

## Research Article

# Application of Multiattention Mechanism in Power System Branch Parameter Identification

Zhiwei Wang,<sup>1</sup> Liguu Weng<sup>1,2</sup>, Min Lu,<sup>3</sup> Jun Liu,<sup>4</sup> and Lingling Pan<sup>4</sup>

<sup>1</sup>Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>2</sup>Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>3</sup>State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, China

<sup>4</sup>China Electric Power Research Institute Co., Ltd., Nanjing 210003, China

Correspondence should be addressed to Liguu Weng; 002311@nuist.edu.cn

Received 9 July 2021; Accepted 17 August 2021; Published 31 August 2021

Academic Editor: Zhijie Wang

Copyright © 2021 Zhiwei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Maintaining accuracy and robustness has always been an unsolved problem in the task of power grid branch parameter identification. Therefore, many researchers have participated in the research of branch parameter identification. The existing methods of power grid branch parameter identification suffer from two limitations. (1) Traditional methods only use manual experience or instruments to complete parameter identification of single branch characteristics, but they are only used to identify a single target and cannot make full use of the historical information of power grid data. (2) Deep learning methods can complete model training through historical data, but these methods cannot consider the constraints of power grid topological structure, which is equivalent to identifying connected power grid branches separately. To overcome these limitations, we propose a novel multitask Graph Transformer Network (GTN), which combines a graph neural network and a multiattention mechanism to construct our model. Specifically, we input the global features and topology information of branch nodes into our GTN model. In the process of parameter identification, the multihead attention mechanism is used to fuse the branch feature information of different subspaces, which highlights the importance of different branches and enhances the ability of local feature extraction. Finally, the fitting and prediction of each branch feature are completed through the decoding layer. The experiment shows that our proposed GTN is superior to other machine learning methods and deep learning methods and can still realize accurate branch parameter identification under various noise conditions.

## 1. Introduction

Accurate identification of branch parameters is very important for the development of modern power systems [1]. Solving the problem of intelligent identification of steady-state branch parameters of the power grid, realizing efficient deployment of power grid regulation and control system, is conducive to providing guarantee for online safe and stable operation of large power grid [2]. Stable and effective management in power systems depends on accurate prediction of future branch parameters in different time ranges. Most of the existing power grid branch parameter identification methods are mainly model-driven, with relatively low identification

accuracy and poor reliability. Reliable and effective power grid branch parameter identification technology can be applied to the online application of transmission systems such as state estimation and power flow calculation, so as to improve the reliability of power grid transmission and the credibility of dispatching auxiliary decision-making and support the correctness of power grid analysis and decision-making. This greatly improves the practical level of the whole application of dispatching automation systems and is of great significance to promote sustainable development and harmonious society.

For many years, researchers have put forward various methods of branch parameter identification. These methods are mainly divided into four categories. (1) Theoretical

calculation method: in the long-term practical work, the line parameters are obtained from the design manual and product catalog according to the experience value or approximate calculation. However, due to environmental factors and changes in operating conditions, theoretical calculations cannot reflect the changes in the real parameters of transmission lines. (2) Parameter measurement method: the transmission line parameter measurement method is used to test the transmission line on the spot by using additional measuring devices in the state of power-on or power-off. (3) Estimation of line parameters based on SCADA (Supervisory Control and Data Acquisition): the line parameter estimation based on SCADA uses the field operation data, which is unified identification and estimation of the line parameters of the whole network. However, this method is difficult to measure the influence of measurement errors at different locations on the estimation accuracy of parameters of a single line and the mutual influence of parameter estimation results of different lines. (4) Estimation of line parameters based on PMU (Phasor Measurement Unit): the line parameter identification based on PMU can decouple the line to be identified from other elements of the power grid and then separate decoupling, thus effectively identifying individual line parameters. However, at present, the coverage of PMU devices is not wide enough and the cost of PMU devices is too high, so this method has not been popularized.

In the research of parameter identification, there are many works combined with machine learning. In order to maintain the stability of the power grid, Eskandarpour and Khodaei [3] proposed to make use of knowledge discovery methods and statistical machine learning for predicting the risk of failures for components and systems. Wang et al. [4] chose Random Forest (RF) as the basic classifier of AdaBoost to carry out feature construction engineering, which improved the detection accuracy of the model. Although machine learning approaches have witnessed the progress of power system branch parameter identification, there are some issues that need to be tackled in ML-based parameter identification for the branch of power systems. First of all, the robustness of the traditional least square method is not reliable. When the input data contain too much noise or introduce noise in the measurement process, the identification result of the least square method will become very poor. Secondly, methods like support vector regression (SVR) increase the dimension of input data and carry out predictive regression on input data in high-dimensional space. However, SVR depends on the selection of parameters and kernel function and is greatly affected by data. Finally, an integrated method like RF, which determines the final prediction result by voting of each tree, but when regression is carried out, it is difficult to get the final prediction result, cannot make predictions beyond the range of training set data, which may lead to overfitting in some specific noise data modeling, and this problem has been verified [5].

In recent years, deep learning has developed rapidly. Particularly, deep neural networks have made great progress in the fields of computer vision, natural language processing, and speech recognition [6]. The convolution kernel used in the traditional deep neural network is Convolutional Neural

Networks (CNN) [7], and the convolution method is shown in Figure 1(a). The processed data are Euclidean data such as image data [8] and speech data [9]. However, as far as the power transmission system is concerned, the number of grid nodes is numerous and irregular, and the data structure is shown in Figure 1(b). For such non-Euclidean data, there are few deep learning models that can be applied to deal with this type of data. Researchers tried to use a fully connected neural network (FCN) to deal with the task of power grid branch parameter identification, thus incorporating massive historical data to predict the development trend of power grid branch parameters. However, the general FCN model cannot consider the topological structure of transmission systems, and with the increase of the number of layers, the prediction results are easy to be overfitted, and the model becomes difficult to train, which limits the performance of the model and makes the prediction results inaccurate.

In this work, we aim to accurately identify the parameters of the power grid branches by adopting the latest graph neural network model [10] and the multihead attention mechanism [11]. Instead of stacking multiple hidden layers between input and output, this work adopts the structure of Graph Transformer [12]. This method can take the adjacency matrix of graph structure data and graph data as input and completely depends on the attention mechanism to describe the relationship between input and output. The introduction of an attention mechanism makes the proposed model pay more attention to the global feature information and avoids the repeated convolutional process of deep nets, leading to the proposed model better expresses branch information.

