



Event Detection and Identification in Distribution Networks Based on Invertible Neural Networks and Pseudo Labels

Fan Yang¹, Zenan Ling^{2,3}, Yuhang Zhang¹, Xing He^{1*}, Qian Ai¹ and Robert C. Qiu⁴

¹Department of Electrical Engineering, State Energy Smart Grid Research and Development Center, Shanghai Jiaotong University, Shanghai, China, ²Key Laboratory of Machine Perception (MoE), School of EECS, Peking University, Beijing, China, ³Pazhou Laboratory, Guangzhou, China, ⁴School of Electronic Information and Communication, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:

Bo Yang,
Kunming University of Science and
Technology, China

Reviewed by:

Fang Shi,
Shandong University, China
Yunfei Ma,
Zhejiang University, China

*Correspondence:

Xing He
hexing_hx@126.com

Specialty section:

This article was submitted to
Smart Grids,
a section of the journal
Frontiers in Energy Research

Received: 20 January 2022

Accepted: 07 February 2022

Published: 17 March 2022

Citation:

Yang F, Ling Z, Zhang Y, He X, Ai Q and
Qiu RC (2022) Event Detection and
Identification in Distribution Networks
Based on Invertible Neural Networks
and Pseudo Labels.
Front. Energy Res. 10:858665.
doi: 10.3389/fenrg.2022.858665

Anomalous event detection and identification are important to support situational awareness and security analysis in power grids. Particularly, the distribution network is with complicated topology, variable load behaviors, and integration of nonlinear distributed generators (DGs), which is difficult to implement complete modeling mathematically. With the deployment of advanced measurement devices such as μ PMUs in distribution networks, massive data containing rich system status information becomes available. In this paper, a framework for event detection, localization, and classification is studied to extract event features from measurements in distribution networks. Specifically, a method based on an invertible neural network (INN) is employed to model the complex distributions of normal-state measurements offline in a flexible way. It then establishes explicit likelihoods as the indicator to enable real-time event detection. Furthermore, a Jacobian-based method is utilized for spatial localization. Finally, as the events in practical power grids are mostly recorded unlabeled, the pseudo label (PL) based approach, superior in the separating ability for events under a low labeling rate, and is used to implement event classification. Several typical types of events simulated in the IEEE 34-bus system and real-world cases in a low-voltage system verify the effectiveness and superiorities of the framework.

Keywords: distribution network, data-driven, event detection and localization, event classification, invertible neural network, pseudo labels

1 INTRODUCTION

In power grids, anomalous events refer to incidents that violate well-defined normal operating conditions. The detection and identification of them are important to support situational awareness and security analysis in power grids. In distribution networks, anomalous events are mainly composed of short-circuit faults and tripping events, which can cause the voltages and currents to exceed limits, be out of allowed ranges, and generate asymmetries. Lack of monitoring to these events could fail to make necessary and immediate responses, decreasing the safety, reliability, and quality of power supply, and even leading to more serious contingencies (Samuelsson et al., 2006). Therefore, accurately detecting events, identifying their locations, and determining their classifications are essential, so that the system status can be comprehensively assessed and proper actions can be taken before any sporadic event escalates to worse effects.

Traditional model-based approaches for event recognition are usually aimed at a certain event signal or topology. Event characteristics are analyzed based on different levels of assumptions and simplifications (Wang et al., 2018) (Wei et al., 2021). However, these approaches are difficult to model each type of event completely and accurately, and are not adaptable to power systems' complex, and changeable operation status (Song et al., 2015).

To cope with the complexity and uncertainty of system operations, constructing smart distribution networks has been accelerated, which aims to improve real-time monitoring, situational awareness, and rapid control. With the background, the large-scale deployment of measuring devices, such as μ PMUs, has been promoted, and allowing for the real-time transmission of massive data in distribution networks. Data-driven approaches of event analysis utilize the rich information contained in signals, relying on no assumptions or simplifications of the system modeling. They can generally provide better robustness to the variations of systems' topologies and operations, thus having an extensive application prospect.

In literature, various data-driven approaches have been applied in the area of event analysis. The principal component analysis (PCA) is used in (Xie et al., 2014) to reduce the dimension during the feature extraction for event detection. In (Ahmed et al., 2021), event detection, localization, and classification are implemented by utilizing the deep autoencoder (DAE). The features of cascading events are analyzed and trained by a shallow convolutional neural network (CNN) in (Li and Wang, 2019). In (Wang et al., 2019), the measurements at the normal state are modeled by a one-class support vector machine (OCSVM) hence realizing the event detection. An enhanced long short-term memory (LSTM) network is used in (Li et al., 2021) to implement the fast event detection of a system containing renewable energy. In (Liu et al., 2019), an approach is proposed based on the local outlier factor (LOF) to detect and locate events using reduced PMU data. In (He et al., 2019), invisible power usage events are detected by high-dimensional statistics in random matrix theory (RMT). In (Pandey et al., 2020), density-based spatial clustering is applied to classify events into short circuit faults and those caused by a significant imbalance of active and reactive powers, and by identifying the types of disturbed measurements.

However, how to appropriately use the online measurements and realize the event detection, localization, and classification in a more effective way deserves further consideration. For the existing data-driven approaches, some limitations exist:

- 1) Feature selection is not paid attention to, especially for the event classification. Various measurements exhibit different characteristics, but they are usually utilized without more considerations of applicability. For example, voltage magnitudes are utilized in (Tong et al., 2021) or together with current magnitudes in (Wilson et al., 2020), but their changes are indefinite and can confuse events on some occasions.
- 2) Parameters or thresholds are required to be preset, and they are strongly depended on by some methods (Xie et al., 2014;

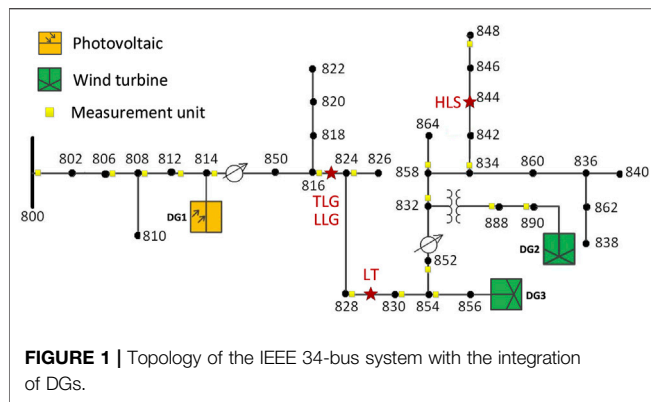
Wang et al., 2019; Ahmed et al., 2021). The optimal settings are hard to adapt to all datasets.

- 3) Unlike transmission networks, statistical properties of the fluctuated measurements in distribution networks cannot be approximated as a Gaussian distribution or other typical distributions. More nonlinearities and uncertainties are exhibited, so the theoretical basis of many methods is invalid.
- 4) Measurements of practical power systems exhibit significant imbalance, which means the measurements obtained at normal states are significantly larger than those obtained at anomalous states. Besides, only a few events are identified and labeled by operators (about 2%) (Wilson et al., 2020). It hinders the use of supervised approaches (Li and Wang, 2019; Yadav et al., 2019; Li et al., 2021), while unsupervised approaches (Pandey et al., 2020; Wilson et al., 2020; Ahmed et al., 2021) can only make identifications roughly.

To cope with the above problems, a semi-supervised framework is studied and employed for event detection, localization, and classification in distribution networks by taking advantage of invertible neural networks (INNs) and pseudo labels (PLs). Offline training is conducted using the INN in (Kingma and Dhariwal, 2018) to learn the distribution of measurements obtained at normal states. The explicit likelihoods can be calculated for event detection, and an input-output Jacobian is utilized for event localization. Then a CNN-and-PL-based approach is explored for event classification. Contributions of this paper are summarized as follows.

- 1) Based on INNs, the framework can effectively model the complex distributions of measurements obtained at normal states, so as to detect events reliably, and sensitively in distribution networks.
- 2) The event classification is based on accurate event localization, so the exact signal features around the event location can be utilized, supporting the more precise, and reliable event classification. Further, the combination of voltages/currents and differential currents/voltages is utilized and verified to possess an enhanced ability to distinguish between several principal events in DG-integrated distribution networks.
- 3) The event analysis, especially the event classification, under the low labeling rate of measurements is figured out by the CNN-and-PL-based approach. The significant advantages over other approaches in solving this problem have been verified in distribution networks.

The rest of this paper is organized as follows. In **Section 2**, the characteristics of various kinds of measurements are illustrated when different events occur. Requirements for event analysis are also discussed. In **Section 3**, a semi-supervised framework is studied for event detection, localization, and classification in distribution networks with the integration of DGs. Case studies are conducted in **Section 4**, where both simulated and real-world data are utilized to make the verifications. Finally, conclusions are given in **Section 5**.



2 PROBLEM FORMULATION

Different events will make voltages, currents, or other measurements exhibit different characteristics. Selecting various measurements or their combinations to carry out event analysis will make variable influences on the sensibility and reliability. In this section, considering the characteristics of distribution networks, the representative features of different kinds of measurements are analyzed, and a specific combination is selected for event classification. In addition, the limitations of some typical methods to learn and model the behaviors of real-world measurements are illustrated, and requirements of methods for event detection and classification are discussed.

