Electrical Power and Energy Systems 146 (2023) 108805

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Transformer-based deep learning model for forced oscillation localization

Mustafa Matar [a], Pablo Gill Estevez [b], Pablo Marchi [b], Francisco Messina [b], Ramadan Elmoudi [c], Safwan Wshah [d,*]

[a] Department of Electrical and Biomedical Engineering, The University of Vermont, 33 Colchester Avenue, Burlington, 05401, Vermont, USA
[b] School of Engineering, Universidad de Buenos Aires and the CSC-CONICET, Buenos Aires, Argentina
[c] New York Power Authority (NYPA), Buffalo, NY, USA
[d] Department of Computer Science, The University of Vermont, 33 Colchester Avenue, Burlington, 05401, Vermont, USA

ARTICLE INFO

Keywords:
Forced oscillations
Phasor measurement unit (PMU)
Dissipating energy
Deep learning
Transformer-based deep learning

ABSTRACT

Accurately locating Forced Oscillations (FOs) source(s) in a large-scale power system is a challenging task, and an important aspect of power system operation. In this paper, a complementary use of Deep Learning (DL)-based and Dissipating Energy Flow (DEF)-based methods are proposed to localize forced oscillation source(s) using data from Phasor Measurement Units (PMUs), by tracing the forced oscillations source(s) on the branch level in the power system network. The robustness, effectiveness and speed of the proposed approach is demonstrated in a WECC 240-bus test system, with high renewable integration in the system. Several simulated cases were tested, including non-gaussian noise, partially observable system, and operational topology variations in the system which correspond to real-world challenges. Timely localization of forced oscillation at an early stage provides the opportunity for taking remedial reaction. The results show that without the information of system operational topology, the proposed method can achieve high localization accuracy in only 0.33 s.

1. Introduction

Forced Oscillations (FOs) have become a major concern that poses a threat to the security and stability of large-scale interconnected power systems [1]. They can be caused by grid abnormal conditions such as periodic disturbances, malfunctioning controllers [2], insufficient damping of power systems, [3], periodic system disturbances, equipment failure, inadequate control designs [4], or cyclic loads [5]. Consequently, this phenomenon results in detrimental effects [6], such as reduction in the power transfer limit, potential equipment damage, power quality issues, system collapse or even large scale power outages [1,7].

FOs with frequencies higher than 1 Hz tend to be local and may be seen in a few locations near the FO source. FOs with frequencies less than 1 Hz may interact with natural oscillatory modes of the power grid and can cause wide-area oscillations across an interconnected power system [1]. For example, a 0.25 Hz FO in Alberta, Canada in 2005 led to 200 MW resonant oscillations on California Oregon Intertie (COI), 1100 miles away from the FO source [1].

Mitigating sustained oscillations starts by locating the source(s) and then disconnecting it from the network [8]. Therefore, localizing the FO source(s) is required without delay to support operational decisions to prevent further damages.

FO is still a very challenging task. Fortunately, with the increasing number of installed Phasor Measurement Units (PMUs), it is feasible to monitor FOs, particularly with a sampling rate of 30–60 samples/s, which is sufficient to capture most of FOs characteristics [6]. Advanced algorithms need to monitor PMU's data to localize and mitigate the causes. Wang Kernis et al. provided a survey of the most common methods and their limitations over the last decade [4]. Current methods can be classified into six major categories include traveling wave analysis, damping torque-based, mode shape estimation based, energy-based analysis, machine learning and deep learning.

Traveling wave-based method [9], utilizes the arrival time delay of the disturbance traveling wave, using the sampled data and the waveform similarity method to localize the source(s) of FOs associated with an abnormal wave speed, assuming that wave speeds are nearly the same throughout the grid network. This method not only requires the installation of additional equipment at many designed locations, but it also has the potential to misjudge the source(s) location when wave speeds vary significantly across the network [4].

Damping torque-based method was proposed for single machine system by [10], and multi-machine systems in [11,12]. Damping torque-based method has a clear physical meaning and could be applied locally

* Corresponding author.
E-mail address: Safwan.Wshah@uvm.edu (S. Wshah).

https://doi.org/10.1016/j.ijepes.2022.108805
Received 12 July 2022; Received in revised form 3 October 2022; Accepted 13 November 2022
Available online 24 November 2022
0142-0615/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Acronyms

BN	Batch Normalization
CNN	Convolutional Neural Network
COI	California Oregon Intertie
DEF	Dissipating Energy Flow
DL	Deep Learning
DTL	Deep Transfer Learning
FC	Fully Connected
FO	Forced Oscillation
FSST	Fourier Synchrosqueezing Transform
ISO	Independent System Operator
ISO-NE	Independent System Operator of New England
LN	Layer Normalization
LSTM	Long Short-Term Memory
ML	Machine Learning
ODE	Ordinary Differential Equation
PMU	Phasor Measurement Unit
POI	Point of Interconnection
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SST	Synchrosqueezing Transform
TF	Time-Frequency
WECC	Western Electricity Coordinating Council

for each individual generator. However, this method may fail to work under some FOs cases, and it requires detailed information including the rotor angle, rotor speed and electrical torque of generators and estimates the damping torque K_{md} for each single generator, which are not directly available [13] and difficult to gather from a wide-area perspective [4]. In addition, this method is only valid when the speed deviations for all other generators are zeros, while it is not always the case [14].

In [15] a mode shape-based method presented the relative magnitude and phasing of an oscillation through the grid. This method used PMUs measurements assuming full observability of the system, which is not practical [16]. In addition, it may fail or produce incorrect results for weakly damped, or FOs cases.

Among all these methods, the Dissipating Energy Flow (DEF) methods which utilizes every branch of the power system network where PMU data is available, has shown a promising performance [17]. As an example, they have been recently adopted by Independent System Operator of New England (ISO-NE) [18]. However, these methods are not able to differentiate between the actual source(s) bus and the one with a significant negative damping contribution [19]. In addition, these methods are still subject to performance improvement as they could not handle all the cases in the 2021 IEEE-NASPI Oscillation Source Location contest [20] and they need a large time window interval.

Machine Learning (ML) methods showed very promising performance in localizing FOs. For example, Ensemble learning approach based on data mining [21], and multivariate time series classification [19] have been applied on Western Electricity Coordinating Council (WECC) 179-bus systems. However, these approaches have many practical limitations. It requires full observability of the system and detailed information, i.e., rotor angle and rotor angle speed, which may not available.

Deep Learning (DL) has emerged as a powerful tool and achieved great success in power system applications. Recently, it has been applied for FOs localization. A two-stage Deep Transfer Learning (DTL) was proposed by [22] to convert the FOs localization problem into an image recognition problem, which utilizes a VGG16 network trained first on areas and then fine-tuned on generators. Another DL-based approach was proposed by [23], which utilizes convolutional Long Short-Term Memory (LSTM) that input row and pre-processed features to their model. They interface all possible source data to their model, resulting in an enormous model that needs to be trained on the given grid configurations.

The current DL-based approaches suffer from major issues that prevent them from being scalable or adaptable to grid variations such as: (1) They are topology dependent and work only on a fixed topology. In case of minor or major topology variations, they need to be reconstructed and retrained from scratch, which is not practical. (2) Both are designed to receive data from all network buses. For large bus systems, the model needs to handle a very large input data which limits their scalability to large networks. (3) They need very large simulated data for training. (4) They assume that every possible source bus should be equipped with a PMU, and that is not practical. (5) They are designed as a classification problem with the output represented using one-hot coding. This will only allow them to localize one source. In reality, FO could have multiple simultaneous sources.

To overcome the aforementioned issues in DL approaches, this work proposes a new Transformer-based DL approach which is robust, fast, work with presence of minor or even major topology variations without the need for retraining and can localize multiple FOs source(s). Our proposed methodology works at the branch-level to localize the oscillation source and needs little amount of data for training comparing to the other methods. It requires raw PMU measurements, and high level features to trace the FO source(s) back. Our proposed approach is built on the advantages of DEF methods by using DEF features to train our models along with the raw data and oscillation flow direction. The main contributions are summarized as follows:

(1) We propose a new Transformer-based DL approach to accurately localize FOs source(s) with low computational burden using a relatively small window of PMU measurements.

(2) The proposed approach has reliable and robust performance in the presence of real-world scenarios, such as topology variation, load variations, and partially observable system. Moreover, it also shows robustness to measurement noise typically observed in PMU data.