The main contributions of this paper are as follows:

- (i) We propose a novel multitask Graph Transformer Network (GTN). The encoding layer of the network is constrained by the grid structure, and the multiattention mechanism is used to consider the feature information of different branches. Based on fusing global information, important feature information and node information are fully captured. As far as we know, we are the first to use Graph Transformer to capture features from power transmission systems and apply them to the task of power grid parameter identification.
- (ii) The decoding layer uses the fully connected layers as the decoding structure and decodes the branch feature information fused in the coding layer according to the task information of different branches. The module can decode multiple branches in the power grid loop at the same time. The experimental results of our proposed model have higher accuracy and robustness because of the combination of topology information and global information.
- (iii) Compared with the machine learning models and deep learning models, the model we proposed has better performance. In addition, the Graph Transformer structure performs well in the face of noise and data loss.

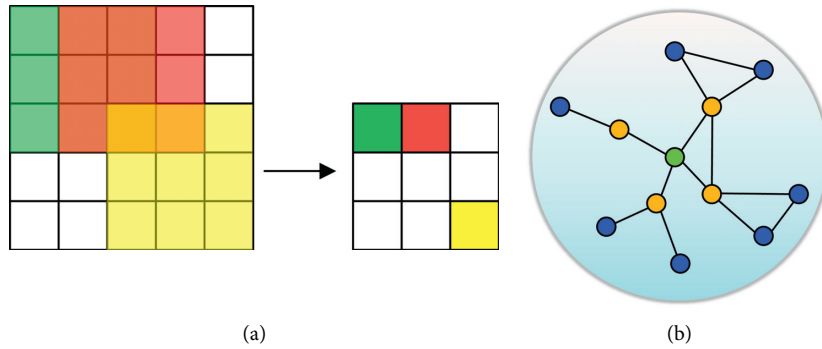


FIGURE 1: Euclidean structured data and non-Euclidean structured data. (a) Schematic diagram of CNN convolution. (b) Graph structured data.

The rest of this paper is arranged as follows. In Section 2, we introduce the development history of the branch parameter identification in the past decades. In Section 3, we introduced how to combine a graph neural network with the multihead attention mechanism. In Section 4, we introduce and analyze the experimental results. Finally, in Section 5, we summarize the above work and point out the shortcomings of this work.

## 2. Related Work

*2.1. Method for Acquiring Transmission Line Parameters.* In the past decades, researchers have put forward various methods to solve the problem of parameter identification of the power grid branches. These studies can generally be divided into the following four categories:

- (1) Theoretical calculation method: the theoretical calculation of line parameters is based on Carson's model [13]. The resistance, reactance, and susceptance are calculated according to the formula by using the physical parameters such as self-geometric mean distance, mutual geometric mean distance, and wire material of the line and combining with the external environmental factors such as soil moisture and air temperature. However, the electromagnetic model of transmission lines is greatly simplified by the theoretical calculation method, the influence of uncertain factors such as temperature and wire sag [14] does not be considered, and the calculation results are inconsistent with the actual situation. In addition, due to environmental factors and changes in operating conditions, theoretical calculations cannot reflect the changes in real parameters of transmission lines.
- (2) Parameter measurement method: the transmission line parameter measurement methods are a group of technologies to test the transmission line on the spot by using additional measuring devices in the state of power-on or power-off, which can be divided into instrument methods, digital methods, and injection measurement methods. Instrument methods realize the measurement of various states of the line by using

various instruments such as voltmeter, ammeter, power meter, and frequency meter under the power-off state and then calculating the parameters according to the corresponding formula after manual reading. Crotti et al. [15] proposed to establish a new measurement framework, which was used to realize the traceability measurement of PQ parameters in the power grid system when there was interference from the power grid system. However, due to the instrument problems, it is still impossible to accurately identify the parameters. The principle of the instrument method is simple and easy to operate, but there are inaccuracies in human readings and environmental interference. Digital methods improved the experimental data of instrument method by using a single-chip micro-computer and digital signal processing technology and improved the measurement accuracy, but it does not fundamentally change the shortcomings of traditional measurement methods in actual voltage operation environment. Injection measurement methods can be implemented when the electrical powers are "off" or incomplete "on." Based on the time pulse provided by GPS, it measures the manually added synchronous voltage and current signals and calculates the corresponding parameters through the transmission line model. Nezhadi et al. [16] proposed a new method to use stationary wavelets to denoise current and voltage signals. In the frequency range where signal energy is greater than noise energy, accurate impedance estimation can be realized by using signal injection. Ye et al. [17] declared that asynchronous time should be introduced into two-terminal fault record information, and the asynchronous time should be solved by the electric quantity constraint equation. Based on the modified synchronized voltage and current phasors at both ends, the steady-state parameters of the transmission line were determined to achieve the result of parameter identification. The injection measurement method is complicated to operate, which requires additional experimental devices, and it is difficult to reflect the true conditions of line parameters under different working conditions and operating environments.

- (3) Estimation of line parameters based on SCADA: state estimation is an important part of the Energy Management System (EMS), which often leads to unsatisfactory estimation results due to inaccurate parameters, so SCADA data is used to estimate line parameters. It mainly includes two categories: augmented state estimation and measurement residual sensitivity analysis. Debs's work [18] proposed a recursive filtering type algorithm, which proved the feasibility of parameter estimation in power systems. Do Coutto Filho et al. [19] put forward an offline processing method for branch parameters of the suspicious power grid, which could complete branch parameter identification by temporarily eliminating the participation of suspicious parameters in the process of state estimation until the suspicious parameters are corrected. Stacchini de Souza et al. [20] proposed a method of network parameter estimation and correction which is based on a genetic algorithm, combined the genetic algorithm and branch power to complete system state estimation. In Chen et al. [21], a method based on long short-term memory (LSTM) and autoencoder (AE) neural network is introduced to assess sequential condition monitoring data of the wind turbine. Parameter estimation based on SCADA data uses the field operation data to identify and estimate the line parameters of the whole network uniformly. Because the dimension of state quantity is increased, parameter estimation is carried out by equation redundancy, which may lead to numerical instability. In addition, measurement configuration needs to be fully considered to satisfy observability, and it is difficult to measure the estimation accuracy of measurement errors at different locations for a single line parameter.
- (4) Estimation of line parameters based on PMU: compared with theoretical calculation, traditional measurement, and state estimation, PMU measurement can decouple a single line from the whole network and identify it independently. Ding et al. [22] proposed the method of window sliding total least squares, PMU data of sliding window are used for parameter identification, and the influence of white noise is effectively overcome by minimizing the sum of squares of errors in the window. Zhao et al. [23] developed and implemented an online PMU-based transmission line (TL) parameter identification system (TPIS), which could consider transmission tower geometries, conductor dimension, estimates of line length, conductor sags, and so on to improve the accuracy of parameter identification. Asprou and Kyriakides [24] reported that a methodology was proposed for identifying and estimating the erroneous transmission line parameters using measurements provided by PMU and estimated states provided by a state estimator. However, in the process of practical application, there are inevitable errors in PMU measurement data, and

there is a certain gap between the identification results and the theoretical values, which leads to the problem of credibility and availability of the identification results. Therefore, the related factors affecting the identification results need to be further studied.