2.1 Selection of Measurements

Three-phase voltages and currents are usually used for event detection in data-driven approaches, as they effectively reflect the operating status and can be directly obtained by online monitoring devices. However, limitations exist when inappropriately using these measurements for event classification.

Some work utilizes voltage magnitudes for event classification (Tong et al., 2021), and some combine the voltages with currents (Wilson et al., 2020). In this section, the characteristics of these measurements are analyzed when four typical events happen in the IEEE 34-bus system, including three-line-to-ground fault (TLG), line-to-line-to-ground fault (LLG), heavy load switching-in event (HLS), and line trip (LT). The topology is shown in **Figure 1** with positions of assumed events marked. Three DGs are integrated into the system, i.e., a photovoltaic (PV) at Bus 814, two doubly-fed induction generators (DFIGs) at Bus 856 and Bus 890. For LLG, disturbed phases are set as phases A and B, and the LT is assumed as a three-phase event. A heavy load of 0.35 MW is switched in at Bus 844 for the HLS. The outputs of the PV at Bus 814, the DFIG at Bus 890, and the DFIG at Bus 856 are 0.25, 0.776, and 0.703 MW, respectively. In this situation, the penetration rate of DGs is 48.78%.

Changes of measurements are listed in **Table 1**. For phase A, magnitudes of voltages at both ends (U_{a1} and U_{a2}), currents (I_a), differential currents (ΔI_a) on the disturbed branch, and differential voltages (ΔU_a) are listed. Herein, ΔI_a is calculated by the sum of current phases at both ends, and ΔU_a is the voltage

difference between the voltage phases at the two ends. They reflect the leakage current and the voltage drop on the branch, respectively. Curves of $T - U_{a1} - I_a$, $T - U_{a1} - \Delta I_a$, and $T - U_{a1} - \Delta U_a$ are plotted. It can be observed that only voltage and current magnitudes cannot identify some certain events like HLS and LT. This is because the integration of DGs and the branches existing between two measurement units will make the power flow and the caused voltage drop uncertain on various conditions, including various capacities and positions of DGs, line parameters, load levels, imbalance degrees, and disturbance intensities of events, etc. To this end, only voltage or current magnitudes cannot perform well in event classification. According to the theoretical analysis and comprehensive simulations, a combination of three-phase voltages, currents, differential currents, and differential voltages is demonstrated to be capable of effectively distinguishing between TLG, LLG, HLS, and LT. The characteristics of these measurements under the four events are summarized in **Table 1**. Therefore, in this paper, such measurement combination will serve as the selected features to implement the event classification.

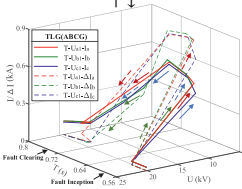
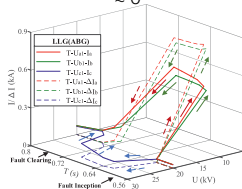
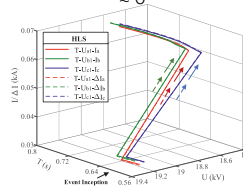
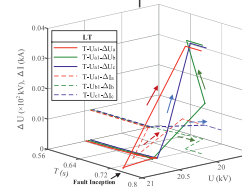
2.2 Requirements for Event Detection and Classification

2.2.1 Event Detection

Figure 2 shows a typical topology of a medium and low voltage distribution network, where online monitoring data is collected from measurement units distributed in the network. **Figure 3A** shows three-phase voltage magnitudes recorded at load-side transformers in region A. The sampling interval between every two measurements is 15 min. Since voltage magnitudes are closely related to load levels, curves in **Figure 3A** exhibit a typical daily pattern, i.e., low voltage in the day and early night for the heavy load, whereas high voltage at midnight for the light load. In addition, voltage measurements show different details between days: fluctuation amplitudes, shapes, and presence of spikes, etc., which are caused by load switching and changes of operating states. The complex, nonlinear, and dynamic characteristics make the modeling of real-world measurements challenging. As a result, methods extracting simple features for event detection malfunction in some situations.

Here, a DAE-based approach (Ahmed et al., 2021) and a PCA-based approach (Xie et al., 2014) are utilized to detect the faults marked in **Figure 3A**. **Figure 3B** shows their detection indicators, i.e., Z-score and mean absolute error (MAE). In **Figure 3B**, Z-score identifies the fault on April 5th with a significant voltage drop but misses the fault on April 4th. This is because the simple structure of DAE cannot model complex distributions of real-world measurements effectively, and the indicator is not sensitive enough. Besides, the detection threshold (a constant, i.e., three) set in (Ahmed et al., 2021) is questionable because a fixed threshold is hard to be appropriate for all situations. In **Figure 3B**, MAE is significantly affected by a pre-defined parameter, i.e., cumulative variance percentage (CVP). When the CVP is selected as 98.5%, 99%, and 99.5%, PCA cannot accurately detect the two faults in **Figure 3A**. PCA is a linear

TABLE 1 | Characteristics of measurements in different events.

Events	TLG (ABCG)	LLG (ABG)	HLS	LT
Signals ¹	$R = 10\Omega, R = 50\Omega$	$R = 10\Omega, R = 50\Omega$		
U_{a1}	↓(Inception) ↑(Clearing) ²	↓ ↑	Indefinite	Indefinite
U_{a2}	↓ ↑	↓ ↑	Indefinite	Indefinite
I_a	↑ ↓	↑ ↓	Indefinite	Indefinite
U_{b1}	↓ ↑	↓ ↑	Indefinite	Indefinite
U_{b2}	↓ ↑	↓ ↑	Indefinite	Indefinite
I_b	↑ ↓	↑ ↓	Indefinite	Indefinite
U_{c1}	↓ ↑	↓ ↑	Indefinite	Indefinite
U_{c2}	↓ ↑	↓ ↑	Indefinite	Indefinite
I_c	↑ ↓	↑ ↓	Indefinite	Indefinite
ΔI_a	↑ ↓	↑ ↓	↑	Indefinite
ΔI_b	↑ ↓	↑ ↓	↑	Indefinite
ΔI_c	↑ ↓	↑ ↓	↑	Indefinite
ΔU_a	↑ ↓	≈ 0	≈ 0	↑
ΔU_b	↑ ↓	≈ 0	≈ 0	↑
ΔU_c	↑ ↓	≈ 0	≈ 0	↑
Example				

¹ U_{a1} and U_{a2} denote voltage magnitudes of phase A at both ends; I_a and ΔI_a denote currents and differential currents of phase A on the disturbed branch; ΔU_a denotes differential voltages and is calculated by $U_{a1} - U_{a2}$.

²Symbols of ↓ and ↑ denote the decrease and increase of signals after the inception and clearing of events. "Indefinite" means the change of signal is uncertain.

dimension reduction method and cannot effectively deal with nonlinear measurements. Also, a proper CVP is hard to find in advance for all datasets. To this end, two aspects require attention for event detection algorithms in distribution networks: 1) the ability to model complex and nonlinear real-world measurements; 2) the robustness to pre-defined parameters.

2.2.2 Event Classification

Supervised approaches for event classification are dependent on large amounts of labeled data for training, such as (Li et al., 2021) and (Yadav et al., 2019). However, only about 2% of the total number of recorded events are labeled by the operators in a hand-crafted way (Wilson et al., 2020), which hinders their practical applications. Unsupervised approaches require no

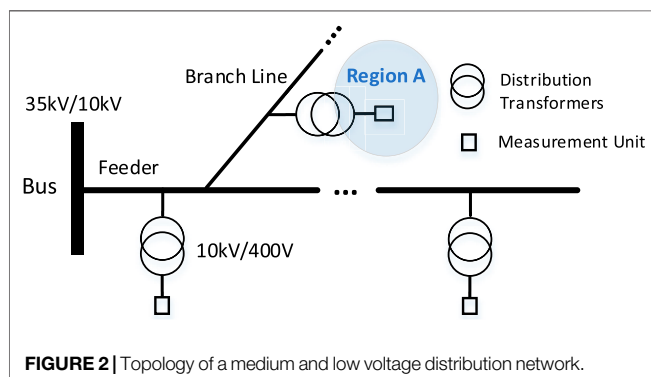
prior labeling of samples, but can only classify events roughly. Examples include (Wilson et al., 2020) and (Ahmed et al., 2021), which can only distinguish the number of disturbed phases but cannot further determine the specific type of events. Besides, active and reactive events are identified in (Ahmed et al., 2021) and (Pandey et al., 2020) simply by the category of disturbed measurements. In contrast, semi-supervised approaches simultaneously utilize labeled and unlabeled data, and thus they can realize refined classification with only a limited number of labeled samples. Therefore, semi-supervised approaches are preferable for event classification in practical applications.

3 EVENT DETECTION, LOCALIZATION, AND CLASSIFICATION BASED ON INVERTIBLE NEURAL NETWORKS AND PSEUDO LABELS

In this section, a framework is introduced for event detection, localization, and classification based on INNs and PLs. Event detection and localization are realized by INNs, and a PL-based is utilized to classify the events with measurements obtained at disturbed locations.

