2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) | 978-1-6654-5556-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/SmartGridComm57358.2023.10333935

Early Alarm: Robust Event Analysis for Power Systems using 1-D Fully Convolutional Network

Chengcheng Li, Wei Wang†, Zhihao Jiang, Lin Zhu, Jinyuan Sun, Yilu Liu, Hairong Qi
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN
{cli42, zjiang13}@vols.utk.edu, wangweiutk@gmail.com, {lzhu12, jysun, liu, hqi}@utk.edu

***Abstract*—This work presents a novel deep learning model for early, accurate, and robust detection, recognition, and temporal localization of multi-type events in large-scale power systems. The proposed method develops a unified 1-D fully convolutional network (FCN) model that takes time series of raw frequency signals measured from a power system as input, extracts distinguishing features, and predicts at every temporal point in the time series if an event is happening and what the type of the event is. Compared to existing methods, the proposed model eliminates the necessity for hand-crafted feature extraction or complicated data pre-processing, can flexibly handle input signals of arbitrary length, and precisely infer the event occurrence time. Most importantly, the model is intentionally trained with incomplete patterns, such that it is more robust to partial features of an event which is common in real-world online recognition, resulting in early alarm for power system failures. Extensive experimental results demonstrate that the proposed method achieves superior performance to the state-of-the-art, and also shows strong robustness to noise and system oscillations.**

I. Introduction

In today's power grid systems, the ever-increasing scale and complexity demand the capability of early alarm for situational awareness, such that an event can be captured at its early stage, and the corresponding safety operations can be taken properly and timely to prevent potentially cascading events or large-scale outages from happening [1]. As a result, early, accurate, and robust detection and recognition of various system events become vitally important. Early efforts on event detection and recognition were mostly model-based [2], [3]. These methods, although effective, heavily rely on the control-theoretic modeling of power systems that are tightly coupled with many influential factors pertaining to different environmental changes or types of equipment used, making these models very difficult to design and generalize. On the other hand, the large-scale deployment of various monitoring devices in power systems, such as the phasor measurement unit (PMU) [4] and the frequency disturbance recorders (FDR) [5], greatly facilitates the acquisition of a large amount of measurement data at low cost. As a result, numerous data-driven event analysis approaches have surfaced [6], [7].

Machine learning is essentially a data-driven approach, and has been popularly used to solve the event detection problem in power systems [8]–[10]. Most of these methods first extract features from the raw measurement data and then feed the features to a classifier for event detection or recognition purpose. A variety of signal processing techniques have been employed for feature extraction, including, for example, Fourier transform, Wavelet transform, Stockwell transform, and sparse coding [8], [11]–[14]. In the meanwhile, numerous general-purpose classifiers have been adopted for classification purposes, with the support vector machine [9] and neural networks [15] being the two most popularly used. However, extracting a set of effective features is a non-trivial task and needs deep domain knowledge.

More recently, deep learning techniques [16], also referred to as deep neural networks (DNN), have brought unprecedented advantages to the event analysis for power grids. For example, Li *et al.* [17] proposed to extract features with physical interpretations and then employ the Convolutional Neural Network (CNN) for fault localization. Wang *et al.* [18] developed two CNN models to detect events, each of which handles one type of measurement signal. Yu *et al.* [19] devised a scheme that extracts statistical features with Wavelet transform and then employs a CNN for fault classification, phase identification, and location detection.

Although effective, these methods still have not fully leveraged the superior feature extraction capability of deep learning. One essential drawback of existing methods is that they all conduct the event detection and recognition tasks based on segments of the signal of a "fixed" window size. Compared to the early ad hoc approaches [20]–[22] where the magnitude of frequency change is detected at each point in the time series, the segment-based analysis could be much more tolerant to noises and oscillations for recognition purpose [23]. However, it also suffers from three drawbacks. *First*, the window size is hard to determine, especially for events of different types. Even events of the same type could present significant intra-class variation due to the complexity of the system and the degree of power changes, which might result in diverse patterns of different lengths. *Second*, the fixed window-size approaches need to wait until the complete time series within the window is collected before performing the recognition, which inevitably delays the response time of the system. *Third*, since an effective event recognition approach is supposed to tolerate signal segments that present only partial features, the adjacent segments, *e.g.*, seg_1, seg_2, and seg_3 in Fig. 1 may all be detected and recognized with the correct event label,

This work is supported in part by NSF CNS-2038922 and in part by NSF/DOE EEC-1041877. † Corresponding author, work done while at the University of Tennessee.

making it difficult to estimate the exact occurrence time of the event. This might be the reason why existing works mainly address the problems of *event detection* or *event recognition* but cannot perform *occurrence time localization* well.