(3) The proposed approach is trained one time under several cases and does not require any re-training or fine-tuning for topology and load variation, which could happen during the system operation, which makes it suitable for practical online application.

(4) Comparing to the FOs localization literature, the proposed approach significantly restricts the required information such as generator models, rotor angle and rotor angle speed to localize the FO source(s), resulting in an efficient and easily implementation.

(5) The proposed approach is able to reliably detect multiple oscillation sources simultaneously.

The remainder of this paper is organized as follows. The following section provides mathematical formulation of the problem. In Section 3, the steps of the proposed approach are described in detail. A case study on the WECC 240-bus system are presented in Section 4. Experiments setup is detailed in Section 5. In Section 6 all experimental results and findings in are reported. Discussion of the results are drawn presented in Section 7. In Section 8 the main conclusions are summarized.

2. Mathematical formulation of the problem

The dynamic behavior of a power system can be represented by a n-sets of first order nonlinear Ordinary Differential Equations (ODEs) [24]:

$$\dot{x}_i = f_i(x_1, x_2 ..., x_n; u_1, u_2 ..., u_r; t); \quad i = 1, 2, \dots n \tag{1}$$

where n is the order of the system, r is the number of inputs. This can be written in the following form by using vector-matrix notation:

$$\dot{x} = \mathbf{f}(x, u, t) \tag{2a}$$

$$\mathbf{y} = \mathbf{g}(x, u) \tag{2b}$$

where the *state vector* $\boldsymbol{x} \in \mathbb{R}^n$, *input vector* $\boldsymbol{u} \in \mathbb{R}^r$, the derivative of a *state vector* $\boldsymbol{x}$ with respect to given time t is denoted by $\dot{\boldsymbol{x}}$, the *output vector* $\boldsymbol{y} \in \mathbb{R}^m$, m is potential output measurements, and $\boldsymbol{g}$ is the algebraic equations of the network.

Let i^* denote the source of forced oscillation in the system. The input i^* varies periodically, i.e., it can be considered as the superposition of h distinct frequency components. The amplitudes, frequencies and phase displacements of these frequency components comprise the sets $A = \{a_k\}$, $\Omega = \{\omega_k\}$ and $\Phi = \{\phi_k\}$, respectively, for all $\boldsymbol{k} \in \{1, 2 \ldots h\}$. Therefore, we can write the i^* input with periodical injection as:

$$u_i^*(t) = \sum_{k=1}^{h} a_k sin(w_k t + \phi_k) \tag{3}$$

As a result, sustained oscillations will be then triggered over the grid. We term the generator/load associated with i^* as the FO source. In other words, the FO localization problem is equivalent to pinpointing the FO source thought power system using available PMU measurements as input to the proposed approach. Due to the complexity of power system dynamics, the power system model may not have a fixed topology during system operations. Therefore, we propose a branch-level approach that is scalable and reliable under different real-world scenarios, as described in the following sections.

3. Proposed approach

This paper proposes a new DL-based approach that localizes the oscillation source by building an oscillation flow map at the branch level, as shown in Fig. 1. In case of a detected FO, the system reads all PMU data from each monitored branch to determine the FO direction at the branch level. It starts by buffering the data for a certain length, then calculating FO frequency and DEF features. Finally, it uses the pre-trained DL model to specify the oscillation direction from the calculated features and the raw data.

The Transformer-based DL model is trained on several features, including raw PMU measurements, i.e., active power and reactive power, FO frequency, and flow of the dissipating energy. A fixed-length sliding window was used to split the data into smaller time periods with a specific size as discussed in Section 3.3. The inference stage is applied once the FOs detected in the grid by any detection tools such as GE PhasorPoint [25], this stage could be done in online fashion. This stage consists of several components: (i) PMU data gathering stage, including extraction, curation, and processing. (ii) Characterize the oscillation by identifying the oscillation frequency. (iii) Calculate the flow of the dissipating energy for each monitored branch. (iv) Localize FO by passing buffered data and the calculated features to the pre-trained Transformer-based DL model to determine FO propagation at each branch to build a comprehensive visualization map that will allow network operators to trace the FO source(s), similar to [8,17,26] methods.

The input features to the proposed Transformer-based DL model consists of the normalized vector of active power (P_{MW}), reactive power (Q_{MVAR}), FO frequency (f_{Hz}), and flow of the dissipated energy (W_{pu}^D). The input window size should preferably include 20–40 periods of oscillations [8]. Considering that the FO frequencies are less than 1 Hz, the window size is set to be fixed-size of 70 s, even though a smaller window size can also work as well but with less robustness. The proposed approach relaxes the assumption of accurate detection of the beginning of FO, by allowing the use of window that includes measurements collected before the FO occurs. The output of the proposed approach is a visualization map, which indicates the FO propagation direction at each branch. The source(s) of oscillations is localized in system elements producing FO waves in the network.

The observability of the System by PMUs determines the resolution of localization. The proposed approach is based on tracing the FO flow through the power system network. Installing PMUs at power plant Point of Interconnection (POI) and monitoring each generator enables source localization up to a specific generator. In other words, when a power system is fully observable, the approach can point to the exact source of oscillation. Similarly, when a power system is partially observable, control room operators need to know whether the source(s) of oscillation is located inside or outside of control area to take proper course of actions [8]. The proposed approach allows for accurately localizing the suspect area containing the FO source(s). Fig. 2 illustrates these two cases where arrows represent the oscillation flow through the power system network.

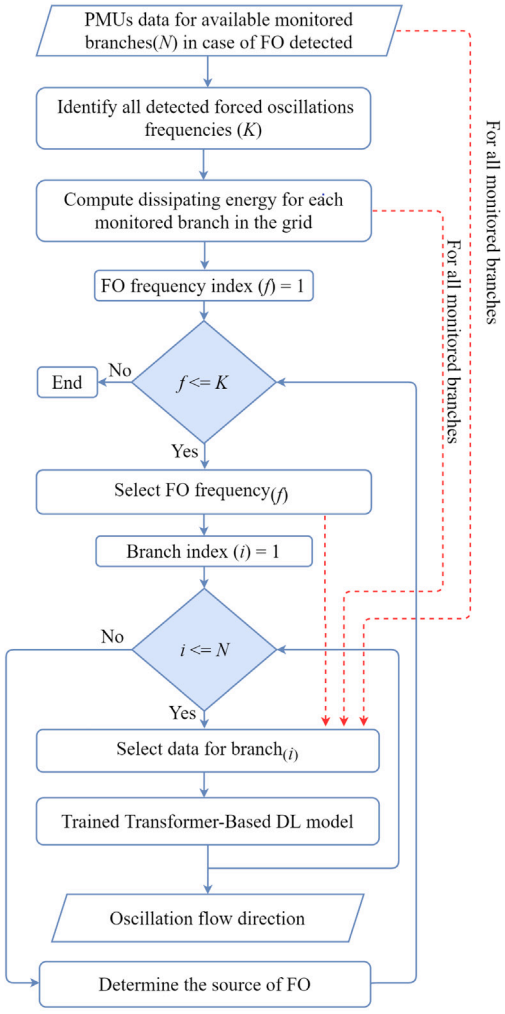

Fig. 1. Flowchart of the proposed FO localization approach. In case of FO has been detected, the approach will read all the available PMU data to determine FO propagation at each branch to build a comprehensive visualization map that will allow network operators determine the FO source(s). The DL model is used after being trained offline for one time. (- - ›) Direction of data flow, (⟶) Direction of process flow.

3.1. FO frequency identification

Monitored branches PMU measurements, namely voltages angles and currents angles, should be unwrapped using a jump threshold of 150°. Then, any missing PMU data points (NaN) and outliers should be replaced with interpolated data. Finally, low frequency trends are extracted by calculating 300 order median filter with 10 s equivalent window.