3.1 Likelihood-Based Event Detection

Likelihoods measure the probability that a sample belongs to a certain distribution. If a sample follows the distribution, the

**FIGURE 2** | Topology of a medium and low voltage distribution network.

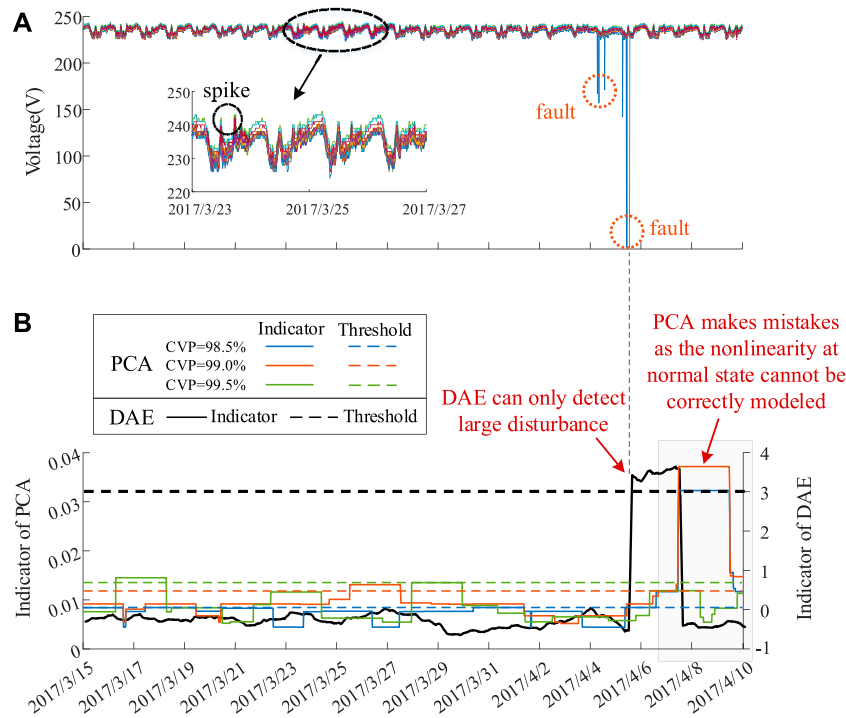


FIGURE 3 | (A) Real-world measurements of three-phase voltage magnitudes obtained at load-side transformers; **(B)** detection results of DAE and PCA under different CVPs.

likelihood is high, and vice versa (Myung, 2003). In power grids, normal measurements are abundant whereas there is little anomalous data. A straightforward idea for event detection is that distributions of normal measurements are first learned and parameterized. At monitoring time, likelihoods of unseen measurements are calculated under the learned distribution. Low likelihoods indicate the occurrence of events.

Assume that $\mathbf{Z} \in \mathbb{R}^D$ is the random variable representing distributions of normal measurements, i.e., the target distribution we need to model. Let $\mathbf{Y} \in \mathbb{R}^D$ be a random variable with a known and tractable probability density function (PDF) $p_Y(\mathbf{y})$ and $\mathbf{Z} = \mathbf{f}(\mathbf{Y})$, where \mathbf{f} is an invertible function. Using the change of variables formula (Dinh et al., 2014), one can compute the PDF of the random variable \mathbf{Z} by

$$p_Z(\mathbf{z}) = p_Y(\mathbf{g}(\mathbf{z})) \left| \det \frac{\partial \mathbf{g}}{\partial \mathbf{z}} \right|, \quad (1)$$

where \mathbf{g} is the inverse of \mathbf{f} , $\frac{\partial \mathbf{g}}{\partial \mathbf{z}}$ is the Jacobian of \mathbf{g} , \det means determinant calculation, and $|\cdot|$ means absolute value operation. In Eq. 1, the function \mathbf{f} “pushes forward” the base density $p_Y(\mathbf{y})$ to a more complex density $p_Z(\mathbf{z})$.

Further, assume that the base density $p_Y(\mathbf{y})$ and the function \mathbf{f} are parameterized by vectors ϕ and θ . Given a set of normal measurements (denoted as $\mathcal{D} = \{\mathbf{z}_i\}_{i=1}^M$), we can perform a likelihood-based estimation of parameters $\Theta = (\theta, \phi)$ by Eq. 1. Note that in this case, only normal measurements $\mathcal{D} = \{\mathbf{z}_i\}_{i=1}^M$ can be observed, whereas parameters $\Theta = (\theta, \phi)$ need to be estimated. The log-likelihood is formulated as

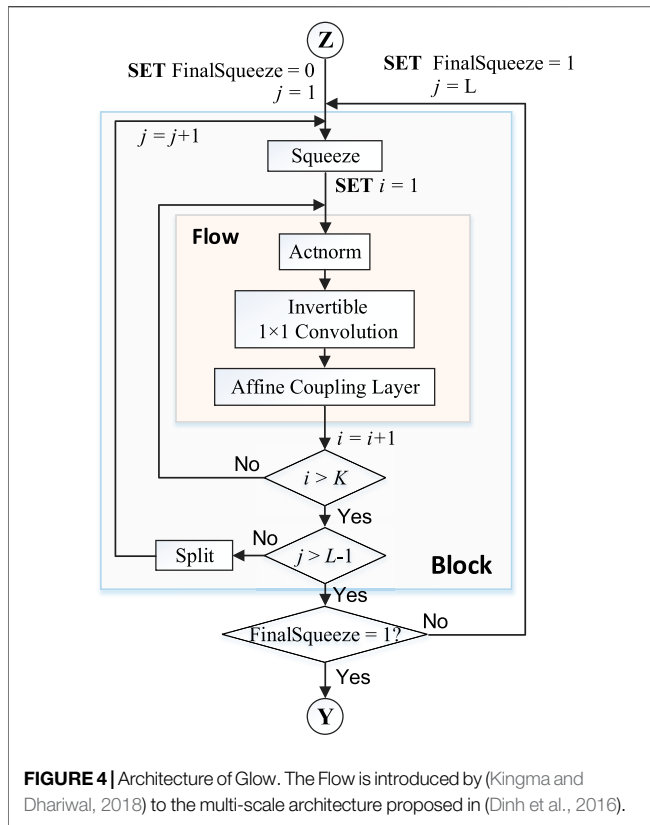
$$\begin{aligned} \log p(\mathcal{D} | \Theta) &= \sum_{i=1}^M \log p_Z(\mathbf{z}_i | \Theta) \\ &= \sum_{i=1}^M \log p_Y(\mathbf{g}(\mathbf{z}_i | \theta) | \phi) + \log \left| \det \frac{\partial \mathbf{g}(\mathbf{z}_i | \theta)}{\partial \mathbf{z}_i} \right|, \end{aligned} \quad (2)$$

where the first term is the log-likelihood of normal measurements under the base density, and the second term (frequently called the log-determinant or volume correction) accounts for the change of volume induced by the transformation \mathbf{g} .

The main procedure for event detection includes two steps. Firstly in the training phase, parameters of the function \mathbf{f} (i.e., θ) and the base density $p_Y(\mathbf{y})$ (i.e., ϕ) are adjusted to maximize the log-likelihood $\log p(\mathcal{D} | \Theta)$, so that distributions of normal measurements can be well modeled. Secondly for online applications, the learned model assigns different likelihoods to unseen measurements by Eq. 2, and low likelihoods indicate the occurrence of events. It is noted that to obtain explicit log-likelihoods $\log p(\mathcal{D} | \Theta)$ in Eq. 2, the existence of \mathbf{g} is necessary. That is, the transformation function \mathbf{f} needs to be invertible. INN is an appropriate tool that allows for this requirement and thus is natural for likelihood-based event detection.

3.2 Invertible Neural Networks

INNs can model complex distributions from a simple base distribution via a set of invertible and differentiable transformations. Hence, they process remarkable representation abilities for complex, nonlinear measurements obtained in the real world. For INNs, efficient



calculation of log-determinant is particularly important because they are repeatedly computed in Eq.2 during training. In this paper, we utilize a computationally efficient model named Glow despite various architectures of INNs (Kingma and Dhariwal, 2018). Glow introduces Flow (Kingma and Dhariwal, 2018) to the multi-scale architecture proposed in (Dinh et al., 2016). In Figure 4, inputs (i.e., normal measurements \mathbf{Z}) are first squeezed by the squeeze layer to permute the dimension. Subsequently there are K Flows, and each Flow contains three components:

- Actnorm layer: Actnorm is short for activation normalization. It performs an affine transformation of inputs using a scale and bias parameter, such that the outputs per channel have zero mean and unit variance.
- Invertible 1×1 convolution: Permutation of dimensions is necessary for flows to ensure that dimensions can affect each other after sufficient steps of the Flow. A 1×1 convolution with an equal number of input and output channels is equivalent to a permutation operation of dimensions and can be computationally efficient (Kingma and Dhariwal, 2018). The log-determinant of an invertible 1×1 convolution of an $h \times w \times c$ tensor \mathbf{h} with $c \times c$ weight matrix \mathbf{W} is

$$\log \left| \det \left(\frac{d \text{conv2D}(\mathbf{h}; \mathbf{W})}{d\mathbf{h}} \right) \right| = h \cdot w \cdot \log |\det(\mathbf{W})|. \quad (3)$$

The cost of computing $\det(\mathbf{W})$ is $\mathcal{O}(c^3)$, but can be reduced to $\mathcal{O}(c)$ by parameterizing \mathbf{W} directly in its LU decomposition.