*2.2. Graph Neural Network.* In recent years, graph neural network (GNN) has demonstrated its efficiency in social networking, link prediction, traffic flow prediction, and other fields. To some extent, parameter identification of transmission lines in power transmission systems can also be regarded as a special graph node regression prediction. Zhou et al. [25] showed that when dealing with graph structure data, graph convolution neural network had unique advantages, which could consider both node features and node topology, and aggregated the information of adjacent nodes by using graph convolution kernel, and these convolution kernels could extract local features by end-to-end training. In other words, through the adjacency matrix constructed previously, the graph convolution neural network can obtain local features by aggregating the feature information of neighboring nodes.

Graph convolution neural network was first proposed by Scarselli et al. [26], in which the computation of graph convolution is defined in Fourier domain, while Kipf and Welling [10] proposed that first-order ChebShev polynomial could be used to generate graph convolution kernel approximately, which greatly improved the computational efficiency of graph convolution neural network. However, the feature information obtained by these methods still depends on Laplace feature related to a graph structure. In recent years, GAT [27] (graph attention network), GraphSAGE [28] (graph sample and aggregate), and other graph neural networks had appeared one after another. They have a common feature; that is, they assign different importance to different nodes in the neighborhood by using the attention mechanism and have achieved relatively good results. In addition, when FCN is used to process data, the number of layers of the model is too shallow to train and fit the desired model effect, while the number of layers of the model is too deep to easily lead to overfitting. This inspired us to use the attention mechanism to create a model; that is to say, we can use the attention mechanism to describe the relationship between input and output completely instead of traditional convolution. This can avoid overfitting of the model due to too deep layers, and the attention mechanism makes the model itself can pay attention to important nodes and feature information through learning.

*2.3. Multihead Attention Mechanism.* The structure of the multihead attention mechanism was first proposed by Vaswani et al. [11], and it was applied in natural language processing (NLP) [29, 30] firstly. Through the attention mechanism, the network emphasizes the regions of interest in the way of dynamic weighting and suppresses those regions with irrelevant backgrounds at the same time. With the weak improvement of CNN's indicators in the fields of

visual inspection and classification in recent years, the multihead attention mechanism, as a convolution structure different from CNN, shines brilliantly in the field of computer vision. For example, Dosovitskiy et al. [31] put forward the ViT model, abandoned the traditional CNN model, fully utilized the attention mechanism, applied Transformer to image classification, and achieved good classification results. Carion et al. [32] combined the common CNN and transformer architecture, took CNN as the backbone to learn the 2D representation of the input image, then used the transformer to supplement the position encoding of the input image, and finally directly predicted the detection results. Based on the above work, DETR (Detection Transformer) model was proposed. Zheng et al. [33] proposed a semantic segmentation model named Segmentation Transformer (SETR), which used Vision Transformer (ViT) as the encoder of images and then added a CNN decoder to complete the prediction of semantic graphs. The above papers show that dividing the model into multiple headers and forming multiple subspaces can make the model pay attention to different aspects of information. In other words, multihead attention can make the network capture richer feature information and finally combine the outputs by concatenating. In this paper, the multihead attention model can obtain different position information from multiple subspaces to obtain more comprehensive information.

**2.4. Multitask Learning.** Multitask learning is a kind of transfer learning, which aims to use the knowledge learned from other tasks in the target task when doing multiple tasks, so as to improve the effectiveness of the target task [34, 35]. Multitask learning can make the model adapt to multiple task scenarios, which can effectively increase the anti-interference ability of the model. There are two modes of multitask learning, as shown in Figures 2(a) and 2(b). They are hard sharing of hidden layer parameters and soft sharing of hidden layer parameters, respectively.

- (i) Hard sharing of parameters: multiple tasks share the same hidden layer of the network but do different tasks near the output of the network
- (ii) Soft sharing of parameters: different tasks use different networks, but the network parameters of different tasks use L1 regularization or L2 regularization as constraints to encourage parameter similarity

The model in this paper adopts parameter hard sharing, which was beneficial to reduce the risk of overfitting [36]. When the tasks we learn at the same time are more, the model we proposed can capture the same representation of the more tasks, resulting in an overfitting risk. Through multitask learning, we hope to predict the parameters of multiple branches at the same time and avoid overfitting through this learning method, so as to improve the robustness of the model.

### 3. Proposed Algorithm

In this section, we first define the branch parameter identification of transmission systems. Then, we introduce the technical details of our proposed model.

**3.1. Problem Statement.** Given the features of the power grid branch, the goal is to predict the true values of line susceptance  $b$  and branch conductance  $g$  of each branch. In this paper, a multitask Graph Transformer Network is designed to achieve that. By connecting the transformer nodes in the power system, we construct a graph  $G(V, E)$  composed of vertex set  $V$  and edge set  $E$  representing the connectivity between points. Assuming that the power transmission network has  $N$  transformer nodes, for line  $k$ , we express the input features of the distribution system as follows:  $(P_i^k, P_j^k, Q_i^k, Q_j^k, U_i^k, U_j^k, y^k) \in X_k^7$ , in which  $i$  and  $j$  represent the nodes at both ends of the  $k$ -th branch, then  $P_i^k$  and  $P_j^k$  represent the active power at both ends of the  $k$ -th branch, and similarly,  $Q_i^k$  and  $Q_j^k$  represent reactive power at both ends of the branch, while  $U_i^k$  and  $U_j^k$  represent both ends of the branch, and  $y^k$  represents the susceptance to the ground of the  $k$ -th branch. According to equations (1) and (2), which are derived from  $\pi$ -type equivalent circuits, we can calculate the label values of line susceptance  $b$  and branch conductance  $g$ .