Now, to identify the frequency of interest of the forced oscillation f_{Hz}, short-time Fourier transform based Synchrosqueezing Transform

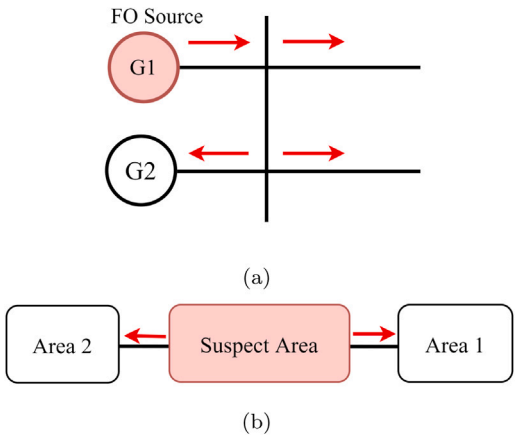

Fig. 2. The oscillation flow pattern interpretation. (a) PMU monitors point of connection of a power plant. (b) localization of suspect area non-observable by PMUs.

(SST) of active power flow P and reactive power flow Q of each available branch [27] were calculated. A unique Time-Frequency (TF) representation was calculated from individual Fourier Synchrosqueezing Transform (FSST) coefficients.

$$\mathbf{MSST}(\Omega) = \sqrt{\sum_{i=1}^{N} |\mathbf{SST_P^i}(\Omega)|^2 + |\mathbf{SST_Q^i}(\Omega)|^2} \tag{4}$$

where P^i and Q^i are the measurements of the active power flows and the reactive power flows of branches $i = 1, \ldots, N$. Application of identification algorithm on the resulting TF representation was used to identify each signal oscillatory component [28].

3.2. Dissipating energy flow

The concept of DEF [17] is based on the notion that the FO source is responsible for a dissipation of energy across the grid. DEF calculates the dissipating energy flow in the network branches, which is equivalent to the energy dissipated by a damping torque. The value and sign of the rate of change of dissipating energy have a physical interpretation as the amount and direction of the flow. DEF in-branch ij is expressed as follows, and the physical meaning of the equation is shown in [26].

$$W_{ij}^D \cong \int (\Delta P_{ij} d\Delta\Theta_i + \Delta Q_{ij} \frac{d(\Delta V_i)}{V_i^*}) \tag{5}$$

where $V_i^* = \tilde{V}_i + \Delta V_i$, and V_i^* is the average voltage in the studied period, ΔP_{ij} and ΔQ_{ij} are deviations from the steady-state values of the active and reactive power flow in branch ij, ΔV_i, $\Delta\Theta_i$ are deviations from steady-state values of voltage and voltage angle at bus i. Eq. (5) was proved in [29]. A discrete-time version of Eq. (6) has the following form:

$$\begin{aligned} W_{ij}^D[t+1] = W_{ij}^D[t] + \Delta P_{ij}[t](\Delta f_i[t+1] - \Delta f_i[t]) \\ + \Delta Q_{ij}[t] \frac{\Delta V_i[t+1] - \Delta V_i[t]}{V_i^*[t]} \end{aligned} \tag{6}$$

where t represents the time instant. DEF in each branch is either increasing or decreasing with time, indicating that dissipating energy flows into the bus from the source or from the bus to a system.

3.3. Transformer-based DL model

DL-based approaches [22,23] are proposed as alternate approach to localize the FO source(s) through the power system network. The Transformer model was developed for natural language processing and computer vision applications. It was introduced as a new sequence learning architecture with significant improvements in different applications over recurrent neural networks [30]. Transformer relies completely on attention mechanisms, with no use of Recurrent Neural Network (RNN) or Convolutional Neural Network (CNN). In this work, a Transformer-based model has been proposed to localize the FO source(s), which has shown superior performance compared with traditional architectures. The proposed Transformer-based model, Hyper-Parameters optimization, and evaluation metrics are described below.

The proposed Transformer-based model is shown in Fig. 3, which comprises of offline training and an online inference phase. The proposed Transformer-based model incorporated skip connections [31] and residual connections [32] with multi-headed attention module [30], combined with an encoder portion of transformer-based deep learning model [30] for forced oscillation localization.

Residual connections and skip connections are used throughout the network to facilitate the optimization of deep neural networks [33,34] to work better with multivariate time-series data, to speed up convergence, enable training of much deeper models, and to solve vanishing gradients issues. The residual connections allow the input to bypass and expand the path to avoid losing original information. The output of the transformer block and the data over a residual connection are added together to form the input to the next transformer block. The output of each transformer block is added together through skip connections to form the output of the entire transformer block.

As time series data is continuous, the embedding layer was built by replacing the standard embedding layer in [30] with a Fully Connected (FC) layer and Layer Normalization (LN) [35] which normalizes the activations along the feature direction. This overcomes the cons of Batch Normalization (BN) by removing the dependency on batches.

Since Transformer encoders is a feed-forward architecture that is insensitive to the ordering of input, positional encoding was used to make the model aware of the sequential nature of the time series. In this work, the positional encoding with the sine and cosine functions [30] was used.

The multi-head attention layer is the core module of transformer encoder, this layer allows the model to jointly attend to information from different representations, and determines how much attention should be paid to useful inputs when determining an output [30]. The input of the multi-headed attention layer is obtained by applying independent linear projection functions (LN) on the sequential features:

$$Q = LN_q(P_{MW}, Q_{MVAR}, W_{pu}^D, f_{Hz}) \tag{7a}$$

$$K = LN_k(P_{MW}, Q_{MVAR}, W_{pu}^D, f_{Hz}) \tag{7b}$$

$$V = LN_v(P_{MW}, Q_{MVAR}, W_{pu}^D, f_{Hz}) \tag{7c}$$

where K, V, and Q denotes the key, value, and query features respectively. These three parameters are fed to the first encoder block. The multi-headed attention layer attempts to map the query to a set key–value pairs with respect to an output to produce the attention matrix. The operation consists of a dot product of the query, Q with all keys and a division by $\sqrt{d_k}$, where and d_k is the dimension of queries and keys, and applying a sofmax function over the result as shown in Eq. (8) to obtain the weights of attention on the values.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{8}$$

Instead of performing the operation in Eq. (8) once to produce a single matrix, the operation can be repeated multiple times in parallel to allow the model to acquire information from several representation sub-spaces at several positions jointly, and the resulting matrices can be concatenated into a larger matrix as shown in Eq. (9).

$$MultiHead(Q, K, V) = concat(head_1, \ldots head_h)W^o \tag{9}$$

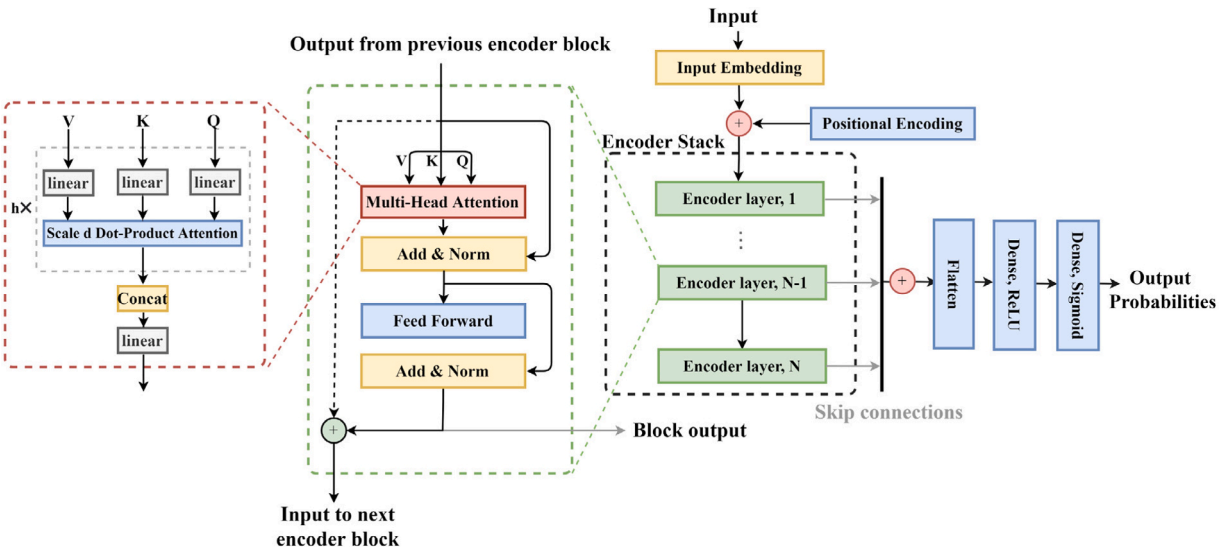

Fig. 3. The overall architecture of Transformer-based DL model.

where $head_i = Attention(QW_i^Q, KW_i^k, VW_i^V)$, $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_V}$, and $W_i^O \in \mathbb{R}^{nd_V \times d_{model}}$, h is the number of heads, d_{model} is embedding size, $d_k = d_v = d_{model}/h$.