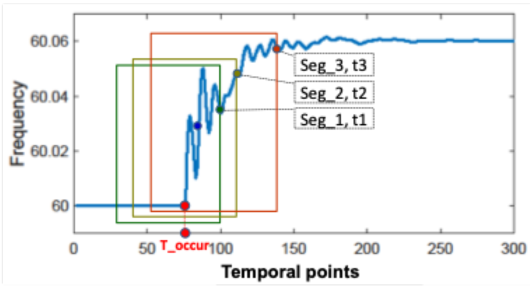


Fig. 1. Illustration of event detection and temporal localization using fixed window-size, where $seg1$, $seg2$, $seg3$ are three cropped segments of fixed size at time t_1, t_2, t_3, respectively. All of them could be recognized as an abnormal event using a classifier after proper training. However, none of them is able to provide the precise event occurrence time, marked as "T_occur".

There are a few existing works that attempted to solve the event detection, recognition, and temporal localization problems in a unified framework [11], [14], however, they are still based on analysis of complete fixed-length signal segments and do not leverage the superior feature extraction capacity from deep learning for early alarm. Furthermore, most of the existing event analysis works only focus on three event types, *i.e.*, generator trip, line trip, and load shedding, but often exclude oscillation due to the indistinct feature it carries and the interference of system noises, although oscillation is a common event type in power systems.

In this paper, we tackle the above challenging issues by presenting a holistic (*i.e.*, end-to-end) DNN-based approach for simultaneous detection, recognition, and temporal localization of all the four types of events (including oscillation) without the constraint of fixed window size. To our best knowledge, this is the first attempt in the literature. The backbone of our model is a 1-D fully convolutional network (FCN). The concept of FCN has been revived since [24], which converts the classic CNN [25]–[27] into an FCN for the problem of image semantic segmentation, *i.e.*, pixel-wise class labeling. Although originated from CNN [28], [29], FCN carries its unique characteristics that would greatly advance the state-of-the-art event analysis for power systems. The proposed model takes in the streaming time series of signals and outputs predictions if an event is happening and what kind of event it is, at every single point in the time series - hence capable of indicating the exact occurrence time of the event. The advantage of the proposed method is multi-fold.

1) *Multi-tasking*: The proposed approach addresses event detection, recognition, and temporal localization within a single deep network, which can be easily trained end-to-end.
2) *Early detection*: Due to the point-wise prediction ability, the proposed method does not need to wait for a complete pattern to appear for the recognition, hence the ability of early detection.
3) *Accuracy and Robustness*: Benefiting from the learnable feature extraction capability in DNN, the extracted features are of much higher discrimination than handcrafted feature extractors. Therefore, the proposed approach can distinguish different kinds of events, including oscillations, more accurately with stronger robustness.
4) *Adaptivity*: The proposed 1-D FCN model can handle arbitrary-length time series, eliminating the demand for predefined window size, and run predictions efficiently.

II. Preliminary: CNN and FCN

CNN has become the de facto standard for signal and image processing and served as the backbone of many deep learning models. CNN is composed of a sequence of layers, each of which transforms one volume of activations to another through a differentiable function. Three typical layers are used to build CNN architectures: the convolution layer, the pooling layer, and the fully-connected layer. The convolution layer is used to generate responses to different aspects/objects in the data, the pooling layer is used to shrink the activations for multi-scale feature abstraction, and the fully-connected layer feeds the overall feature into a classifier and predicts the final class label. The convolution kernels in the convolution layer and the weights used in the fully-connected layer are updated through the back-propagated loss via gradient, thus CNN has superior feature extraction capability to the previous hand-crafted features and more robust recognition accuracy.

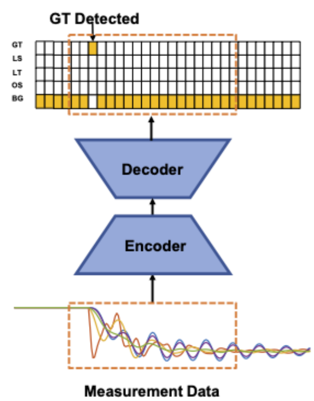


Fig. 2. Framework of FCN-based event analysis, which predicts dense point-wise class labels for the given signal, where a generator trip (GT) was identified.

In contrast, FCN removes the last fully-connected layer, making the whole network composed of only convolution layers and pooling layers. Due to the down-scale in pooling layers, the output size will be much smaller than the input data. In essence, FCN predicts the class labels in the last convolution layer for each local area of the input data. Using deconvolution layers or simply up-scaling is able to restore the output to the same size as the input. As shown in Fig. 2, an FCN is composed of two main modules, the feature extraction part, referred to as the *encoder*, and the deconvolution part, referred to as the *decoder*. It is able to predict a class label for each unit of the input data. In other words, the FCN could be interpreted as employing a CNN classifier to predict a class label for each segment cropped by a sliding window that centered at every unit in the signal. Therefore, FCN naturally operates on inputs with arbitrary size and produces an output with the same or linearly down-sampled dimension.