- Affine coupling layer: Glow follows the computationally efficient affine coupling layer introduced in (Dinh et al., 2014), which consists of split and concatenation, a nonlinear mapping, and a permutation.

In Figure 4, the squeeze layer, K flows, and the split layer (reverse of the squeeze layer) are collectively called a block. The multi-scale architecture contains $L - 1$ whole blocks and one block without the split layer. Finally, the output of the multi-scale architecture are known random variables \mathbf{Y} . More details of Glow can be found in (Dinh et al., 2014; Dinh et al., 2016; Kingma and Dhariwal, 2018).

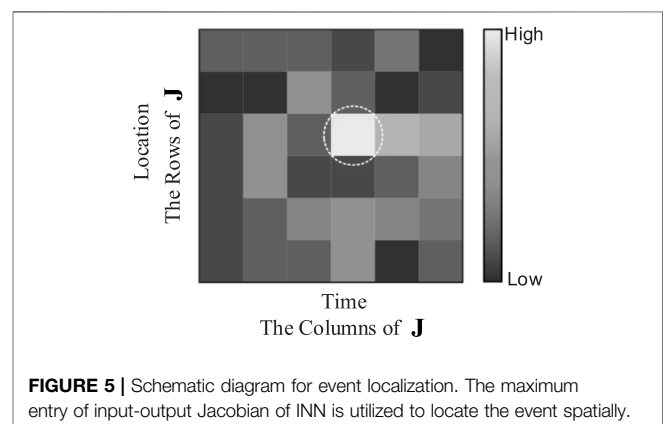
3.3 Event Localization Using Input-Output Jacobian

For practical applications, online measurements (such as three-phase voltage magnitudes) truncated by moving windows are obtained as input samples of INNs, so that explicit likelihoods can be calculated in real-time for situational awareness. Let the column vector $\mathbf{x}_t \in \mathbb{C}^N$ contain measurement variables of N monitoring channels at sampling point t , i.e., $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})^H$. When the length of the moving window is set as T , the observation matrix $\mathbf{X}_t \in \mathbb{C}^{N \times T}$ is generated as

$$\mathbf{X}_t = (\mathbf{x}_{t-T+1}, \mathbf{x}_{t-T+2}, \dots, \mathbf{x}_t). \quad (4)$$

Denote the likelihood estimated by the trained INN as P_Θ . As is described in Section 3.1, the trained INN assigns lower likelihoods to abnormal samples than normal ones. For moving windows, once the likelihood is lower than a decision boundary (DB), events are deduced to occur, and it requires further analysis.

To spatially locate the detected event, an input-output Jacobian is calculated by the trained INN, so that the monitoring channel that contributes the most to the low likelihood can be determined. Note that $x_{i,k}$ contained in Eq.4 is the measurement obtained in the i th monitoring channel at the k th sampling point. Then we can measure the contribution of $x_{i,k}$ to the output by



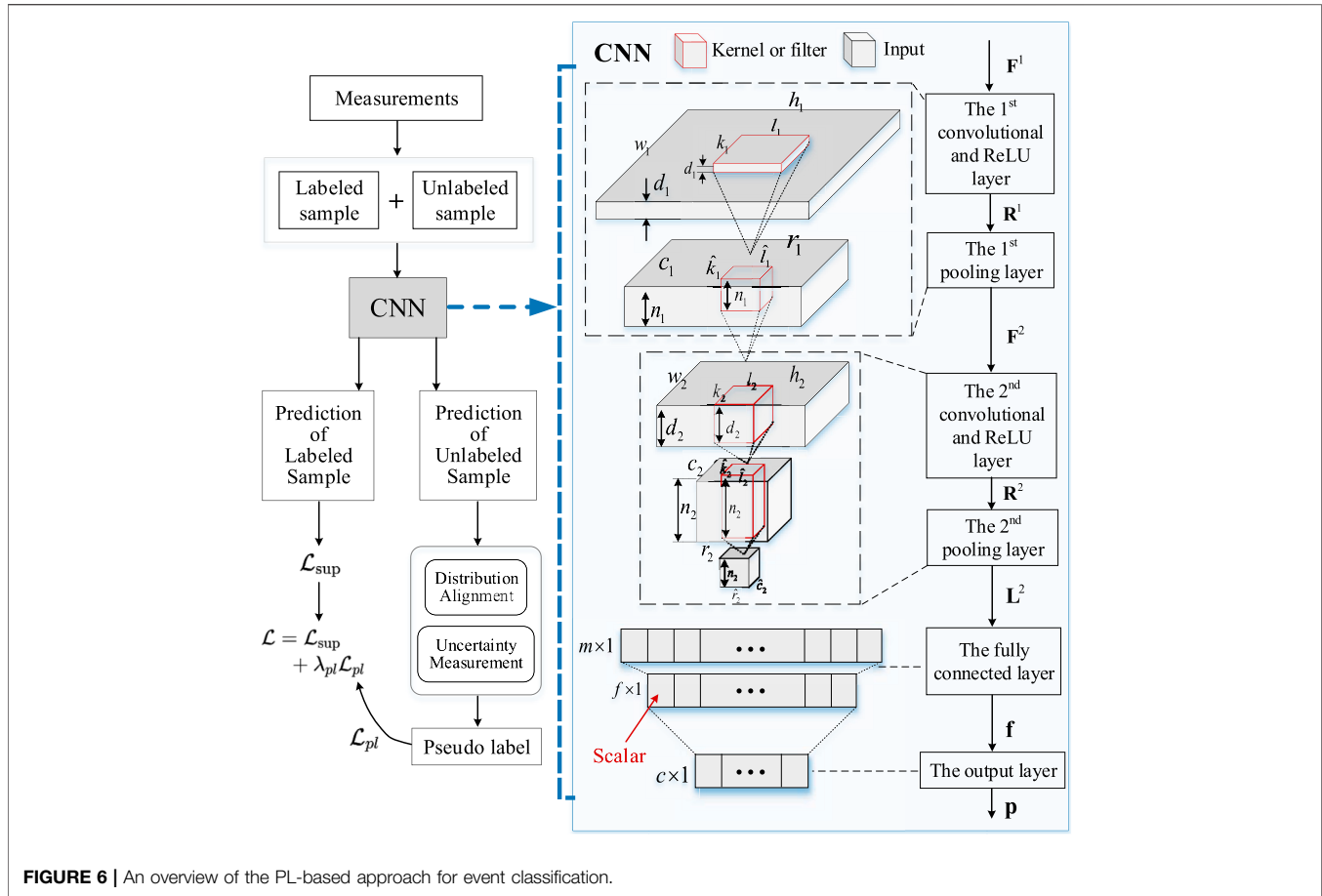


FIGURE 6 | An overview of the PL-based approach for event classification.

$$\mathbf{J} = \frac{\partial P_{\Theta}}{\partial \mathbf{X}}, \quad (5)$$

where P_{Θ} is the output likelihood, \mathbf{X} is the input (observation matrix) with entries $x_{i,k}$, and \mathbf{J} is the input-output Jacobian whose entry $j_{i,k}$ measures the contribution of $x_{i,k}$ to P_{Θ} , $i \in (1, \dots, N)$, $k \in (1, \dots, T)$. If the norm of $j_{i,k}$ is small, the entry $x_{i,k}$ only affects P_{Θ} slightly. Otherwise, the entry $x_{i,k}$ has a large impact on P_{Θ} , if the norm of $j_{i,k}$ is large. This inspires us to find $x_{i,k}$ contributing the most to the low likelihood by

$$(\eta, \tau) = \arg \max_{(i,k)} |j_{i,k}|, \quad (6)$$

where η and τ indicate the spatial location and the occurring time of the event. **Figure 5** gives a schematic diagram for event localization.

3.4 Event Classification Based on Pseudo Labels

According to **Section 2.1**, voltages/currents and differential currents/voltages are appropriate features for event classification. **Figure 6** gives an overview of the PL-based approach, which is semi-supervised with only part of the samples labeled. Let $\mathcal{X} = \{(x_b, y_b) | b \in (1, \dots, B)\}$ denote a

batch of B labeled samples, where x_b denotes samples, and y_b denotes labels. Let $\mathcal{U} = \{u_b | b \in (1, \dots, \mu B)\}$ denote a batch of μB unlabeled samples, where μ determines the relative size of \mathcal{X} and \mathcal{U} . The target is to optimize the following two losses:

- the supervised loss \mathcal{L}_{sup} on labeled samples;
- the pseudo-labeling loss \mathcal{L}_{pl} on unlabeled samples.