Next, the output of the multi-head attention layer is fed into a fully connected feed-forward network, which consists of two linear transformations with a Rectified Linear Unit (ReLU) activation, and linear activation respectively. After obtaining the output of each encoding layer, they are added to each other, then fed it to flatten layer, dense with ReLU, and dense with sigmoid activation function, turning them into output probability value.

3.3.1. Hyper-parameters optimization

The selection of appropriate hyper-parameters are often necessary to maximizes the performance of any deep learning-based model, several hyper-parameters were considered to tune. In this work, Random Search [36] was used to find the network optimal hyper-parameters combinations. The optimal hyper-parameters for optimizer type, learning rate, model dimension, number of heads, and number of encoder blocks are Adam [37], 1×10^{-5}, 64, 3, and 3 respectively. All the experiments proposed Transformer-based model was carried out using Tensorflow v2.8.0.

3.3.2. Evaluation metrics

Several evaluation metrics are used to comprehensively evaluate the performance of the proposed approach. Precision, recall, and F1-score are the most important metrics [38]. The values of these metrics were calculated during the testing process. The formulas that are used in these computations are shown in Eq. (10).

$$Precision = \frac{TP}{TP + FP} \tag{10a}$$

$$Recall = \frac{TP}{TP + FN} \tag{10b}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10c}$$

where TP, FP, TN, FN are the true positive, false positive, true negative, and false negative respectively. In addition, area Under an receiver operating characteristic (ROC) curve (AUC), which is a collective measure of sensitivity $(TP/(TP + FN))$ and specificity $(TN/(TN + FP))$ over a wide range of all possible threshold values.

4. Case study on WECC 240-Bus System

The 2021 IEEE-NASPI co-hosted Oscillation Source Location (OSL) contest [20] aims to help the academia and vendors further develop and improve the available OSL tools, and help utilities identify and evaluate OSL tools for online use. The testing power system (WECC 240-Bus System) consists of North, South, California, and Mexico areas. It all has 243 buses, 146 generating units at 56 power plants (including 109 synchronous machines and 37 renewable generators), 122 power transformers, 7 switched shunts, 329 transmission lines, and 139 loads. Each generator is represented by the detailed model and equipped with an exciter and governor [39]. A short-circuit fault and/or a line tripping event may happen as an oscillation initiator or as a result of a forced oscillation. There are 16 cases, which includes single and multiple FO sources with different frequencies, where each event contains a 30 s leading window before the FO event and 60 s time window after that, total 90 s of data.

The cases provided in the 2021 IEEE-NASPI OSL contest [20] were used in this research work, these cases has been synthetically generated using a WECC 240-bus test system using Powertech Lab's transient security assessment tool (TSAT) software [40], with a 30 Hz sampling rate which is sufficient to capture electromechanical oscillations. The topology of this network model shown in [41], developed by National Renewable Energy Laboratory (NREL) [42], modified by the OSL committee [20].

There are several challenges concerning the FO events in real-world, contest committee taken it into account such as: (1) Synthetic PMU measurements provide partial observability to the system in all four areas which includes only 23 out of 56 power plants; (2) Missing PMU data, packets of one or more sequential samples make up each missing sample. The size and distribution of missing packets can vary significantly over time and between utilities. Thus, it is difficult to model a distribution of missing samples; (3) Simultaneous occurrence of system disturbances to reflect real-world challenges; (4) Additive White Gaussian Noise (AWGN) was added to the load during simulation to mimic random load fluctuations; (5) EPRI's PMU Emulator was used to process the synthetic PMU to simulate PMU device performance, which may have resulted in corrupted or substandard measurements, a mix of P Class (2-cycle window) and M Class (6-cycle window) PMUs were used. EPRI's PMU emulator [43] was used to process the synthetic PMU data to introduce data quality problems [44].

5. Experimental setup

Several test setups have been conducted in this work to evaluate the scalability, generalization capability and robustness of the proposed Transformer-Based approach to system variations. In addition, ablation studies have been conducted to evaluate the effectiveness of the proposed Transformer-Based approach. These experiments have been applied to the 2021 IEEE-NASPI OSL contest cases and generated cases from WECC 240-Bus system load and topology variation as described below.

To efficiently train the proposed Transformer-Based model, we extracted a large amount of training data by splitting 15 FO cases from [20] into smaller time periods of a specific size using a fixed-length sliding window.

5.1. Scalability studies

IEEE-NASPI Case Study Without Variations: In this experiment we are training and evaluating our proposed Transformer-Based approach on the cases described in Section 4 without any variations. Because we only have 16 FO cases in the contest dataset, the leave-one-out cross-validation technique was used here to decide on the best model to use in the following experiments and to estimate the model's performance on unseen data, where the number of folds is set to be equal to the number of cases in the dataset. Thus, the learning algorithm is applied once for each case, using all other cases as a training set and using the selected case as a testing case. This experiment can be computationally expensive, but it estimates the skill of a DL model on unseen data and ensures the generalization ability of the model and its applicability to a variety of scenarios.

Influence of Load and Small Topology Variation: During operation, topology variation of the power system could happen for several reasons i.e. faults, scheduled maintenance.

In this experiment, the effectiveness and robustness of the proposed approach are studied against a topology variation. The scenario of a small system topology change to create six challenging testing cases is that transmission lines were set to be outage separately, varying the load values through the power system network.

The proposed Transformer-Based approach has been trained one time under several FO cases from the contest dataset, and the testing performed on several load and topology variation cases in this and following experiment. In other words, the proposed Transformer-Based approach does not require any re-training or fine-tuning for topology and load variation.

Influence of Major Topology Variation: In addition for the mentioned reasons in previous experiment, generators model topologies might not be known beforehand, owing to the unpredictable switching of power system stabilizers. Thus, it is highly recommended to have a FO localization method that does not heavily rely on the availability of the first-principle model and topology information of the power system network [16].

In this experiment, the effectiveness and robustness of the proposed approach are studied against major variation of the topology. Two scenario of major system topology change to create four challenging testing cases are considered, one is removing whole area from the service which make the problem of localization FO significantly challenging since a set of generators out of operation, and another is disconnecting Bus with all the branches connected to them.

5.2. Ablation studies

The purpose of this experiment is to measure the effectiveness and the contribution of the proposed model modifications to the overall approach and features importance. Two kinds of ablation studies, namely, model ablation studies and feature ablation studies were conducted. Each ablation study involves training and evaluating a model with one or more of its components removed. Similarly, a feature ablation study involves training and evaluating the model using a different subset of features in the dataset as described below.

Transformer Model Ablation: This ablation study was conducted to provide insights into the relative contribution of each modification on the original Transformer architectural to the performance of proposed Transformer-Based model by removing the modified components from the model (e.g., skip connection and/or residual connections).

Feature Ablation: In order to verify the choice of features to the proposed Transformer-Based model and understand the influence of omitting specific features on the model performance, a feature ablation study was conducted by calculating AUC when leaving each feature group out. The difference between the performance of the model with all of the features and the performance with those features excluded is an indicator of how impactful the features are. Four kinds of features are considered, including active power, reactive power, frequency, and dissipated energy. The model's performance evaluated on the branch level using three testing cases: a case from the testing data set, a case with small topology variation (Branch# 6301-6103-2 was set to be outage and the load at bus# 4203 decreased by 10%), and a case with major topology variation (Colorado zone was removed) were used in this experiment.

5.3. Comparison of DL models

In this study a comparison of DL models in the context of FO localization is presented. The proposed Transformer-based model performance is compared with well-known traditional CNNs, LSTM architectures in [45] after modifying them to fit the problem as described below. The training data, and testing cases are the same for all the evaluated architectures, to make the comparison fair. The performance metrics for comparison of DL models are AUC and F1 scores.

CNN: CNNs architecture is adapted from [45]. This architecture consist of two convolutional layers interleaved with maximum pooling operations followed by two fully connected layers. The last layer (output layer) was replaced with layer with size of one which gives the direction of FO.