For *offline* event analysis with complete time series, there is not much difference between an FCN and a CNN-based classification on segments cropped by a sliding window. Both could generate dense class labels for each unit in the signal. However, FCN does this in a much more efficient way. We take the example of two segments cropped from a fix-size sliding window at two adjacent units with a large overlap. Although the two segments only differ slightly, each of which needs to pass through the CNN layers for feature extraction and class label prediction separately, resulting in quite a lot of repeated computation. In contrast, FCN computes multi-scale features for the whole signal first, so that the features computed in a low-level layer could be used by the upper-level layer directly, and thus no redundant computation exists.

For *online* event analysis, the advantage of FCN becomes even more prominent. As mentioned earlier, the event occurrence time could not be precisely determined from a sliding window cropped segment. For example, in a real monitoring system, the inspection will be applied to a segment of the latest streaming data, *e.g.*, seg_3 in Fig. 1. For seg_3 at t_3, the event pattern is already well-formed, such that it can be successfully recognized and the alert triggered. However, the true occurring time T_{occur} cannot be derived from the recognition of seg_3. In contrast, given seg_3 in real inspection, FCN will scan each unit in seg_3, and output the highest confidence at the true point of T_{occur}. Especially, if the current moment is even earlier at t_1 in Fig. 1, where the pattern of the event may just start partially appearing toward the tail. For existing pattern recognition models trained by complete event patterns, the latest segment seg_1 may not be recognized as an event but instead as noise. Since FCN is trained in a point-wise manner and intentionally trained with incomplete patterns, it repeatedly inspects each unit in seg_1 and the classification at the unit T_{occur} could have high confidence of being an abnormal event.

III. Methodology

A. Problem Formulation

Since any event occurring in a power system will cause certain frequency changes and oscillations until the system becomes stable at another frequency level, in the context of this paper, we use the time series of frequency signals as model input. However, the algorithm developed can be generalized to detect anomalies from other time series. Suppose there are M monitoring devices (e.g., FDR, PMU) installed with each bus hosting one such device, where the time-series signals can be collected. We conduct event analysis in a power system on all four typical types, namely, generator trip (GT), line trip (LT), load shedding (LS), and oscillation (OS). In addition, we add the normal status as the fifth label. The measurement data collected from these M devices over a certain duration can be represented as a matrix $\mathbf{X} \in R^{M\times T}$, as shown in Eq. 1, where T is the number of the units/samples recorded in the time series. Namely, the matrix $\mathbf{X}$ denotes an M-channel 1-D time series, with each column, $\mathbf{x}_j$ $(j = 1, \ldots, T)$, denoting a temporal point (a vector) from the M observation sites.

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,j} & \ldots & x_{1,T} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,j} & \ldots & x_{2,T} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{M,1} & x_{M,2} & \ldots & x_{M,j} & \ldots & x_{M,T} \end{bmatrix} \tag{1}$$

Given the observation $\mathbf{X}$, we formulate the event detection, recognition, and temporal localization into a unified point-wise classification problem, which can be perfectly implemented by an FCN deep learning model. The general framework of the proposed method is shown in Fig. 2. Inspired by [24] where FCN was used to generate a dense prediction for pixel-wise image segmentation, we employ FCN to produce a point-wise prediction for the given signal.

Mathematically, the proposed FCN model can be formulated as $\mathbf{Y} = f(\mathbf{X}|\mathbf{\Theta}, D)$, where $\mathbf{X}^{M\times T}$ is the measurement data, $\mathbf{Y} \in R^{5\times T}$ is the corresponding output categorical prediction. Each column vector in $\mathbf{Y}$, $\mathbf{y}_j$, denotes the probability of the event belonging to the 5 classes considered (GT, LT, LS, OS, and normal) at time j. f represents the mapping function of the prediction model with with well-designed neural network structure (will be elaborated in section **??**) and $\mathbf{\Theta}$ being the parameters that need to be optimized in training. $D = \{\mathbf{X}_i, \mathbf{Y}_i\}_{i\in\mathcal{I}}$ is the training dataset consisting of pairs of measurement data, $\mathbf{X}_i$, and their corresponding point-wise event type, $\mathbf{Y}_i$, and $\mathcal{I}$ is the set of indices of training samples.

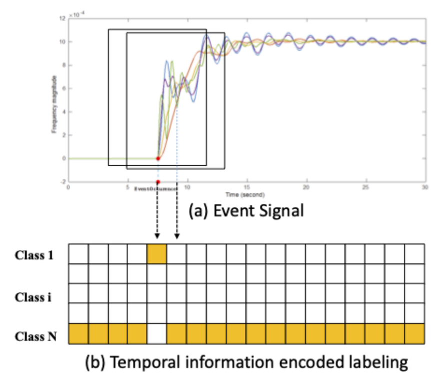


Fig. 3. Point-wise and temporal information encoded ground truth annotation.