Both labeled and unlabeled samples are trained with a shared backbone of CNN with cross-entropy loss. For c -class classification, the supervised loss is calculated as

$$\mathcal{L}_{sup} = \frac{1}{B} \sum_{b=1}^B H(y_b, \mathbf{p}(y | x_b)) \quad (7)$$

with $H(y_b, \mathbf{p}(y | x_b)) = -\sum_{i=1}^c y_b^i \log(p_i(y | x_b))$, where $\mathbf{p}(y | x_b)$ is the prediction vector with $p_i(y | x_b)$ indicating the probability of assigning x_b to class i , $i = 1, 2, \dots, c$, $\sum_{i=1}^c p_i(y | x_b) = 1$, y_b^i indicates the one-hot encoding of assigning y_b to class i , $y_b^i \in \{0, 1\}$. Similarly, the pseudo-labeling loss is penalized over unlabeled samples u_b using PLs p_b by c -class classification, which is defined as

$$\mathcal{L}_{pl} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} H(p_b, \mathbf{p}(y | u_b)) \quad (8)$$

with $H(p_b, \mathbf{p}(y|u_b)) = -\sum_{i=1}^c p_b^i \log(p_i(y|u_b))$.

For typical PL-based methods, the p_b of an unlabeled sample u_b is directly obtained by the prediction vector $\mathbf{p}(y|u_b)$ (Lee, 2013). However, pseudo labeling and re-training are realized in the same network, which suffers from model homogenization and is easy to be trapped in a local minimum. Therefore, distribution alignment and uncertainty measurement are utilized to refine the classification method.

- **Distribution alignment:** Inspired by (Berthelot et al., 2019), prediction vectors are normalized to make category distributions homogeneous. Specifically, a running average of prediction vectors is calculated for unlabeled samples and denoted as $\hat{\mathbf{p}}$. Then for a given unlabeled sample u_b , its prediction vector $\mathbf{p}(y|u_b)$ is scaled by the ratio $\hat{\mathbf{p}}(y|u_b) = \mathbf{p}(y|u_b)/\hat{\mathbf{p}}$, and the obtained PL is \hat{p}_b .
- **Uncertainty measurement:** To enhance the performance of classification, only samples with high-precision PLs are selected for re-training. Here, the maximum entry of $\hat{\mathbf{p}}(y|u_b)$ measures the uncertainty. Only samples with $\max \hat{\mathbf{p}}(y|u_b)$ larger than a pre-set threshold (τ) are used for re-training.

In summary, our modified pseudo-labeling loss is formulated as

$$\mathcal{L}_{pl} = \frac{1}{\mu_B} \sum_{b=1}^{\mu_B} \mathbf{1}(\max \hat{\mathbf{p}}(|u_b|) > \tau) H(\hat{p}_b, \hat{\mathbf{p}}(y|u_b)), \quad (9)$$

where $\mathbf{1}$ is an indicator function, and the loss function is

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda_{pl} \mathcal{L}_{pl}, \quad (10)$$

where λ_{pl} denotes the balancing factor that controls the weight of the pseudo-labeling loss.

3.5 Convolution Neural Networks

To make this paper self-contained, a brief introduction is given for the CNN classifier in this section. As is shown in **Figure 6**, the CNN we construct here consists of 2 convolutional layers, 2 Rectified linear units (ReLU) layers, 2 pooling layers, a fully connected layer, and an output layer. The input is a 3-dimensional volume $\mathbf{X} \in \mathbf{R}^{w \times h \times d}$ with width w , height h , and depth d . The output is a prediction vector of c classes, and the class with the highest probability indicates the type of the event. Let $\mathbf{X}^i \in \mathbf{R}^{w_i \times h_i \times d_i}$ denote the i th input of the convolutional layer. Let $\mathbf{W}^{i,j} \in \mathbf{R}^{k_i \times l_i}$ be the j th kernel for the i th layer. Each kernel is moved along the width and height directions of \mathbf{X}^i to perform the dot product in the overlapping part. If the kernel is moved beyond the dimension of \mathbf{X}^i , zeros are padded to the border of \mathbf{X}^i to match the size of the kernel. The convolution results of n_i kernels are stacked together into an output $\mathbf{C}^i \in \mathbf{R}^{c_i \times r_i \times n_i}$. Then, \mathbf{C}^i is fed into the i th ReLU layer with $\mathbf{R}^i = \max(\mathbf{C}^i, 0)$, where $\max(\cdot)$ is performed on each entry of \mathbf{C}^i . Then the maximum pooling layer further reduces the size of \mathbf{R}^i . Let the size of the pooling filter be $\hat{k}_i \times \hat{l}_i$. The filter is moved along the width and height directions of \mathbf{C}^i in each depth layer, and only the maximum entry within the filter remains. The output is \mathbf{L}^i , and it becomes the input of the $(i+1)$ -th convolutional layer, i.e., $\mathbf{X}^{i+1} = \mathbf{L}^i$.

After the second pooling layer, the output \mathbf{L}^2 is reshaped into a vector $\mathbf{q} \in \mathbf{R}^m$ and then input into the fully connected layer. Denote the output of the fully connected layer as $\mathbf{f} \in \mathbf{R}^l$, and finally, the prediction vector $\mathbf{p} \in \mathbf{R}^c$ can be computed by $\mathbf{p} = g((\mathbf{W}^o)^T \mathbf{f} + \mathbf{b}^o)$, where $\mathbf{W}^o \in \mathbf{R}^{l \times c}$ and $\mathbf{b}^o \in \mathbf{R}^c$ denote the output weights and bias, $g(\cdot)$ is a softmax function with $g(x) = \frac{e^x}{\sum e^x}$. The prediction vector \mathbf{p} includes probabilities of c classes for the input \mathbf{X} , and the highest probability indicates the classified class of \mathbf{X} .

Based on the research in **Section 3**, the flowchart of the framework for event detection, localization, and classification is presented in **Figure 7**.

4 CASE STUDIES

In this section, the framework for event detection, localization, and classification is validated with both simulated data and real-world online monitoring data. Comparisons with other approaches are also given in this section.

4.1 Case Studies on Event Detection and Localization

4.1.1 Simulated Data

The INN-based method is tested with the IEEE 34-bus system shown in **Figure 1** for event detection. According to different distances to generators, several event locations are set for TLG, LLG, HLS, and LT, as is shown in **Table 1**, where FCT, FLL, and GR represent fault clear time, fault location in line, and ground resistance. Three-phase voltage magnitudes are measured by 17 measurement units, and the total dimensionality of measurement variables is 51. Simulated data is generated with PSCAD. The simulation step is set as 50 μ s, and the phasor is calculated for every cycle in the 50 Hz system. The simulation time of each sample is set as one second. Gaussian noise with a signal-to-noise ratio (SNR) of 50 dB is added to mimic normal fluctuations. Finally, a total of 2000 normal samples of size 51×50 are utilized for training, whereas the test set contains 1,600 samples, and 400 of them are anomalous.

Figure 8 shows the detection result, i.e., likelihood distributions for both normal and abnormal samples in the test set. It can be observed that the trained INN assigns lower likelihoods to abnormal samples than normal ones, which verifies the feasibility of likelihoods serving for classification. Then a DB can be naturally designed to distinguish abnormal samples. It is noted that in this case, the lowest likelihood for abnormal samples is $-9,892$. For an intuitive comparison, we just show samples with likelihoods larger than -2 in **Figure 8**.

4.1.2 Real-World Data

In this part, online monitoring data obtained from a distribution network in Hangzhou city of China is used to validate the approach. The distribution network contains 200 feeder lines with 8,000 load-side transformers. Here, the measurements in **Figure 3A** are utilized for analysis. The feeder line contains 14 load-side transformers, and the total dimensionality of