$$g^k = \frac{(P_i^k + P_j^k) \left[ P_i^{k^2} + (Q_i^k + U_i^{k^2} y^k)^2 \right]}{\left( U_i^{k^2} \left[ (P_i^k + P_j^k)^2 + [Q_i^k + Q_j^k + (U_i^{k^2} + U_j^{k^2}) y^k]^2 \right] \right)}, \quad (1)$$

$$b^k = \frac{-[Q_i^k + Q_j^k + (U_i^{k^2} + U_j^{k^2}) y^k] \left[ P_i^{k^2} + (Q_i^k + U_i^{k^2} y^k)^2 \right]}{\left( U_i^{k^2} \left[ (P_i^k + P_j^k)^2 + [Q_i^k + Q_j^k + (U_i^{k^2} + U_j^{k^2}) y^k]^2 \right] \right)}. \quad (2)$$

The inputs of our proposed multitask Graph Transformer Network are feature matrix  $X \in R^{n \times 7}$  and adjacency matrix  $A$ . The features of the input data contain  $N$  nodes, and each node contains the above seven features. If a power transmission system topology contains  $M$  branches, each branch needs to calculate the true value of the corresponding line susceptance  $b$  and branch conductance  $g$ .

**3.2. Traditional Machine Learning Model.** Traditional machine learning models can be used for parameter identification of transmission system branches. The most typical one is the linear regression method, which minimizes the sum of squares of errors. Dividing the data into the training set and the test set, calculate the sum of squares  $Q$  of the total error of the training data and get the linear regression model.

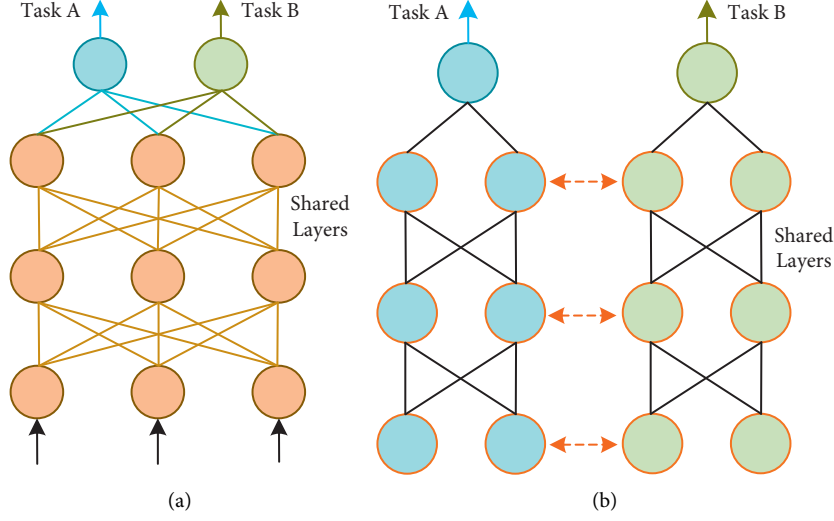


FIGURE 2: Two learning modes of multitask learning. (a) Hard parameter sharing for multitask learning. (b) Soft parameter sharing for multitask learning.

The linear regression model is applied to the test set to verify the quality of the model. As far as the linear regression method is concerned, its effect is very close to the true value without noise and other types of interference, such as node data loss. However, as far as the actual transmission system is concerned, noise interference and data loss often occur in the process of collecting data. When this happens, the linear regression model becomes unsuitable because of its poor robustness. When there is a little noise in the data, the prediction results will deviate greatly. In addition to the linear regression method, we will compare with some classical machine learning methods, including SVR (support vector regression), RF (Random Forest), and deep learning method FCN, to show the superiority of our proposed model.

**3.3. Overall Framework.** Our goal is to learn more fusion information by making full use of local and global structures, so as to make the prediction results more robust and accurate.

As shown in Figure 3, the multitask Graph Transformer consists of two parts: the encoding part and the decoding part. In Figure 3 encoding part, it takes the feature  $h^l$  and adjacency matrix  $A$  of graph structure data of power grid topology nodes as inputs and pays attention to different branch information and feature information in different subspaces by using the multihead attention mechanism. Finally, we concatenate these different subspaces, so that the information learned before is fused and input into the coding part. The structure of the coding part is shown by the decoder in the figure, which is composed of  $m$  parallel two-layer fully connected layers, and  $m$  represents  $m$  branches in the distribution system, using the branch network of different branches to fit the characteristics of different branches and achieve the purpose of accurate identification of branch parameters.

**3.4. Application of Multiattention Mechanism.** The multi-head attention mechanism has played an important role in many fields, including NLP and computer vision. Therefore, we consider applying the multihead attention mechanism to the parameter identification of transmission system branches combined with a graph neural network. The specific implementation of the multihead attention mechanism is shown in the encoder part in Figure 3. Firstly, the node feature  $h^l$  and the adjacency matrix  $A$  are considered as input data:

$$q_{c,i}^{(l)} = W_{c,q}^{(l)} h_i^{(l)} + b_{c,q}^{(l)}, \quad (3)$$

$$k_{c,j}^{(l)} = W_{c,k}^{(l)} h_j^{(l)} + b_{c,k}^{(l)}, \quad (4)$$

$$\alpha_{c,ij}^{(l)} = \frac{\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} \rangle}{\sum_{u \in N(i)} \langle q_{c,i}^{(l)}, k_{c,u}^{(l)} \rangle}, \quad (5)$$

$$V_{c,j}^{(l)} = W_{c,v}^{(l)} h_j^{(l)} + b_{c,v}^{(l)}. \quad (6)$$

According to equations (3)–(6),  $h^l$  is the output features of each layer. When  $l=0$ ,  $h^0$  is the original input data. In order to introduce the input data into different subspaces, firstly, the target node features are divided into source node features  $h_i^{(l)}$  and pointing node features  $h_j^{(l)}$  according to the adjacency matrix. The source node feature  $h_i^{(l)}$  and the pointing node feature  $h_j^{(l)}$  are, respectively, converted into the query vector  $q_{c,i}^{(l)}$  and the key vector  $k_{c,j}^{(l)}$  by using linear functions. In the above formula,  $W_{c,q}^{(l)}$ ,  $W_{c,k}^{(l)}$ ,  $b_{c,q}^{(l)}$ , and  $b_{c,k}^{(l)}$  are all trainable weight coefficients. In equation (5),  $\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} \rangle = \exp(q^T k / \sqrt{d})$  represents the ratio of dot product functions of  $q$  and  $k$  to  $\sqrt{d}$ ,  $d$  represents the number of hidden neurons in each subspace (that is, the head), and  $\alpha_{c,ij}^{(l)}$  represents the attention coefficient of a branch relative to the central node