A key point that makes our FCN model different from previous works is the way the annotation is generated. As illustrated in Fig. 3, in the temporal-aware label matrix $\mathbf{Y}^{5\times T}$, only one single unit in the time series at the occurrence time was annotated as one of the four abnormal states, while the other units, even the ones during the state transition are all annotated as normal states. Using this annotation, the model can be trained to achieve superior discrimination for the pattern at the true occurrence time, and thus the model achieves the best capability to localize the actual time when an event starts to occur. Briefly, the advantages of our FCN model include:

it 1) has outstanding feature learning capability, 2) is able to handle arbitrary-sized inputs, and 3) can generate point-wise dense predictions with temporal information encoded.

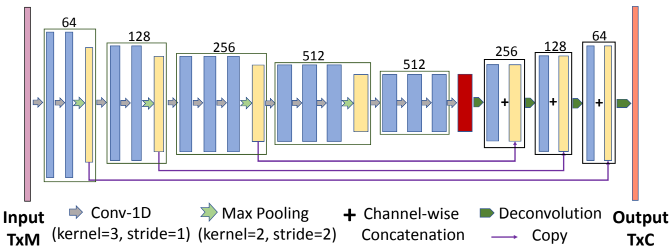


Fig. 4. The 1-D FCN network architecture. The model builds upon small convolution layers of kernel size 3, with skip-connection between parts of encoder and decoder, yielding output of the same size as input with accurate point-wise labeling.

B. Network Design

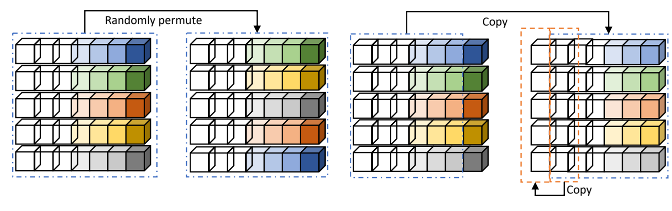
Since the frequency signals under study are 1-D time series with M channels, 1-D convolution layer is adopted as the basic block of our FCN model. Inspired by VGG [30] and FCN for image semantic segmentation [24], our neural network is also built with small convolution filters, *e.g.*, conv layers of kernel size 1×3 and stride 1, max-pooling layers of window size 2 and stride of 2. The up-sampling part in the decoder is carried out with learnable transposed convolution. To further refine the temporal detection in the dense output, skip-connections are used to combine features from corresponding layers in the encoder and decoder. As frequently used in dense prediction tasks, skip connections [31] help to enable feature reusability and stabilize model training and convergence. The details of our network are shown in Fig 4.

C. Training Details

1) Loss Function and Parameters Optimization.: Since the training set is severely skewed where most event samples belong to the "normal event" category, the cross-entropy loss is utilized to measure the discrepancy between the ground truth and the actual output of the model such that the class-wise training loss can be balanced by the weighted cross-entropy loss. A less-than-1 scaling factor is applied for the normal-event class, and the weighted cross-entropy loss is defined as Eq. 2,

$$\mathcal{L} = -\sum_{i}^{\mathcal{I}} \sum_{j}^{T} \sum_{c}^{C} s_c \times \mathbf{Y}_{i,j}^{c} \times \log f(\mathbf{X}_{i,j})^{c}, \tag{2}$$

where $\mathbf{Y}_{i,j}^{c}$ and $f(\mathbf{X}_{i,j})^{c}$ are the ground-truth and predicted probabilities of the j-th temporal point belonging to class c in point-wise label $\mathbf{Y}_i$ respectively, C is the number of classes considered, i.e., $C = 5$ in this work. s_c is the scaling factor for class c, which is set as 0.1 for the class of normal status and 1 for all the other event classes (GT, LT, LS, and OS). Parameters of the model are initialized using He's normal initialization [32]. Model is trained for 60 epochs from scratch using Adam optimizer with a learning rate of $2e-4$. The batch size was set as 16, and weight decay was set as $1e-5$ to reduce over-fitting. All these hyper-parameters are determined based on thorough empirical studies.



(a) Channel Shuffling (b) Temporal Shifting

Fig. 5. Illustration of channel shuffling and temporal shifting for training data augmentation.

2) Data Augmentation.: Training a deep learning model usually requires a relatively large dataset, otherwise, overfitting might become a potential issue. However, in this work, the training samples for the four event types can be very limited. To mitigate this problem, data augmentation is conducted, including channel shuffling and temporal shifting. As shown in Fig. 5(a) where each of the M channels is represented in a unique color, channel shuffling is a random permutation of the M channels for each training signal. Fig. 5(b) shows the temporal shifting where we shift the fixed-length training signals from left to right along the temporal axis and pad the missing values at the beginning with zero, resulting in events represented by *partial* pattern. The temporal shifting largely improves the model capacity in early detection.