Random search [36] was applied to find the best performing CNN architecture hyper-parameters which is as follows: The two convolutional layers consist of 256 and 512 filters, respectively. These convolutional layers employ a one-dimensional convolutional kernel with the size of input samples and half of the input samples, respectively. A down sampling process by a factor of two is applied using a max-pooling layer to compress the features. The two fully connected layers consist of 1024 and 256 hidden neurons, respectively. Each layer is followed by a dropout layer with a drop rate set to 0.25 to prevent overfitting.

The vanishing gradient issue is raised in this architecture with long temporal data. In which the gradient faded out as the model got deeper prevents these from capturing the long temporal relationships in the input data. Due to the long input measurement of 70 s at a sampling rate of 30 samples per second, making their models' parameters very hard to optimize.

LSTM: LSTM architecture is adapted from [45]. This architecture was first presented to prevent backpropagated errors from vanishing or exploding, but they fail to remove it completely.

This architecture consists of a single-layer LSTM. An LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. Forget gates were introduced to avoid long-term dependency. However, LSTMs are prone to overfitting and it is difficult to apply the dropout algorithm to curb this issue.

Table 1
Leave-one-out cross-validation results.

Fold#	Testing case	Actual oscillation source(s)		Identified oscillation source(s)				
		Area	Source(s)	Avg. FO Freq (Hz)	F1	AUC	Area	Source(s)
1	Case 1	South	1431	0.821	0.991	1.000	South	1431
2	Case 2	California	2634	1.194	0.978	0.998	California	2634
3	Case 3*	South	1131	0.379	0.996	1.000	South	1131, 1032
4	Case 4*	California	3831	0.379	0.987	0.998	California	3831, 3836, 3835, 2434, 3433
5	Case 5*	North	4231	0.723	0.996	0.997	North	4231
6	Case 6	North	7031	1.267	0.897	0.982	North	7031
7	Case 7	California	2634	0.379	0.996	1.000	California	2634
8	Case 8	North	6333	0.614	0.979	1.000	North	6333
9	Case 9	North	6533, 4131	0.762	0.991	0.999	North	6533, 4131
10	Case 10*	California	3931	1.218	0.979	0.999	California	3931
		North	6335	0.614	0.990	1.000	North	6335
11	Case 11	North	4009	0.614	0.996	1.000	North	4009
12	Case 12	North	6335	0.920	0.996	1.000	North	6335
13	Case 13	North, California	4010, 2619	0.614	0.991	1.000	North, California	4010, 2619
14	A1	North	6333	0.572	0.996	1.000	North	6333
15	A2′	North	6333	0.572	0.991	1.000	North	6333
16	A3″	North	6333	0.572	1.000	1.000	North	6333

Table 2
Model ablation results.

Architecture	Evaluation metric	
	AUC	F1
The proposed approach	0.9953	0.9704
Removing skip connections	0.9814	0.9588
Removing residual connections	0.9839	0.9542
Removing skip and residual connections	0.9784	0.9330

5.4. Influence of non-Gaussian noise

PMU measurements involve noise, which may affect or even disable certain PMU applications. This noise can change from time to time due to the aging process, instruments, communication channels, etc.

Recent practical studies [46,47] revealed that the noise in PMU measurements tend to follow non-Gaussian, thick-tailed distributions such as student-t and Laplace distributions. Therefore, it would be valuable in this experiment to investigate and evaluate the performance of the proposed approach immune to non-Gaussian PMU measurement noise with Laplace distribution. The index of signal-to-noise ratio (SNR) is used in this section to describe the ratio between the power of the desired output signal (meaningful input) and the background noise (unwanted input), which is defined as follows:

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{11}$$

where P is average power. Laplace distribution is employed to corrupt the original PMU measurements, with typical SNR range 30–50 dB [48].

5.5. Time efficiency

Timely FOs localization at an early stage provides the opportunity for taking remedial reaction. So, this experiment measures the computational time of the proposed Transformer-Based approach.

6. Results and findings

In this section the results of the five experimental setups are demonstrated, along with discussion of the results.

6.1. Scalability studies

IEEE-NASPI Case Study Without Variations: The proposed approach was tested on WECC 240-Bus system cases. In order to estimate the skill of the model and generalization ability as discussed in 5.1. Table 1 shows Leave-One-Out Cross-Validation results including Frequency of each FO, the FO source(s), and the identified source(s).

This experiment shows that the proposed approach could generalize to unseen scenarios. FO frequency of case A1 is 0.379 Hz before t = 30 s, 0.614 Hz from t = 35–60 s, and 0.725 Hz after t = 65 s The network has a mix of P/M class PMU. Case A2′ is similar to case A1, but with 100% P class PMU. Case A3″ is similar to Case A1, but with 100% M class PMU.

It is important to highlight the power of the proposed approach to localize FO with partially observably system. This is to account for the fact that today's power systems, and for a long period in the future, will have only partial observability by PMUs [44].

The FO in both case 3*, and case 4* resonate with the lowest frequency inter-area mode. The system in case 5 has different modes at: 0.614 Hz, 0.708 Hz, 0.741 Hz and 0.78 Hz without creating resonance. There are two FOs sources in case 10, each resonates with a natural mode. Case 3, 4, and 5, and one of the two sources in case 10 are not directly measured by PMU. These cases can be challenging for methods requiring direct measurements at the source.

Fig. 4 shows case 3 and 5, noted that all the oscillations flowed out of the shaded area in red, as explained in Section 3 the source localized in this area. The identification that the source is localized outside/inside of Independent System Operator (ISO) control area is the best possible result using the available PMUs measurements.

Influence of Load and Small Topology Variation: In this experiment, the effectiveness and robustness of the proposed approach are studied against topology variation. We tested the proposed Transformer-Based approach, which trained on several FO cases from the contest dataset without any topology variation included, on several cases with load and topology variation. Transmission lines were set to be outage separately and loads changes were introduced to construct six testing cases. One of the testing cases, FO resonates with a regional inter-area mode at 0.614 Hz. PMU is available where the source is located. Transmission line 4101-4102-2 was set to be outage, and load at bus 6104 was changed from 1700 MW to 1000 MW. According to the testing results, F-1 score, and AUC at branch level are 0.953, 0.991 respectively. The FO localization accuracy maintains 100%.

Influence of Major Topology Variation: Two testing cases were created by removing whole areas, including all generators in the area, which make major changes in the topology, such as removing Colorado zone, Alberta, and British zones, as shown in Fig. 5 where the shaded

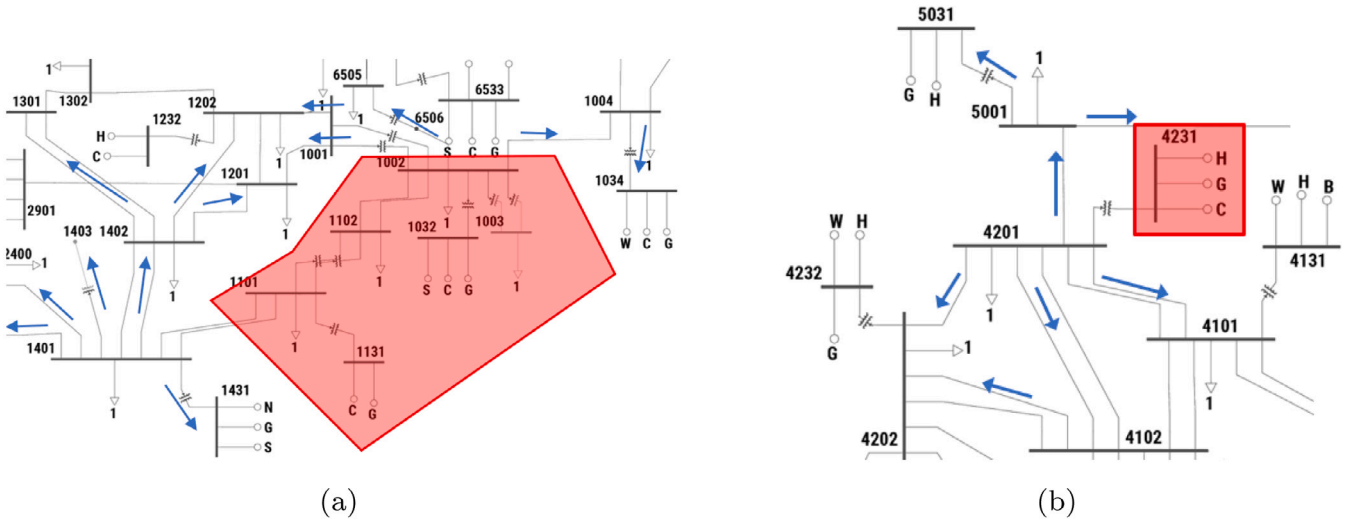

Fig. 4. FO flows in (a) Case 3 (b) Case 5.

