaptive control, mathematical morphology, evolutionary computation, power quality, and power system control and operation.