IV. Experiments and Results

Extensive experiments are conducted to evaluate the model performance from different aspects, including (1) model performance for event detection, recognition, and temporal localization, (2) modal capacity for early detection, (3) the effect of the number of channels (or sensing devices) used, (4) model sensitivity to noise, (5) model robustness in handling the oscillation event, and (6) the computation analysis. In the following, we first detail the experimental setup and performance metrics used.

A. Experimental Setup

1) Dataset: The proposed model evaluation is conducted on a benchmark system, the Northeast Power Coordinating Council (NPCC) testbed, which is a reduced PSS/E [33] simulation model of real systems covering the whole or parts of ISO-NE, NYISO, PJM, MISO, and IESO [34]. The NPCC model has 140 buses, 48 machines, 230 branches, and 28GW loads. It keeps the characteristics of real systems and represents the backbone transmission of the northeast region of the Eastern Interconnection [35]. The dataset of event instances from the NPCC testbed includes 91 GT, 93 LS, and 96 LT cases. Besides, 60 oscillation cases are also included to study model robustness to oscillations, which are generated by applying a three-phase bus fault at every single bus. The length of each instance is around 30 seconds recording the event pattern during the system state transition. The sampling rate is 10Hz, resulting in about 300 samples in the recording of each instance. The occurrence time of each sample varies in a range from the 1st second to the 12th second in the recording. With the proposed data augmentation, we further extend the occurrence time into a wider range for early event detection. The dataset is randomly divided into training and testing sets with 70% and 30% instances, respectively. It is guaranteed that the instances in testing are not seen during the training.

2) Implementation: We implement all the experiments with TensorFlow [36] on NVIDIA GeForce GTX 2080ti and Intel(R) Core(TM) i7-6850K CPU@3.60GHz. The source code and pre-trained model will be released after the publication.

3) Metrics: For a fair comparison, we adopt the same evaluation metrics used in [11], [14], including the detection accuracy (DA), false alarm rate (FA), event pattern recognition rate (EPR), and occurrence time deviation (OTD).

- Detection accuracy (DA): Suppose the event actually occurs at time T_{occur}. If the model can detect the event within the range of $[T_{occur} - \delta, T_{occur} + \delta]$, where δ is a very small value, then we consider the model correctly detects the event; otherwise, it is a false alarm. DA denotes the ratio between the number of correctly detected events and the number of total events. The effect of δ on DA will be further evaluated.
- False alarm rate (FA): FA is the ratio between the number of falsely detected events and the number of total events.
- Event pattern recognition rate (EPR): EPR calculates the percentage of the correctly classified events out of the total correctly detected events.
- Occurrence time deviation (OTD): OTD indicates the relative deviation between the derived occurrence time and the actual occurrence time.

B. Experimental Results

We conduct six sets of experiments to provide a comprehensive evaluation of the proposed model for the tasks of event detection, recognition, and temporal localization. In the following, we first compare the proposed model with state-of-the-art approaches using the four metrics explained above. We then study the modal capacity in terms of early detection, the number of channels (or sensors) used, sensitivity to noise, and oscillation detection. We also provide computational analysis in the end.

1) Comparison with the state-of-the-art: We compare the performance of the proposed model with two recent works, a sparse coding method, CSC [11], and a deep learning-based method, FED [18]. The former (CSC) method was applied for event detection, recognition, and temporal localization simultaneously and has achieved state-of-the-art accuracy [37], while the latter (FED) method was for event recognition only. For event detection and temporal localization, the parameter δ controls the tolerance to imprecision.

We first set $\delta = 0$ for zero tolerance to imprecise temporal localization. The corresponding results are shown in the first row of Table I. We can observe that the CSC method cannot detect any events with this high precision, none of the derived occurrence time matches the ground truth, resulting in DA rate as 0. Under this rigorous condition, our method still achieves a 98.81% DA rate, with the FA rate being only 14.28%, and this is why the compared method's FA rate can be 102.08%. Note that one single event might be detected as multiple false alarms. The FED method only works for event recognition, with the EPR rate being 92.86%, which is not as high as the 100% EPR from our model.

We also relax the value of δ to 0.1, 0.2, 1.0, 3.5 sec, respectively, so the event detection will be treated as correct if the derived occurrence time falls into the interval of $T_{occur} \pm \delta$, which also indicates the OTD is less than δ. The results are shown in Row 2 through Row 5 of Table I. We can observe that our method achieved a 100% DA rate with $\delta = 0.1$ without any false alarm. In contrast, the CSC method only achieves a 37.5% DA rate. If increase the δ to ≥ 0.2, the DA rate for CSC could be boosted significantly, but the FA rate is still above 0. For event pattern recognition, both CSC and our method achieve very good accuracy. In the subsequent experiments, we fix δ at 0.1 sec unless otherwise specified.