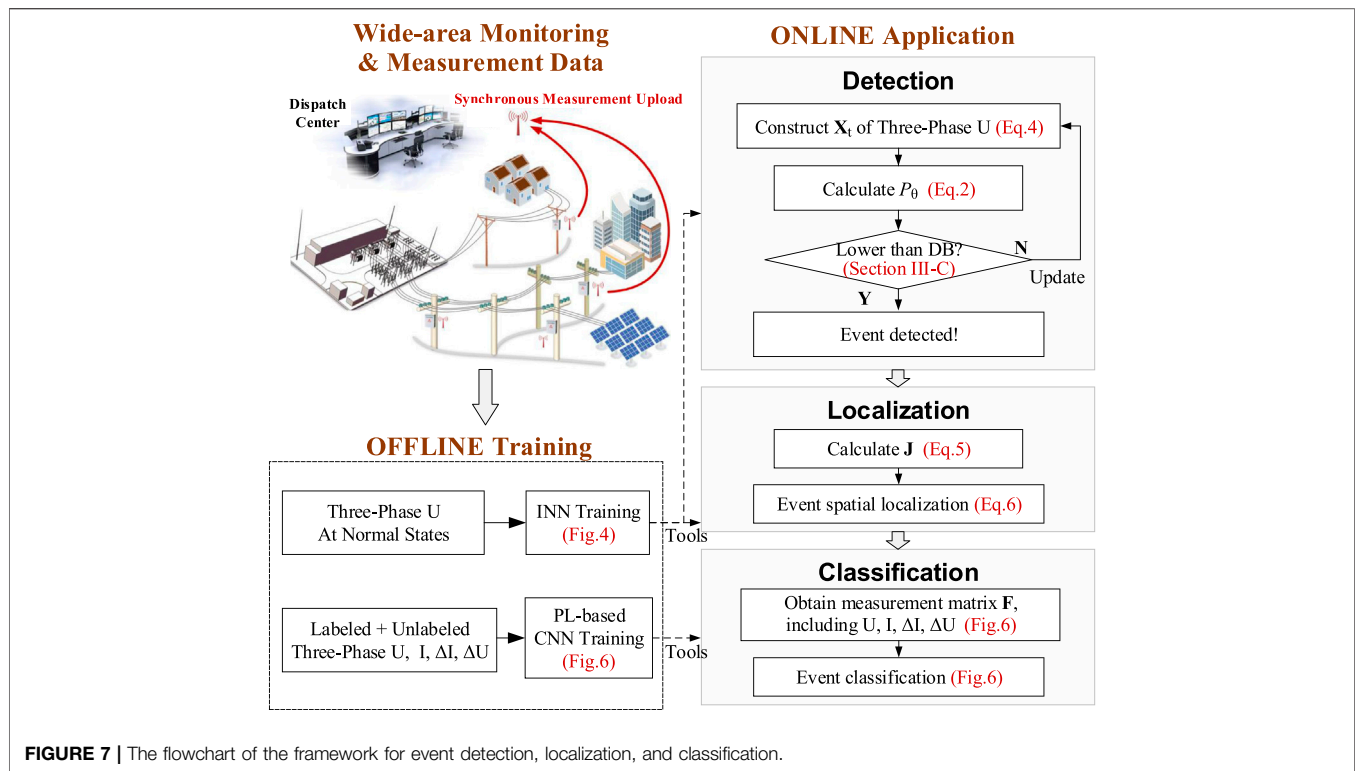
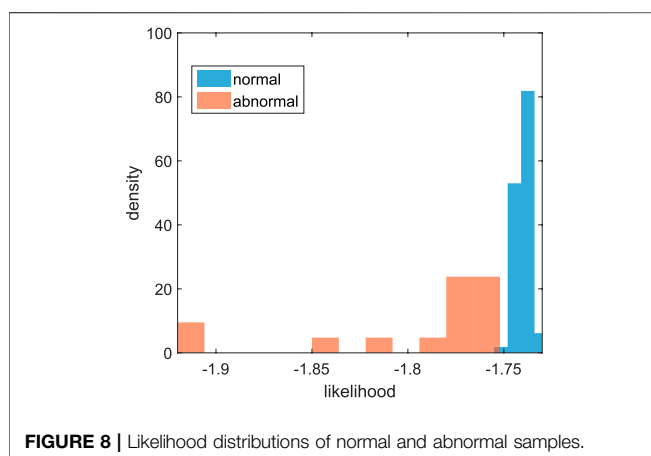


TABLE 2 | Different events simulated in the IEEE 34-bus system.

Event	Simulation setting	Event location
TLG	FCT 0.1 s, FLL 0.1–0.9 p.u., GR 0–50 Ω	Branch 816–824, Branch 828–830, Branch 832–858, Branch 812–814
LLG	FCT 0.1 s, FLL 0.1–0.9 p.u., GR 0–50 Ω	Branch 816–824, Branch 852–854, Branch 888–890, Branch 832–858
HLS	0.3–0.6 MW	Bus 844, Bus 888, Bus 860, Bus 810
LT	FLL 0.1–0.9 p.u.	Branch 828–830, Branch 816–824, Branch 832–858, Branch 888–890



three-phase voltage magnitudes is 42. The online monitoring data were sampled during 2017/3/1 00:00:00 ~2017/4/9 23:45:00. Amongst, normal measurements during 2017/3/1 00:00:00 ~2017/3/14 23:45:00 are utilized to train the INN. The

remaining obtained during 2017/3/15 00:00:00 ~2017/4/9 23:45:00 are tested. A continuously moving window of size 42×192 is utilized to truncate the datasets. Raw measurements of the test set and the likelihood curve obtained by the trained INN are shown in **Figure 9A,B**, respectively. The DB is determined as the minimum value of likelihoods obtained in the training set. On April 4th and April 5th, multiple events occurred successively, and measurements of the 2 days are zoomed in as **Figure 9A**. It can be observed that likelihoods in **Figure 9B** first drops below the DB slightly on April 4th and then drops significantly on April 5th, indicating a more serious event on April 5th.

Further, the observation matrix truncated on April 5th is utilized for event localization. The input-output Jacobian is presented as a 3-D map in **Figure 10**. The maximum entry of the Jacobian is circled and the Location Index is determined as 29, indicating the B-phase of the 10-th transformer, and which matches the event records. In this case, three-phase voltages are obtained at load-side transformers. However, on some feeders in distribution networks, only line-to-line voltages can be acquired for the economy. In this situation, the localization accuracy may be reduced, but the disturbed location can still be

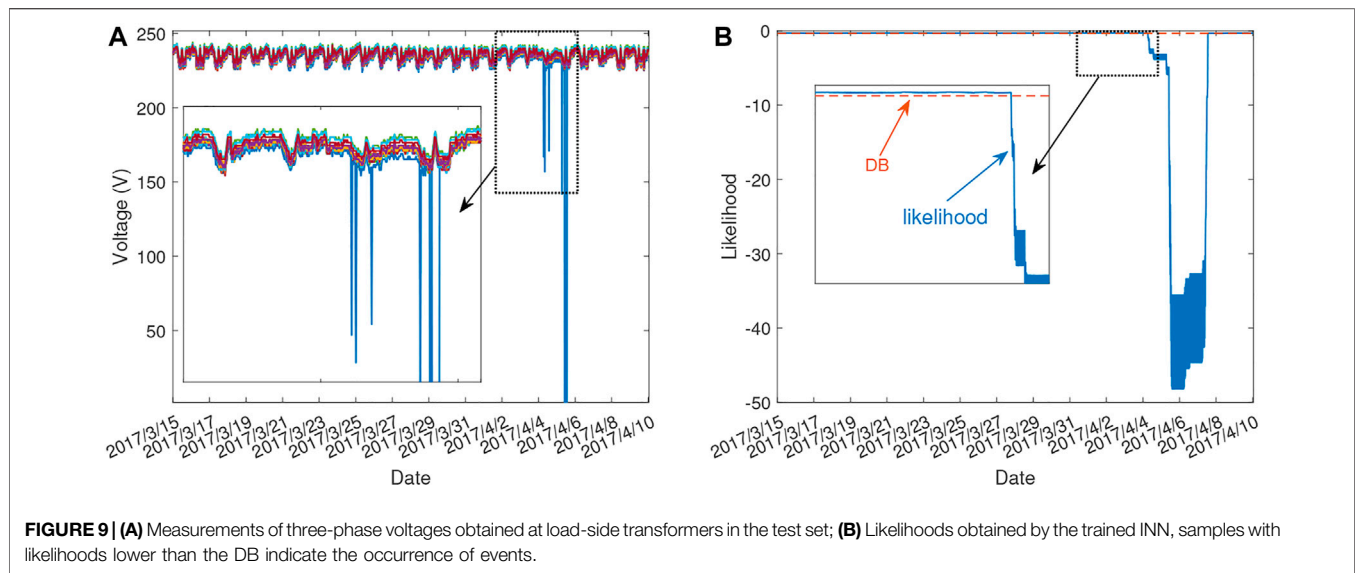


FIGURE 9 | (A) Measurements of three-phase voltages obtained at load-side transformers in the test set; **(B)** Likelihoods obtained by the trained INN, samples with likelihoods lower than the DB indicate the occurrence of events.

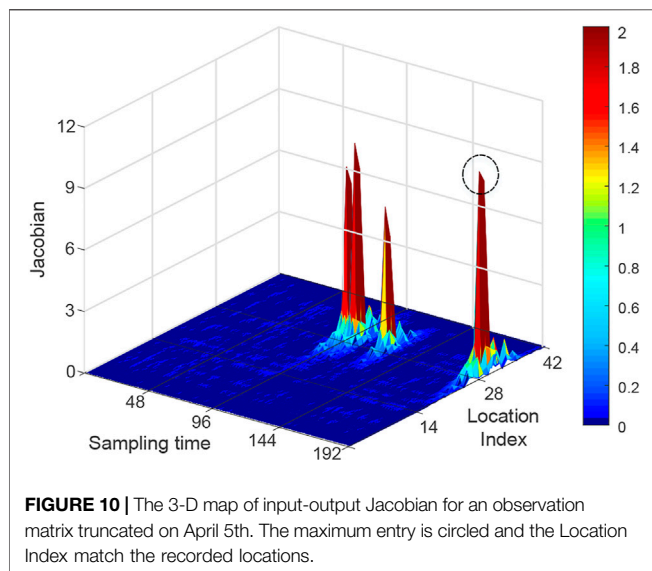


FIGURE 10 | The 3-D map of input-output Jacobian for an observation matrix truncated on April 5th. The maximum entry is circled and the Location Index match the recorded locations.

determined as the nearby position where three-phase voltages can be acquired (e.g., substations, switching stations, and load-side transformers).