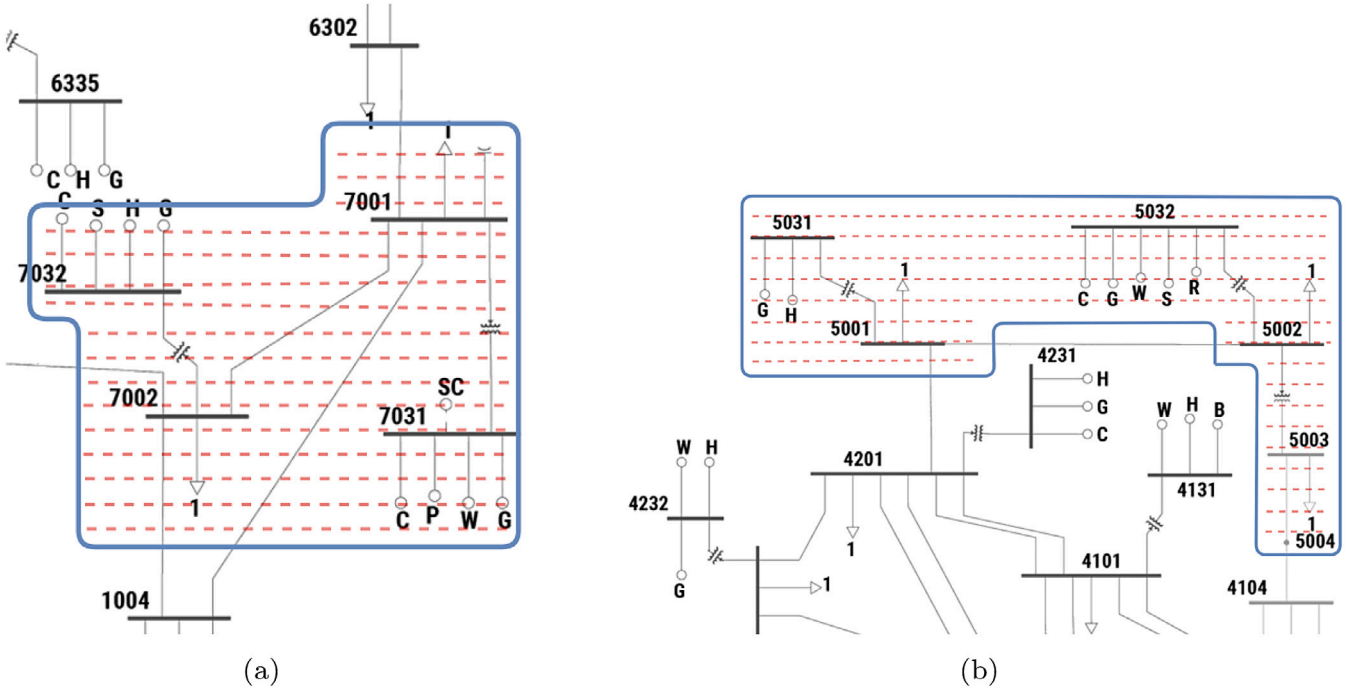

Fig. 5. Topology variation cases (a) Removing Colorado zone (b) Removing Alberta and British zones.

area was removed from the system, and two testing cases were created by disconnecting Bus with all the branches linked to them. According to the testing results, average F-1 score, and AUC at branch level are 0.95, 0.988 respectively. The FO localization accuracy maintains 100%. This experiment shows that the proposed approach is robust enough to handle topology variation of the power system, which make it suitable for real-world applications.

6.2. Ablation studies

Transformer Model Ablation: A comprehensive study was performed to highlight the importance of each component in the proposed approach. As shown in Table 2, four different structures were assessed by removing the components one by one when building the proposed approach to measure their effects on the model. This experiment shows that skip connections and residual connections were useful to improve the model performance.

Feature Ablation: were conducted in which the features were removed one by one when building the proposed approach to measure their effects on the proposed approach. Case 1, Case 2, and Case 14 were used as testing cases to evaluate the performance, whereas Case 2 and 14 were not included in the training data for this experiment.

The results, in terms of F1 score and AUC score at branch level, which were the chosen evaluation metric for this experiment, shows that the best performing was when the approach trained on all available features as shown in Fig. 6. Even though training the approach on all features excludes dissipating energy achieved good results on the same topology, it failed in terms of topology variation.

This experiment provide insights into the relative contribution of each modification on the original Transformer architectural to the performance of proposed Transformer-Based model. It shown how impactful the features are, and justified the used of it.

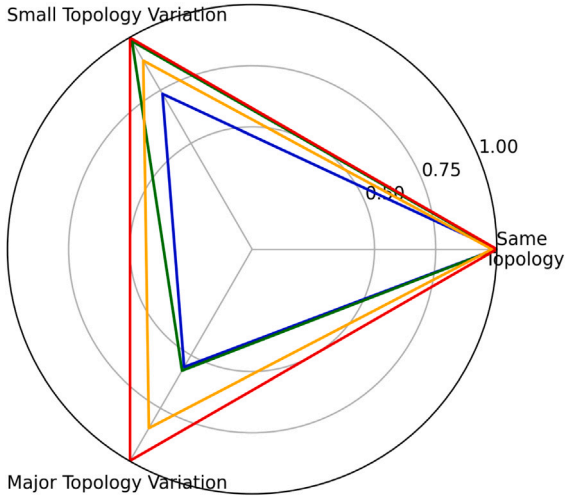

Fig. 6. AUC for different training features at branch level. (——) active, reactive power; (——) active, reactive power, and oscillation frequency; (——) active, reactive power, and dissipating energy; (——) all features.

Table 3
Comparison of performance of different DL models.

DL architecture	Training accuracy (%)	Validation accuracy (%)
Transformer-based (Proposed)	99.94	99.37
CNN	92.63	88.31
LSTM	83.42	77.68

6.3. Comparison of DL models

In this study a quantitative comparison of DL models in the context of FO localization is presented. The well-known CNNs, LSTM in [45] have also been implemented to compare with the results from the proposed Transformer-based model.

Table 3 shows the accuracies of the proposed Transformer-based model and two different DL-based models used for the performance comparison. There was a significant difference between the proposed model and CNNs, LSTM in terms of training and validation accuracy. The comparison results have shown comparable and better performance for he proposed Transformer-based approach without topology variation. In addition, scalability of these models were tested on topology variation cases, but it showed that these DL-models are not scalable nor robust to the topology variations. From this study, we can see that the proposed Transformer-based model has superiority in performance, robustness and scalability.

6.4. Influence of non-Gaussian noise

The proposed approach is tested under different non-Gaussian noise levels with Laplace distribution to the original measurements to ensure that it could localize FO source(s) accurately in the actual power system. The proposed approach shows excellent anti-noise performance due to the contribution of the attention mechanism. This mechanism emphasizes the features, and reducing the noise interference of PMUs measurements. The performance of branch-level localization with varying noise levels is shown in Fig. 7.

6.5. Time efficiency

An essential metric to consider while running deep learning models for online applications is the computation time. A Windows PC with Intel(R) Core(TM) i7-9700 @3.00 GHz, 32 GB memory, NVIDIA GeForce RTX 2080Ti was used to run the trained models.

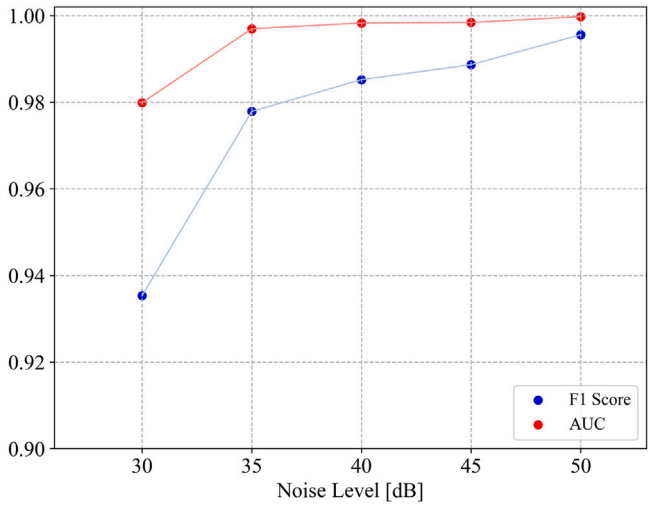

Fig. 7. Localization performance of the proposed approach with different level of noise.