TABLE I
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART.

Interval δ (s)	Method	DA (%)	FA (%)	EPR (%)
0.0	CSC	0.0	102.08	NA
	FED	NA	NA	92.86
	Ours	98.81	14.28	100
0.1	CSC	37.5	63.89	100
	Ours	**100**	**0**	100
0.2	CSC	96.53	4.86	100
	Ours	100	0	100
1.0	CSC	100	1.39	100
	Ours	100	0	100
3.5	CSC	100	1.39	100
	Ours	100	0	100

2) Study of early detection: This experiment explores how early the proposed approach can achieve accurate detection, recognition, and temporal localization by feeding incomplete signal segments into our model. We intentionally truncate the input time series after event occurrence such that the event lasted for 1, 2, 2.5, 5, and 10 seconds, respectively. The experimental results are reported in Table II. We observe that we can detect 50% of events as early as in the first second. Also, the DA rate could be improved to 97.6% once the event lasted for 2 seconds. Furthermore, with more data streaming in, the OTD value of the derived occurrence time could also be reduced.

TABLE II
PERFORMANCE ON EARLY DETECTION WITH $\delta = 0.1 sec$.

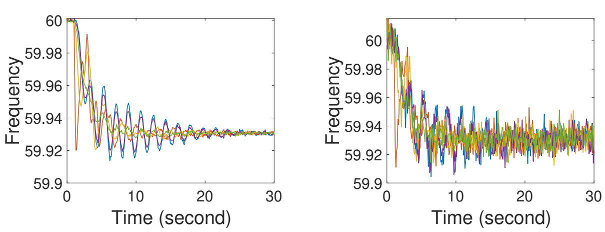
Event Duration (s)	DA (%)	FA (%)	EPR (%)	OTD (s)
1.0	50	0	100	0.04
2.0	97.62	0	100	0.04
2.5	100	0	100	0.02
5.0	100	0	100	0.01
10.	100	0	100	0.01

3) Study of number of channels: This experiment studies the model robustness to the number of input channels (or monitoring devices), M, used. Intuitively, the more monitoring devices used, the more tolerant the model is to noise. In the experiments, the default value of M is 5. Here, we vary the number of channels, M, to 1, 3, and 5, respectively and re-train the model and evaluate the event analysis accuracy. All the experimental results are reported in Table III. We can observe that fewer channels carry less information, making it a bit more challenging for event analysis. However, with 1 channel only, the DA rate and FA rate just degrade slightly with the EPR rate kept at the same 100%. In addition, the OTD value is just slightly increased with only one monitoring device used. The results demonstrate the strong feature extraction capability of the proposed deep learning model so that we can safely make operational decisions for the power system when fewer sensing devices are available.

TABLE III
ACCURACY FOR THE DIFFERENT NUMBER OF CHANNELS OR MONITORING DEVICES USED.

Channels	DA (%)	FA (%)	EPR (%)	OTD (s)
1	98.57	1.19	100	0.04
3	99.60	0.71	100	0.01
5	100	0	100	0.01

4) Sensitivity to noise: This experiment evaluates the model from the perspective of robustness to noise. We manually add different levels of Gaussian noise to the signals. In order to get an intuition of noise intensity, we show the noisy signals with different signal-to-noise ratios (SNR) in Fig. 6. Experimental results are shown in Table IV. It is observed that the proposed method can maintain almost the same level of accuracy when the SNR is as low as 30 dB. However, the CSC method only achieves 95.65% DA and 29.13% FA rates, respectively, with SNR being 40 dB [11].



(a) SNR=40 (b) SNR=20

Fig. 6. Illustration of noisy signals with different SNRs as model input.

5) Robustness to oscillation: The oscillation event has been recognized as a vital concern for the stability of power systems. Although there have been extensive studies focusing on oscillation itself, most existing works for event analysis choose to exclude it. The challenges of oscillation analysis lie in several aspects. First of all, oscillations can be excited by other types of events, including even non-obvious disturbances or normal power system operations [38]. Second, although it is commonly observed that oscillations are associated with generator trip, load shedding, or line trip events, their effects on these three types of event analyses are quite different. For GT and LS, the influence of oscillation is small since it will not change the general pattern (or frequency profile) of GT/LS and thus can be treated as noise. However, oscillation presents a non-ignorable challenge to LT, whose frequency change profile can be quite similar to those of oscillations.