4.1.3 Comparisons With Other Approaches

In this part, the INN-based approach is compared with other approaches for event detection, including DAE (Ahmed et al., 2021), PCA (Xie et al., 2014), Gaussian mixture model (GMM) (Catterson et al., 2010), OCSVM (Wang et al., 2019), and K-means (Ozgonenel et al., 2012). Assume that positive samples are abnormal samples with events, whereas negative samples are those obtained at normal states. In order to evaluate the performance of the approaches, four categories of samples are generated according to genuine types and detection results:

- True Positive (TP): abnormal samples (positive samples) that are detected as anomalies (positives);
- False Positive (FP): normal samples (negative samples) that are detected as anomalies (positives);
- False Negative (FN): abnormal samples (positive samples) that are detected to be normal (negatives);
- True Negative (TN): normal samples (negative samples) that are detected to be normal (negatives).

Precision measures the detection accuracy and is given by

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (11)$$

Recall is defined as the number of positives the model claims compared to the actual number of positives there are throughout the data. It is given by

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (12)$$

Different precision and recall values are achieved when different DBs are set to distinguish between normal and abnormal samples. The higher the precision and recall values, the better the detection performance of one approach. However, a higher recall value generally corresponds to a lower precision value. Therefore, precision-recall curves (PRCs) generated under different DBs are utilized for a comprehensive evaluation of approaches, and we compute the area under the PRC, termed the AP by

$$AP = \int_0^1 p(r)dr, \quad (13)$$

where “p” denotes precision, “r” denotes recall. The higher the AP is, the better the detection performance, $AP \in [0, 1]$. Here, the calculation of AP for the comparison approaches is introduced as follows.

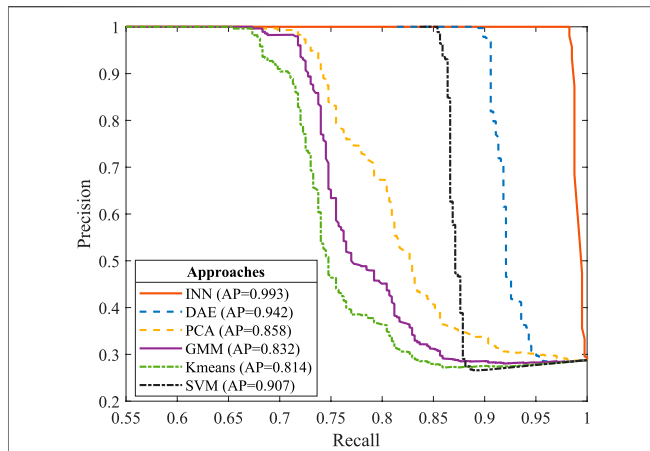


FIGURE 11 | PRCs of different approaches with simulated data in the IEEE 34-bus system. APs of INN are larger than others, which verifies the advantages of INN.

- **DAE:** Observation matrices obtained at normal states are utilized for training. The loss function is the reconstruction error (RE) of input samples and is calculated by MSE as $\frac{1}{m} \sum_{i=1}^m (x_i - \hat{x}_i)^2$, where m is the number of entries in the observation matrix, x_i and \hat{x}_i are true values and predicted values of entries, respectively. A sample is considered to be abnormal if the RE is larger than the DB.
- **PCA:** PCA is a classical dimensional reduction method. Given an observation matrix $\mathbf{X} \in \mathbb{C}^{N \times T}$ obtained in normal states, the covariance matrix is $\mathbf{C} = \mathbf{X}\mathbf{X}^T$. Calculate the eigenvalues and eigenvectors of \mathbf{C} and rearrange the eigenvalues in decreasing order. Out of the N eigenvalues, select the largest m satisfying $\sum_{i=1}^m \lambda_i / \sum_{i=1}^N \lambda_i \geq \kappa$, where κ is the CVP and $m < N$. PMUs corresponding to the m largest eigenvalues are called “pilot PMUs”, and the remaining $(N - m)$ PMUs are “non-pilot PMUs”. Form the base matrix $\mathbf{X}_B \in \mathbb{C}^{m \times T}$ using measurements of pilot PMUs. Select a non-pilot PMU with measurements $\mathbf{x} \in \mathbb{C}^{1 \times T}$, and the linear regression coefficients of \mathbf{x} on \mathbf{X}_B can be calculated as $\mathbf{v} = (\mathbf{X}_B \mathbf{X}_B^T)^{-1} \mathbf{X}_B \mathbf{x}^T$. For a newly observed matrix \mathbf{X}^{new} , the predicted value of non-pilot is obtained as $\hat{\mathbf{x}}^{new} = \mathbf{v}^T \cdot \mathbf{X}_B^{new}$. The MAE $\frac{1}{T} \sum_{i=1}^T |\mathbf{x}^{new} - \hat{\mathbf{x}}^{new}|$ serves as the detection indicator. A sample is seen as abnormal if the MAE is larger than the DB.
- **GMM:** GMM is a clustering-based method that approximates complex distributions with a linear superposition of multiple Gaussian distributions. For GMM, the number of clustering categories is pre-designed, and assume that normal samples are clustered with smaller category indices. A sample is considered to be abnormal if the category index is larger than the DB.
- **OCSVM:** OCSVM learns a hyperplane to enclose normal samples. Signed distances to the separating hyperplane are positive for an inlier and negative for an outlier. A sample is considered to be abnormal if the signed distance is smaller than the DB.

TABLE 3 | APs of different approaches with real-world data.

Approaches	INN	DAE	PCA	GMM	K-means	OCSVM
AP	0.988	0.759	0.374	0.135	0.231	0.725

- **K-means:** Samples are clustered by k centers. Assume that normal samples are clustered with smaller indices, and a sample is considered to be abnormal if the category index is larger than the DB.

Both simulated data and real-world data are utilized for comparison. For simulated data, the training set and test set are the same as in **Section 4.1.1**. For real-world data, 50 feeder lines with 120 event records during 2017/3/20 00:00:00 ~2017/4/9 23:45:00 are analyzed. A moving window with 96 sampling points is utilized to truncate the datasets. For the simulated data, PRCs, and APs of different approaches are shown in **Figure 11**. For the real-world data, APs of different approaches are calculated and given in **Table 3**. It can be observed that INN achieves the highest AP for both simulated data and real-world data. For DAE, PCA, GMM, K-means, and OCSVM, AP is significantly lower for real-world data than for simulated data. This is because real-world data exhibits complex and nonlinear properties, which is more difficult to model than simulated data. Specifically, PCA is a linear dimension reduction approach and is not applicable for nonlinear measurements. DAE is a nonlinear generalization of PCA. However, it is vulnerable to sporadic spikes and random fluctuations because of the simple structure. K-means, GMM, and SVM are strongly dependent on pre-designed parameters, whose optimal settings are hard to find for all datasets. INN, by contrast, and is capable of modeling and characterizing complex distributions without empirical settings or assumptions. As a result, it outperforms other approaches, especially in dealing with real-world datasets.

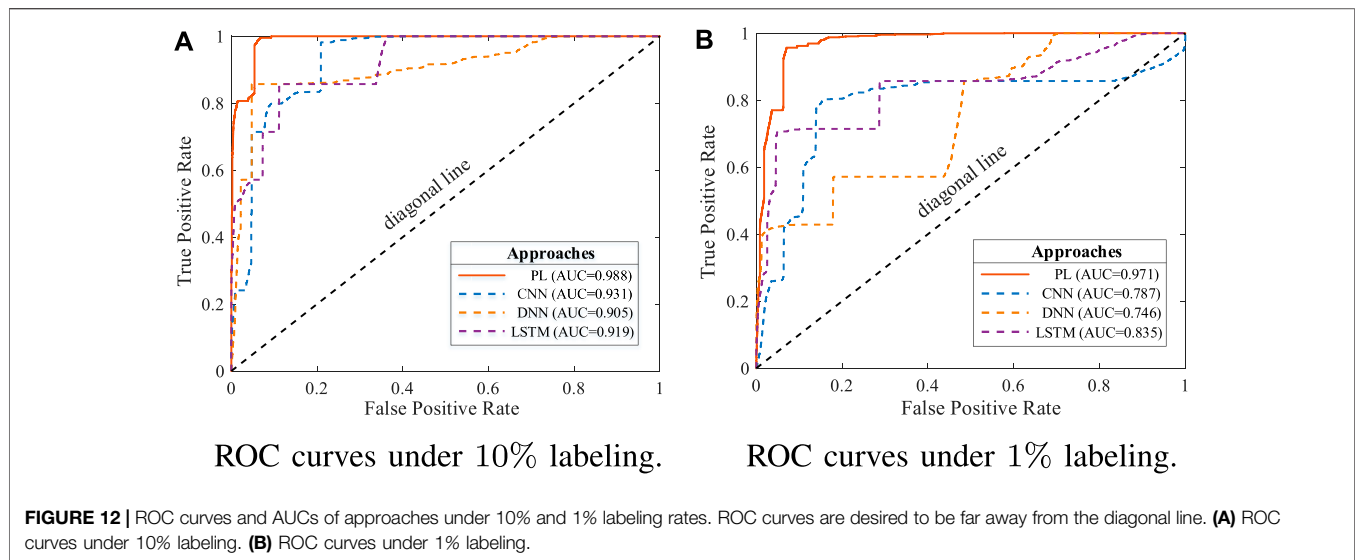
4.2 Case Studies on Event Classification

In this section, the PL-based approach is compared with other approaches for event classification, including CNN (Li and Wang, 2019), deep neural network (DNN) (Yadav et al., 2019), and LSTM (Li et al., 2021). Different events are generated as in **Table 2**. Received operational characteristics (ROC) and the area under the ROC curve (AUC) can measure the capability of a classifier to distinguish between multiple classes and they serve as evaluation metrics. For events of type i , the ROC is calculated by assuming type i as the positive class, and all others as negative classes. Then the average ROC is defined by TPR_{aver} against FPR_{aver} with

$$\text{TPR}_{\text{aver}} = \frac{\sum_{i=1}^n \text{TP}_i}{\sum_{i=1}^n (\text{TP}_i + \text{FN}_i)}, \quad (14)$$

$$\text{FPR}_{\text{aver}} = \frac{\sum_{i=1}^n \text{FP}_i}{\sum_{i=1}^n (\text{FP}_i + \text{TN}_i)}, \quad (15)$$

where n is the number of classes. The average ROC curve is desired to be far away from the diagonal line, and it indicates



an enhanced separating ability for different events. The AUC reveals the capability of a classifier quantitatively and $AUC \in [0, 1]$. A larger AUC indicates better performance in classification.