The identifying frequency and calculation of dissipating energy features take an average of 32.24 ms for all branches in the system. For transformer-based deep learning, the total average computing time is 301.2 ms for all branches to solve one whole case which make the proposed approach suitable for online applications. Our approach is fastest compared with ~1 s in [22], ~5 s in [8], and ~4.4 min (without parallelization) or ~33 s (if parallelized on 8 multi-core processors) using DEF-based solution [49].

7. Discussion

Unlike the current deep learning that works on a fixed topology, we propose a branch-level DL-based approach that builds an oscillation propagation map to help network operators localize the FO source. We investigated our approach under different settings to check its scalability and adaptability to real-time applications.

The proposed Transformer-based Deep learning model shows high performance in capturing temporal dependencies compared to traditional DL approaches (CNN and RNN). We refer the high performance to the attention mechanism, which allows the model to understand which parts of the input are important and how relevant each part of the input is to the other parts of the input data [30,50].

As shown in Section 6.1, our approach demonstrates that we can handle major topology variations without retraining our DL models, which indicates that our system can be applied to a new topology settings with minimal supervision. Also, our model showed robustness in handling minor topology variations, which indicates its applicability to real-time scenarios.

The proposed approach indicates robustness against non-Gaussian PMU noise with Laplace distribution, as shown in Section 6.4. Real-life PMU measurements involve noise [22], which may affect or even disable certain PMU applications. This noise can change from time to time due to aging processes, instruments, communication channels, etc. Recent practical studies [46,47] revealed that the noise in PMU measurements tends to follow non-Gaussian distributions such as student-t and Laplace distributions.

As shown in the ablation study implemented in Section 6.2, we find that including FO frequency and DEF in addition to the raw PMU data results in better performance. This finding is consistent with prior research showing that more information helps DL models' scalability.

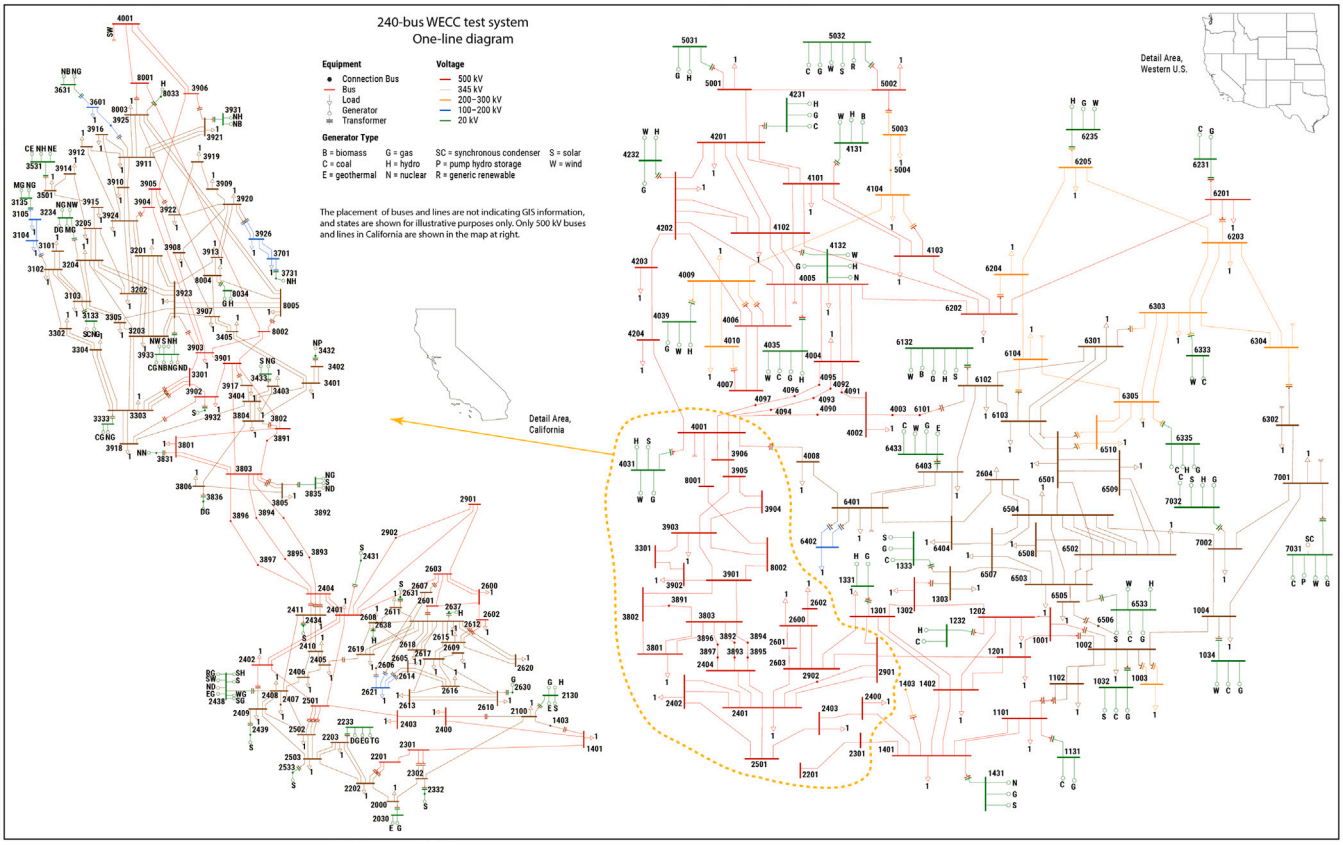

Fig. A.8. One-line diagram of 240-bus WECC system.

8. Conclusion

In this paper we proposed a new branch-level Transformer-based approach inspired by DEF-based methods to localize FO. Our model handles significant shortcomings of the current DL approaches. Our approach is scalable and can handle topology variations and PMU noise. The efficacy of the proposed approach is illustrated in a WECC 240-bus test system, which includes a high renewable integration in the system.

The proposed approach does not require extensive training data, information on system models, or grid topology, resulting in an efficient and easily deployable solution for online operation. Most importantly, it does not need to be re-trained for any topology variations.

The proposed approach achieved high speed, high localization performance even with the presence of non-Gaussian noise, a partially observable system, and operational topology variations which reflect real-world challenges. Future work will focus on identifying the controller type of oscillation source(s).

CRediT authorship contribution statement

Mustafa Matar: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Data curation. **Pablo Gill Estevez:** Data Curation, Software. **Pablo Marchi:** Data Curation, Writing - review & editing. **Francisco Messina:** Data curation, Writing – review & editing. **Ramadan Elmoudi:** Writing – review & editing. **Safwan Wshah:** Conceptualization, Methodology, Supervision, Validation, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is available to the public on IEEE-NASPI Oscillation Source Location (OSL) Contest.

Acknowledgment

Computations were performed on the Vermont Advanced Computing Core supported in part by the National Science Foundation (NSF), United States award No. OAC-1827314.

Appendix. Model of the 240-bus WECC test system

A one-line diagram of the 240-bus WECC test system is shown in Fig. A.8. The 240-bus WECC test system models can be downloaded from NREL's test case repository [51].

References

[1] NERC. Forced oscillation monitoring & mitigation. Atlanta, GA, USA: North American Electric Reliabilty Corporation; 2017.
[2] Chen C, Yang H, Wang W, Mandich M, Yao W, Liu Y. Harmonic transmission characteristics for ultra-long distance AC transmission lines based on frequency-length factor. Electr Power Syst Res 2020;182:106189.
[3] Liu G, Venkatasubramanian V. Oscillation monitoring from ambient PMU measurements by frequency domain decomposition. In: 2008 IEEE international symposium on circuits and systems. IEEE; 2008, p. 2821–4.
[4] Wang B, Sun K. Location methods of oscillation sources in power systems: a survey. J Mod Power Syst Clean Energy 2017;5(2):151–9.
[5] Yu Y, Min Y, Chen L, Ju P. The disturbance source identification of forced power oscillation caused by continuous cyclical load. In: 2011 4th International conference on electric utility deregulation and restructuring and power technologies (DRPT). IEEE; 2011, p. 308–13.
[6] Zhou N, Dagle J. Initial results in using a self-coherence method for detecting sustained oscillations. IEEE Trans Power Syst 2014;30(1):522–30.