In this experiment, we include oscillation as one of the event categories and report the prediction accuracy for each category separately. From the results of clean data shown in the first row of Table IV, we can see that the oscillation only degraded the DA rate of LT slightly from 100% to 96.7%, and the DA rate for oscillation itself also achieved 100%. The mean FA rate was increased a bit to 1.08%, and the OTD value is trivial that less than 0.001 seconds.

TABLE IV
EXPERIMENTAL RESULTS ON CLASS-WISE ANALYSIS ACCURACY WITH DIFFERENT LEVELS OF GAUSSIAN NOISE (MEASURED IN SNR) PRESENTED.

SNR	Type	DA (%)	FA (%)	EPR (%)	OTD (s)
Clean	GT	100	0	100	0
	LS	100	0.77	100	0
	LT	96.77	0	100	0
	OS	100	6	100	0.006
	Mean	99.02	1.08	100	0.00099
60 dB	GT	100	0	100	0
	LS	100	0.77	100	0
	LT	96.77	0	100	0
	OS	100	6	100	0.0067
	Mean	99.02	1.08	100	0.00099
50 dB	GT	100	0.33	100	0
	LS	100	0.77	100	0
	LT	96.77	0	100	0
	OS	100	5.33	100	0.013
	Mean	99.02	1.08	100	0.00198
40 dB	GT	100	0	100	0
	LS	100	0.38	100	0
	LT	96.77	0	100	0
	OS	100	2.66	100	0.013
	Mean	99.02	0.49	100	0.00198
30 dB	GT	96.33	3	100	0.0034
	LS	98.46	2.69	100	0.0658
	LT	96.77	0	100	0
	OS	99.33	1.33	100	0.1476
	Mean	97.45	1.76	100	0.04
25 dB	GT	79.67	0	100	0.0085
	LS	89.61	1.54	100	0.128
	LT	96.77	4.19	100	0
	OS	58.67	8.67	100	0.3283
	Mean	84.31	2.94	100	0.0708
20 dB	GT	47.00	9.66	100	0.18
	LS	96.54	4.62	100	0.2057
	LT	96.77	11.61	100	0
	OS	35.33	37.33	100	0.432
	Mean	70.49	13.04	100	0.13

6) Computation analysis: In this last set of experiments, we measure the model latency for both training and inference, where we run 5 trials and 100 trials for training and inference, respectively. The average time for training with 60 epochs is 16.3 min, and the average inference time for a sample with a length of 30 sec is only 0.017 sec, which significantly surpasses the sampling frequency of 10 Hz. Hence, our model can easily achieve real-time performance for event detection in real power systems.

V. Conclusion

This work presented a new approach for event monitoring in situational awareness of power systems. The proposed model enables multi-tasks within a single fully convolutional network, including event detection, recognition, and temporal localization. The approach takes advantage of the excellent feature extraction ability of deep learning, which helps to achieve superior performance as compared to other existing works. More importantly, due to our training strategy, the proposed model is also very sensitive to partial features, and presents prominent effectiveness for the early detection of abnormal events.

References

[1] Mathaios Panteli and Daniel S Kirschen. Situation awareness in power systems: Theory, challenges and applications. *Electric Power Systems Research*, 122:140–151, 2015.

[2] M Aldeen and R Sharma. Robust detection of faults in frequency control loops. *IEEE Transactions on Power Systems*, 22(1):413–422, 2007.

[3] Xi-Fan Wang, Yonghua Song, and Malcolm Irving. Small-signal stability analysis of power systems. *Modern Power Systems Analysis*, pages 489–542, 2008.

[4] Xianda Deng, Desong Bian, Di Shi, Wenxuan Yao, Zhihao Jiang, and Yilu Liu. Line outage detection and localization via synchrophasor measurement. In *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pages 3373–3378. IEEE, 2019.

[5] Z. Zhong, C. Xu, B. Billian, L. Zhang, S. Tsai, R. Conners, Centeno V., A. Phadke, and Y. Liu. Power system frequency monitoring network (fnet) implementation. *IEEE Trans. Power Systems*, 20:1914–1921, 2005.

[6] Shutang You, Jiahui Guo, Gefei Kou, Yong Liu, and Yilu Liu. Oscillation mode identification based on wide-area ambient measurements using multivariate empirical mode decomposition. *Electric Power Systems Research*, 134:158–166, 2016.

[7] Shengyuan Liu, Yuxuan Zhao, Zhenzhi Lin, Yilu Liu, Yi Ding, Li Yang, and Shimin Yi. Data-driven event detection of power systems based on unequal-interval reduction of pmu data and local outlier factor. *IEEE Transactions on Smart Grid*, 2019.

[8] V Thiyagarajan and NP Subramaniam. Wavelet approach and support vector networks based power quality events recognition and categorisation. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1667–1671. IEEE, 2016.

[9] Peter GV Axelberg, Irene Yu-Hua Gu, and Math HJ Bollen. Support vector machine for classification of voltage disturbances. *IEEE Transactions on power delivery*, 22(3):1297–1303, 2007.