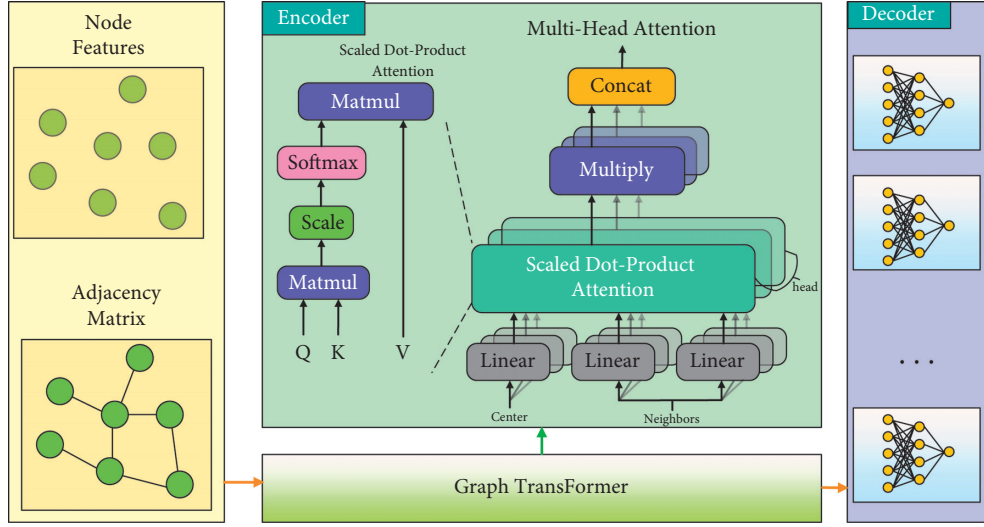


FIGURE 3: Multitask Graph Transformer Network structure.

in the  $c$ -th subspace. In addition, as shown in equation (6), the node-pointing feature  $h_j^{(l)}$  is transformed into a value vector  $V_{c,j}^{(l)}$  by using a linear function.

$$\hat{h}_i^{(l)} = \left\| \left[ \sum_{j \in N_i} (\alpha_{c,j}^{(l)} V_{c,j}^{(l)}) \right] \right\|, \quad (7)$$

$$r_i^{(l)} = W_r^{(l)} \hat{h}_i^{(l)} + b_r^{(l)}, \quad (8)$$

$$h_i^{(l+1)} = \hat{h}_i^{(l)} + r_i^{(l)}. \quad (9)$$

In equation (7),  $\|$  represents the operation of concatenating multiple subspaces. At first, the value vector  $V_{c,j}^{(l)}$  is multiplied by the attention coefficient, and then the information pointing to the node feature  $j$  is transmitted to the source node  $i$  according to the adjacency matrix  $A$  to form the source node feature  $\hat{h}_i^{(l)}$ . Then, the source node feature  $h_i^{(l)}$  is transformed into the source node feature  $r_i^{(l)}$  by using a linear function in equation (8), and the source node feature  $h_i^{(l+1)}$  of the next layer is obtained by adding the new source node features  $r_i^{(l)}$  and  $h_i^{(l)}$  in equation (9). The above is the realization process of the multiattention mechanism.

**3.5. Multitask Regression Model.** In our proposed GTN model, we use a hard parameter sharing mechanism, and the specific implementation of the multitask regression model is shown in the decoder part in Figure 3. According to Figure 3, we can find that the decoding part of the multitask Graph Transformer Network model proposed in this paper realizes decoding through multiple two fully connected layers. The encoding layer in the figure fuses rich feature and semantic information in different subspaces by taking the topology of power grid and node feature information as input and fuses the feature information of different subspaces by concatenating, which ensures that the encoding layer fuses global

information as the input of the decoding layer. As a branch of the power grid system, each branch has its own characteristics, which realize decoding through the fully connected layers and complete the task of parameter identification of power grid branches. Each branch network can fit the branch characteristics according to the branch characteristics, so as to achieve the purpose of accurate prediction.

## 4. Experiment Result

**4.1. Dataset.** Our data set comes from the actual grid line data collected by China Electric Power Research Institute, and the collection frequency is once every minute. The data set contains 8460 sets of data; there are 17 lines that need to be identified. We selected data of seven days, including 6000 sets of data as our training data, 1000 sets of data as test data, and the remaining 1460 sets of data as verification data. Figure 4 shows the topology information of collected data, which shows the connection mode between nodes.

**4.2. Baseline and Noise Settings.** In order to prove that our model can simulate the branch parameters which are closest to the real results under various error conditions, we added three kinds of noises to the original data and compared the identification results without noise and with noise as follows: (1) Gaussian noise: according to the method proposed by Brown [37], we added two kinds of Gaussian noises to the node features, which made their SNR reach 50 dB and 30 dB, respectively. (2) Node loss: in the actual distribution system, there are often cases where a node line is damaged and data cannot be collected. In order to simulate this problem in model training, we decided to simulate the loss of grid nodes, randomly select one node from each group of data, and set its characteristics to 0. (3) Loss of node features: in the process of collecting circuit data, it is common for a sensor to be damaged, and it often happens that a branch current or voltage cannot be collected. In order to simulate the

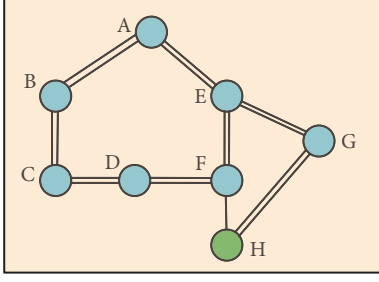


FIGURE 4: Topological structure diagram of transmission system power grid branch.

occurrence of this situation, we randomly select one of the seven features of each group of data and set it as 0, so as to compare the situation that no data can be collected during the actual operation of the power grid.