[7] Zhang QF, Luo X, Litvinov E, Dahal N, Parashar M, Hay K, Wilson D. Advanced grid event analysis at ISO New England using PhasorPoint. In: 2014 IEEE PES general meeting| conference & exposition. IEEE; 2014, p. 1–5.

[8] Maslennikov S, Wang B, Litvinov E. Dissipating energy flow method for locating the source of sustained oscillations. Int J Electr Power Energy Syst 2017;88:55–62.

[9] Dong Q, Liang J, Yan X, Yang R. Locating method of disturbance source of low frequency oscillation in large scale power grid. In: Zhongguo dianji gongcheng xuebao(proceedings of the chinese society of electrical engineering), vol. 32. (1):Chinese Society for Electrical Engineering; 2012, p. 78–83.

[10] Demello FP, Concordia C. Concepts of synchronous machine stability as affected by excitation control. IEEE Trans Power Appar Syst 1969;88(4):316–29.

[11] Gao Y, Liu D-c, Huang G-b, Shi Q-y. Locating method of disturbance source of forced power oscillation based on prony anyasis. In: 2012 China international conference on electricity distribution. IEEE; 2012, p. 1–4.

[12] Li Y, Huang Y, Liu J, Yao W, Wen J. Power system oscillation source location based on damping torque analysis. Power Syst Prot Control 2015;43(14):84–91.

[13] Zuhaib M, Rihan M, Saeed MT. A novel method for locating the source of sustained oscillation in power system using synchrophasors data. Prot Control Modern Power Syst 2020;5(1):1–12.

[14] Zheng X, et al. The complex torque coefficient approach's applicability analysis and its realization by time domain simulation. Proc-Chine Soc Electr Eng 2000;20(6):1–4.

[15] Al-Ashwal N, Wilson D, Parashar M. Identifying sources of oscillations using wide area measurements. In: Proceedings of the CIGRE US national committee 2014 grid of the future symposium, houston, vol. 19. 2014.

[16] Huang T, Freris NM, Kumar P, Xie L. A synchrophasor data-driven method for forced oscillation localization under resonance conditions. IEEE Trans Power Syst 2020;35(5):3927–39.

[17] Chen L, Min Y, Hu W. An energy-based method for location of power system oscillation source. IEEE Trans Power Syst 2012;28(2):828–36.

[18] Maslennikov S, Litvinov E. ISO New England experience in locating the source of oscillations online. IEEE Trans Power Syst 2020;36(1):495–503.

[19] Meng Y, Yu Z, Lu N, Shi D. Time series classification for locating forced oscillation sources. IEEE Trans Smart Grid 2021;12(2):1712–21. http://dx.doi.org/10.1109/TSG.2020.3028188.

[20] Wang B, Maslennikov S. IEEE-NASPI oscillation source location contest-case development and results. Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States); 2021.

[21] Banna HU, Solanki SK, Solanki J. Data-driven disturbance source identification for power system oscillations using credibility search ensemble learning. IET Smart Grid 2019;2(2):293–300.

[22] Feng S, Chen J, Ye Y, Wu X, Cui H, Tang Y, Lei J. A two-stage deep transfer learning for localisation of forced oscillations disturbance source. Int J Electr Power Energy Syst 2022;135:107577.

[23] Talukder S, Liu S, Wang H, Zheng G. Low-frequency forced oscillation source location for bulk power systems: A deep learning approach. In: 2021 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2021, p. 3499–504.

[24] Kundur P. Power systems stability and control. In: Conference proceedings. New York: McGraw-Hi1l; 1994.

[25] [Online] Available: https://www.ge.com/digital/applications/transmission/phasorpoint, Accessed: 2022-27-05.

[26] Maslennikov S, Wang B, Litvinov E. Locating the source of sustained oscillations by using PMU measurements. In: 2017 IEEE power & energy society general meeting. IEEE; 2017, p. 1–5.

[27] Estevez PG, Marchi P, Galarza C, Elizondo M. Non-stationary power system forced oscillation analysis using synchrosqueezing transform. IEEE Trans Power Syst 2020;36(2):1583–93.

[28] Gill Estevez P, Marchi P, Galarza CG, Messina F. Forced oscillation identification and filtering from multi-channel time-frequency representation. IEEE Trans Power Syst 2022;1. http://dx.doi.org/10.1109/TPWRS.2022.3172850.

[29] Moon Y-H, Cho B-H, Lee Y-H, Kook H-J. Derivation of energy conservation law by complex line integral for the direct energy method of power system stability. In: Proceedings of the 38th IEEE conference on decision and control (cat. no. 99ch36304), vol. 5. IEEE; 1999, p. 4662–7.

[30] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;30.

[31] Oord Avd, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K. Wavenet: A generative model for raw audio. 2016, arXiv preprint arXiv:1609.03499.

[32] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.

[33] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI conference on artificial intelligence. 2017.

[34] Matar M, Xu B, Elmoudi R, Olatujoye O, Wshah S. A deep learning-based framework for parameters calibration of power plant models using event playback approach. IEEE Access 2022;10:72132–44. http://dx.doi.org/10.1109/ACCESS.2022.3188313.

[35] Ba JL, Kiros JR, Hinton GE. Layer normalization. 2016, arXiv preprint arXiv:1607.06450.

[36] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13(2).

[37] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.

[38] Junker M, Hoch R, Dengel A. On the evaluation of document analysis components by recall, precision, and accuracy. In: Proceedings of the fifth international conference on document analysis and recognition. ICDAR'99 (Cat. No. PR00318). IEEE; 1999, p. 713–6.

[39] Wang B, Kenyon RW, Tan J. Developing a pscad model of the reduced 240-bus wecc test system. Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States); 2022.

[40] Transient security assessment tool (TSAT). British Columbia, Canada: Powertech Labs Inc., URL https://powertechlabs.com/tsat/.

[41] Price JE, Goodin J. Reduced network modeling of WECC as a market design prototype. In: 2011 IEEE power and energy society general meeting. IEEE; 2011, p. 1–6.

[42] Yuan H, Biswas RS, Tan J, Zhang Y. Developing a reduced 240-bus WECC dynamic model for frequency response study of high renewable integration. In: 2020 IEEE/PES transmission and distribution conference and exposition (T&D). IEEE; 2020, p. 1–5.

[43] Myrda P, Farantatos E, Patel M, Hooshyar H. Summary of EPRI synchrophasor related activities. Oct, 2019.

[44] Maslennikov S, Wang B. Creation of simulated test cases for the oscillation source location contest. Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States); 2022.

[45] Wshah S, Shadid R, Wu Y, Matar M, Xu B, Wu W, Lin L, Elmoudi R. Deep learning for model parameter calibration in power systems. In: 2020 IEEE international conference on power systems technology (POWERCON). 2020, p. 1–6. http://dx.doi.org/10.1109/POWERCON48463.2020.9230531.

[46] Wang S, Zhao J, Huang Z, Diao R. Assessing Gaussian assumption of PMU measurement error using field data. IEEE Trans Power Deliv 2017;33(6):3233–6.

[47] Huang C, Thimmisetty C, Chen X, Stewart E, Top P, Korkali M, Donde V, Tong C, Min L. Power distribution system synchrophasor measurements with non-Gaussian noises: Real-world data testing and analysis. IEEE Open Access J Power Energy 2021;8:223–8.

[48] Kummerow A, Dirbas M, Monsalve C, Nicolai S, Bretschneider P. Robust disturbance classification in power transmission systems with denoising recurrent autoencoders. Sustain Energy, Grids Netw 2022;32:100803.

[49] Gill Estevez P, Marchi P, Galarza CG. 2021 IEEE-NASPI oscillation source location contest. School of Engineering, University of Buenos Aires (FIUBA), URL https://www.naspi.org/sites/default/files/2021-10/D3S8_02_estevez_fiuba.pdf.

[50] Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: A survey. ACM Comput Surv 2020.

[51] Test case repository for high renewable study. 2022, [Online] Available: https://www.nrel.gov/grid/test-case-repository.html, Accessed: 2022-04-05.