[10] Ananya Chakraborty and Ratan Mandal. A novel technique employing dwt-based envelope analysis for detection of power system transients. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 346–350. IEEE, 2017.

[11] Yang Song, Wei Wang, Zhifei Zhang, Hairong Qi, and Yilu Liu. Multiple event detection and recognition for large-scale power systems through cluster-based sparse coding. *IEEE Transactions on Power Systems*, 32(6):4199–4210, 2017.

[12] Guo-Sheng Hu, Jing Xie, and Feng-Feng Zhu. Classification of power quality disturbances using wavelet and fuzzy support vector machines. In *2005 International conference on machine learning and cybernetics*, volume 7, pages 3981–3984. IEEE, 2005.

[13] Wei Wang, Li He, Penn Markham, Hairong Qi, and Yilu Liu. Detection, recognition, and localization of multiple attacks through event unmixing. In *IEEE Smart Grid Comm*, pages 73–78. IEEE, 2013.

[14] Wei Wang, Li He, Penn Markham, Hairong Qi, Yilu Liu, Qing Charles Cao, and Leon M Tolbert. Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid. *IEEE Transactions on Smart Grid*, 5(4):1654–1664, 2014.

[15] Silvio Simani and Cesare Fantuzzi. Fault diagnosis in power plant using neural networks. *Information Sciences*, 127(3-4):125–136, 2000.

[16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[17] Wenting Li, Deepjyoti Deka, Michael Chertkov, and Meng Wang. Real-time fault localization in power grids with convolutional neural networks. *arXiv preprint arXiv:1810.05247*, 2018.

[18] Weikang Wang, He Yin, Chang Chen, Abigail Till, Wenxuan Yao, Xianda Deng, and Yilu Liu. Frequency disturbance event detection based on synchrophasors and deep learning. *IEEE Transactions on Smart Grid*, 2020.

[19] James JQ Yu, Yunhe Hou, Albert YS Lam, and Victor OK Li. Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks. *IEEE TRANSACTIONS ON SMART GRID*, 10(2), 2019.

[20] Fahmida N Chowdhury and Jorge Luman Aravena. A modular methodology for fast fault detection and classification in power systems. *IEEE transactions on control systems technology*, 6(5):623–634, 1998.

[21] KM Silva, Benemar A Souza, and Nubia SD Brito. Fault detection and classification in transmission lines based on wavelet transform and ann. *IEEE Transactions on Power Delivery*, 21(4):2058–2063, 2006.

[22] T. Xia, H. Zhang, R. Gardner, J. Bank, J. Dong, J. Zuo, Y. Liu, L. Beard, P. Hirsch, G. Zhang, and R. Dong. Wide-area frequency based event location estimation. In *IEEE PESGM*, pages 1–7, 2007.

[23] Augusto Santiago Cerqueira, Danton Diego Ferreira, Moisés Vidal Ribeiro, and Carlos Augusto Duque. Power quality events recognition using a svm-based method. *Electric Power Systems Research*, 78(9):1546–1552, 2008.

[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[26] Zi Wang, Chengcheng Li, and Xiangyang Wang. Convolutional neural network pruning with structural redundancy reduction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14913–14922, 2021.

[27] Chengcheng Li, Zi Wang, and Hairong Qi. Fast-converging conditional generative adversarial networks for image synthesis. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2132–2136. IEEE, 2018.

[28] Chengcheng Li, Zi Wang, and Hairong Qi. Online knowledge distillation with history-aware teachers. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022.

[29] Chengcheng Li, Zi Wang, and Hairong Qi. Online knowledge distillation by temporal-spatial boosting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 197–206, 2022.

[30] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[31] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[33] Intrduction of pss@e. http://w3.siemens.com/smartgrid/global/en/products-systems-solutions/software-solutions/planning-data-management-software/planning-simulation/pages/pss-e.aspx.

[34] Intrduction of large scale testbed at curent, availabel online. http://curent.utk.edu/files/3713/6811/8508/Fact_sheet_tomsovic.pdf.

[35] Wenyun Ju, Junjian Qi, and Kai Sun. Simulation and analysis of cascading failures on an npcc power system test bed. In *Power & Energy Society General Meeting, 2015 IEEE*, pages 1–5. IEEE, 2015.

[36] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, and Craig Citro et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[37] Yang Song. Multi-event Analysis for Large-scale Power System through Cluster-based Sparse Coding, 2017. [Online; accessed October-2018].

[38] Ke Zhang, Yanzhu Ye, Lang Chen, Yingchen Zhang, R Matthew Gardner, and Yilu Liu. Fnet observations of low frequency oscillations in the eastern interconnection and their correlation with system events. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–8. IEEE, 2011.