In order to prove the validity of our proposed model, we adopt the following methods as baselines:

- (1) Linear regression: the least square method is usually used as a common method in engineering. Because of its simple principle and a small amount of calculation, the least square method is often used in engineering. However, because its parameters are small and the global information cannot be considered, when a considerable amount of noise appears in the data set, the accuracy of parameter identification by the least square method will drop a lot, so its robustness is poor and it cannot achieve the purpose of accurately identifying branch parameters.
- (2) SVR: support vector regression machine is a machine learning method for regression tasks based on a support vector machine. Similarly, the kernel function is used to map features to high-dimensional space and regress them, but it partly depends on the integrity of training data and the choice of the kernel function.
- (3) RF: Random Forest is a classical algorithm in machine learning. It is a combination of multiple decision trees and depends on each decision tree to make a prediction about the target task. Finally, the final average value is obtained by averaging the predicted values of all decision trees. Its advantage is that, for unbalanced data, it can balance errors and maintain prediction accuracy when features are lost. Similarly, when the Random Forest is faced with noisy data, it will be overfitted, which cannot achieve the purpose of accurate identification.
- (4) FCN: fully connected neural network is one of the most commonly used neural networks in deep learning. It can constantly update the weights of its neurons by training and learning to identify different branch parameters. But for fully connected neural networks, overfitting is a fatal weakness. In the face of missing data or loud noise, the performance of the model cannot be fully developed.

**4.3. Evaluation Indicators and Parameter Settings.** In the model evaluation, it is usually necessary to determine the

evaluation index to measure the quality of the model experiment. In order to evaluate the quality of our model, considering that our task is a kind of linear regression, we decided to use MAE, MSE, and RMSE as evaluation indexes of the model.

MAE is also called mean absolute error, and its calculation formula is as follows:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)}|, \quad (10)$$

MSE is also called mean square error, and its calculation formula is as follows:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)})^2. \quad (11)$$

RMSE is also called root mean square error, and its calculation formula is as follows:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)})^2}, \quad (12)$$

where  $m$  represents the number of test sets,  $y_{\text{test}}^{(i)}$  represents the true value of the  $i$ -th branch in the test set, and  $\hat{y}_{\text{test}}^{(i)}$  represents the predicted value of the  $i$ -th branch in the test set. In the comparison diagram of model training in Figure 5, we choose MAE and RMSE as evaluation indicators.

The parameters of the model are set as follows: (1) LinearRegression: as a basic algorithm commonly used in parameter identification, it will be included in the basic model. (2) SVR: the radial basis convolution kernel (RBF) of SVR is set with  $C = 100$  and  $\gamma = 0.1$  and SVR in the scikit-learn library [38] in python is used to realize the support vector regression machine in this paper. (3) RF: the number of trees is set to 300, the minimum sample number of each leaf is set to 35, and the minimum sample number required for splitting is set to 3. After five cross-verifications, we determined the superparameters of SVR and RF. (4) FCN: as the most commonly used baseline model in deep learning, in order to prevent overfitting, our fully connected neural network has two layers, and its hidden neurons are 512 and 256, respectively. In the FCN model, we use the linear activation function (ReLU) as the activation function.

RMSE is used as the evaluation index in Tables 1 and 2. By comparing the experimental data table and combining the model training diagram, we can find the following.

When there is no noise or little noise, the least square method performs well and is simple and easy to use. However, the actual situation is often not ideal. We can find that the accuracy of the linear regression method drops rapidly when the noise is added to the experimental data; especially when the signal-noise ratio reaches 30 dB, the effects of other models become rather poor. As for other machine learning algorithms, although some models perform well in some of the above tasks, the accuracy of these models is not up to our requirements, which is due to the limitations of the machine learning model itself. However, on the basis of considering the relationship between