For the training set and test set, numbers of cases for each event are 400 and 150, respectively. For a fair comparison, rates of labeled samples are set as 10% and 1% for CNN, DNN, LSTM, and PL. **Figure 12** shows ROC curves and corresponding AUCs of different approaches under 10% and 1% labeling rate. It can be observed that the PL-based approach obtains the largest AUC, especially under a low labeling rate of 1%. This benefits from the re-training process using samples with high-precision PLs. In this way, the rate of labeled samples becomes higher after epochs of training, and the PL-based approach achieves the effect comparable to supervised learning in the test set. Therefore, the PL-based approach outperforms the CNN, DNN, and LSTM-based approaches under a low labeling rate.

5 CONCLUSION

In this paper, a framework is presented for event detection, localization, and classification in distribution networks to realize real-time situational awareness and event analysis. Key findings are summarized as follows.

REFERENCES

- 1) The INN-based approach outperforms others in event detection with a higher AP due to INN's superior ability in modeling complex, nonlinear measurements.
 - 2) Based on feature analysis of several principal events, including TLG, LLG, HLS, and LT, we verify that a combination of voltages/currents and differential currents/voltages possesses distinctive characteristics for different events and is appropriate for event classification.
 - 3) For event classification, the PL-based approach shows superiority over CNN, DNN, and LSTM-based approaches, and the AUC is increased by 10% under a low labeling rate (1%).
- DATA AVAILABILITY STATEMENT**
- The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.
- AUTHOR CONTRIBUTIONS**
- FY: conceptualization, writing—original draft. ZL: methodology, formal analysis. YZ: methodology, visualization. XH: writing—review and editing. QA: methodology, supervision. RQ: visualization, supervision.
- Augmentation Anchoring,” arXiv preprint arXiv:1911.09785, accessed 13 Feb 2020 2019.
- Catterson, V. M., McArthur, S. D. J., and Moss, G. (2010). Online Conditional Anomaly Detection in Multivariate Data for Transformer Monitoring. *IEEE Trans. Power Deliv.* 25 (4), 2556–2564. doi:10.1109/tpwrd.2010.2049754
- Dinh, L., Krueger, D., and Bengio, Y., “Nice: Non-linear Independent Components Estimation,” arXiv preprint arXiv:1410.8516, accessed 30 Oct 2014 2014.
- Dinh, L., Sohl-Dickstein, J., and Bengio, S., “Density Estimation Using Real Nvp,” arXiv preprint arXiv:1605.08803, accessed 27 May 2016 2016.
- Ahmed, A., Sajan, K. S., Srivastava, A., and Wu, Y. (2021). Anomaly Detection, Localization and Classification Using Drifting Synchronphasor Data Streams. *IEEE Trans. Smart Grid* 12 (4), 3570–3580. doi:10.1109/tsg.2021.3054375
- Berthelot, D., Carlini, N., Cubuk, E. D., Kurakin, A., Sohn, K., Zhang, H., et al. “Remixmatch: Semi-supervised Learning with Distribution Alignment and

- He, X., Chu, L., Qiu, R. C., Ai, Q., Ling, Z., and Zhang, J. (2019). Invisible Units Detection and Estimation Based on Random Matrix Theory. *IEEE Trans. Power Syst.* 35 (3), 1846–1855.
- Kingma, D. P., and Dhariwal, P. (2018). “Glow: Generative Flow with Invertible 1x1 Convolutions,” in *Advances in Neural Information Processing Systems*. Montréal, Canada: Curran Associates, 10215–10224.
- Lee, D.-H. (2013). “Pseudo-label: The Simple and Efficient Semi-supervised Learning Method for Deep Neural Networks,” in Workshop on challenges in representation learning, Atlanta, Georgia, June 16–21 (ICML), 896.
- Li, W., and Wang, M. (2019). Identifying Overlapping Successive Events Using a Shallow Convolutional Neural Network. *IEEE Trans. Power Syst.* 34 (6), 4762–4772. doi:10.1109/tpwrs.2019.2914774
- Li, Z., Liu, H., Zhao, J., Bi, T., and Yang, Q. (2021). Fast Power System Event Identification Using Enhanced Lstm Network with Renewable Energy Integration. *IEEE Trans. Power Syst.* 36 (5), 4492–4502. doi:10.1109/tpwrs.2021.3064250
- Liu, S., Zhao, Y., Lin, Z., Liu, Y., Ding, Y., Yang, L., et al. (2019). Data-driven Event Detection of Power Systems Based on Unequal-Interval Reduction of Pmu Data and Local Outlier Factor. *IEEE Trans. Smart Grid* 11 (2), 1630–1643.
- Myung, I. J. (2003). Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.* 47 (1), 90–100. doi:10.1016/s0022-2496(02)00028-7
- Ozgonenel, O., Thomas, D., Yalcin, T., and Bertizlioglu, I. N. (2012). “Detection of Blackouts by Using K-Means Clustering in a Power System,” in 11th IET International Conference on Developments in Power Systems Protection (DPSP 2012, Birmingham, UK, 23–26 April 2012 (IET), 1–6. doi:10.1049/cp.2012.0079
- Pandey, S., Srivastava, A., and Amidan, B. (2020). A Real Time Event Detection, Classification and Localization Using Synchrophasor Data. *IEEE Trans. Power Syst.* 35 (6), 4421–4431. doi:10.1109/tpwrs.2020.2986019
- Samuelsson, O., Hemmingsson, M., Nielsen, A. H., Pedersen, K. O. H., and Rasmussen, J. (2006). Monitoring of Power System Events at Transmission and Distribution Level. *IEEE Trans. Power Syst.* 21 (2), 1007–1008. doi:10.1109/tpwrs.2006.873014
- Song, J., Cotilla-Sanchez, E., Ghanavati, G., and Hines, P. D. (2015). Dynamic Modeling of Cascading Failure in Power Systems. *IEEE Trans. Power Syst.* 31 (3), 2085–2095. doi:10.1109/TPWRS.2015.2439237
- Tong, H., Qiu, R. C., Zhang, D., Yang, H., Ding, Q., and Shi, X. (2021). Detection and Classification of Transmission Line Transient Faults Based on Graph Convolutional Neural Network. *CSEE J. Power Energ. Syst.* 7 (3), 456–471. doi:10.17775/CSEEJPES.2020.04970
- Wang, X., Gao, J., Wei, X., Zeng, Z., Wei, Y., and Kheshti, M. (2018). Single Line to Ground Fault Detection in a Non-effectively Grounded Distribution Network. *IEEE Trans. Power Deliv.* 33 (6), 3173–3186. doi:10.1109/tpwrd.2018.2873017
- Wang, Z., Fu, Y., Song, C., Zeng, P., and Qiao, L. (2019). Power System Anomaly Detection Based on Ocsvm Optimized by Improved Particle Swarm Optimization. *IEEE Access* 7, 181580–181588. doi:10.1109/access.2019.2959699
- Wei, M., Zhang, H., Shi, F., Chen, W., and Terzija, V. (2021). Nonlinearity Characteristic of High Impedance Fault at Resonant Distribution Networks: Theoretical Basis to Identify the Faulty Feeder. *IEEE Trans. Power Deliv.* doi:10.1109/tpwrd.2021.3074368
- Wilson, A. J., Reising, D. R., Hay, R. W., Johnson, R. C., Karrar, A. A., and Daniel Loveless, T. (2020). Automated Identification of Electrical Disturbance Waveforms within an Operational Smart Power Grid. *IEEE Trans. Smart Grid* 11 (5), 4380–4389. doi:10.1109/tsg.2020.2990079
- Xie, L., Chen, Y., and Kumar, P. R. (2014). Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis. *IEEE Trans. Power Syst.* 29 (6), 2784–2794. doi:10.1109/tpwrs.2014.2316476
- Yadav, R., Raj, S., and Pradhan, A. K. (2019). Real-time Event Classification in Power System with Renewables Using Kernel Density Estimation and Deep Neural Network. *IEEE Trans. Smart Grid* 10 (6), 6849–6859. doi:10.1109/tsg.2019.2912350

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Ling, Zhang, He, Ai and Qiu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.