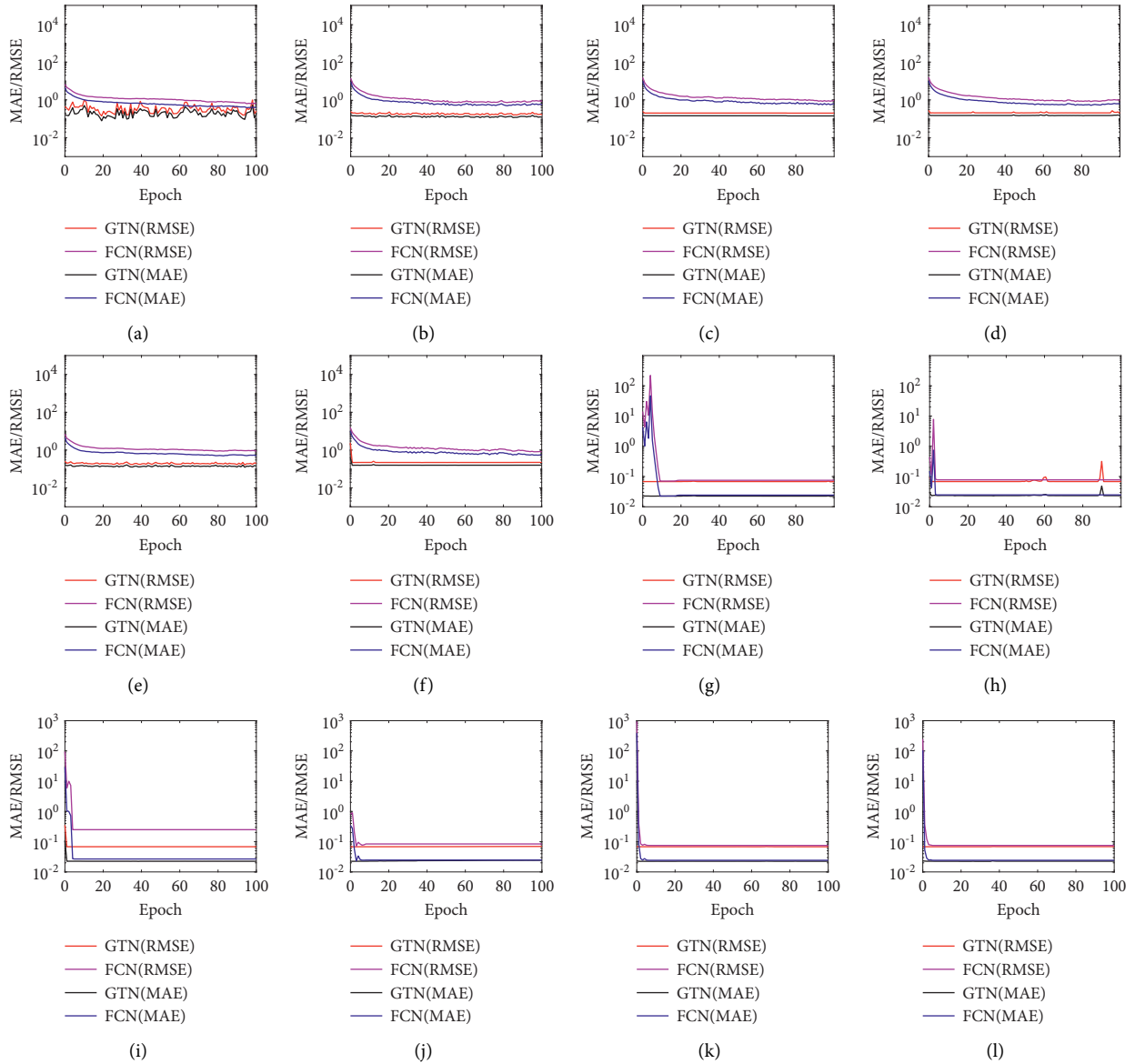


FIGURE 5: Fitting curves of FCN and GTN on target parameters under different noise correction conditions. RMSE and MAE are selected by indicators. (a-f) Line susceptance  $b$ . (g-i) Branch conductance  $g$ . (f) The superposition of noise 50 dB, dropNode, and dropData online susceptance  $b$ . (i) The superposition of noise 50 dB, dropNode, and dropData on branch conductance  $g$ .

TABLE 1: Experimental results of grid dataset on index line susceptance  $b$ .

Line susceptance $b$	None	50 dB	30 dB	dropData	dropEdge	50 dB + DD + DE
LinearRegression	<b>0.0006</b>	<b>0.1339</b>	0.4402	0.1273	0.1862	0.2569
SVR	0.1329	0.2161	1.4923	<b>0.1181</b>	0.1780	0.2399
RF	0.0785	0.1558	0.2298	0.1795	0.1898	0.2335
XGBoost	0.0856	0.1732	0.2431	0.2124	0.2373	0.2542
lightGBM	0.1324	0.1988	0.3562	0.3122	0.3242	0.3541
KNN	0.0943	0.1524	0.2345	0.2462	0.2145	0.2451
Bagging	0.0541	0.1451	0.2331	0.2143	0.2364	0.2442
FCN	0.6314	0.7603	0.8430	0.8129	0.8724	0.8133
GTN	0.1413	0.1658	<b>0.1992</b>	0.1696	<b>0.1652</b>	<b>0.2188</b>

Here we choose RMSE as the comparative index. The highest values for the different metrics are highlighted in bold.

TABLE 2: Experimental results of grid dataset on index branch conductance  $g$ .

Branch conductance $g$	None	50 dB	30 dB	dropData	dropEdge	50 dB + DD + DE
LinearRegression	<b>0.0052</b>	<b>0.0134</b>	0.1019	0.2138	0.2922	0.3481
SVR	0.0985	0.1097	0.1583	0.0885	0.3196	0.3110
RF	0.0710	0.0714	0.1258	0.0769	0.3232	0.3192
XGboost	0.0754	0.0983	0.1642	0.1013	0.1257	0.1343
lightGBM	0.1043	0.1342	0.1871	0.1222	0.1465	0.1543
KNN	0.0951	0.1011	0.1498	0.1228	0.1028	0.1435
Bagging	0.0772	0.0992	0.1345	0.1173	0.1274	0.1371
FCN	0.0706	0.0777	0.2494	0.0723	0.0751	0.0751
GTN	0.0672	0.0682	<b>0.0680</b>	<b>0.0678</b>	<b>0.0680</b>	<b>0.0681</b>

The order of RMSE index used by branch conductance  $g$  is  $10^{-7}$ . Here, we choose RMSE as the comparative index. The highest values for the different metrics are highlighted in bold.

topological structure and multisource data, the accuracy of the proposed model does not decrease a lot because of increasing noise. It is robust to resist the influence of noise. In addition, by comparing with FCN, we can find that the gradient of our proposed model drops rapidly and tends to converge after the 10th generation epoch, and the accuracy has not changed much, which shows the superiority of our model in deep learning-based parameter identification algorithms.

*4.4. The Practical Application of Our Proposed Method.* In the actual power grid transmission operation, parameter identification, as the basis of power grid regulation and control systems, has always been a hot topic of research. Most of the existing power grid branch parameter identification methods are model-driven, which have low identification accuracy and poor reliability and perform poorly when there is noise in actual power grid operation. From the experimental comparison results, it can be found that the model proposed by us has high prediction accuracy, excellent performance, and good robustness in the case of adding various noises, because of considering the topological structure constraints of power grid branches and paying attention to key branches and feature information by multihead attention mechanism. Compared with the traditional parameter identification method, the effect is improved. If the predicted model is deployed to the terminal of the power grid dispatching center, the predicted results of the model can effectively solve the problem of intelligent identification of steady-state branch parameters of the power grid, improve the reliability level of the analysis results of the dispatching system, and more effectively guarantee the online safe and stable operation of the large power grid.

*4.5. Discussion.* Number of headers in Graph Transformer: the number of headers in Graph Transformer represents the number of subspaces predicted by the model. The larger the number of subspaces, the richer the information of model fusion, but the more parameters of the model, the slower the process of model training. Choosing an appropriate number of heads is a problem that needs to be solved. At present, only an appropriate number of heads is selected through multiple experiments. In addition, because there are too many parameters in Graph Transformer, the model

parameters can be reduced by model pruning or neural network architecture search in future research, which makes the model lighter while ensuring the accuracy of the model. This is more conducive to the deployment of the model to the terminal of power grid dispatching center and improves the reliability and real-time performance of power grid branch prediction.

How to identify branch parameters of different magnitude? As far as the branch parameter identification task in this paper is concerned, the order of magnitude of line susceptance  $b$  and branch conductance  $g$  is quite different. If the simple method of loss function addition is adopted, the model will ignore the accuracy of branch conductance  $g$  and mainly focus on the accuracy of line susceptance  $b$ ; therefore, we adopted the approach that separately identified the two targets to avoid this situation. In future research, we can set a dynamic weight value to give different weight values to the line susceptance  $b$  and branch conductance  $g$  of the same line. By training an attention-based neural network, we can suppress a large number of targets and promote a small number of targets. This method will be able to identify different levels of branch parameters at the same time.

## 5. Conclusion

In this work, we propose a novel multitask Graph Transformer Network (GTN) to identify the branches of the power grid. GTN uses Graph Transformer to construct the input of graph data and abandons the traditional convolution while model learning features, and fully makes use of attention mechanism to realize the aggregation of branch features. Specifically, in the training process, the model can fuse rich global information by setting different subspaces. In addition, the attention mechanism can enhance the extraction of local information, highlight the importance of different neighbor nodes, and increase their influence by giving relatively important branches and high weights features. GTN aims to complete the task of power grid parameter identification by using the topological constraints and connections of the power grid structure. Experiments on the actual data collected by China Electric Power Research Institute show that our proposed GTN model can cope well under different noise conditions because of the integration of global information. Compared with the traditional model, the robustness of the model is improved, and the identification accuracy is also

improved, which provides a comprehensive guarantee for power grid operation and dispatching.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the project of SGCC “Data and Model Driven Parameter Identification for the Branches in Power Grid” (5108-2020190227A-0-0-00)

## References

- [1] J. Suonan and J. Qi, “An accurate fault location algorithm for transmission line based on  $r-l$  model parameter identification,” *Electric Power Systems Research*, vol. 76, no. 1–3, pp. 17–24, 2005.
- [2] J. Sun, M. Xia, and Q. Chen, “A classification identification method based on phasor measurement for distribution line parameter identification under insufficient measurements conditions,” *IEEE Access*, vol. 7, pp. 158732–158743, 2019.
- [3] R. Eskandarpour and A. Khodaei, “Machine learning based power grid outage prediction in response to extreme events,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3315–3316, 2016.
- [4] D. Wang, X. Wang, Y. Zhang, and L. Jin, “Detection of power grid disturbances and cyber-attacks based on machine learning,” *Journal of Information Security and Applications*, vol. 46, pp. 42–52, 2019.
- [5] M. R. Segal, *Machine Learning Benchmarks and Random Forest Regression*, Kluwer Academic Publishers, London, UK, 2004.
- [6] M. Xia, X. Zhang, W. A. Liu, L. Weng, and Y. Xu, “Multi-stage feature constraints learning for age estimation,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2417–2428, 2020.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [9] N. Kanda, X. Lu, and H. Kawai, “Maximum a posteriori based decoding for ctc acoustic models,” in *Interspeech*, pp. 1868–1872, International Speech Communication Association, Baixas, France, 2016.
- [10] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, <https://arxiv.org/abs/1609.02907>.
- [11] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” 2017, <https://arxiv.org/abs/1706.03762>.
- [12] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, “Graph transformer networks,” 2019, <https://arxiv.org/abs/1911.06455>.
- [13] J. R. Carson, “Wave propagation in overhead wires with ground return,” *Bell System Technical Journal*, vol. 5, no. 4, pp. 539–554, 1926.
- [14] J. S. Thorp, A. G. Phadke, S. H. Horowitz, and M. M. Begovic, “Some applications of phasor measurements to adaptive protection,” *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 791–798, 1988.
- [15] G. Crotti, H. van den Brom, E. Mohns et al., “Measurement methods and procedures for assessing accuracy of instrument transformers for power quality measurements,” in *Proceedings of the 2020 Conference on Precision Electromagnetic Measurements (CPEM)*, pp. 1–2, IEEE, Denver, CO, USA, August 2020.
- [16] M. M. A. Nezhadi, H. Hassanpour, and F. Zare, “Grid impedance estimation using low power signal injection in noisy measurement condition based on wavelet denoising,” in *Proceedings of the 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS)*, pp. 81–86, IEEE, Shahrood, Iran, December 2017.
- [17] C. Ye, S. Feng, P. Xu, and J. Liu, “Transmission line parameter identification considering non-synchronized time of fault recording information,” in *Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 1749–1753, IEEE, Xi’an, China, May 2018.
- [18] A. Debs, “Estimation of steady-state power system model parameters,” *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 5, pp. 1260–1268, 1974.
- [19] M. B. Do Coutto Filho, J. C. Stacchini de Souza, and E. B. M. Meza, “Off-line validation of power network branch parameters,” *IET Generation, Transmission & Distribution*, vol. 2, no. 6, pp. 892–905, 2008.
- [20] J. C. Stacchini de Souza, M. B. Do Coutto Filho, and E. B. M. Meza, “Treatment of multiple network parameter errors through a genetic-based algorithm,” *Electric Power Systems Research*, vol. 79, no. 11, pp. 1546–1552, 2009.
- [21] H. Chen, H. Liu, X. Chu, Q. Liu, and D. Xue, “Anomaly detection and critical scada parameters identification for wind turbines based on lstm-ae neural network,” *Renewable Energy*, vol. 172, pp. 829–840, 2021.
- [22] L. Ding, T. Bi, and D. Zhang, “Transmission line parameters identification based on moving-window tls and pmu data,” vol. 3, pp. 2187–2191, in *Proceedings of the 2011 International Conference on Advanced Power System Automation and Protection*, vol. 3, , IEEE, Beijing, China, October 2011.
- [23] X. Zhao, H. Zhou, D. Shi, H. Zhao, C. Jing, and C. Jones, “On-line pmu-based transmission line parameter identification,” *CSEE Journal of Power and Energy Systems*, vol. 1, no. 2, pp. 68–74, 2015.
- [24] M. Asprou and E. Kyriakides, “Identification and estimation of erroneous transmission line parameters using pmu measurements,” *IEEE Transactions on Power Delivery*, vol. 32, no. 6, pp. 2510–2519, 2017.
- [25] J. Zhou, G. Cui, Z. Zhang et al., “Graph neural networks: a review of methods and applications,” 2018, <https://arxiv.org/abs/1812.08434>.
- [26] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.
- [27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” 2017, <https://arxiv.org/abs/1710.10903>.

- [28] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," 2017, <https://arxiv.org/abs/1706.02216>.
- [29] M. D. Yandell and W. H. Majoros, "Genomics and natural language processing," *Nature Reviews Genetics*, vol. 3, no. 8, pp. 601–610, 2002.
- [30] M. Xia, W. A. Liu, K. Wang, X. Zhang, and Y. Xu, "Non-intrusive load disaggregation based on deep dilated residual network," *Electric Power Systems Research*, vol. 170, pp. 277–285, 2019.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," 2020, <https://arxiv.org/abs/2010.11929>.
- [32] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision*, pp. 213–229, Springer, Basel, Switzerland, November 2020.
- [33] S. Zheng, J. Lu, H. Zhao et al., "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," 2020, <https://arxiv.org/abs/2012.15840>.
- [34] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [35] M. Xia, W. A. Liu, K. Wang, W. Song, C. Chen, and Y. Li, "Non-intrusive load disaggregation based on composite deep long short-term memory network," *Expert Systems with Applications*, vol. 160, no. 6, Article ID 113669, 2020.
- [36] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, <https://arxiv.org/abs/1706.05098>.
- [37] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in PMU data," in *Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, Boston, MA, USA, July 2016.
- [38] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.